Data Transmission over Bandlimited Channels

This lecture describes limits on the maximum symbol and information rate for band-limited channels.
After this lecture you should be able to: determine if a channel meets the Nyquist no-ISI criteria and, if so, the maximum signalling rate without ISI; determine the maximum error-free information rate over the BSC and AWGN channels; determine the specific conditions under which these two limits apply. You should be able to perform computations involving the OFDM symbol rate, sampling rate, block size and guard interval.

Introduction

All practical channels are band-limited – either low-pass or band-pass. There are two theorems, the Nyquist no-ISI criteria and Shannon's capacity theorem, that provide some guidance about maximum data rates that can be achieved over a bandlimited channel.

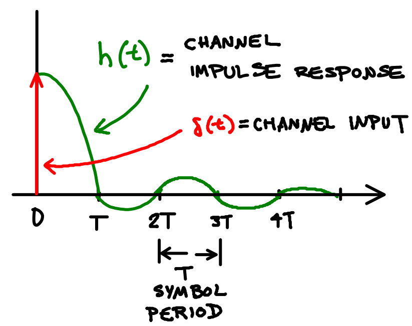
Inter-Symbol Interference

Bandwidth-limited low-pass channels attenuate higher-frequency components of a signal. This "rounds off" pulse shapes which increases their rise and fall times and extends their durations. Each symbol then extends into subsequently-transmitted symbols. This causes one symbol to interfere with subsequently-transmitted symbols. This is called inter-symbol interference (ISI).

Exercise 1: Draw a square pulse of duration T. Draw the pulse after it has passed through a linear low-pass channel that results in rise and fall times of $T/3$. Draw the output for an input pulse of the opposite polarity. Use the principle of superposition to draw the output of the channel for a positive input pulse followed by a negative input pulse.

Nyquist no-ISI Criteria in Time

Consider a system that transmits symbols that are infinitely-short pulses of different amplitudes. A low-pass channel will limit the rise time of these impulses and cause them to be extended in time. However, if the response of the channel to these impulses is zero after one symbol period then the impulse will not cause ISI to the next impulse. And if the channel impulse response passes through zero at all future multiples of the symbol period then the impulses will not interfere with subsequent impulses. This is the Nyquist no-ISI condition stated in the time domain.



Exercise 2: What is the impulse response of a channel that does not alter its input? Does this impulse response meet the Nyquist condition? Will it result in ISI?

An example of an impulse response that meet this criteria is the $\text{sinc}()$ function:

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

which has value 1 at $t = 0$ and 0 at multiples of T.

Exercise 3: Draw the impulse response of a channel that meets the Nyquist condition but is composed of straight lines.

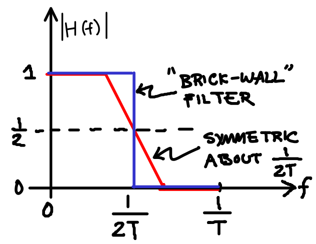
Nyquist no-ISI Criteria in Frequency

It is possible to determine the conditions for a channel's transfer function to result in no ISI. A common way of stating this condition is that the channel's frequency response must have odd symmetry around half of the symbol frequency[1] ($\frac{1}{2T}$):

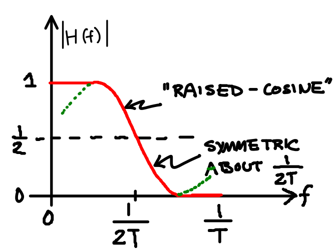
$$H(\frac{1}{2T} + f) + H^*(\frac{1}{2T} - f) = 1 \text{ for } 0 \leq |f| \leq \tfrac{1}{2T}$$

[1]The asterisk indicates complex conjugate. This can be ignored for real(izable) baseband channels.

Just as there could be many impulse responses that are zero at multiples of the symbol period, there are many transfer functions that result in no ISI. For example, the following two straight-line transfer functions meet the no-ISI condition[2]:



The "brick-wall" filter (blue) has a response that is 1 below half of the symbol rate ($\frac{1}{2} \cdot \frac{1}{T} = \frac{1}{2T}$ where $\frac{1}{T}$ is the symbol rate) and zero above that. Although such a filter would have the minimum bandwidth required to meet the Nyquist condition for a symbol period T, it is not physically realizable and has other problems as described below. The filter with the straight-line transfer function is more practical but still difficult to implement. A common and more practical transfer function is the so-called raised-cosine function which is a half-cycle of a cosine function offset to have a minimum value of zero and centered around half of the symbol rate:



Note that it is the symmetry around the frequency $\frac{1}{2T}$ that ensures there will be no ISI rather than the exact filter shape. Thus we are free to implement other transfer functions if they make the implementation easier.

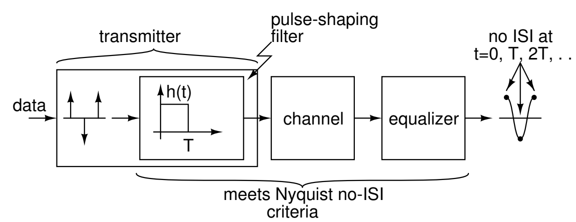
Exercise 4: Draw the magnitude of a raised-cosine transfer function that would allow transmission of impulses at a rate of 800 kHz with no interference between the impulses.

Often we cannot control over the impulse response or transfer function of the channel and we need to add filtering at the transmit or receive sides of the channel so that the overall transfer function meets the Nyquist criteria. This is called equalization and is described below.

Pulse-Shaping Filter

Note that the no-ISI criteria ensures that a channel produces no ISI when transmittign *impulses*, not for the square pulses typically used by line codes.

However, we can treat the transmitter as including a filter that converts impulses to pulses. We then consider that the overall channel includes this (im)pulse-shaping filter. So for transmitters that generate pulses it is the combination of this hypothetical impulse-shaping filter and the channel that has to meet the Nyquist criteria:



Exercise 5: Draw the impulse response of a filter than converts input impulses to pulses of duration T? Draw the signal after the pulse-shaping filter in the diagram above.

Equalization

To avoid ISI, the total channel response including transmit filters, the channel and the receiver filter(s) have to meet the Nyquist no-ISI condition.

When the channel by itself doesn't meet the no-ISI conditions, the transmitter and/or receiver can use a filter called an equalizer that modifies the overall transfer function to ensure the no-ISI condition is met.

Transmitter and receiver filters typically have other functions beside equalization. For example, the transmit filter may limit the bandwidth of the transmitted signal to limit interference to users on adjacent channels. The receiver filter may filter out noise and interference from adjacent channels and thus improve the SNR and SIR. The communication system designer would design the transmitter and receiver filters to meet both the filtering and equalization requirements.

[2]For simplicity we only show one component (the real or imaginary portion) of the transfer function.

A common situation is a flat channel where interference is not an issue. In this case a reasonable approach is to put half of the filtering at the transmitter and half at the receiver. In order to achieve an overall raised cosine transfer function, each side has to use a "root raised cosine" (RRC) transfer function. The product of the two filters is thus the desired raised-cosine function which meets the no-ISI condition.

Equalizers also have to compensate for the (imaginary) pulse-shaping filter. Since the pulse-shaping filter has a low-pass ($\text{sinc}(f)$) shape, the equalizer response has more gain at higher frequencies that a true raised-cosine function [3].

Excess Bandwidth

Channels can have different transitions between the passband and the stopband of the transfer function while still meeting the no-ISI conditions.

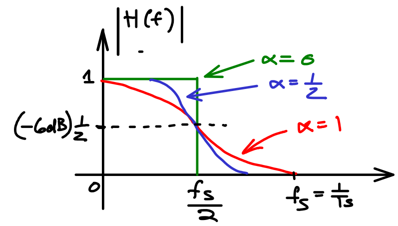
A parameter, α, which defines how much wider the channel is than the minimum is called the "excess bandwidth" parameter. It is defined as:

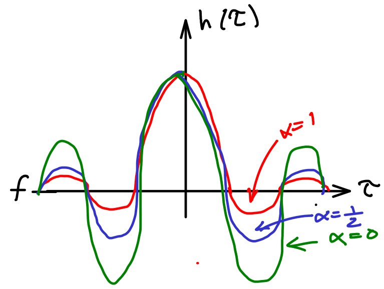
$$\alpha = \frac{\text{total bandwidth} - \text{minimum bandwidth}}{\text{minimum bandwidth}}$$

for example, if B is the maximum channel bandwidth (frequency at which $H(f) = 0$) and T is the symbol period so that $\frac{1}{2T}$ is the minimum possible bandwidth, then $\alpha = \frac{B-1/(2T)}{1/(2T)} = 2BT - 1$.

Why would we make the bandwidth larger than necessary? The value of α affects the shape of the impulse response. Larger values of α result in less overshoot and make the received pulse more "square" and this in turn makes the receiver less sensitive to variations in when the receiver samples the received signal.

The following diagram shows how α for a raised-cosine transfer function affects the impulse response:





Larger values of excess bandwidth (wider bandwidth channels) results in less "ringing" of the impulse response which in turn reduces the amount of ISI near the sampling point. This makes the receiver less sensitive to errors in where (when) it samples the received signal.

Nyquist Criteria and Bit Rate

Note that the symbol rate limitations defined by the Nyquist criteria *do not* determine the maximum bit rate that can be achieved over a channel – they only determine the maximum *symbol* rate *without ISI*.

We can increase the bit rate by increasing the number of bits per symbol (e.g. by using multiple levels). For example, using symbols each of which could be at one of 1024 levels we can transmit 10 bits per symbol.

Exercise 6: A channel has a 3 kHz bandwidth and meets the Nyquist non-ISI conditions with $\alpha = 1$. How many levels are required to transmit 24 kb/s over this channel using multi-level signalling?

We can also design receivers that recover the transmitted data even in the presence of ISI. For example, if we know the symbols that have been transmitted in the past and we know the channel impulse response then we can predict the ISI and subtract it from the current received symbols. This is called decision-feedback equalization (DFE).

[3] Sometimes called "sinc compensation."

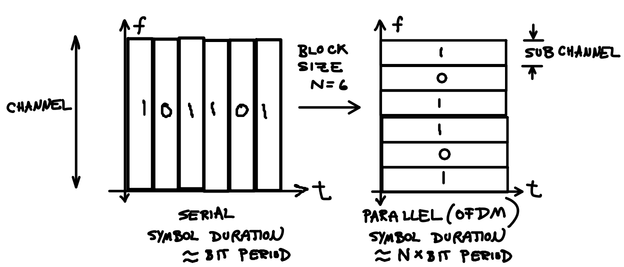
Adaptive Equalizers

In many communication systems the transfer function of the channel cannot be predicted ahead of time. One example is a modem used over the public switched telephone network (PSTN). Each phone call will result in a channel that includes different "loops" and thus different frequency responses. Another example is multipath propagation in wireless networks. The channel impulse response changes as the receiver, transmitter or objects in the environment move around.

To compensate for the time-varying channel impulse response the receiver can be designed to adjust the receiver equalizer filter response using various algorithms.

OFDM

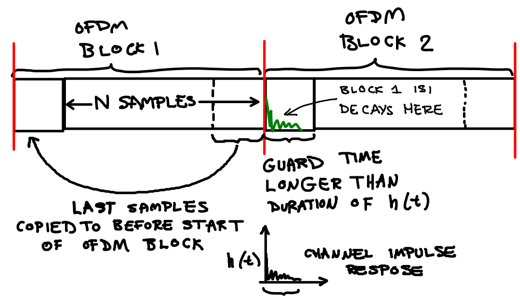
An alternative to equalization is a technique called Orthogonal Frequency Division Multiplexing (OFDM). An OFDM transmitter collects a group of N symbols at a time and uses them to modulate N "subcarriers" (modulation is covered in another course). These subcarriers are transmitted in parallel over the same time duration that would have been required to transmit the N symbols serially. The net effect is to reduce the symbol rate by a factor N but with no impact on the overall bit rate.



We usually insert a "guard time" (or "guard interval") in-between symbols. Its duration is longer than the duration of the channel impulse response.

Since no data is transmitted during the guard time, this reduces the average data rate. However, the guard time is typically a small fraction of the OFDM symbol duration and so the impact on the overall throughput is relatively small.

Rather than transmitting nothing during the guard interval, a small part of the end of the block of N samples is copied to the start of the symbol and transmitted during the guard time. This is called a "cyclic" or "periodic" extension.



The value of N is typically a power of 2 to allow efficient implementation using the Fast Fourier Transform (FFT) algorithm.

OFDM has become more popular than adaptive equalization because it is simpler to implement and more robust. This is partly because it is not necessary to estimate the channel to correct for ISI. OFDM is used by most ADSL, WLAN and 4G cellular standards.

Exercise 7: The 802.11g WLAN standard uses OFDM with a sampling rate of 20 MHz, with $N = 64$ and guard interval of $0.8\mu s$. What is the total duration of each OFDM block, including the guard interval? How long is the guard time?

Shannon Capacity

The Shannon Capacity of a channel is the information rate above which it is not possible to transmit data with an arbitrarily low error rate.

One example of a channel is the Binary Symmetric Channel (BSC). This channel transmits discrete bits (0 or 1) with a bit error probability (BER) of p. The capacity of the BSC in units of information bit per "channel use" (transmitted bit) is :

$$C = 1 - (-p \log_2 p - (1-p) \log_2 (1-p))$$

which is 0 for $p = 0.5$ (when each transmitted bit is equally likely to be received right or wrong) and 1 when $p = 0$ (the error-free channel) or when $p = 1$ (a perfectly inverting channel).

Exercise 8: What is capacity of a binary channel with a BER of $\frac{1}{8}$ (assuming the same BER for 0's and 1's)?

For a continuous channel corrupted by Additive White Gaussian Noise (AWGN), the capacity can be shown to be:

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

where C is the capacity (b/s), B is the bandwidth (Hz) and $\frac{S}{N}$ is the signal to noise (power) ratio.

The Shannon limit does not say that you can't transmit data faster than this limit, only that if you do, you can't reduce the error rate to an arbitrarily low value. However, in practice, attempting to transmit at information rates above capacity results in high error rates.

Exercise 9: Can we use compression to transmit information faster than the (Shannon) capacity of a channel? To transmit data faster than capacity? Explain.

Shannon's work also does not specify how to achieve capacity. However, Shannon's work does hint that using error-correcting codes with long codewords (to be discussed later) should allow us to achieve arbitrarily-low error rates as long as we limit the information rate to less than the channel capacity.

Exercise 10: What is the channel capacity of a 4 kHz channel with an SNR of 30dB?

Some systems using modern forward error-correcting (FEC) codes such as Low Density Parity Check (LDPC) codes can communicate at very low error rates over AWGN channels with SNRs only a fraction of a dB more than the minimum required by the capacity theorem.

Exercise 11: What do the Nyquist no-ISI criteria and the Shannon Capacity Theorem limit? What channel parameters determine these limits?