Module
4
Signal Representation and Baseband Processing

Lesson 21

Nyquist Filtering and Inter Symbol Interference

After reading this lesson, you will learn about:

- ***Power spectrum of a random binary sequence;***
- ***Inter symbol interference (ISI);***
- ***Nyquist filter for avoiding ISI;***
- ***Practical improvisation of ideal Nyquist filter;***
- ***Raised Cosine (RC) filter and Root–Raised Cosine (RRC) filtering;***

Nyquist's sampling theorem plays a significant role in the design of pulse-shaping filters, which enable us to restrict the bandwidth of information-bearing pulses. In this lesson, we start with a short discussion on the spectrum of a random sequence and then focus on the concepts of Nyquist Filtering. Specifically, we develop an idea about the narrowest (possible) frequency band that will be needed for transmission of information at a given symbol rate.

Consider a random binary sequence shown in **Fig.4.21.1 (a)** following a common style (NRZ: Non-Return-to-Zero Pulses) for its representation. Note that a binary random sequence may be represented in several ways. For example, consider **Fig.4.21.1 (b).** The impulse sequence of **Fig.4.21.1 (b)** is an instantaneously sampled version of the NRZ sequence in **Fig.4.21.1 (a)** with a sampling rate of one sample/pulse. The information embedded in the random binary sequence of **Fig.4.21.1 (a)** is fully preserved in the impulse sequence of **Fig.4.21.1 (b)**. Verify that you can easily read out the logical sequence only by looking at the impulses. So, the two waveforms are equivalent so far as their information content is concerned. The obvious difference, from practical standpoint, is the energy carried by a pulse.

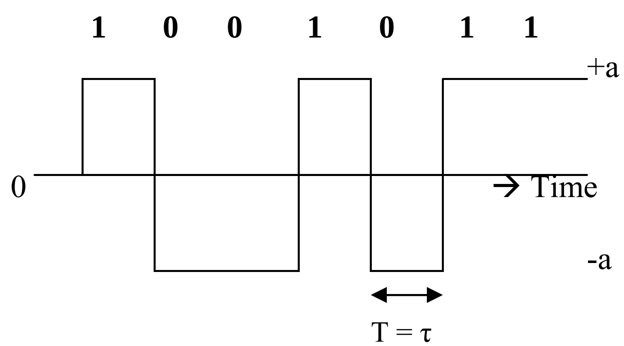

Fig.4.21.1(a) *Sketch of a NRZ (Non-Return-to-Zero) waveform for a random binary sequence*

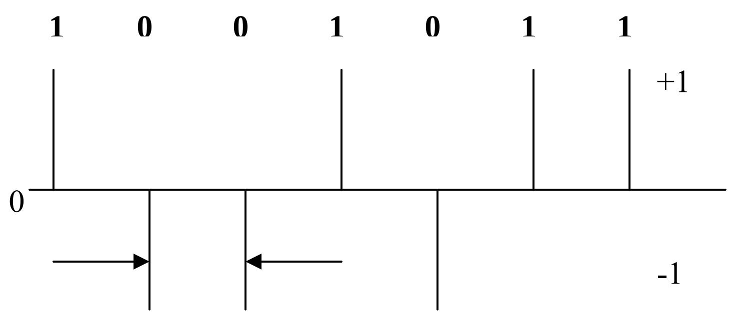

Fig.4.21.1 (b): *Instantaneously sampled version of the NRZ sequence in* **Fig.4.21.1 (a)**

Power Spectrum of Random Binary Sequence

Let us consider the NRZ representation of **Fig.4.21.2(a)** where in a pulse is of height 'a' and duration T_b. We wish to get an idea about the spectrum of such sequences. The sequence may be viewed as a sample function of a random process, say X(t). So, our approach is to find the ACF of X(t) first and then take its Fourier Transform. Now, the starting instant of observation need not synchronize with the start time of a pulse. So, we assume that the starting time of the first pulse (i.e. the initial delay)'t_d' is equally likely to lie anywhere between 0 and T_b, i.e.,

$$p_n\left(t_\theta\right)=\begin{cases}\dfrac{1}{T_b},0\le t_d\le T_b\\ 0,elsewhere\end{cases} \qquad 4.21.1$$

Next, the '0'-s and '1'-s are equally likely. So, we can readily note that, E[X(t)] = 0.

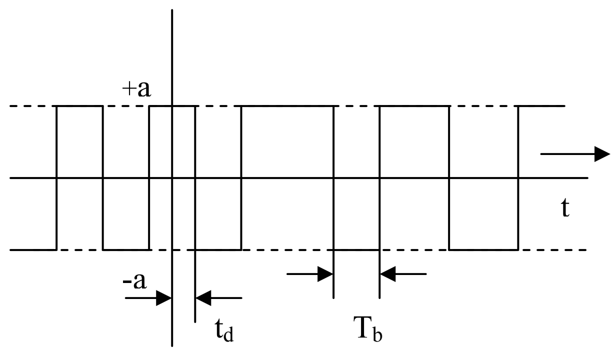

Fig. 4.21.2(a): *A random NRZ sequence representing an information sequence*

ACF of X(t)

Let R_x (t_k,t_i) denote the ACF of X(t).

$$\therefore R_x\,(t_k,t_i) = E[X(t_k)\cdot X(t_i)] \qquad 4.21.2$$

Two distinct cases are to be considered: a) when the shift, i.e. $|t_k - t_i|$ is greater than the bit duration T_b and b) when $|t_k - t_i| \le T_b$.

Case-I: Let, $|t_k - t_i| > T_b$.

In this case, X (t_k) and X (t_i) occur in different bit intervals and hence they are independent of each other. This implies,

$$E\,[X\,(t_k)\cdot X\,(t_i)] = E\,[X\,(t_i)]\cdot E\,[X\,(t_k)] = 0; \qquad 4.21.3$$

Case-II: $|t_k - t_i| < T_b$. For simplicity, let us set $t_k = 0$ and $t_i < t_k = 0$.

In this case, the random variables $X(t_k)$ and $X(t_i)$ occur in the same pulse interval iff $t_d < T_b - |t_k - t_i|$. Further, both $X(t_k)$ and $X(t_i)$ are of same magnitude 'a' and same polarity.

Thus, we get a conditional expectation:

$$E\left[X(t_k)\cdot X(t_i)\Big|_{t_d}\right]=\begin{cases}a^2, t_d < T_b-|t_k-t_i|\\ 0, elsewhere\end{cases} \qquad 4.21.4$$

Averaging this result over all possible values of t_d, we get,

$$E\left[X(t_k)\cdot X(t_i)\right]=\int_0^{T_b-|t_k-t_i|} a^2 p(t_d)\,dt_d$$

$$=\int_0^{T_b-|t_k-t_i|}\frac{a^2}{T_b}dt_d=a^2\left(1-\frac{|t_k-t_i|}{T_b}\right),\quad |t_k-t_i|\le T_b$$

4.21.5

By similar argument, it can be shown that for other values of t_k , the ACF of a binary waveform is a function of the time shift $\tau = t_k - t_i$. So, the autocorrelation function $R_x(\tau)$ can be expressed as [**Fig.4.21.2(b)**]:

$$R_x(\tau)=\begin{cases}a^2\left(1-\dfrac{|\tau|}{T_b}\right), |\tau|<T_b\\ 0, |\tau|\ge T_b\end{cases} \qquad 4.21.6$$

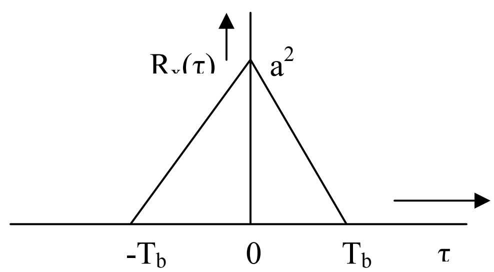

Fig. 4.21.2(b): *Auto Correlation Function $R_x(\tau)$ for a random binary waveform*

Now the power spectral density of the random process X(t) can be obtained by taking the Fourier Transform of $R_x(\tau)$:

$$S_x(f)=\int_{-T_b}^{T_b} a^2\left(1-\frac{|\tau|}{T_b}\right)\exp(-j2\pi f\tau)\,d\tau$$

$$=a^2T_b\,\mathrm{sinc}^2(fT_b) \qquad 4.21.7$$

A rough sketch of $S_x(f)$ is shown in **Fig. 4.21.2(c)**. Note that the spectrum has a peak value of '$a^2.T_b$'. The spectrum stretches towards $\pm\infty$ and it has nulls at $\pm\,{}^{1}\!/_{n.T_b}$. A normalized version of the spectrum is shown in **Fig. 4.21.2(d)** where the amplitude is normalized with respect to the peak amplitude and the frequency axis is expressed in terms of 'fT_b'.

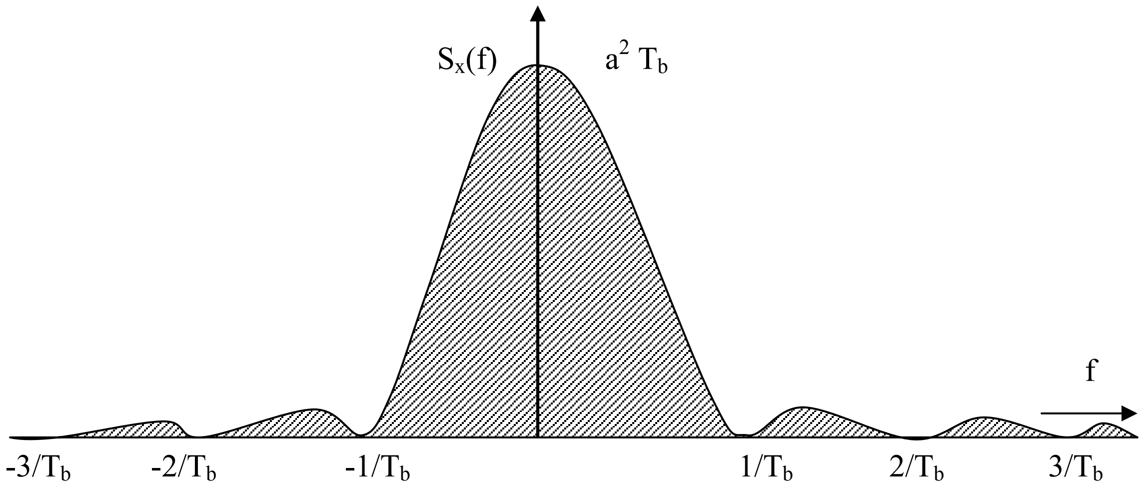

Fig. 4.21.2(c) : *A sketch of the power spectral density, $S_x(f)$ for a random binary NRZ pulse sequence*

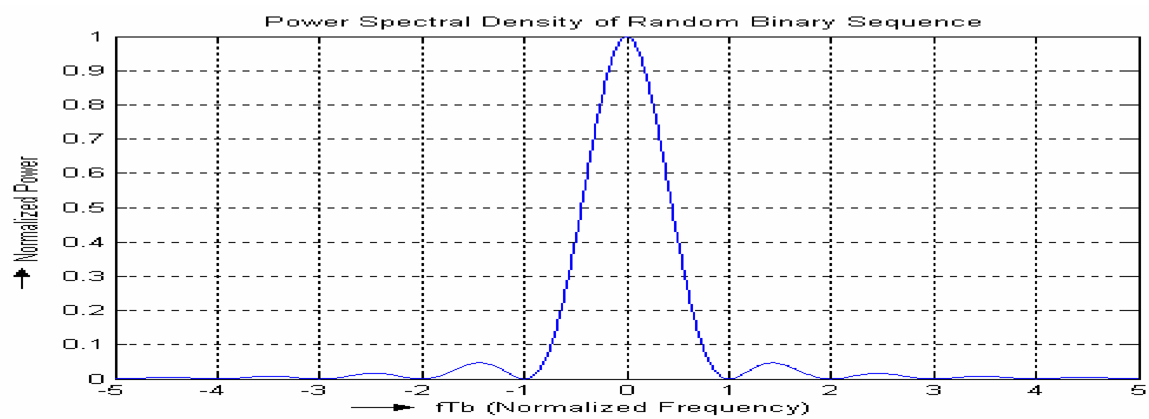

Fig. 4.21.2(d) : *Normalized power spectral density, $S_x(f)$ for a random binary NRZ pulse sequence*

The wide stretch of the spectrum is understandable as the time pulses are sharply limited within a bit duration. But then, if such a random NRZ sequence is used to modulate a carrier sinusoid, one can easily imagine that the modulated signal spectrum will also have an infinite width and hence, such a modulated signal cannot be transmitted without distortion though a wireless channel, which is band-limited. If the modulated signal is forced through a band limited channel without appropriate spectral shaping, the spectrum of the modulated signal at the receiver will be truncated due to the band pass characteristic of the channel. From the principle of time-frequency duality, one can now guess that the energy of a transmitted pulse will no more be limited within its typical duration 'T_b'.

Alternatively, the energy of one pulse will spill over the time slot of one or more subsequent pulses, causing *Inter Symbol Interference (ISI).* So, over a specific pulse duration 'T_b', the receiver will collect energy due to one desired and multiple undesired pulses. A typically vulnerable situation is when a negative pulse appears in a string of positive pulses or vice versa. In general, the received signal becomes more vulnerable to noise and upon demodulation; the information sequence may be erroneous. The extent of degradation in the quality of received information depends on the time spread of energy of a transmitted pulse and how this effect of ISI is addresses in the receiver.

Another reason why sharp rectangular pulses, even though designed following Gramm-Scmidt orthogonalization procedure, are not good for band-limited channels is that, one is simply not allowed to use the full bandwidth that may be presented by a physical channel. Specifically, most wireless transmissions must have a priori approval from concerned regulatory authority of a country. It is mandatory that signal transmission is done precisely over the narrow portion of the allocated bandwidth so that the adjacent bands can be allocated for other transmission schemes.

Further, to conserve bandwidth for various transmission applications, the narrowest feasible frequency slot only may be allocated. So, it is necessary to address the issue of ISI in general and the issue of small transmission bandwidth by shaping pulses. There are several equalization techniques available for addressing the issue of ISI. Many of these techniques probe the physical channel and use the channel state information in devising powerful adaptive equalizers. In this course, we skip further discussion on equalization and focus only on the issue of pulse shaping for reduction in transmission bandwidth.

Nyquist Filter for avoiding ISI

Let us recollect the second part of Nyquist's sampling theorem for low pass signals which says that a signal band limited to B Hz can be recovered from a sequence of uniformly spaced and instantaneous samples of the signal taken at least at the rate of 2B samples per second. Following this theorem, we may now observe that the impulse sequence of **Fig.4.21.1 (b)**, which contains information '1001011', can be equivalently described by an analog signal, band limited to ½ x 1sample/pulse. That is, if the bit rate is 1 bit/sec and we sample it at 1 sample/sec, the minimum bandwidth necessary is ½ Hz! An ideal low-pass filter with brick-wall type frequency response and having a cutoff of 0.5 Hz will generate an equivalent analog waveform (pulse sequence) when fed with the random impulse sequence.

Recollect that the impulse response of an ideal low-pass filter is a sinc function [y = sinc(x), **Fig. 4. 21.3**]. We note that a) sinc(0) = 1.0 and the impulse response has nulls at $x = \pm 1, \pm 2, \ldots$ sec, b) null-to-null time-width of the prominent pulse is 2 sec, c) the pulse is symmetric around t = 0 and d) the peak amplitude of the pulse is of the same polarity as that of the input impulse. As the filter is a linear network, the output analog waveform is simply superposition of sinc pulses, where the peaks of adjacent sinc pulses are separated in time by 1 sec, which is the sampling interval.

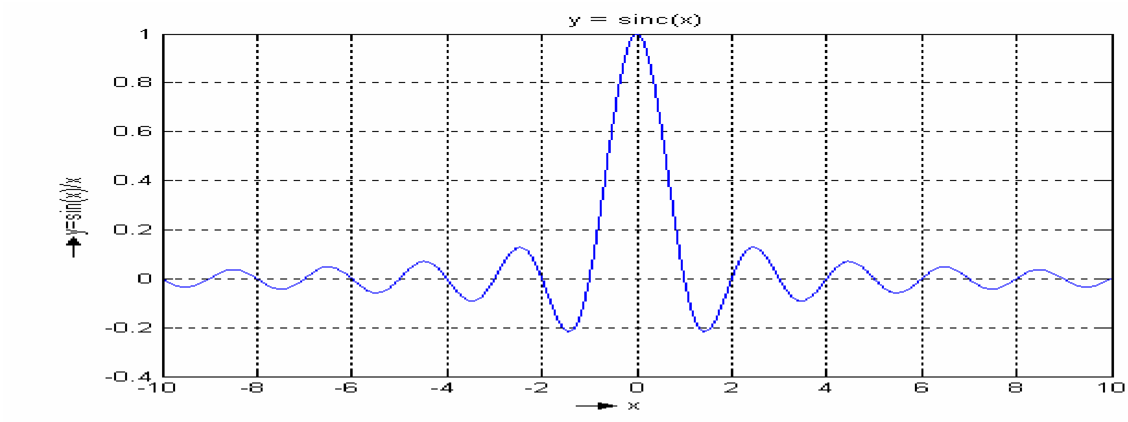

Fig. 4. 21.3(a) *Plot of y = sinc(x)*

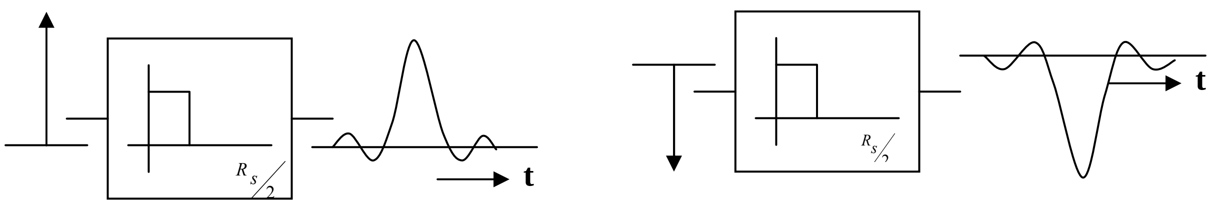

Fig. 4. 21.3(b) *Sketch of output of Nyquist filter for positive and negative impulses*

In general, if the information symbol rate is R_s symbols/sec, i.e. the symbol interval is $T_s = {}^{1}\!/_{R_s}$ second, the single-sided bandwidth of the low-pass filter, known popularly as the equivalent Nyquist Bandwidth, is $B_N = {}^{R_s}\!/_{2}$ Hz.

A simple extension of the above observations implies that for instantaneous samples of random information-bearing pulse (or symbol) sequence (@ 1 sample per symbol) will exhibit nulls at $\pm nT_s$ seconds at the filter output. Now, assuming an ideal noise less baseband channel of bandwidth B_N and zero (or fixed but known) delay, we can comment that the same filter output waveform will appear at the input of the receiver. Though the waveform will appear much different from the initial symbol sequence, the information (embedded in the polarity and magnitude) can be retrieved without any effect of the other pulses by sampling the received baseband signal exactly at peak position of each shaped pulse as at those time instants all other pulses have nulls. This ideal brick-wall type filter is known as the Nyquist Filter for zero-ISI condition. **Fig. 4.21.4** highlights some important features of Nyquist Filter.

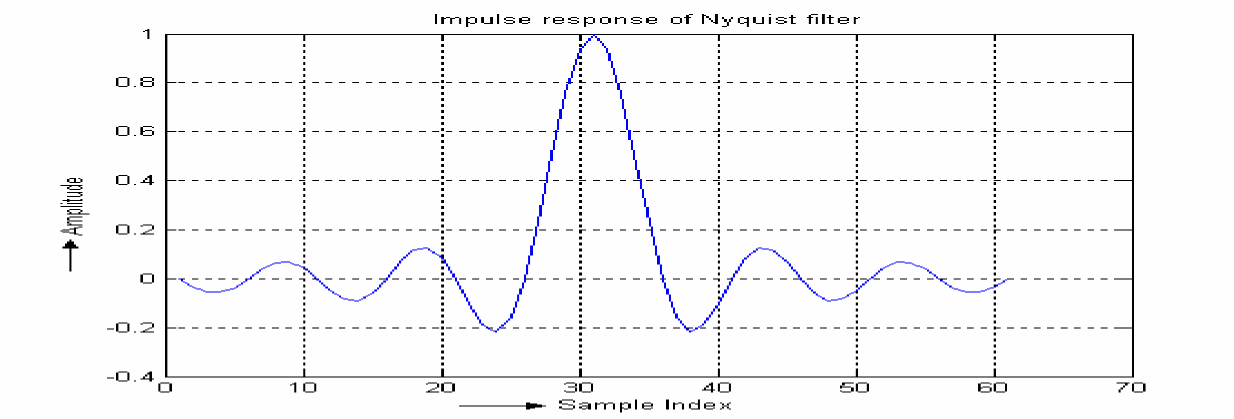

Fig. 4.21.4 *Typical impulse response of a Nyquist filter [5 samples /symbol duration have been used to obtain the responses using a digital computer. ±32 sample duration around the normalized peak value of 1.0 is shown]*

Mathematical Explanation

Let us consider an ideal lowpass filter with single-sided bandwidth W. It is usually convenient to deal with the normalized impulse response h(t) of the filter where in,

$$h(t)=\text{sinc }(2Wt) = \sin(2\pi Wt)\big/2\pi Wt \qquad 4.21.8$$

Now if an impulse of strength 'A' is applied at the input of the filter at $t = t_1$, the filter o/p may be expressed as,

$$y(t) = Ah(t- t_1)=A.\text{sinc } 2W (t- t_1) \qquad 4.21.9$$

An information-carrying symbol sequence may be represented as

$$\sum_{i=0}^{\infty} A_i.\delta(t-t_i), \text{ where } A_i = \pm A \text{ and } t_i = i.T_s \qquad 4.21.10$$

The response of the low-pass filter to this sequence is,

$$y(t) = \sum_{i=0}^{\infty} A_i.\sin c\{2W(t-t_i)\} \qquad 4.21.11$$

Now, if we set $W = R_s/2 = 1\big/2.T_s$ and sample the output of the filter at $t = m.T_s$, it is easy to see that,

$$y(t = m.T_s) = \sum_{i=0}^{m} A_i.\text{sinc}2W(m - i)T_s = \sum_{i=0}^{m} A_i.\text{sinc}(m - i) = A_m \qquad 4.21.12$$

So, we obtain the peak of the m-th pulse clean and devoid of any interference from previous pulses.

Practical improvisations:

The above discussion on Nyquist bandwidth is ideal because neither a low-pass filter (LPF) with brick-wall type response can be realized physically nor can an ideal impulse sequence be generated to represent a discrete information sequence. Further, note that if there is any error in the sampling instant (which, for a practical system is very likely to occur from time to time due to the effects of thermal noise and other disturbances), contribution from the other adjacent pulses will creep into the sample value and will cause Inter Symbol Interference. For example, the second lobe peak is only about 13 dB lower compared to the main lobe peak and the decay of the side-peaks of a sin x/x function is not very rapid with increase in x. So, the contribution of these peaks from adjacent symbols may be significant. Fortunately, the desired features of pulse epoch and zero crossings are not unique to a 'sinc' pulse. Other pulse shapes are possible to design with similar features, though the bandwidth requirement for transmitting a discrete information sequence will be more compared to the corresponding Nyquist bandwidth (B_N). Two such relevant constructs are known as a) Raised Cosine (RC) filter and b) Root-Raised Cosine (RRC) filter.

Raised Cosine Filter

Let us denote a normalized pulse shape which avoids ISI as x(t). then,

$$x(t)\big|_{t=\pm nT_s} = 0, \text{n} \neq 0 \text{ and } x(0) = 1 \qquad 4.21.13$$

Some of the practical requirements on x(t) are the following:

(a) Energy in the main pulse is as much as possible compared to the total energy distributed beyond the first nulls around the main peak. This ensures better immunity against noise at the receiver for a given signal transmission power. So, it is desired that the magnitde of the local maxima of the i-th pulse of x(t) between iT≤ t < (i+1)T decreases monotonically and rapidly with time.

(b) The pulse shape x(t) should be so chosen that some error in instants of sampling at the receiver does not result in appreciable ISI. These two requirements are usually depicted in the form of a mask in technical standards.

Out of several mathematical possibilities, the following amplitude mask is very useful for application on the ideal Nyquist filter impulse response h(t):

$$m(t) = \frac{\cos\left(\beta\pi \, {}^{t}\!/_{T_s}\right)}{1-(4\beta^2t^2/T_s^{\ 2})} \qquad 4.21.14$$

The resulting pulse shape is known as a Raised Cosine pulse with a *roll-off* of 'β'(0∠ β ≤1). The Raised Cosine pulse is described as:

$$p_{RC}(t) = \text{sinc}(t/T_s).\ \frac{\cos\left(\beta\pi \, {}^{t}\!/_{T_s}\right)}{1-(4\beta^2t^2/T_s^{\ 2})}$$

4.21.15

The normalized spectrum of Raised Cosine pulse is:

$$H(f) = 1, \text{ for } |f| \le \frac{(1-\beta)}{2T_s},$$
$$= \cos^2 \frac{\pi T_s}{2\beta}\left(|f| - \frac{(1-\beta)}{2T_s}\right), \quad \text{for } \frac{(1-\beta)}{2T_s} \le |f| \le \frac{(1+\beta)}{2T_s}$$
$$= 0, \quad \text{for } |f| > \frac{(1+\beta)}{2T_s} \qquad 4.21.16$$

Figs. 4.21.5 (a)-(c) highlight features of a Raised Cosine (RC) filter. The roll off factor 'β' is used to find a trade off between the absolute bandwidth that is to be used and the difficulty (in terms of the order of the filter and the associated delay and inaccuracy) in implementing the filter. The minimum usable bandwidth is obviously the Nyquist bandwidth, B_N (for β =0) for which the filter is unrealizable in practice while a maximum absolute bandwidth of $2B_N$ (for β = 1.0) makes it much easier to design the filter. β lies between 0.2 and 0.5 for most of the practical systems where transmission bandwidth is not a luxury. Use of analog components was dominant earlier though the modern trend is to use digital filters. While digital FIR structure usually ensures better accuracy and performance, IIR structure is also used to reduce the necessary hardware.

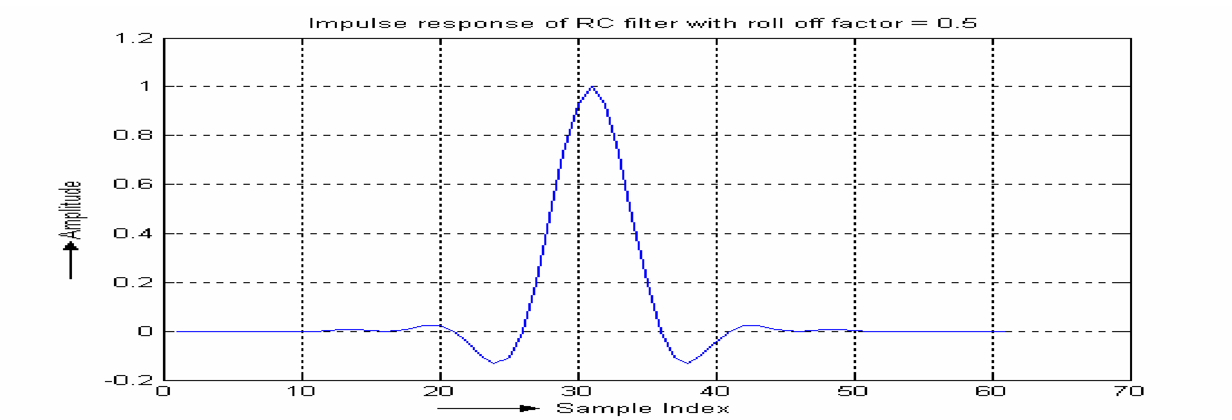

Fig. 4.21.5 (a) *Typical impulse response of a Raised Cosine filter with a roll off factor β = 0.5[5 samples /symbol duration have been used to obtain the responses using a digital computer. ±32 sample duration around the normalized peak value of 1.0 is shown]*

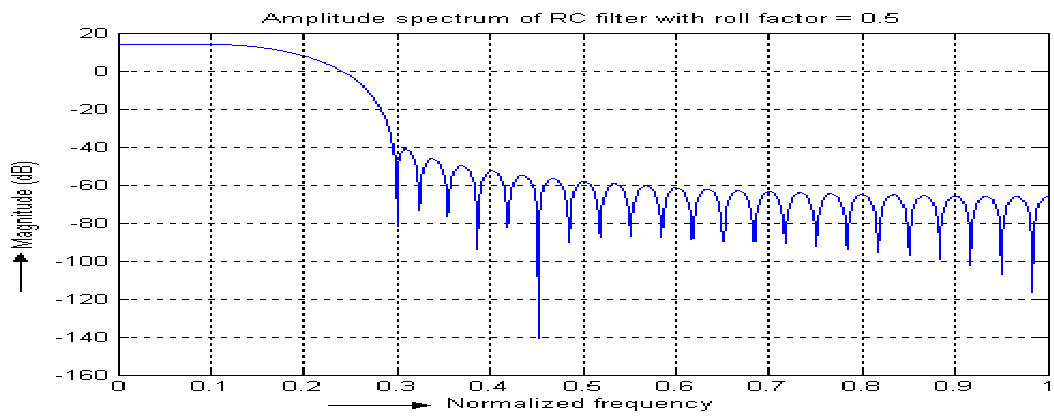

Fig. 4.21.5 (b) *Typical amplitude spectrum of a Raised Cosine filter with a roll off factor β= 0.5. The magnitude is not normalized. Multiply the normalized frequency values shown above by a factor of 5 to read the frequency normalized to symbol rate. For example, i) 0.1(from the above figure)x5 = 0.5 = (1 -β).R_s, where β=0.5 and R_s = 1. The transition band of the filter starts here and ii) 0.3(from the above figure)x5 = 1.5 = (1 +β).R_s. The stop band of the filter starts here.*

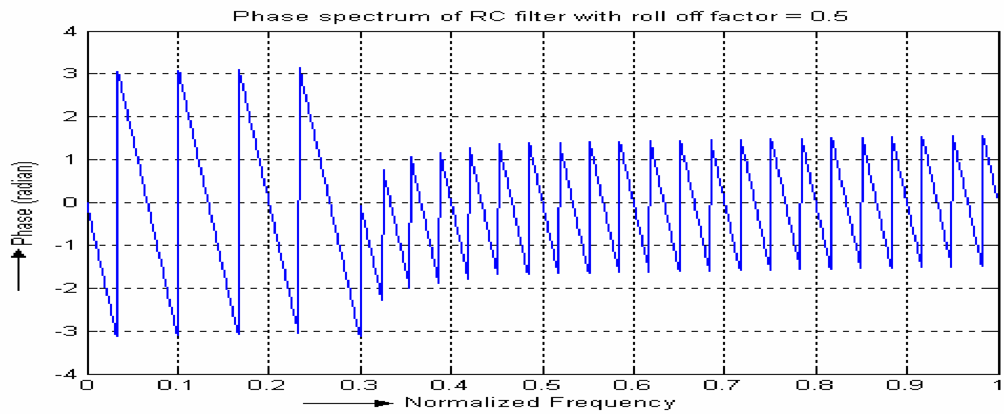

Fig. 4.21.5 (c) *Typical phase spectrum of a Raised Cosine filter with a roll off factor β = 0.5. Multiply the normalized frequency values shown above by a factor of 5 to read the frequency normalized to symbol rate.*

The side lobe peaks in the impulse response of a Raised Cosine filter decreases faster with time and hence results in less ISI compared to the ideal Nyquist filter in case of sampling error in the receiver.

There is another interesting issue in the design of pulse shaping filters when it comes to applying the concepts in a practical communication transceiver. From our discussion so far, it may be apparent that the pulse-shaping filter is for use in the transmitter only, before the signal is modulated by a carrier or launched in the physical channel. However, there is always a need for an equivalent lowpass filter in the receiver to eliminate out-of-band noise before demodulation and decision operations are carried out. This purpose is accomplished in a practical receiver by splitting a Raised Cosine filter in two parts. Each part is known as a Root Raised Cosine (RRC) filter. One RRC filter is placed in the transmitter while the other part is placed in the receiver. The transmit RRC filter does the job of pulse shaping and bandwidth-restriction fully while not ensuring the zero-ISI condition completely. In case of a linear time invariant physical channel, the receiver RRC filter, in tandem with the transmit RRC filter, fully ensures zero-ISI condition. Additionally, it filters out undesired out-of-band thermal noise. On the whole, this approach ensures zero-ISI condition in the demodulator where it is necessary and it also effectively ensures that the equivalent noise-bandwidth of the received signal is equal to the Nyquist bandwidth B_N. The overall complexity of the transceiver is reduced without any degradation in performance compared to a system employing a RC filter in the transmitter and a different out-of-band noise eliminating filter in the receiver. **Figs. 4.21.6 (a)-(c)** summarize some features of a Root-Raised Cosine filter.

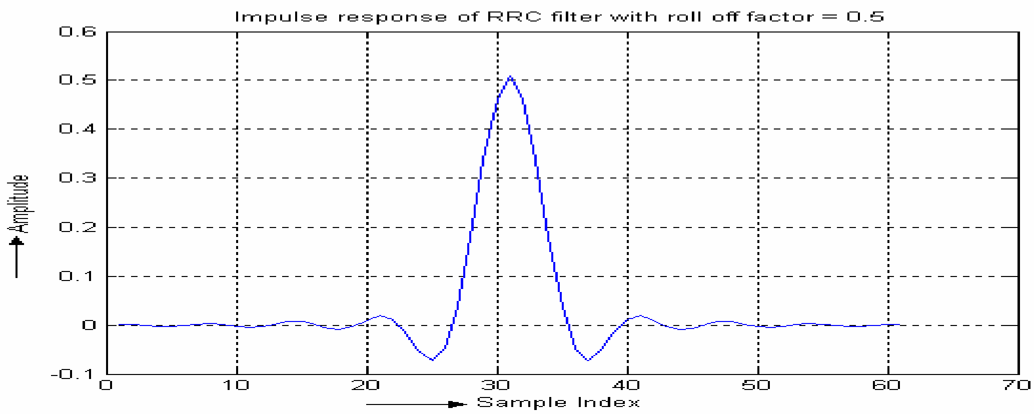

Fig. 4.21.6 (a) *Typical impulse response of a Root Raised Cosine filter with a roll off factor β = 0.5[5 samples /symbol duration have been used to obtain the responses using a digital computer. ±32 sample duration around the normalized peak value of 0.5 is shown]*

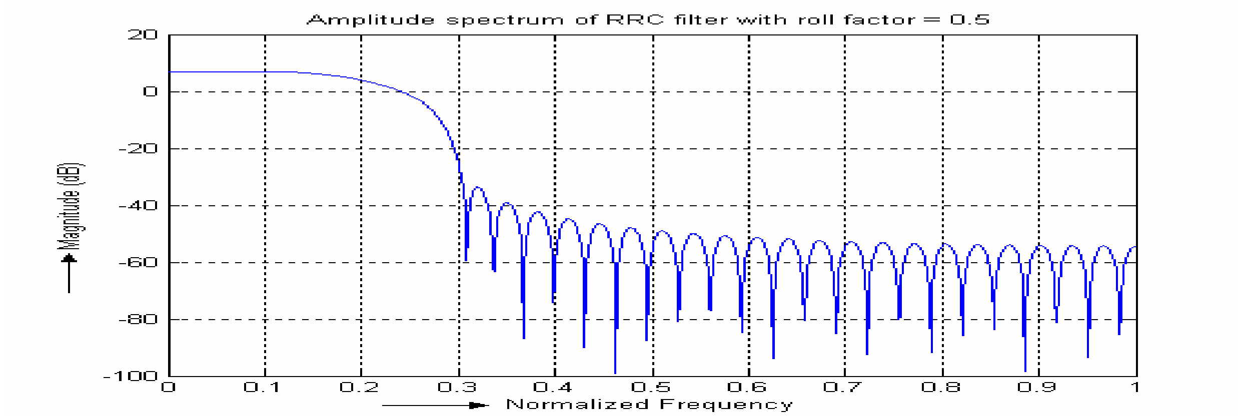

Fig. 4.21.6 (b) *Typical amplitude spectrum of a Raised Cosine filter with a roll off factor β = 0.5. Multiply the normalized frequency values shown above by a factor of 5 to read the frequency normalized to symbol rate.*

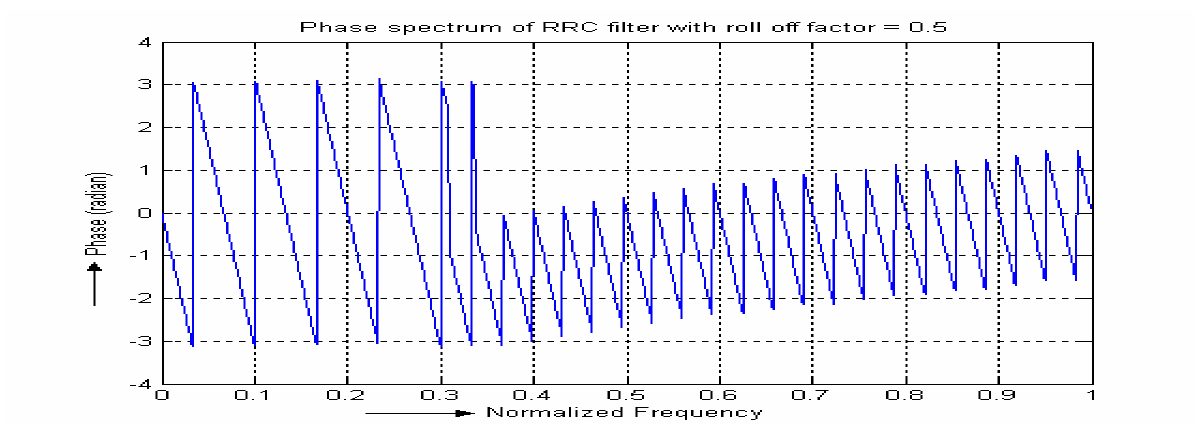

Fig. 4.21.6 (c) *Typical phase spectrum of a Raised Cosine filter with a roll off factor β = 0.5. Multiply the normalized frequency values shown above by a factor of 5 to read the frequency normalized to symbol rate.*

We will mention another practical issue related to the design and implementation of pulse-shaping filters. Usually, an information sequence is presented at the input of a pulse shaping filter in the form of pulses (.e.g. bipolar NRZ) and the width of such a pulse is not negligibly small (as in that case the energy per pulse would be near zero and hence some pulses may even go unrecognized). This finite non-zero width of the pulses causes a distortion in the RC or RRC pulse shape. The situation is somewhat analogous to what happens when we go for flat-top sampling of a band-limited analog signal instead of instantaneous sampling. The finite pulse-width results in amplitude fluctuation and again introduces ISI. It needs a relatively simple pulse magnitude correction which is commonly referred as '$\frac{x}{\sin x}$ amplitude compensation' to restore the quality of shaped pulses. The transfer function of the RRC shaping filter in the transmitter is scaled appropriately to incorporate the amplitude correction.

Problems

Q4.21.1) Mention possible causes of Inter Symbol Interference (ISI) in a digital communication receiver.

Q4.21.2) Sketch the power spectrum of a long random binary sequence when the two logic levels are represented by +5V and 0V.

Q4.21.3) Sketch the frequency response characteristics of an ideal Nyquist low pass filter.

Q4.21.4) What are the practical difficulties in implementing an ideal Nyquist low pass filter?