also be introduced among the parameters that force a tradeoff in the selection of a modulation scheme.

3.5 Capacity of the AWGN channel

We now evaluate the capacity of the AWGN channel; specifically, we examine the Gaussian channel and a code built out of one-dimensional (i.e., real) elementary signals. For every channel use, the input is x and the output is the real random variable $y = x + z$. Assume initially that no constraint is put on the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, except for a constraint on the energy of the input signal, which has the form $\mathbb{E}\, x^2$. Since $z \perp\!\!\!\perp x$, we have (see Appendix A for the relevant definitions):

$$\mathrm{H}(y \mid x) = \mathrm{H}(x + z \mid x) = \mathrm{H}(z \mid x) = \mathrm{H}(z) \tag{3.32}$$

and hence

$$\begin{aligned} \mathrm{I}(x; y) &= \mathrm{H}(y) - \mathrm{H}(y \mid x) \\ &= \mathrm{H}(y) - \mathrm{H}(z) \end{aligned} \tag{3.33}$$

Now (Theorem A.3.1, Appendix A),

$$\mathrm{H}(z) = \frac{1}{2} \log 2\pi e\, \mathbb{E}\, z^2 \tag{3.34}$$

and, since $\mathbb{E}\, z = 0$,

$$\mathbb{E}\, y^2 = \mathbb{E}\, (x + z)^2 = \mathbb{E}\, x^2 + \mathbb{E}\, z^2 \tag{3.35}$$

Thus, the entropy of $\mathcal{Y}$ is bounded above by $\frac{1}{2} \log 2\pi e (\mathbb{E}\, x^2 + \mathbb{E}\, z^2)$, and in conclusion

$$\begin{aligned} \mathrm{I}(x; y) &\leq \frac{1}{2} \log 2\pi e (\mathbb{E}\, x^2 + \mathbb{E}\, z^2) - \frac{1}{2} \log 2\pi e\, \mathbb{E}\, z^2 \\ &= \frac{1}{2} \log \left(1 + \frac{\mathbb{E}\, x^2}{\mathbb{E}\, z^2}\right) \end{aligned} \tag{3.36}$$

and the maximum of $\mathrm{I}(\mathcal{X}; \mathcal{Y})$ is attained when x is a Gaussian random vector with zero mean and variance $\mathbb{E}\, x^2$. This maximum value is the information capacity of the Gaussian channel:

$$C = \frac{1}{2} \log (1 + \mathrm{SNR}) \quad \text{bit/dimension} \tag{3.37}$$

where

$$\mathrm{SNR} \triangleq \frac{\mathbb{E}\, x^2}{\mathbb{E}\, z^2} \tag{3.38}$$

Observation 3.5.1 If x and z are complex, then the maximum value of the mutual information is achieved for x Gaussian, with zero mean, variance $\mathbb{E}|x|^2$, and independent real and imaginary parts. Moreover, it is convenient to express C in bit/dimension pair:

$$C = \log(1 + \text{SNR}) \quad \text{bit/dimension pair} \tag{3.39}$$

where now $\text{SNR} \triangleq \mathbb{E}|x|^2/\mathbb{E}|z|^2$.

Observation 3.5.2 The SNR (3.38) can be given different expressions as follows. Assume x to be N-dimensional. The signal variance is $\mathcal{E}$, while the noise variance is $NN_0/2$. Since the Shannon bandwidth of a signal is $W = N/2T$, we may write

$$\text{SNR} = \frac{\mathcal{E}}{NN_0/2} = \frac{\mathcal{E}/T}{N_0 W} = \frac{\mathcal{P}}{N_0 W} \tag{3.40}$$

Recalling (3.28), we can also express the SNR in the form

$$\text{SNR} = \frac{\mathcal{E}_b R_b}{N_0 W} \tag{3.41}$$

Since SNR$= 2\mathcal{E}/(NN_0)$, we see that, as $N \to \infty$, if $\mathcal{E}/N_0$ remains constant then the number of bits per dimension expressed by C tends to zero, because SNR$\to 0$. The number NC of bits that can be reliably transmitted over N dimensions tends to the constant limit $\log(e)\mathcal{E}/N_0$. We shall return on this in Section 3.5.1.

Sketch of the proof of the capacity theorem

The capacity (3.37) is also the maximum achievable rate for the channel. A fundamental theorem of Information Theory (Appendix A) shows that there exists a sequence of codes with rate C and block length n such that, as $n \to \infty$, the error probability tends to 0. Here we provide a qualitative summary of the proof, in the form originally given by Shannon.

As we are considering one-dimensional elementary signals and code words with length n, the dimensionality of the signal constellation is n. Observe now that the volume of a n-dimensional sphere Σ_n with radius r is proportional to r^n; thus, the volume of the shell between $r - \epsilon$ (with $0 < \epsilon < r$) and r is proportional to $r^n - (r - \epsilon)^n$. The ratio between the volume of the shell and the volume of the sphere is

$$\frac{r^n - (r - \epsilon)^n}{r^n} = 1 - \left(1 - \frac{\epsilon}{r}\right)^n$$

and tends to 1 as $n \to \infty$, no matter what the thickness ϵ of the shell is. This phenomenon, called *sphere hardening*, is summarized by saying that the volume of a n-dimensional sphere tends to concentrate near its surface as $n \to \infty$.

Next, consider a set of code words $\mathbf{x}$ whose components are subject to the energy constraint $\mathbb{E}x^2 \leq \mathcal{E}$, and let the received vector be $\mathbf{y} = \mathbf{x} + \mathbf{z}$. Let us first apply the sphere-hardening concept to the noise vector $\mathbf{z}$. As n grows to infinity, due to the law of large numbers, the squared length of vector $\mathbf{z}$ tends to a constant value:

$$\sum_{i=1}^{n} z_i^2 \approx n\mathbb{E}z^2 = nN_0/2$$

where z_i are the independent, equally distributed random components of $\mathbf{z}$. Sphere hardening assures that, while fluctuations of the length of $\mathbf{z}$ are possible, they tend to vanish as $n \to \infty$, so that $\mathbf{x}+\mathbf{z}$ lies on the surface of the sphere $\Sigma_n(\mathbf{x})$, centered at $\mathbf{x}$ and with radius $\sqrt{nN_0/2}$. Thus, signals differing by a Euclidean distance less than $\sqrt{nN_0/2}$ cannot be detected without ambiguity. Conversely, $\mathbf{x}$ can be detected with vanishingly small ambiguity if $\Sigma_n(\mathbf{x})$ is disjoint from the spheres associated with the other code words: in fact, $\Sigma_n(\mathbf{x})$ is contained in the Voronoi region of $\mathbf{x}$.

Further, consider the received vector $\mathbf{y}$. Its squared length tends to

$$\sum_{i=1}^{n} y_i^2 \approx n\mathbb{E}y^2 = n(\mathbb{E}x^2 + \mathbb{E}z^2) \leq n(\mathcal{E} + N_0/2)$$

and consequently $\mathbf{y}$ lies within a sphere with radius $\sqrt{n(\mathcal{E} + N_0/2)}$. In these conditions, the maximum number of disjoint spheres $\Sigma_n(\mathbf{x})$ that can be accommodated inside the sphere with radius $\sqrt{n(\mathcal{E} + N_0/2)}$ is no more than the ratio of the volumes

$$\frac{[n(\mathcal{E} + N_0/2)]^{n/2}}{[n(N_0/2)]^{n/2}} = (1 + \text{SNR})^{n/2}$$

This is the number $|\mathcal{S}|$ of code words. Thus, the rate of the code is

$$C = \frac{\log |\mathcal{S}|}{n} = \frac{1}{2}\log(1 + \text{SNR}) \quad \text{bit/dimension} \tag{3.42}$$

This "sphere-packing" argument also shows that we cannot hope to send information at a rate greater than C with low probability of error.

3.5.1 The bandlimited Gaussian channel

Assume now that the transmission of signal $\mathbf{x}$ takes a time T. Assuming as usual that the dimensionality of the constellation is $N = 2WT$, with W its (Shannon-)

bandwidth occupancy, we transmit $2W$ dimensions per second. Thus, using (3.40), Equation (3.37) can be rewritten in the form

$$C = W \log \left(1 + \frac{\mathcal{P}}{N_0 W}\right) \quad \text{bit/s} \tag{3.43}$$

which expresses the capacity of the bandlimited AWGN channel.

Notice that, as $W \to \infty$, we have

$$C \to \frac{\mathcal{P}}{N_0} \log e \quad \text{bit/s} \tag{3.44}$$

which shows how capacity grows linearly with signal power, rather than logarithmically as in (3.37). Equation (3.44) also shows that, for a given $\mathcal{P}/N_0$, the capacity remains bounded even though W (and hence the number of signal dimensions) grows without bounds. This occurs because $\mathcal{P}$ is fixed, and hence the power per Hz tends to zero.

If (3.41) is used, we obtain

$$C = W \log \left(1 + \frac{\mathcal{E}_b R_b}{N_0 W}\right) \quad \text{bit/s}$$

Since for reliable transmission we must have $R_b < C$, we require that

$$\frac{R_b}{W} < \log \left(1 + \frac{\mathcal{E}_b R_b}{N_0 W}\right)$$

Solving this inequality for the minimum allowable $\mathcal{E}_b/N_0$, we obtain

$$\frac{\mathcal{E}_b}{N_0} > \frac{2^{R_b/W} - 1}{R_b/W}$$

as plotted in Figure 3.3. The curve in this figure demarcates the region in which arbitrary low $P(e)$ can be reached: for any given R_b/W there exists a minimum value of $\mathcal{E}_b/N_0$ that must be exceeded if arbitrarily high reliability must be achieved. Notice that, as W increases, the required $\mathcal{E}_b/N_0$ approaches the lower limit

$$\lim_{W \to \infty} \frac{2^{R_b/W} - 1}{R_b/W} = \ln 2 \Leftrightarrow -1.6 \text{ dB}$$

Moreover, as $R_b/W > 2$, that is, when bandwidth is constrained, the energy-to-noise ratio required for reliable transmission increases dramatically. The region

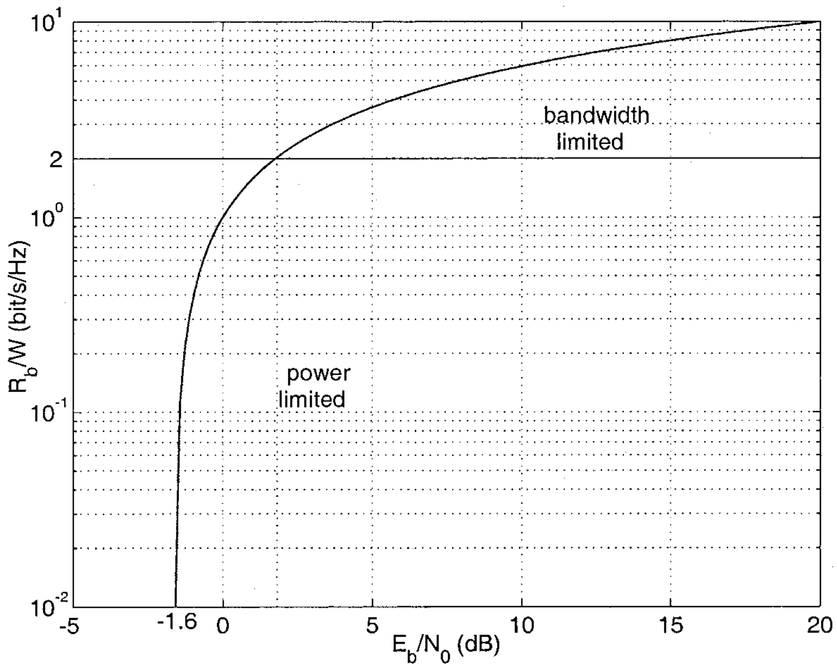

Figure 3.3: *Capacity limits for the bandlimited AWGN channel.*

where $R_b/W > 2$ (more than 2 bit/s/Hz, or equivalently more than 1 bit per dimension) is usually referred to as the *bandwidth-limited region,* and the region where $R_b/W < 2$ as the *power-limited region.* Figure 3.3 suggests that if the available power is severely limited, then we should compensate for this limitation by increasing the bandwidth occupancy, while the cost of a bandwidth limitation is an increase in the transmitted power.

Example 3.6

In this example we exhibit explicitly an M-ary signal constellation that, with no bandwidth constraint, has an error probability that tends to 0, as $M \to \infty$, provided that $\mathcal{E}_b/N_0 > \ln 2$, and hence shows the best possible behavior asymptotically. This is the set of M orthogonal, equal-energy signals defined by

$$(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 0, & i \neq j \\ \mathcal{E}, & i = j \end{cases} \tag{3.45}$$

This signal set has dimensionality $N = M$. Due to the special symmetry of this signal set, the Voronoi regions of the signals are all congruent (more on this *infra*, in Section 3.6), and hence the error probability $P(e \mid \mathbf{x}_i)$ is the same for all transmitted