Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

EEU44C18 / EEP55C28
Digital Wireless Communications
Lecture 3: Digital Modulation and Pulse Shaping

Dr. Arman Farhang (arman.farhang@tcd.ie)
2023

Review of Lecture 2

- Analogue modulation and demodulation.
- Complex baseband model of analogue modulation.
- Channel modelling.

Outline

- Represent the signal space diagram of a digital signal for every modulation scheme.
- Analyse the robustness of the different amplitude/phase modulation schemes.
- Estimate the effects that carrier timing offset and carrier frequency offset have on the space diagram of the demodulated digital signal.
- Describe the Nyquist criterion for non-dispersive channel transmission.
- Analyse the consequences of using ideal pulse shaping and raised cosine pulse shaping.

Digital Modulation

- In Lecture 1, it was seen that if the complex baseband representation of a signal is used, the modulated signal may be written as

$$s(t) = \Re(s_l(t)e^{j2\pi f_c t}) \tag{1}$$

where $s_l(t) = x(t) + jy(t)$ and $x(t)$ and $y(t)$ may be individually demodulated.

- In particular, $x(t)$ is called the in phase component and $y(t)$ is called the quadrature component of the signal.
- By expansion of this, we see that this is amplitude modulation

$$s(t) = x(t)\cos(2\pi f_c t) + y(t)\sin(2\pi f_c t) \tag{2}$$

or equivalently phase modulation

$$s(t) = a(t)\cos(2\pi f_c t + \phi(t)) \tag{3}$$

where $a(t) = \sqrt{x^2(t) + y^2(t)}$ and $\phi(t) = \tan^{-1}(y(t)/x(t))$

Digital Modulation

- Quadrature modulated signals may be considered either as:
  - Amplitude modulation with $x(t)$ amplitude modulating $\cos(2\pi f_c t)$, and $y(t)$ amplitude modulating $\sin(2\pi f_c t)$.
  - Joint amplitude and phase modulation with amplitude $a(t)$ and phase $\phi(t)$. (Note that when $a(t) = \text{constant}$, this may be considered as phase modulation).
  - The all-digital model has two important characteristics:
    - The most important time instant is the sampling instant.
    - At the sampling instant, both amplitude and phase take on discrete values (only a certain value within a set can be taken).
  - Each transmitted (possibly complex) digital/discrete value is called a **symbol**.

Digital Modulation

- Using the transmitter structure seen in Lecture 1, it was seen how by using the complex baseband model the analogue modulation and DAC module can be omitted. In addition, on the receiver side, the analogue demodulation and ADC can also be omitted in the study of the wireless communication system.

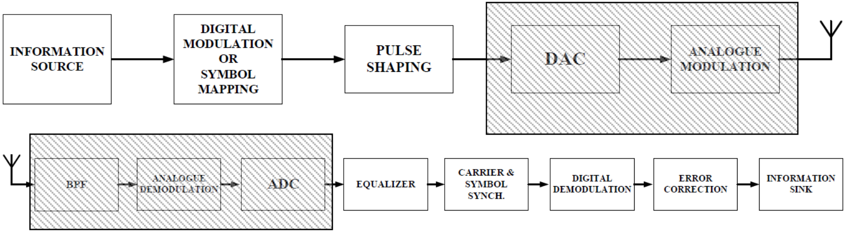


Figure: Wireless digital communication system equivalent model when using the complex baseband signal representation

Digital Modulation

- On the transmitter side, every T seconds, the system sends K bits of information through the channel at a data rate $R = K/T$ bits per second (bps). Prior to being transmitted through the channel, the information bits are fed into the digital modulator and pulse shaping module.
- Digital modulation is the action of dividing the information bit stream into finite length bit words and encoding them into one of several possible transmitted signals.
- On the other side, the digital demodulator decodes the received signal by matching the received value with the "closest" value of all possible transmitted signals. In this way, the receiver minimizes the probability of detection error.

Geometric Representation of Signals

Signal Space

- The geometric representation of signals is based on vector representation.
- In vector representation a basis set formed by N linearly independent unitary orthogonal vectors is required to represent any other vector. This basis is called a N-dimensional Hilbert space.
- In the signal space, the unitary vectors basis is replaced by a set of N linearly independent "orthonormal" signals $\phi_1(t), \phi_2(t), \ldots, \phi_N(t)$.
  - Orthogonality: the inner product between any two of them is equal to zero.

$$< \phi_i(t), \phi_j(t) >= \int_0^T \phi_i(t)\phi_j(t)dt = 0 \tag{4}$$

  - Normalised in energy

$$E =< \phi_i(t), \phi_i(t) >= \int_0^T \phi_i^2(t)dt = 1 \tag{5}$$

Geometric Representation of Signals

Signal Space

- In digital modulation, a signal spaced formed by two signals $(\phi_1(t), \phi_2(t))$ is necessary.
- A 2-dimensional Hilbert space is particularly known as a Euclidean space.
- In the previous lecture it was already seen that the signals $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are orthogonal to each other. Therefore, they can be used as the basis signals for the Euclidean space.
- In order to make them orthonormal they energy should be unity. The energy of a sinusoidal signal is given by

$$E = A^2 T/2 \tag{6}$$

  where A is the amplitude of the sinusoid.

Geometric Representation of Signals

Signal Space

- An orthonormal basis is then achieved by choosing:

$$\phi_1(t) = \sqrt{\frac{2}{T}}\cos(2\pi f_c t) \tag{7}$$

$$\phi_2(t) = \sqrt{\frac{2}{T}}\sin(2\pi f_c t) \tag{8}$$

- The signal space diagram where the two signals above are used as basis is called signal space diagram.
- Any other signal can be represented by its projections over those two basis functions.

$$s(t) = s_1(t)\phi_1(t) + s_2(t)\phi_2(t) \tag{9}$$

Geometric Representation of Signals
Signal Space

Signal space

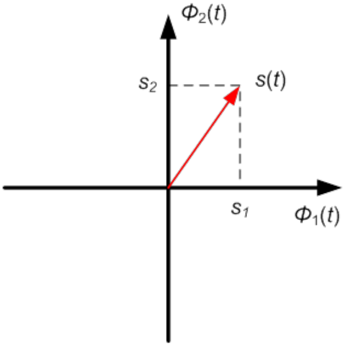


Figure: Signal space representation

Signal Space Representation

Digital Modulation Schemes

- In a digital system, the given transmitted signal $s(t)$ can only take a set of M finite values at every sample instant $\mathbf{s}_i \in \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M\}$. When using a signal space to represent $s(t)$, each of the M possible values will be represented by its projection over the two basis functions as in (3.8), like a vector

$$\mathbf{s}_i = (s_{i_1}, s_{i_2}) \tag{10}$$

- The representation of each of the M values on the signal space is called **constellation point** or **symbol**.
- The set of M possible transmitted symbols is called **signal constellation** or **signal space representation**.

Digital Modulation Schemes

Signal Constellation

- Example of signal constellation

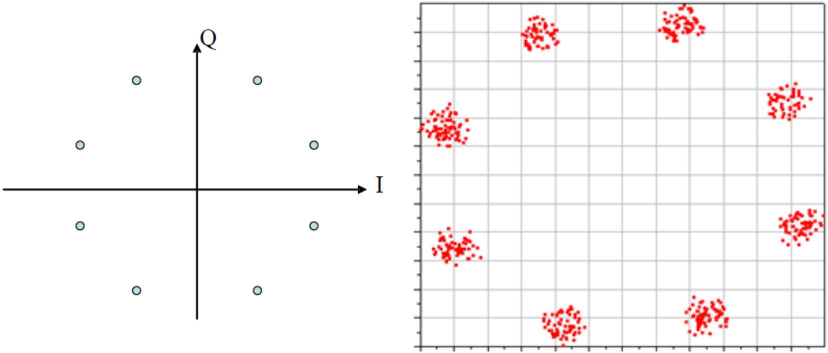


Figure: 8-PSK Constellation diagram at transmitter(left) and receiver (right)

Digital Modulation Schemes

Signal Constellation

- Consequently, digital modulation is equivalent to choosing appropriate mappings of a set of bits onto a symbol of the signal space at each *sampling instant*. This is why digital modulation is also called **digital mapping** or **symbol mapping**.
- In the absence of distortion, noise, carrier and symbol timing error offset, the signal space diagram of the transmitter and the signal space diagram at the receiver are the same.
- Another way to represent the constellation diagrams is using **eye diagrams**.
- The “wider” and the “higher” the opening of the eyes, the stronger the signal will be against jitter and noise respectively.

Digital Modulation Schemes

Signal Constellation

- Eye diagram of a binary PAM signal is created by overlapping and displaying multiple segments of the received signal over the duration of a few symbol periods.
- The result is known as an eye diagram or eye pattern due to its resemblance to the human eye.

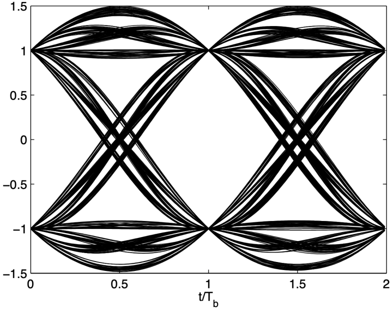


Figure: Binary PAM eye diagram

Digital Modulation Schemes

Signal Constellation

- In the digital system, the signal space diagram is a representation of the signal at the sampling instant.
- Signal space diagrams give the value of the signal at the sampling instant only.
  - Phase and amplitude information may be read directly from this diagram.
- The amplitude of every symbol (or distance between the symbol and the constellation origin) is calculated in the same way as a vector length would be

$$|\mathbf{s}_i| = \sqrt{s_{i_1}^2 + s_{i_2}^2} \tag{11}$$

- The distance between two constellation points

$$|\mathbf{s}_i - \mathbf{s}_j| = \sqrt{(s_{i_1} - s_{j_1})^2 + (s_{i_2} - s_{j_2})^2} \tag{12}$$

- The distance between two points in signal space is directly related to the difference in voltage between two samples.

Digital Modulation Schemes
Signal Constellation

- The minimum distance between two points is an important indicator of the system performance
  - large distance $\rightarrow$ lower probability of making errors.
- Average energy of transmitted signal

$$E = \frac{1}{M}\sum_{i=1}^{M} A_i^2 \tag{13}$$

  where M is the number of symbols in the constellation and A_i is the amplitude of the ith symbol.
- For the same average energy, the bigger the constellation the smaller the distance between the points. Consequently, the smaller the robustness of the modulation scheme against noise and jitter.
- The example in the following slide demonstrates this point.

Digital Modulation Schemes
Signal Constellation

- Example of two signal constellations with same average energy.

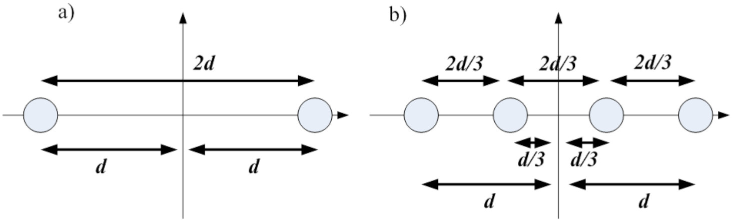


Figure: Symbol separation in two constellations with the same average energy

- Distance in signal space diagrams may be used in conjunction with knowledge of noise and distortion in a communication system to calculate the probability of making an error.

Digital Modulation Schemes

- Digital modulation schemes based on amplitude/phase modulation of the carrier signal are called linear modulation schemes.
- Popular linear modulation schemes
  - Pulse Amplitude Modulation (PAM): information encoded in amplitude only.
  - Phase Shift Keying (PSK): information encoded in phase only.
  - Quadrature Amplitude Modulation (QAM): information encoded in both amplitude and phase.
- A number of n bits are mapped into each symbol. Therefore, the number of symbols is $M = 2^n$.
- Considering an input bit rate R into the modulator, in its output a symbol rate (f_{sym}) equal to R/n symbols per second (sym/s) or bauds (Bd) will be obtained.

Digital Modulation Schemes

M-ary PSK

- In M-ary PSK modulation, the in-phase and quadrature components are interrelated in such a way that the envelope remains constant. This results in a circular constellation.
- Without any pulse shape applied to the signal, for all PSK modulations the bandwidth (90% of signal power) of the bandpass signal is equal to

$$W = 2f_{\mathrm{sym}} \tag{14}$$

- If the signal is centred at baseband, the bandwidth is $W = f_{\mathrm{sym}}$.

Digital Modulation Schemes
BPSK

- Constellation: uses two phases separated by 180 of one carrier signal.

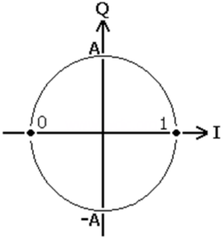


Figure: Signal space representation of BPSK

- The time domain representation.

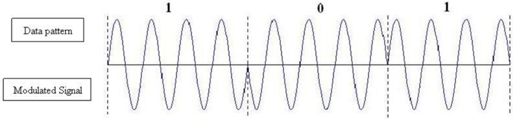


Figure: Waveforms of BPSK modulation

Digital Modulation Schemes
QPSK

- Constellation: uses 4 phases.

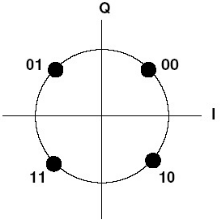


Figure: Signal space representation of QPSK

- The time domain representation.

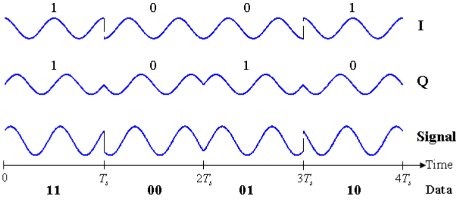


Figure: Waveforms of QPSK modulation

Digital Modulation Schemes
PAM & QAM

- PAM/QAM modulation corresponds to the transmission of one or more bits by the mapping different combinations of bits to different amplitude levels.
- Examples of PAM/QAM constellation

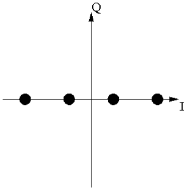


Figure: Signal constellation of 4-PAM

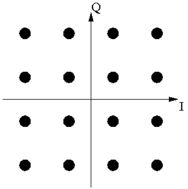


Figure: Signal constellation of 16-QAM

- As for PSK, for PAM and QAM modulations the bandwidth of the bandpass signal without any pulse shaping applied is $W = 2f_{\mathrm{sym}}$.

Digital Demodulation

Amplitude/Phase

- The addition of noise during the signal transmission produces a vertical and horizontal shift of the symbols from their ideal position in the signal space.
- Considering that all symbols have the same probability of being transmitted, optimum **decision regions** can be set in the signal space to decide the value of a received symbol.

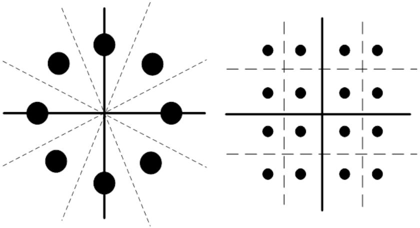


Figure: Signal decision regions for 8-PSK and 16 QAM modulation schemes

Digital Demodulation
Amplitude/Phase

- The higher the signal energy, the larger the decision regions.
- Therefore, the probability of error in the symbol detection becomes smaller.

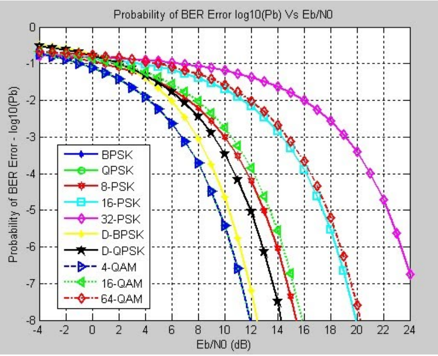


Figure: Digital modulation schemes performance comparison

Digital Demodulation
Amplitude/Phase

- As it can be seen in the previous figure, the bigger the average bit energy (E_b) versus the noise energy (N_0), the smaller the bit-error-rate (BER) for each modulation.
- In addition, for a same E_b/N_0, the modulation schemes with smaller amount of symbols present a lower BER than modulation schemes with a high number of symbols.

Carrier Timing and Frequency Offsets

- The demodulation performed previously which recovers the signal modulated on $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ made the assumption that a perfect copy of both $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ is available at the receiver. In practice, this is an invalid assumption for two reasons:
  - Carrier timing offset: The sinusoids at the receiver are offset in time by some set value τ from those in the transmitter.
  - Carrier frequency offset: The sinusoids at the receiver do not have the same frequency as those in the transmitter.

Carrier Timing Offset

- Imagine that, when attempting to demodulate using $\cos(2\pi f_c t)$, the receiver is offset in time by τ seconds relative to the transmitter. In this case we are demodulating
$$s(t) = x(t)\cos(2\pi f_c t) + y(t)\sin(2\pi f_c t) \tag{15}$$
by
$$\cos(2\pi f_c(t+\tau)) = \cos(2\pi f_c t + \psi) \tag{16}$$
where $\psi = 2\pi f_c \tau$.
- The resulting signal (after a LPF) is
$$x(t)\cos(\psi) - y(t)\sin(\psi) \tag{17}$$
- Similarly, demodulating by $\sin(2\pi f_c t)$ gives the quadrature component as
$$x(t)\sin(\psi) + y(t)\cos(\psi) \tag{18}$$

Carrier Timing Offset

- The in-phase and quadrature components are:

$$(x(t)\cos(\psi) - y(t)\sin(\psi)) + j\,(x(t)\sin(\psi) + y(t)\cos(\psi)) \tag{19}$$

  which is equal to

$$(x(t) + jy(t))\,e^{j\psi} \tag{20}$$

- Interpretation: A carrier offset error in the receiver corresponds to an anti-clockwise rotation of the signal space by $\psi = 2\pi f_c \tau$.

Example: A BPSK digital signal is transmitted over an ideal, noise-free channel at 2.4 GHz and received using a carrier offset in time by 52 ps ($1\text{ps} = 10^{-12}$ s). The sampling is perfect. Sketch the signal space diagram for the received constellation.

Carrier Frequency Offset

- Imagine that when attempting to demodulate with $\cos(2\pi f_c t)$, we are off by f_0 Hz. In this case we are demodulating by:

$$\cos(2\pi(f_c + f_0)t) = \cos(2\pi f_c t + \psi(t)) \tag{21}$$

  where $\psi(t) = 2\pi f_0 t$, a function of t.
- Following the same steps as for carrier timing offset, we can write the in-phase and quadrature components as

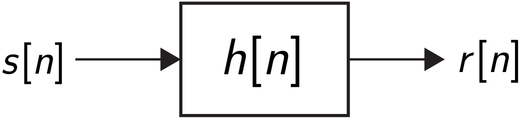
$$(x(t)\cos(\psi(t)) - y(t)\sin(\psi(t))) + j\,(x(t)\sin(\psi(t)) + y(t)\cos(\psi(t))) \tag{22}$$

  which is equal to $(x(t) + jy(t))\,e^{j\psi(t)}$.
- A frequency offset between the transmitter and receiver corresponds to an anti-clockwise rotation of the signal space by $\psi(t) = 2\pi f_0 t$.
  - $\psi(t)$ is a function of time.
  - The signal space diagram is further rotated at each sample.
  - The larger the frequency offset, the faster the rotation.

Nyquist Criterion for Distortionless Transmission

Time domain

- Given a digital system with input $s[n]$ and $r[n]$ output.



- A transmission is distortionless if the transmitted signal does not suffer ISI and therefore the input and output are equal, i.e. $r[n] = s[n]$ (without considering AWGN).

Nyquist Criterion for Distortionless Transmission
Time domain

- In the time domain, the condition for the output signal, $r[n]$, to be equal to the input signal, $s[n]$, is

$$r[n] = s[n] \text{ if } h[n] = \delta[n] \tag{23}$$

- In other words:

$$h[n] = h[nT] = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \tag{24}$$

- Consequently, from the time domain perspective, the Nyquist criterion states that the requirement for a distortionless channel is that the channel impulse response is equal to zero at the ideal sampling points nT.

Nyquist Criterion for Distortionless Transmission

Frequency domain

- The Fourier transform of the necessary channel impulse response to obtain a distortionless transmission is

$$h[n] = h[nT] = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \stackrel{\mathcal{F}}{\Longrightarrow} \sum_{k=-\infty}^{\infty} H(f - \frac{k}{T}) = T \tag{25}$$

  where T is the sampling period.
- In the frequency domain, this is equivalent to a flat frequency response
  - Note that before sampling, the frequency response is bandlimited but after sampling aliasing occurs and the frequency response is no longer bandlimited.
  - The frequency response of interest is the frequency response of the digital system and is thus the aliased version obtained by sampling.

Pulse-Shaping

- Digital data cannot be directly applied to the channel.
- The data is filtered before transmission.
- The filter should satisfy Nyquist criterion for distortionless transmission.
- By the Sampling Theorem, the sampling rate of our digital data is what determines its bandwidth.

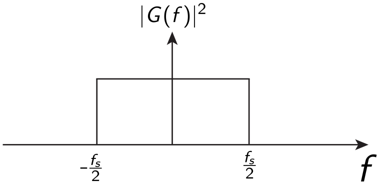
Ideal Pulse-Shaping

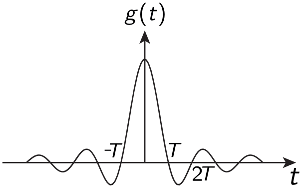
- Ideal low-pass filter (also called the brick-wall filter or Nyquist filter) with cut-off frequency of $f_s/2$, where f_s is the sampling frequency chosen as $f_s = f_{sym}$.
- The ideal low-pass filter satisfies the Nyquist Criterion for distortionless transmission.
- The ideal brick-wall filter has response

$$G(f) = \begin{cases} T & -\frac{f_s}{2} \leq f \leq \frac{f_s}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$g(t) = \mathrm{sinc}(2Wt) = \frac{\sin(2\pi Wt)}{2\pi Wt}$$

where $W = f_s/2$.





Ideal Pulse-Shaping

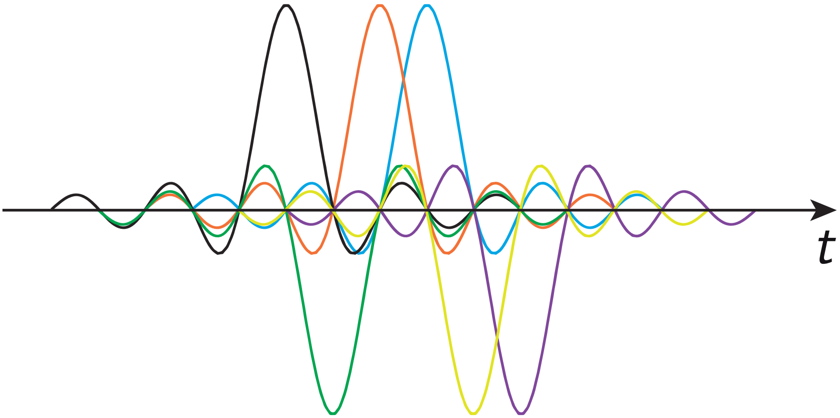


Figure: Sinc functions corresponding to digital signal 101100

Ideal Pulse-Shaping

- When using a brick-wall low-pass filter, the bandwidth of the bandpass transmitted digital is

$$W = f_{\mathrm{sym}} \tag{26}$$

- The brick-wall low-pass filter has a few characteristics that make it highly undesirable:
  - It is infinite in time, therefore it is necessary to approximate it by truncation.
  - It is non-causal so it introduces a delay into the system.
  - It is **highly sensitive to timing error** which might be fatal on the receiver to perform symbol detection.

Pulse-Shaping
Raised-Cosine Filter

- The only frequency response which fits in our bandwidth and satisfies all our constraints is the brick wall response.
- We introduce the concept of excess bandwidth:
  - The bandwidth of our signal is directly related to our sampling/symbol rate.
  - If we allow the shaping pulse to have a bandwidth greater than half the sampling frequency, $f_s/2$ then we have more degrees of freedom in our choice of pulse shape.
    - The overall bandwidth of the signal is increased, but the symbol rate is not.
    - This is termed excess bandwidth.
  - The raised cosine filter is a filter which satisfies the requirements of a pulse shaping filter with the use of excess bandwidth.
- With a raised cosine filter, the bandwidth of the bandpass transmitted digital is

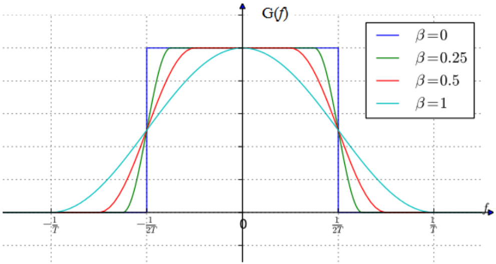
$$W = f_{sym}(1 + \beta) \tag{27}$$

Pulse-Shaping

Raised-Cosine Filter

- Frequency response

$$G(f) = \begin{cases} T & 0 \le |f| \le \frac{1-\beta}{2T} \\ \frac{T}{2} + \left\{1 + \cos\left[\frac{\pi T}{\beta}\left(|f| - \frac{1-\beta}{2T}\right)\right]\right\} & \frac{1-\beta}{2T} \le |f| \le \frac{1+\beta}{2T} \\ 0 & |f| \ge \frac{1+\beta}{2T} \end{cases} \tag{28}$$

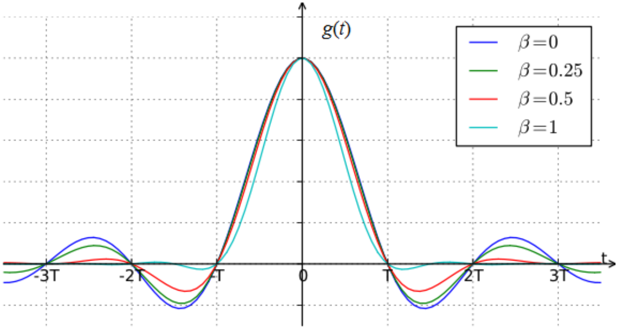


Pulse-Shaping

Raised-Cosine Filter

- Time domain

$$g(t) = \mathrm{sinc}(\frac{t}{T})\frac{\cos(\frac{\pi\beta t}{T})}{1-\frac{4\beta^2 t^2}{T^2}} \tag{29}$$



Pulse-Shaping

Raised-Cosine Filter

- The case of $\beta = 1$ (corresponding to 100% excess bandwidth) is called the full-cosine rolloff and has some useful properties:
  - The width of the pulse at half its full amplitude is equal to the symbol rate.
  - The time response crosses zero not only at $t = kT$, but also at $t = kT + \frac{T}{2}$
- These properties are useful in extracting timing information from the signal.
- Other important properties of RCF
  - Although, like the brick-wall filter, RCF has a non-causal time domain response, it falls off as $1/k^3$ and so we can get a good approximation by introducing a delay and truncating.
  - In this case, timing error leads to finite (hopefully small) error.

Pulse-Shaping
Raised-Cosine Filter

- A symbol-spaced filter can only control the filter response up to the half-sampling frequency $f_S/2$, therefore there are two implementation options for the RCF.
  - Since we are using bandwidth beyond this frequency, it is necessary to implement the pulse-shaping filter as either an analogue filter applied to the signal after the ADC.
  - In order to use a digital filter an option is to increase the sample rate of the data by interpolating. After the digital modulator $f_s = f_{\mathrm{sym}}$, after the interpolator $f_s = Nf_{\mathrm{sym}}$, where N is the interpolation factor. Note that increasing the sample rate of a signal does not affect its bandwidth.

Pulse-Shaping

Channel Model

- The discrete channel model is the sampled version of **all filtering** which occurs between the discrete input and discrete output of the all-digital model.
  - All filtering typically comprises: transmit filter, “physical” channel response and receive filter.
  - The transmit and receive filters are each designed such that the analogue transmission is appropriately pulse shaped.
- The discrete channel is modelled as

$$h[k] = h(kT) \tag{30}$$

  with $h(t) = g(t) * \hat{h}(t)$ where $g(t)$ is the response of the transmit and receive filters and $\hat{h}(t)$ is the “physical” channel. In a lot of systems, $g(t)$ is equivalent to the pulse-shaping filter response.
- The overall discrete model has not changed - the channel has been redefined to include all filtering.

Pulse-Shaping
Receiver Side

- In wireless communications, pulse shaping can be carried out only on the transmitter side (as it has been seen so far) or on both transmitter and receiver sides.
- The second option provides certain advantages that will be analysed in Lecture 4. For this configuration the filter on the transmitter side is called "pulse shape filter" ($p(t)$) and the filter on the receiver side is called "matched filter" ($p^*(-t)$).
- When cascaded
$$g(t) = p(t) * p^*(-t) \tag{31}$$
- As in the previous case, the overall system model does not change.