Chapter 10

Gaussian Channel

In communication theory it is often assumed that the transmitted signals are distorted by some noise. The most common noise to assume is additive Gaussian noise, i.e. the so called Additive White Gaussian Noise channel, AWGN. Even though the noise in reality is more complex, this model is very efficient when simulating for example background noise or amplifier noise. Then the model can be complemented by e.g. impulse noise or other typical noise models that are out there. In this chapter we will have a closer look at AWGN channels and see how the previous theory applies here. We will derive a fundamental limit of the signal to noise ration (SNR) specifying when it is not possible to achieve reliable communication.

10.1 Gaussian channel

In a communication system data is often represented in a binary form. However, binary digits are elements in a discrete world, and, in all cases, we need to represent it in a continuous form. It can be as a magnetization on a hard disc, polarization of a transistor in a flash memory or signals in a cable or an antenna. To model this we need to consider continuous variables instead of the discrete we have had so far in our channel models. The noise added on the channel is typically Gaussian (i.e. Normal distributed) and represents for example the background noise, amplifier noise in the transceivers and signals from other communication systems working in the same frequency bands.

A signal in a digital communication system can be represented as by a continuous random variable. This value can be decomposed in two parts added together

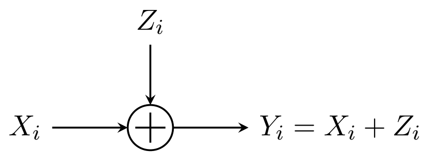
$$Y = X + Z$$

where X is the information carrier component and Z noise component. The average power allocated by the variable X is defined as the second moment,

$$P = E\big[X^2\big]$$

We are now ready to define a Gaussian channel as follows.

Definition 30 *A* ***Gaussian channel*** *is a time-discrete channel with input X and output $Y = X + Z$, where Z models the noise and is Normal distributed, $Z \in \mathcal{N}(0, \sqrt{N})$.*



The communication signalling is limited by a power constraint on the transmitter side,

$$E\big[X^2\big] \leq P$$

□

Without the power constraint in the definition we would be able to choose as many signal alternatives as far apart as we like. Then we would be able to transmit as much information as we like in a single channel use. With the power constraint we get a more realistic system where we need to find other means that increasing the power to get a higher information throughput over the channel.[1] To see how much information is possible to transmit over the channel we again maximizes the mutual information between the transmitted variable X and the received variable Y, with the side condition that the power is limited by P.

Definition 31 *The* ***information capacity*** *for a Gaussian channel is*

$$C = \max_{f(x),P} I(X;Y)$$

□

As before when calculating the capacity we can use

$$I(X;Y) = H(Y) - H(Y|X)$$

Then the second term is

$$H(Y|X) = H(X + Z|X) \stackrel{(a)}{=} H(Z|X) \stackrel{(b)}{=} H(Z)$$

In (a) we used that conditioned on X, $X + Z$ is a known shift in position of Z which does not change the entropy. To get equality (b) we note that X and Z are independent (the noise in the transmission is independent of the transmitted symbol). This means the information over the channel can be viewed as the difference in entropy between the received symbol and the noise,

$$I(X;Y) = H(Y) - H(Z)$$

[1]Before Shannon published his work it was the common knowledge that to get higher throughput it was necessary to increase the power of the transmitted signals. What Shannon showed was that there are ways to reach high communication rates at maintained power.

With the statistics if the noise known to be a normal distribution with zero mean and variance N, we can get

$$H(Z) = \tfrac{1}{2}\log(2\pi e N)$$

We also know from the previous chapter that for a given mean and variance, the Gaussian distribution maximizes the entropy. So, maximizing $H(Y)$ over all distributions of X gives

$$\max_{f(x),P} H(Y) \leq \tfrac{1}{2}\log(2\pi e\sigma^2)$$

where we get equality for $Y \in \mathcal{N}(0,\sigma)$. Since $Y = X + Z$, we can use that the sum of two Gaussian variables is again Gaussian. Then, by letting $X \in \mathcal{N}(0,\sqrt{P})$, we get the desired distribution on Y, where $\sigma^2 = P - N$. Hence, the information capacity is given by

$$\begin{aligned}
C &= \max_{f(x),P} I(X;Y) \\
&= \max_{f(x),P} H(Y) - H(Z) \\
&= \frac{1}{2}\log(2\pi e(P+N)) - \frac{1}{2}\log(2\pi e N) \\
&= \frac{1}{2}\log\Big(\frac{2\pi e(P+N)}{2\pi e N}\Big) \\
&= \frac{1}{2}\log(1+\frac{P}{N})
\end{aligned}$$

We formulate this as a theorem.

Theorem 44 *The mutual information for a Gaussian channel is maximized for $X \sim \mathcal{N}(0,\sqrt{P})$, as*

$$\begin{aligned}
C &= \max_{f(x),P} I(X;Y) \\
&= \max_{f(x),P} H(Y) - H(Z) \\
&= \frac{1}{2}\log\Big(1+\frac{P}{N}\Big)
\end{aligned}$$

□

Similar to the discrete case we want to consider which code rates that can give a code where the error probability goes to zero as the length of the codewords tend to infinity. To start with we define the terminology of *achievable code rate* as below.

Definition 32 *A code rate is achievable if there exists a $(2^{nR}, n)$ code that satisfy the power constraint, such that the probability of error P_e tends to zero. The* ***channel capacity*** *is the supremum of all achievable rates.* □

In a similar way as for the discrete case we can formulate the channel coding theorem.[2]

[2] In this text the proof for the channel coding theorem for the Gaussian channel is omitted.

Theorem 45 *The Channel capacity of a Gaussian channel with power constraint P and noise variance N is*

$$C = \frac{1}{2}\log\Big(1 + \frac{P}{N}\Big)$$

□

The terminology *signal to noise ratio*, SNR, is often used for the relation between the signal power and the noise power. In this case the signal power is P while the noise has the power $E\big[Z^2\big] = N$. Hence, in this case $\mathrm{SNR} = \frac{P}{N}$. Depending on the topic and what type of system considered, there are many different ways to define the SNR. It is often important to be aware of the specific definition used in a text.

10.2 Parallel Gaussian Channels

In some cases there can be several independent parallel Gaussian channels used by the same communication system. In Figure 10.1 there are n such parallel channels. Each of the channels has a power constraint $P_i = E\big[X_i^2\big]$ and a noise variance N_i. The total power is $P = \sum_i P_i$.

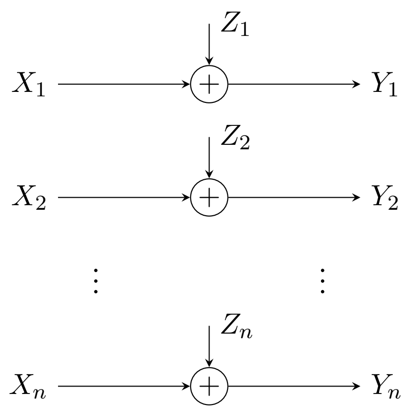


Figure 10.1: N independent parallel Gaussian channels.

Since we consider n independent Gaussian channels, the mutual information between the transmitted vector $\boldsymbol{X} = X_1, \dots, X_n$ and the received vector $\boldsymbol{Y} = Y_1, \dots, Y_n$ can be written as

$$\begin{aligned} I(\boldsymbol{X}, \boldsymbol{Y}) = I(X_1, \dots, X_n; Y_1, \dots, Y_n) &\overset{(a)}{\leq} \sum_{i=1}^{n} I(X_i; Y_i) \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} \frac{1}{2}\log\Big(1 + \frac{P_i}{N_i}\Big) \end{aligned}$$

where we have equality in (a) if the variables X_i are independent and in (b) if they are Gaussian. Since $\sigma_X^2 = E[X^2] = P$, we can maximize the mutual information for the set of P_i by using independent Gaussian varibale $X_i \in \mathcal{N}(o, \sqrt{P})$. To get the capacity we maximize the above expression with respect to P_i. With the additional constraint $\sum_i P_i = P$ we can use the Lagrange multiplier to achieve this. The maximization function is then given by

$$J = \sum_{i=1}^{n} \frac{1}{2} \log\Big(1 + \frac{P_i}{N_i}\Big) + \lambda\Big(\sum_{i=1}^{n} P_i - P\Big)$$

Setting the derivative equal to zero we get

$$\frac{1}{2\ln 2}\frac{1}{P_i + N_i} + \lambda = 0$$

or, equivalently,

$$P_i + N_i = -\frac{1}{\lambda 2\ln 2} = B$$

Here, the constant B is independent of the sub-channel, i. We end up with an equation system with $n+1$ variables (P_1,. . .,P_n,B) and the same number of equations.

$$\begin{cases} P_i = B - N_i, & \forall i \\ \sum_i P_i = P \end{cases}$$

This is solvable, and $P_i = \frac{1}{n}P + \frac{1}{n}\sum_j N_j - N_i$. However, we do have another requirement, from the reality, saying that $P_i \geq 0$. From the Kuhn-Tucker conditions[3] we can rewrite the first set equations, and remaining the optimality, as

$$P_i = \big(B - N_i\big)^+$$

where

$$(x)^+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

The above modification Means that some channels may have too much noise and should not be used at all. We summarize the derivations as a theorem.

Theorem 46 *Given k independent parallel Gaussian channels with noise variance N_i, $i = 1, \ldots, k$, and a restricted total transmitted power, $\sum_i P_i = P$.*

The capacity is given by

$$C = \frac{1}{2}\sum_{i=1}^{n} \log\Big(1 + \frac{P_i}{N_i}\Big)$$

where

$$P_i = (B - N_i)^+, \qquad (x)^+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and B is such that $\sum_i P_i = P$. □

[3]The Kuhn-Tucker method can be seen as a generalization of the Lagrangian multiplier method, often used in non-linear optimization. It is also known under the name Karush-Kuhn-Tucker.

This method is often referred to as *water filling*, which can be seen in the next example.

Example 60 Assume we have a system with four independent Gaussian channels with noise variance $N_1 = 2$, $N_2 = 4$, $N_3 = 6$a and $N_4 = 3$. The total power used in transmission is restricted to $P = 6$. The condition $P_i = B - N_i$ is equivalent to

$$B = P_i + N_i$$

and, hence we get for all four sub-channels

$$\begin{aligned} 4B &= P_1 + N_1 + P_2 + N_2 + P_3 + N_3 + P_4 + N_4 \\ &= \underbrace{P_1 + P_2 + P_3 + P_4}_{P=6} + \underbrace{N_1 + N_2 + N_3 + N_4}_{15} = 21 \end{aligned}$$

This gives $B = \frac{21}{4}$. Using this result would require $P_3 = \frac{21}{4} - 6 = -\frac{3}{4}$, which is not possible. The conclusion of this is that the 3rd sub-channel has too much noise and cannot be used if optimizing according to the algorithm.

In a second attempt to find an optimal distribution of the available power we turn off sub-channel 3 and use the other three. Similar as above we get

$$3B = \underbrace{P_1 + P_2 + P_4}_{P=6} + \underbrace{N_1 + N_2 + N_4}_{9} = 15$$

and $B = \frac{15}{3} = 5$. Hence, we get the power distribution

$$P_1 = 3 \quad P_2 = 1 \quad P_3 = 0 \quad P_4 = 2$$

and the capacity is

$$C = \frac{1}{2}\log\Big(1 + \frac{3}{2}\Big) + \frac{1}{2}\log\Big(1 + \frac{1}{4}\Big) + \frac{1}{2}\log\Big(1 + \frac{2}{3}\Big) \approx 1.1904\text{bit/Channel use}$$

In Figure 10.2 a graphical interpretation of the power allocation is shown. The noise level for each sub-channel act as surface of a landscape. Then the power is poured in the landscape as water and the surface will stay at the level B. The depth of the power for each sub-channel is the amount of power used. If, as in sub-channel 3, the water level does not reach above the noise level, this sub-channel should not be used. This interpretation is the origin of the terminology *water filling*.

10.3 Band-limited Gaussian channel

Often in a communication system the signalling is allowed to occupy a certain bandwidth. therefore, it is interesting to consider signals with a limited bandwidth. To start with the signal is sampled, and for that purpose we need to use the sampling theorem. Then each sample can be considered to be transmitted over a Gaussian channel.

To start with, a band limited signal is a signal where frequency contents is limited inside a bandwidth W. For example, speech is normally located within the frequency bandwidth 0-4 kHz. By modulating the signal it can be shifted up in frequency and located in a

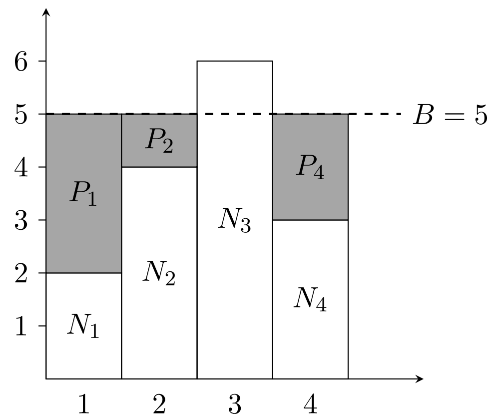


Figure 10.2: Water filling principle in Example 60.

higher band. Still it occupies 4 kHz bandwidth. In this way it is possible to allocate several bands of 4 kHz after each other, and in principle it we can pack one voice signal each 5kHz in the frequency band.

To transmit e.g. a voice signal we can either use analogue technology and transmit it as it is. But it is easier to process the signal if it is sampled and converted to digital data. Then it is possible to use a suitable source coding algorithm to reduce the redundancy. Following this there should also be a channel code to protect the information from channel errors. In this way it is possible to achieve much better quality at a lower transmission cost (i.e. bandwidth). Sampling the signal means taking the value from the continuous signal at periodic time values. Setting the sampling frequency to F_s, meaning there should be F_s samples each second. If the continuous time signal $x(t)$ is sampled with frequency F_s the sample values are given by

$$x_n(n) = x\big(\tfrac{n}{F_s}\big)$$

For a band limited signal with a bandwidth of W, the sampling theorem states that $F_s \geq 2W$ to be able to reconstruct the original signal. So, a voice signal that is band limited to $W = 4$ kHz should be sampled with at least $F_s = 8$ kHz.

The next theorem is the celebrated sampling theorem, introduced by Harry Nyquist in 1928 [19], and further improved by Shannon in [22]. Actually, Nyquist studied the number of pulses that can be transmitted over a certain bandwidth, which can be seen as the dual of the sampling theorem as we know it today, and what is given next.

Theorem 47 *Let $x(t)$ be a band limited signal, $f_{\max} \leq W$. If the signal is sampled with $F_s = 2W$ samples per second to form the sequence $x(\frac{n}{2W})$, it can be reconstructed with*

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right) \mathit{sinc}\left(t - \frac{n}{2W}\right)$$

where

$$\mathit{sinc}(t) = \frac{\sin(2\pi W t)}{2\pi W t}$$

□

The sinc-function is 1 for $t = 0$ and 0 for $t = k/2W$, $k \neq 0$. If the sampling frequency is less than $2W$ the reconstructed signal will be distorted due to aliasing and perfect reconstruction is not possible. The sampling frequency $F_s = 2W$ is often called the *Nyquist rate*. For a formal proof of the sampling theorem we refer to a basic course in signal processing, e.g. [20].

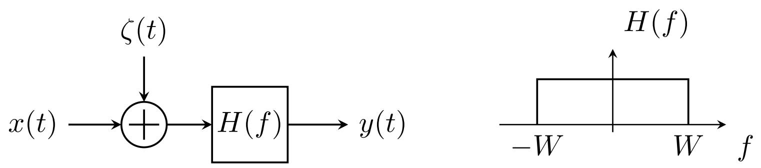
We are now ready to define a channel model for band limited signals. Assume we have a signal with highest frequency content $f_{\max} = W$, giving the required bandwidth W. Then, sampling the signal at the Nyquist rate we have the sampling frequency $F_s = 2W$. The sampling time, i.e. the time between two samples, is

$$T_s = \frac{1}{F_s} = \frac{1}{2W}$$

Sampling the signal $x(t)$ gives the sampled sequence

$$x_n = x(nT_s) = x\big(\tfrac{n}{2W}\big)$$

Definition 33 *A band limited Gaussian channel consists of a band limited input signal $x(t)$, where $f_{\max} = W$, additive white Gaussian noise $\zeta(t)$, and an ideal low-pass filter, as in the following figure.*



□

Since the signal $x(t)$ is band limited in the bandwidth W it passes the ideal filter without changes. The meaning of that the noise is white is that the power spectral density function occupies all frequencies with a constant value. This value is normally set to

$$R_\zeta(f) = \frac{N_0}{2}, \quad f \in \mathbb{R}$$

After the filtering we get the signal $z(t) = \zeta(t) * h(t)$, which is also band limited with power spectral density

$$R_z(f) = \begin{cases} \frac{N_0}{2}, & -W \leq f \leq W \\ 0, & \text{otherwise} \end{cases}$$

The corresponding auto correlation function is the inverse Fourier transform of the power spectral density function. In this case the noise auto correlation function is

$$r_z(\tau) = \frac{N_0}{2}\text{sinc}(\tau)$$

To get a time discrete sequence the received signal is sampled at the Nyquist rate, $F_s = 2W$. Then the auto-correlation sequence for the noise becomes

$$r_z\big(\tfrac{n}{2W}\big) = \begin{cases} \frac{N_0}{2}, & n = 0 \\ 0, & \text{otherwise} \end{cases}$$

This imply that the resulting sampled noise is normal distributed with zero mean and variance $N_0/2$,

$$z_n \in \mathcal{N}\big(0, \tfrac{N_0}{2}\big)$$

Hence, we can use the previous theory for the Gaussian channel. For each sample transmitted the capacity is

$$C = \frac{1}{2}\log\Big(1 + \frac{2\sigma_x^2}{N_0}\Big) \text{ bit/sample}$$

The power of the transmitted signal is constraint to P. That means each transmitted sample has the energy $\sigma_x^2 = \frac{P}{2W}$, which gives the capacity per sample

$$C = \frac{1}{2}\log\Big(1 + \frac{P}{N_0 W}\Big) \text{ bit/sample}$$

With $F_s = 2W$ samples every second, the achievable bit rate becomes

$$C = W\log\Big(1 + \frac{P}{N_0 W}\Big) \text{ bit/second}$$

We formulate this result as a theorem.

Theorem 48 *Let $x(t)$ be a band limited signal, $f_{\max} \leq W$, and $z(t)$ noise with power spectra $R_z(f) = N_0/2$, $|f| \leq W$. The channel*

$$y(t) = x(t) + z(t)$$

has the capacity (in bit/second)

$$C = W\log\Big(1 + \frac{P}{N_0 W}\Big)$$

□

Example 61 Today the main dominating technology for fixed Internet access is through ADSL (Asymmetric Digital Subscriber Line), which can give bit rates up to 26 Mb/s. But there is also a generation change ongoing and the Internet providers are upgrading to VDSL (Very high speed DSL) equipment. This will enable bit rates of up to the order of 150 Mb/s. The advantage with DSL technology is that it reuses the old telephone lines to access the housholds, so it is a relativly cheap technology to roll out. Comparing with optical networks in the access link (fibre to the home, FttH) where they must dig a new infrastructure of optical fibres to each house, this is an economically feasible technology.

In both ADSL and VDSL the speech signals are in the band $0-4$ kHz and the data signals are positioned from 25 kHz up to 2.2 MHz for ADSL and 17 MHz for VDSL (depending on which band-plan is used). To do capacity calculations on the VDSL band we neglect the speech band and assume that the available bandwidth is $W = 17$ MHz. The signalling level is set by the regulators (standardized by ITU-T) to -60 dBm/Hz.[4]

[4]Often the unit dBm/Hz is used for PSD. This means the power level expressed in mW, normalized with the bandwidth and expressed in dB, i.e. $P_{\text{dBm/Hz}} = 10\log_{10}(P_{\text{mW/Hz}})$.

The absolute maximum that is possible to transmit can be found when the noise is as low as possible. The thermal noise, or the Johnson-Nyquist noise, is the noise generated in electrical circuits. The thermal noise is typically white and at room temperature about -174dBm/Hz. We can now calculate the power and the noise variance as

$$P = 10^{-60/10} \cdot W \quad [\text{mW}]$$

$$N_0 = 10^{-174/10} \quad [\text{mW/Hz}]$$

Then the capacity is

$$C_{-174} = W \log\Big(1 + \frac{P}{N_0 W}\Big) = 17 \cdot 10^6 \log\Big(1 + \frac{10^{-60/10}}{10^{-174/10}}\Big) = 644 \text{ Mb/s}$$

While this is a maximum theoretical possible bit rate, in real cables it turn out that the noise level is higher. This is due to all other disturbances that ocure since the telephone cable act as an antenna. In a real situation the noise level is somewhere around -145 dBm/Hz. With htis level instead of the thermal noise we achive

$$C_{-174} = 17 \cdot 10^6 \log\Big(1 + \frac{10^{-60/10}}{10^{-145/10}}\Big) = 480 \text{ Mb/s}$$

In the real VDSL system the theoretical maximum bit rate is about 250 Mb/s but in practice it is about 150 Mb/s.

In many communication systems of today a wide bandwidth is used to get high bit-rates. The channel is typically not constant over the entire band, but there are variations both in noise level and signal attenuation. One popular method to signal over such channels is with OFDM (Orthogonal Frequency Division Multiplexing) modulation. This is used in e.g. WLAN (802.11), xDSL, DVB-T (digital TV) and the down link of LTE (Long Term Evolution). Then the band width is divided into several sub-bands that can be used independently of each other. Then we have a case very similar to the parallel Gaussian channels, which led to the water filling algorithm. To get a similar result we use the same approach with Lagrangian multiplier and set up the optimization function for n sub-channels with bandwidth W_Δ as

$$J = \sum_i C_i + \lambda\Big(\sum_i P_i - P\Big) = \sum_i W_\Delta \log\Big(1 + \frac{P_i}{N_{0,i} W_\Delta}\Big) + \lambda\Big(\sum_i P_i - P\Big)$$

Setting its derivative with respect to P_i equal to zero and solving for P_i we get

$$P_i = -\frac{W_\Delta}{\lambda \ln 2} - N_{0,i} W_\Delta$$

The first term is constant and can be assigned as

$$B = -\frac{W_\Delta}{\lambda \ln 2}$$

As in the previous case we can not have negative powers and we need to use the Kuhn-Tucker argument to achieve the water-filling algorithm,

$$\begin{cases} P_i = \big(B - N_{0,i} W_\Delta\big)^+ \\ \sum_i P_i = P \end{cases} \tag{10.1}$$

In many situations the signal is attenuated during the transmission. This is normally modelled by a filter affecting the signal $x(t)$ and the received signal is instead

$$y(t) = x(t) * g(t) + z(t)$$

After sampling the received signal power is $P_{\text{rec}} = P|G|^2$, where $|G(f)| = G$ is assumed to be constant over the considered bandwidth. Then the capacity becomes

$$C = W \log\Big(1 + \frac{P|G|^2}{N_0 W}\Big)$$

For the OFDM type of channel the water filling argument corresponding to (10.1) becomes

$$\begin{cases} P_i = \Big(B - \frac{N_{0,i} W_\Delta}{|G_i|^2}\Big)^+ \\ \sum_i P_i = P \end{cases}$$

where G_i is assumed to be constant over the sub-channel. However the attenuation can vary between sub-channels and in this case both the noise level and attenuation level can be considered to be frequency dependent over the bandwidth.

10.4 Fundamental Shannon limit

One of the most famous results from Information theory is the fundamental limit, or Shannon limit, that sets requirements on the signal to noise ratio for reliable communication. We will in this section derive this limit in the general case without restrictions. To reach this limit it is required that the coding rate goes to zero, and we will also consider the case when the coding rate is fixed.

Consider a band limited Gaussian channel with bandwidth W and noise level N_0. If the transmitted power constraint is P, then the capacity is given by

$$C = W \log\Big(1 + \frac{P}{N_0 W}\Big) \text{ bit/second}$$

If we not have any other constraints we would like to use as much bandwidth as possible. In theory the available bandwidth is infinite, and therefore we let $W \to \infty$ in the formula.

$$\begin{aligned} C_\infty &= \lim_{W\to\infty} W \log\Big(1 + \frac{P/N_0}{W}\Big) \\ &= \lim_{W\to\infty} \log\Big(1 + \frac{P/N_0}{W}\Big)^W \\ &= \log e^{P/N_0} = \frac{P/N_0}{\ln 2} \end{aligned}$$

Assigning the achieved bit rate as R_b, it is required that this is not more than the capacity, $C_\infty > R_b$. Further, assume that the signalling time is T_s and that in each signal k information bits are transmitted. Then

$$PT_s = E_s = E_b k \tag{10.2}$$

where E_s is the average energy per transmitted symbol and E_b is the average energy per information bit. The variable E_b is a very important number since it is something that can be compared between different systems, without having the same number of bits per symbol or even the same coding rate. Then a very system independent signal to noise ratio is SNR $= E_b/N_0$. From (10.2) we can write the energy per bit as

$$E_b = \frac{PT_s}{k}$$

Considering the ration between C_∞ and the bit rate R_b we get

$$\frac{C_\infty}{R_b} = \frac{P/N_0}{\ln 2}\frac{T_s}{k} = \frac{E_b/N_0}{\ln 2} > 1$$

where we used that $R_b = \frac{k}{T_s}$ and that for reliable communication we require $C_\infty > R_b$. Rewriting the above, we can conclude that for reliable communication we require the SNR to be

$$\frac{E_b}{N_0} > \ln 2 = 0.69 = -1.59 \text{ dB}$$

The value -1.6 dB is the well known Shannon limit and constitute a hard limit for when it is possible to achieve reliable communication. If the SNR is less than this limit it is not possible to reach error probability that tends to zero, independent of what system is used.

10.4.1 Limit for fix coding rate

In the above calculations there are no limits on either bandwidth or coding rate. In fact, it will require coding rate that goes to zero and a computational complexity that goes to infinity to reach this limit. In this section we will derive a corresponding limit for fixed code rate. First, assume that a codeword consists of N samples and that there are K information bits in it, giving a $(2^K, N)$ code with rate

$$R = \frac{K}{N}$$

The duration of time for a codeword can then be set to T, and assuming the Nyquist rate, this is coupled to the number of samples through $N = TF_s = 2WT$ sample/codeword. The information bit rate is the number of information bits in a codeword over the duration,

$$R_b = \frac{K}{T} = 2W\frac{K}{N} = 2WR$$

Similarly, in each codeword we use an average energy of KE_b, and the corresponding power is

$$P = \frac{KE_b}{T}$$

With this at hand, we can rewrite the SNR of the capacity formula as

$$\frac{P}{N_0W} = \frac{KE_b}{TN_0W} = 2WR\frac{E_b}{N_0W} = 2\frac{E_b}{N_0}R$$

Then, since the bit rate is less than the capacity, we get

$$R_b = 2WR < W \log\Big(1 + 2\frac{E_b}{N_0}R\Big)$$

which gives

$$1 + 2\frac{E_b}{N_0}R > 2^{2R}$$

or, equivalently,

$$\frac{E_b}{N_0} > \frac{2^{2R} - 1}{2R}$$

Using, e.g. a code with rate $R = \frac{1}{2}$ we can see that the limit is now shifted to

$$\frac{E_b}{N_0} > 1 = 0 \text{ dB}$$

To get a better communication environment we need to decrease the code rate. It can only be decreased down to zero, where we see that the bound has a limit value in

$$\frac{E_b}{N_0} > \lim_{R\to 0} \frac{2^{2R} - 1}{2R} = \lim_{R\to 0} \frac{2 \cdot 2^{2R} \ln 2}{2} = \ln 2 = -1.59 \text{ dB}$$

where we used l'Hospital's rule to derive the limit. From this we can see that to reach the limit -1.59 dB we need to have the code rate approaching zero.

10.4.2 Limit for QAM

SNR gap

10.4.3 Limit for MIMO

10.5 Coding gain and shaping gain

To be done.