Digital Communication Systems

S Le Goff

October 2023

School of Engineering

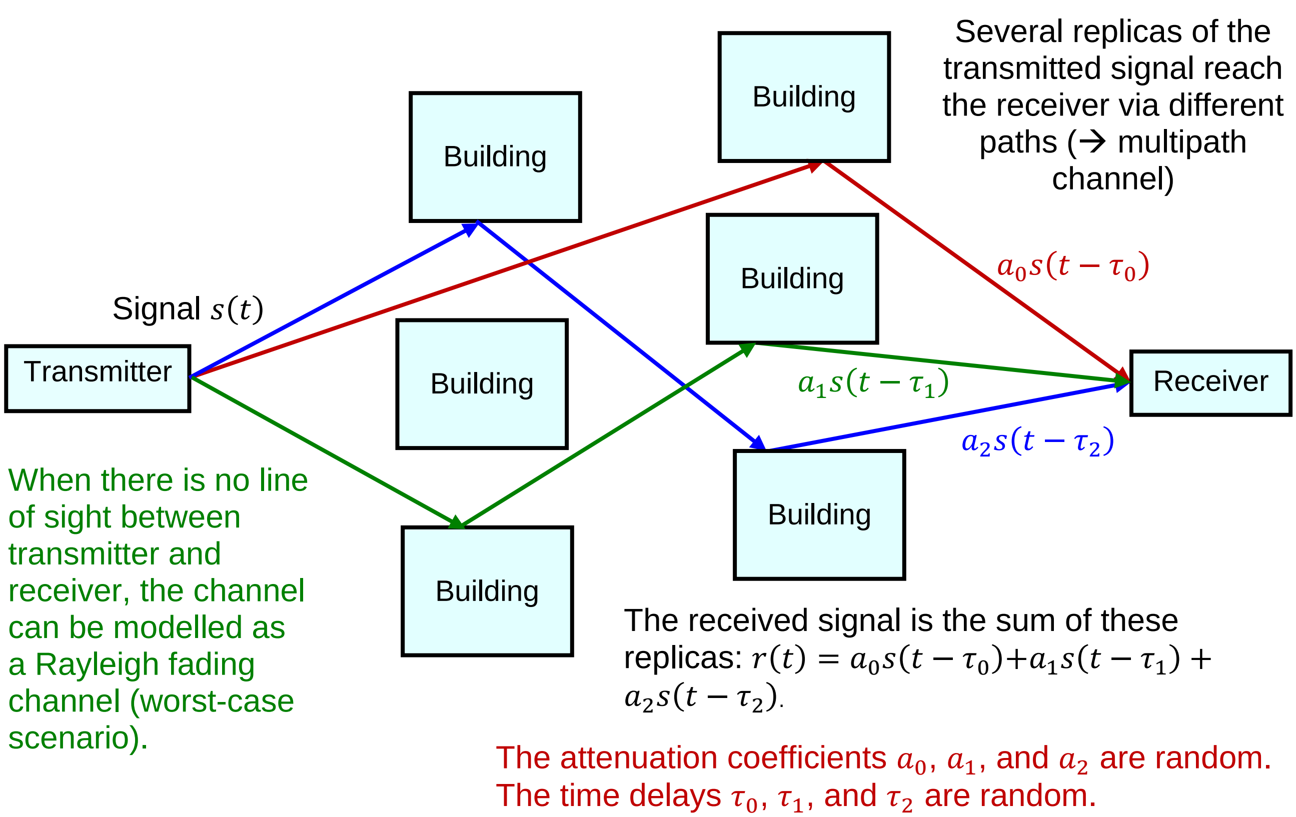
Newcastle University

Contents

10. Digital Communications over Multipath Channels

Throughout this module, we have so far exclusively focused on communications over AWGN channels. However, in the case of wireless communications, the AWGN channel is not the most realistic channel model. Instead, wireless communication channels are often more accurately modelled as fading channels.

In wireless communications, the transmitted signal reaches the receiver through several paths. As a result, the receiver must process several randomly-attenuated and randomly-delayed versions of this signal as illustrated in the figure below. The interference between these versions of the same signal at the receive antenna can be either destructive, meaning that the signal versions tend to cancel each other, or constructive, which occurs when the resulting combined signal is stronger than each version taken separately.

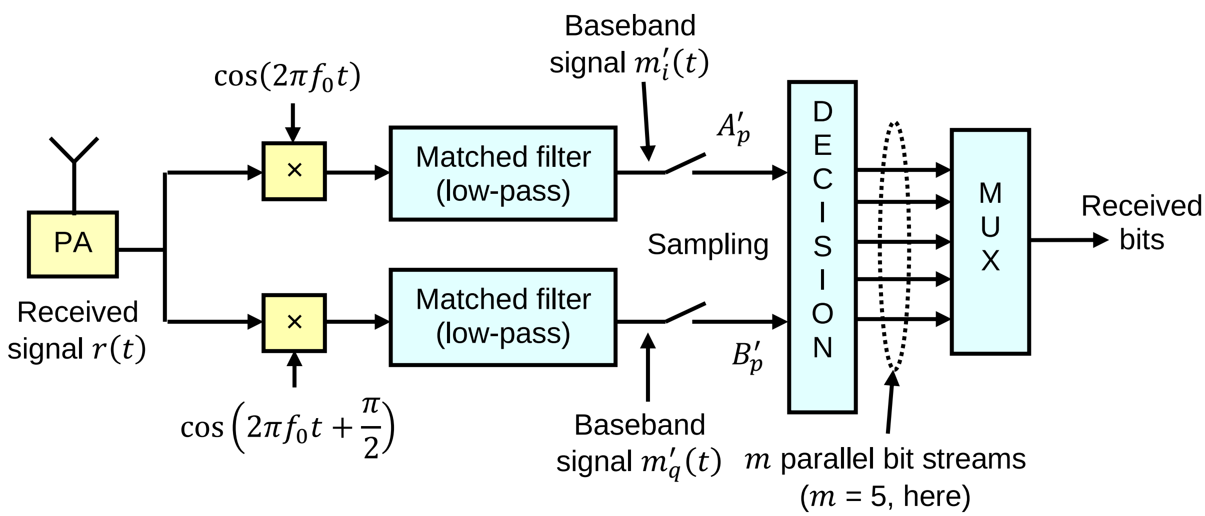


Obviously, we are more particularly worried about the destructive nature of the interference since its detrimental effect can sometimes make the signal completely undetectable by the receive antenna.

Because signal transmission happens via several paths, such channel is said to be a *multi-path channel*.

In this chapter, we are going to revisit the principles of operation of a digital communication receiver by assuming that the latter is used in conjunction with a multipath channel.

We have represented below the generic structure of a digital communication receiver.



The received signal must be amplified since the signal power received by the antenna is usually quite low.

A significant amount of noise is added to the received signal at this stage. This noise originates from the movement of electrons inside the connections and components of the radio-frequency (RF) power amplifier used at the front-end of the radio receiver. This amplifier is often referred to as a *low-noise amplifier*.

It is necessary for a low-noise amplifier to boost the desired signal power while adding as little noise and distortion as possible in this process so that the retrieval of the transmitted symbols C_k is possible in the later stages of the receiver. This is in fact at that stage that a white Gaussian noise process $n(t)$ is added to the received signal.

In radio channels, the additive white Gaussian noise (AWGN) is not the only disturbance that affects the transmitted signal $s(t)$. In fact, many wireless channels are subject to interference that is even more detrimental to communications integrity than the AWGN:

(1) Interference among different users. We talk here about phenomena such as multi-user interference and co-channel interference that are beyond the scope of a first course on digital communications.

(2) Interference among several randomly-attenuated and randomly-delayed versions of the same signal. This kind of interference is typically present in fading channels. This is precisely the topic of this chapter.

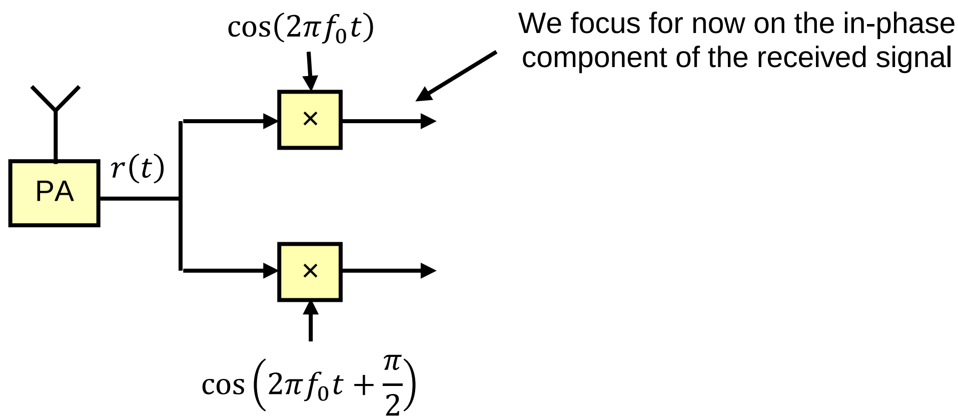
(3) Interference among successive symbols carried by the same signal. This is known as *inter-symbol interference* (ISI). This issue will also be addressed in this chapter as many fading channels are affected by ISI.

Throughout this chapter, we will assume that the channel is a multipath channel with L paths, i.e., the received signal $r(t)$ present at the low-noise amplifier output is given by

$$r(t) = \sum_{l=0}^{L-1} a_l s(t - \tau_l) + n(t),$$

where $n(t)$ is a zero-mean Gaussian random noise process with constant power spectral density (PSD), i.e., $\Phi_n(f) = \frac{N_0}{2}$ for all frequency values. The quantities a_l and τ_l represent the attenuation and the delay of the transmitted signal when the latter travels via path $l \in \{0, 1, \dots, L-1\}$.

• **Signal present at the in-phase mixer output**



First, we are going to focus on the in-phase component of the receiver.

To obtain the base-band signal $m_i'(t)$, which is an estimate of the base-band signal $m_i(t)$ present at the transmitter side, one has first to demodulate the signal $r(t)$ by mixing it with a locally-generated carrier signal $\cos(2\pi f_0 t)$.

The expression of the received signal can be written as follows:

$$r(t) = \sum_{l=0}^{L-1} a_l \sum_{k=0}^{+\infty} A_k\, h(t - kT - \tau_l) \cos\big(2\pi f_0 (t - \tau_l)\big) - B_k h(t - kT - \tau_l) \sin\big(2\pi f_0 (t - \tau_l)\big) + n(t).$$

The signal at the in-phase mixer output is then expressed as

$$r(t) \cdot \cos(2\pi f_0 t) = \sum_{l=0}^{L-1} a_l \sum_{k=0}^{+\infty} \Big[A_k \cos\big(2\pi f_0 (t - \tau_l)\big) \cos(2\pi f_0 t) - B_k \sin\big(2\pi f_0 (t - \tau_l)\big) \cos(2\pi f_0 t)\Big]\, h(t - kT - \tau_l) + n(t) \cos(2\pi f_0 t).$$

We remember that

$$\cos(a)\cos(b) = \frac{1}{2}\cdot[\cos(a+b)+\cos(a-b)]$$

and
$$\sin(a)\cos(b) = \frac{1}{2}\cdot[\sin(a+b)+\sin(a-b)].$$

We thus obtain the following expression:

$$r(t)\cdot\cos(2\pi f_0 t) = \frac{1}{2}\sum_{l=0}^{L-1}\sum_{k=0}^{+\infty} a_l\Big[A_k\cos\big(2\pi f_0(2t-\tau_l)\big) + A_k\cos(2\pi f_0\tau_l) - B_k\sin\big(2\pi f_0(2t-\tau_l)\big) + B_k\sin(2\pi f_0\tau_l)\Big]\,h(t-kT-\tau_l) \;+ n(t)\cos(2\pi f_0 t).$$

We now need to find the characteristics of the noise $n_i'(t) = n(t)\cos(2\pi f_0 t)$ present at the in-phase mixer output. We must remember that, at any time t, $n(t)$ is a random variable as $n(t)$ is a random process, whereas $\cos(2\pi f_0 t)$ is simply a number as $\cos(2\pi f_0 t)$ is a deterministic signal.

The noise process $n_i'(t)$ has the same distribution as the noise process $n(t)$ since multiplying a random variable by any number does not have any effect on the shape of its probability density function.

Recall that $n(t)$ is Gaussian with a mean equal to zero. Therefore, $n_i'(t)$ is also Gaussian with a mean m given by

$$m = E\{n_i'(t)\} = E\{n(t)\}\cdot E\{\cos(2\pi f_0 t)\} = E\{n(t)\}\cdot\cos(2\pi f_0 t) = 0.$$

We also need to know whether $n_i'(t)$ is a white noise process. One way to answer this question consists of determining its autocorrelation function.

The autocorrelation function, $\Gamma_{n_i'}(t)$, of the random process $n_i'(t)$ can be computed as follows:

$$\Gamma_{n_i'}(t) = E_\tau\{n_i'(\tau)\cdot n_i'(\tau-t)\} = E_\tau\{n(\tau)\cdot n(\tau-t)\}\cdot E_\tau\{\cos(2\pi f_0\tau)\cdot\cos(2\pi f_0[\tau-t])\}$$

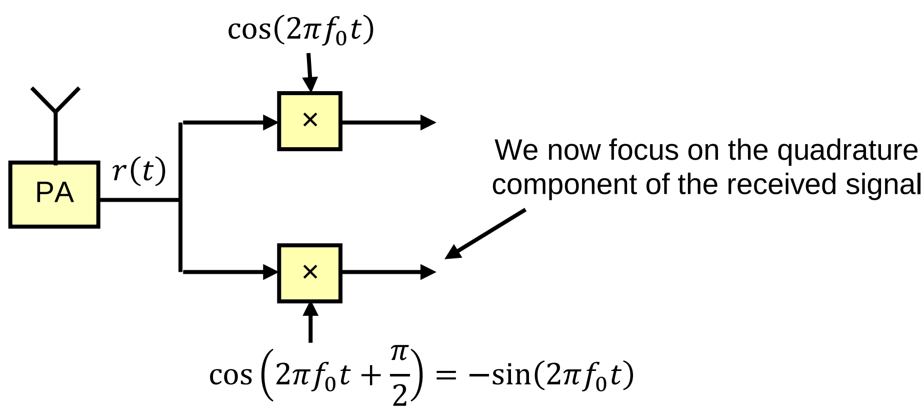
$$= \frac{1}{2}\cdot\Gamma_n(t)\cdot[E_\tau\{\cos(2\pi f_0 t)\} + E_\tau\{\cos(2\pi f_0[2\tau-t])\}] = \frac{1}{2}\cdot\Gamma_n(t)\cdot\cos(2\pi f_0 t)$$

$$= \frac{N_0}{4} \cdot \delta(t) \cdot \cos(2\pi f_0 t) = \frac{N_0}{4} \cdot \delta(t) \cdot \cos(0) = \frac{N_0}{4} \cdot \delta(t).$$

This is the autocorrelation function of a white noise process with a power spectral density $\Phi_{n_i'}(f) = \frac{N_0}{4}$ for all frequencies. Hence, $n_i'(t)$ is a white random process.

• **Signal present at the quadrature mixer output**

We are now going to focus on the quadrature component of the receiver.



To obtain the base-band signal $m_q'(t)$, which is an estimate of the base-band signal $m_q(t)$, present at the transmitter side, one has first to demodulate the signal $r(t)$ by mixing it with a locally-generated carrier signal $\cos\left(2\pi f_0 t + \frac{\pi}{2}\right) = -\sin(2\pi f_0 t)$.

Recall that the expression of the received signal is given by

$$r(t) = \sum_{l=0}^{L-1} a_l \sum_{k=0}^{+\infty} A_k\, h(t - kT - \tau_l) \cos\big(2\pi f_0 (t - \tau_l)\big) - B_k h(t - kT - \tau_l) \sin\big(2\pi f_0 (t - \tau_l)\big) + n(t).$$

The signal at the quadrature mixer output is then expressed as

$$-r(t) \cdot \sin(2\pi f_0 t) = \sum_{l=0}^{L-1} \sum_{k=0}^{+\infty} -a_l A_k\, h(t - kT - \tau_l) \cos\big(2\pi f_0 (t - \tau_l)\big) \sin(2\pi f_0 t) + a_l B_k h(t - kT - \tau_l) \sin\big(2\pi f_0 (t - \tau_l)\big) \sin(2\pi f_0 t) - n(t) \sin(2\pi f_0 t).$$

We remember that

$$\cos(a)\sin(b) = \frac{1}{2}\cdot[\sin(a+b) - \sin(a-b)]$$

and
$$\sin(a)\sin(b) = \frac{1}{2}\cdot[\cos(a-b) - \cos(a+b)].$$

We thus obtain

$$-r(t)\cdot\sin(2\pi f_0 t) = \frac{1}{2}\sum_{l=0}^{L-1}\sum_{k=0}^{+\infty} a_l\big[-A_k\sin\big(2\pi f_0(2t-\tau_l)\big) - A_k\sin(2\pi f_0\tau_l) - B_k\cos\big(2\pi f_0(2t-\tau_l)\big) + B_k\cos(2\pi f_0\tau_l)\big]\, h(t-kT-\tau_l) - n(t)\sin(2\pi f_0 t).$$

We need to find the characteristics of the noise $n'_q(t) = -n(t)\sin(2\pi f_0 t)$ present at the quadrature mixer output. We must remember that, at any time t, $n(t)$ is a random variable as $n(t)$ is a random process, whereas $\sin(2\pi f_0 t)$ is simply a number as $\sin(2\pi f_0 t)$ is a deterministic signal.

The noise process $n'_q(t)$ has the same distribution as the noise process $n(t)$ since multiplying a random variable by any number does not have any effect on the shape of its probability density function.

Recall that $n(t)$ is Gaussian with a mean equal to zero. Therefore, $n'_q(t)$ is also Gaussian with a mean m given by

$$m = E\{n'_q(t)\} = -E\{n(t)\}\cdot E\{\sin(2\pi f_0 t)\} = -E\{n(t)\}\cdot\sin(2\pi f_0 t) = 0.$$

We also need to know whether $n'_q(t)$ is a white noise process. To do so, we are going to determine its autocorrelation function.

The autocorrelation function, $\Gamma_{n'_q}(t)$, of the random process $n'_q(t) = -n(t)\sin(2\pi f_0 t)$ can be computed as follows:

$$\Gamma_{n'_q}(t) = E_\tau\{n'_q(\tau)\cdot n'_q(\tau-t)\} = E_\tau\{n(\tau)\cdot n(\tau-t)\}\cdot E_\tau\{\sin(2\pi f_0\tau)\cdot\sin(2\pi f_0[\tau-t])\}$$

$$= \frac{1}{2}\cdot\Gamma_n(t)\cdot[E_\tau\{\cos(2\pi f_0 t)\} - E_\tau\{\cos(2\pi f_0[2\tau-t])\}] = \frac{1}{2}\cdot\Gamma_n(t)\cdot\cos(2\pi f_0 t)$$

$$= \frac{N_0}{4} \cdot \delta(t) \cdot \cos(2\pi f_0 t) = \frac{N_0}{4} \cdot \delta(t) \cdot \cos(0) = \frac{N_0}{4} \cdot \delta(t).$$

This is the autocorrelation function of a white noise process with a power spectral density $\Phi_{n'_q}(f) = \frac{N_0}{4}$ for all frequencies. Hence, $n'_q(t)$ is a white random process.

- **Are $\boldsymbol{n'_i(t)}$ and $\boldsymbol{n'_q(t)}$ independent random processes?**

To answer this question, we must first realise that the random processes $n'_i(t)$ and $n'_q(t)$ are jointly Gaussian. In other words, any linear combination of them is Gaussian. It is indeed easy to see that a random process defined as

$$\alpha\, n'_i(t) + \beta\, n'_q(t) = n(t)\,[\alpha \cos(2\pi f_0 t) - \beta \sin(2\pi f_0 t)],$$

where α and β designate two arbitrary real numbers, is Gaussian.

If two jointly Gaussian random processes are uncorrelated, then they are independent. So, to prove the independence of $n'_i(t)$ and $n'_q(t)$, we now need to show that these two jointly Gaussian random processes are uncorrelated.

The correlation between the random processes $n'_i(t)$ and $n'_q(t)$ can be determined by showing that their cross-correlation function, $\Gamma_{n'_i,n'_q}(t)$, is such that

$$\Gamma_{n'_i,n'_q}(t) = E_\tau\{n'_i(\tau) \cdot n'_q(\tau - t)\} = E_\tau\{n'_i(\tau)\} \cdot E_\tau\{n'_q(\tau - t)\}.$$

As $n'_i(t)$ and $n'_q(t)$ have means equal to zero, we have $E_\tau\{n'_i(\tau)\} = E_\tau\{n'_q(\tau - t)\} = 0$ here. So, to prove that $n'_i(t)$ and $n'_q(t)$ are uncorrelated, we must show that $\Gamma_{n'_i,n'_q}(t) = 0, \forall t$.

The cross-correlation function $\Gamma_{n'_i,n'_q}(t)$ can be computed as follows:

$$\Gamma_{n'_i,n'_q}(t) = E_\tau\{n'_i(\tau) \cdot n'_q(\tau - t)\} = -E_\tau\{n(\tau) \cdot n(\tau - t)\} \cdot E_\tau\{\cos(2\pi f_0 \tau) \cdot \sin(2\pi f_0 [\tau - t])\}$$

$$= -\frac{1}{2} \cdot \Gamma_n(t) \cdot [E_\tau\{\sin(2\pi f_0 [2\tau - t])\} - E_\tau\{\sin(2\pi f_0 t)\}] = \frac{1}{2} \cdot \Gamma_n(t) \cdot \sin(2\pi f_0 t)$$

$$= \frac{N_0}{4} \cdot \delta(t) \cdot \sin(2\pi f_0 t) = \frac{N_0}{4} \cdot \delta(t) \cdot \sin(0) = 0.$$

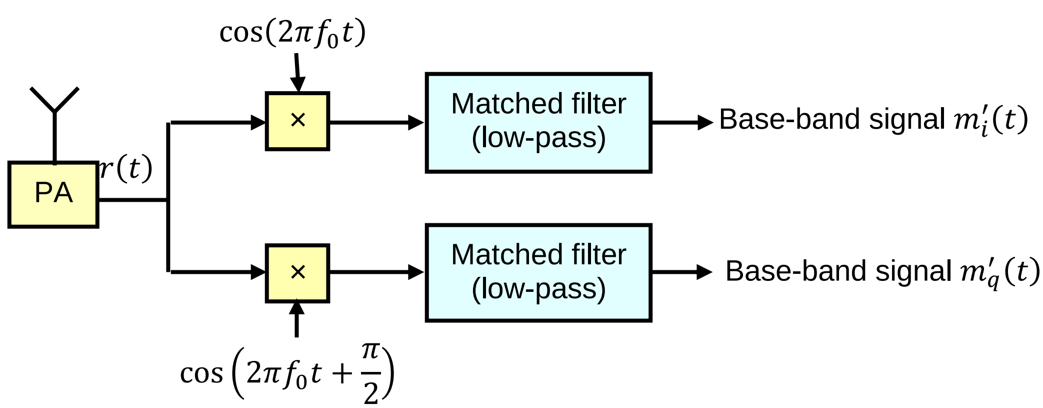
As the cross-correlation function $\Gamma_{n_i', n_q'}(t)$ is equal to zero for all values of t, we can state that the random processes $n_i'(t)$ and $n_q'(t)$ are uncorrelated. As they are also jointly Gaussian, we then conclude that they are independent.

• **Expression of the in-phase baseband signal $m_i'(t)$**

The signal present at the in-phase mixer output then goes through a low-pass matched filter with an impulse response $h(T_p - t)$ and a transfer function $H^*(f) \cdot e^{-j2\pi f T_p}$.

At the in-phase filter output, we obtain the base-band signal $m_i'(t)$ given by

$$m_i'(t) = \frac{1}{2} \sum_{l=0}^{L-1} \sum_{k=0}^{+\infty} a_l [A_k \cos(2\pi f_0 \tau_l) + B_k \sin(2\pi f_0 \tau_l)]\, h(t - kT - \tau_l) * h(T_p - t) + n_i'(t) * h(T_p - t).$$



Let us focus first on the first term in this expression. It can easily be shown that the Fourier transform of the term $h(t - kT - \tau_l) * h(T_p - t)$ is given by

$$H(f) \cdot e^{-j2\pi f(kT+\tau_l)} \cdot H^*(f) \cdot e^{-j2\pi f T_p} = H(f) \cdot H^*(f) \cdot e^{-j2\pi f(kT+\tau_l+T_p)} = G(f) \cdot e^{-j2\pi f(kT+\tau_l+T_p)},$$

with $G(f) = |H(f)|^2$.

This result is the Fourier transform of a term $g(t - kT - \tau_l - T_p)$, with $g(t) = h(t) * h(-t)$.

We then obtain

$$m_i'(t) = \frac{1}{2}\sum_{l=0}^{l=L-1}\sum_{k=0}^{+\infty} a_l[A_k\cos(2\pi f_0\tau_l) + B_k\sin(2\pi f_0\tau_l)]\, g\big(t - kT - \tau_l - T_p\big) + n_i'(t) * h\big(T_p - t\big).$$

Remarkably, this equation can be re-written in a more concise and elegant form as

$$m_i'(t) = \frac{1}{2}\sum_{l=0}^{L-1}\sum_{k=0}^{+\infty} a_l \Re\big[C_k e^{-j2\pi f_0\tau_l}\big] g\big(t - kT - \tau_l - T_p\big) + n_i'(t) * h\big(T_p - t\big),$$

which is equivalent to

$$m_i'(t) = \frac{1}{2}\sum_{k=0}^{+\infty}\sum_{l=0}^{L-1} \Re\big[C_k a_l e^{-j2\pi f_0\tau_l} g\big(t - kT - \tau_l - T_p\big)\big] + n_i'(t) * h\big(T_p - t\big),$$

which yields

$$m_i'(t) = \Re\left[\sum_{k=0}^{+\infty}\sum_{l=0}^{L-1}\frac{1}{2} C_k a_l e^{-j2\pi f_0\tau_l} g\big(t - kT - \tau_l - T_p\big)\right] + n_i''(t).$$

The fact that the pulse $h(t)$ has been replaced by another pulse $g(t)$ in the transmitted signal component merely reflects the distortion effect due to the low-pass matched filter.

Let us now determine the characteristics of the noise $n_i''(t) = n_i'(t) * h\big(T_p - t\big)$ present at the matched filter output.

The noise process $n_i''(t)$ is not white because the matched filter is also a low-pass filter. However, $n_i''(t)$ is Gaussian with zero-mean since linear filtering of a zero-mean Gaussian process produces another zero-mean Gaussian process.

The variance of $n_i''(t)$ is not infinite because $n_i''(t)$ is not a white noise process. We thus need to evaluate the variance $n_i''(t)$. To do so, one can start with the definition of the variance of a random process: $\sigma^2 = E\{[n_i''(t)]^2\} - [E\{n_i''(t)\}]^2 = E\{[n_i''(t)]^2\}$ as $E\{n_i''(t)\} = 0$.

As the term $E\{[n_i''(t)]^2\}$ is also the value of the autocorrelation function $\Gamma_{n_i''}(t)$ of $n_i''(t)$ at time $t = 0$, we have $\sigma^2 = \Gamma_{n_i''}(0)$.

We also know that $\Gamma_{n_i''}(t)$ is the inverse Fourier transform of the power spectral density $\Phi_{n_i''}(f)$. Therefore, we can write $\sigma^2 = \Gamma_{n_i''}(0) = \int_{-\infty}^{+\infty} \Phi_{n_i''}(f)df$.

We can use the expression giving the PSD of a random process at the output of a linear filter $H^*(f) \cdot e^{-j2\pi f T_p}$: $\Phi_{n_i''}(f) = \Phi_{n_i'}(f) \cdot \left|H^*(f) \cdot e^{-j2\pi f T_p}\right|^2 = \Phi_{n_i'}(f) \cdot |H(f)|^2 = \frac{N_0}{4} G(f)$.

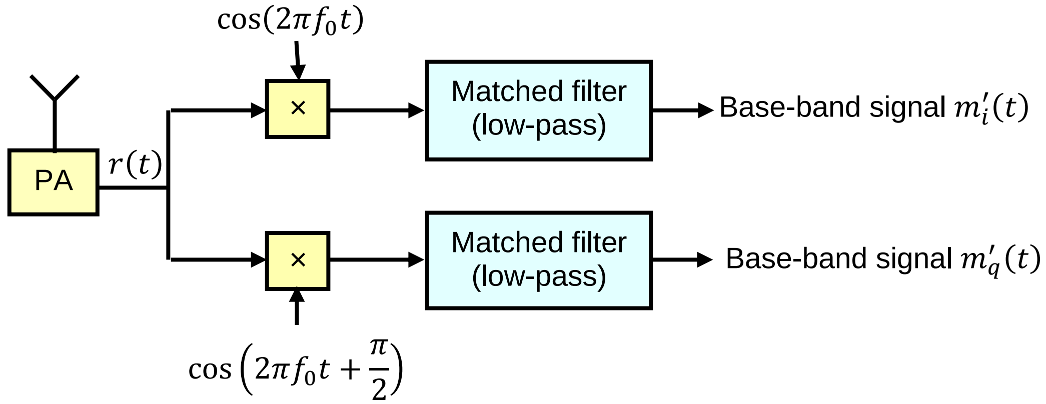
We thus have $\sigma^2 = \int_{-\infty}^{+\infty} \Phi_{n_i''}(f)df = \frac{N_0}{4} \cdot \int_{-\infty}^{+\infty} G(f)df$.

As $G(f)$ is the Fourier transform of $g(t)$, it is easy to show that the term $\int_{-\infty}^{+\infty} G(f)df$ is simply the value of $g(t)$ at time $t = 0$: $g(0) = \int_{-\infty}^{+\infty} G(f)df = E_h$.

This finally leads to the expression $\sigma^2 = \frac{N_0}{4} g(0) = \frac{N_0 E_h}{4}$, where E_h designates the energy of the pulse $h(t)$.

• **Expression of the quadrature baseband signal $m_q'(t)$**

Like its in-phase counterpart, the signal present at the quadrature mixer output goes through a low-pass matched filter with an impulse response $h(T_p - t)$ and a transfer function $H^*(f) \cdot e^{-j2\pi f T_p}$.



At the quadrature matched filter output, we obtain the base-band signal $m_q'(t)$ given by

$$m_q'(t) = \frac{1}{2}\sum_{l=0}^{L-1}\sum_{k=0}^{+\infty} a_l[-A_k\sin(2\pi f_0\tau_l) + B_k\cos(2\pi f_0\tau_l)]\, h(t-kT-\tau_l) * h\left(T_p - t\right) + n_q'(t) * h\left(T_p - t\right).$$

As for the first term in this expression, it can easily be shown that the Fourier transform of the term $h(t-kT-\tau_l) * h\left(T_p - t\right)$ is given by

$$H(f)\cdot e^{-j2\pi f(kT+\tau_l)}\cdot H^*(f)\cdot e^{-j2\pi f T_p} = H(f)\cdot H^*(f)\cdot e^{-j2\pi f\left(kT+\tau_l+T_p\right)} = G(f)\cdot e^{-j2\pi f\left(kT+\tau_l+T_p\right)},$$

with $G(f) = |H(f)|^2$.

This result is the Fourier transform of a term $g\left(t-kT-\tau_l-T_p\right)$, with $g(t) = h(t) * h(-t)$.

We then obtain

$$m_q'(t) = \frac{1}{2}\sum_{l=0}^{l=L-1}\sum_{k=0}^{+\infty} a_l[-A_k\sin(2\pi f_0\tau_l) + B_k\cos(2\pi f_0\tau_l)]\, g\left(t-kT-\tau_l-T_p\right) + n_q'(t) * h\left(T_p - t\right).$$

This equation can be re-written in a more concise and elegant form as

$$m_q'(t) = \frac{1}{2}\sum_{l=0}^{L-1}\sum_{k=0}^{+\infty} a_l\Im\left[C_k e^{-j2\pi f_0\tau_l}\right] g\left(t-kT-\tau_l-T_p\right) + n_q'(t) * h\left(T_p - t\right),$$

which is equivalent to

$$m_q'(t) = \frac{1}{2}\sum_{k=0}^{+\infty}\sum_{l=0}^{L-1}\Im\left[C_k a_l e^{-j2\pi f_0\tau_l} g\left(t-kT-\tau_l-T_p\right)\right] + n_q'(t) * h\left(T_p - t\right),$$

which yields

$$m_q'(t) = \Im\left[\sum_{k=0}^{+\infty}\sum_{l=0}^{L-1}\frac{1}{2}C_k a_l e^{-j2\pi f_0\tau_l} g\left(t-kT-\tau_l-T_p\right)\right] + n_q''(t).$$

The noise $n_q''(t) = n_q'(t) * h\left(T_p - t\right)$ present at the matched filter output is no longer white, but still Gaussian with a mean of zero since linear filtering of a zero-mean Gaussian process produces another zero-mean Gaussian process.

The variance of $n_i''(t)$ is not infinite. Instead, we can show that it is given by $\sigma^2 = \frac{N_0 E_h}{4}$. The demonstration of this result is identical to that developed in the previous section.

• **Expression of the complex baseband signal $m'(t) = m'_i(t) + jm'_q(t)$**

It is possible to combine $m'_i(t)$ and $m'_q(t)$ into a single complex baseband signal $m'(t)$ whose expression is

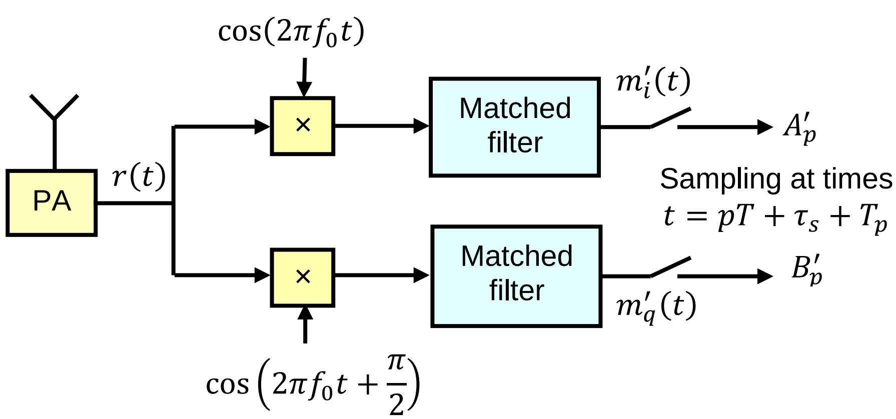
$$m'(t) = m'_i(t) + jm'_q(t) = \sum_{k=0}^{+\infty}\sum_{l=0}^{L-1}\frac{1}{2}C_k a_l e^{-j2\pi f_0 \tau_l} g\left(t - kT - \tau_l - T_p\right) + n''(t).$$

where $n''(t) = n''_i(t) + jn''_q(t)$ designates a complex zero-mean Gaussian noise process with a variance $2\sigma^2 = \frac{N_0 E_h}{2}$. This means that the variances of $n''_i(t)$ and $n''_q(t)$ are $\sigma^2 = \frac{N_0 E_h}{4}$.

From now on, we are going to adopt this notation as it is not only very elegant but also allows us to jointly process both in-phase and quadrature components.

• **Sampling of the complex baseband signal $m'(t)$**

The complex baseband signal $m'(t)$ is then sampled at times $t = pT + \tau_s + T_p$, where p is an integer, to obtain estimates $C'_p = A'_p + jB'_p$ of the complex symbols $C_p = A_p + jB_p$ that were transmitted at times $t = pT$. These sampling times have been determined by the receiver to account for the transmission delays $\tau_0, \tau_1, \ldots, \tau_{L-1}$.



Assuming, for simplicity's sake, that the delay τ_0 corresponds to the fastest path whereas the delay τ_{L-1} is associated with the slowest path, we typically have $\tau_0 \le \tau_s \le \tau_{L-1}$ because it makes sense to sample the baseband signal any time after the arrival of the fastest path and before that of the slowest path.

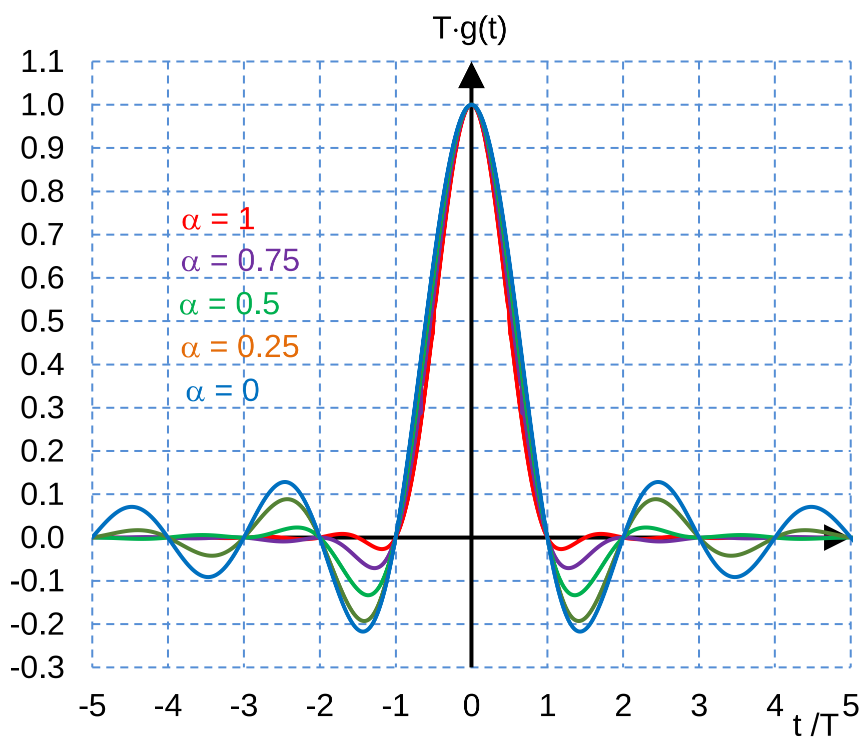
At time $t = pT + \tau_s + T_p$, we can thus write

$$C'_p = m'(pT + \tau_s + T_p) = \sum_{k=0}^{+\infty} \sum_{l=0}^{L-1} \frac{1}{2} C_k a_l e^{-j2\pi f_0 \tau_l} g\big((p-k)T + \tau_s - \tau_l\big) + n''(pT + \tau_s + T_p).$$

Now, we need to remember that the pulse $g(t)$ satisfies the Nyquist criterion, i.e., we have $g\big((p-k)T\big) = 0$ for any $p \neq k$. Typically, $g(t)$ is a raised-cosine pulse expressed as

$$g(t) = \frac{1}{T} \cdot \mathrm{sinc}\left(\frac{\pi t}{T}\right) \cdot \frac{\cos\left(\frac{\pi \alpha t}{T}\right)}{1 - \frac{4\alpha^2 t^2}{T^2}},$$

where the parameter α, called *roll-off factor*, can take any value in the range from 0 to 1.



In the expression of C'_p, we notice that the L terms $g\big((p-k)T+\tau_s-\tau_l\big)$ cannot all be equal to zero when $p \neq k$, since we have $\tau_s \neq \tau_l$ for at least $(L-1)$ paths. This implies that the multipath channel can generate inter-symbol interference (ISI) because a sample C'_p, which should only depend on the symbol C_p transmitted at time $t = pT$, is also going to be dependent on the other adjacent symbols $\dots$, C_{p-2}, C_{p-1}, C_{p+1}, C_{p+2}, $\dots$ in the transmitted sequence. This ISI is going to significantly degrade the error performance of our communication system.

• **Flat fading channel (delay spread $\boldsymbol{\Delta\tau \ll T}$)**

However, if we assume that $|\tau_s - \tau_l| \ll T$ for all paths, then we have

$$g\big((p-k)T+\tau_s-\tau_l\big) \sim g\big((p-k)T\big),\ \ \forall\, l\ \in \{0,\dots,L-1\}.$$

Under this assumption, we can thus write

$$C'_p = m'\big(pT+\tau_s+T_p\big) \sim \sum_{k=-\infty}^{+\infty} \frac{1}{2} C_k g\big((p-k)T\big) \sum_{l=0}^{L-1} a_l e^{-j2\pi f_0 \tau_l} + n''\big(pT+\tau_s+T_p\big),$$

which is equivalent to

$$C'_p \sim \frac{E_h}{2} C_p \sum_{l=0}^{L-1} a_l e^{-j2\pi f_0 \tau_l} + n''\big(pT+\tau_s+T_p\big).$$

In this case, the ISI can be considered as negligible since the sample C'_p almost only depends on the transmitted symbol C_p.

The L conditions $|\tau_s - \tau_l| \ll T$, $\forall\, l\ \in \{0,\dots,L-1\}$, can be combined in a single condition: the delay spread, $\Delta\tau$, defined as the time difference between the slowest path (corresponding here to $l = L-1$) and the fastest path (corresponding here to $l = 0$), is much smaller than the symbol period T, i.e., $\Delta\tau = \tau_{L-1} - \tau_0 \ll T$.

A multipath channel for which the delay spread is such that $\Delta\tau \ll T$ is said to be *non-frequency-selective* or, more generally, *flat*. Later, we will see that the only effect of the channel in this case, apart from the addition of the usual Gaussian noise, is to randomly amplify or attenuate the transmitted symbols. Hence, when $\Delta\tau \ll T$, the channel does not generate any linear filtering effect on the transmitted signal $s(t)$, i.e., does not filter out any particular frequency components from $s(t)$.

As a summary, for a flat multi-path channel, the sample $C_p^{'}$ obtained at time $t = pT + \tau_s + T_p$ is given by

$$C_p^{'} = m'(pT + \tau_s + T_p) \sim \frac{E_h}{2} C_p \sum_{l=0}^{L-1} a_l e^{-j2\pi f_0 \tau_l} + n_p,$$

where n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{N_0 E_h}{2}$, meaning that the variances of its real and imaginary parts are $\sigma^2 = \frac{N_0 E_h}{4}$.

Example --

Consider transmission over a two-path wireless channel with $\tau_0 = 0$ and $\tau_1 = \Delta\tau$, where $\Delta\tau$ clearly denotes the delay spread of this multi-path channel.

The expression of the complex baseband signal $m'(t)$ generated by the matched filters is given by

$$m'(t) = \sum_{k=0}^{+\infty} \sum_{l=0}^{1} \frac{1}{2} C_k a_l e^{-j2\pi f_0 \tau_l} g(t - kT - \tau_l - T_p) + n''(t).$$

This equation can be developed as

$$m'(t) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k a_0 e^{-j2\pi f_0 \tau_0} g(t - kT - \tau_0 - T_p) + \sum_{k=0}^{+\infty} \frac{1}{2} C_k a_1 e^{-j2\pi f_0 \tau_1} g(t - kT - \tau_1 - T_p) + n''(t).$$

By introducing the notations $\overline{a_0} = a_0 e^{-j2\pi f_0 \tau_0}$ and $\overline{a_1} = a_1 e^{-j2\pi f_0 \tau_1}$ to lighten the expression of $m'(t)$ and remembering that $\tau_0 = 0$ and $\tau_1 = \Delta\tau$, we obtain

$$m'(t) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k \overline{a_0} g(t - kT - T_p) + \sum_{k=0}^{+\infty} \frac{1}{2} C_k \overline{a_1} g(t - kT - \Delta\tau - T_p) + n''(t).$$

By sampling this baseband signal at time $t = pT + \tau_s + T_p$, with $\tau_s = \tau_0 = 0$, we obtain an estimate $C_p^{'}$ of the transmitted symbol C_p:

$$C_p^{'} = m'\left(pT + \tau_s + T_p\right) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k \overline{a_0} g\left((p-k)T\right) + \sum_{k=0}^{+\infty} \frac{1}{2} C_k \overline{a_1} g\left((p-k)T - \Delta\tau\right) + n''\left(pT + \tau_s + T_p\right).$$

Using the fact that $g(t)$ is a raised-cosine pulse, we obtain

$$C_p^{'} = \frac{1}{2} C_p \overline{a_0} g(0) + \sum_{k=0}^{+\infty} \frac{1}{2} C_k \overline{a_1} g\left((p-k)T - \Delta\tau\right) + n_p,$$

which can be further developed as

$$C_p^{'} = \frac{1}{2} C_p\left(\overline{a_0} g(0) + \overline{a_1} g(-\Delta\tau)\right) + \cdots + \frac{1}{2} C_{p-2} \overline{a_1} g(2T - \Delta\tau) + \frac{1}{2} C_{p-1} \overline{a_1} g(T - \Delta\tau) + \frac{1}{2} C_{p+1} \overline{a_1} g(-T - \Delta\tau) + \frac{1}{2} C_{p+2} \overline{a_1} g(-2T - \Delta\tau) + \cdots + n_p.$$

This expression shows that the sample $C_p^{'}$ depends on the symbol C_p, as it should be, but also on the other transmitted symbols ... $C_{p-2}, C_{p-1}, C_{p+1}, C_{p+2}$ ... Such inter-symbol interference (ISI) can be very detrimental to the transmission reliability.

To remove (almost all of) this ISI, we must ensure that $g\left((p-k)T - \Delta\tau\right) \sim 0$ for $p \neq k$. In other words, we must have ... $g(2T - \Delta\tau) \sim g(T - \Delta\tau) \sim g(-T - \Delta\tau) \sim g(-2T - \Delta\tau) \sim$ .... ~ 0. All these conditions can be jointly satisfied if the delay spread is such that $\Delta\tau \ll T$.

Provided that $\Delta\tau \ll T$, we can finally write $C_p^{'} \sim \frac{1}{2} C_p g(0)(\overline{a_0} + \overline{a_1}) + n_p$, which indicates the absence of ISI in this case.

Each term $\bar{a_l} = a_l e^{-j2\pi f_0 \tau_l}$, $l \in \{0, \dots, L-1\}$, represents the outcome of a complex random variable whose probability density function is unknown. Recall that the quantities a_l and τ_l represent the attenuation and the delay of the transmitted signal when the latter travels via path $l \in \{0, 1, \dots, L-1\}$.

By applying the *central-limit theorem*, we can state that the sum $\sum_{l=0}^{L-1} \bar{a_l}$ is a Gaussian sample, even though each term $\bar{a_l}$ does not necessarily follow a Gaussian distribution. This result is valid provided that all individual terms $\overline{a_l}$, $l \in \{0, \dots, L-1\}$, are independent and follow the same distribution, and the number L of paths is assumed to be sufficiently large. All these are realistic assumptions in radio channels.

• Flat fading channel model ($\Delta\tau \ll T$)

We can thus write that, for a flat fading channel, the sample $C_p^{'}$ obtained by sampling the baseband signal at time $t = pT + \tau_s + T_p$ is given by

$$C_p^{'} = w_p \frac{E_h}{2} C_p + n_p,$$

where $w_p = w_{p,1} + jw_{p,2} = \sum_{l=0}^{L-1} \bar{\alpha}_l$ denotes a complex Gaussian sample with a mean and a variance that will be specified later, whereas $n_p = n_{p,1} + jn_{p,2}$ is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{N_0}{2} g(0) = \frac{N_0 E_h}{2}$.

We can see that the symbol C_p transmitted at time $t = pT$ is affected by both an additive noise component n_p as well as a multiplicative noise component w_p. The fact that each symbol C_p is multiplied, and thus possibly attenuated, by a random number w_p is the reason why such a channel is referred to as a *flat fading channel*.

If we divide the expression of the sample $C_p^{'}$ by a factor $\frac{2}{E_h}$ for simplicity's sake, we obtain the following expression:

$$C_p^{'} = w_p C_p + n_p,$$

where n_p is now a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \left(\frac{2}{E_h}\right)^2 \frac{N_0 E_h}{2} = \frac{2N_0}{E_h}$.

We are now going to introduce the signal-to-noise ratio (SNR) in this equation to make the latter more informative. Recall that, for an AWGN channel, we previously showed that the average SNR per symbol, $\frac{E_s}{N_0}$, is given by

$$\frac{E_s}{N_0} = \frac{\gamma}{2} \frac{E_h}{N_0},$$

whereas the average SNR per information bit, $\frac{E_b}{N_0}$, is given by

$$\frac{E_b}{N_0} = \frac{1}{m}\frac{E_s}{N_0} = \frac{\gamma}{2m}\frac{E_h}{N_0},$$

where γ is a constellation parameter defined as $\gamma = E_{C_p}\left\{\left|C_p\right|^2\right\}$ and m designates the number of bits per symbol.

Over a flat fading channel, we must use the same expressions as above for the average SNRs because the effect of multi-path propagation on the received signal component is neutral *on average*. In other words, the fading effect is not supposed to reduce or increase the *average* energy, E_s, of the received signal component present at the power amplifier output. This average energy E_s must thus be identical to that of an equivalent AWGN channel.

Therefore, we can now write

$$C_p^{'} = w_p C_p + n_p,$$

where n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{2N_0}{E_h} = \gamma\left(\frac{E_s}{N_0}\right)^{-1} = \frac{\gamma}{m}\left(\frac{E_b}{N_0}\right)^{-1}$.

Note that, hereafter, we will generally use the SNR per information bit, $\frac{E_b}{N_0}$, rather than the SNR per symbol as it is the most common form of SNR employed in communication engineering.

It is often preferable to adopt a polar notation for the fading sample w_p. We thus obtain

$$C_p^{'} = h_p e^{j\theta_p} C_p + n_p,$$

where $h_p = \sqrt{w_{p,1}^2 + w_{p,2}^2}$ and $\theta_p = \tan^{-1}\left(\frac{w_{p,2}}{w_{p,1}}\right)$ designate the magnitude and argument (phase) of w_p, respectively.

This expression clearly shows the effect of fading: It does rotate the constellation by a random angle θ_p and, at the same time, compresses or expands it by a random coefficient h_p.

The magnitude of the complex symbol C_p, which represents the signal component in the expression of $C_p^{'}$, is multiplied by a random fading sample h_p. This implies that the average SNR is in fact multiplied by a random coefficient h_p^2.

The case "$h_p > 1$" corresponds to a constellation expansion, which has a beneficial effect on the transmission reliability as it is equivalent to an increase in SNR by a factor $h_p^2 > 1$.

On the other hand, the case "$h_p < 1$" corresponds to a constellation compression, which has a detrimental effect on the transmission reliability. In fact, it is equivalent to an SNR reduction since this SNR is multiplied by a factor $h_p^2 < 1$. With fading values close to zero ($h_p \sim 0$), the performance degradation can even become very severe as the samples $C_p^{'}$ are then almost entirely made up of noise ($C_p^{'} \sim n_p$). In this case, the multipath channel is said to be in a state of *deep fade*.

The big problem with the flat fading channel is that it rotates and expands or compresses the constellation in a random fashion. In such case, it is impossible to define suitable boundary lines for the detection of the transmitted symbols C_p. As a result, the communication system is simply not going to work properly.

• **Flat fading channel model with knowledge of the rotation angle θ_p**

For a few simple constellations, such as BPSK and QPSK, the boundary line issue can be fixed by only providing to the receiver an estimate of the rotation angle θ_p. For these constellations, knowledge of the magnitude h_p is not required at the receiver side.

We could show that the argument θ_p is uniformly distributed between 0 and 2π. In general, provided that the fading channel varies slowly with time, the receiver is able to estimate the value of this angle with sufficient accuracy, and hence compensate for it by multiplying the sample $C_p^{'}$ by $e^{-j\theta_p}$. Thus, by keeping the same notation as before, assuming perfect knowledge of the rotation angle at the receiver side, the channel model can be further simplified to

$$C_p^{'} = h_p C_p + n_p,$$

where the new additive Gaussian noise sample n_p has been obtained by rotating the previous Gaussian noise sample by $e^{-j\theta_p}$, which does not change anything to its mean and variance.

• Rotating a complex zero-mean Gaussian noise sample results in another complex zero-mean Gaussian noise with the same variance.

Let us demonstrate that rotating a complex zero-mean Gaussian noise sample by an arbitrary angle θ_p does result in another complex zero-mean Gaussian noise sample with identical variance.

Assume that $n_p = n_{p,1} + j n_{p,2}$ is a complex zero-mean Gaussian noise sample with variance $2\sigma^2$. Also, assume, that the real and imaginary parts of that noise are independent, meaning that we can write $E\{n_{p,1} n_{p,2}\} = E\{n_{p,1}\} E\{n_{p,2}\}$.

The new rotated noise sample is given by $n_p e^{-j\theta_p} = [n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p] + j[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p]$. What are the characteristics of that complex noise sample?

First, let us prove that $[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p]$ and $[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p]$ are independent.

To do so, we first need to demonstrate that they are jointly Gaussian. In other words, we must show that any linear combination of them is Gaussian. It is easy to see that the quantity

$$\alpha\,[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p] + \beta\,[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p]$$
$$= n_{p,1}[\alpha\cos\theta_p - \beta\sin\theta_p] + n_{p,2}[\alpha\sin\theta_p + \beta\cos\theta_p],$$

where α and β designate two arbitrary real numbers, is Gaussian as the sum of the two independent Gaussian samples $n_{p,1}[\alpha\cos\theta_p - \beta\sin\theta_p]$ and $n_{p,2}[\alpha\sin\theta_p + \beta\cos\theta_p]$ is also Gaussian.

The mean of $\left[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right]$ is given by $E\left\{n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right\} = E\left\{n_{p,1}\right\}\cos\theta_p + E\left\{n_{p,2}\right\}\sin\theta_p = 0$, as $E\left\{n_{p,1}\right\} = E\left\{n_{p,2}\right\} = 0$. In the same way, we can easily show that the mean of $\left[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p\right]$ is also equal to zero.

We know that, if two jointly Gaussian random variables are uncorrelated, then they are independent. So, to prove the independence of $\left[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right]$ and $\left[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p\right]$, we now need to show that these two jointly Gaussian samples are uncorrelated.

The correlation between these samples, which have been shown to have a mean equal to zero, can be determined by evaluating their cross-correlation as follows:

$$E\left\{\left[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right]\cdot\left[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p\right]\right\}$$
$$= -E\left\{n_{p,1}^2\right\}\cos\theta_p\,\sin\theta_p + E\left\{n_{p,2}^2\right\}\cos\theta_p\,\sin\theta_p$$
$$= -\sigma^2\cos\theta_p\,\sin\theta_p + \sigma^2\cos\theta_p\,\sin\theta_p = 0.$$

The random processes $\left[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right]$ and $\left[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p\right]$ are uncorrelated and jointly Gaussian. They are therefore independent.

The variance of $\left[n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right]$ is given by

$$E\left\{\left(n_{p,1}\cos\theta_p + n_{p,2}\sin\theta_p\right)^2\right\}$$
$$= E\left\{n_{p,1}^2\right\}\cos^2\theta_p + E\left\{n_{p,2}^2\right\}\sin^2\theta_p + 2\,E\left\{n_{p,1}\right\}E\left\{n_{p,2}\right\}\cos\theta_p\,\sin\theta_p$$
$$= \sigma^2\left(\cos^2\theta_p + \sin^2\theta_p\right) = \sigma^2.$$

This is the same variance as the samples $n_{p,1}$ and $n_{p,2}$. In the same way, we can demonstrate that the variance of $\left[-n_{p,1}\sin\theta_p + n_{p,2}\cos\theta_p\right]$ is also equal to σ^2.

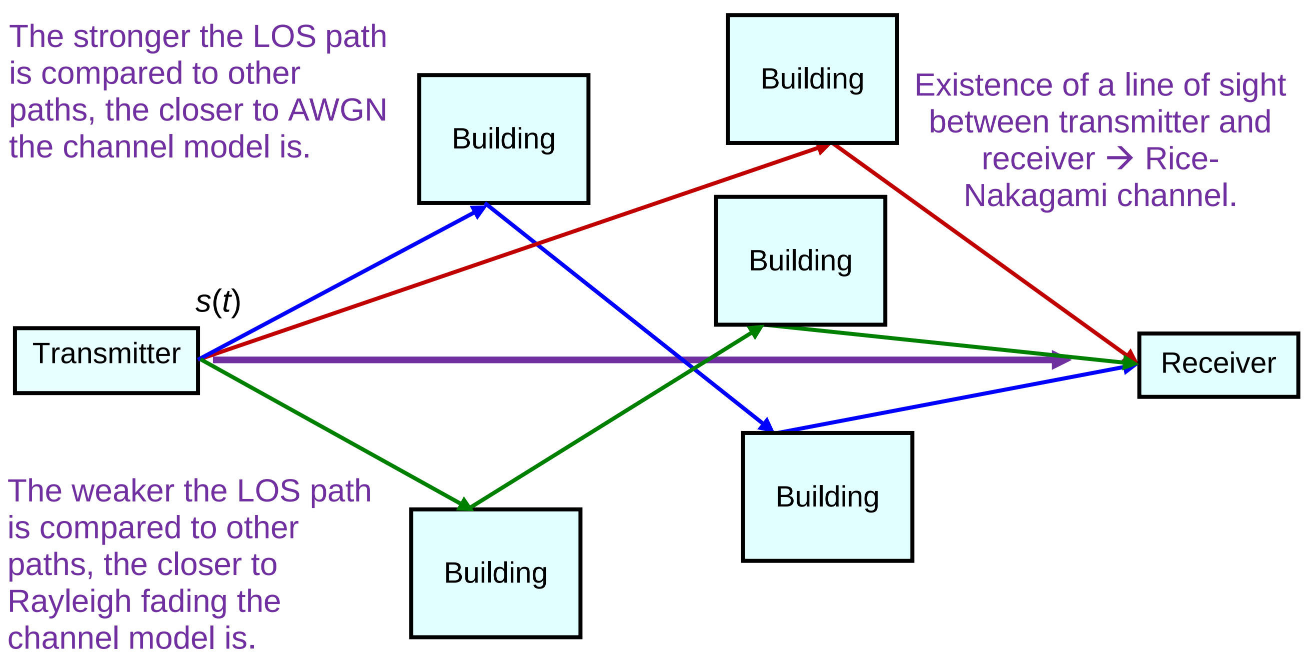
From now on, unless specified otherwise, we are always going to assume that the receiver is able to compensate for the constellation rotation θ_p. In the context of a flat fading channel, the estimate $C_p^{'}$ of a transmitted symbol C_p will thus be given by

$$C_p^{'} = h_p C_p + n_p,$$

where n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \gamma\left(\frac{E_S}{N_0}\right)^{-1} = \frac{\gamma}{m}\left(\frac{E_b}{N_0}\right)^{-1}$.

• Rice-Nakagami model for flat fading channels (Rice-Nakagami flat fading channel)

If one of the paths is significantly stronger than the others, which typically happens when there is a line-of-sight between transmitter and receiver, the channel can be modelled as a Rice-Nakagami channel, meaning that each fading sample h_p follows a Rice-Nakagami distribution.



When there is a line of sight between transmitter and receiver, we have $E\{w_{p,1}\} = m \neq 0$ and $E\{w_{p,2}\} = 0$. In this case, the fading sample h_p follows a Rice-Nakagami distribution given by

$$P_{h_p}(x) = \frac{x}{\sigma_p^2} \cdot e^{-\frac{x^2+m^2}{2\sigma_p^2}} \cdot I_0\left(\frac{mx}{\sigma_p^2}\right) \text{ for } x > 0,$$

and $$P_{h_p}(x) = 0 \text{ for } x < 0.$$

In this expression, the quantity σ_p^2 denotes the variance of the Gaussian samples $w_{p,1}$ and $w_{p,2}$. The function $I_0(\cdot)$ is the zero-order Bessel function of the first kind.

The fading channel must always be normalized so that $E\{h_p^2\} = 1$ for reasons that will be clarified later. This implies that the variance σ_p^2 is given by $\sigma_p^2 = \frac{1-m^2}{2}$ because we have $E\{h_p^2\} = E\{w_{p,1}^2 + w_{p,2}^2\} = E\{w_{p,1}^2\} + E\{w_{p,2}^2\} = \sigma_p^2 + m^2 + \sigma_p^2 = 2\sigma_p^2 + m^2$.

The normalized probability density function of the fading sample is thus given by

$$P_{h_p}(x) = \frac{2x}{1-m^2} \cdot e^{-\frac{x^2+m^2}{1-m^2}} \cdot I_0\left(\frac{2mx}{1-m^2}\right) \text{ for } x > 0,$$

and $$P_{h_p}(x) = 0 \text{ for } x < 0.$$

• **Rayleigh model for flat fading channels (Rayleigh flat fading channel)**

However, when no path dominates the other ones, the channel is then modelled as a Rayleigh fading channel, i.e., each fading sample h_p follows a Rayleigh distribution. The Rayleigh fading model corresponds to a worst-case scenario in the sense that there is no line of sight between transmitter and receiver. This is the channel model that is the most often considered by communication engineers.

When there is no line of sight between transmitter and receiver, we have $E\{w_{p,1}\} = E\{w_{p,2}\} = 0$. In this case, the fading sample h_p follows a Rayleigh distribution given by

$$P_{h_p}(x) = \frac{x}{\sigma_p^2} \cdot e^{-\frac{x^2}{2\sigma_p^2}} \text{ for } x > 0,$$

and $$P_{h_p}(x) = 0 \text{ for } x < 0,$$

where the quantity σ_p^2 denotes the variance of the Gaussian samples $w_{p,1}$ and $w_{p,2}$.

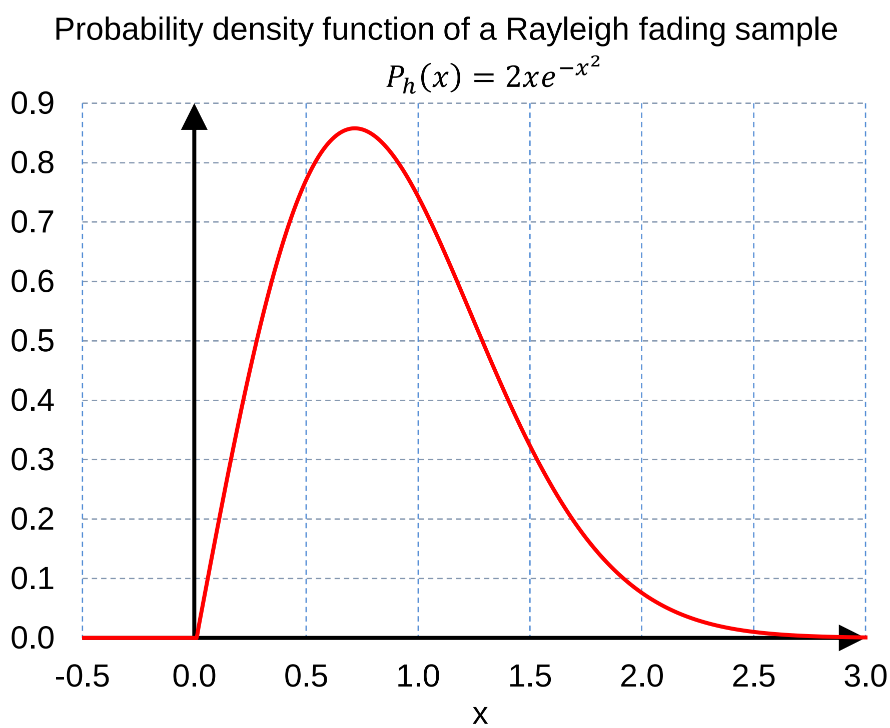
The fading channel must always be normalized so that $E\{h_p^2\} = 1$. This implies that the variance σ_p^2 is given by $\sigma_p^2 = \frac{1}{2}$ because we have $E\{h_p^2\} = E\{w_{p,1}^2 + w_{p,2}^2\} = E\{w_{p,1}^2\} + E\{w_{p,2}^2\} = \sigma_p^2 + \sigma_p^2 = 2\sigma_p^2$.

The normalized probability density function of the fading sample is thus given by

$$P_{h_p}(x) = 2x\, e^{-x^2} \text{ for } x > 0,$$

and

$$P_{h_p}(x) = 0 \text{ for } x < 0.$$



• Flat fading channels with perfect channel state information (CSI)

We have seen that, if the receiver has the knowledge of the rotation angle θ_p, a flat fading channel can be modelled using the following equation:

$$C_p' = h_p C_p + n_p,$$

where h_p is a fading sample that follows either a Rice-Nakagami or a Rayleigh distribution, whereas n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \gamma\left(\frac{E_s}{N_0}\right)^{-1} = \frac{\gamma}{m}\left(\frac{E_b}{N_0}\right)^{-1}$.

For all constellations, excepted the simplest ones such as BPSK and QPSK, the presence of the multiplicative noise h_p is going to prevent a communication system from working properly because it affects the boundary lines used for symbol recovery in an unpredictable way.

This is why it is crucial to provide the receiver with the knowledge of the fading sample value so that the system can compensate for the amplification or attenuation effect due to this fading sample.

We generally assume that the receiver knows the exact values of the successive samples h_p. In such case, the communication system is said to operate with *perfect channel state information* (CSI). This assumption is not as unrealistic as it seems because there are practical techniques (*channel estimation* algorithms) that can be used to accurately calculate in real time the values of samples h_p.

In general, the communication scheme cannot work properly without the knowledge of h_p at the receiver side, thus implying that CSI is a must in many practical applications.

Let us assume that the value of h_p is known at the receiver side.

We can then normalize the received sample $C_p' = h_p C_p + n_p$ by dividing it by h_p. We thus obtain a new channel sample that is still called C_p' but now expressed as

$$C_p' = C_p + n_p$$

where $n_p = n_{p,1} + jn_{p,2}$ is now a zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{\gamma}{h_p^2 m}\left(\frac{E_b}{N_0}\right)^{-1} = \frac{\gamma}{m}\left(h_p^2\frac{E_b}{N_0}\right)^{-1}$.

This new equation is identical to that of an AWGN channel and indicates that, at the scale of one particular symbol C_p only, the flat fading channel with perfect CSI can be viewed as an AWGN channel for which the SNR is $h_p^2 \frac{E_b}{N_0}$.

Note that this statement is actually true whether there is perfect CSI or not. The only difference that CSI makes, and it is a big one in practice, is that, without CSI, the receiver does not know the value of the SNR $h_p^2 \frac{E_b}{N_0}$ allocated to the received sample, while, with perfect CSI, the receiver knows it and can thus exploit this knowledge to improve transmission reliability. The quantity $h_p^2 \frac{E_b}{N_0}$ is often referred to as the instantaneous SNR, and is obviously different from the average SNR $\frac{E_b}{N_0}$.

However, the expression $C_p^{'} = C_p + n_p$ does not mean that we have managed to convert our flat fading channel into an AWGN channel by using CSI.

In an AWGN channel, the SNR $\frac{E_b}{N_0}$ is a constant for all received estimates $C_p^{'}$, which does provide a certain degree of predictability in the reliability of the channel estimates $C_p^{'}$. For instance, operating at high SNR over an AWGN channel is going to guarantee reliable communications at (almost) all times.

The situation is very different in a flat fading channel because, in such channel, the SNR fluctuates with time in a random fashion, depending on the value of h_p^2. The communication system may operate with a high average SNR per information bit, $\frac{E_b}{N_0}$, but a particular symbol may still see an instantaneous SNR, $h_p^2 \frac{E_b}{N_0}$, that is much lower than $\frac{E_b}{N_0}$. When this happens, the transmission reliability is going to be greatly degraded, in spite of the high average SNR, because the bit error probability is going to be unacceptably high as long as the instantaneous SNR remains low.

In any case, a fading channel must be defined so that the fading effect does not affect the average SNR value in our communication system. The fading samples make the SNR fluctuate around its average value but are not supposed to amplify or reduce this average value. That is why we must make sure to have $E\left\{h_p^2\frac{E_b}{N_0}\right\} = E\{h_p^2\}\frac{E_b}{N_0} = \frac{E_b}{N_0}$, thus implying that $E\{h_p^2\} = 1$.

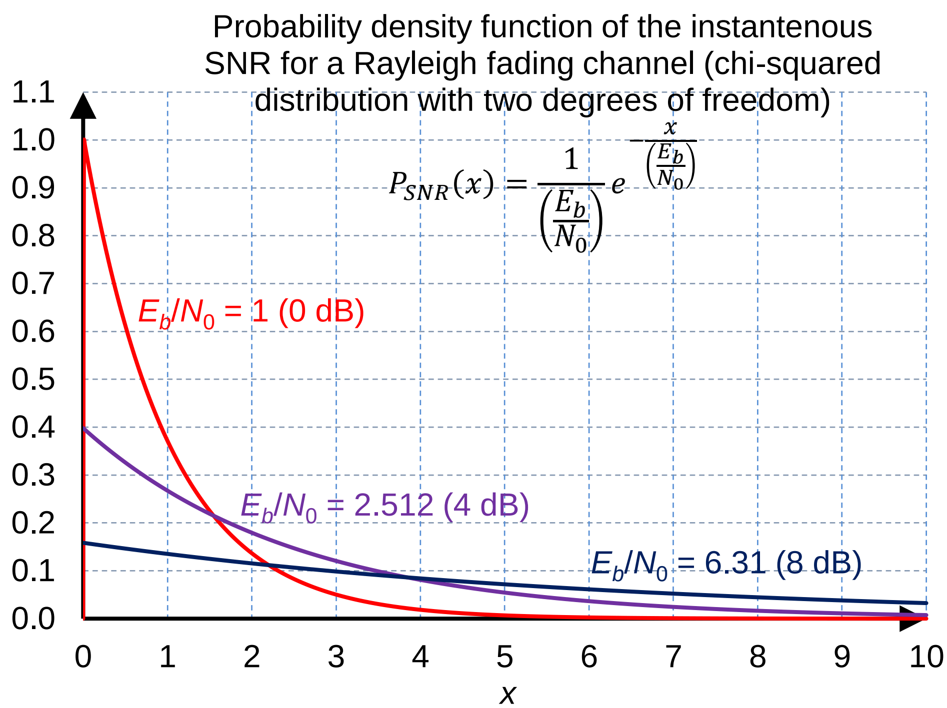
Recall that, to generate a normalized Rayleigh fading sample h_p, one must first generate two independent zero-mean Gaussian samples $w_{p,1}$ and $w_{p,2}$ with a variance $\sigma_p^2 = \frac{1}{2}$, and then use the following equation:

$$h_p = \sqrt{w_{p,1}^2 + w_{p,2}^2}.$$

When the Rayleigh distribution is normalized, we could show that the probability density function of the instantaneous SNR per information bit, $h_p^2\frac{E_b}{N_0} = \left(w_{p,1}^2 + w_{p,2}^2\right)\frac{E_b}{N_0}$, is a chi-squared distribution with two degrees of freedom that is expressed as

$$P_{SNR}(x) = \frac{1}{\left(\frac{E_b}{N_0}\right)} e^{-\frac{x}{\left(\frac{E_b}{N_0}\right)}} = \left(\frac{E_b}{N_0}\right)^{-1} e^{-x\left(\frac{E_b}{N_0}\right)^{-1}},$$

where $\frac{E_b}{N_0}$ designates the average SNR per information bit.



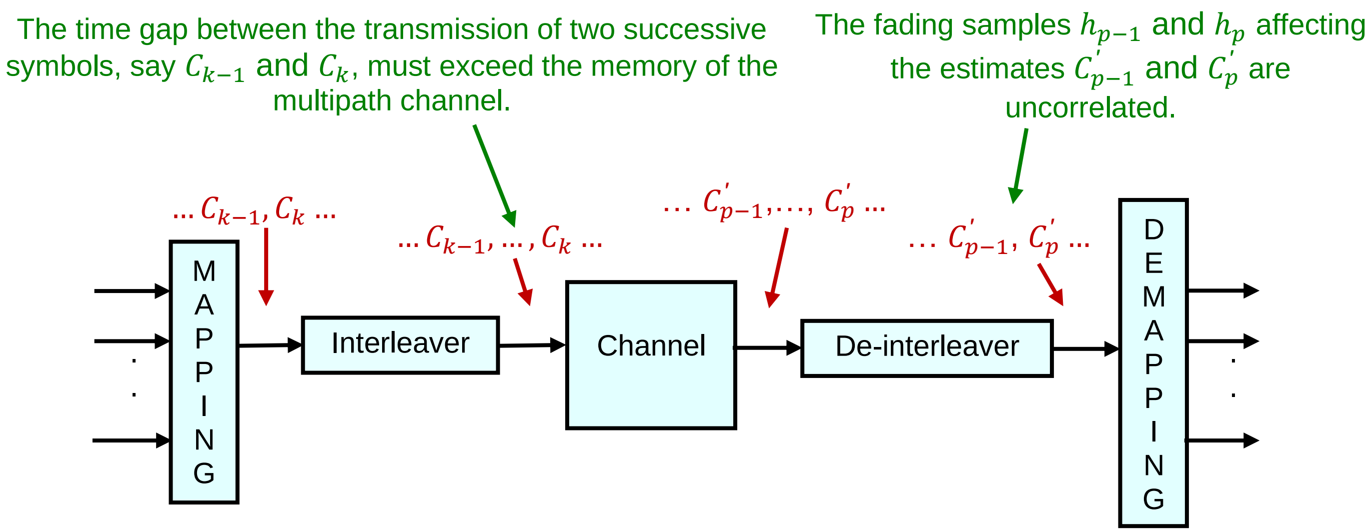
- **Memoryless flat fading channels**

The successive fading samples … $h_{p-2}, h_{p-1}, h_p, h_{p+1}, h_{p+2}$ … are generally strongly correlated because, in most cases, the multipath channel does not change significantly during a symbol interval T. In fact, many multipath channels tend to vary slowly with time. Therefore, unlike AWGN channels, fading channels are inherently not memoryless.

The memory present in fading channels can constitute a major issue in practice since it implies that the channel may enter a state of deep fade and stay in it for a long time. When it happens, several consecutive frames of transmitted symbols can be lost because the instantaneous SNR remains too low for too long.

A straightforward solution to this issue would be to avoid having significant correlation between successive symbols. In this way, it is extremely unlikely that successive symbols in any frame would all be affected by a deep fade during transmission.

In many applications, a fading channel can be made memoryless by inserting an interleaving function in the communication system, as shown in the figure below.

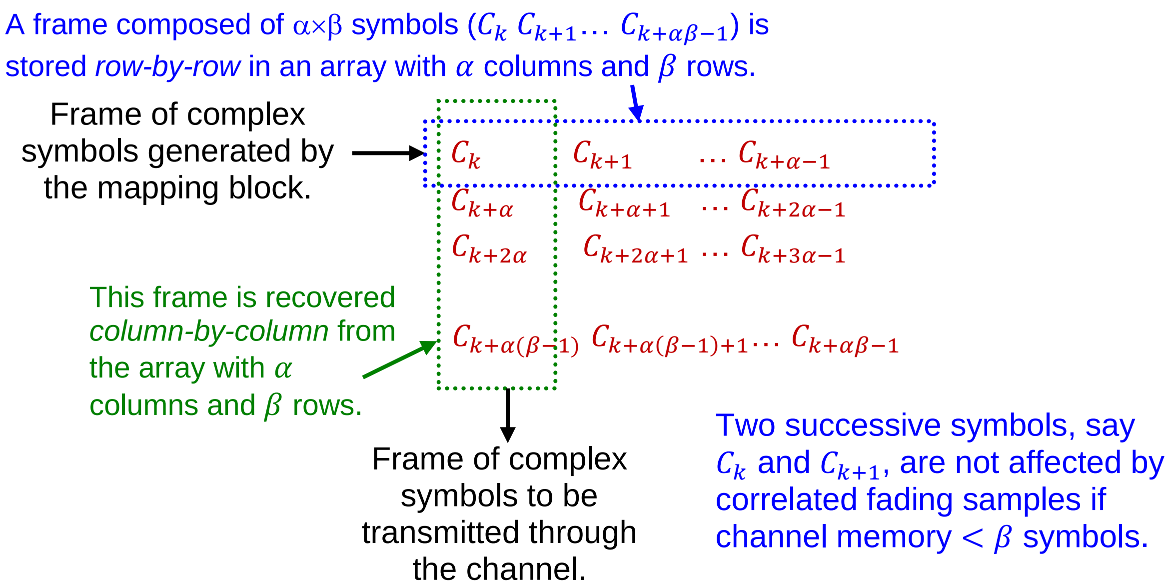


The interleaving technique consists of *scrambling* the sequence of successive symbols before transmitting it. If the size of the interleaving function is sufficiently large, then two successive symbols, say C_{k-1} and C_k, can in fact be transmitted at very different times and thus be affected by independent fading samples despite the inherent memory of the channel.

The benefit of making a fading channel memoryless is that it becomes extremely unlikely for an eventual deep fade to affect many successive received samples. The error-correction capabilities of an eventual channel decoder can then be optimized.

The figure below shows how an interleaving function can be implemented to make a multipath channel memoryless. Frames of symbols are stored row-by-row in a two-dimensional (2-D) array. The symbols are then recovered column-by-column for transmission through the multipath channel.

Assume that the channel memory does not exceed the length of a column. As two successive symbols are always transmitted in separate columns, this assumption implies there cannot be any correlation between the fading samples affecting two successive symbols. The multipath channel can thus be considered as memoryless.



If a multipath channel changes very slowly, then the 2-D array used for interleaving needs to be very large to guarantee independence between received samples. Hence, the use of interleaving may not be a practical option in some applications for which the transmission delay or frame size are limited.

If it is impractical to make the fading channel memoryless, communication engineers must then adopt other strategies to cope with the memory present in the channel. In fact, when the fading sample h_p remains constant for a long time, the fading channel can be seen as an AWGN channel with a SNR equal to $h_p^2 \frac{E_b}{N_0}$.

Under such scenario, one can think, for instance, of using an *adaptive modulation and coding* approach: communication systems can be designed so that they are able to adapt themselves in real time to the channel conditions. Whenever the SNR changes significantly, these systems can switch to a different modulation scheme as well as a different error-correcting technique.

The goal is to keep the bit error probability constant (typically, $P_{eb} \sim 10^{-5}$) and independent from the SNR while achieving, at all times, the highest possible bit rate. Recall that the bit rate, often expressed in Mbits per second (Mbps), is a measure of the transmission speed.

If the SNR collapses to a very low level, the system can employ a very power-efficient modulation scheme, such as QPSK, combined with a near-capacity error-correcting code, such as a turbo or LDPC (low-density parity-check) code, with a low coding rate. In this way, the bit error probability could be kept unchanged despite the drop in SNR, but the price to pay would be a reduction in bit rate.

On the other hand, if the SNR increases to a very high level, the communication system can now switch to a bandwidth-efficient modulation scheme, such as 64- or 256-QAM, combined with a high coding rate channel code. As a result, one would obtain an increase in bit rate while keeping the bit error probability unchanged.

• **Frequency-selective fading channels ($\boldsymbol{\Delta\tau}$ NOT $\ll \boldsymbol{T}$)**

In today's wireless communication systems, it is often unrealistic to ensure that the delay spread $\Delta\tau = \tau_{L-1} - \tau_0$ is much smaller than the symbol period T. The need for ever-increasing bit and symbol rates requires ever-decreasing symbol durations, which makes it impossible to assume that $|\tau_s - \tau_l| \ll T$ for all paths in many systems.

In this case, we can no longer assume that $g\big((p-k)T + \tau_s - \tau_l\big) \sim g\big((p-k)T\big)$, $\forall\, l \in \{0, \dots, L-1\}$. Hence, the flat fading model is no longer valid, and we must develop a new model by starting from the original expression obtained by sampling the baseband signal $m'(t)$ at time $t = pT + \tau_s + T_p$:

$$C_p^{'} = m'\big(pT + \tau_s + T_p\big) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k \sum_{l=0}^{L-1} \overline{a}_l g\big((p-k)T + \tau_s - \tau_l\big) + n_p,$$

where $\overline{a_l} = a_l e^{-j2\pi f_0 \tau_l}$ and n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{N_0 E_h}{2}$.

Example --

To better understand this complicated expression, consider the example of the transmission of three successive symbols C_0, C_1, and C_2 at times $t = 0$, $t = T$, and $t = 2T$, respectively.

Assume transmission over a two-path wireless channel with $\tau_0 = 0$ and $\tau_1 = \frac{3T}{2}$. The delay spread is thus $\Delta\tau = \tau_1 - \tau_0 = \frac{3T}{2}$, and is even greater than the duration T of a symbol.

The expression of the complex baseband signal generated by the matched filters is given by

$$m'(t) = \frac{1}{2}C_0\overline{a_0}g(t - \tau_0 - T_p) + \frac{1}{2}C_0\overline{a_1}g(t - \tau_1 - T_p) + \frac{1}{2}C_1\overline{a_0}g(t - T - \tau_0 - T_p) + \frac{1}{2}C_1\overline{a_1}g(t - T - \tau_1 - T_p) + \frac{1}{2}C_2\overline{a_0}g(t - 2T - \tau_0 - T_p) + \frac{1}{2}C_2\overline{a_1}g(t - 2T - \tau_1 - T_p) + n''(t).$$

which leads to

$$m'(t) = \frac{1}{2}\left[C_0\overline{a_0}g(t - T_p) + C_0\overline{a_1}g\left(t - \frac{3T}{2} - T_p\right) + C_1\overline{a_0}g(t - T - T_p) + C_1\overline{a_1}g\left(t - T - \frac{3T}{2} - T_p\right) + C_2\overline{a_0}g(t - 2T - T_p) + C_2\overline{a_1}g\left(t - 2T - \frac{3T}{2} - T_p\right)\right] + n''(t),$$

If we sample, for instance, this baseband signal at time $t = 2T + \tau_s + T_p$, with $\tau_s = \tau_0 = 0$, we obtain an estimate C_2' of the transmitted symbol C_2 expressed as follows:

$$C_2' = \frac{1}{2}\left[C_0\overline{a_0}g(2T) + C_0\overline{a_1}g\left(2T - \frac{3T}{2}\right) + C_1\overline{a_0}g(2T - T) + C_1\overline{a_1}g\left(2T - T - \frac{3T}{2}\right) + C_2\overline{a_0}g(2T - 2T) + C_2\overline{a_1}g\left(2T - 2T - \frac{3T}{2}\right)\right] + n''(2T + T_p),$$

which is equivalent to

$$C_2' = \frac{1}{2}\left[C_0\overline{a_0}g(2T) + C_0\overline{a_1}g\left(\frac{T}{2}\right) + C_1\overline{a_0}g(T) + C_1\overline{a_1}g\left(-\frac{T}{2}\right) + C_2\overline{a_0}g(0) + C_2\overline{a_1}g\left(-\frac{3T}{2}\right)\right] + n_2,$$

thus yielding

$$C_2' = \frac{1}{2}\left[C_0\overline{a_1}g\left(\frac{T}{2}\right) + C_1\overline{a_1}g\left(-\frac{T}{2}\right) + C_2\overline{a_0}g(0) + C_2\overline{a_1}g\left(-\frac{3T}{2}\right)\right] + n_2.$$

We can finally write $C_2' = \frac{1}{2}C_0\overline{a_1}g\left(\frac{T}{2}\right) + \frac{1}{2}C_1\overline{a_1}g\left(-\frac{T}{2}\right) + \frac{1}{2}C_2\left[\overline{a_0}g(0) + \overline{a_1}g\left(-\frac{3T}{2}\right)\right] + n_2$.

This expression shows that the sample C_2' depends on the symbol C_2, as it should be, but also on the other two symbols C_0 and C_1: there is inter-symbol interference (ISI), and that is detrimental to the transmission reliability.

Let us go back to the general case. By sampling the base-band signal at time $t = pT + \tau_s + T_p$, we obtain a channel estimate C_p' which is expressed as

$$C_p' = m'\left(pT + \tau_s + T_p\right) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k \sum_{l=0}^{L-1} \overline{a_l} g\left((p-k)T + \tau_s - \tau_l\right) + n_p.$$

where $\overline{a_l} = a_l e^{-j2\pi f_0 \tau_l}$ and n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{N_0 E_h}{2}$.

By developing this expression, we obtain

$$C_p' = \cdots + \frac{1}{2} C_{p-1} \sum_{l=0}^{L-1} \overline{a_l} g(T + \tau_s - \tau_l) + \frac{1}{2} C_p \sum_{l=0}^{L-1} \overline{a_l} g(\tau_s - \tau_l) + \frac{1}{2} C_{p+1} \sum_{l=0}^{L-1} \overline{a_l} g(-T + \tau_s - \tau_l) + \ldots + n_p.$$

This shows the presence of inter-symbol interference (ISI) as the sample C_p' clearly depends on the successive transmitted symbols $\ldots, C_{p-1}, C_p, C_{p+1}, \ldots$.

A multipath channel for which the delay spread $\Delta\tau$ is NOT much smaller than the symbol duration T is said to be *frequency selective* because such channel, unlike a flat fading channel, has a linear filtering effect on the transmitted signal $s(t)$. This is clearly indicated by the expression of the channel output C_p' as a function of the channel inputs $\ldots, C_{p-1}, C_p, C_{p+1}, \ldots$ which is that of a digital finite impulse response filter.

By applying the central-limit theorem, we can state that the terms ... , $\sum_{l=0}^{L-1} \overline{a_l} g(T + \tau_s - \tau_l)$, $\sum_{l=0}^{L-1} \overline{a_l} g(\tau_s - \tau_l)$, $\sum_{l=0}^{L-1} \overline{a_l} g(-T + \tau_s - \tau_l)$, ... are all complex Gaussian samples.

We can thus write that, for a frequency-selective multi-path channel, the sample $C_p^{'}$ taken at time $t = pT + \tau_s + T_p$ is given by

$$C_p^{'} = \cdots + w_{p-1}\frac{E_h}{2}C_{p-1} + w_p\frac{E_h}{2}C_p + w_{p+1}\frac{E_h}{2}C_{p+1} \ldots + n_p,$$

which can also be written as

$$C_p^{'} = \sum_{m=-\infty}^{+\infty} w_{p+m}\frac{E_h}{2}C_{p+m} + n_p,$$

where the quantities $w_{p+m} = w_{p+m,1} + jw_{p+m,2} = \sum_{l=0}^{L-1}\overline{a}_l\frac{g(-mT+\tau_s-\tau_l)}{g(0)} = \sum_{l=0}^{L-1}\overline{a}_l\frac{g(-mT+\tau_s-\tau_l)}{E_h}$ are complex Gaussian samples, whereas $n_p = n_{p,1} + jn_{p,2}$ is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{N_0E_h}{2}$.

If we normalize the expression of the sample $C_p^{'}$ for simplicity's sake, we obtain the expression of the channel estimate $C_p^{'}$ for a frequency-selective fading channel:

$$C_p^{'} = \sum_{m=-\infty}^{+\infty} w_{p+m}C_{p+m} + n_p,$$

where n_p is now a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \left(\frac{2}{E_h}\right)^2\frac{N_0E_h}{2} = \frac{2N_0}{E_h} = \gamma\left(\frac{E_s}{N_0}\right)^{-1} = \frac{\gamma}{m}\left(\frac{E_b}{N_0}\right)^{-1}$.

We notice that the pulses $g(t)$ are such that $\frac{g(-mT+\tau_s-\tau_l)}{g(0)} \to 0$ as $m \to -\infty$ and $m \to +\infty$. Therefore, in practice, the sum does not contain an infinite number of terms and the sample $C_p^{'}$ can then be expressed as

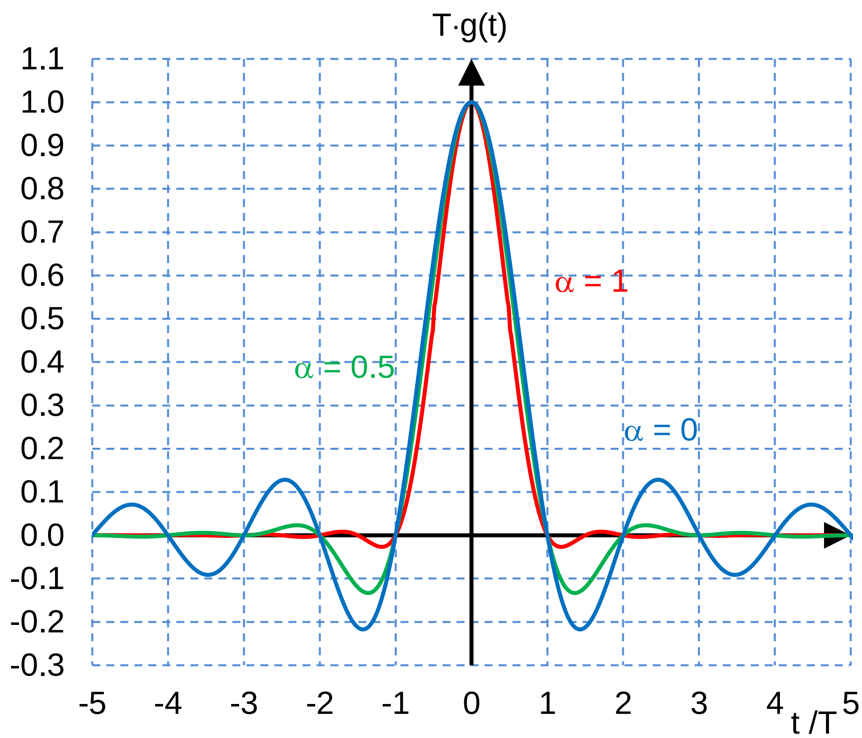
$$C_p^{'} = \sum_{m=-M}^{+M} w_{p+m}C_{p+m} + n_p,$$

where $2M+1$ is the number of taps in the frequency-selective multipath channel.

A polar notation can also be adopted for the fading samples w_{p+m}:

$$C_p^{'} = \sum_{m=-M}^{+M} h_{p+m}e^{j\theta_{p+m}}C_{p+m} + n_p,$$

where $h_{p+m} = \sqrt{\left(w_{p+m,1}\right)^2 + \left(w_{p+m,2}\right)^2}$ and $\theta_{p+m} = \tan^{-1}\left(\frac{w_{p+m,2}}{w_{p+m,1}}\right)$ designate the magnitude and argument, or phase, of w_{p+m}, respectively.



The arguments θ_{p+m} are uniformly distributed between 0 and 2π, whereas the fading samples h_{p+m} follow either a Rice-Nakagami or a Rayleigh distribution depending on whether the radio channel has a line of sight or not.

• Why do we need to normalize fading samples?

The fading samples must be normalized so that $\sum_{m=-M}^{+M}\left(h_{p+m}\right)^2 = 1$. This normalization allows for a fair comparison with other systems and other types of channels.

To clarify this point, let us consider the three following equations for the channels that have been studied:

(1) AWGN channel: $C_p' = C_p + n_p,$

(2) Flat fading channel: $C_p' = h_p e^{j\theta_p} C_p + n_p,$

(2) Frequency-selective fading channel: $C_p' = \sum_{m=-M}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} + n_p.$

In all three equations, the complex noise sample n_p is Gaussian with a mean equal to zero and a variance expressed as $2\sigma^2 = \frac{\gamma}{m}\left(\frac{E_b}{N_0}\right)^{-1}$. Recall that this expression means that the variance of both the real and imaginary parts of the noise sample is given by $\sigma^2 = \frac{\gamma}{2m}\left(\frac{E_b}{N_0}\right)^{-1}$.

Equivalently, the variance of the Gaussian noise sample n_p can also be expressed as a function of the SNR per symbol, $\frac{E_s}{N_0}$, rather than the SNR per information bit, $\frac{E_b}{N_0}$. In this case, we have $2\sigma^2 = \gamma\left(\frac{E_s}{N_0}\right)^{-1}$.

The parameter E_s has been defined throughout these lecture notes as the average energy per symbol available at the RF power amplifier output inside the receiver.

However, some authors define E_s as the average energy per transmitted symbol C_p, i.e., the average amount of energy allocated for the transmission of one particular symbol C_p. As a matter of fact, this is not a meaningful definition in the real world. What truly matters in a practical communication system is not the level of signal that was radiated by the transmit antenna, but instead the signal level available at the low-noise power amplifier output inside the receiver. Obviously, one can expect that both are going to be related, but it makes more sense to focus on the received signal as this is the one that is actually processed by the receiver.

Defining the quantity E_s as the average signal energy embedded in each channel sample $C_p^{'}$ is therefore the correct thing to do. This does not change anything for the AWGN channel as we have assumed for simplicity's sake that the signal component generated by the low-noise power amplifier was identical to the signal generated by the transmitter. Both have been referred to as the signal $s(t)$. But, in the context of multipath channels, our more realistic definition allows us to clarify the need for a normalization of the fading samples.

No matter what channel model we decide to use, it is crucial to ensure that the same average energy E_s is embedded in each channel sample $C_p^{'}$. Remember that fading, whether flat or frequency selective, is not supposed to amplify or attenuate the average SNR because that is not the result of signal propagation via multiple paths.

For any type of channel, the SNR $\frac{E_s}{N_0}$ does represent the average level of signal present in each channel sample $C_p^{'}$ with respect to the level of additive Gaussian noise which is measured by the parameter N_0. Since the additive noise has the same variance $2\sigma^2 = \gamma \cdot \left(\frac{E_s}{N_0}\right)^{-1}$ in our three channel models, we must ensure that the average signal energy embedded in each channel sample $C_p^{'}$ is also identical, and equal to E_s, for these different channel models.

On an AWGN channel, the signal energy E_s embedded in a sample $C_p^{'}$ is a constant and simply represents the energy allocated to each symbol C_p.

Over a flat fading channel, the SNR of each received sample does fluctuate depending on the channel conditions. The equation $C_p^{'} = h_p e^{j\theta_p} C_p + n_p$ clearly indicates that the signal energy E_s embedded in each sample $C_p^{'}$ is a random number $\left(h_p\right)^2 E_s$.

As the fading effect is not supposed to increase or decrease the average signal energy, we must always ensure that $E\left\{\left(h_p\right)^2 E_s\right\} = E\left\{\left(h_p\right)^2\right\} E_s = E_s$, i.e., we must have $E\left\{\left(h_p\right)^2\right\} = 1$.

As for the frequency-selective channel, the equation $C_p^{'} = \sum_{m=-M}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} + n_p$ shows that the total signal energy embedded in $C_p^{'}$ is now spread out over $(2M + 1)$ consecutive symbols.

The total amount of energy embedded in $C_p^{'}$ is given by $\sum_{m=-M}^{+M} \left(h_{p+m}\right)^2 E_s = E_s \sum_{m=-M}^{+M} \left(h_{p+m}\right)^2$. Once again, the fading effect is not supposed to increase or decrease the average signal energy. Consequently, we must ensure that $E\left\{E_s \sum_{m=-M}^{+M} \left(h_{p+m}\right)^2\right\} = E_s \cdot E\left\{\sum_{m=-M}^{+M} \left(h_{p+m}\right)^2\right\} = E_s \cdot \sum_{m=-M}^{+M} E\left\{\left(h_{p+m}\right)^2\right\} = E_s$, i.e., we must have $\sum_{m=-M}^{+M} E\left\{\left(h_{p+m}\right)^2\right\} = 1$.

• Converting a frequency-selective fading channel into a flat fading channel

The error performance of communication systems over frequency-selective channels is very poor due to the presence of ISI.

With an AWGN or a flat fading channel, we easily show that the bit error probability $p_{eb} \to 0$ as $\frac{E_S}{N_0} \to +\infty$, which is the least one can expect with any communication scheme.

However, this is not even true for a frequency-selective channel for which it is impossible to achieve an error probability equal to zero as $\frac{E_S}{N_0} \to +\infty$.

The reason is that, for such channel, the expression $C_p^{'} = \sum_{m=-M}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} + n_p$ can be rewritten as $C_p^{'} = h_p e^{j\theta_p} C_p + \sum_{m=-M,\neq 0}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} + n_p$. This indicates that the noise component in each received sample $C_p^{'}$ is in fact given by $\sum_{m=-M,\neq 0}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} + n_p$.

As $\frac{E_S}{N_0} \to +\infty$, we see that $\sum_{m=-M,\neq 0}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} + n_p \to \sum_{m=-M,\neq 0}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m}$, because we then have $\sum_{m=-M,\neq 0}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} \gg n_p$.

The fact that $\sum_{m=-M,\neq 0}^{+M} h_{p+m} e^{j\theta_{p+m}} C_{p+m} \neq 0$ implies that the noise cannot be made negligible when $\frac{E_S}{N_0} \to +\infty$, which shows up as an error floor in the error probability versus SNR curve. The existence of such error floor is obviously very detrimental to the performance of any communication system.

Ii is therefore crucial to always try to convert any frequency-selective channel into an equivalent flat fading channel which displays a much better error performance. Three techniques that are traditionally used to do so are equalisation, orthogonal frequency-division multiplexing (OFDM), or the rake receiver structure (if spreading codes are employed).

We are going to explain below the basic principles of OFDM.

• Orthogonal frequency-division multiplexing (OFDM)

One way to suppress ISI is to extend the duration of each symbol so that it becomes much greater than the delay spread $\Delta\tau$. OFDM technology allows for the duration of each symbol to be increased by a factor $N \gg 1$. By ensuring that $N \cdot T \gg \Delta\tau$, we can guarantee the absence of ISI in our communication link.

Frequency-division multiplexing (FDM) systems have been around for decades. In these systems, the sequence of successive complex symbols C_k of duration T is converted into N parallel streams in which the duration of each symbol becomes $N \cdot T$ seconds.

The demultiplexing operation converts a single high symbol rate stream into N low symbol rate parallel streams. The symbols $C_{k,n}, n \in \{0, \dots, N-1\}$, present on the $(n+1)$-th stream, are then transmitted using a carrier wave with a frequency $f_{0,n}$. The modulated signal for this particular stream can be expressed using complex notation as

$$s_n(t) = \Re\left[\sum_{k=0}^{+\infty} C_{k,n} h(t - kNT) e^{j2\pi f_{0,n} t}\right].$$

The total signal produced by the transmitter is the sum of all these modulated signals and is thus given by

$$s(t) = \sum_{n=0}^{N-1} s_n(t) = \Re\left[\sum_{n=0}^{N-1} \sum_{k=0}^{+\infty} C_{k,n} h(t - kNT) e^{j2\pi f_{0,n} t}\right].$$

At the receiver side, the symbols $C_{k,n}, n \in \{0, \dots, N-1\}$, are recovered by simply mixing the received signal with the carrier $e^{j2\pi f_{0,n} t}$ and then using a low-pass matched filter followed by a sampling device, as it is traditionally done.

Assuming the use of root raised cosine filters at both the transmitter and receiver sides, the bandwidth required for transmitting a signal $s_n(t)$ is given by $B_n = \frac{1+\alpha}{NT}$, and the separation between two adjacent carrier frequencies $f_{0,n}$ and $f_{0,n+1}$ must be, at least, $\frac{1+\alpha}{NT}$.

Hence, the bandwidth, B_{FDM}, required for transmitting the combined signal produced by the transmitter is, in practice, given by

$$B_{FDM} = NB_n + (N-1)B_g = \frac{1+\alpha}{T} + (N-1)B_g,$$

where B_g denotes the guard band that separates two adjacent spectra.

The FDM system bandwidth B_{FDM} is thus higher than the bandwidth $B_{SC} = \frac{1+\alpha}{T}$ that is required in a traditional single-carrier communication scheme due to the need for guard bands in practice. The presence of the term $(N-1)B_g$ in the equation above indicates that the difference between the single-carrier and the FDM systems can in fact be quite significant for large values of N.

In any case, the key advantage of a multicarrier system compared to its single-carrier counterpart is the absence of ISI on the parallel streams due to the low symbol rates on these streams.

OFDM works in a similar fashion as FDM, but with a few fundamental differences.

Firstly, the symbols $C_{k,n}, n \in \{0, \dots, N-1\}$, are transmitted using simple square pulses rather than root raised cosine pulses, which simplifies the implementation of the communication system.

Secondly, the carrier frequencies $f_{0,n}$ are in the form f_0, $f_0 + \frac{1}{NT}$, $f_0 + \frac{2}{NT}$, $f_0 + \frac{3}{NT}$, …, $f_0 + \frac{N-1}{NT}$. We are going to see that these frequencies are orthogonal, meaning that it is possible to *attach* N different streams of symbols $C_{k,n}, n \in \{0, \dots, N-1\}$, to each of them before transmission, mix (sum) the N resulting modulated signals over the wireless channel, and then separate these streams in the receiver.

Because an OFDM system employs square pulses, the power spectral density of each modulated signal $s_n(t) = \Re\left[\sum_{k=0}^{+\infty} C_{k,n} h(t - kNT) e^{j2\pi f_{0,n} t}\right]$ can be expressed using a sine cardinal function as follows:

$$S_n(f) = K \cdot \mathrm{sinc}^2\left(\pi NT\left(f - f_{0,n}\right)\right),$$

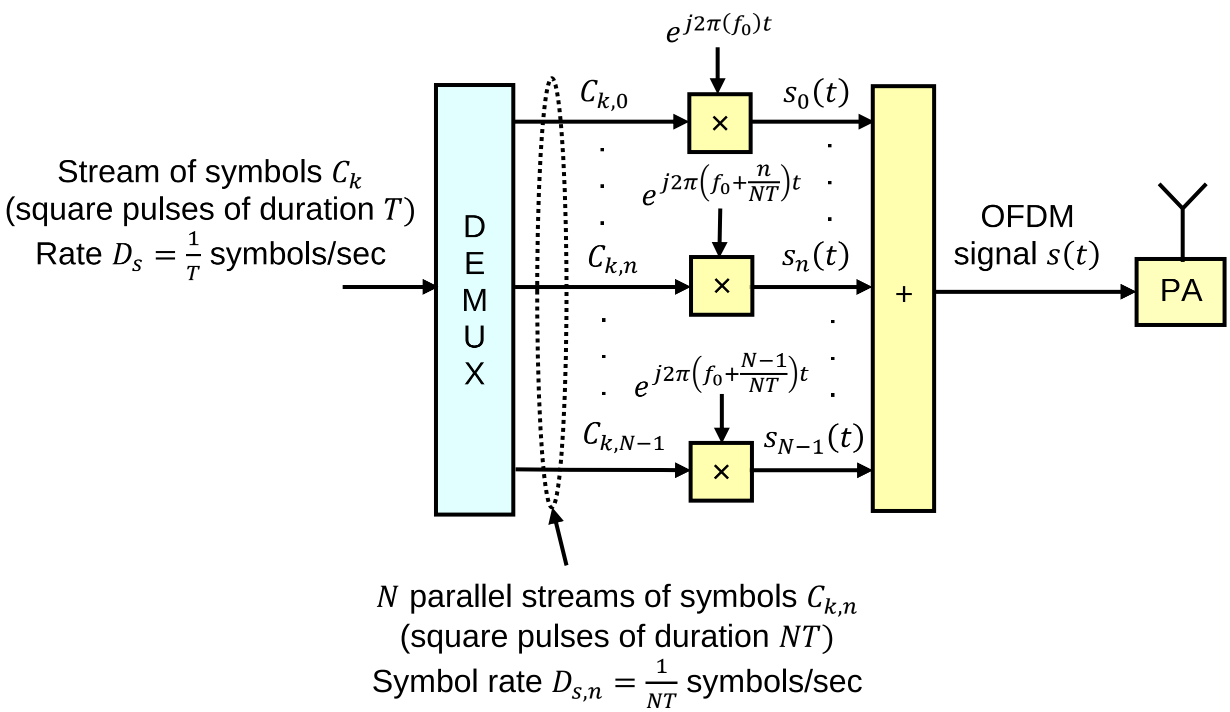
with $f_{0,n} = f_0 + \frac{n}{NT}, n \in \{0, \dots, N-1\}$.

The total signal produced by the OFDM transmitter is the sum of the modulated signals $s_n(t)$:

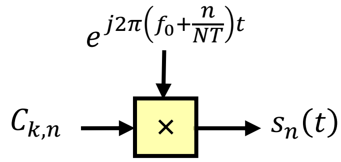
$$s(t) = \sum_{n=0}^{N-1} s_n(t) = \Re\left[\sum_{n=0}^{N-1} \sum_{k=0}^{+\infty} C_{k,n} h(t - kNT) e^{j2\pi f_{0,n} t}\right] = \Re\left[\sum_{n=0}^{N-1} \sum_{k=0}^{+\infty} C_{k,n} h(t - kNT) e^{j2\pi\left(f_0 + \frac{n}{NT}\right)t}\right],$$

which can be also written as $s(t) = \Re\left[\sum_{n=0}^{N-1} \sum_{k=0}^{+\infty} C_{k,n} h(t - kNT) e^{j2\pi\left(\frac{n}{NT}\right)t} e^{j2\pi f_0 t}\right]$.

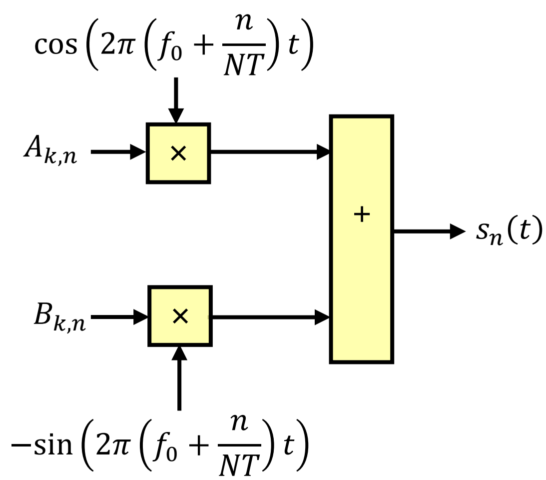
The terms $e^{j2\pi\left(\frac{n}{NT}\right)t}, n \in \{0, \dots, N-1\}$, represent the sub-carriers associated with the N parallel streams.



Note that, in the figure above, the notation

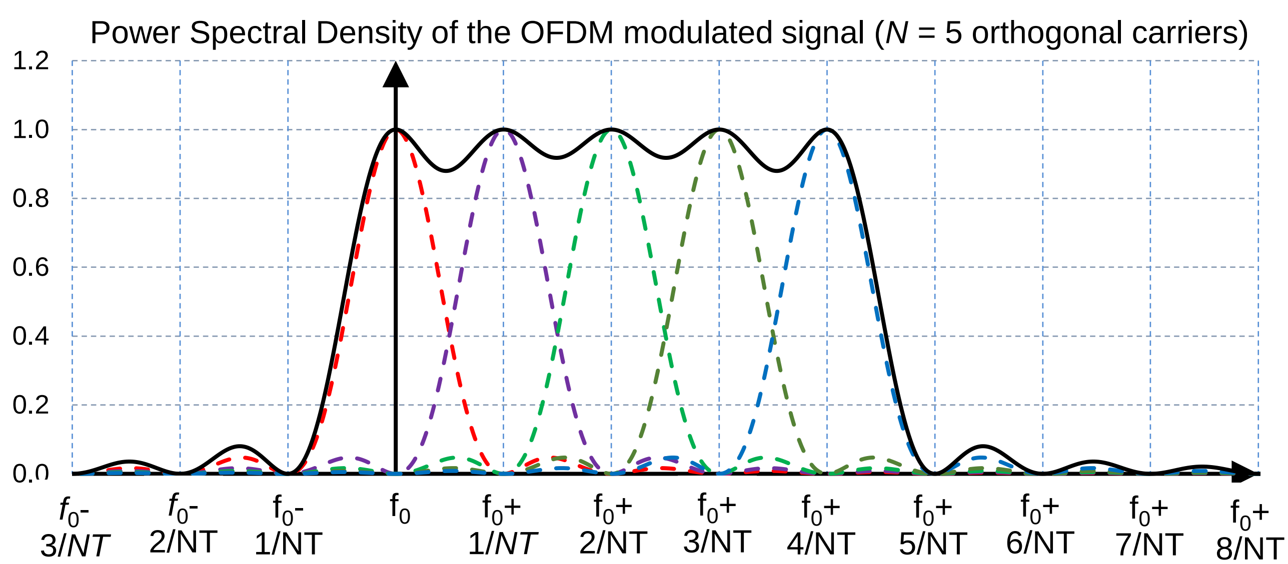


is equivalent to



.

As an illustration, we have shown below the power spectral density (PSD) of the modulated signals $s_n(t)$ in the case where the number of sub-carriers is $N = 5$. The total PSD, i.e., that of the combined signal $s(t) = \sum_{n=0}^{N-1} s_n(t)$, is also shown.



We notice that there is a significant overlap of the individual spectra $S_n(f)$. The overall spectrum of the combined OFDM signal $s(t)$ can be seen as ranging, approximately, from $f_0 - \frac{1}{NT}$ to $f_0 + \frac{N-1}{NT} + \frac{1}{NT}$. Therefore, the bandwidth required for the transmission of this signal is

$$B_{OFDM} \sim \frac{N+1}{NT} \sim \frac{1}{T}.$$

The OFDM system bandwidth B_{OFDM} is thus lower than both bandwidths $B_{SC} = \frac{1+\alpha}{T}$ and $B_{FDM} = \frac{1+\alpha}{T} + (N-1)B_g$ needed by equivalent single-carrier and FDM systems, respectively. OFDM is thus a bandwidth-efficient technique.

The overlap between individual spectra overlap could certainly cause very significant interference between parallel streams that are adjacent in the frequency domain. However, it does not happen because, at frequencies $f_{0,n} = f_0 + \frac{n}{NT}, n \in \{0, \dots, N-1\}$, the magnitudes of all other individual spectra is zero.

Such a communication scheme is therefore said to operate using orthogonal frequencies and that is the reason it is referred to as orthogonal FDM (OFDM). This orthogonality between sub-carrier frequencies $f_{0,n}, n \in \{0, \dots, N-1\}$, is going to allow for the separation of the individual streams at the receiver side.

To explain how the parallel streams are separated at the receiver side, consider a particular time interval ranging from $t = kNT$ to $(k+1)NT$. Assuming the square pulses have unit amplitude, the expression of the OFDM signal transmitted during this time interval is

$$s(t) = \sum_{n=0}^{N-1} s_n(t) = \Re\left[\sum_{n=0}^{N-1} C_{k,n} e^{j2\pi\left(f_0 + \frac{n}{NT}\right)t}\right],$$

which can also be written as

$$s(t) = \sum_{n=0}^{N-1} A_{k,n} \cos\left(2\pi\left(f_0 + \frac{n}{NT}\right)t\right) - B_{k,n} \sin\left(2\pi\left(f_0 + \frac{n}{NT}\right)t\right),$$

since we have $C_{k,n} = A_{k,n} + jB_{k,n}$.

Assume noise-free transmission for simplicity's sake as our goal here is only to show how the various real-valued symbols $A_{k,n}$ and $B_{k,n}, n \in \{0, \dots, N-1\}$, which are all combined in a single OFDM signal, can be separated after reception.

Assume we want to recover a particular symbol $A_{k,p}$. We first need to mix the received signal $r(t) = s(t)$ with the sinusoidal function $\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right)$. The resulting signal is given by

$$r(t)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right) = \sum_{n=0}^{N-1} A_{k,n}\cos\left(2\pi\left(f_0 + \frac{n}{NT}\right)t\right)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right) - B_{k,n}\sin\left(2\pi\left(f_0 + \frac{n}{NT}\right)t\right)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right),$$

which is equivalent to

$$r(t)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right) = \sum_{n=0}^{N-1} \frac{A_{k,n}}{2}\left[\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right) + \cos\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)\right] - \frac{B_{k,n}}{2}\left[\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right) + \sin\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)\right].$$

Then, this signal is integrated between $t = kNT$ and $(k+1)NT$:

$$\int_{kNT}^{(k+1)NT} r(t)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right)dt = \int_{kNT}^{(k+1)NT} \sum_{n=0}^{N-1} \frac{A_{k,n}}{2}\left[\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right) + \cos\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)\right] - \frac{B_{k,n}}{2}\left[\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right) + \sin\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)\right]dt,$$

which yields

$$\int_{kNT}^{(k+1)NT} r(t)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right)dt = \sum_{n=0}^{N-1} \frac{A_{k,n}}{2}\int_{kNT}^{(k+1)NT} \cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt + \sum_{n=0}^{N-1} \frac{A_{k,n}}{2}\int_{kNT}^{(k+1)NT} \cos\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)dt - \sum_{n=0}^{N-1} \frac{B_{k,n}}{2}\int_{kNT}^{(k+1)NT} \sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt - \sum_{n=0}^{N-1} \frac{B_{k,n}}{2}\int_{kNT}^{(k+1)NT} \sin\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)dt.$$

.

For $n \neq p$, we notice that both terms $\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)$ and $\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)$ are sinusoidal functions with a period $\frac{NT}{|n-p|}$. The integral of such periodic functions over a time window equal to NT, i.e., between times $t = kNT$ and $t = (k+1)NT$, is equal to zero because NT is a multiple of the period of $\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)$ and $\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)$.

For $n = p$, we have $\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right) = \cos(0) = 1$ and $\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right) = \sin(0) = 0$. The integral of a constant function of unit amplitude between $t = kNT$ and $t = (k+1)NT$, is equal to NT, whereas the integral of zero over the same time interval is obviously zero.

We also notice that both terms $\cos\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)$ and $\sin\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)$ are sinusoidal functions with a period $\frac{1}{2f_0+\frac{n+p}{NT}} = \frac{NT}{2NTf_0+n+p}$. The integral of such periodic functions between $t = kNT$ and $t = (k+1)NT$ is equal to zero because NT is a multiple of the period of $\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)$ and $\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)$. This is easily understood once we realise that the quantity $(2NTf_0 + n + p)$ is an integer since the product Tf_0 is itself an integer.

Therefore, we can finally write $\int_{kNT}^{(k+1)NT} r(t)\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right)dt = A_{k,p}\frac{NT}{2}$., which clearly shows that, by mixing the received OFDM signal $r(t)$ with the sinusoidal function $\cos\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right)$ and then performing an integration between $t = kNT$ and $t = (k+1)NT$, the receiver is able to recover the symbol $A_{k,p}$.

To recover the symbol $B_{k,p}$, we first mix the received signal $r(t) = s(t)$ with the sinusoidal function $-\sin\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right)$. The resulting signal is given by

$$-r(t)\sin\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right) = \sum_{n=0}^{N-1} -A_{k,n}\cos\left(2\pi\left(f_0 + \frac{n}{NT}\right)t\right)\sin\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right) +$$
$$B_{k,n}\sin\left(2\pi\left(f_0 + \frac{n}{NT}\right)t\right)\sin\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right),$$

which is equivalent to

$$-r(t)\sin\left(2\pi\left(f_0 + \frac{p}{NT}\right)t\right) = \sum_{n=0}^{N-1} \frac{A_{k,n}}{2}\left[\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right) - \sin\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)\right] +$$
$$\frac{B_{k,n}}{2}\left[\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right) - \cos\left(2\pi\left(2f_0 + \frac{n+p}{NT}\right)t\right)\right].$$

Then, this signal is integrated between $t = kNT$ and $(k+1)NT$:

$$-\int_{kNT}^{(k+1)NT} r(t)\sin\left(2\pi\left(f_0+\frac{p}{NT}\right)t\right)dt = \int_{kNT}^{(k+1)NT}\sum_{n=0}^{N-1}\frac{A_{k,n}}{2}\left[\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)-\sin\left(2\pi\left(2f_0+\frac{n+p}{NT}\right)t\right)\right]+\frac{B_{k,n}}{2}\left[\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)-\cos\left(2\pi\left(2f_0+\frac{n+p}{NT}\right)t\right)\right]dt,$$

which yields

$$-\int_{kNT}^{(k+1)NT} r(t)\sin\left(2\pi\left(f_0+\frac{p}{NT}\right)t\right)dt = \sum_{n=0}^{N-1}\frac{A_{k,n}}{2}\int_{kNT}^{(k+1)NT}\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt - \sum_{n=0}^{N-1}\frac{A_{k,n}}{2}\int_{kNT}^{(k+1)NT}\sin\left(2\pi\left(2f_0+\frac{n+p}{NT}\right)t\right)dt + \sum_{n=0}^{N-1}\frac{B_{k,n}}{2}\int_{kNT}^{(k+1)NT}\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt - \sum_{n=0}^{N-1}\frac{B_{k,n}}{2}\int_{kNT}^{(k+1)NT}\cos\left(2\pi\left(2f_0+\frac{n+p}{NT}\right)t\right)dt.$$

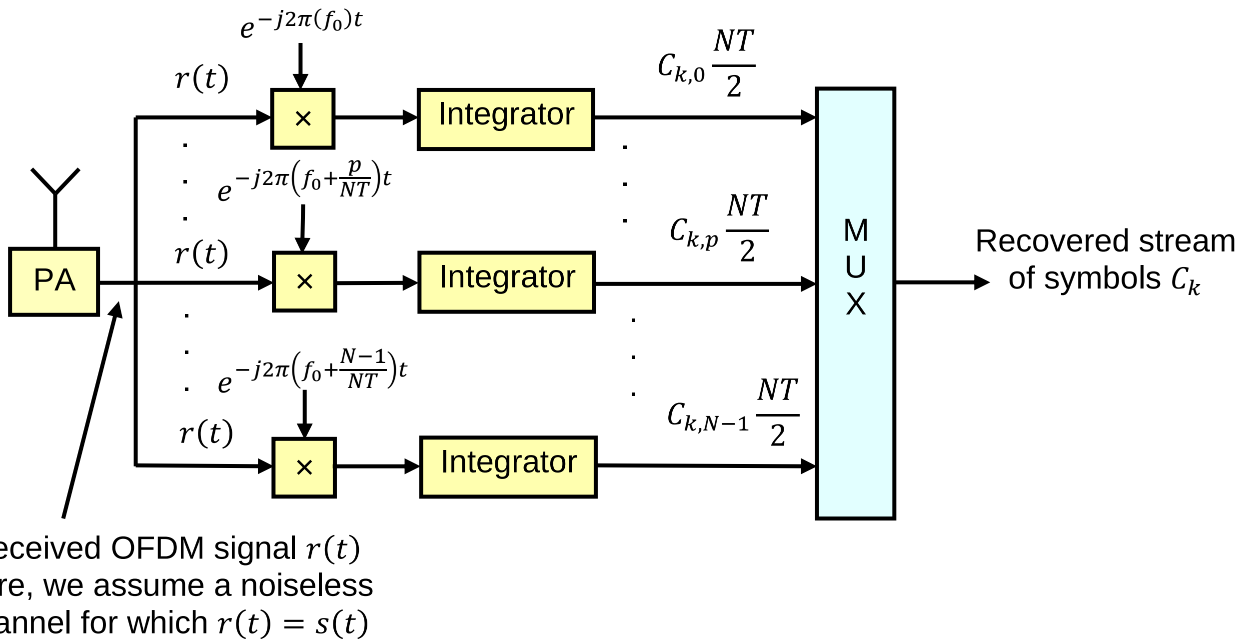
.

In the same way as was previously done, we can show that

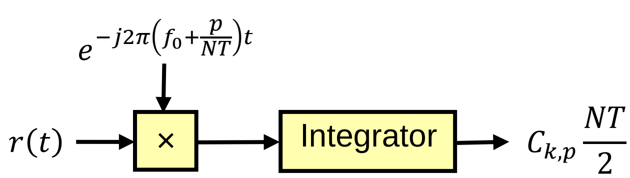
- $\int_{kNT}^{(k+1)NT}\sin\left(2\pi\left(2f_0+\frac{n+p}{NT}\right)t\right)dt = \int_{kNT}^{(k+1)NT}\cos\left(2\pi\left(2f_0+\frac{n+p}{NT}\right)t\right)dt = 0$ for any value of p,
- $\int_{kNT}^{(k+1)NT}\sin\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt = 0$ for any value of p,
- $\int_{kNT}^{(k+1)NT}\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt = 0$ for any value of $p \neq n$,
- $\int_{kNT}^{(k+1)NT}\cos\left(2\pi\left(\frac{n-p}{NT}\right)t\right)dt = \int_{kNT}^{(k+1)NT}\cos(0)dt = \int_{kNT}^{(k+1)NT}1\,dt = NT$ for $p = n$,

We can finally write $-\int_{kNT}^{(k+1)NT} r(t)\sin\left(2\pi\left(f_0+\frac{p}{NT}\right)t\right)dt = B_{k,p}\frac{NT}{2}$., which clearly shows that, by mixing the received OFDM signal $r(t)$ with the sinusoidal function $-\sin\left(2\pi\left(f_0+\frac{p}{NT}\right)t\right)$ and then performing an integration between $t = kNT$ and $t = (k+1)NT$, the receiver is able to recover the symbol $B_{k,p}$.

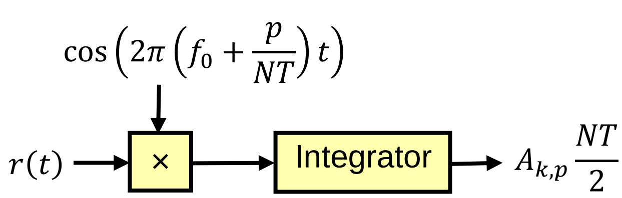
The general structure of an OFDM receiver is shown below.

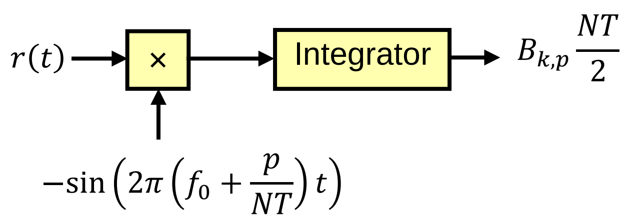


Note that, in the figure above, the notation



is equivalent to





• Practical implementation of OFDM using IDFT and DFT

In today's wireless communication systems, the number N of carriers can be very large. The issue is that the implementation of a very large number of mixers in each transmitter and an equally large number of mixers and integrators in each receiver is clearly not a realistic option. We thus need to find a more practical way to implement the OFDM technology. A possible technique to do so consists of using an inverse discrete Fourier transform (IDFT) at the transmitter side and a corresponding discrete Fourier transform (DFT) at the receiver side.

We are now going to revisit the previous explanations of OFDM technology by showing how to implement the latter using IDFT and DFT. To this end, we consider hereafter a (relatively) slow-changing multipath channel with no line of sight. In other words, the fading samples are assumed to follow a Rayleigh distribution and remain constant over the duration of a large number (N) of consecutive symbols. These are realistic assumptions in many practical wireless systems.

We have previously seen that, in OFDM communication systems, the sequence of complex symbols to be transmitted is first broken into successive vectors of N symbols by using a demultiplexing function. Let us focus on one of these vectors and adopt the following notation: A vector is composed of N symbols $C_k, k \in \{0, \dots, N-1\}$. Each of these vectors is then used to compute a new vector of N complex symbols $\bar{C}_p$ by using inverse discrete Fourier transform (IDFT):

$$\bar{C}_p = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} C_k \cdot e^{j\frac{2\pi kp}{N}}, p \in \{0, \dots, N-1\}.$$

The modulated signal generated by the OFDM transmitter for this vector of N symbols $\bar{C}_p, p \in \{0, \dots, N-1\}$, can be expressed using complex notation as

$$s(t) = \Re\left[\sum_{p=0}^{N-1} \bar{C}_p h(t - pT) e^{j2\pi f_0 t}\right],$$

which can be further developed as

$$s(t) = \Re\left[\frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} \sum_{k=0}^{N-1} C_k\, h(t - pT) e^{j\left(2\pi f_0 t + \frac{2\pi kp}{N}\right)}\right].$$

This equation can also be written as

$$s(t) = \Re\left[\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} C_k \sum_{p=0}^{N-1} h(t-pT)e^{j\left(2\pi f_0 t+\frac{2\pi kp}{N}\right)}\right].$$

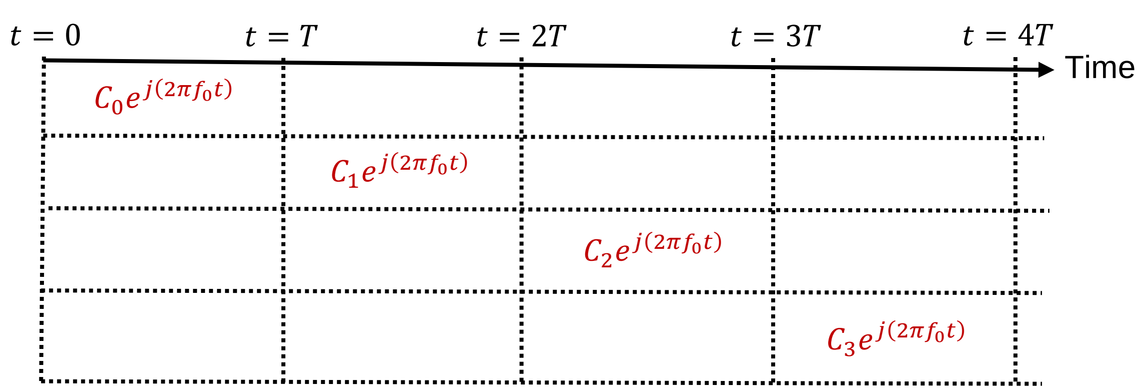
Recall that, without IDFT prior to transmission, the expression of the modulated signal would be

$$s(t) = \Re\left[\sum_{k=0}^{N-1} C_k h(t-kT)e^{j2\pi f_0 t}\right].$$

To better understand the difference between a traditional transmitter and an OFDM transmitter, let us consider a simple example for which the symbol-carrying pulse $h(t)$ is a square pulse equal to the unit between $t = 0$ and T, and equal to zero outside this interval. Let us also focus on a particular symbol C_k transmitted between $t = kT$ and $t = (k+1)T$.

In a traditional communication system relying on square pulses, the expression of the modulated signal generated by the transmitter between $t = kT$ and $t = (k+1)T$ is given by $s(t) = \Re\left[C_k e^{j2\pi f_0 t}\right]$. The time spent to transmit symbol C_k is T seconds.

As illustrated in the drawing below when $N = 4$, a traditional system therefore operates similarly to a time-division multiple access (TDMA) system for which each symbol is given its own time slot and shares the same sine wave $e^{j(2\pi f_0 t)}$ with all other symbols.



In an OFDM communication system, the expression of the modulated signal generated by the transmitter between $t = kT$ and $t = (k+1)T$ is given by

$$s(t) = \frac{1}{\sqrt{N}} \Re\left[C_k \sum_{p=0}^{N-1} h(t-pT) e^{j\left(2\pi f_0 t + \frac{2\pi k p}{N}\right)}\right].$$

This equation shows that a symbol C_k is transmitted using the sine wave $e^{j(2\pi f_0 t)}$ between $t = 0$ and T, $e^{j\left(2\pi f_0 t + \frac{2\pi k}{N}\right)}$ between $t = T$ and $2T, e^{j\left(2\pi f_0 t + \frac{4\pi k}{N}\right)}$ between $t = 2T$ and $3T$, etc. This indicates that the time spent to transmit symbol C_k is NT seconds.

The N-fold increase in symbol duration constitutes a great improvement compared to a traditional communication system because this is precisely what allows the OFDM technology to suppress ISI.

Recall that a traditional communication scheme does exhibit ISI when it cannot satisfy the condition $\Delta\tau \ll T$, which is often the case in today's high-data rate applications.

In contrast, an OFDM system is not affected by ISI since the condition required to suppress it becomes $\Delta\tau \ll NT$, with $N \gg 1$. In most cases, the value of N can be adjusted to ensure that this inequality is satisfied.

As illustrated in the drawing below when $N = 4$, the time spent to transmit any symbol is $NT = 4T$ seconds. Each symbol C_k uses all available time slots but must obviously share them with all other symbols.

All signals are mixed during transmission over the radio channel. The task of the OFDM receiver consists of separating these signals to allow for the recovery of the symbols C_k.

The separation of symbols C_k at the receiver side is made possible by the fact that they use different patterns of sine waves, in a way that is somewhat similar to frequency-division multiple access (FDMA).

$t=0$ $t=T$ $t=2T$ $t=3T$ $t=4T$ → Time

$C_0 e^{j(2\pi f_0 t)}$	$C_0 e^{j(2\pi f_0 t)}$	$C_0 e^{j(2\pi f_0 t)}$	$C_0 e^{j(2\pi f_0 t)}$
$C_1 e^{j(2\pi f_0 t)}$	$C_1 e^{j\left(2\pi f_0 t+\frac{\pi}{2}\right)}$	$C_1 e^{j(2\pi f_0 t+\pi)}$	$C_1 e^{j\left(2\pi f_0 t+\frac{3\pi}{2}\right)}$
$C_2 e^{j(2\pi f_0 t)}$	$C_2 e^{j(2\pi f_0 t+\pi)}$	$C_2 e^{j(2\pi f_0 t)}$	$C_2 e^{j(2\pi f_0 t+\pi)}$
$C_3 e^{j(2\pi f_0 t)}$	$C_3 e^{j\left(2\pi f_0 t+\frac{3\pi}{2}\right)}$	$C_3 e^{j(2\pi f_0 t+\pi)}$	$C_3 e^{j\left(2\pi f_0 t+\frac{\pi}{2}\right)}$

The signal transmitted between $t=pT$ and $t=(p+1)T$ is $\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} C_k e^{j\left(2\pi f_0 t+\frac{2\pi kp}{N}\right)} = \frac{1}{2}\sum_{k=0}^{3} C_k e^{j\left(2\pi f_0 t+\frac{\pi}{2}kp\right)}$. It is the sum of all terms in the corresponding column (the normalization factor $\frac{1}{\sqrt{N}}$ has not been shown in the drawing).

In this example, we see that the transmitted signal is the sum of $N=4$ sine waves with a frequency f_0 and different phase shifts at times $t=kT$. These patterns of phase shifts are orthogonal. This crucial property will be used by the receiver to separate the four sine waves after transmission and recover the original symbols C_k.

In our example, the patterns $\left\{e^{j\left(\frac{2\pi kp}{N}\right)}\right\}$ of phase shifts are as follows:

- For transmitting C_0: $\left\{e^{j\left(\frac{2\pi kp}{N}\right)}\right\} = \left\{e^{j0}, e^{j0}, e^{j0}, e^{j0}\right\}$.
- For transmitting C_1: $\left\{e^{j\left(\frac{2\pi kp}{N}\right)}\right\} = \left\{e^{j0}, e^{j\frac{\pi}{2}}, e^{j\pi}, e^{j\frac{3\pi}{2}}\right\}$.
- For transmitting C_2: $\left\{e^{j\left(\frac{2\pi kp}{N}\right)}\right\} = \left\{e^{j0}, e^{j\pi}, e^{j0}, e^{j\pi}\right\}$.
- For transmitting C_3: $\left\{e^{j\left(\frac{2\pi kp}{N}\right)}\right\} = \left\{e^{j0}, e^{j\frac{3\pi}{2}}, e^{j\pi}, e^{j\frac{\pi}{2}}\right\}$.

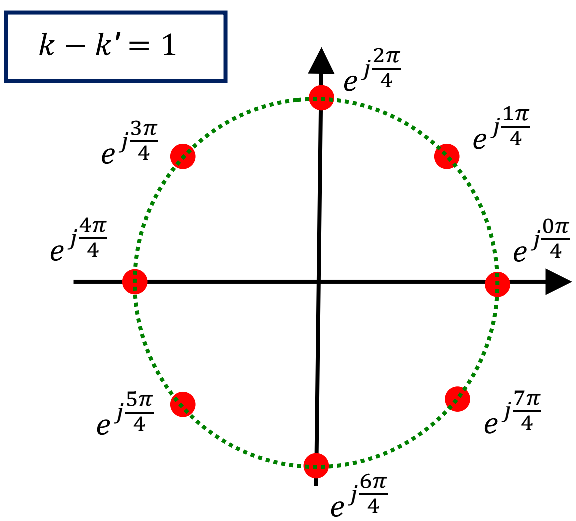
Two phase patterns $\left\{e^{j\left(\frac{2\pi kp}{N}\right)}\right\}$ and $\left\{e^{j\left(\frac{2\pi k'p}{N}\right)}\right\}$ are said to be orthogonal because, for any $k, k' \in \{1, \dots, N-1\}$, we have

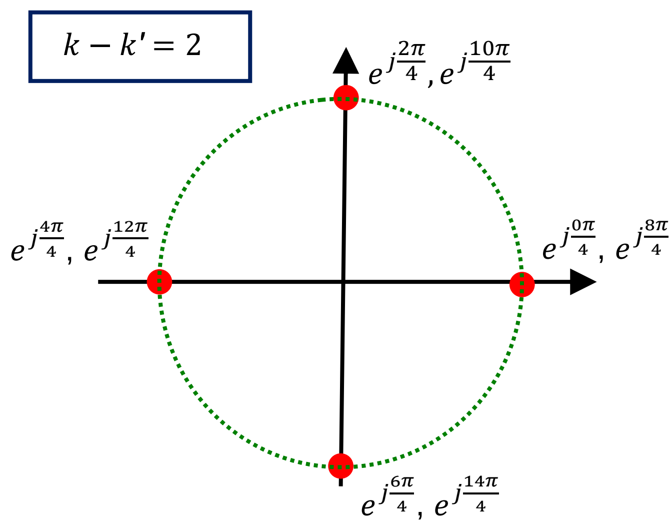
$$\sum_{p=0}^{N-1} e^{j\frac{2\pi kp}{N}} \left(e^{j\frac{2\pi k'p}{N}}\right)^* = \sum_{p=0}^{N-1} e^{j\frac{2\pi(k-k')p}{N}} = 0 \text{, when } k \neq k',$$

and $$\sum_{p=0}^{N-1} e^{j\frac{2\pi(k-k')p}{N}} \neq 0 \text{, when } k = k'.$$

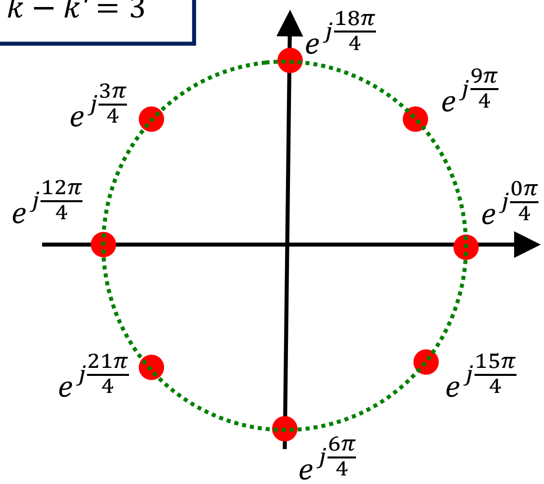
As an illustration of orthogonality, we show in the figures below an example with $N = 8$. It can be graphically seen that we indeed have $\sum_{p=0}^{7} e^{j\frac{\pi}{4}p(k-k')} = 0$, for $k - k' = 1, 2. \dots, 7$.

As both k and k' range from 0 to $N - 1 = 7$, the quantity $k - k'$ can actually range from -7 to $+7$. However, to scan all possibilities, we do not need to consider the negative values of $k - k'$ because having $\sum_{p=0}^{7} e^{j\frac{\pi}{4}p(k-k')} = 0$ implies that $\left(\sum_{p=0}^{7} e^{j\frac{\pi}{4}p(k-k')}\right)^* = \sum_{p=0}^{7} e^{-j\frac{\pi}{4}p(k-k')} = \sum_{p=0}^{7} e^{j\frac{\pi}{4}p(k'-k)} = 0$.

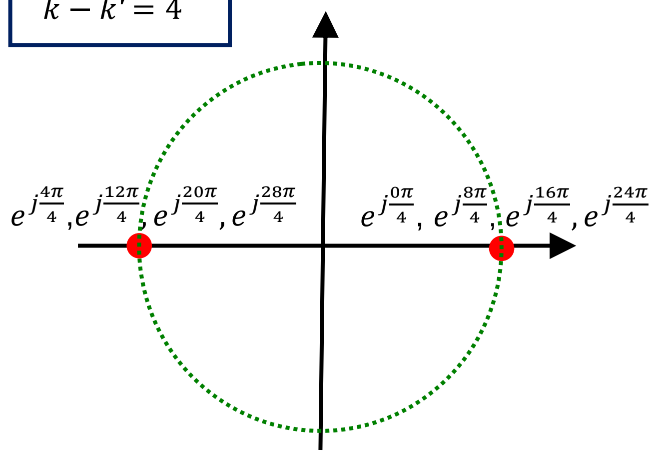




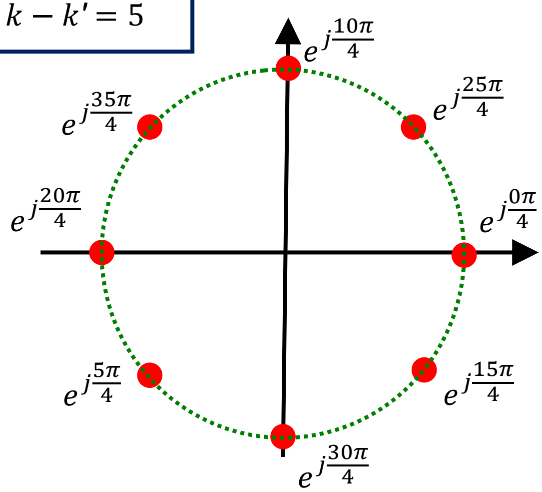
$k - k' = 3$



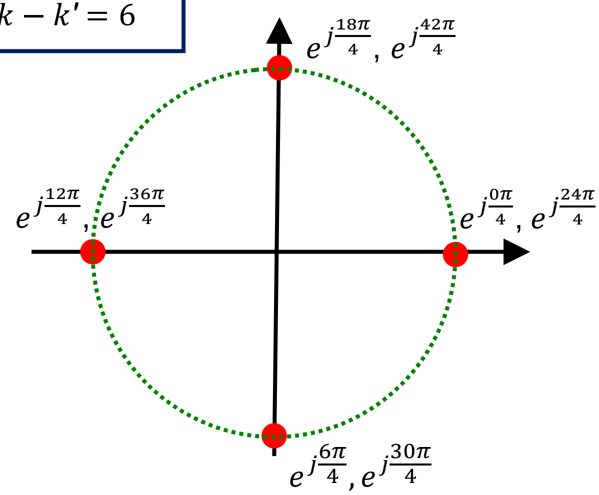
$k - k' = 4$



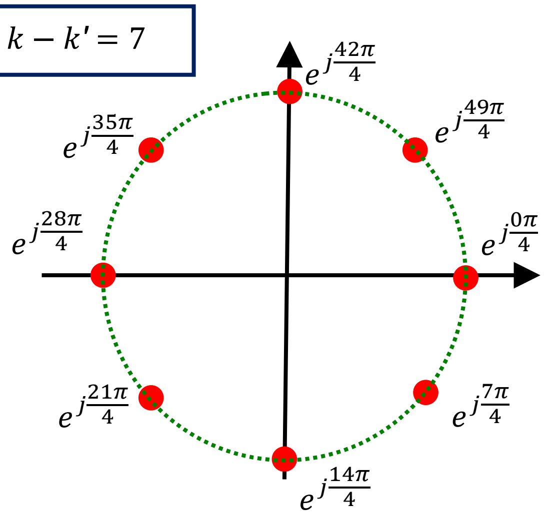
$k - k' = 5$



$k - k' = 6$



$k - k' = 7$



After demodulation, i.e., removal of the term $e^{j(2\pi f_0 t)}$ from the received signal, the receiver can recover all symbols C_k by performing a discrete Fourier transform (DFT) of the channel estimates representing the N symbols $\bar{C}_p, p \in \{0, \ldots, N-1\}$.

Assume for now noiseless transmission. Using DFT, the receiver computes a new vector of symbols of N samples $\bar{\bar{C}}_{k\prime}$ as follows:

$$\bar{\bar{C}}_{k\prime} = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} \bar{C}_p \cdot e^{-j\frac{2\pi lp}{N}},\ k' \in \{0, \ldots, N-1\}.$$

This expression can be written as

$$\bar{\bar{C}}_{k'} = \frac{1}{N} \sum_{p=0}^{N-1} \sum_{k=0}^{N-1} C_k \cdot e^{j\frac{2\pi kp}{N}} \cdot e^{-j\frac{2\pi k'p}{N}},$$

which is equivalent to

$$\bar{\bar{C}}_{k'} = \frac{1}{N} \sum_{k=0}^{N-1} C_k \sum_{p=0}^{N-1} e^{j\frac{2\pi(k-k')p}{N}}.$$

Let us develop this expression to better understand it:

$$\bar{\bar{C}}_{k'} = \frac{1}{N}\left(C_0 \sum_{p=0}^{N-1} e^{j\frac{2\pi\left(-k'\right)p}{N}} + C_1 \sum_{p=0}^{N-1} e^{j\frac{2\pi\left(1-k'\right)p}{N}} + \cdots + C_{N-1} \sum_{p=0}^{N-1} e^{j\frac{2\pi\left(N-1-k'\right)p}{N}} \right).$$

As previously seen, all the terms $\sum_{p=0}^{N-1} e^{j\frac{2\pi\left(-k'\right)p}{N}}$, $\sum_{p=0}^{N-1} e^{j\frac{2\pi\left(1-k'\right)p}{N}}, \ldots, \sum_{p=0}^{N-1} e^{j\frac{2\pi\left(N-1-k'\right)p}{N}}$ are equal to zero, except the one corresponding to $k = k'$ since $\sum_{p=0}^{N-1} e^{j\frac{2\pi(0)p}{N}} = N$. We can thus write

$$\bar{\bar{C}}_{k'} = \frac{1}{N}(NC_{k'}) = C_{k'}, \text{ for any } k' \in \{0, \ldots, N-1\}.$$

This result means that the vector of samples $\bar{\bar{C}}_{k'}$ is identical to the original vector of symbols C_k. The receiver has therefore been able to recover the symbols C_k by exploiting the orthogonality between the different patterns of phase shifts assigned to these symbols.

Let us now revisit the operation of the OFDM receiver by using a more realistic scenario.

Let us assume that the N complex symbols $\bar{C}_p$ are transmitted over a frequency-selective fading channel. Consider, without loss of generality, that the fading samples follow a Rayleigh distribution. In other words, the channel is said to be a frequency-selective Rayleigh fading channel. Also assume that the channel changes slowly so that the fading samples for all taps remain constant during the transmission of a vector of N symbols $\bar{C}_p$.

At the receiver side, the estimates $\bar{C}_p'$ of symbols $\bar{C}_p$ can thus be expressed as follows:

$$\bar{C}_p' = \sum_{m=-M}^{+M} w_m \bar{C}_{p+m} + n_p,$$

where the quantity $w_m = w_{m,1} + jw_{m,2}$ is a complex zero-mean Gaussian sample with a variance $2\sigma_m^2$, representing the fading affecting tap m, whereas n_p is a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \gamma \cdot \left(\frac{E_s}{N_0}\right)^{-1}$. Note that the fading samples must be normalised so that $\sum_{m=-M}^{+M} E\{(h_m)^2\} = 1$, which is equivalent to $\sum_{m=-M}^{+M} \left(E\left\{\left(w_{m,1}\right)^2\right\} + E\left\{\left(w_{m,2}\right)^2\right\}\right) = \sum_{m=-M}^{+M} 2\sigma_m^2 = 1$.

We can thus write

$$\bar{C}_p' = \sum_{m=-M}^{+M} w_m \left(\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} C_k \cdot e^{j\frac{2\pi k(p+m)}{N}}\right) + n_p = \frac{1}{\sqrt{N}} \sum_{m=-M}^{+M} \sum_{k=0}^{N-1} C_k \cdot w_m \cdot e^{j\frac{2\pi k(p+m)}{N}} + n_p,$$

which can be rewritten as

$$\bar{C}_p' = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} C_k \cdot e^{j\frac{2\pi kp}{N}} \sum_{m=-M}^{+M} w_m \cdot e^{j\frac{2\pi km}{N}} + n_p.$$

We notice that each term $w_m \cdot e^{j\frac{2\pi km}{N}}$ is a complex zero-mean Gaussian sample with a variance $2\sigma_m^2$. Hence, the term $w_k = \sum_{m=-M}^{+M} w_m \cdot e^{j\frac{2\pi km}{N}}$ is a zero-mean Gaussian sample with a variance given by $\sum_{m=-M}^{+M} 2\sigma_m^2 = 1$.

We thus have $\bar{C}_p' = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} w_k C_k \cdot e^{j\frac{2\pi kp}{N}} + n_p$.

This equation indicates that the (normalised) frequency-selective channel has just been converted into a (normalised) flat fading channel because any reference to multiple taps is no longer present in it.

At the receiver side, we perform a discrete Fourier transform using the N estimates $\bar{C}_p^{'}$, $p \in \{0, \ldots, N-1\}$, in order to produce a vector of N samples $\bar{\bar{C}}_{k'}$:

$$\bar{\bar{C}}_{k'} = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} \bar{C}_p^{'} \cdot e^{-j\frac{2\pi lp}{N}},\ k' \in \{0, \ldots, N-1\}.$$

We can then write

$$\bar{\bar{C}}_{k'} = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} \left(\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} w_k C_k \cdot e^{j\frac{2\pi kp}{N}} + n_p \right) \cdot e^{-j\frac{2\pi lp}{N}},$$

which is equivalent to

$$\bar{\bar{C}}_{k'} = \frac{1}{N} \sum_{p=0}^{N-1} \sum_{k=0}^{N-1} w_k C_k \cdot e^{j\frac{2\pi(k-l)p}{N}} + \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} n_p \cdot e^{-j\frac{2\pi lp}{N}}.$$

We obtain $\bar{\bar{C}}_{k'} = \frac{1}{N} \sum_{k=0}^{N-1} w_k C_k \sum_{p=0}^{N-1} e^{j\frac{2\pi(k-l)p}{N}} + \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} n_p \cdot e^{-j\frac{2\pi lp}{N}}$.

Due to the orthogonality of the subcarriers, we can write

$$\sum_{p=0}^{N-1} e^{j\frac{2\pi(k-l)p}{N}} = 0 \text{ when } k \neq l,$$

and

$$\sum_{p=0}^{N-1} e^{j\frac{2\pi(k-l)p}{N}} = N \text{ when } k = l,$$

Hence, the equation $\bar{\bar{C}}_{k'} = \frac{1}{N} \sum_{k=0}^{N-1} w_k C_k \sum_{p=0}^{N-1} e^{j\frac{2\pi(k-l)p}{N}} + \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} n_p \cdot e^{-j\frac{2\pi lp}{N}}$ becomes

$$\bar{\bar{C}}_{k'} = w_l C_l + n_l,$$

where $n_l = \frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} n_p \cdot e^{-j\frac{2\pi lp}{N}}$ is a zero-mean Gaussian noise sample with a variance $2\sigma^2 = \gamma \cdot \left(\frac{E_S}{N_0}\right)^{-1}$ because each term $n_p \cdot e^{-j\frac{2\pi lp}{N}}$ is a zero-mean Gaussian sample with the same variance.

As a conclusion, the OFDM technique has allowed us to convert a slow-fading frequency-selective Rayleigh fading channel into a flat Rayleigh fading channel, which is going to tremendously improve the error performance of our communication system.

11. Assessing the Error Performance of Uncoded and Coded Digital Communications Schemes using Computer Simulations

In a previous chapter, we managed to obtain a generic mathematical expression for approximating the error probability for any modulation scheme. This expression has been obtained assuming transmission over an AWGN channel.

However, there are plenty of practical situations where finding an accurate approximation of the error probability at the receiver output is very difficult, or sometimes even impossible.

For instance, this may be the case with a communication system using error-correcting coding, also known as channel coding or error-control coding, in association with the modulation scheme so as to improve the error performance. In practice, all modern communication schemes employ an error-correcting code.

This is also the case when the channel is not AWGN. For example, finding an accurate general expression of the error probability over a flat fading channel is very challenging because some assumptions that were made to simplify the error probability derivation over an AWGN channel are no longer valid over a fading channel.

For many of these systems, the only way to obtain an accurate approximation of the error probability at the receiver output often consists of using computer simulations.

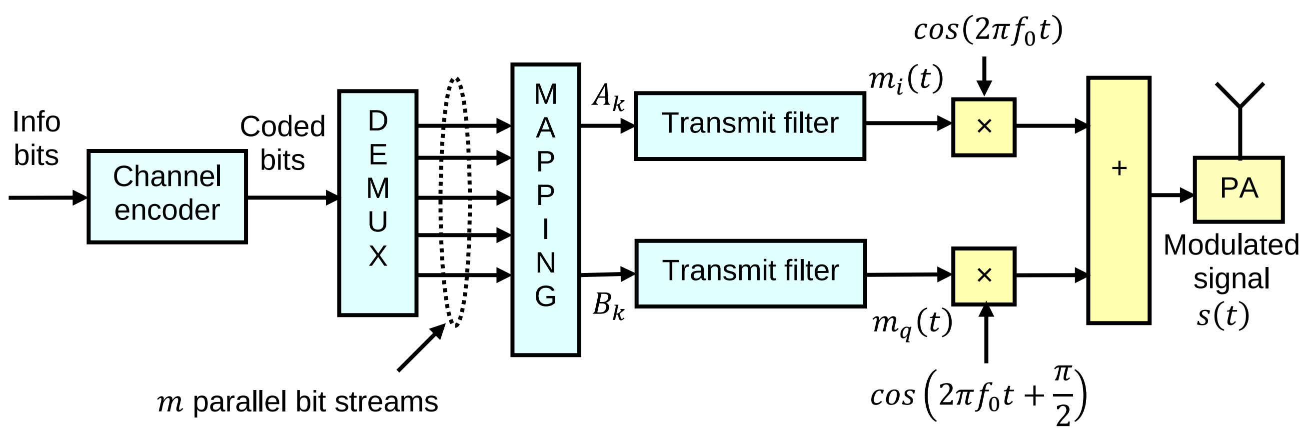
In this chapter, we are going to explain how to simulate the error performance of a communication system that is protected against transmission errors by an error-correcting code.

Channel coding is a very rich topic that is beyond the scope of this introductory module on digital communications. We will thus not be able to provide the reader with any significant details about the various principles and techniques, such as channel capacity, block codes, convolutional codes, turbo codes, and LDPC codes, that are today at the core of channel coding. However, we will still study in class the concept of convolutional codes and Viterbi decoding in order to better understand the MATLAB experiments. If some readers are interested in knowing more about channel codes, they are invited to read a separate set of lecture notes written by yours truly for another module (EEE8099 - Information Theory and Coding).

- **Coded modulation schemes – The transmitter**

We have depicted below the generic block diagram of a coded modulation scheme, i.e., a communication system combining a modulation scheme with a channel code.

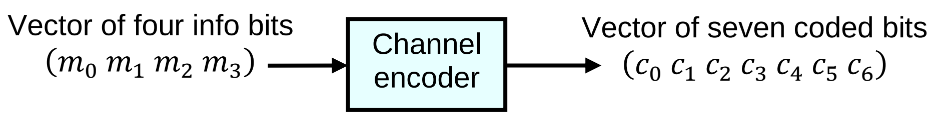
The block diagram of a transmitter with channel encoding is very similar to that of an uncoded modulation scheme. However, in coded modulation systems, the information bits are encoded by a channel encoder, which can typically be a convolutional, turbo, or low-density parity-check (LDPC) encoder. The coded bits thus generated are then transmitted using the classical transmitter structure already studied in previous chapters.



The channel encoding process basically consists of introducing some clever dependency between transmitted bits: the value of any coded bit depends on the values of the preceding and succeeding coded bits. In other words, with channel coding, only some specific binary sequences can be transmitted, whereas other sequences are forbidden. When this technique is properly implemented, it allows for the correction of most erroneous bits at the receiver side and thus decreases the bit error probability without any increase in SNR and bandwidth.

An example of channel encoder is shown in the figure below.

Example of an error-correcting encoder with a coding rate $R_c = \frac{4}{7}$



The one-to-one correspondence between a vector of four info bits and a vector of seven coded bits is defined using seven linear equations

$$c_0 = m_0$$
$$c_1 = m_1$$
$$c_2 = m_2$$
$$c_3 = m_3$$
$$c_4 = m_0 \oplus m_2 \oplus m_3$$
$$c_5 = m_0 \oplus m_1 \oplus m_2$$
$$c_6 = m_1 \oplus m_2 \oplus m_3$$

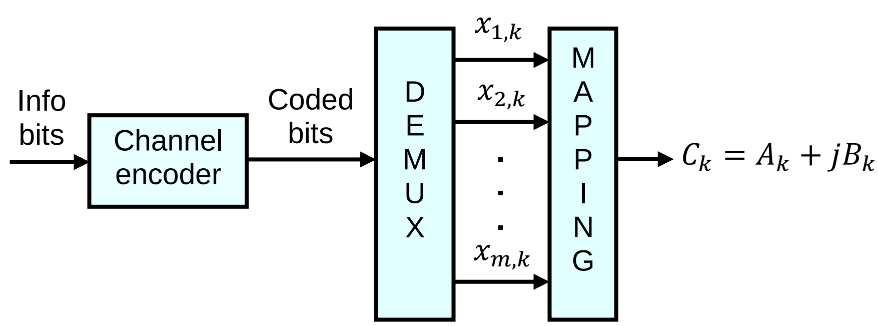
There are $2^4 = 16$ possible vectors of info bits. Thus, only 16 vectors of coded bits can be generated among the $2^7 = 128$ possible ones → Only 16 specific binary sequences can exist at the encoder output. This knowledge can be exploited at the receiver side to correct transmission errors.

In this example, the sequence of information bits is broken into successive vectors of four bits. The encoder processes each four-bit vector to generate a corresponding vector of seven coded bits. The processing simply consists of applying a set of seven linear equations given in the figure above.

The successive seven-bit vectors form the sequence of coded bits to be transmitted over the channel. One can see that the bits c_4, c_5, and c_6 entirely depend on the bits c_0, c_1, c_2, and c_3. Those three bits are thus redundant in terms of information content. Their presence however offers some degree of protection against transmission errors.

Error correction cannot be performed in an uncoded system because the latter only transmit information bits. These bits are inherently independent, which implies that the value of any of them does not depend at all on the values of the other ones. Therefore, no information bit can tell us anything about the preceding and succeeding ones. In such case, the receiver does not have the capability to correct any erroneous bit.

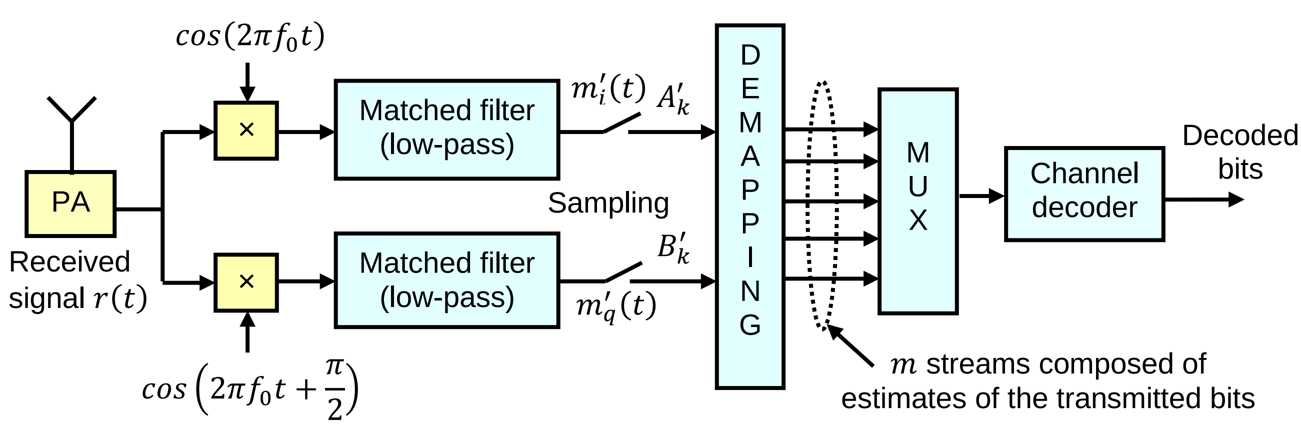
The coded bit stream is de-multiplexed into m parallel bit streams. Let $(x_{1,k}, x_{2,k}, \dots, x_{m,k})$ denote the vector of m parallel coded bits at time kT, where k is an integer and T is the duration of each transmitted symbol.



The mapping operation converts this m-bit vector $(x_{1,k}, x_{2,k}, \dots, x_{m,k})$ into a pair of real-valued symbols A_k and B_k, or equivalently a complex symbol $C_k = A_k + jB_k$ that can take $M = 2^m$ different values. Traditionally, we employ Gray mapping since such labeling technique minimizes the bit error probability for a given symbol error probability.

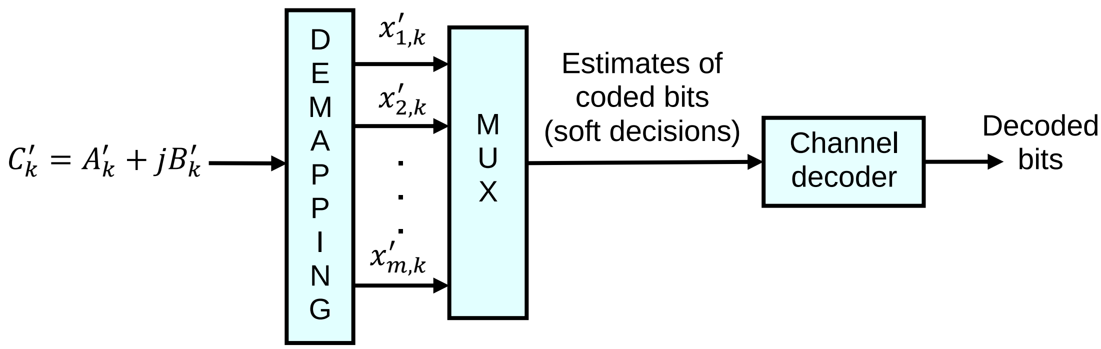
At the mapping block output, we thus have two streams composed of real-valued symbols A_k and B_k, or equivalently a single stream composed of complex symbols $C_k = A_k + jB_k$.

- **Coded modulation schemes – The receiver**



At the receiver side, an estimate $C'_k = A'_k + jB'_k$ of the corresponding transmitted complex symbol C_k is available, at each time kT, at the sampling device output.

In a coded modulation system, the sample C'_k is processed by a *de-mapping block* whose task is to convert C'_k into a vector $(x'_{1,k},\ x'_{2,k},\ \ldots,\ x'_{m,k})$ of real-valued samples that estimate the transmitted m-bit vector $\left(x_{1,k}, x_{2,k}, \ldots, x_{m,k}\right)$. Such de-mapping block does not exist in uncoded communication systems where it is replaced by a simple decision block.

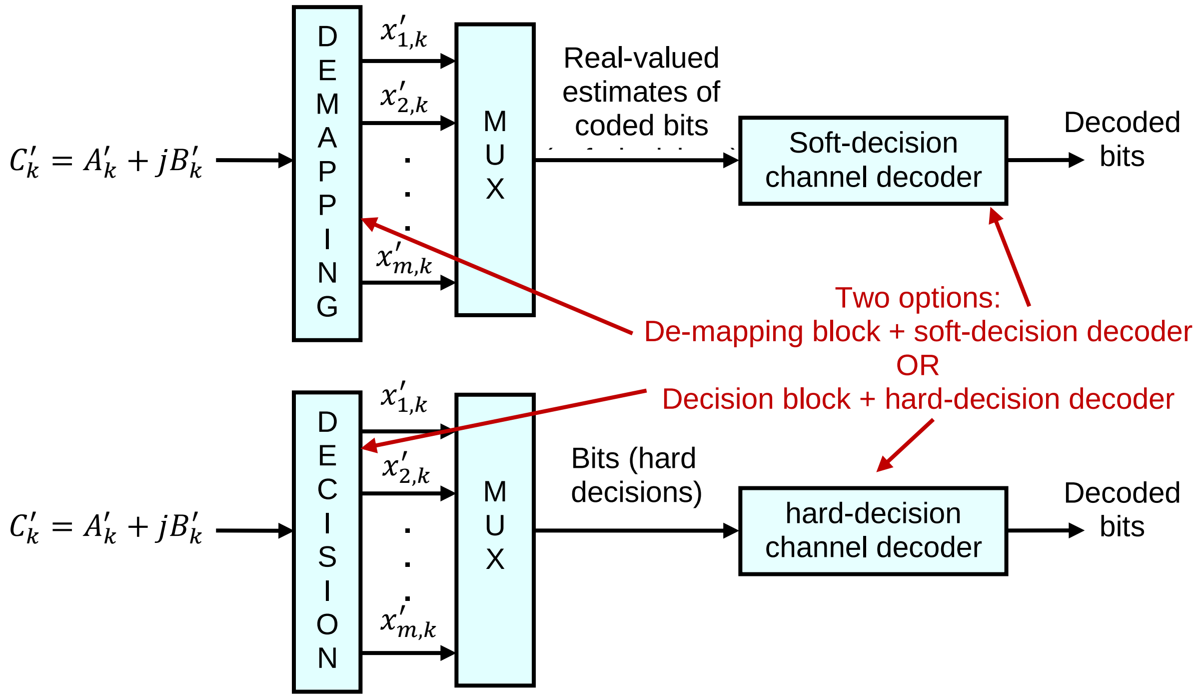


Why do we need to use a de-mapping block in coded systems?

We must first understand that, in most applications, error-correcting decoders operate using real-valued estimates of the transmitted coded bits. In other words, channel decoders must be fed with estimates of the transmitted coded bits.

These estimates are often referred to as *soft decisions* by communication engineers, and the decoders processing them are thus called *soft-decision channel decoders*.

Note that it could also be possible to feed channel decoders with bits representing the transmitted coded sequence. These bits, called hard decisions, would be generated by employing a decision block at the sample device output, exactly like in uncoded systems. The channel decoders that are designed to process hard decisions are thus referred to as hard-decision channel decoders.



The "de-mapping block + soft-decision decoder" option has much better error-correction capabilities than the "decision block + hard-decision decoder"
→ If you have the choice, always go for the "de-mapping block + soft-decision decoder" option.

The fact is that soft-decision decoders have a much better error-correction capability than their hard-decision counterparts. Therefore, they are *de facto* the preferred choice for designing communication systems.

Throughout this module, we will only consider from now on the use of soft-decision channel decoders because hard-decision decoders are, nowadays, only employed in very few applications.

The estimates of the transmitted coded bits are generally not available at the sampling device output because the latter only generates real-valued estimates A'_k and B'_k of the transmitted symbols A_k and B_k.

The task of the de-mapping block therefore consists of converting each pair of estimates A'_k and B'_k into m real-valued estimates $x'_{1,k}$, $x'_{2,k}$, …, and $x'_{m,k}$ of the coded bits $x_{1,k}, x_{2,k}, \dots,$ and $x_{m,k}$ associated with symbols A_k and B_k. These soft decisions $x'_{1,k}$, $x'_{2,k}$, …, and $x'_{m,k}$ are computed using the corresponding complex sample $C'_k = A'_k + jB'_k$.

Before proceeding further, let us drop the time index k as the time is not relevant throughout the following calculations and let us, from now on, adopt the following notations:

- $C = A + jB$ denotes the transmitted complex symbol and $(x_1, x_2, \dots, x_m)$ is the m-bit vector associated with it.

- $C' = A' + jB'$ is the corresponding sample available at the de-mapping block input. For an AWGN channel, we can write $A' = A + n_1$ and $B' = B + n_2$, where n_1 and n_2 designate two independent Gaussian noise samples.

- $(x'_1, x'_2, \dots, x'_m)$ is the vector of samples produced by the de-mapping block. Each sample x'_i, $i \in \{1, 2, \dots, m\}$, is an estimate of the transmitted coded bit x_i.

The estimates x'_i are real numbers and they can be positive or negative. The closer to zero the value of x'_i is, the less certain the receiver is regarding the value of the corresponding transmitted bit x_i. In the case where $x'_i = 0$, the receiver actually informs the channel decoder that it does not have any idea about the value of x_i. On the other hand, the farthest from zero the value of x'_i is, the more certain the receiver is regarding the value of the corresponding transmitted bit x_i.

Over an AWGN channel, the estimate x'_i, $i \in \{1, 2, \dots, m\}$, of the coded bit x_i is computed using the generic expression

$$x'_i = \frac{\sigma^2}{2} \ln\left(\frac{\Pr\{x_i = 0 | C'\}}{\Pr\{x_i = 1 | C'\}}\right),$$

where σ^2 denotes the variance of the noise samples n_1 and n_2.

After a few mathematical manipulations, we obtain

$$x_i' = \frac{\sigma^2}{2}\ln\left(\frac{\sum_{C\in\Delta_{0,i}}\exp\left\{-\frac{d^2(C',C)}{2\sigma^2}\right\}}{\sum_{C\in\Delta_{1,i}}\exp\left\{-\frac{d^2(C',C)}{2\sigma^2}\right\}}\right),$$

where $d^2(C',C) = (A'-A)^2 + (B'-B)^2$ designates the square of the Euclidean distance between the sample C' and a particular symbol C in the constellation, whereas $\Delta_{j,i}$, $j = 0$ or 1 and $i \in \{1, 2, \ldots, m\}$, represents the set of all symbols C whose mapping label is j in the i-th position.

In this expression, we assume that the channel decoder uses the following convention: a negative estimate x_i' is indicative of a coded bit equal to one, whereas a positive estimate x_i' is indicative of a coded bit equal to zero. We adopt this convention here as it is the one used by the Viterbi decoder that you will consider in your MATLAB programs.

For example, a soft decision $x_i' = 0.1$ indicates that the receiver believes that the coded bit x_i is positive, i.e., equal to zero with the convention defined above. However, the fact that x_i' is in fact *hardly positive* also indicates that the belief of the receiver is very weak, and the coded bit x_i is in fact almost as likely to be a one as a zero. This is very precious information that the soft-decision channel decoder can exploit to provide optimal error correction.

Let us consider another example. An estimate $x_i' = -3.2$ indicates that the receiver believes that the coded bit x_i is equal to one. The fact that x_i' is *very negative* also indicates that the belief of the receiver is very strong, meaning that the coded bit x_i is much more likely to be a one than a zero. Once again, this is very precious information that can be exploited by the channel decoder.

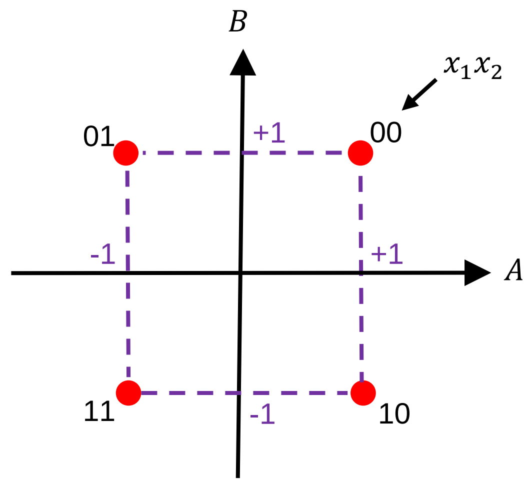
As an illustration, consider the very simple case of BPSK modulation with the following mapping: $x_1 = 0 \to A = +1$ and $x_1 = 1 \to A = -1$. The de-mapper output is given by

$$x_1' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-1)^2}{2\sigma^2} + \frac{(A'+1)^2}{2\sigma^2}\right) = A',$$

This result shows that, for BPSK, no de-mapping block is in fact needed since the estimate x_1' of the bit x_1 is already available at the sampling device output.

Example ---

Consider the 4-QAM constellation with Gray mapping shown below. Find the two de-mapping equations for this modulation scheme.



The sets $\Delta_{j,i}$, $j = 0$ or 1 and $i \in \{1, 2\}$, are defined as follows:
For the coded bit x_1: $\Delta_{0,1} = \{+1 + j, -1 + j\}$ and $\Delta_{1,1} = \{+1 - j, -1 - j\}$;
For the coded bit x_2: $\Delta_{0,2} = \{+1 + j, +1 - j\}$ and $\Delta_{1,2} = \{-1 + j, -1 - j\}$.

The computation of the first estimate x_1' is thus done as follows:

$$x_1' = \frac{\sigma^2}{2} \ln\left(\frac{\exp\left\{-\frac{(A'-1)^2+(B'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2+(B'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'-1)^2+(B'+1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2+(B'+1)^2}{2\sigma^2}\right\}}\right),$$

which can be written as

$$x_1' = \frac{\sigma^2}{2} \ln\left(\frac{\exp\left\{-\frac{(B'-1)^2}{2\sigma^2}\right\}\cdot\left[\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}\right]}{\exp\left\{-\frac{(B'+1)^2}{2\sigma^2}\right\}\cdot\left[\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}\right]}\right),$$

which finally leads to

$$x_1' = \frac{\sigma^2}{2} \ln\left(\frac{\exp\left\{-\frac{(B'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(B'+1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(B'-1)^2}{2\sigma^2} + \frac{(B'+1)^2}{2\sigma^2}\right) = B'.$$

The computation of the second estimate x_2' is performed as follows:

$$x_2' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-1)^2+(B'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'-1)^2+(B'+1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+1)^2+(B'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2+(B'+1)^2}{2\sigma^2}\right\}}\right),$$

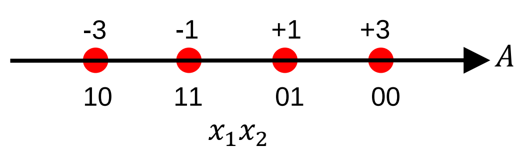
which can be written as
$$x_2' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}\cdot\left[\exp\left\{-\frac{(B'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(B'+1)^2}{2\sigma^2}\right\}\right]}{\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}\cdot\left[\exp\left\{-\frac{(B'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(B'+1)^2}{2\sigma^2}\right\}\right]}\right),$$

thus leading to
$$x_2' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-1)^2}{2\sigma^2}+\frac{(A'+1)^2}{2\sigma^2}\right) = A'.$$

These results clearly show that, for 4-QAM, no de-mapping block is needed since the estimates of both bits x_1 and x_2 are already available at the sampling device output.

Example --

Consider the 4-ASK constellation with Gray mapping depicted below. Find the two de-mapping equations for this modulation scheme.



The sets $\Delta_{j,i}$, $j = 0$ or 1 and $i \in \{1, 2\}$, are defined as follows:

- For the coded bit x_1: $\Delta_{0,1} = \{+1, +3\}$ and $\Delta_{1,1} = \{-3, -1\}$,
- For the coded bit x_2: $\Delta_{0,2} = \{-3, +3\}$ and $\Delta_{1,2} = \{-1, +1\}$.

The first de-mapping expression is given by

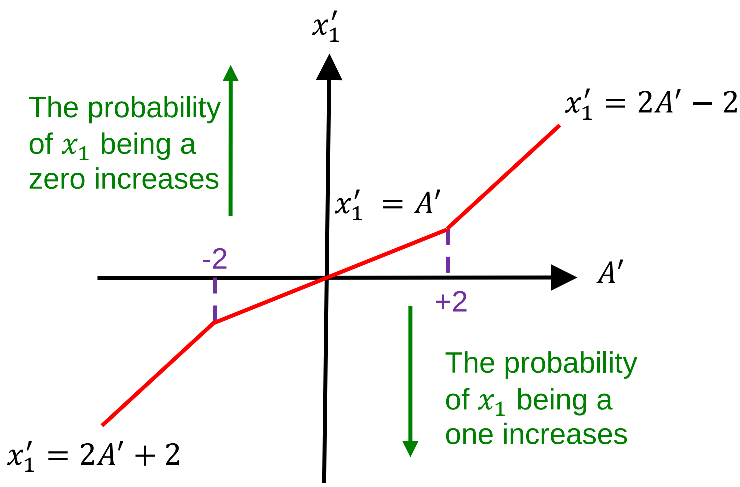
$$x_1' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+3)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right).$$

This expression is quite complicated, and it is thus worth simplifying it. This can be done as follows:

If $A' > +2$, then $x_1' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-3)^2}{2\sigma^2} + \frac{(A'+1)^2}{2\sigma^2}\right) = 2A' - 2.$

If $|A'| \leq +2$, then $x_1' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-1)^2}{2\sigma^2} + \frac{(A'+1)^2}{2\sigma^2}\right) = A'.$

If $A' < -2$, then $x_1' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+3)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-1)^2}{2\sigma^2} + \frac{(A'+3)^2}{2\sigma^2}\right) = 2A' + 2.$



The second de-mapping expression is given by

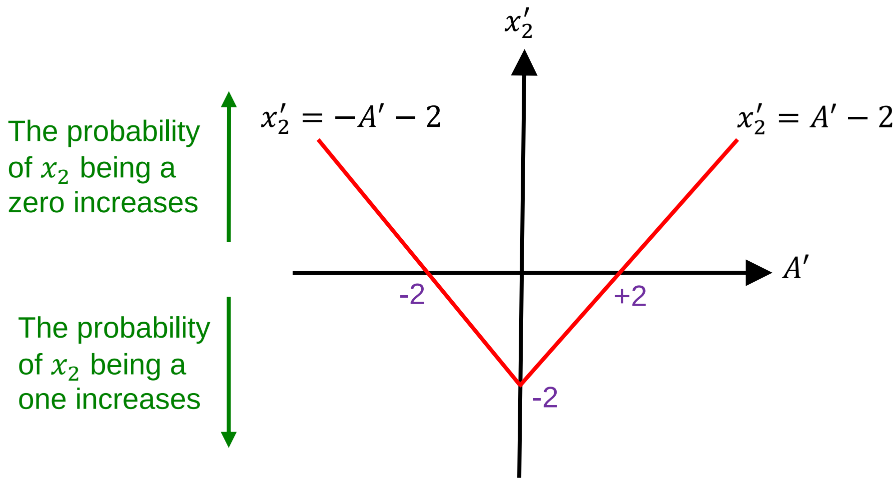
$$x_2' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+3)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right).$$

This expression can be simplified as follows:

If $A' > 0$, then $x_2' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'-1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-3)^2}{2\sigma^2} + \frac{(A'-1)^2}{2\sigma^2}\right) = A' - 2.$

If $A' < 0$, then $x_2' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'+3)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+1)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'+3)^2}{2\sigma^2} + \frac{(A'+1)^2}{2\sigma^2}\right) = -A' - 2.$

The last two equations can be combined to form the following expression, applicable for all values of the received sample A': $x_2' \sim |A'| - 2$.



- **De-mapping equations for a flat fading channel with perfect channel state information**

Recall that the estimate $C' = A' + jB'$ is the complex sample available at the de-mapping block input. For a flat fading channel, we can write $A' = hA + n_1$ and $B' = hB + n_2$, where h is a fading sample that follows either a Rice-Nakagami distribution or a Rayleigh distribution, whereas n_1 and n_2 designate two independent Gaussian noise samples.

Over a flat fading channel with perfect channel state information (CSI), the de-mapping expressions must be different from those employed over an AWGN channel. Those expressions must take into account the fact that the fading sample values are known at the receiver side. In other words, their knowledge can be exploited by the de-mapping block in order to achieve optimal error performance.

Over a flat fading channel with perfect CSI, the estimate x_i', $i \in \{1, 2, \ldots, m\}$, of the coded bit x_i is computed using the generic expression

$$x_i' = \frac{\sigma^2}{2} \ln\left(\frac{\Pr\{x_i = 0 | h, C'\}}{\Pr\{x_i = 1 | h, C'\}}\right),$$

where σ^2 denotes the variance of the noise samples n_1 and n_2.

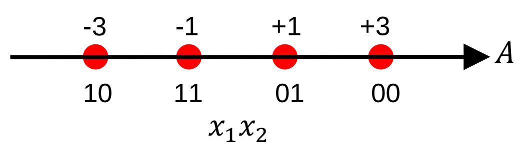
After a few mathematical manipulations, we obtain

$$x_i' = \frac{\sigma^2}{2}\ln\left(\frac{\sum_{C\in\Delta_{0,i}}\exp\left\{-\frac{d^2(C',hC)}{2\sigma^2}\right\}}{\sum_{C\in\Delta_{1,i}}\exp\left\{-\frac{d^2(C',hC)}{2\sigma^2}\right\}}\right),$$

where $d^2(C',hC) = (A'-hA)^2 + (B'-hB)^2$ designates the square of the Euclidean distance between the sample C' and a particular symbol hC in the constellation, whereas $\Delta_{j,i}$, $j = 0$ or 1 and $i \in \{1,2,\ldots,m\}$, represents the set of all symbols C whose mapping label is j in the i-th position.

Example ---

Consider the 4-ASK constellation with Gray mapping depicted below. Find the two de-mapping equations for this modulation scheme.



The sets $\Delta_{j,i}$, $j = 0$ or 1 and $i \in \{1,2\}$, are defined as follows:

- For the coded bit x_1: $\Delta_{0,1} = \{+1,+3\}$ and $\Delta_{1,1} = \{-3,-1\}$,
- For the coded bit x_2: $\Delta_{0,2} = \{-3,+3\}$ and $\Delta_{1,2} = \{-1,+1\}$.

The first de-mapping expression is given by

$$x_1' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3h)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'-h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+3h)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+h)^2}{2\sigma^2}\right\}}\right).$$

This expression is quite complicated, and it is thus worth simplifying it. This can be done as follows:

If $A' > +2h$, then $x_1' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+h)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-3h)^2}{2\sigma^2} + \frac{(A'+h)^2}{2\sigma^2}\right) = 2h(A'-h).$

If $|A'| \leq +2h$, then $x_1' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+h)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-h)^2}{2\sigma^2} + \frac{(A'+h)^2}{2\sigma^2}\right) = hA'$.

If $A' < -2h$, then $x_1' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+3h)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-h)^2}{2\sigma^2} + \frac{(A'+3h)^2}{2\sigma^2}\right) = 2h(A'+h)$.

The second de-mapping expression is given by

$$x_2' = \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3h)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+3h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'-h)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(A'+h)^2}{2\sigma^2}\right\}}\right).$$

This expression can be simplified as follows:

If $A' > 0$, then $x_2' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'-3h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'-h)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'-3h)^2}{2\sigma^2} + \frac{(A'-h)^2}{2\sigma^2}\right) = h(A'-2h)$.

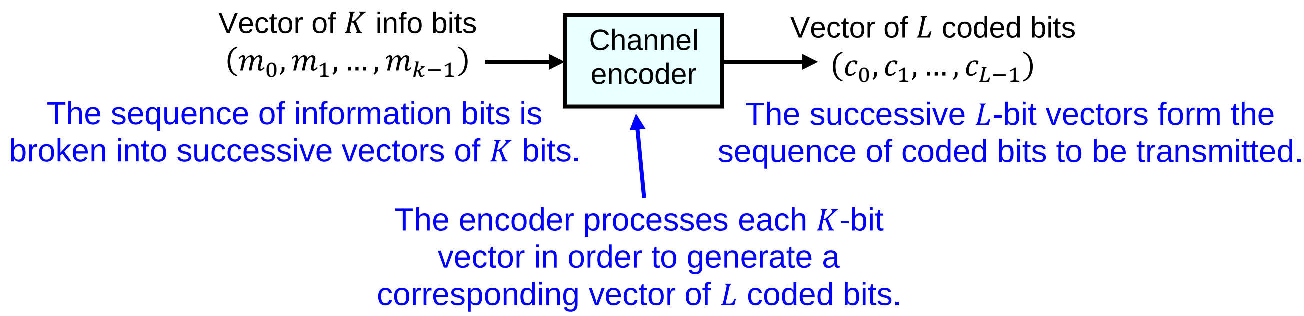
If $A' < 0$, then $x_2' \sim \frac{\sigma^2}{2}\ln\left(\frac{\exp\left\{-\frac{(A'+3h)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(A'+h)^2}{2\sigma^2}\right\}}\right) = \frac{\sigma^2}{2}\left(-\frac{(A'+3h)^2}{2\sigma^2} + \frac{(A'+h)^2}{2\sigma^2}\right) = h(-A'-2h)$.

The last two equations can be combined to form the following expression, applicable for all values of the received sample A': $x_2' \sim h(|A'| - 2h)$.

- **Computer simulations of coded modulation schemes**

Computer simulations can be used to assess the error performance of a coded modulation scheme with great accuracy. The first stage consists of generating a stream of N information bits with $\Pr\{0\} = \Pr\{1\} = \frac{1}{2}$. Then, this stream is encoded by the channel encoder.

At the encoder output, the length of the corresponding binary sequence is $\frac{N}{R_c}$ coded bits, where R_c is the coding rate of the error-correcting code. The coding rate of a channel code is the ratio between the number of information bits and the number of coded bits generated by encoding these information bits. The coding rate is always less than the unit because each information bit generates more than one coded bit. So we always have $R_c < 1$.



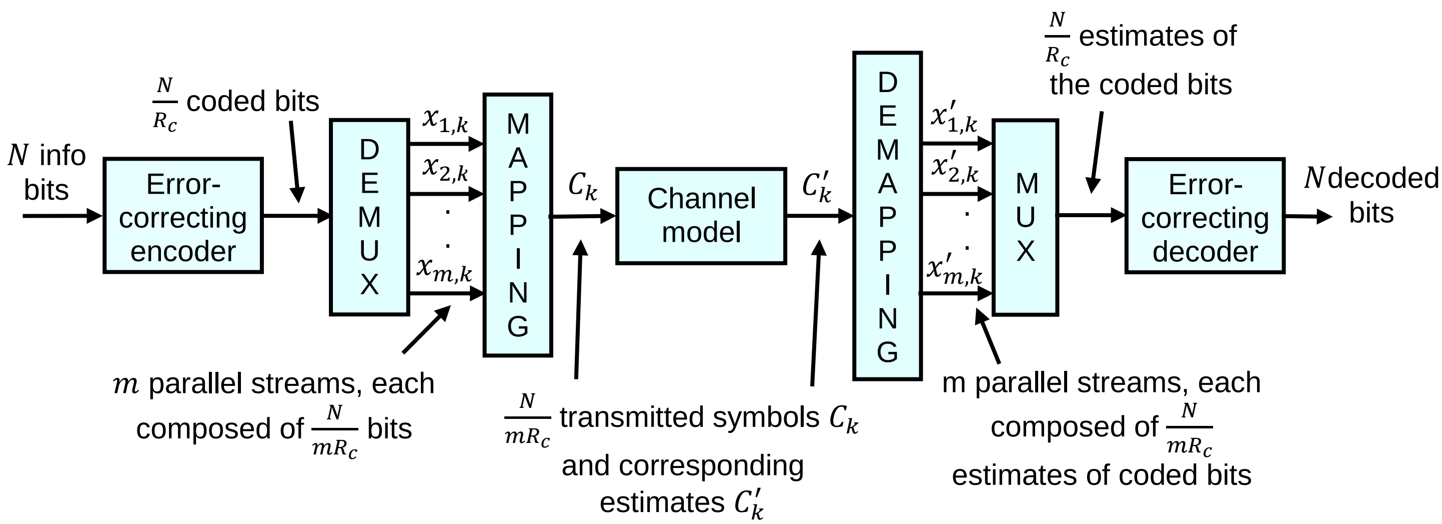
A sequence of N info bits is converted into a sequence of $\frac{N}{R_c}$ coded bits, where $R_c = \frac{K}{L}$ designates the coding rate. The fact that $R_c < 1$ implies that a coded system must transmit more bits than its uncoded counterpart. Those redundant bits are used at the receiver side to detect and then correct the

After de-multiplexing of the coded bit stream, the k–th vector composed of m coded bits is mapped onto a constellation symbol C_k, i.e., a pair of real-valued symbols A_k and B_k.

Note that each transmitted constellation symbol C_k carries the information corresponding to m coded bits, i.e., only mR_c information bits. Thus, the spectral efficiency of the whole communication system is no longer m bits/s/Hz as in the uncoded case, but instead it is reduced to mR_c bits/s/Hz.

The use of channel coding thus reduces the spectral efficiency by a factor equal to the coding rate of the channel code. That is the main drawback of coded systems. In fact, this is the price to pay for an improved bit error performance.

This is a crucial point that must always be considered when comparing the error performance of different coded modulation systems. Error performance comparisons between different communication systems are meaningful only when all schemes have identical spectral efficiencies.



Then, the transmission of constellation symbols $C_k = A_k + jB_k$ over the channel is simulated by computing the channel outputs A'_k and B'_k as a function of A_k and B_k.

We often consider the following channel model: $A'_k = h_k A_k + n_{1,k}$ and $B'_k = h_k B_k + n_{2,k}$, where $n_{1,k}$ and $n_{2,k}$ are two independent zero-mean Gaussian noise samples with a variance $\sigma^2 = \frac{\gamma}{2}\left(\frac{E_s}{N_0}\right)^{-1}$.

For an AWGN channel, we simply assume that $h_k = 1, \forall k$, whereas, for a (normalized) flat Rayleigh fading channel, we assume that the quantity h_k follows a Rayleigh distribution. In this case, the sample h_k is generated using the expression

$$h_k = \sqrt{w_{k,1}^2 + w_{k,2}^2},$$

where $w_{k,1}$ and $w_{k,2}$ denote two independent zero-mean Gaussian samples with a variance $\sigma_k^2 = \frac{1}{2}$.

Since the error performance of the system is ultimately a function of the SNR per information bit, $\frac{E_b}{N_0}$, we need to specify in any computer simulation program the relation between σ^2 and $\frac{E_b}{N_0}$ which is given by

$$\sigma^2 = \frac{\gamma}{2}\left(\frac{E_s}{N_0}\right)^{-1} = \frac{\gamma}{2}\left(\frac{mR_cE_b}{N_0}\right)^{-1} = \frac{\gamma}{2mR_c}\left(\frac{E_b}{N_0}\right)^{-1},$$

since, in a coded modulation system, each complex symbol C_k carries mR_c information bits.

If the SNR is expressed in decibels (dB), we use the following equation instead:

$$\sigma^2 = \frac{\gamma}{2mR_c}10^{-\frac{1}{10}\cdot\left(\frac{E_b}{N_0}\right)_{\mathrm{dB}}}.$$

Example 1 --

Coded 4-QAM combined to a channel code with a coding rate $R_c = \frac{1}{2}$ (spectral efficiency $\eta = 1$ bit/s/Hz). The constellation symbols are such that $C_k \in \{+1+j, -1+j, -1-j, +1-j\}$. We have $\gamma = 2$ and $m = 2$, which yields $\sigma^2 = 10^{-0.1\cdot\left(\frac{E_b}{N_0}\right)_{\mathrm{dB}}}$.

Example 2 --

Coded QPSK combined to a channel code with $R_c = \frac{1}{2}$ ($\eta = 1$ bit/s/Hz). The symbols A_k and B_k are located on the unit-energy circle: $C_k \in \{+1, +j, -1, -j\}$. We have $\gamma = 1$ and $m = 2$, which yields $\sigma^2 = \frac{1}{2} \times 10^{-0.1\cdot\left(\frac{E_b}{N_0}\right)_{\mathrm{dB}}}$.

Example 3 --

Uncoded 4-QAM ($\eta = 2$ bits/s/Hz) with symbols $C_k \in \{+1+j, -1+j, -1-j, +1-j\}$. We have $R_c = 1$, $\gamma = 2$ and $m = 2$, which yields $\sigma^2 = \frac{1}{2} \times 10^{-0.1\cdot\left(\frac{E_b}{N_0}\right)_{\mathrm{dB}}}$.

Example 4 --

Coded 8-PSK combined to a channel code with $R_c = \frac{2}{3}$ ($\eta = 2$ bits/s/Hz). The eight symbols are located on the unit-energy circle: $C_k \in \left\{+1, \frac{+1+j}{\sqrt{2}}, +j, \frac{-1+j}{\sqrt{2}}, -1, \frac{-1-j}{\sqrt{2}}, -j, \frac{+1-j}{\sqrt{2}}\right\}$. We have $\gamma = 1$ and $m = 3$, which yields $\sigma^2 = \frac{1}{4} \times 10^{-0.1 \cdot \left(\frac{E_b}{N_0}\right)_{\text{dB}}}$.

Example 5 ---

Uncoded 8-PSK ($\eta = 3$ bits/s/Hz) with $C_k \in \left\{+1, \frac{+1+j}{\sqrt{2}}, +j, \frac{-1+j}{\sqrt{2}}, -1, \frac{-1-j}{\sqrt{2}}, -j, \frac{+1-j}{\sqrt{2}}\right\}$. We have $R_c = 1$, $\gamma = 1$ and $m = 3$, which yields $\sigma^2 = \frac{1}{6} \times 10^{-0.1 \cdot \left(\frac{E_b}{N_0}\right)_{\text{dB}}}$.

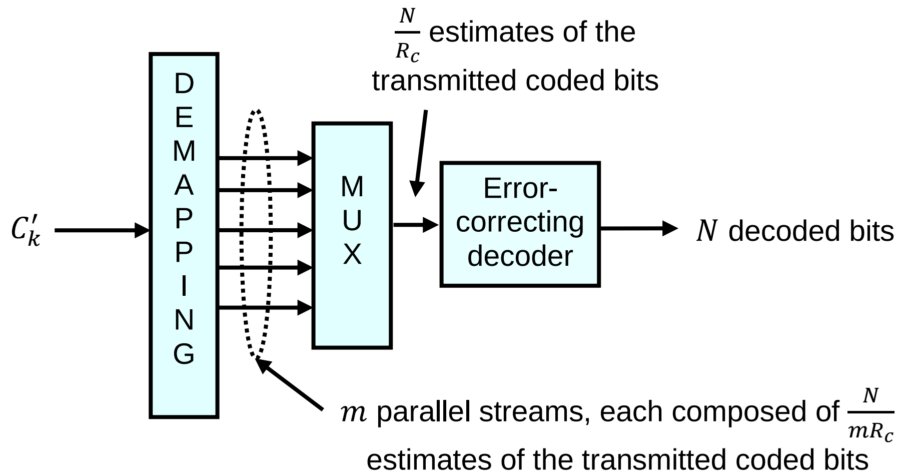
Example 6 ---

Coded 16-QAM combined to an error-correcting code with $R_c = \frac{3}{4}$ ($\eta = 3$ bits/s/Hz). The constellation symbols are such as $A_k, B_k \in \{-3, -1, +1, +3\}$. We have $\gamma = 10$ and $m = 4$, which yields $\sigma^2 = \frac{5}{3} \times 10^{-0.1 \cdot \left(\frac{E_b}{N_0}\right)_{\text{dB}}}$.

Example 7 ---

Uncoded 16-QAM (η = 4 bits/s/Hz) with $A_k, B_k \in \{-3, -1, +1, +3\}$. We have $R_c = 1$, $\gamma = 10$ and $m = 4$, which yields $\sigma^2 = \frac{5}{4} \times 10^{-0.1 \cdot \left(\frac{E_b}{N_0}\right)_{\text{dB}}}$.

At the receiver side, the k-th channel estimate $C'_k = A'_k + jB'_k$ is processed by the de-mapping block so as to produce the k-th vector composed of m real-valued samples $x'_{i,k}$, $i \in \{1, 2, \dots, m\}$. These samples estimate the m coded bits $x_{i,k}, i \in \{1, 2, \dots, m\}$, that were associated with the transmitted complex symbol $C_k = A_k + jB_k$ at the transmitter side. Finally, after multiplexing, the samples are fed into the soft-decision channel decoder.



At this stage, we can compute the bit error rate (BER) which is a measure of the bit error probability P_{eb}. To do so, the sequence of information bits is compared with the decoded sequence: an error event is counted every time a decoded bit is different from the corresponding transmitted information bit. Finally, the BER is obtained by dividing the total number of errors by the number N of transmitted information bits.

- END -