Digital Communication Using GNU Radio

Prof Kumar Appaiah

Department of Electrical Engineering

Indian Institute of Technology Bombay

Week-03

Lecture-13

Welcome to this lecture on digital communication using GNU radio. In this lecture, we are going to discuss more about modulation and pulse shaping and how different pulse shapes affect the performance when you perform digital modulation. So, as far as the outline goes, we will be particularly concerned about the spectral occupancy of modulated signals. We will then discuss the concept of Nyquist ISI-free modulation which ensures that your symbols when they are detected at the receiver do not undergo any inter-symbol interference. We will then discuss minimum bandwidth Nyquist pulses and excess bandwidth and explore what trade-offs are involved in their design and use. Finally, we will look at raised cosine and root raised cosine pulses and how they offer an excellent trade-off to utilize the ISI-free signaling while also having other benefits.

Spectral occupancy is something which is of importance as we have discussed earlier. It is important for you to restrict the signal to be within a frequency range that is allotted to you especially if you are being provided this on a license or under any other constraint imposed by the system. So, let us look at our standard digital communication waveform. We have

$$s(t)=\sum_{n=-\infty}^{\infty} b[n]g_{TX}(t-nT) \quad .$$

This is the standard approach of using the pulse

$g_{TX}(t)$

to send the symbols

$b[0]$, $b[1]$, $b[2]$ etc.

separated by T symbols sorry T seconds in between each pair of symbols. So, you are essentially sending one symbol every T seconds using the pulse

$g_{TX}(t)$.

Now, when it comes to exploring the how much spectrum this particular signaling approach involves using, we will assume initially that

$b[n]$

are zero mean and its uncorrelated sequence that is

$b[n]$

is a random process which has zero mean and is uncorrelated. In other words,

$\mathrm{E}(b[n]b[n-k])=\delta[k]$

as you have seen earlier possibly with some scaling. In such a situation, the power spectral density of

$s(t)$

is given by

$\frac{\mathrm{E}[|(b[n])^2|]}{T}|g_{TX}(f)|^2$.

So, this is the Fourier transform of

$g_{TX}(t)$.

This particular equation is of essence because what this says is that the spectral usage of your transmit signal

$s(t)$

is nearly entirely determined by the Fourier transform of your

$g_{TX}(t)$.

In other words, if you design your

$g_{TX}(t)$

to have a very small spectral footprint, you are going to use a very small amount of spectrum and if your

$g_{TX}(t)$

uses a large amount of spectrum, then by corollary your

$s(t)$

will also use the similar spectrum. So, it is essential for you to design your

$g_{TX}(t)$

or if you look at the Fourier domain

$g_{TX}(f)$

to occupy only the amount of spectrum that is necessary. So, let us start simple.

The rectangular pulse is the most convenient to design and analyze and as you have already seen it is very easy for you to imagine rectangular pulses because your symbol takes that value for T seconds, then again for the next block of T seconds it takes a different value and so on. The problem is as you have seen in the previous slide, your

$g_{TX}(t)$

essentially determines the spectrum which you use and a rectangular pulse has a Fourier transform which is a sinc which is a very wide footprint and rectangular pulses are generally unsuitable for bandwidth constrained usage. Even otherwise the wide spectrum means that it undergoes some transformations when it goes through practical channels which may or may not be desirable. The other extreme is the sinc. The sinc is another interesting signal which is band limited and in fact you can actually space sincs exactly T apart and if you sample them at integer multiples of T you will get the original symbols, but the problem is that the decay is 1 upon t.

This will have some practical implications as we will see, but the sinc essentially dies down very very gradually and this poses many practical issues which may or may not be desirable. So the question which we want to ask is what are the best pulses for bandwidth constrained signaling because rectangular is not good because it occupies too much spectrum. The best spectral signal sinc is not good because it has other issues like it requires a long temporal footprint. So what are the best pulses that we can consider for bandwidth constrained signaling? Before we talk about those pulses let us just try to understand the Nyquist ISI free signaling criterion. Let us first look at a simple model of our communication system.

Let us assume that your symbols

$b[k]$

are being sent at the rate of one such symbol every T seconds. Then they essentially are shaped through a transmit filter

$$g_{TX}(t) \quad .$$

This is the same as constructing your transmit signal

$$s(t) = \sum_{k=-\infty}^{\infty} b[k] g_{TX}(t-kT) \quad .$$

Typically there is also this channel filter. The channel can be a wire channel or it can be air or it can be an optical fiber and there are many such possible channels which themselves may have their own impulse response or frequency response that can affect your signal.

Similarly you have a receive filter that is at the receiver before you get back your symbols you will filter them so that you try to maybe undo the effect of the channel or try to just make some process so that you are able to get the symbols more reliably. We will see this in subsequent lectures. That yields

$$z(t)$$

and this

$$z(t)$$

is sampled at the rate of one sample every T seconds to get

$$z(kT) \quad .$$

So the overall channel that you have or overall system that you have which takes you from this

$$s(t)$$

or rather this

$$b[k]$$

through

$$z(kT)$$

is given by

$$x(t) = g_{TX}(t) * g_c(t) * g_{RX}(t) \quad .$$

So now we ask the question what is the condition on

$x(t)$

so that

$z(kT)$

directly gives us

$b[k]$.

Of course we are ignoring any delays which may happen that is something which can be taken care of but the question which we are asking is when can we say that

$b[k]=z(kT)$

despite the fact that there are three filters in between. It turns out that in the time domain picture this is when

$x(mT)=1$ if $m=0$ and $x(mT)=0$ otherwise.

This will ensure that

$b[k]=z(kT)$.

In other words you can think about signals which satisfy this. A simple example would be a sinc.

A sinc let us say if you say

$\mathrm{sinc}(t/T)$

will be equal to 1 when t is zero but it will be 0 for any other integer multiple of T. If you put t as capital T capital 2T minus 3T and so on it is going to be zero. So sinc is one of the examples of symbols that satisfy this. If you now look at this

$x(mT)$

as a discrete time signal this essentially corresponds to the signal

$\delta(n)$ or $\delta(m)$

it does not matter. What is the Fourier domain interpretation of this? If you sample

$x(t)$

at the rate of one symbol every T seconds that should yield a flat result.

You should get a flat spectrum that is. That is you know that whenever you sample a continuous time signal in the Fourier domain you are going to get copies. These copies should be flat that makes complete sense because the DTFT of

$\delta(n)$

is actually 1. So this is the interpretation that is in the time domain if you sample the signal at integer symbols of capital T you should get one only when $m=0$. In the frequency domain you get copies and if you basically sample and look at the spectrum or a Fourier transform of the sampled signal it should be flat.

Let's try to put this in a more practical sense by looking at some simple examples. Let's look at the Nyquist ISI-free signaling criterion for a signal that we actually already know and use. So what is the minimum bandwidth Nyquist pulse? That is what is the spectrum of that particular signal which uses the least bandwidth and yet satisfies the Nyquist criterion. It turns out that that is if you look at this particular spectrum this is minus one by 2T t one by 2T and the value is T it is just a constant. If you now shift this by T and repeat it this is which is what happens if you sample at the rate T you get this you get this you get this you get this and this essentially yields a flat spectrum.

In other words if you are going to have a flat spectrum then you are not going to have the ISI issue what pulse does this corresponds to this exactly corresponds to the sinc pulse of course there is a T scaling but it corresponds to the sinc pulse. In practice however the sinc pulse is seldom used for two reasons. One is that the tight synchronization requirement can actually result in large ISI if the sampling time is not perfect. That is if you have this kind of behavior where it falls you know very slowly then if you have a neighboring symbol which may look like this and if you now let me draw the axis also if you now sample exactly at this point then the contribution of this red sinc is zero if you start sampling over here the contribution of this red sinc is not zero and you will have one more sinc because let us say you will have yellow sinc let us say which is also present so the yellow sinc is also coming and adding up oops sorry so yellow sinc is also going to come and add up the red sinc green sinc all these will add up and these are essentially going to cause these are essentially going to cause you a lot of inter symbol interference so this one is going to get added up and then this one is also going to get added up and this will cause a lot of inter symbol interference even though there is no other imperfection or noise. So the compromise is to make the sinc or whatever pulse you have die faster which you can achieve by using excess bandwidth.

So in order to avoid this tight synchronization requirement we want a pulse that has the property that it dies much faster that is it undergoes something like this that is it should not have too much of an impact two or three symbols down. Let us see how we can achieve that. This is where the raised cosine pulse comes in like in all other aspects of our engineering this is an engineering tradeoff here the spectrum is actually expanded in a controlled fashion that is your spectrum unlike this rectangle is going to actually start coming down much sooner let us say we take the red curve which is a is 0.25 it starts coming down sooner and then goes beyond 1 upon 2T and only then dies. So what does this correspond to this s of f is defined as T for mod f less than 1 minus a upon 2T t upon 2 times 1 minus sine mod f minus 1 upon 2T pi t upon a over here and 0 otherwise.

So why is this called raised cosine it is because over here this looks like a cosine it is like it is actually a cosine it is a 1 minus sine is actually you know cos square so you can actually just look at it as a raised cosine and what does this achieve this particular spectrum has many of the desirable properties over and above sinc for example in this case first is it looks like when a is close to 0 this is exactly a sinc but as you make a larger and larger the spectrum becomes fatter and fatter so potentially you may have in the time domain the signal becoming narrower and narrower may have smaller decay. Let us see whether that is the case. So this is actually a raised cosine for a equal to let us say 0.5 so what happens is this you have

$$s(t)=\text{sinc}(t/T)\frac{\cos(\pi at/T)}{1-(2at/T)^2}$$

this is the inverse Fourier transform of the

$$S(f)$$

that you saw in the previous slide. Now you may say that this looks just like sinc in fact it has sinc on top of it but actually what happens is there is a

$$\cos(\pi at/T)$$

which is fine but the denominator there is a

$$(2at/T)^2$$

so a sinc is actually

$$\frac{\sin(\pi t/T)}{\pi t/T}$$

so the sinc is essentially proportional to

$$1/t$$

and for large t this is proportional to

$$1/t^2$$

so this is effectively proportional to

$$1/t^3$$

for large enough t and what does this mean when you look farther and farther away the amplitude is going to be much lower than that of a sinc so if you look carefully at this red curve you can clearly see that when compared to the sync the raised cosine dies much faster in fact its amplitude seems you can even ignore this amplitude after 2T while for the sinc its actually pretty high in fact for the sinc it goes down as $1/t$ so may need to go at least 10 or 11 times T into the time axis for you to have a small enough amplitude this way the raised cosine gives you a much much more advantage in the time domain in that it does not cause too much of issue when you go even 2 or 3 symbols away unlike the sinc now one more interesting thing is that you will see that the raised cosine actually seems to honor the zero crossing criteria that is

$x(mT)=1$ only for $m=0$

or

$$s(mT)$$

in this case so then can you be sure that the raised cosine is also a Nyquist pulse in fact that turns out to be true as we will see and the key advantage is that

$$1/t^3$$

is the rate of decay and the impact of current symbol on future symbols even if you sample incorrectly is lower and you have lower penalty for incorrect sampling locations and you have you also satisfy the raised cosine you know the ISI free criterion why is that the case so if you briefly look back at the expression you have to take this particular expression and you have to then shift it by T and then just see what happens if you shift this particular curve by T you will get something like this and this should be added these you know this and the corresponding curve should be added if you add them up the question is does it become 1 because we want the sampled signal to have a Fourier transform which is flat at 1 so the question is does this happen it turns out that if you actually shift the s of f which corresponds to the rate root race rather the raised cosine then this particular curve the red curve which is at 1 upon minus 1 of you know exactly as $1/T$ and the red curve which is at $-1/T$ so this just to you know be sure I am going to this part is actually $1/T$ this part is actually $-1/T$ if you now look at these two points so if you look at these points over here there is no problem it is flat over here there is no problem it is flat the question is what happens in this part similarly on the other part it turns out that that spectrum with that 1 minus sign expression was designed carefully in such a way that if you actually add those two up with the appropriate expressions they will add up to 1 that is if you evaluate

$$\sum_{m=-\infty}^{\infty} S(f+m/T)$$

you will end up getting 1 for all values of f which is exactly what you need because that is what the Nyquist ISI free criterion says.

So now what is the main takeaway from the raised cosine? The raised cosine is an interesting pulse because it allows you to trade off the bandwidth usage you end up using more bandwidth over here unlike a sinc which uses only this much bandwidth but as a return what you get is that you get a time domain pulse which decay is much faster in fact it decays as $1/t^3$ and has several other advantages naturally there is a penalty in the bandwidth usage so you do not want to use too much of bandwidth because then that will make your spectral footprint larger so you use the correct amount of bandwidth in order to use these kinds of pulses. So just a practical note in practice whenever people perform these kinds of signaling the spectrum always uses guard bands in other words when you say 5 megahertz it is 5 megahertz but because you are using a raised cosine you may spill over up to 5.5 megahertz so the next channel or next users signal will actually start at 6 megahertz and go to 11 megahertz so it is like a 1 megahertz of gap for you to accommodate these kind of raised cosine pulses and because the raised cosine its Fourier transform has a slight expansion. So in this way the raised cosine has several advantages and we will actually evaluate this on GNU radio to confirm that these advantages actually carry through. So now what is the root raised cosine pulse? If you recall our system model wherein we had

$$b[k]$$

which went up to

$$z(t)$$

and we got

$$z(mT)$$

we actually had this

$$x(t)$$

that took care of all the transmit pulse shaping the channel and the receive pulse shaping which should be our receive pulse shaping.

So if you look at this particular

$x(t)$

we want this to actually we want to design this in a way where the Nyquist criterion is satisfied. So you need to choose

$g_{TX}(t)$

and choose

$g_{RX}(t)$.

Typically

$g_c(t)$

is not something which is in your control it is medium dependent and for example if you use a copper cable or a coaxial cable depending on how much bandwidth you use

$g_c(t)$

may be flat or

$g_c(t)$

may be looking like an impulse

$g_c(t)$

may have some other impulse response. So it is highly dependent on the medium. So if you have a particular medium you have to account for

$g_c(t)$

typically you may try to design your

$g_{TX}(t)$ and $g_{RX}(t)$

to accommodate for it you know if it is possible at all.

Sometimes what people do is they say let's take

$g_c(t)=\delta(t)$

for our design and then later we will compensate for it. So compensation for it is called equalization which we will see in a later class but for now we are going to assume that

$$g_c(t)$$

is not there or in other words that

$$g_c(t) = \delta(t) \quad .$$

If you take

$$g_c(t) = \delta(t)$$

for simplicity then one option for us is to choose both

$$g_{TX}(t) \quad \text{and} \quad g_{RX}(t)$$

to be the same pulse

$$g(t) \quad .$$

This has many advantages because your design is essentially design effort is hard because both the

$$g_{TX}(t) \quad \text{and} \quad g_{RX}(t)$$

are the same but this also has another major advantage in that

$$g_{TX}(t) \quad \text{and} \quad g_{RX}(t)$$

are matched to each other. But the question is we want a raised cosine pulse effective we want

$$x(t)$$

to be a raised cosine.

So how do we do that? If we look at the Fourier transform of

$$g(t)$$

and let it be

$$G(f) = \sqrt{S(f)}$$

where

$$S(f)$$

is the Fourier transform of the raised cosine pulse. So in other words the root raised cosine pulse is obtained by taking the inverse Fourier transform of the square root of the Fourier transform of the raised cosine. So in other words the square root is not taken in the time domain it is taken in the frequency domain. So you take your raised cosine pulses Fourier transform you take its square root as you know as

$$\sqrt{S(f)}$$

find its inverse Fourier transform and that will give you the root raised cosine pulse. Now if you look at

$$g_{TX}(t) * g_{RX}(t)$$

the Fourier transforms will be

$$(\sqrt{S(f)})(\sqrt{S(f)}) = S(f)$$

which is our raised cosine pulse and you know that

$$S(f)$$

satisfies the condition that summation

$$\sum_{-\infty}^{\infty} S(f + m/T) = 1$$

that is it satisfies the Nyquist ISI free criterion.

So it is actually useful for us to choose

$g_{TX}(t)$ and $g_{RX}(t)$

to be the same root raised cosine pulses where the root is taken in the frequency domain. In other words if you take this

$$\sqrt{S(f)}$$

is essentially obtained by taking square root of the you take square root of this or let us not show it let us assume that this is just square roots in this case there is not much difference. So you take the

$$\sqrt{S(f)}$$

where the

$S(f)$

was the old one you take its corresponding time domain pulses

$g_{TX}(t)$ and $g_{RX}(t)$

and the resulting spectrum is

$S(f)$.

So the root raised cosine is something which is very commonly used and this root raised cosine is something which we will explore in the next GNU radio lecture as well. In practice of course as we mentioned

$g_c(t)$

cannot be ignored so the effective system is

$x(t)$

and even if you have chosen

$g_{TX}(t)$ and $g_{RX}(t)$

to be equal reason we have written

$g_{TX}(t)$

is because we chose

$g_{TX}(t)=g_{RX}(t)$

then you are going to get the raised cosine convolved with

$g_c(t)$.

So we have to compensate for

$g_c(t)$

somehow and here we call it equalization because we need to have some other particular system or filter that will equalize the channel. So what does equalization mean essentially we want something which results in something which is flat. Unfortunately we do not get something which is flat we get something which is like this. So an equalizer is something which essentially offers just a compensation so it does something like this so that the product is flat.

So it essentially equalizes the gain. So equalization is something which we will see when we discuss demodulation and that is something which we will see soon. But now the simplification which we use is we will design an equalizer so we will assume that our bandwidth in our bandwidth of interest it is actually flat which means we can ignore

$g_c(t)$

or there is someone who has designed the equalizer within our bandwidth of interest which means our raised cosine essentially works. To summarize pulse shaping is important and necessary for honoring bandwidth constraints and like ISI free signaling pulses such as sinc have some advantages of course we discussed that sinc has some disadvantages in particular because of its very very slow decay. So we looked at the raised cosine pulse which is a tradeoff over sinc. It uses more spectrum but it is less susceptible to jitter etc.

because of the fact that it decays very quickly. In practical scenarios we need equalization to compensate for channels in the band of interest that is we need to be able to filter so that the effect of the $g_c(t)$ is also compensated. In the next lecture we will actually have a GNU radio demonstration of the root raised cosine pulses and how it impacts your design of digital communication transmission systems. Thank you.