For the probability of picking codeword $\mathbf{c}_m = (a_{j_{m1}}, a_{j_{m2}}, \ldots, a_{j_{mn}})$, use the product distribution on the components

$$p(\mathbf{c}_m) = \prod_{l=1}^{n} p(c_{ml}) = \prod_{l=1}^{n} p(a_{j_{ml}})$$

This means that each component of each codeword is randomly chosen independently and with replacement. Finally, for the single-letter probability distribution, use the probability distribution that achieves channel capacity, $p(a_j) = p_j$.

Let E denote the expectation operator with respect to the random-code selection. Then

$$E[p_{e|m}] \le \varepsilon + \sum_{m' \ne m} E\left[\sum_{\mathbf{y}} \phi(\mathbf{c}_{m'}, \mathbf{y}) Q(\mathbf{y}|\mathbf{c}_m)\right]$$

$$\le \varepsilon + \sum_{m' \ne m} Pr[(\mathbf{x}, \mathbf{y}) \in \mathscr{T}_{xy}(\delta)]$$

where now the probability is over the ensemble of codes and $\mathbf{c}_m$ is replaced by $\mathbf{x}$ as a reminder that it is now a realization of the random variable associated with the random-code selection. All terms in the sum are now the same, so we have

$$E[p_{e|m}] \le \varepsilon + (M-1) Pr[(\mathbf{x}, \mathbf{y}) \in \mathscr{T}_{xy}(\delta)]$$

Because $\mathbf{x}$ is a randomly chosen codeword other than the codeword actually transmitted and $\mathbf{y}$ is the response of the channel to the randomly selected codeword that is transmitted, $\mathbf{x}$ and $\mathbf{y}$ are realizations of independent random variables. In step 3, we shall show that the second term can be made smaller than ε.

Before giving the next step of the proof, we will drop the m on the left side of the inequality based on the following argument. The right side of the bound no longer depends on m. Let $Pr[\mathbf{c}_m]$ be the probability of using the mth codeword. The average probability of error of the code is $p_e = \sum_m Pr[\mathbf{c}_m] p_{e|m}$. Over the ensemble, the expected value of the average probability of error is now bounded as follows:

$$E[p_e] = E\left[\sum_m Pr[\mathbf{c}_m] p_{e|m}\right]$$

$$= \sum_m Pr[\mathbf{c}_m] E[p_{e|m}]$$

$$\le \varepsilon + (M-1) Pr[(\mathbf{x}, \mathbf{y}) \in \mathscr{T}_{xy}(\delta)]$$

where $\mathbf{x}$ and $\mathbf{y}$ are independently selected n-tuples.

Step 3: All that remains is to evaluate $Pr[(\mathbf{x}, \mathbf{y}) \in \mathscr{T}_{xy}(\delta)]$, given that $\mathbf{x}$ and $\mathbf{y}$ are independent with probability $p(\mathbf{x})$ and $q(\mathbf{y})$ respectively. This is the probability that the pair $(\mathbf{x},\mathbf{y})$ turns out to be jointly typical even

though $\mathbf{x}$ and $\mathbf{y}$ are generated independently. But

$$Pr[(\mathbf{x}, \mathbf{y}) \in \mathscr{T}_{xy}(\delta)] = \sum_{(\mathbf{x},\mathbf{y}) \in \mathscr{T}_{xy}(\delta)} p(\mathbf{x}) q(\mathbf{y})$$

$$\le |\mathscr{T}_{xy}(\delta)| e^{-n[H(X)-\delta]} e^{-n[H(Y)-\delta]}$$

The inequality follows because, if $(\mathbf{x},\mathbf{y}) \in \mathscr{T}_{xy}(\delta)$, then $\mathbf{x}$ and $\mathbf{y}$ are each typical individually. Then using Theorem 5.5.2 to bound $|\mathscr{T}_{xy}(\delta)|$ and replacing $M-1$ with M gives

$$(M-1) Pr[(\mathbf{x}, \mathbf{y}) \in \mathscr{T}_{xy}(\delta)] \le M e^{n[H(X,Y)+\delta]} e^{-n[H(X)-\delta]} e^{-n[H(Y)-\delta]}$$

$$= e^{nR} e^{-n[I(X;Y)-3\delta]}$$

$$= e^{-n[C-R-3\delta]}$$

Because $R < C$ and δ is an arbitrary positive number, we can choose δ to satisfy the inequality $C - R - 3\delta > 0$. Hence for sufficiently large n,

$$(M-1) Pr[(\mathbf{x},\mathbf{y}) \in \mathscr{T}_{xy}(\delta)] < \varepsilon$$

Step 4: We have shown that for large enough blocklength n, the average probability of block decoding error over all codes satisfies

$$E[p_e] < 2\varepsilon$$

There must be at least one code at least as good as the average. Consequently, there exists at least one code whose probability of block decoding error satisfies

$$p_e < 2\varepsilon$$

Because ε is arbitrary, we can replace 2ε with ε to complete the proof of the theorem. □

5.6 THE RANDOM-CODING BOUND

The channel coding theorem of the previous section shows that codes with arbitrarily small probability of block decoding error exist at any rate smaller than the channel capacity C. However, we made no attempt to estimate how large the blocklength needed to be to attain a specified probability of decoding error. In this section we shall re-prove the channel coding theorem using an alternative and more delicate method that describes a little of the dependence between blocklength and probability of decoding error.

This time the derivation will be based on the maximum-likelihood decoder. Recall that the decoder with minimum probability of error estimates the transmitted codeword by comparing the a posteriori distributions $P(\mathbf{c}_m|\mathbf{y})$ and selecting that m for which $P(\mathbf{c}_m|\mathbf{y})$ is largest for the particular output $\mathbf{y}$ that was received. In the proof of the next theorem, we shall use instead the decoding rule that m is selected if $Q(\mathbf{y}|\mathbf{c}_m) > Q(\mathbf{y}|\mathbf{c}_{m'})$ for all $m' \ne m$. This is

potentially a decoder with a larger probability of error, but when all codewords are used with equal probability, it is equivalent to the minimum probability-of-error decoder. Because the theorem we are proving is an upper bound on p_e, it is permissible to use the maximum-likelihood decoding rule, even if the codewords are not used with equal probability. The probability of error of the optimum decoding rule can only be smaller.

□ **Theorem 5.6.1** Let $R = (1/n)\log M$. There exists a data transmission code of size M and blocklength n for the memoryless channel with block transition matrix $Q(\mathbf{y}|\mathbf{x})$ whose probability of block decoding error p_e satisfies

$$p_e \le \min_{s\in[0,1]} \min_{p(\mathbf{x})} \left\{ e^{nsR} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} p(\mathbf{x}) Q(\mathbf{y}|\mathbf{x})^{1/(1+s)} \right]^{1+s} \right\}$$

Proof Let $\{\mathbf{c}_0, \ldots, \mathbf{c}_{M-1}\}$ denote the set of codewords, and let the decoding rule be that the mth codeword is decoded when $\mathbf{y}$ is received if

$$Q(\mathbf{y}|\mathbf{c}_m) > Q(\mathbf{y}|\mathbf{c}_{m'})$$

for all $m' \neq m$. In case several values of m show the same maximum, we allow any arbitrary method of breaking the tie, but for bounding the probability of error, we assume that an error is then always made. The set of $\mathbf{y}$ that decode into $\mathbf{c}_m$ is

$$\mathcal{U}_m = \{\mathbf{y}: Q(\mathbf{y}|\mathbf{c}_m) > Q(\mathbf{y}|\mathbf{c}_{m'}) \quad \text{for all } m' \neq m\}$$

The characteristic function of this set $\phi_m(\mathbf{y})$ is by definition equal to 0 if $\mathbf{y} \notin \mathcal{U}_m$ and equal to 1 if $\mathbf{y} \in \mathcal{U}_m$. The first step in the proof is to bound $\phi_m(\mathbf{y})$ as follows:

$$1 - \phi_m(\mathbf{y}) \le \left\{ \sum_{m' \neq m} \left[\frac{Q(\mathbf{y}|\mathbf{c}_{m'})}{Q(\mathbf{y}|\mathbf{c}_m)} \right]^{1/(1+s)} \right\}^s \qquad \text{for all } s > 0$$

This inequality is true if $\phi_m(\mathbf{y}) = 1$ because the right side is nonnegative. It is also true if $\phi_m(\mathbf{y}) = 0$ because then there is at least one m' with $Q(\mathbf{y}|\mathbf{c}_m) \le Q(\mathbf{y}|\mathbf{c}_{m'})$. For this value of m'

$$1 \le \frac{Q(\mathbf{y}|\mathbf{c}_{m'})}{Q(\mathbf{y}|\mathbf{c}_m)}$$

The right side is greater than 1, so it can be raised to any positive power and will still be greater than 1. Hence

$$1 \le \left[\frac{Q(\mathbf{y}|\mathbf{c}_{m'})}{Q(\mathbf{y}|\mathbf{c}_m)} \right]^{1/(1+s)}$$

for all $s > 0$. The equality is preserved if any number of nonnegative terms are added to the right side, and still preserved if the resulting sum is raised to any positive power. Therefore the above bound follows.

Given that the mth codeword is transmitted, the probability of block decoding error is given by

$$\begin{aligned} p_{e|m} &= \sum_{\mathbf{y}\notin\mathcal{U}_m} Q(\mathbf{y}|\mathbf{c}_m) \\ &= \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{c}_m)[1 - \phi_m(\mathbf{y})] \\ &\le \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{c}_m) \left\{ \sum_{m'\neq m} \left[\frac{Q(\mathbf{y}|\mathbf{c}_{m'})}{Q(\mathbf{y}|\mathbf{c}_m)} \right]^{1/(1+s)} \right\}^s \\ &= \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)} \left[\sum_{m'\neq m} Q(\mathbf{y}|\mathbf{c}_{m'})^{1/(1+s)} \right]^s \end{aligned}$$

We are now ready to randomly select a code. Define a probability distribution on the set of codes such that the codewords are selected independently, with probability of selecting $\mathbf{x}$ as a codeword equal to $p(\mathbf{x})$. Thus the probability of the code $\{\mathbf{c}_0, \ldots, \mathbf{c}_{M-1}\}$ is given by

$$Pr[\{\mathbf{c}_0, \ldots, \mathbf{c}_{M-1}\}] = \prod_{m=0}^{M-1} p(\mathbf{c}_m)$$

For the moment, the probability distribution $p(\mathbf{x})$ will be arbitrary. The expected value $p_{e|m}$ over the ensemble is the ensemble average

$$\begin{aligned} E[p_{e|m}] &\le E\left\{ \sum_{\mathbf{y}} Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)} \left[\sum_{m'\neq m} Q(\mathbf{y}|\mathbf{c}_{m'})^{1/(1+s)} \right]^s \right\} \\ &= \sum_{\mathbf{y}} E\left\{ Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)} \left[\sum_{m'\neq m} Q(\mathbf{y}|\mathbf{c}_{m'})^{1/(1+s)} \right]^s \right\} \\ &= \sum_{\mathbf{y}} E[Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)}] E\left[\sum_{m'\neq m} Q(\mathbf{y}|\mathbf{c}_{m'})^{1/(1+s)} \right]^s \end{aligned}$$

where the first step follows because the expectation of a sum equals the sum of the expectations. The second step follows because the codewords are selected independently and the first term depends only on $\mathbf{c}_m$, while the second term does not depend on $\mathbf{c}_m$. Now suppose $s \le 1$. Then t^s is a concave function of t, so Jensen's inequality is applicable. That is,

$$E[t^s] \le [E(t)]^s$$

Therefore

$$\begin{aligned} E[p_{e|m}] &\le \sum_{\mathbf{y}} E[Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)}] \left[E \sum_{m'\neq m} Q(\mathbf{y}|\mathbf{c}_{m'})^{1/(1+s)} \right]^s \\ &= \sum_{\mathbf{y}} E[Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)}] \left\{ \sum_{m'\neq m} E[Q(\mathbf{y}|\mathbf{c}_{m'})^{1/(1+s)}] \right\}^s \end{aligned}$$

The M codewords are selected with identical distributions. Therefore $E[Q(\mathbf{y}|\mathbf{c}_m)^{1/(1+s)}]$ does not depend on m, so that

$$E[p_{e|m}] \le \sum_{\mathbf{y}} E[Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}]\{(M-1)E[Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}]\}^s$$
$$= (M-1)^s \sum_{\mathbf{y}} \{E[Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}]\}^{1+s}$$

We can replace $M-1$ with M without violating the inequality. Write the expectation explicitly in terms of $p(\mathbf{x})$:

$$E[p_{e|m}] \le M^s \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}\right]^{1+s}$$

The right side no longer depends on m, so we can remove the conditioning on m. Specifically, if $Pr[\mathbf{c}_m]$ is the probability of using the mth codeword, then $p_e = \sum_m Pr[\mathbf{c}_m]p_{e|m}$. Over the ensemble, the average probability of error has the expected value

$$E[p_e] = E\left[\sum_m Pr[\mathbf{c}_m]p_{e|m}\right]$$
$$= \sum_m Pr[\mathbf{c}_m]E[p_{e|m}]$$
$$\le M^s \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}\right]^{1+s}$$

Because the expected value over the ensemble of codes satisfies this bound, there must be at least one code that itself satisfies the bound. Hence there exists a code whose probability of block decoding error satisfies

$$p_e \le M^s \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}\right]^{1+s}$$

This is true for any $s \in [0, 1]$ and for any $\mathbf{p}$, and so holds even if we choose s and $\mathbf{p}$ to make the right side smallest. □

Theorem 5.6.1 is expressed in terms of block probabilities on the input blocks. Our next step is to express the bound in terms of per-letter probabilities. The bound will be given in terms of the random-coding exponent, which is given in the following definition.

□ **Definition 5.6.2** The *random-coding exponent*† is given by

$$E_r(R) = \max_{0 \le s \le 1} \max_{\mathbf{p}} \left[-sR - \log \sum_{k=0}^{K-1} \left(\sum_{j=0}^{J-1} p_j Q_{k|j}^{1/(1+s)}\right)^{1+s}\right]$$ □

†The use of E to denote the random-coding exponent should not be confused with the use of E to denote the expectation.

In Chapter 10, we will find that $E_r(R)$ is always positive for R smaller than C and is always equal to zero for R larger than C. A sketch of $E_r(R)$ for a binary symmetric channel with crossover probability 0.01 is shown in Fig. 5.10. The region where $E_r(R)$ is a straight line is the region where the maximum over s is achieved at $s = 1$. It turns out that for every channel, $E_r(R)$ has such a straight line whose slope is -1.

□ **Theorem 5.6.3** Let p_e be the probability of error of the best block code of size M and blocklength n for a memoryless channel and let $R = (1/n) \log M$. Then

$$p_e \le e^{-nE_r(R)}$$

Proof From the proof of Theorem 5.6.1, we have that

$$p_e \le e^{nsR} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})^{1/(1+s)}\right]^{1+s}$$

for every $\mathbf{p}$ and for all $s \in [0, 1]$. For a memoryless channel, $Q(\mathbf{y}|\mathbf{x})$ is given by the product distribution

$$Q(\mathbf{y}|\mathbf{x}) = \prod_{l=1}^{n} Q(y_l|x_l)$$

so we will choose $\mathbf{p}$ to be a product distribution as well:

$$p(\mathbf{x}) = \prod_{l=1}^{n} p(x_l)$$

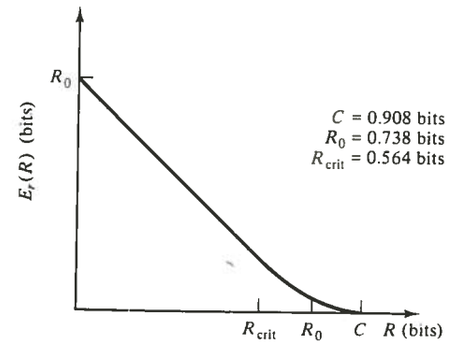

Figure 5.10 A random-coding exponent for block codes for a binary symmetric channel, $p = 0.01$.

We need not bother to prove that a product distribution gives the tightest bound because the inequality holds for any choice of $p(\mathbf{x})$. Therefore

$$p_e \le e^{nsR} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} \prod_{l=1}^{n} p(x_l)\ Q(y_l|x_l)^{1/(1+s)} \right]^{1+s}$$

The next step is to replace the sum of products with a product of sums. Notice that the sum over all $\mathbf{x}$ is a sum over all vectors $(x_1, x_2, \ldots, x_n)$. Similarly, the sum over all $\mathbf{y}$ is a sum over all vectors $(y_1, y_2, \ldots, y_n)$. That is,

$$p_e \le e^{nsR} \sum_{y_1} \sum_{y_2} \cdots \sum_{y_n} \left[\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \prod_{l=1}^{n} p(x_l) Q(y_l|x_l)^{1/(1+s)} \right]^{1+s}$$

Now apply the general rule†

$$\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \left[\prod_{l=1}^{n} A(x_l) \right] = \prod_{l=1}^{n} \left[\sum_{x_l} A(x_l) \right]$$

twice to interchange the sums and the product. This gives

$$p_e \le e^{nsR} \prod_{l=1}^{n} \left\{ \sum_{y_l} \left[\sum_{x_l} p(x_l) Q(y_l|x_l)^{1/(1+s)} \right]^{1+s} \right\}$$

For each l, the term in braces is the same. Replace $p(x_l)$ and $Q(y_l|x_l)$ with the abbreviated notation p_j and $Q_{k|j}$, now independent of l, to give the product of n identical terms. Therefore

$$p_e \le e^{nsR} \left[\sum_{k} \left(\sum_{j} p_j Q_{k|j}^{1/(1+s)} \right)^{1+s} \right]^{n}$$

Picking s and $\mathbf{p}$ to make this bound the tightest gives

$$p_e \le e^{-nE_r(R)}$$

to complete the proof of the theorem. □

We can state as a corollary to Theorem 5.6.3 Shannon's second coding theorem, which was already given as Theorem 5.5.3. This gives us an alternative proof of the channel coding theorem, and some new insights.

□ **Corollary 5.6.4 (Shannon's second coding theorem)** Suppose $R < C$, Then for any $\varepsilon > 0$ there exist a blocklength n and a code of blocklength n and rate R whose probability of block decoding error satisfies

$$p_e \le \varepsilon$$

†This is a more general form of the obvious expression

$$\left(\sum_j A_j \right) \left(\sum_k B_k \right) = \sum_j \sum_k A_j B_k$$

Proof We shall prove in Theorem 10.1.2 and Corollary 10.1.6 that $E_r(R)$ is positive for all R smaller than channel capacity. Therefore, in the theorem, pick n satisfying

$$p_e \le e^{-nE_r(R)} \le \varepsilon$$

Such an n always exists if $R < C$ because $E_r(R) > 0$. □

5.7 THE CHANNEL RELIABILITY FUNCTION

The random-coding exponent gives a bound on the probability of error as a function of blocklength. It is suggestive of the error exponent for hypothesis testing that was studied in Section 4.6. That error exponent was found to be asymptotically tight with blocklength.

It is natural to ask how the behavior of the average probability of decoding error of block codes depends asymptotically on blocklength, and whether the random-coding exponent is tight. These are difficult questions, which we introduce briefly in this section and return to more fully in Chapter 10. In that chapter, we shall expend a great deal of effort in learning how, for good codes, the probability of block decoding error $p_e(\mathscr{C})$ behaves as a function of the code's rate and blocklength, and in learning something about the structure of good codes.

The probability of block decoding error when codeword $\mathbf{c}_m$ is the transmitted codeword is

$$p_{e|m} = \sum_{\mathbf{y} \in \mathscr{Y}_m^c} Q(\mathbf{y}|\mathbf{c}_m)$$

If the M codewords are used with equal probability, then the average probability of decoded block error for the code $\mathscr{C}$ is

$$p_e(\mathscr{C}) = \sum_{m=0}^{M-1} \frac{1}{M} \sum_{\mathbf{y} \in \mathscr{Y}_m^c} Q(\mathbf{y}|\mathbf{c}_m)$$

Even when the code $\mathscr{C}$ is known, this expression is impractical to evaluate unless the code is very small, and small codes are not very good. Further, we do not know how to find good codes of large blocklength. Nevertheless, we will not be discouraged, but will press on and try to get some idea of how $p_e(\mathscr{C})$ for an optimal code $\mathscr{C}$ depends on the blocklength and rate of the code. Remarkably, we can get a partial answer even though we do not know how to find the optimal codes.

Let $p_e(n,R)$ denote the smallest possible probability of block decoding error of any code of blocklength n and rate R. That is,

$$p_e(n, R) = \min_{\mathscr{C}} p_e(\mathscr{C})$$

where the minimum is over all codes of blocklength n and rate R. The channel reliability at rate R is defined as

$$E^*(R) = \lim_{n \to \infty} \left[-\frac{1}{n} \log p_e(n, R) \right]$$

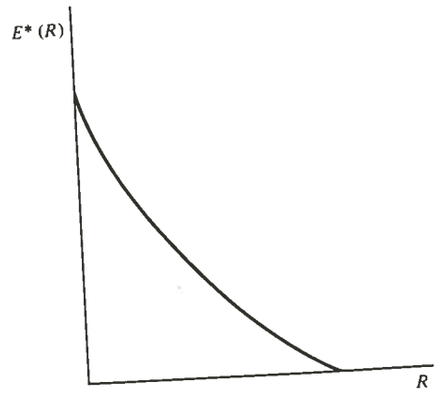

Figure 5.11 Typical reliability function for channel codes.

providing the limit exists. If the limit does exist,† the asymptotic behavior in n of $p_e(n, R)$ is given by

$$p_e(n, R) = e^{-nE^*(R) + o(n)}$$

where $o(n)$ is a term that goes to zero as n goes to infinity.

A channel reliability function $E^*(R)$ is defined for each channel **Q**. A sketch of a typical $E^*(R)$ is shown in Fig. 5.11. The function is monotonically decreasing in R and is conjectured to be convex. The true function $E^*(R)$ is known only for rates greater than that rate at which the random-coding bound $E_r(R)$ deviates from a straight line of slope -1. This is the *critical rate* R_{crit}. The rate R_{crit} occurs at the largest R where $E^*(R)$ has a slope of -1. For lower rates, only bounds on $E^*(R)$ are known.

The *cutoff rate* R_0, another parameter used to characterize a channel, is defined as the rate at which that tangent to $E_r(R)$ (or to $E^*(R)$) of slope -1 intersects the R axis. This rate can be expressed as

$$R_0 = \max_{\mathbf{p}} \left[-\log \sum_k \left(\sum_j p_j Q_{kj}^{1/2} \right)^2 \right]$$

Whereas the channel capacity of a channel **Q** is the rate beyond which it is impossible to communicate over the channel, the cutoff rate of **Q** is widely believed to be the rate beyond which it is very expensive to communicate over the channel. This belief stems from the fact that for some kinds of decoders for data transmission codes, known as sequential decoders, the complexity of the decoders grows very rapidly as the code rate is increased above R_0. However, a general proof of this significance for R_0 has never been discovered, and

†If the limit does not exist, then the lower and upper bounds described in this chapter become lower bounds on the limit infimum and upper bounds the limit supremum, respectively.

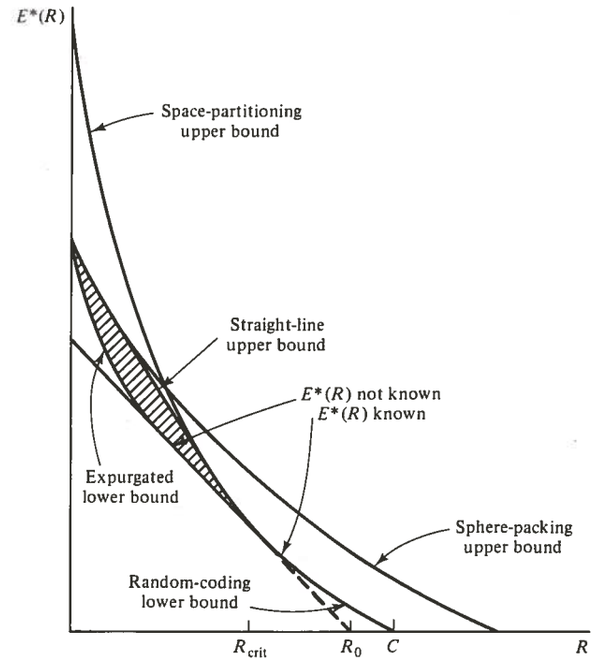

Figure 5.12 Bounds on the reliability function.

examples of channels are known for which the cutoff rate clearly has no meaning. (See Problem 5.26.)

The $E^*(R)$ function, to the extent it is known, is established by proving five bounds, three upper and two lower. These correspond respectively to three lower bounds and two upper bounds on the probability of block decoding error. These are the *random-coding* lower bound, which was derived in Section 5.6; the *expurgated* lower bound; the *space-partitioning*† upper bound; the *sphere-packing* upper bound; and the *straight-line* upper bound. Derivations of the latter four bounds will be found in Chapter 10; they are all rather difficult. Figure 5.12 gives a sketch of the five bounds. Figure 5.13 gives the bounds evaluated for a specific example—a binary symmetric channel with symbol error probability equal to 0.1.

The upper bounds on $E^*(R)$, corresponding to lower bounds on p_e, are derived by identifying a binary hypothesis-testing problem in the decoding problem and applying the methods of Chapter 4. A lower bound on the probability of error of the hypothesis-testing problem then is used to infer a lower bound on the probability of error of the decoding problem.

†This is the bound usually called the "sphere-packing bound." We apologize for the nonstandard terminology, which is selected here so that the latter bound, proved by packing spheres, can be referred to as the sphere-packing bound.

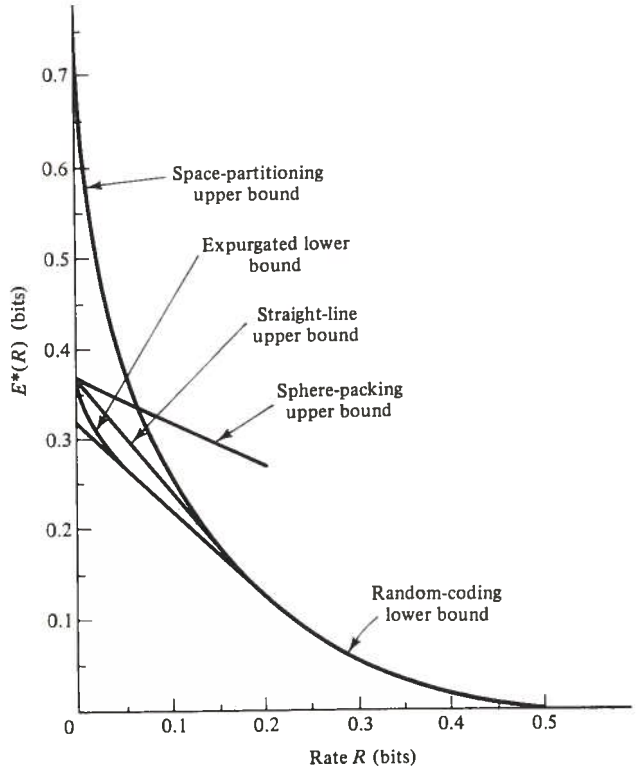

Figure 5.13 Bounds on the reliability function for a binary symmetric channel, $p = 0.1$.

For high-rate codes, the lower bound on p_e is called the *space-partitioning bound*. The nib of its derivation is the hypothesis-testing problem given by

$$H_0: \mathbf{c}_m = \text{transmitted codeword}$$

$$H_1: \mathbf{c}_m \neq \text{transmitted codeword}$$

where $\mathbf{c}_m$ is chosen to be one of the more difficult codewords to decode.

For low-rate codes, the lower bound on p_e is called the *sphere-packing bound*. The nib of its derivation is a different hypothesis-testing problem given by

$$H_0: \text{of } \mathbf{c}_m \text{ and } \mathbf{c}_{m'},\ \mathbf{c}_m = \text{transmitted codeword}$$

$$H_1: \text{of } \mathbf{c}_m \text{ and } \mathbf{c}_{m'},\ \mathbf{c}_{m'} = \text{transmitted codeword}$$

where $\mathbf{c}_m$ and $\mathbf{c}_{m'}$ are chosen to be two of the more difficult codewords to distinguish.

Speaking very loosely, and only from the nature of the derivations of the bounds, it appears that the performance of good codes of large blocklength