Data Transmission over Bandlimited Channels

In many cases the channel bandwidth limits the maximum symbol rate or data rate. This lecture describes two ways we can estimate the maximum symbol or data rate that can be transmitted over a band-limited channel.

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 an AWGN channel; 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) and the channel bandwidth limits the maximum data rate. We will study 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.

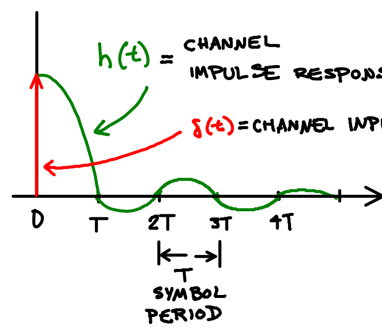
Inter-Symbol Interference

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 interference 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 as (infinitely-)short pulses of different amplitudes. These are called "impulses." A low-pass channel will limit the rise time of the impulses and cause them to be "smeared" in time. However, if the response of the channel to these impulses is zero at multiples of the symbol period then one impulse will not cause ISI to subsequent impulses if we sample at these zero-crossing times. This is called the Nyquist no-ISI criteria.



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

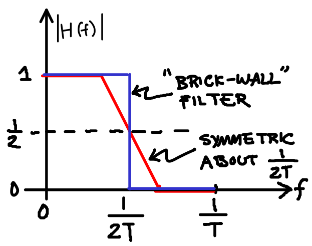
It is possible to derive the characteristics of the channel's frequency-domain transfer function that result in no ISI. One way of stating this condition is that the channel's frequency response have odd symmetry around half of the symbol frequency:

$$H(\frac{1}{2T} + f) + H(\frac{1}{2T} - f) = 1 \text{ for } 0 < |f| < \frac{1}{2T}$$

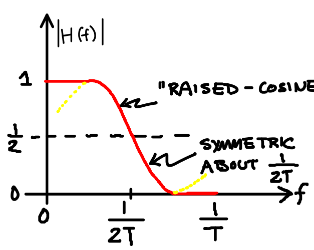
Note that this condition applies to the complex frequency response. Thus both the real and imaginary parts of $H(f)$ need to have this symmetry[1].

Often we have little 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.

Just as there could be many impulse responses that are zero at multiples of the symbol period, there are many no-ISI transfer functions. 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 ($\frac{1}{2} \times \frac{1}{T} = \frac{1}{2T}$ $\frac{1}{T}$ is the symbol rate) and zero above that. Although such a filter would have the minimum overall bandwidth required 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 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 $1/2T$ that ensures there will be no ISI rather than the exact filter shape. Thus we are free to implement other transfer functions, possibly arbitrary ones, 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.

Pulse-Shaping Filter

Note that the no-ISI criteria applies for a channel that produces no ISI for *impulses*, not for the square pulses typically used by line codes. Since practical systems don't transmit impulses, the Nyquist criteria cannot be used to evaluate the channel.

Instead, we can pretend that the transmitter includes a filter that converts input 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? What is the shape of the frequency response of this filter? *Hint: the Fourier transform of a pulse of duration T is* $\frac{sin(\pi f T)}{\pi f}$. What is the "bandwidth" of this filter – when is it first zero? How does this compare to the "bandwidth" of the raised-cosine filter above?

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:

$$\alpha = \frac{\text{total bandwidth}}{\text{minimum possible bandwidth}} - 1$$

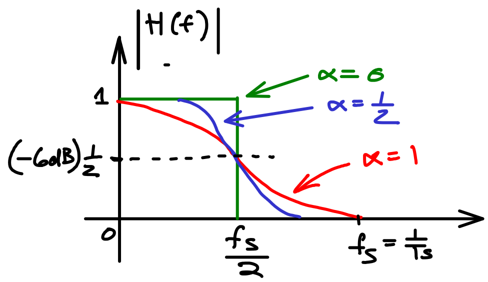
Exercise 6: What is the possible range of values of α?

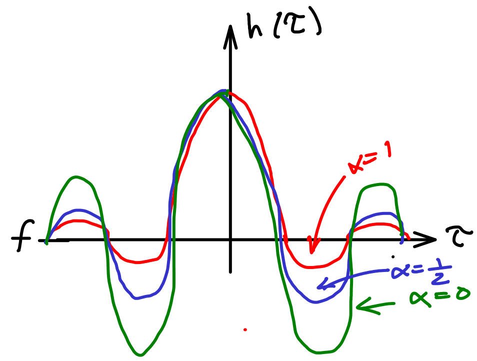
Why would we make the bandwidth larger than necessary? The value of α has an impact on the shape of the impulse response and this in turn affects the sensitivity of the receiver to errors in when the receiver samples the received signal.

[1] While we often only show the magnitude of the frequency response, the phase response is usually linear and this can be used to determine the ratio of real to imaginary components.

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

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.

Equalization

To avoid ISI, the total channel response including the pulse-shaping filters, 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 modify 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.

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.

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.

Nyquist Criteria and Bit Rate

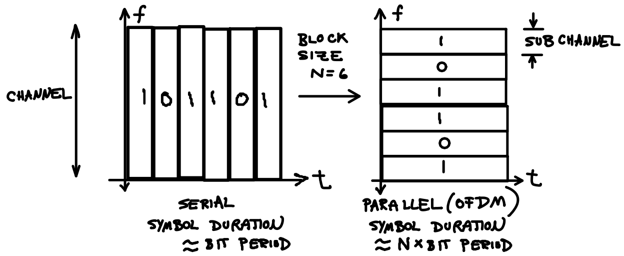
Note The symbol rate limitations defined by the Nyquist criteria *do not* determine the maximum bit rate that can be achieved over a channel – they only determine the maximum *symbol* rate *without ISI*.

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

We can also design receivers that recover the transmitted data even in the presence of ISI. For example, if we know the data that was transmitted previously and we know the channel impulse response then we can predict the ISI and subtract it from the received signal.

OFDM

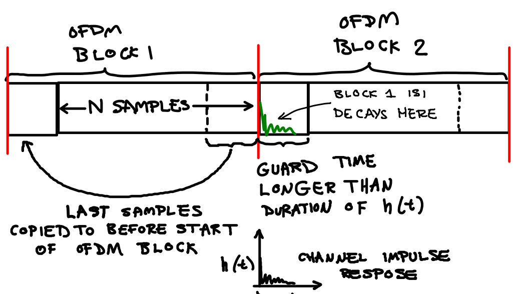
An alternative to equalization is a technique called Orthogonal Frequency Division Multiplexing (OFDM). An OFDM transmitter groups N symbols and uses them to modulate N "subcarriers" (modulation is covered in another course). These subcarriers are transmitted in parallel over the same time duration that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



We usually insert a "guard time" (or "guard interval") in-between symbols that is longer than the duration of the channel impulse response. Typically this will be several times the symbol duration that would have been used without OFDM.

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 is relatively small.

Rather than transmitting nothing during the guard interval, a small part of the end of the block of N samples are 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 modern 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 many guard samples are used?

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.

Different types of channels will have different capacities. For the Additive White Gaussian Noise (AWGN) channel the capacity is:

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

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 8: Can we use compression to transmit data faster than the Shannon 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 information rate to less than the channel capacity.

Exercise 9: 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 10: What are some differences between the signalling rate limit imposed by the Nyquist no-ISI criteria and the Shannon Capacity Theorem? For example, what do they limit and what parameters determine these limits?