3 Pulse Shaping and Matched Filtering

Up to now we have only considered the transmission of complex data symbols over an AWGN channel. Of course, in reality there is no such thing as a complex symbol coming out of the antenna. Instead, the symbols have to be transformed into a time-continuous waveform. This transformation is called *modulation*.

3.1 Modulation

3.1.1 Transmission at Radio Frequency (RF)

The principle of digital communication systems is that only a finite number M of different symbols can be transmitted. Each symbol a_i is mapped onto a pulse $g_i(t)$, $i = 1, \ldots, M$. Since only finitely many pulses are possible, the receiver can recover the transmitted data if the signal is not too heavily distorted by the channel.

In this lab course, we only consider *linear* modulation schemes, where the pulses $g_i(t)$ can be expressed as scaled versions of the same base pulse: $g_i(t) = a_i g(t)$. Hence, the magnitude of the data symbol determines the amplitude of the transmitted pulse, and the argument determines its phase. (An example of a *nonlinear* modulation scheme is frequency shift keying, where the information lies in the frequency of the transmitted signal.)

Let T denote the duration of one data symbol, i.e., a new pulse is transmitted each T seconds. Note that the pulse $g(t)$ can still be longer than T; the pulses are allowed to overlap. The signal $s(t)$ can then be written as the superposition of the pulses, scaled by the data symbols and delayed by multiples of T:

$$s(t) = \sum_{n=-\infty}^{\infty} a[n]g(t - nT). \tag{3.1}$$

The signal $s(t)$ is called a complex baseband signal, because it is complex-valued and its spectrum is centered at $f = 0$. In order to transmit the signal over the air, it must be transformed into a real-valued signal and its spectrum must be shifted to a carrier frequency $f_c \gg B$, where
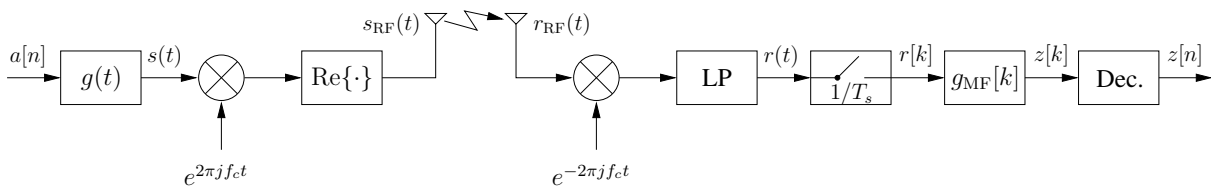

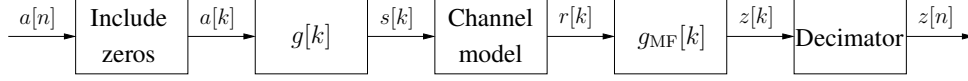
Figure 3.1: Block diagram of a digital transmission system

Figure 3.2: Equivalent baseband system

B is the bandwidth of the signal. This operation can be written as follows:

$$\begin{aligned} s_{\mathrm{RF}}(t) &= \mathrm{Re}\{s(t)e^{2\pi j f_c t}\} \\ &= \mathrm{Re}\{s(t)\}\cos(2\pi f_c t) - \mathrm{Im}\{s(t)\}\sin(2\pi f_c t). \end{aligned} \tag{3.2}$$

The real part of the signal is modulated onto a cosine wave, and the imaginary part onto a sine wave. Due to the orthogonality of sin and cos, the two signals can be separated by the receiver.

The receiver converts the received signal $r_{\mathrm{RF}}(r)$ back into the baseband by multiplying it with $e^{-2\pi j f_c t}$ and filtering the result with a lowpass filter in order to eliminate the spectral components around $2f_c$. The baseband signal $r(t)$ is then sampled with the sampling rate $1/T_s$, so that the remaining signal processing can be done by digital processors. The sampling duration is chosen to be shorter than the symbol duration, $T_s = T/L$, where L is the oversampling factor. We will explain later why the signal must be oversampled. The remaining task of the receiver is now to recover the transmitted data by means of the received data samples $r[k]$. (We use the time index k in order to denote samples of duration T_s, and the index n for data symbols of duration T).

3.1.2 Equivalent Baseband Model

We have seen that a description of the received samples is quite inconvenient if the digital/analog conversion and the upconversion to radio frequency is taken into account. Fortunately, it is possible to describe the whole system as a time-discrete model in the complex baseband (as long as one is not interested in the effects of the RF components, of course).

A block diagram of this simplified model is shown in Figure 3.2. First, the symbol stream is converted to sample rate by inserting $L - 1$ zeros between the symbols. (This implies that L is an integer, which can generally be assumed. In the lab, we use an oversampling factor of $L = 4$.) The sequence $a[k]$ is then processed by the pulse shaping filter $g[k]$. The time-discrete transmitted signal $s[k]$ is then transmitted over a time-discrete baseband channel and filtered by the receiver filter $g_{\mathrm{MF}}[k]$. Now the sample stream can be converted back to symbol rate by simply discarding three out of four samples. If the channel is an AWGN channel, which we assume up to now, the received symbols can then be written as $z[n] = a[n] + w'[n]$, where $w'[n]$ is the filtered noise.

3.2 Pulse Shaping Filters

3.2.1 Ideal Lowpass

We have not yet discussed the question how the pulse $g(t)$ (or its time discrete version $g[k] = g(kT_s)$) should look like. In order to use the available spectrum as efficiently as possible, we would like to use a pulse whose spectrum is as narrow as possible. In order to transmit at a symbol rate $R = 1/T$, the bandwidth B of the signal must be at least as large as R. The

optimum filter would therefore be an ideal lowpass with the spectrum $G(f) = \text{rect}(f/R)$, whose impulse response is known to be $g(t) = \text{si}(\pi t/T)$, with $\text{si}(x) = \sin(x)/x$.

With an ideal lowpass, the transmitted signal can be written as

$$s(t) = \sum_{n=-\infty}^{\infty} a[n]\,\text{si}\left(\pi\frac{t-nT}{T}\right). \tag{3.3}$$

It is important to note that $g(t)$ becomes zero at multiples of the symbol duration T, except of $t = 0$:

$$g(nT) = 0 \quad \forall n \neq 0. \tag{3.4}$$

This is called the Nyquist criterion. If we sample this signal at the correct time instances $t = nT$, only the n-th pulse contributes to this sample, and we have recovered the transmitted symbols (except of noise).

The ideal lowpass however suffers from the drawback that the impulse response converges quite slowly towards zero. If we want to implement $g[k]$ as a digital *finite impulse response* (FIR) filter, we need to implement many filter taps before the impulse response can be considered to be zero. (Strictly speaking, the impulse response of the ideal lowpass is infinitely long, so we can always only approximate its behavior). Another problem is that we must sample the signal exactly at the ideal time instants. If we have a slight timing error, as will always be the case in practical systems, the contribution of the other pulses is not zero anymore and our received symbols are distorted by *inter-symbol interference* (ISI). (The timing problem will be discussed in the next chapter.) Due to these two reasons, we are looking for a filter that still fulfills the Nyquist criterion, but whose impulse response converges to zero more rapidly.

3.2.2   Raised Cosine Filter

Such a filter is the *raised cosine* (RC). Its spectrum is given as

$$G_{\text{RC}}(f) = \begin{cases} 1 & |f| \leq \frac{R}{2}(1-\alpha) \\ \frac{1}{2} + \frac{1}{2}\cos\left(\frac{\pi}{2\alpha}\left(\frac{|f|}{R/2} - 1 + \alpha\right)\right) & \frac{R}{2}(1-\alpha) < |f| < \frac{R}{2}(1+\alpha) \\ 0 & |f| \geq \frac{R}{2}(1+\alpha) \end{cases} \tag{3.5}$$

The parameter α is the rolloff factor and determines the filter bandwidth. The two-sided bandwidth of the RC is $B = R(1+\alpha)$. Thus, the bandwidth of the RC is by a fraction of α larger than that of the ideal lowpass. Figure 3.3 shows the transfer function and the impulse response of the raised cosine filter for $\alpha \in \{0.2, 0.5, 0.8\}$. We see that the bandwidth increases with increasing rolloff factor, but the impulse response converges faster to zero. For $\alpha = 0$, the raised cosine becomes an ideal lowpass. Note that the raised cosine fulfills the Nyquist criterion, i.e. becomes zero at multiples of the symbol duration.

3.2.3   Matched Filtering and Root Raised Cosine Filter

In the receiver, the signal is filtered by another filter. It can be shown that in case of an AWGN channel, the SNR after the receive filter is maximized if this filter is the complex conjugated and reverted (in time direction) transmit filter [2]:

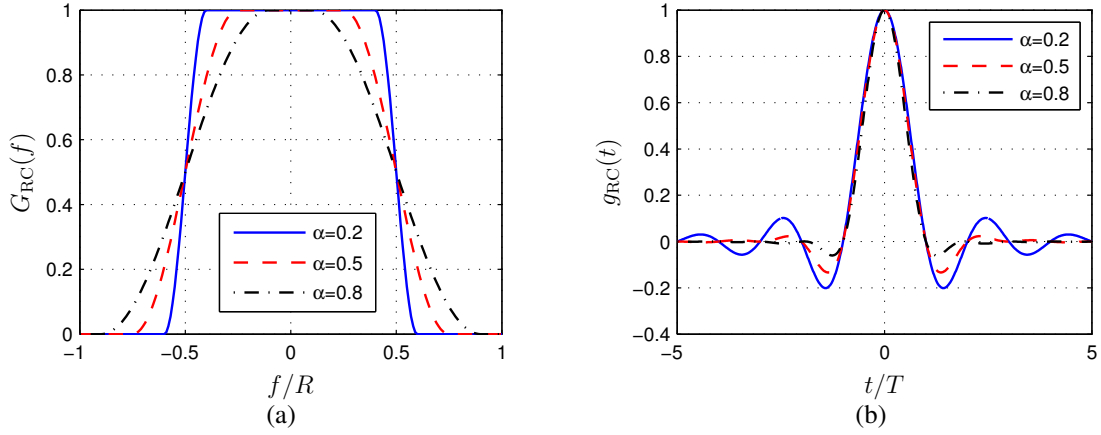$$g_{\text{MF}}(t) = g^*(-t). \tag{3.6}$$

Figure 3.3: Transfer function (a) and impulse response (b) of the raised cosine filter

This filter is matched to the transmit filter and is therefore called the *matched filter* (MF). The total effective filter of the transmission system is then the combination of transmit and receive filter $g(t) * g_{\mathrm{MF}}(t)$, where $*$ denotes convolution. It is important to note that this effective filter (and not the individual filters) must fulfill the Nyquist criterion. We can achieve this if both filters have a transfer function that is equal to the square root of that of the raised cosine filter. Such a filter is therefore called a *root raised cosine* (RRC). The combination of both RRC filters then becomes a raised cosine and thus fulfills the Nyquist criterion. Furthermore, since the filters are real-valued and symmetric, the RRC is its own matched filter.

The impulse response of the RRC filter is

$$g_{\mathrm{RRC}}(t) = \frac{\frac{4\alpha}{\pi}\cos\left(\pi(1+\alpha)\frac{t}{T}\right) + (1-\alpha)\,\mathrm{si}\left(\pi(1-\alpha)\frac{t}{T}\right)}{1 - \left(4\alpha\frac{t}{T}\right)^2}. \tag{3.7}$$

Note that $\mathrm{si}(\pi x) = \mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$.

3.2.4 Summary

When we use the RRC both as pulse shaping filter and matched filter, the effective filter fulfills the Nyquist criterion (3.4) and makes ISI-free reception possible. Also, the RRC converges faster to zero than an ideal lowpass and therefore suffers less from the drawbacks of the ideal lowpass discussed above.

The output of the matched filter can be written as

$$z(t) = \sum_{n=-\infty}^{\infty} a[n]g_{\mathrm{RC}}(t - nT) + w'(t) \tag{3.8}$$

or in time-discrete samples as

$$z[k] = \sum_{n=-\infty}^{\infty} a[n]g_{\mathrm{RC}}[k - nL] + w'[k], \tag{3.9}$$

where the noise term $w'[k]$ denotes the AWGN $w[k]$ after the matched filter. Due to (3.4), every L-th sample corresponds to a data symbol, $z[k = nL] = g_{RC}[0]\,a[n]$ (in the following we assume that the filters are normalized so that $g_{RC}[0] = 1$) and hence the transmitted data can easily be recovered.

Unfortunately, (3.9) only holds if the receiver is perfectly synchronized to the incoming signal, which is never the case in a real system. In the next chapters, we will therefore examine some synchronization algorithms.

3.3   Your Tasks

1. For the first task we provide an already shaped signal on the website. The system model is slightly different as we now use a oversampling factor of $L = 4$. The pulses are RRC shaped with a roll-off factor of $\alpha = 0.22$. Add an AWGN to the system. Afterwards write a matched filter (MF) to demodulate the pulses. Demap the symbols and decode the image. Keep the number of receive filter taps as a variable. The frame synchronization that you implemented in the previous exercise **can not** be used with oversampled signals. For this reason, a modified frame synchronization function is provided on the website.

2. Now implement a complete baseband transmission system to measure the BER:

    i.) Generate a random bitstream.

    ii.) Convert to QPSK symbols.

    iii.) Upsample by a factor $L = 4$.

    iv.) Shape RRC pulses with a roll-off factor of $\alpha = 0.22$ and 41 filter taps.

    v.) Add AWGN.

    vi.) Apply matched filter (MF) with a variable number of filter taps.

    vii.) Revert upsampling (i.e., downsample by a factor $L = 4$).

    viii.) Demap the symbol stream.

    ix.) Calculate the BER.

    A longer FIR receive filter (i.e., a filter having more taps) will lead to a lower bit error rate, but will also require more multiplications, which increases the energy consumption of the (battery driven) device. Your task is to determine the minimum length of the MF that exceeds a specified BER performance. By comparing the transmitted bits with your received bits, calculate the BER at an SNR of 8 dB for different MF lengths. Choose the minimum MF length so that the BER is better than $7 \cdot 10^{-3}$, i.e. the implementation loss is less than 0.2 dB (see Figure 1.5).[1]

    Note: Convolutions introduce heads and tails. Take special care to have a correct symbol alignment for BER calculations.

---

[1]At 8 dB, the best uncoded BER that you can achieve is $6 \cdot 10^{-3}$, and a BER of $7 \cdot 10^{-3}$ is ideally achieved at 7.8 dB.