0122229 现代数字通信与编码理论 October 1, 2010
XDU, Fall 2010 Lecture Notes

Chapter 3 Performance Bounds of Coded Communication Systems

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.

In this chapter we will study the bounding techniques used for performance analysis of coded communication systems, with emphasis on the understanding of Bhattacharyya bound and Gallager bound. In particular, Gallager bound provides a better upper bound on the ML decoding error probability for a specific code and ensemble of random/structured codes. (Gallager bound was introduced in [Gallager65] as an efficient tool to determine the error exponents of the ensemble of random codes, providing informative results up to the ultimate capacity limit.)

We will also show that it is possible to approach capacity on arbitrary DMCs with coding.

It should be pointed out that 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.

Note: The materials presented in Sections 3.3 through 3.7 are based largely on Massey's lecture notes.

3.1 Mathematical Preliminaries

3.1.1 Basic Inequalities

- Markov Inequality

The Markov inequality states that if a non-negative random variable Y has a mean $\mathsf{E}[Y]$, then the probability that the outcome exceeds any given number satisfies

$$P(Y \geq y) \leq \frac{\mathsf{E}[Y]}{y} \tag{3.1}$$

- Chebyshev Inequality

Let Z be an arbitrary random variable with finite mean $\mathsf{E}[Z]$ and finite variance σ_Z^2.

Define Y as the non-negative random variable $Y = (Z - \mathsf{E}[Z])^2$. The $\mathsf{E}[Y]$ = σ_Z^2. Applying (3.1),

$$P\left\{(Z - \mathsf{E}[Z])^2 \geq y\right\} \leq \frac{\sigma_Z^2}{y}$$

Replacing y with ε^2 (for any $\varepsilon>0$), this becomes the well-known Chebyshev inequality

$$P\{|Z-\mathsf{E}[Z]|\geq \varepsilon\} \leq \frac{\sigma_Z^2}{\varepsilon^2} \qquad (3.2)$$

- ■ Chernoff Bound (or Exponential bound)

Let $Y = \exp(sZ)$ for some arbitrary random variable Z that has a moment generating function $g_Z(s)=\mathsf{E}[\exp(sZ)]$ over some open interval of real values of s including s=0. Then, for s in that interval, (3.1) becomes

$$P\{\exp(sZ)\geq y\} \leq \frac{g_Z(s)}{y}$$

Letting $y = \exp(s\delta)$ for some constant δ, we have

$$\begin{aligned} P(Z\geq\delta) &\leq e^{-s\delta}\mathsf{E}[e^{sZ}], \quad \text{for } s\geq 0 \\ P(Z\leq\delta) &\leq e^{-s\delta}\mathsf{E}[e^{sZ}], \quad \text{for } s\leq 0 \end{aligned} \qquad (3.3)$$

The bounds can be optimized over s to get the strongest bound.

- ■ Jensen's Inequality

If f is a convex-∪ function and X is a random variable, then

$$\mathsf{E}[f(X)] \geq f(\mathsf{E}[X]) \qquad (3.4)$$

Moreover, if f is strictly convex, then equality in (3.4) implies that X=$\mathsf{E}[X]$ with probability 1, i.e., X is a constant.

- ■ The Union Bound

For sets A and B, we have

$$P(A\cup B) = P(A)+P(B)-P(A\cap B)$$

Since $P(A\cap B)\geq 0$,

$$P(A\cup B)\leq P(A)+P(B)$$

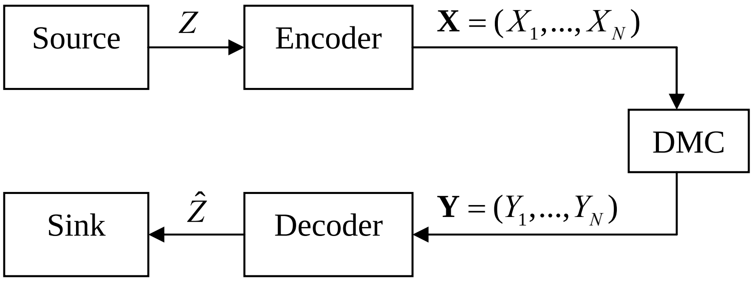
3.2 Block Coding Principles

In the following discussions, we will restrict our attention to block encoders and decoders.

Consider a discrete memoryless channel (DMC) with input alphabet $\mathcal{A}_X=\{a_1,a_2,...,a_K\}$ and output alphabet $\mathcal{A}_Y=\{b_1,b_2,...,b_J\}$. A block code of length N with M codewords for such a channel is a list $(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_M)$, each item of which is an N-tuple of elements from $\mathcal{A}_X$. See Fig. 3.1. We will denote the incoming message index by Z, whose possible values are integers from 1 through $M = 2^L$, where L is the length of input binary data sequence. We assume that when the information sequence corresponding to integer m enters the encoder, the codeword

$\mathbf{x}_m = (x_{m1}, x_{m2}, ..., x_{mN})$ is transmitted.

Let $\mathbf{y} = (y_1, y_2, ..., y_N)$ be the output sequence from the channel corresponding to a codeword input. If message m enters the encoder, $\mathbf{x}_m$ is transmitted, and on the basis of the received sequence **y**, the decoder produces an integer $\hat{m}$. A block error occurs if $\hat{m} \neq m$. We will denote the probability of block decoding error by $P_B(e) = \Pr(\hat{Z} \neq Z)$.


Source
Z
Encoder
$\mathbf{X} = (X_1, ..., X_N)$
DMC
Sink
$\hat{Z}$
Decoder
$\mathbf{Y} = (Y_1, ..., Y_N)$

Figuer 3.1

For a given DMC, given block length N and given code size M, there are $(K^N)^M = K^{MN}$ different block codes since there are K^N choices for each codeword in the list of M codewords. There are M^{J^N} different decoders since there are M choices of $\hat{Z}$ for each of the J^N values of Y. Each such decoder could be used with any such encoder so there are $K^{MN} M^{J^N}$ different coding systems. How do we find a good system from the bewildering array of choices?

■ *Encoding and Decoding Criterion*

The measure of goodness for a block code with M codewords of length N for a given channel is the smallness of the block error probability $P_B(e)$ when the codeword are equally likely and an ML decoder is used. Denote $\hat{Z} = f(Y)$, where f is the decoding function. The ML decoding rule may be described as follows.

The ML decoding rule: For each N-tuple **y** of channel output symbols, the decoder decodes **y** into message m, i.e., $f(\mathbf{y}) = m$, where m is (any one of) the index(es) that maximizes $P_N(\mathbf{y} \,|\, \mathbf{x}_m)$. It can be expressed simply as

$$\hat{m} = f(\mathbf{y}) = \arg\max_m P_N(\mathbf{y} \,|\, \mathbf{x}_m)$$

where $P_N(\mathbf{y} \,|\, \mathbf{x}_m)$ is the probability of receiving a sequence **y** given that the mth codeword is transmitted.

In the case of a DMC used without feedback, this becomes

$$\hat{m} = f(\mathbf{y}) = \arg\max_m \prod_{n=1}^{N} P(y_n \mid x_{m,n})$$

3.3 Codes with Two Codewords – the Bhattacharyya Bound

We will denote the set of received sequences decoded into message m as $\mathcal{D}_m$; i.e.,

$$\mathcal{D}_m = \left\{ \mathbf{y} \in \mathcal{A}_Y^N \mid f(\mathbf{y}) = m \right\}$$

which is called the decision region for message m. Since the output sequence $\mathbf{y}$ is decoded (/mapped) to exactly one message, $\mathcal{D}_m$'s form a collection of disjoint sets whose union is $\mathcal{A}_Y^N$; i.e.,

$$\begin{cases} \mathcal{D}_i \cap \mathcal{D}_j = \phi, & \text{for all } i \neq j \\ \bigcup_i \mathcal{D}_i = \mathcal{A}_Y^N \end{cases}$$

Thus, the probability of decoding error, when message m is sent, is defined as

$$\begin{aligned} P_B(e \mid m) &\equiv \Pr(\hat{Z} \neq Z \mid Z = m) = \Pr(\mathbf{y} \notin \mathcal{D}_m \mid \mathbf{x}_m) \\ &= 1 - \Pr(\mathbf{y} \in \mathcal{D}_m \mid \mathbf{x}_m) = \sum_{\mathbf{y} \notin \mathcal{D}_m} P_N(\mathbf{y} \mid \mathbf{x}_m) \end{aligned} \tag{3.5}$$

The overall probability of decoding error, if the message have a priori distribution Pr(m), is then given by

$$P_B(e) = \sum_{m=1}^{M} \Pr(m) P_B(e \mid m) \tag{3.6}$$

Equations (3.5) and (3.6) apply to any block code and any channel. In particular, the case is simple when $M = 2$. In this case, the error probability, when message 2 is transmitted, is

$$P_B(e \mid 2) = \sum_{\mathbf{y} \in \mathcal{D}_1} P_N(\mathbf{y} \mid \mathbf{x}_2) \tag{3.7}$$

We observe that,

$$\text{for } \mathbf{y} \in \mathcal{D}_1, \quad P_N(\mathbf{y} \mid \mathbf{x}_1) \geq P_N(\mathbf{y} \mid \mathbf{x}_2)$$

for ML decoding. It also implies that $P_N(\mathbf{y} \mid \mathbf{x}_1)^s \geq P_N(\mathbf{y} \mid \mathbf{x}_2)^s$, $0 < s < 1$, and hence that

$$P_N(\mathbf{y} \mid \mathbf{x}_2) \leq P_N(\mathbf{y} \mid \mathbf{x}_2)^{1-s} P_N(\mathbf{y} \mid \mathbf{x}_1)^s \tag{3.8}$$

Substituting (3.8) into (3.7) and letting s=1/2, we have

$$P_B(e \mid 2) \leq \sum_{\mathbf{y} \in \mathcal{D}_1} \sqrt{P_N(\mathbf{y} \mid \mathbf{x}_1) P_N(\mathbf{y} \mid \mathbf{x}_2)} \tag{3.9}$$

Similarly,

$$P_B(e \mid 1) \leq \sum_{\mathbf{y} \in \mathcal{D}_2} \sqrt{P_N(\mathbf{y} \mid \mathbf{x}_1) P_N(\mathbf{y} \mid \mathbf{x}_2)} \tag{3.10}$$

Combining (3.9) and (3.10), we obtain

$$P_B(e|1)+P_B(e|2)\le\sum_{\mathbf{y}}\sqrt{P_N(\mathbf{y}|\mathbf{x}_1)P_N(\mathbf{y}|\mathbf{x}_2)} \tag{3.11}$$

Notice that the summation is now over all **y** and hence is considerably easier to evaluate. Because $P_B(e|1)$ and $P_B(e|2)$ are non-negative, the R.H.S. of (3.11) is an upper bound on each of these conditional error probabilities. Thus, we have

$$P_B(e|m)\le\sum_{\mathbf{y}}\sqrt{P_N(\mathbf{y}|\mathbf{x}_1)P_N(\mathbf{y}|\mathbf{x}_2)}\,,\qquad m = 1,\ 2 \tag{3.12}$$

For the special case of a DMC, it simplifies to

$$P_B(e|m)\le\prod_{n=1}^{N}\sum_{y}\sqrt{P(y|x_{1n})P(y|x_{2n})}\,,\qquad m = 1,\ 2 \tag{3.13}$$

where we have written the dummy variable of summation as y rather than y_n. The bound in (3.12) and (3.13) are known as the *Bhattacharyya bound* on error probability.
Defining

$$d_B=-\log_2\left[\sum_{\mathbf{y}}\sqrt{P_N(\mathbf{y}|\mathbf{x}_1)P_N(\mathbf{y}|\mathbf{x}_2)}\right] \tag{3.14}$$

which we will call the *Bhattacharyya distance* between the two channel input sequences, we can rewrite (3.12) as

$$P_B(e|m)\le 2^{-d_B}\,,\quad m = 1, 2$$

In the special case of a binary-input DMC and the repetition code of length N, (3.13) becomes

$$P_B(e|m)\le\left(\sum_{y}\sqrt{P(y|0)P(y|1)}\right)^N,\qquad m = 1, 2$$

which can be written as

$$P_B(e|m)\le 2^{-ND_B}\,,\quad m = 1, 2$$

where

$$D_B=-\log_2\left[\sum_{y}\sqrt{P(y|0)P(y|1)}\right]$$

<u>Note:</u> The bound in (3.12) can also be interpreted as a Chernoff bound as follows.

$$P_B(e|1)=\Pr\left(\log\frac{P_N(\mathbf{y}|\mathbf{x}_2)}{P_N(\mathbf{y}|\mathbf{x}_1)}\ge 0\,\middle|\,\mathbf{x}_1\text{ sent}\right)$$

$$\le e^{-s\cdot 0}E\left\{\exp\left(s\cdot\log\frac{P_N(\mathbf{y}|\mathbf{x}_2)}{P_N(\mathbf{y}|\mathbf{x}_1)}\right)\right\}=E_{Y|X_1}\left[\frac{P_N(\mathbf{y}|\mathbf{x}_2)}{P_N(\mathbf{y}|\mathbf{x}_1)}\right]^s$$

Hence, $P_B(e|1)\le\sum_{\mathbf{y}}P_N(\mathbf{y}|\mathbf{x}_1)\left[\frac{P_N(\mathbf{y}|\mathbf{x}_2)}{P_N(\mathbf{y}|\mathbf{x}_1)}\right]^s=\sum_{\mathbf{y}}P_N(\mathbf{y}|\mathbf{x}_1)^{1-s}P_N(\mathbf{y}|\mathbf{x}_2)^s$.

3.4 Codes with Many Codewords – the Union Bhattacharyya Bound and Gallager Bound

3.4.1 Union Bhattacharyya Bound

We now consider generalizing the bound in (3.12) to a code $\mathcal{C}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_M\}$ with many codewords. Since the complement of $\mathcal{D}_m$ is given by $\bigcup_{m'\neq m}\mathcal{D}_{m'}$, and noting that $\mathcal{D}_i, i=1,2,\ldots,M$ are disjoint decision regions, we have ((3.5)也可直接写为(3.15))

$$\begin{aligned}P_B(e\,|\,m)&=\Pr\left(\mathbf{y}\in\bigcup_{m'\neq m}\mathcal{D}_{m'}\middle|\mathbf{x}_m\ sent\right)\\&=\sum_{m'\neq m}\Pr\left(\mathbf{y}\in\mathcal{D}_{m'}\,|\,\mathbf{x}_m\ sent\right)\\&=\sum_{\substack{m'=1\\m'\neq m}}^{M}\sum_{\mathbf{y}\in\mathcal{D}_{m'}}P_N(\mathbf{y}\,|\,\mathbf{x}_m)\end{aligned}\qquad(3.15)$$

Note that, for ML decoding,

$$\mathbf{y}\in\mathcal{D}_{m'}\Rightarrow P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})\geq P_N(\mathbf{y}\,|\,\mathbf{x}_m)$$

Since every $\mathbf{y}\in\mathcal{D}_{m'}$ can also be put into the decoding region D_2' of an ML decoder for the code $(\mathbf{x}_1',\mathbf{x}_2')=(\mathbf{x}_m,\mathbf{x}_{m'})$ with only two codewords, we see that

$$\sum_{\mathbf{y}\in\mathcal{D}_{m'}}P_N(\mathbf{y}\,|\,\mathbf{x}_m)\leq\sum_{\mathbf{y}\in D_2'}P_N(\mathbf{y}\,|\,\mathbf{x}_m)=\sum_{\mathbf{y}\in D_2'}P_N(\mathbf{y}\,|\,\mathbf{x}_1')\equiv P_B'(e\,|\,1)\qquad(3.16)$$

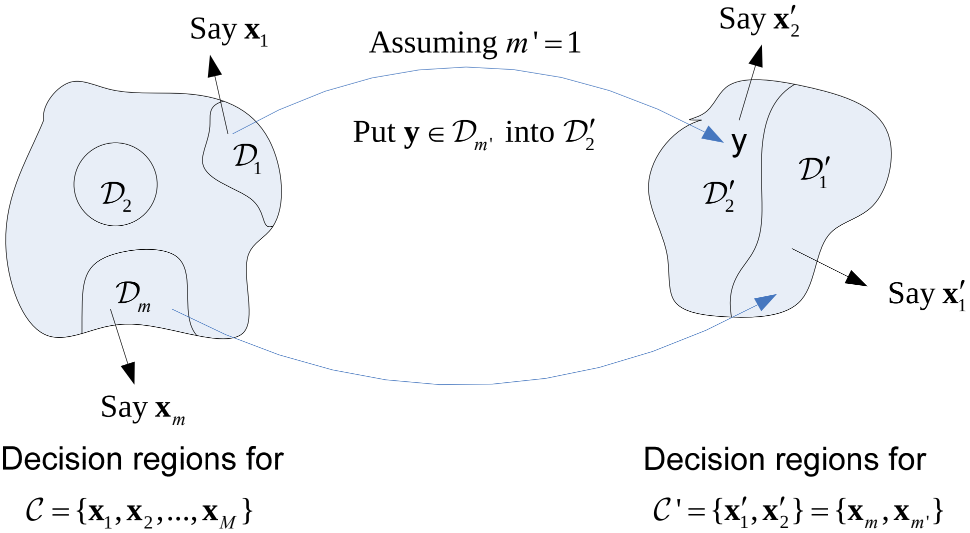


Figure 3.2 Illustration of observation space and decision regions

Invoking the Bhattacharyya bound (cf. (3.10)) in (3.16), we have

$$\sum_{\mathbf{y}\in\mathcal{D}_{m'}} P_N(\mathbf{y}\,|\,\mathbf{x}_m) \le \sum_{\mathbf{y}\in D_2'} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_m)P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}$$

$$\le \sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_m)P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})} \qquad (3.17)$$

Substituting (3.17) into (3.15), we obtain the so-called *union Bhattacharyya bound*:

$$P_B(e\,|\,m) \le \sum_{\substack{m'=1\\m'\neq m}}^{M} \sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_m)P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}\,, \qquad m=1,2,\ldots,M \qquad (3.18)$$

In particular, for memoryless channels,

$$P_B(e\,|\,m) \le \sum_{m'=1}^{M} \prod_{n=1}^{N} \sum_{y} \sqrt{P(y\,|\,x_{mn})P(y\,|\,x_{m'n})}\,, \qquad m=1,2,\ldots,M \qquad (3.19)$$

3.4.2 Gallager Bound

For an ML decoder, we know that

$$\mathbf{y} \notin \mathcal{D}_m \Rightarrow P_N(\mathbf{y}\,|\,\mathbf{x}_{m'}) \ge P_N(\mathbf{y}\,|\,\mathbf{x}_m) \quad \text{for some } m' \neq m.$$

Then we have

$$\sum_{\substack{m'=1\\m'\neq m}}^{M} \left[\frac{P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}{P_N(\mathbf{y}\,|\,\mathbf{x}_m)}\right]^s \ge 1, \quad \text{for any } s>0$$

since at least one term in the summation will itself be at least one and then the sum is at least one. We can raise both sides of the above equation to the power of some non-negative parameter $\rho \ge 0$ and preserve the inequality, i.e.,

$$\mathbf{y} \notin \mathcal{D}_m \Rightarrow \left\{\sum_{\substack{m'=1\\m'\neq m}}^{M} \left[\frac{P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}{P_N(\mathbf{y}\,|\,\mathbf{x}_m)}\right]^s\right\}^{\rho} \ge 1, \quad \text{for any } s>0 \text{ and any } \rho\ge 0 \qquad (3.20)$$

Note that (3.20) can be written as

$$P_N(\mathbf{y}\,|\,\mathbf{x}_m)^{-s\rho}\left[\sum_{\substack{m'=1\\m'\neq m}}^{M} P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})^s\right]^{\rho} \ge 1, \quad \text{any } s>0 \text{ and any } \rho\ge 0 \qquad (3.21)$$

By multiplying each of terms in the summation in (3.5) by the L.H.S. of (3.21), we can see that the error probability is upper-bounded by

$$P_B(e\,|\,m) \le \sum_{\mathbf{y}\notin\mathcal{D}_m} P_N(\mathbf{y}\,|\,\mathbf{x}_m)^{1-s\rho}\left[\sum_{\substack{m'=1\\m'\neq m}}^{M} P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})^s\right]^{\rho} \qquad (3.22)$$

Letting $s = 1/(1+\rho)$ and extending the summation in (3.22) to all **y**, we have the so-called Gallager bound.

- *Gallager Bound*: When the code $\mathcal{C} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ of length N is decoded by an ML decoder, then

$$P_B(e\,|\,m) \le \sum_{\mathbf{y}} P_N(\mathbf{y}\,|\,\mathbf{x}_m)^{\frac{1}{1+\rho}} \left[\sum_{\substack{m'=1\\ m'\ne m}}^{M} P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})^{\frac{1}{1+\rho}} \right]^{\rho}, \qquad m=1,\ldots,M \text{ and any } \rho\ge 0 \tag{3.23}$$

If the channel is memoryless, it becomes

$$P_B(e\,|\,m) \le \prod_{n=1}^{N} \sum_{y} P(y\,|\,x_{mn})^{\frac{1}{1+\rho}} \left[\sum_{\substack{m'=1\\ m'\ne m}}^{M} P(y\,|\,x_{m'n})^{\frac{1}{1+\rho}} \right]^{\rho} \tag{3.24}$$

We can see that, when optimized by choice of ρ, the Gallager bound is strictly tighter than the Bhattacharyya bound except when $\rho = 1$. (It is clear that the union bound (3.18) is a special case of Gallager bound obtained by setting $\rho = 1$.) Notice that unless the code and channel possess a high degree of simplifying symmetry, both bounds are too complicated to calculation in most practical cases.

The Gallager bound is sometimes expressed as

$$P_B(e\,|\,m) \le \sum_{\mathbf{y}} P_N(\mathbf{y}\,|\,\mathbf{x}_m) \left[\sum_{\substack{m'=1\\ m'\ne m}}^{M} \left(\frac{P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}{P_N(\mathbf{y}\,|\,\mathbf{x}_m)} \right)^{s} \right]^{\rho}, \qquad s>0,\ \rho\ge 0 \tag{3.25}$$

3.5 Ensemble Average Performance of Codes with Two Codewords

In this section, we consider random coding and evaluate the average $P_B(e|m)$ for codes with two codewords. Suppose that, for a given channel and given N, one has calculated $P_B(e|m)$ with ML decoding for every length-N block code with two codewords. Note that the error probabilities $P_B(e|m)$ are now (random) variables, dependent on the used specific code. To make the dependence on the code explicit, we write $P_{e|m}(\mathbf{x}_1, \mathbf{x}_2)$ to denote $P_B(e|m)$ for some ML decoder for the particular code $\mathcal{C} = \{\mathbf{x}_1, \mathbf{x}_2\}$.

Let $Q_N(\mathbf{x})$ be an arbitrary probability assignment on the set of channel input sequences of length N. Now consider the ensemble of codes where the codewords are selected independently using the probability assignment $Q_N(\mathbf{x})$. The expected value of $P_B(e|m)$ over the ensemble is then given by

$$\overline{P_B(e\,|\,m)} = \sum_{\mathbf{x}_1} \sum_{\mathbf{x}_2} P_{e|m}(\mathbf{x}_1, \mathbf{x}_2) Q_N(\mathbf{x}_1) Q_N(\mathbf{x}_2) \qquad m = 1, 2$$

By symmetry, $\overline{P_B(e\,|\,1)} = \overline{P_B(e\,|\,2)}$. From (3.12), $\overline{P_B(e\,|\,m)}$ is upper-bounded by

$$\overline{P_B(e\,|\,m)} \le \sum_{\mathbf{x}_1} \sum_{\mathbf{x}_2} \sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_1) P_N(\mathbf{y}\,|\,\mathbf{x}_2)}\, Q_N(\mathbf{x}_1) Q_N(\mathbf{x}_2)$$

$$= \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}_1} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_1)}\, Q_N(\mathbf{x}_1) \right] \left[\sum_{\mathbf{x}_2} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_2)}\, Q_N(\mathbf{x}_2) \right], \quad m=1,\ 2 \tag{3.26}$$

Note that $\mathbf{x}_1$ and $\mathbf{x}_2$ in (3.26) are simply dummy variable of summation, (3.26) may reduced to

$$\overline{P_B(e\,|\,m)} \le \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x})} \right]^2, \quad m{=}1,\ 2 \tag{3.27}$$

To specialize (3.27) to a memoryless channel, we choose

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n) \tag{3.28}$$

That is, we are considering an ensemble in which each letter of each codeword is selected independently with the probability mass function $Q(a_k)$. Then for a DMC, (3.27) becomes

$$\overline{P_B(e\,|\,m)} \le \sum_{y_1} \ldots \sum_{y_N} \left[\prod_{n=1}^{N} \sum_{x_n} Q(x_n) \sqrt{P(y_n\,|\,x_n)} \right]^2$$

$$= \prod_{n=1}^{N} \sum_{y_n} \left[\sum_{x_n} Q(x_n) \sqrt{P(y_n\,|\,x_n)} \right]^2 \tag{3.29}$$

Recognizing that x_n and y_n are now dummy variables of summation, we can rewrite (3.29) as

$$\overline{P_B(e\,|\,m)} \le \left\{ \sum_{y} \left[\sum_{x} Q(x) \sqrt{P(y\,|\,x)} \right]^2 \right\}^N, \quad m = 1,\ 2 \tag{3.30}$$

This is an upper bound on the average error probability over an ensemble of codes with two codewords of length N.

To emphasize the exponential dependence on the code length N, we rewrite (3.30) as

$$\overline{P_B(e\,|\,m)} \le 2^{-N\left\{ -\log_2 \sum_{y} \left[\sum_{x} Q(x) \sqrt{P(y|x)} \right]^2 \right\}}$$

We are free to choose $Q(x)$ so that we obtain the tightest upper bound

$$\overline{P_B(e\,|\,m)} \le 2^{-NR_0}$$

where

$$R_0 = \max_{Q} \left\{ -\log_2 \sum_{y} \left[\sum_{x} Q(x) \sqrt{P(y\,|\,x)} \right]^2 \right\} \tag{3.31}$$

Due to monotonicity of the log function, an equivalent expression for R_0 is

$$R_0 = -\log_2 \left\{ \min_{Q} \sum_{y} \left[\sum_{x} Q(x) \sqrt{P(y\,|\,x)} \right]^2 \right\} \tag{3.32}$$

The quantity that we have denoted R_0 is usually called the *cut-off rate* of the DMC.

For symmetric channels, an equiprobable distribution on the channel input alphabet, $Q(x){=}1/K$, $\forall x \in \mathcal{A}_X$, achieves the extremum. In this case,

$$R_0 = \log_2 K - \log_2 \left\{ \frac{1}{K} \sum_{y} \left[\sum_{x} \sqrt{P(y\,|\,x)} \right]^2 \right\}$$

In particular, for the BSC,

$$R_0 = -\log_2 \sum_y \left[\frac{1}{2} \sum_x \sqrt{P(y \mid x)} \right]^2$$

$$= 1 - \log_2 \left[1 + \sum_y \sqrt{P(y \mid 0) P(y \mid 1)} \right] = 1 - \log_2 \left[1 + 2\sqrt{p(1-p)} \right] \qquad (3.33)$$

3.6 Ensemble Average Performance of Codes with Many Codewords

We now compute the ensemble average (with respect to code selection, keeping *m* fixed) performance for codes with many codewords. Taking expectation in (3.18), we obtain

$$\overline{P_B(e \mid m)} \le \sum_{\substack{m'=1 \\ m' \ne m}}^{M} \mathsf{E} \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y} \mid \mathbf{X}_m) P_N(\mathbf{y} \mid \mathbf{X}_{m'})} \right], \quad m=1,2,\ldots,M \qquad (3.34)$$

If the probability assignments to the encoders satisfy the condition of pairwise independence of codewords, i.e.,

$$Q_N(\mathbf{x}_m, \mathbf{x}_{m'}) = Q_N(\mathbf{x}_m) Q_N(\mathbf{x}_{m'}), \quad \text{all } m' \ne m \qquad (3.35)$$

for all choices of $\mathbf{x}_m$ and $\mathbf{x}_{m'}$, then it follows that

$$\overline{P_B(e \mid m)} \le (M-1) \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) \sqrt{P_N(\mathbf{y} \mid \mathbf{x})} \right]^2, \quad m=1,2,\ldots,M \qquad (3.36)$$

One simple way to get the pairwise independence in (3.35) is to make all the codeword independent; i.e., to assign probability

$$Q_N(\mathbf{x}_1, \ldots, \mathbf{x}_M) = \prod_{m=1}^{M} Q_N(\mathbf{x}_m) \qquad (3.37)$$

to the encoder for $\mathcal{C} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$.

For a DMC, using (3.37) we have

$$\overline{P_B(e \mid m)} \le (M-1) 2^{-NR_0}$$

Since $M - 1 < 2^{NR} \le M$, we have that the ensemble average error probability is upper-bounded by

$$\overline{P_B(e \mid m)} \le 2^{NR} \cdot 2^{-NR_0} = 2^{-N(R_0 - R)} \qquad (3.38)$$

By introducing the *Bhattacharyya exponent* $E_B(R) = R_0 - R$, (3.38) can be written as

$$\overline{P_B(e \mid m)} \le 2^{-NE_B(R)} \qquad (3.39)$$

This is the *random coding union bound* for block codes. The exponent $E_B(R)$ is shown in Fig. 3.3.

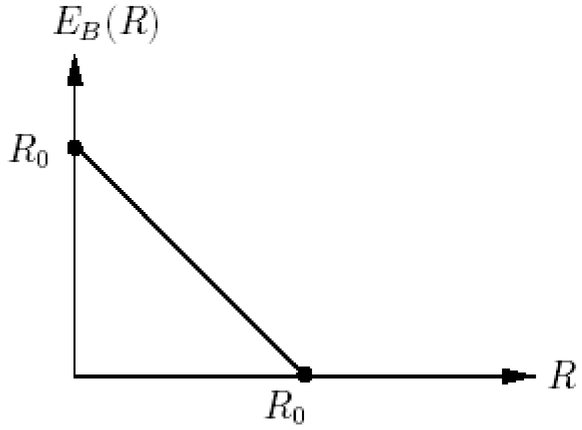


Figure 3.3 The Bhattacharyya exponent of the random coding union bound.

Since R_0 has the same dimension as the code rate, R_0 is often called the *cutoff rate*.

3.7 Gallager's Version of the Coding Theorem for a DMC

We now consider averaging the Gallager bound in (3.23) over an ensemble of codes with many codewords. Rather than bounding $E[P_B(e|m)]$ directly over the ensemble of codes, we begin with bounding the conditional expectation of $P_B(e|m)$ given that the mth codeword is a particular vector **x'**. From (3.23), we have

$$\mathsf{E}\left[P_B(e|m)\,|\,\mathbf{X}_m=\mathbf{x}'\right]\le\sum_{\mathbf{y}}P_N(\mathbf{y}\,|\,\mathbf{x}')^{\frac{1}{1+\rho}}\mathsf{E}\left\{\left[\sum_{\substack{m'=1\\m'\ne m}}^{M}P_N(\mathbf{y}\,|\,\mathbf{X}_{m'})^{\frac{1}{1+\rho}}\right]^{\rho}\,\middle|\,\mathbf{X}_m=\mathbf{x}'\right\},\text{ for all }\rho\ge 0\quad(3.40)$$

Note that $f(\alpha)=\alpha^{\rho}$ is convex-∩ in the interval $\alpha > 0$ when $0\le\rho\le 1$ (since $f''(\alpha)=\rho(\rho-1)\alpha^{\rho-2}\le 0$). By invoking Jensen's inequality, we obtain

$$\mathsf{E}\left[P_B(e|m)\,|\,\mathbf{X}_m=\mathbf{x}'\right]\le\sum_{\mathbf{y}}P_N(\mathbf{y}\,|\,\mathbf{x}')^{\frac{1}{1+\rho}}\left\{\sum_{\substack{m'=1\\m'\ne m}}^{M}\mathsf{E}\left[P_N(\mathbf{y}\,|\,\mathbf{X}_{m'})^{\frac{1}{1+\rho}}\,\middle|\,\mathbf{X}_m=\mathbf{x}'\right]\right\}^{\rho}\qquad(3.41)$$

We choose the same ensemble of codes as we did in the previous section, and assume that the probability assignments satisfy (3.35). Thus,

$$\mathsf{E}\left[P_N(\mathbf{y}\,|\,\mathbf{X}_{m'})^{\frac{1}{1+\rho}}\,\middle|\,\mathbf{X}_m=\mathbf{x}'\right]=\mathsf{E}\left[P_N(\mathbf{y}\,|\,\mathbf{X}_{m'})^{\frac{1}{1+\rho}}\right]$$

$$=\sum_{\mathbf{x}}Q_N(\mathbf{x})P_N(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}}\qquad(3.42)$$

Substituting (3.42) into (3.41) yields

$$\mathsf{E}\left[P_B(e|m)\,|\,\mathbf{X}_m=\mathbf{x}'\right]\le\sum_{\mathbf{y}}P_N(\mathbf{y}\,|\,\mathbf{x}')^{\frac{1}{1+\rho}}(M-1)^{\rho}\left\{\sum_{\mathbf{x}}Q_N(\mathbf{x})P_N(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}}\right\}^{\rho}\qquad(3.43)$$

Therefore,

$$\overline{P_B(e\,|\,m)} = \sum_{\mathbf{x}'} Q_N(\mathbf{x}')\mathrm{E}\left[P_B(e\,|\,m)\,|\,\mathbf{X}_m=\mathbf{x}'\right]$$

$$\leq \sum_{\mathbf{y}}\sum_{\mathbf{x}'} Q_N(\mathbf{x}')P_N(\mathbf{y}\,|\,\mathbf{x}')^{\frac{1}{1+\rho}}(M-1)^{\rho}\left\{\sum_{\mathbf{x}} Q_N(\mathbf{x})P_N(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}}\right\}^{\rho}$$

$$=(M-1)^{\rho}\sum_{\mathbf{y}}\left\{\sum_{\mathbf{x}} Q_N(\mathbf{x})P_N(\mathbf{y}\,|\,\mathbf{x})^{\frac{1}{1+\rho}}\right\}^{1+\rho}, \quad 0\leq\rho\leq 1 \tag{3.44}$$

Inequality (3.44) is *Gallager's celebrated random coding bound* and applies for ML decoding for any block code on any discrete channel. For the special case of a DMC and the choice (3.28) for $Q_N(\mathbf{x})$, (3.44) reduces to

$$\overline{P_B(e\,|\,m)} \leq (M-1)^{\rho}\left\{\sum_{y}\left[\sum_{x} Q(x)P(y\,|\,x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}^{N}, \quad 0\leq\rho\leq 1 \tag{3.45}$$

Considering $M-1<2^{NR}\leq M$, we have

$$\overline{P_B(e\,|\,m)} \leq 2^{NR\rho}\left\{\sum_{y}\left[\sum_{x} Q(x)P(y\,|\,x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}^{N}$$

$$=2^{-N[E_0(\rho,Q)-\rho R]}, \quad 0\leq\rho\leq 1,\ m=1,\ldots,M \tag{3.46}$$

where

$$E_0(\rho,Q) = -\log_2\sum_{y}\left[\sum_{x} Q(x)P(y\,|\,x)^{1/(1+\rho)}\right]^{1+\rho} \tag{3.47}$$

Since (3.46) is valid for each message index m in the code, we see that the average error probability over the messages satisfyies

$$\overline{P_B(e)} = \sum_{m=1}^{M}\Pr(m)\overline{P_B(e\,|\,m)} \leq 2^{-N[E_0(\rho,Q)-\rho R]} \tag{3.48}$$

Since ρ and Q are arbitrary in (3.47), we get the tightest bound by choosing ρ and Q to maximizing $[E_0(\rho, Q)$-ρR]. Thus, we define the *random coding exponent* as

$$E_G(R) = \max_{0\leq\rho\leq 1}\max_{Q}[E_0(\rho,Q)-\rho R] \tag{3.49}$$

which will be also called the *Gallager exponent*. With this definition, we have

$$\overline{P_B(e)} \leq 2^{-NE_G(R)} \tag{3.50}$$

3.7.1 Properties of $E_G(R)$

- It is seen that $E_G(R)\geq E_B(R)=R_0-R$ with equality iff ρ = 1 is maximizing in (3.49). This follows from the fact that $\max_{Q} E_0(1,Q)=R_0$.

- We rewrite (3.49) as $E_G(R) = \max_Q E_G(R,Q)$, where

$$E_G(R,Q) = \max_{0\le\rho\le1}[E_0(\rho,Q) - \rho R] \tag{3.51}$$

We can see that $E_G(R, Q)$ has the following properties. See Fig. 3.4.

a) $E_G(0, Q) = E_0(1, Q)$

b) $E_G(R, Q)$ is linear with slope –1 in the interval $0 \le R \le R_c(Q)$, where

$$R_c(Q) = \left.\frac{\partial E_0(\rho,Q)}{\partial\rho}\right|_{\rho=1}$$

c) $E_G(R, Q)$ is convex-∪ and positive in the interval $0 \le R \le I_Q(X; Y)$, where

$$I_Q(X;Y) = I(X;Y)|_{P_X=Q} = \sum_k\sum_j Q(k)P(j\,|\,k)\log\frac{P(j\,|\,k)}{\sum_i Q(i)P(j\,|\,i)}$$

- It follows immediately from $E_G(R) = \max_Q E_G(R,Q)$ that $E_G(R)$ is just the upper envelope of the curves $E_G(R, Q)$ taken over all Q. It has the following properties. See Fig. 3.5.

a) $E_G(0) = \max_{0\le\rho\le1} E_0(1,Q) = R_0$.

b) $E_G(R)$ is linear with slope –1 in the interval $0 \le R \le R_c$, where $R_c = \max_{Q:R_0-achieving} R_c(Q)$

c) $E_G(R)$ is convex-∪ and positive in the interval $0 \le R \le C$, where C is the capacity of the DMC.

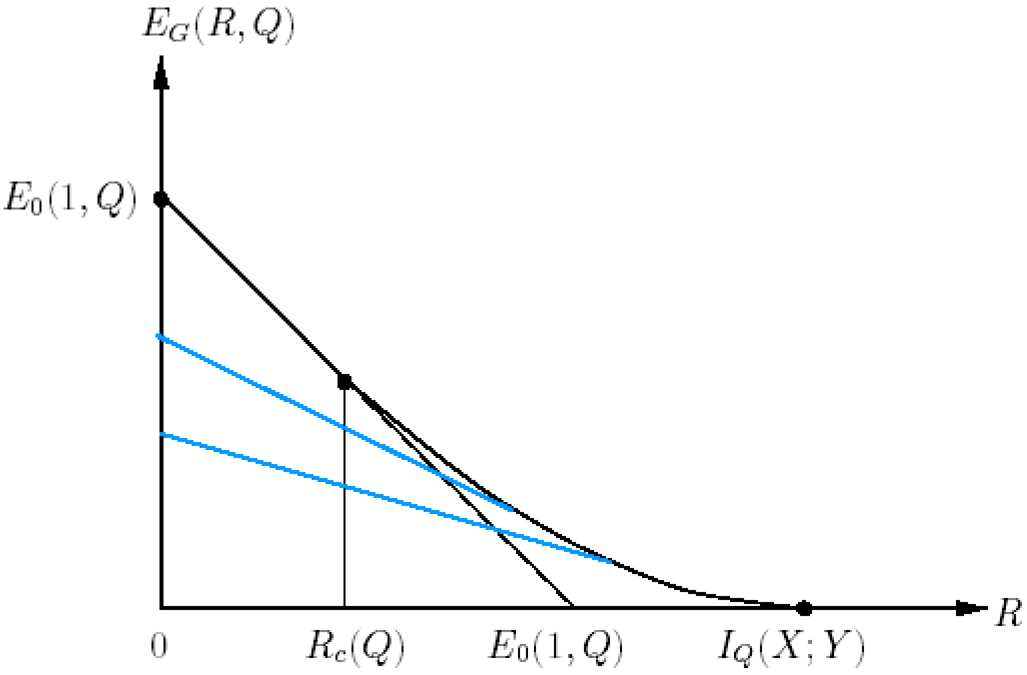


Figure 3.4 The general form of the exponent $E_G(R, Q)$

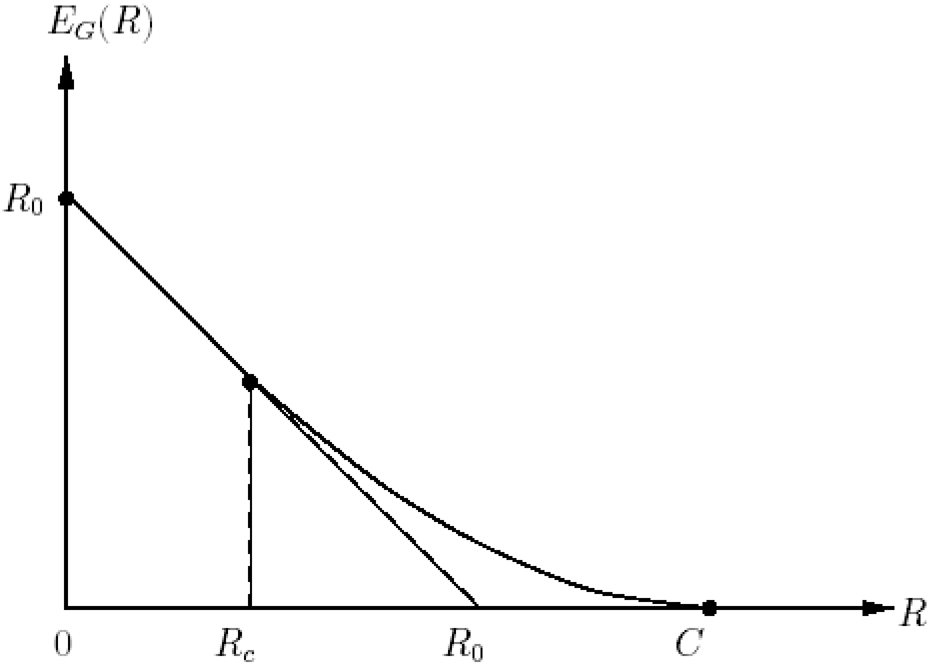
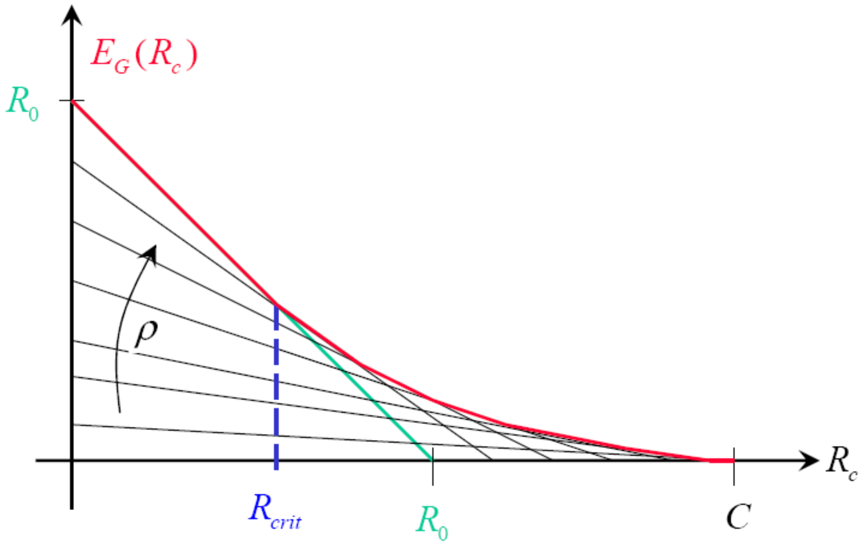


Figure 3.5 The general form of the Gallager exponent $E_G(R)$



Each pair of Q and ρ corresponds to a line with slope -ρ. Up to R_{crit} ρ=1 achieves maximum.

3.7.2 Gallager's *Coding Theorem for a DMC*

Since at least one code in the ensemble has error probability as good as ensemble average (3.50), we then have the fundamental coding theorem for DMCs.

- *Noisy-Channel Coding Theorem*: *Given a DMC with capacity C, there exist block codes with $M=2^{NR}$ codewords of length N, each with fixed rate R < C, for which the block error probability of a ML decoder diminishes to zero exponentially fast in N.*

Similar results are obtainable for channels with discrete input alphabets and continuous output alphabets.

3.7.3 A Summary of Performance Bounds

- For a specific code with two codewords:
  - ***Bhattacharyya bound*** (3.12)
- For a specific code with many codewords:
  - ***Union Bhattacharyya bound*** (3.18)
  - ***Gallager bound*** (3.23)
- For ensembles of random codes with two codewords:
  - ***Cutoff rate*** (3.31)
- For ensembles of random codes with many codewords:
  - ***Random coding union bound:*** *Bhattacharyya exponent* (3.39)
  - ***Gallager's random coding bound*** (3.46): *Gallager exponent* (3.48)

3.8 Random Coding Bound for Trellis Codes

The R_0 and capacity theory is pertinent to the ensemble of trellis codes as well. A detailed discussion can be found in Viterbi and Omura [2]. We briefly present some results below.

Viterbi showed that for the ensemble of (n_0, k_0) convolutional codes of rate R over a DMC, the average bit error probability satisfies

$$\overline{P_b(e)} < c_v(\varepsilon)2^{-N_t E_t(R,\varepsilon)} \tag{3.52}$$

where $N_t = (T+1)n_0$ (T is the memory order of the encoder) is the constraint length, and $c_v(\varepsilon)$ is a constant given by

$$c_v(\varepsilon) = \frac{2^{k_0}}{(1-2^{\varepsilon n_0})^2}$$

Here, ε is an arbitrary positive number. $E_t(R, \varepsilon)$ is a random coding exponent for trellis codes, and given by

$$E_t(R,\varepsilon) = \begin{cases} R_0, & \text{for } R < R_0 - \varepsilon \\ \max\limits_Q E_0(\rho^*,Q), & R_0 - \varepsilon < R < C - \varepsilon \end{cases} \tag{3.53}$$

where ρ^* (which depends Q) is the solution to

$$\frac{E_0(\rho^*,Q)-\varepsilon}{\rho^*} = R$$

and C is the capacity of the DMC.

By choosing ε as small as we like, we see that the best exponent that can be achieved at rate R is

$$E_V(R) = \lim_{\varepsilon \to 0} E_t(R,\varepsilon) \tag{3.54}$$

which will be referred to as *Viterbi exponent*. In other word,

$$E_V(R) = \begin{cases} R_0, & \text{for } R < R_0 \\ \max_Q E_0(\rho^*, Q), & \text{for } R_0 < R < C \end{cases} \tag{3.55}$$

where ρ^* is the solution to $E_0(\rho^*, Q)/\rho^* = R$.

The general form of $E_V(R)$ is sketched in Fig. 3.6. We can see that

$$E_V(R) > E_G(R), \quad \text{for } 0 \le R < C$$

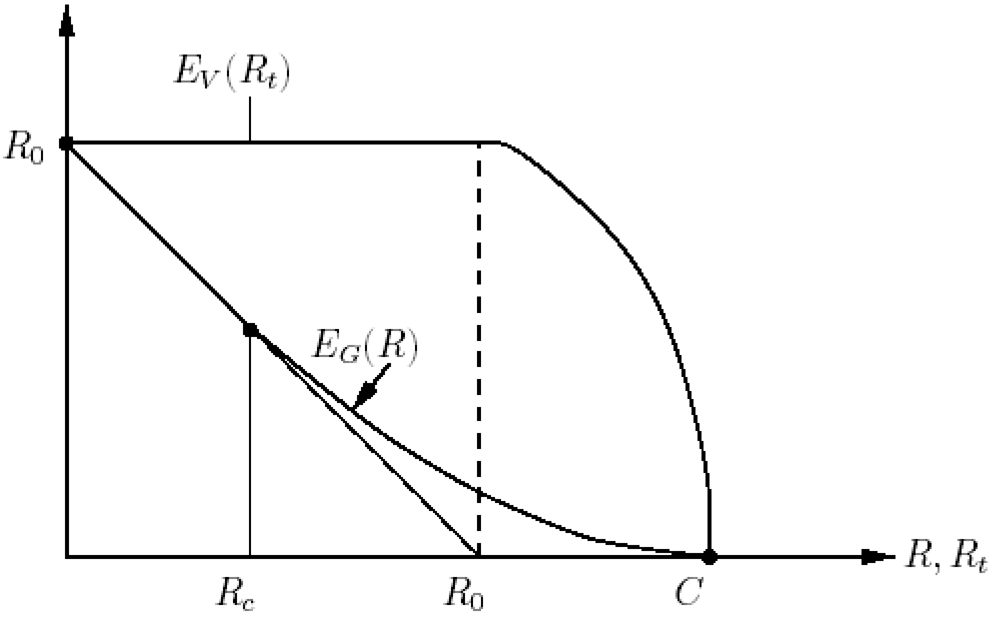


Figure 3.6 The general form of $E_V(R)$ and $E_G(R)$.

Remark:

- Error exponents provide us with a tool for comparing codes.
- It yields a technique for analyzing random ensembles of codes.
- Of course, in comparing coding techniques, other issues such as decoding complexity and decoding delay must be taken into account.
- There is a potentially large coefficient $c_V(\varepsilon)$ in the bound for trellis codes.

3.9 Application of Gallager Bound

We have derived the upper bound on the ML decoding error probability for a particular code and the ensemble of random codes. Furthermore, the Gallager bounds can also be applied to a specific structured code and ensembles of structured codes. It is suitable for both block and bit error probability analysis. The guideline is as follows. The detailed discussion will be provided in Chapter 4.

By partitioning a binary linear block code into constant Hamming-weight subcodes, a union bound over the subcodes yields

$$P_B(e\,|\,0) \le \sum_{d=d_{\min}}^{N} P_{e|0}(d)$$

where $d_{\min}$ is the minimum Hamming distance of the block code of length N, and $P_{e|0}(d)$ is the conditional decoding error probability with respect to the subcode with a constant Hamming weight *d*.

In fact, many existing upper bounds can be regarded as variations of the Gallager bound. The interconnections between them are depicted in Fig. 3.7. See [R1] for details.

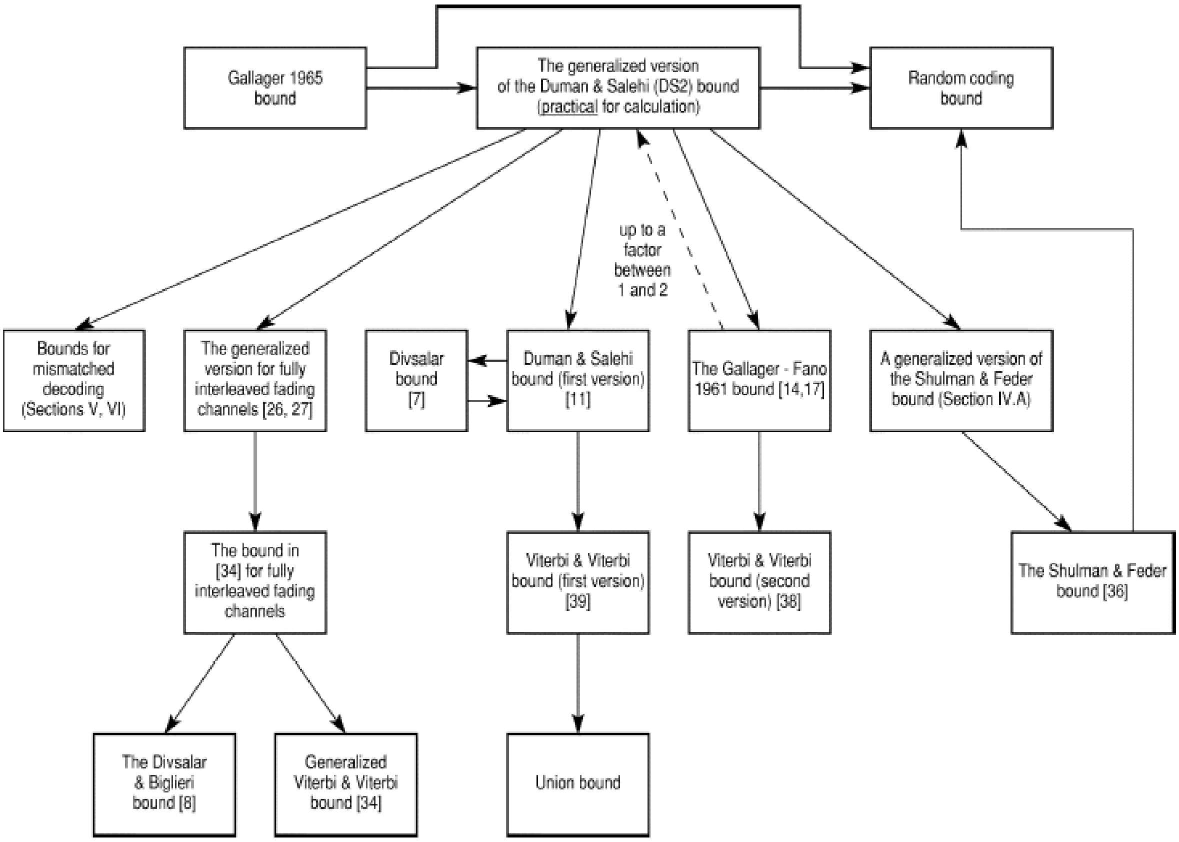

Gallager 1965 bound
The generalized version of the Duman & Salehi (DS2) bound (practical for calculation)
Random coding bound
up to a factor between 1 and 2
Bounds for mismatched decoding (Sections V, VI)
The generalized version for fully interleaved fading channels [26, 27]
Divsalar bound [7]
Duman & Salehi bound (first version) [11]
The Gallager - Fano 1961 bound [14,17]
A generalized version of the Shulman & Feder bound (Section IV.A)
The bound in [34] for fully interleaved fading channels
Viterbi & Viterbi bound (first version) [39]
Viterbi & Viterbi bound (second version) [38]
The Shulman & Feder bound [36]
The Divsalar & Biglieri bound [8]
Generalized Viterbi & Viterbi bound [34]
Union bound

Figure 3.7

C. Schlegel provides in [6, chapter 6] a brief discussion on the error exponent of an 8-PSK constellation over the AWGN channel. Refer to Fig. 3.8, where the uniform input symbol distribution is assumed, and $E_0(\rho)$ = $E_0(\rho$, uniform distribution) and $R_0 = R_0$(uniform distribution). For the AWGN channel,

$$E_0(\rho,\mathbf{Q}) \triangleq -\log_2 \int_y \left(\sum_x Q(x) p(y \mid x)^{1/(1+\rho)} \right)^{1+\rho} dy$$

where $\mathbf{Q} = (q_1, ..., q_M)$ is the probability with which *x* is chosen from a signal set $\mathcal{A}_X$ of size M=|$\mathcal{A}$|, and $p(y|x)$ is the conditional probability of the channel output signal.

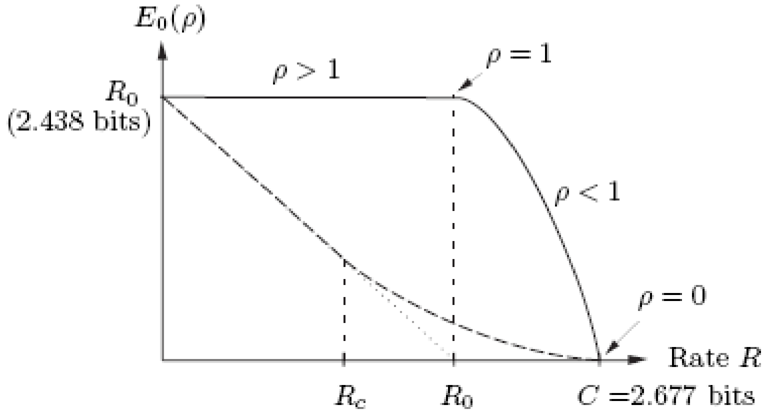


Figure 3.8: Error exponent as a function of the rate R for an 8-PSK constellation on an AWGN-channel at a signal-to-noise ratio of E_s/N_0 = 10dB.

- *Code Complexity*

The last, and somewhat hidden player in the application of coding is complexity. First there is what we might term *code complexity*. In order to approach the Shannon bound, larger and larger codes are required. In fact, Shannon et. al. [] proved the following *lower bound* on the codeword error probability P_B:

$$P_B > 2^{-N\left(E_{sp}(R)+o(N)\right)}; \quad E_{sp}(R) = \max_{\mathbf{Q}} \max_{\rho>1} \left[E_0(\rho,\mathbf{Q}) - \rho R\right]$$

The bound is plotted for rate R = 1⁄2 in Figure 3.9 for BPSK modulation [6], together with selected Turbo coding schemes and classic concatenated methods. The performance of various length codes follows in parallel to the bound, and we see that there is codesize of limiting return at $N \approx 10^4$--10^5, beyond which only small gains are possible. This is the reason why most practical applications of large codes target block sizes no larger than this. On the other hand, codes cannot be shortened much below N = 104 without a measurable loss in performance. Implementations of near-capacity error control systems therefore have to process blocks of 10,000 symbols or more, requiring appropriate storage and memory management.

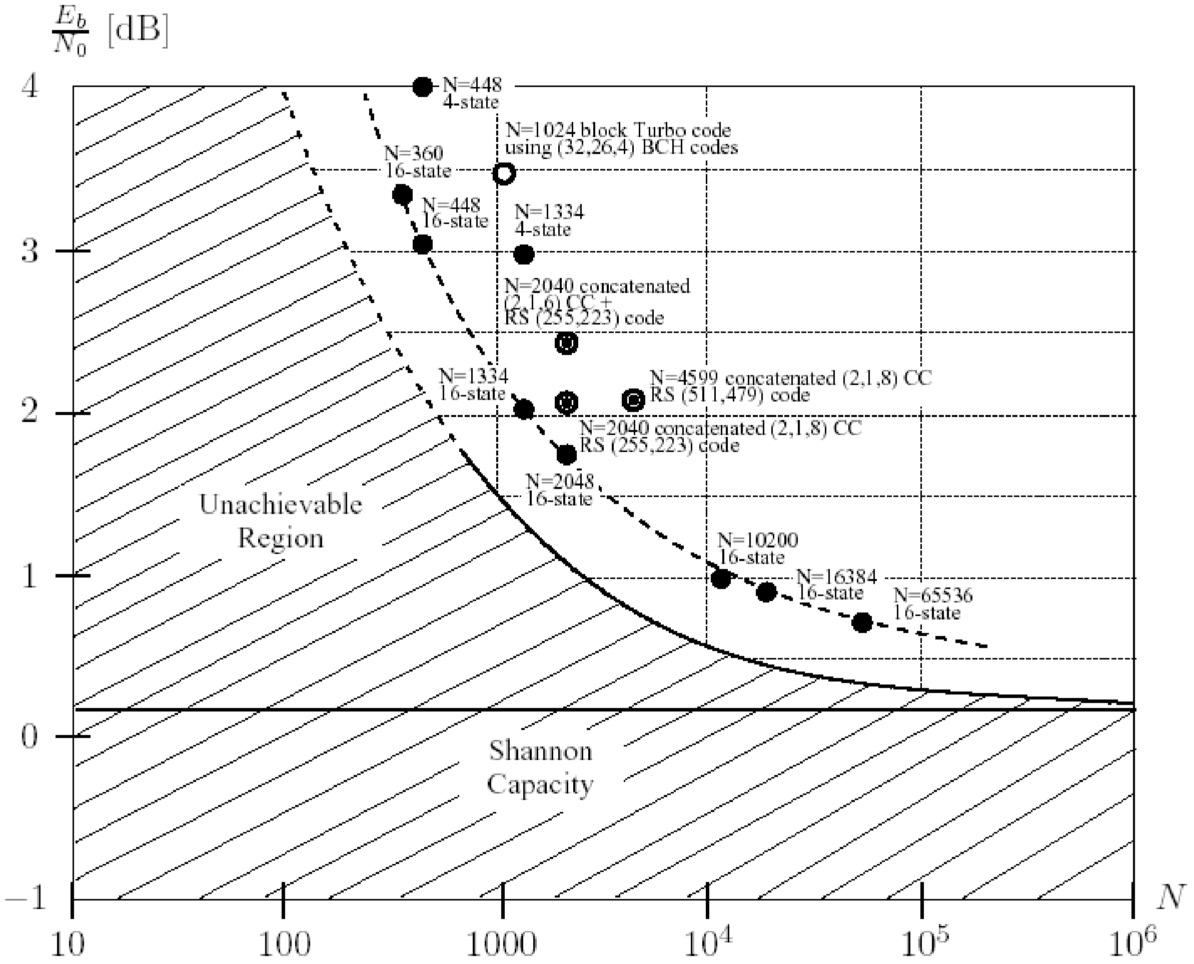


Figure 3.9: Block error rate P_B and sphere-packing lower bound for rate $R = 1/2$ coded example coding systems using BPSK. Turbo codes and selected classic concatenated coding schemes are compared.

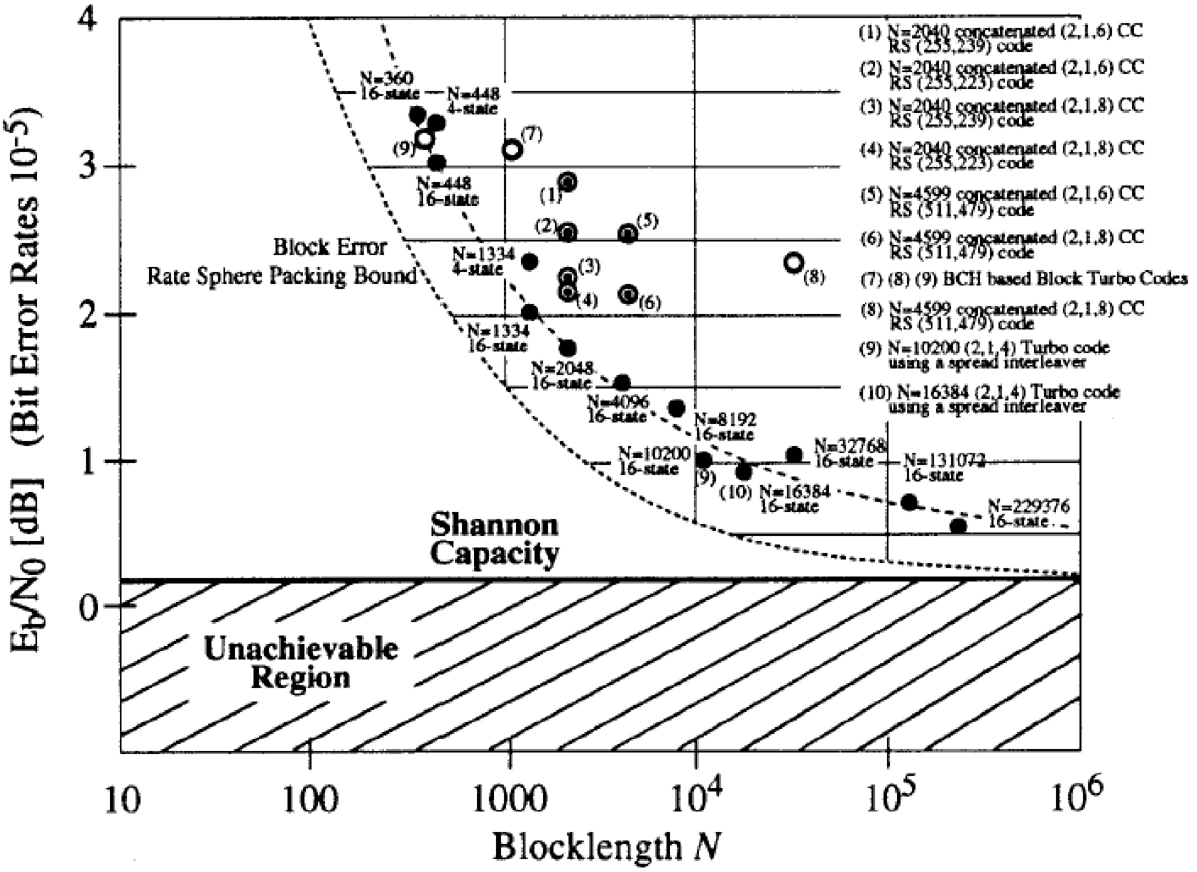


Figure. 3.10 The bit error performance of the same coding schemes as a function of the block length *N*.

The other component of the complexity consideration is the computational complexity; that is, the amount of processing that has to be performed to decode a codeword.

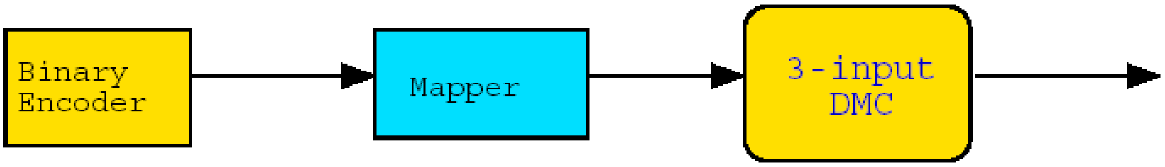
3.10 Achieving Capacity on Arbitrary DMCs with binary Codes

In Chapter 6 of [3], Gallager proved a coding theorem for parity-check codes, which may be addressed as follows.

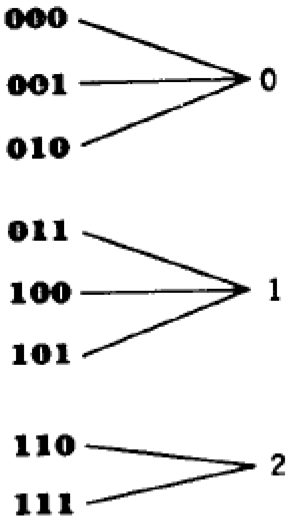
- ***Theorem:*** *Binary linear codes can be used to achieve capacity on an arbitrary discrete memoryless channel.*

Gallager also demonstrated a simple algorithm to generate codewords with the probability distributions required to achieve the results of the coding theorem. However, the problem is that finding decoding algorithm is not simple.

Fig. 3.11 shows an example, where each channel codeword is a sequence of *N* ternary independent digits with the probabilities $Q(0)=3/8$, $Q(1)=3/8$ and $Q(2)=2/8$.



(a)



(b)

Figure 3.11 An example for use of binary codes on a DMC. (a) System model. (b) Mapping rule

Reference

[1] S. G. Wilson, *Digital Modulation and Coding*. Prentice-Hall, 1996.

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

[3] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley and Sons, 1968.

[4] J. L. Massey, *Applied Digital Information Theory*. Course notes. ETH.

[5] J. G. Proakis, *Digital Communications*. 4rd ed. New York: McGraw-Hill, 2001.

[6] C. Schlegel and L. C. Perez, *Trellis and Turbo coding*. NJ: IEEE Press, 2004.

Suggested Reading

[R1] S. Shamai and I. Sason, “Variations on the Gallager bounds, connections and applications,” *IEEE Trans. Inform. Theory*, vol.48, no.12, pp.3029, Dec. 2002.