

Fundamentals of Digital Communication
by Upamanyu Madhow
Cambridge University Press, 2008

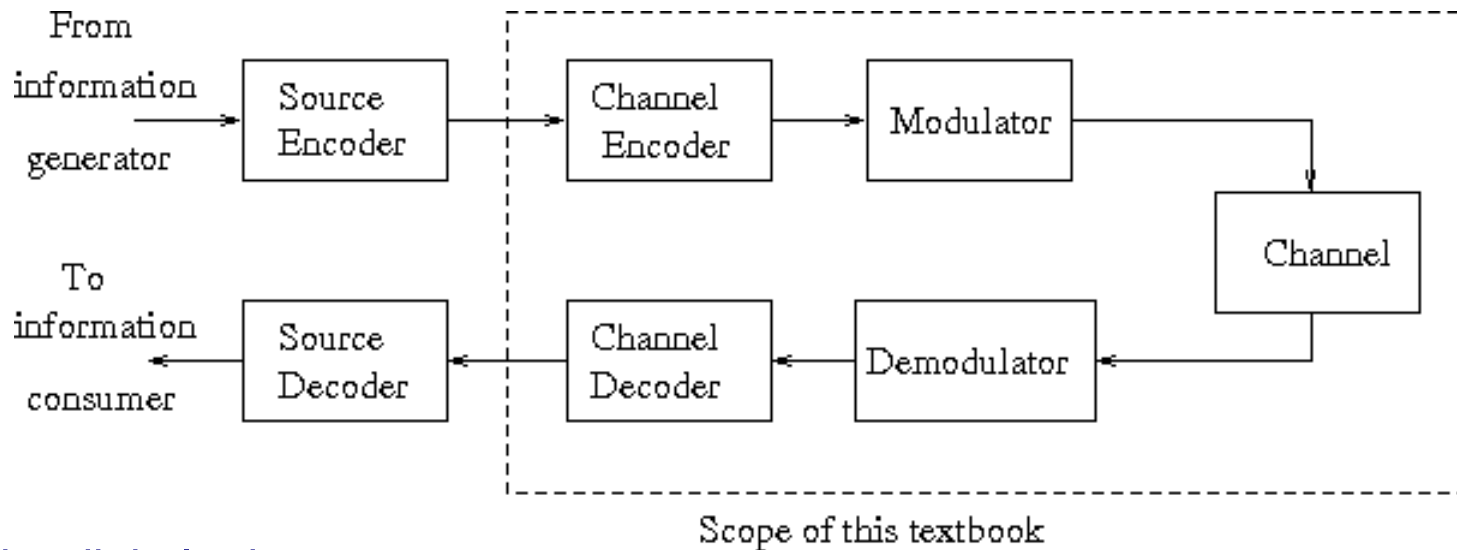
Lecture Outline for Chapters 1 and 2
Copyright by Upamanyu Madhow, 2009

The transition from analog to digital

- Communication: Information transmission between two points
 - in space (telephony, web browsing,...)
 - in time (recording media--CDs, DVDs, hard drives,...)
- Inexorable transition from analog to digital
 - Analog cellular to digital cellular (CDMA, GSM, OFDM)
 - Analog TV/radio to Digital TV/radio
 - LPs to CDs, VHS to DVD
- Content is often analog (speech, image, video)
- Signals sent over physical channels are analog
 - Currents, voltages, EM waves are continuous-valued, continuous-time functions
- So why digital communication?

Why digital?

Block diagram of a digital communication link

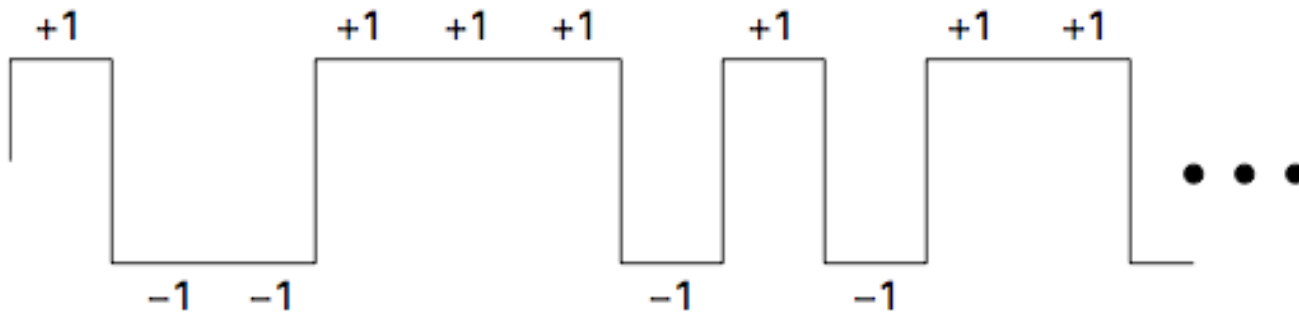


The digital advantage

- **Universality:** Any source can be converted to digital format with prescribed fidelity (economies of scale by multiplexing diverse content on links and networks)
- **Channel-optimized design:** Encoder/decoder and modulator/demodulator can be optimized for channel, without worrying about source characteristics
- **Networking:** Bits can be perfectly regenerated after every link, enabling communication networks (cascading analog links leads to deterioration in signal quality)

Chapter 2: Modulation

- Converting bits into signals that can be sent over (or recorded on) channels



- Example of binary antipodal signaling. rectangular pulse modulated by $+1$ and -1 to send 0s and 1s
(typical convention: map 0 to $+1$, and 1 to -1)
- Sharp transitions in rectangular pulse mean large frequency occupancy: may be OK for wires, but not for bandwidth-constrained wireless channels
- Need a framework for modulator design based on understanding channel constraints (trading off power and bandwidth)

Agenda

- Baseband and passband channels
- Unified modulation design framework
 - Based on complex baseband representation of passband signals and systems
- Let us start with signals and systems review (with comm-centric examples)

Signals and Systems Review

Signals and Systems Review: Outline

- Complex numbers
 - Euler's identity
- Inner product
 - Norm, energy
- Fourier transform
 - Formula
 - Duality (when switching roles of time and freq, change sign of argument)
 - Properties: Convolution/multiplication, Parseval, linearity, time/freq shift
 - Pairs: Delta function/constant, sinc function/boxcar
- Bandwidth

Complex numbers

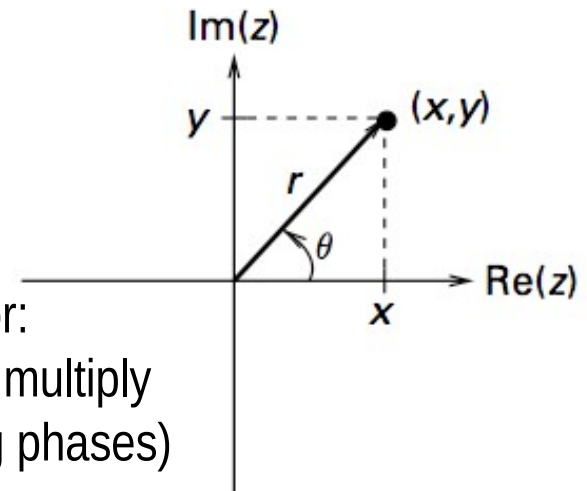
Complex numbers provide a compact way of describing amplitude and phase (and the operations that affect them, such as filtering)

Complex number $z = x + jy$ (x and y real-valued, $j = \sqrt{-1}$.)

Often useful to interpret a complex number as a point in 2-D plane

Cartesian (rectangular) coordinates: $x = \text{Re}(z)$ $y = \text{Im}(z)$

Polar coordinates
 $r = |z| = \sqrt{x^2 + y^2},$
 $\theta = \arg(z) = \tan^{-1} \frac{y}{x}$



But a complex number is more than just a 2D real vector:
mainly because of complex multiplication (one complex multiply
requires four real multiplies, and corresponds to adding phases)

Euler's identity (we use this very often) $e^{j\theta} = \cos \theta + j \sin \theta$

Signal geometry: inner product

The Big Picture: Euclidean geometry is important for communication system designers

1) Continuous-time signals are just like vectors

--standard ideas from Euclidean geometry apply

2) We typically transform continuous-time signals into discrete-time vectors (filtering and sampling) before signal processing

3) Vector manipulation therefore important for both theory and algorithms

Inner product is the key concept in defining signal geometry

$$\langle \mathbf{s}, \mathbf{r} \rangle = \sum_{i=1}^m s[i] r^*[i] = \mathbf{r}^H \mathbf{s}$$

discrete time

$$\langle s, r \rangle = \int_{-\infty}^{\infty} s(t) r^*(t) dt$$

continuous time

(Complex-valued signals needed for our unified framework)

Linearity

$$\begin{aligned} \langle a_1 s_1 + a_2 s_2, r \rangle &= a_1 \langle s_1, r \rangle + a_2 \langle s_2, r \rangle \\ \langle s, a_1 r_1 + a_2 r_2 \rangle &= a_1^* \langle s, r_1 \rangle + a_2^* \langle s, r_2 \rangle \end{aligned}$$

(note that constants in second argument get conjugated when pulled out)

Signal geometry: norm, energy, distance

Signal energy is its inner product with itself: $E_s = ||s||^2 = \langle s, s \rangle = \int_{-\infty}^{\infty} |s(t)|^2 dt$

Norm is the square root of the energy: $||s||$

Distance between two signals is the norm of their difference: $||s_1 - s_2||$

We use these concepts extensively in Chapter 3

--signals sent should be “far enough” apart that we can distinguish them in the presence of noise

--transmit power depends on the energy of the signals sent

Cauchy-Schwartz inequality (file away for later use)

$$|\langle s, r \rangle| \leq ||s|| ||r||$$

Equality if and only if one of the signals is a scalar multiple of the other

Convolution

Communications channels often modeled as LTI systems, so convolution is a key modeling tool: modulated signal convolved with an LTI system, then noise added

Recall the basic definition: $q(t) = (s * r)(t) = \int_{-\infty}^{\infty} s(u)r(t-u)du$

Linearity and shift invariance

$$(a_1 s_1(t - t_1) + a_2 s_2(t - t_2)) * r(t) = a_1 (s_1 * r)(t - t_1) + a_2 (s_2 * r)(t - t_2)$$

(notational abuse in not distinguishing dummy variable is often convenient)

Good engineers love the impulse (delta) function, defined by:

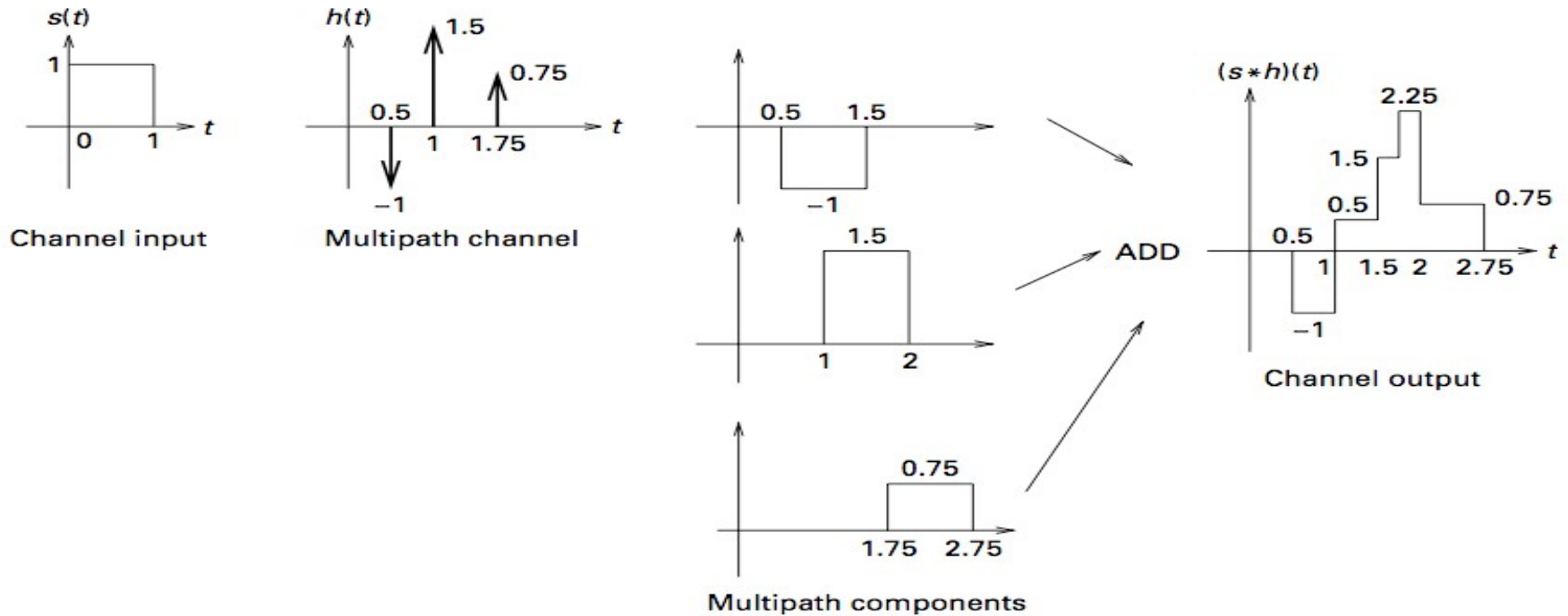
$$\int_{-\infty}^{\infty} \delta(t - t_0) s(t) dt = s(t_0) \quad (\text{visualize as a tall thin pulse of unit area})$$

Convolution with the shifted delta function inserts a time shift:

$$\delta(t - t_0) * s(t) = s(t - t_0).$$

So we can use the delta function to model wireless multipath channels

Example: a multipath channel



$$y(t) = (s * h)(t) = \sum_{i=1}^M a_i s(t - t_i)$$

Output of multipath channel
(before adding receiver noise)

$$h(t) = \sum_{i=1}^M a_i \delta(t - t_i)$$

Multipath channel impulse response

Example: the matched filter

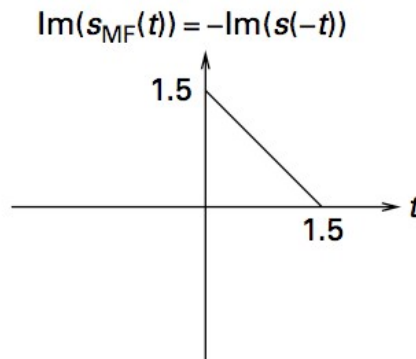
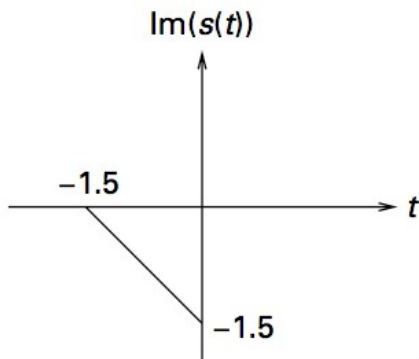
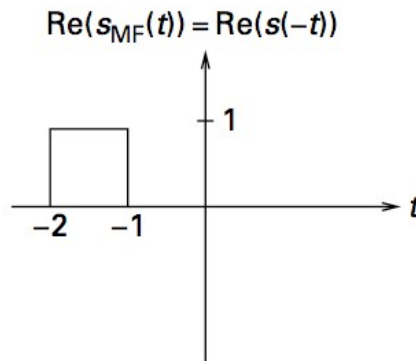
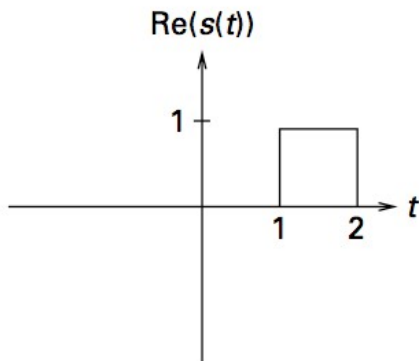
Matched filter for $s(t)$

$$s_{\text{MF}}(t) = s^*(-t)$$

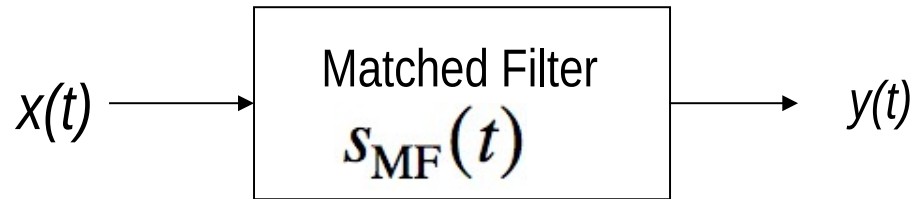
(time domain)

$$S_{\text{MF}}(f) = S^*(f)$$

(frequency domain)



Significance of the matched filter



$$y(t) = (x * s_{MF})(t) = \int_{-\infty}^{\infty} x(u) s_{MF}(t - u) du = \int_{-\infty}^{\infty} x(u) s^*(u - t) du$$

Matched filter performs a template match: correlates input with all possible shifts of the signal $s(t)$

Example: When the input is a shifted version of $s(t)$, the time shift can be estimated by “peak picking” at the output of the matched filter (see Problem 2.5)

Fundamental role of matched filter in communication theory explored in later chapters

Fourier Transform

Fourier Transform

$$s(t) = \int_{-\infty}^{\infty} S(f) e^{j2\pi ft} \, df$$

Inverse Fourier Transform

$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} \, dt$$

Notation for Fourier Transform Pair

$$s(t) \leftrightarrow S(f)$$

Time-Frequency Duality:

if $s(t)$ has Fourier transform $S(f)$, then the signal $r(t) = S(t)$ has Fourier transform $R(f) = s(-f)$.

(So we need to keep track of only half of the Fourier transform pairs)

Delta function \leftrightarrow Constant function (trivial but important pair)

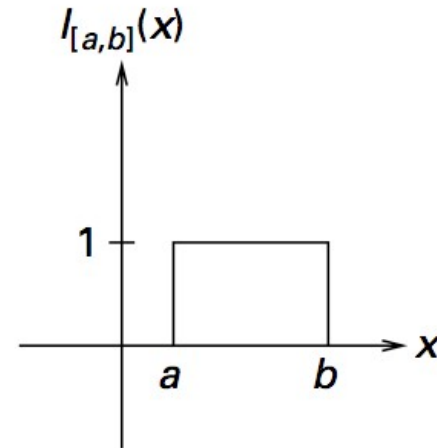
Indicator and boxcar functions

Indicator function of a set A

(very useful for compact notation)

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Boxcar (indicator function of an interval)

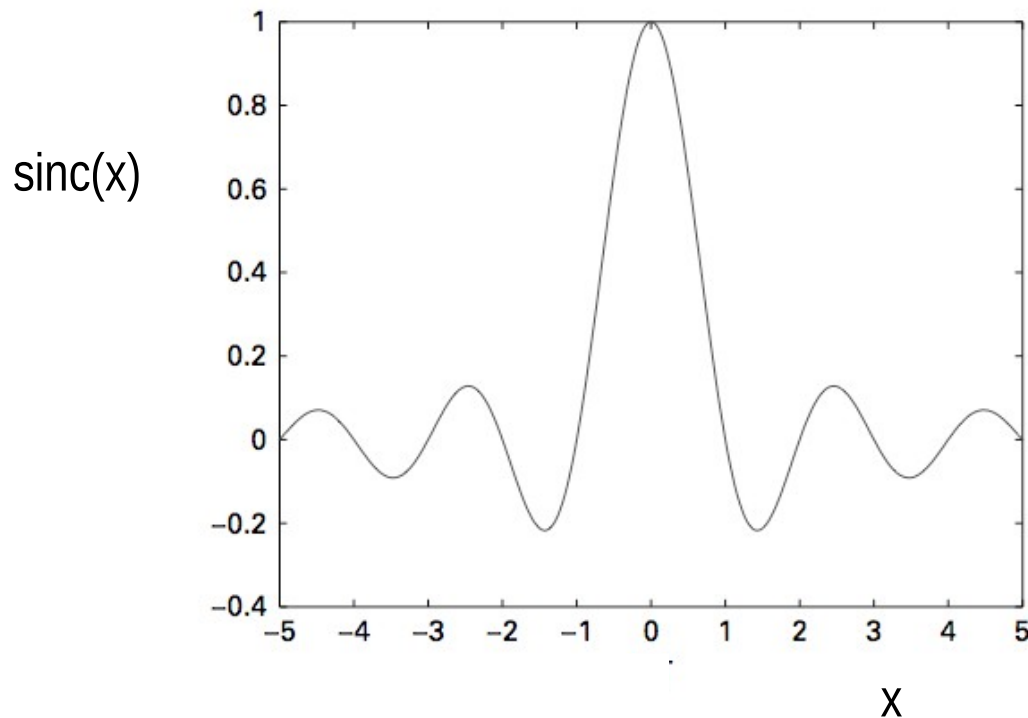


Sinc function

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

($\text{sinc}(0) = 1$, defined as the limit)

$\text{sinc}(x)$, decays as $1/|x|$ with sinusoidal fluctuations

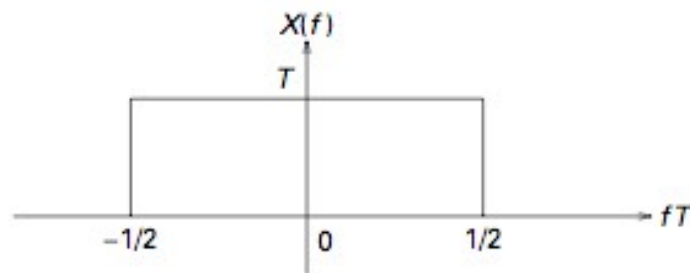


Boxcar and sinc are a Fourier transform pair

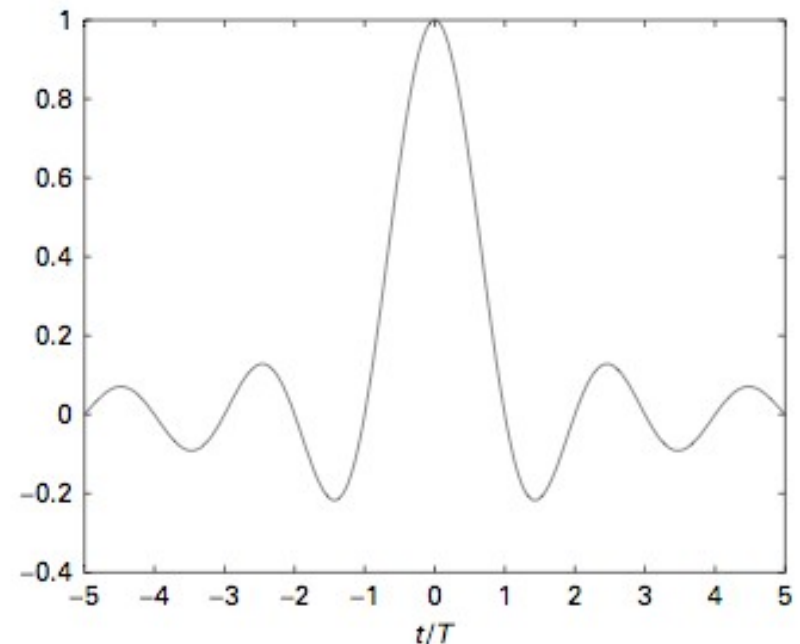
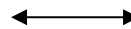
$$s(t) = I_{[-\frac{T}{2}, \frac{T}{2}]}(t) \leftrightarrow S(f) = T \operatorname{sinc}(fT)$$

Timelimited signals with sharp edges decay slowly with frequency: rectangular time domain pulse has a sinc spectrum, which exhibits $1/|f|$ decay

By duality, ideal bandlimited signal corresponds to time domain sinc (slow decay with time has bad implications, we shall see. and leads us to avoid sharp edges in frequency domain in our designs)



(a) Frequency domain boxcar



(b) Time domain sinc pulse

Fourier Transform Properties, I

Convolution \longleftrightarrow Multiplication

$$s(t) = (s_1 * s_2)(t) \leftrightarrow S(f) = S_1(f)S_2(f)$$

$$s(t) = s_1(t)s_2(t) \leftrightarrow S(f) = (S_1 * S_2)(f).$$

Example: Trapezoid \longleftrightarrow Product of sincs

Translation

$$s(t - t_0) \leftrightarrow S(f)e^{-j2\pi ft_0}$$

Time delay leads to freq-dependent phase lags

$$s(t)e^{j2\pi f_0 t} \leftrightarrow S(f - f_0).$$

Frequency shift leads to phase rotation

Scaling

$$s(at) \leftrightarrow \frac{1}{|a|} S\left(\frac{f}{a}\right)$$

Time compression leads to bandwidth expansion

Fourier Transform Properties, II

$$s^*(t) \leftrightarrow S^*(-f)$$

$$s^*(-t) \leftrightarrow S^*(f)$$

Complex conjugation in one domain corresponds to flip and conjugation in the other

Important implication: the spectrum of real-valued signals is conjugate symmetric

$$s(t) = s^*(t) \longleftrightarrow S(f) = S^*(-f)$$

Real part of spectrum is symmetric $\text{Re}(S(f)) = \text{Re}(S(-f))$

Imaginary part of spectrum antisymmetric $\text{Im}(S(f)) = -\text{Im}(S(-f))$

Observation to be used later in understanding passband signals: a real-valued time domain signal is completely described by its Fourier transform for positive frequencies (so we can throw away the negative frequencies when mathematically describing it)

Fourier Transform Properties, III

Parseval's Identity

Inner product can be computed in either time or frequency domain

$$\langle s_1, s_2 \rangle = \int_{-\infty}^{\infty} s_1(t) s_2^*(t) dt = \int_{-\infty}^{\infty} S_1(f) S_2^*(f) df = \langle S_1, S_2 \rangle$$

Specialization: energy can be computed in time or frequency domain

$$E_s = ||s||^2 = \int_{-\infty}^{\infty} |s(t)|^2 dt = \int_{-\infty}^{\infty} |S(f)|^2 df.$$

Exercise: Use Parseval's identity to evaluate $\int_0^{\infty} (\sin^4 x / x^4) dx$

Approach: think of integrating over the entire real line, and then think of what is the Fourier transform pair for the square of a sinc.

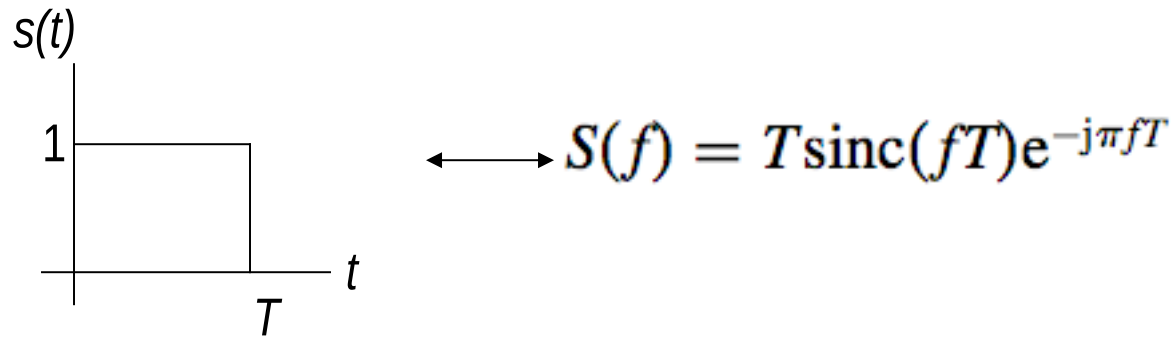
Bandwidth

- Bandwidth of a signal quantifies its frequency occupancy
- **One-sided bandwidth:** We only consider positive frequencies when computing bandwidth for *physical* signals
 - For example, a WiFi signal may occupy a 20 MHz bandwidth, between 2.4-2.42 GHz
 - Physical signals are real-valued (in the time domain)
 - Hence they are conjugate symmetric in the frequency domain, so we can specify them completely by their spectrum over positive frequencies
- We shall also consider complex-valued (in the time domain) signals later
 - Complex envelope of a real-valued passband signal
 - It will turn out that the two-sided bandwidth of the complex envelope equals the physical (one-sided) bandwidth of the passband signal

Aside: reminder on why we need negative frequencies

- We like working with complex exponentials because they are eigenfunctions of LTI systems
 - Need complex exponentials at both positive and negative frequencies to span the space of square integrable signals
 - Real-valued sines and cosines with positive frequencies alone would also work, but these are not eigenfunctions of LTI systems, hence are less convenient
- Physical signals are real-valued (time domain)
 - Hence must satisfy consistency condition of conjugate symmetry (all the information resides in either positive or negative frequencies, hence only need spectrum for one of these)
 - Hence physical bandwidth = one-sided bandwidth

Example: bandwidth of a boxcar



Energy spectral density = magnitude squared of Fourier transform

$$|S(f)|^2 = T^2 \text{sinc}^2(fT)$$

Not strictly bandlimited, but can define ***fractional energy containment bandwidth***

One-sided fractional energy containment bandwidth B (fraction a) satisfies:

$$\int_{-B}^B |S(f)|^2 df = a \int_{-\infty}^{\infty} |S(f)|^2 df$$

Example: $a=0.99$ corresponds to 99% energy containment bandwidth

Boxcar example (contd.)

Useful (and insightful) to normalize: can set T to convenient value (say $T=1$).

Equivalent to defining one unit of time as T .

By scaling relation between time and frequency, if bandwidth for $T=1$ is B_1 , then bandwidth for general T is B_1/T .

Parseval's identity can be used to evaluate energy in whichever domain is more convenient.

$$\int_{-B_1}^{B_1} \text{sinc}^2 f \, df = a \quad \xrightarrow{\text{symmetry}} \quad \int_0^{B_1} \text{sinc}^2 f \, df = \frac{a}{2}$$

Evaluate energy in time domain for boxcar (= 1 for $T=1$)

Numerical results: $B_1=0.85$ for $a=0.9$; $B_1=10.2$ for $a=0.99$.

When we want stricter energy containment, choosing a rectangular time domain pulse is a very bad idea (when we are willing to be sloppy, it's OK.)

Baseband and Passband

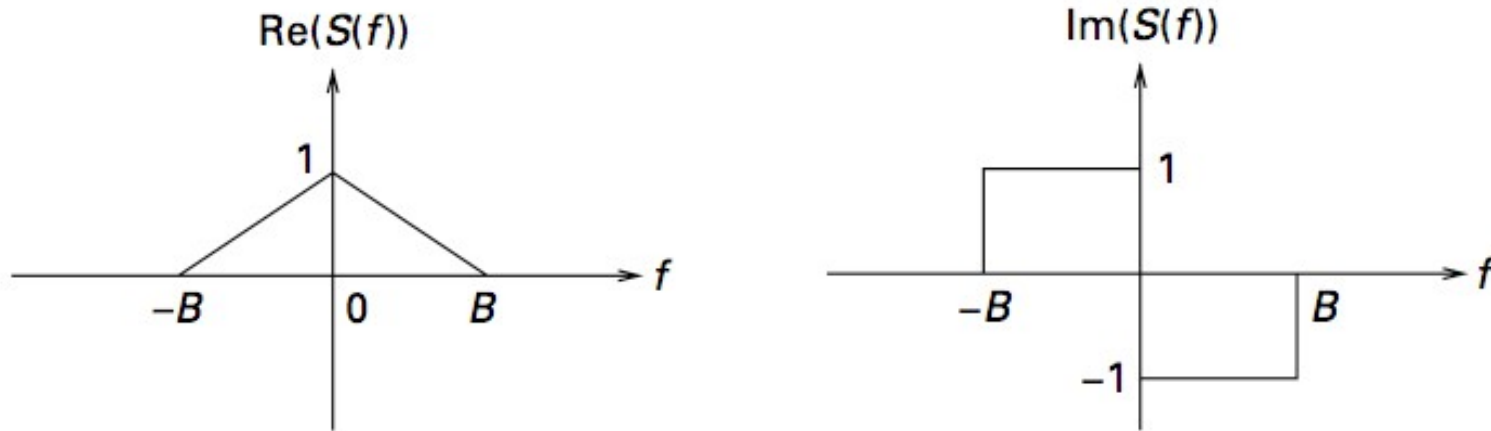
The Complex Baseband Representation

Baseband and Passband Signals/Channels

- Channels often approximated as LTI systems
 - Signal passes through channel, and then noise is added
- Channels allocated/described typically in terms of frequency bands
 - Signals have to be designed for the corresponding frequency band
- Baseband channels/signals
 - Energy concentrated in a frequency band around DC
- Passband channels/signals
 - Energy concentrated in a frequency band away from DC
- Unified treatment of baseband and passband systems
 - Complex baseband representation of passband systems

Baseband Signals

Baseband signals have energy/power concentrated in a band around DC.



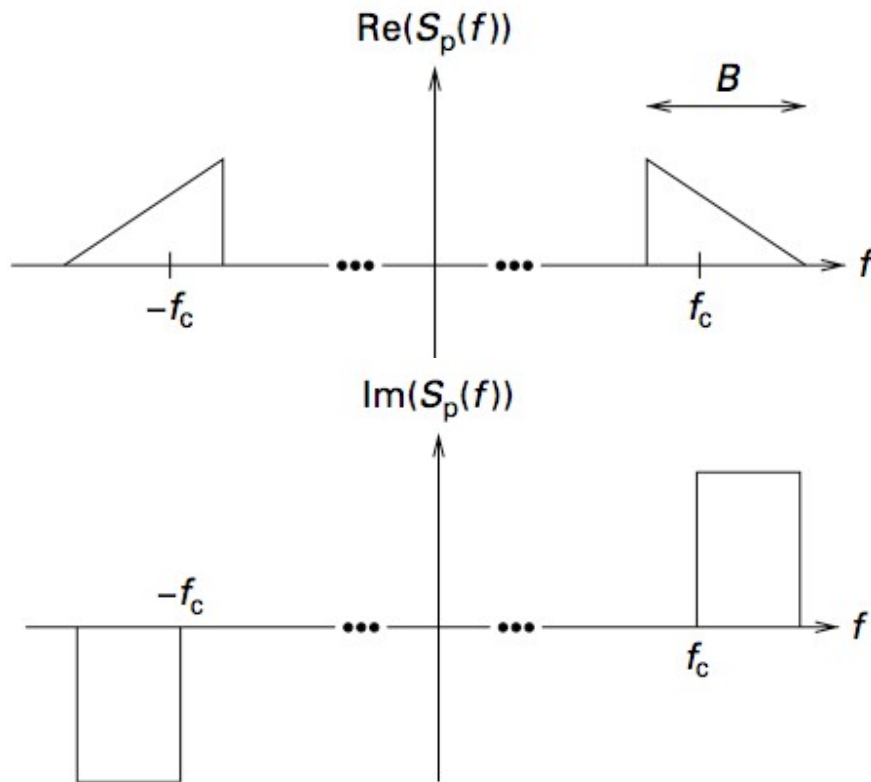
Above: Real baseband signal of bandwidth B ($S(f)$ obeys conjugate symmetry)

For communication over a physical baseband channel, we consider physical (real-valued) baseband signals.

For communication over a physical passband channel (discussion coming up), we consider complex-valued baseband signals which provide a convenient mathematical representation for the corresponding passband signals.

Passband Signals

Passband signals have energy/power concentrated in a band away from DC.



Passband signal of bandwidth B

We only consider physical (real-valued) passband signals, hence their spectra always obey conjugate symmetry

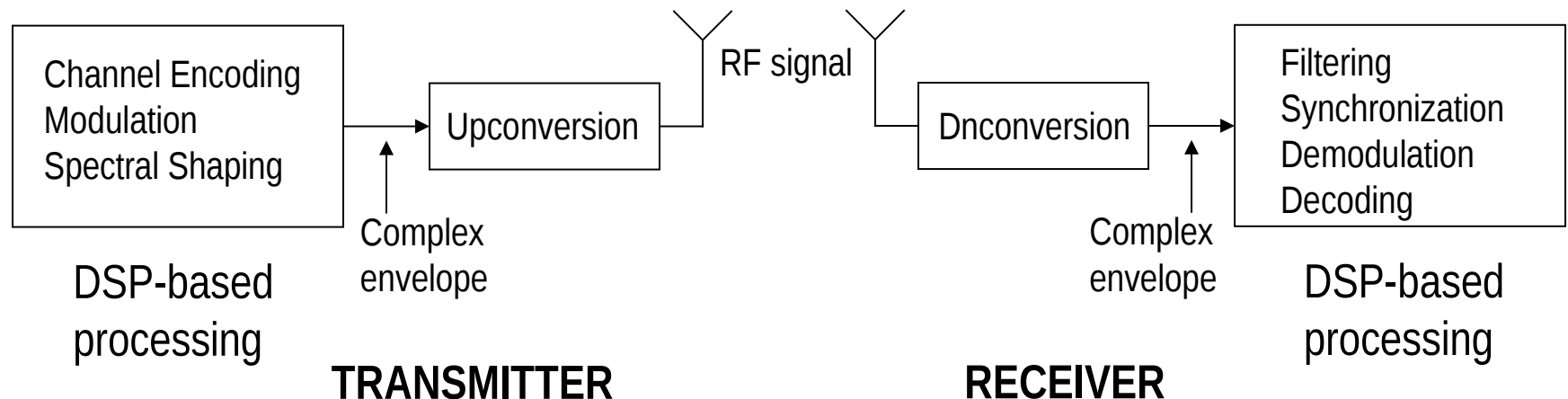
Complex baseband representation: the big picture

- Any real-valued passband signal can be represented by a baseband signal which is in general complex-valued. This is called its **complex envelope**, or **complex baseband representation**.
- The complex envelope carries all the information in the passband signal
- Passband filtering operations can be equivalently performed in complex baseband
- Two-dimensional representation of complex envelope
 - Cartesian coordinates: A pair of real-valued baseband waveforms called the in-phase (I) and quadrature (Q) components
 - Polar coordinates: Envelope and phase waveforms

Modern transceiver architectures are based on complex baseband

- Modern transceivers work with the complex envelope rather than with the passband signal
 - Complex baseband signals can be represented accurately by samples at a reasonable sampling rate
 - Inexpensive to perform complicated digital signal processing (DSP) on the samples: Moore's law scaling
 - This architecture has been responsible for economies of scale in cellular and WiFi

All the action is in complex baseband for a typical wireless transceiver



Two-dimensional representation of passband signals

Any passband signal can be written as (we show this soon):

$$s_p(t) = \sqrt{2}s_c(t) \cos 2\pi f_c t - \sqrt{2}s_s(t) \sin 2\pi f_c t,$$

$s_c(t)$ (“c” for “cosine”)	In-phase (I) component	Real-valued baseband waveforms
$s_s(t)$ (“s” for “sine”)	Quadrature (Q) component	

f_c Reference frequency (chosen somewhere in the passband)

Example (illustrates that baseband/passband signals need not be strictly bandlimited)

$$s_p(t) = \sqrt{2}I_{[0,1]}(t) \cos 300\pi t - \sqrt{2}(1 - |t|)I_{[-1,1]}(t) \sin 300\pi t$$

I component ($f_c = 150$) Q component

$$s_c(t) = I_{[0,1]}(t)$$

$$s_s(t) = (1 - |t|)I_{[-1,1]}(t)$$

Complex envelope

$$s(t) = s_c(t) + js_s(t) \quad \text{Complex envelope}$$

Rectangular coordinates: I and Q components are its real and imaginary parts

Polar coordinates: Envelope and phase

$$e(t) = |s(t)| = \sqrt{s_c^2(t) + s_s^2(t)}, \quad \theta(t) = \tan^{-1} \frac{s_s(t)}{s_c(t)}$$

Three ways to write a passband signal

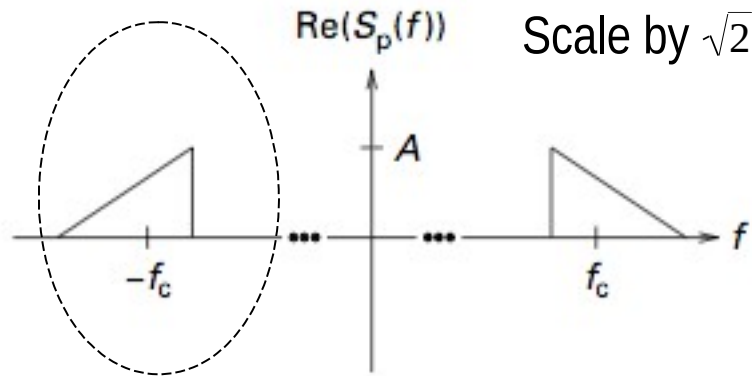
$$s_p(t) = \sqrt{2}s_c(t) \cos 2\pi f_c t - \sqrt{2}s_s(t) \sin 2\pi f_c t, \quad \text{In terms of I and Q}$$

$$s_p(t) = \text{Re}(\sqrt{2}s(t)e^{j2\pi f_c t}) \quad \text{In terms of complex envelope}$$

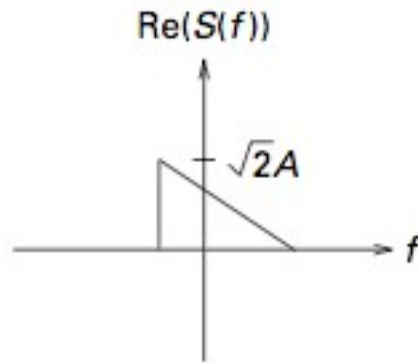
$$s_p(t) = e(t) \cos(2\pi f_c t + \theta(t)) \quad \text{In terms of envelope and phase}$$

Frequency domain construction of complex envelope

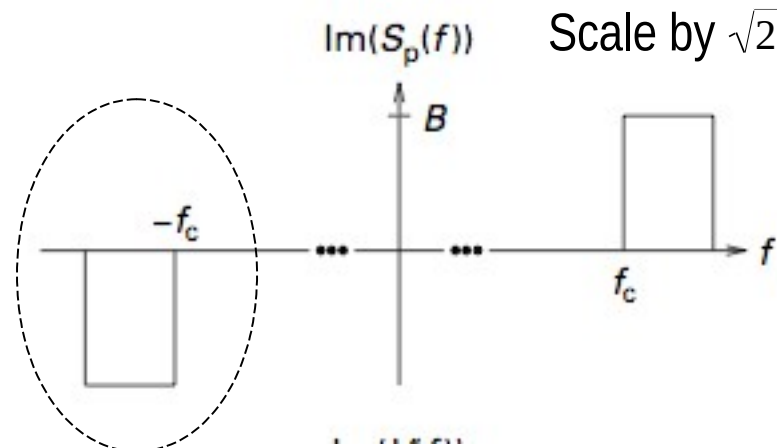
← Move to left by f_c



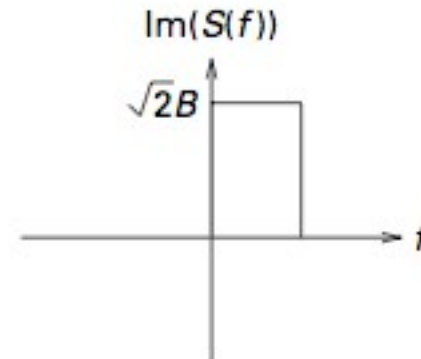
Throw away negative frequencies part



← Move to left by f_c



Throw away negative frequencies part



Let's see why this works

Frequency domain construction (contd.)

By construction, we see that from any real-valued passband signal we can construct a baseband signal. But does this satisfy the time domain relationships we require?

We can construct the positive frequencies part of the passband signal by shifting $S(f)$ to the right by f_c (and scaling down by square root of 2). The negative frequencies part is then determined by conjugate symmetry (conjugate and reflect around origin):

$$S_p(f) = \frac{S(f - f_c) + S^*(-f - f_c)}{\sqrt{2}}, \quad \text{for the given frequency domain construction}$$

But

$$S(f - f_c) \leftrightarrow s(t)e^{j2\pi f_c t} \quad \text{and} \quad S^*(-f - f_c) \leftrightarrow (s(t)e^{j2\pi f_c t})^*$$

This gives us the desired time domain relationship:

$$s_p(t) = [s(t)e^{j2\pi f_c t} + (s(t)e^{j2\pi f_c t})^*] / \sqrt{2} = \sqrt{2} \operatorname{Re}(s(t)e^{j2\pi f_c t})$$

Some technical properties

Orthogonality of I and Q

$\sqrt{2}s_c(t) \cos 2\pi f_c t$ and $\sqrt{2}s_s(t) \sin 2\pi f_c t$ are orthogonal, as long as s_c and s_s are baseband with bandwidth smaller than f_c

Passband inner product in terms of complex baseband inner product:

$$\langle u_p, v_p \rangle = \langle u_c, v_c \rangle + \langle u_s, v_s \rangle = \text{Re}(\langle u, v \rangle)$$

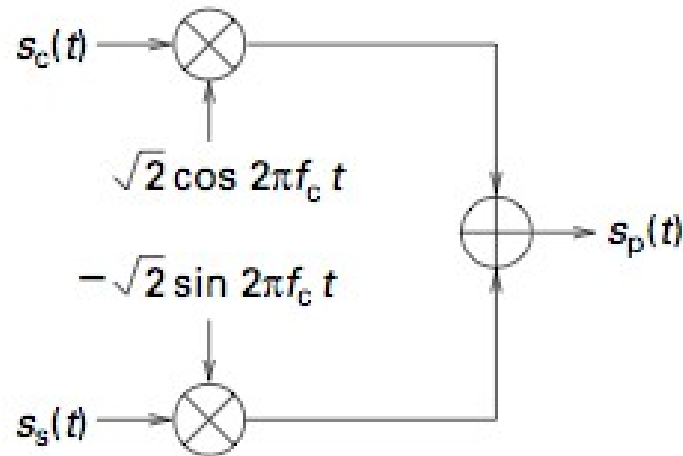
For our scaling, energies of passband signal and complex envelope are equal

Follows from the inner product formula by setting the two signals to be the same

Class exercise: Find the inner product of the two passband waveforms

$$I_{[0,1]}(t) \cos(300\pi t) \text{ and } I_{[1,3]}(t) \sin(300\pi t + \pi/4)$$

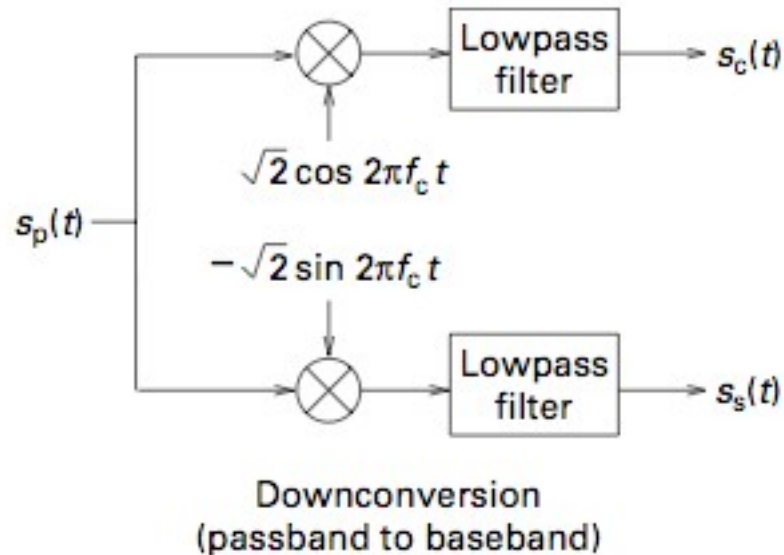
Upconversion from baseband to passband (at transmitter)



Upconversion
(baseband to passband)

Follows directly from representation of passband signal in terms of I and Q components

Downconversion from passband to baseband (at receiver)



Working through what happens on the top branch:

$$\begin{aligned}\sqrt{2}s_p(t) \cos(2\pi f_c t) &= 2s_c(t) \cos^2 2\pi f_c t - 2s_s(t) \sin 2\pi f_c t \cos 2\pi f_c t \\ &= s_c(t) + \underbrace{s_c(t) \cos 4\pi f_c t - s_s(t) \sin 4\pi f_c t}_{\text{Passband signal at } 2f_c \text{ rejected by LPF}}\end{aligned}$$

Passband signal at $2f_c$ rejected by LPF

Encoding information in passband signals

Observation 1: *Information resides in complex envelope*

Variations due to “carrier” term $e^{j2\pi f_c t}$ are very rapid but predictable

Observation 2: *Carrier terms due to I and Q components are orthogonal*

$$\sqrt{2}s_c(t) \cos 2\pi f_c t \quad \text{and} \quad \sqrt{2}s_s(t) \sin 2\pi f_c t$$

are orthogonal, regardless of choice of I and Q components, as long as they are baseband signals with bandwidth less than f_c

Conclusion: *Passband modulation corresponds to encoding information into I and Q components.* I and Q thus provide orthogonal “channels” on which we can potentially send separate messages (hence the name **two-dimensional** modulation)

Crucial caveat: I-Q orthogonality assumes ideal carrier sync. I and Q get “mixed up” under phase and frequency shifts. Receiver must perform carrier synchronization in order to estimate original I and Q.

Linear Modulation

Send a complex number $b = b_c + jb_s = re^{j\theta}$,

by sending the complex baseband signal $s(t)=bp(t)$, where $p(t)$ is a baseband pulse (assume $p(t)$ real-valued for simplicity)

Corresponding passband signal:

$$\begin{aligned} s_p(t) &= \text{Re} \left(\sqrt{2}s(t)e^{j2\pi f_c t} \right) = \sqrt{2} (b_c p(t) \cos 2\pi f_c t - b_s p(t) \sin 2\pi f_c t) \\ &= \sqrt{2}r \cos(2\pi f_c t + \theta). \end{aligned}$$

Amplitude modulation of I and Q by b_c, b_s

Or envelope/phase modulation by $r \geq 0, \theta$

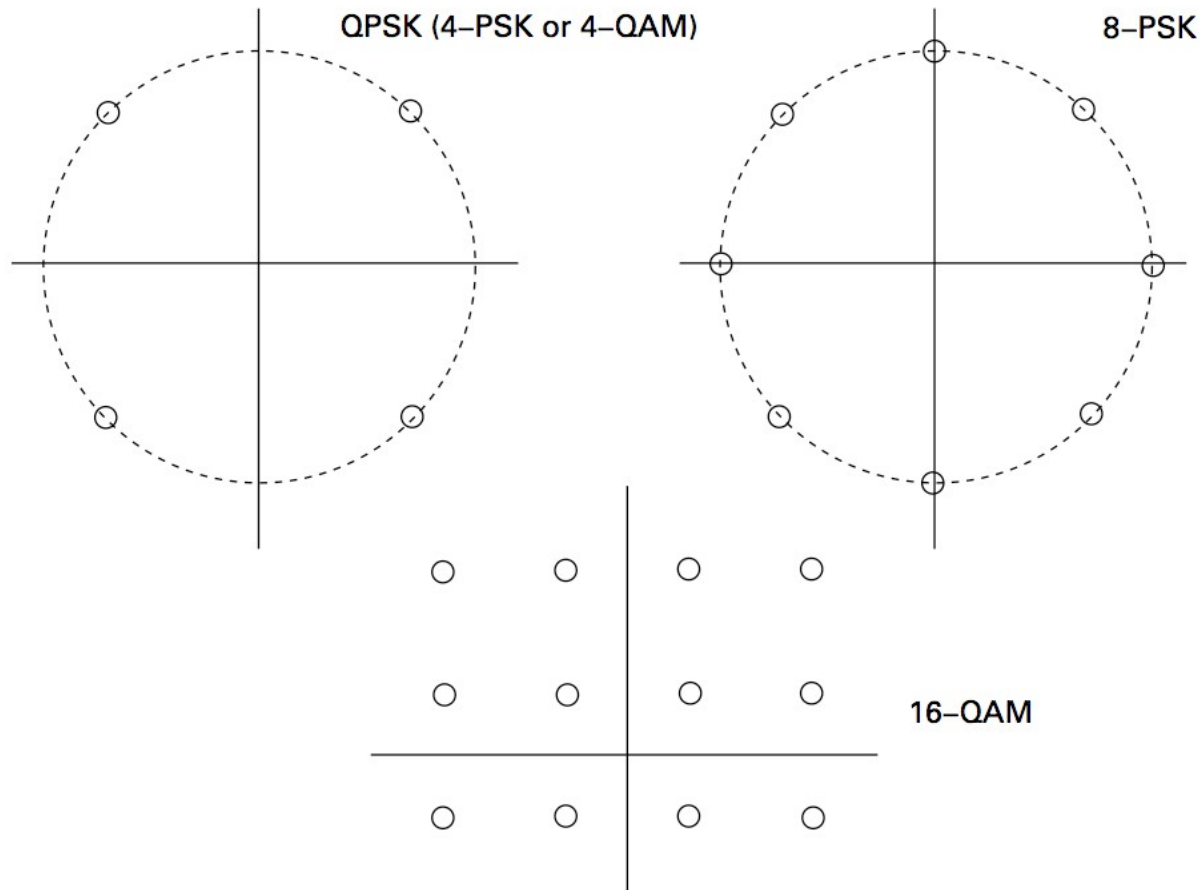
In practice, we send a stream of symbols by sending the complex baseband signal

$$\sum_n b[n]p(t - nT)$$

Some constellations for linear modulation

Symbol b is typically chosen from a finite *constellation*.

Number of bits/symbol = $\log_2 M$, where M is the number of constellation points



Much more on this later...

Most transceiver operations can be performed in complex baseband

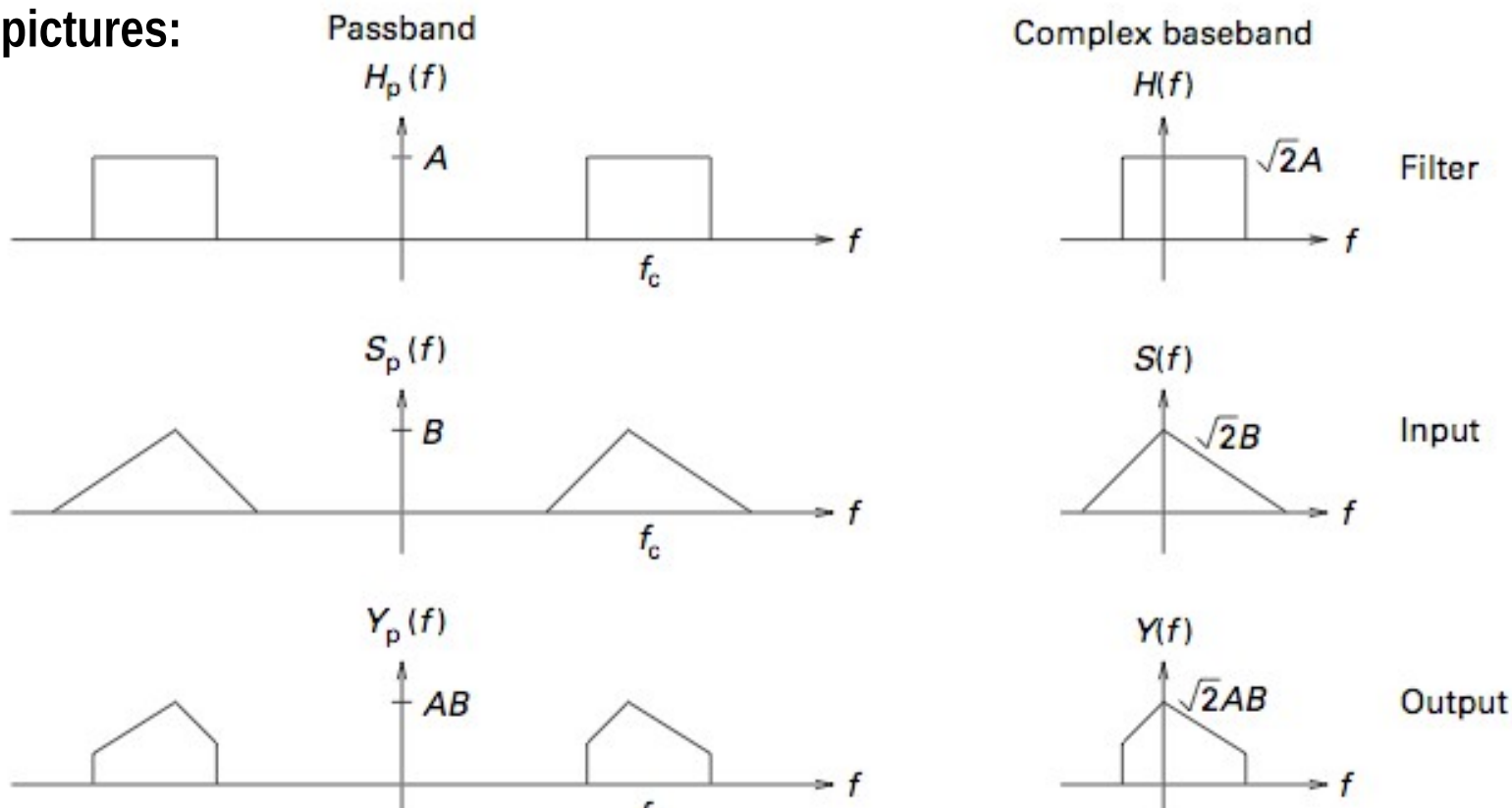
- Filtering
- Carrier frequency/phase correction
- Coherent and noncoherent reception

Passband filtering is equivalent to complex baseband filtering



(except for a square root 2 scale factor)

Proof by pictures:



$$Y(f) = \sqrt{2}Y_+(f + f_c) = \sqrt{2}S_+(f + f_c)H_+(f + f_c) = \frac{1}{\sqrt{2}}S(f)H(f)$$

Filtering in complex baseband

Complex-valued convolution to implement equivalent of passband filtering operation

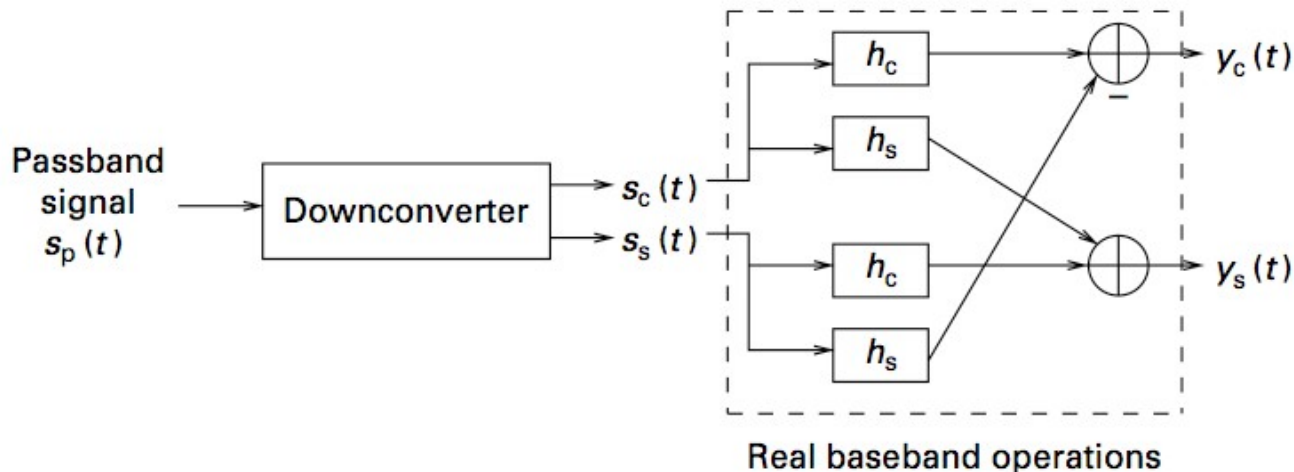
$$y(t) = \frac{1}{\sqrt{2}}(s * h)(t).$$

Requires four real-valued convolutions:

$$y_c = \frac{1}{\sqrt{2}}(s_c * h_c - s_s * h_s), \quad y_s = \frac{1}{\sqrt{2}}(s_s * h_c + s_c * h_s)$$

Downconverter can use sloppy analog passband filter

Sophisticated filtering can be implemented in baseband (square root 2 factors not shown)



Possible in-class exercises

Class exercise: Find the inner product of the two passband waveforms

$$I_{[0,1]}(t)\cos(300\pi t) \text{ and } I_{[1,3]}(t)\sin(300\pi t + \pi/4)$$

Class exercise: Find the convolution of the two passband waveforms

$$I_{[0,1]}(t)\cos(300\pi t) \text{ and } I_{[1,3]}(t)\sin(300\pi t + \pi/4)$$

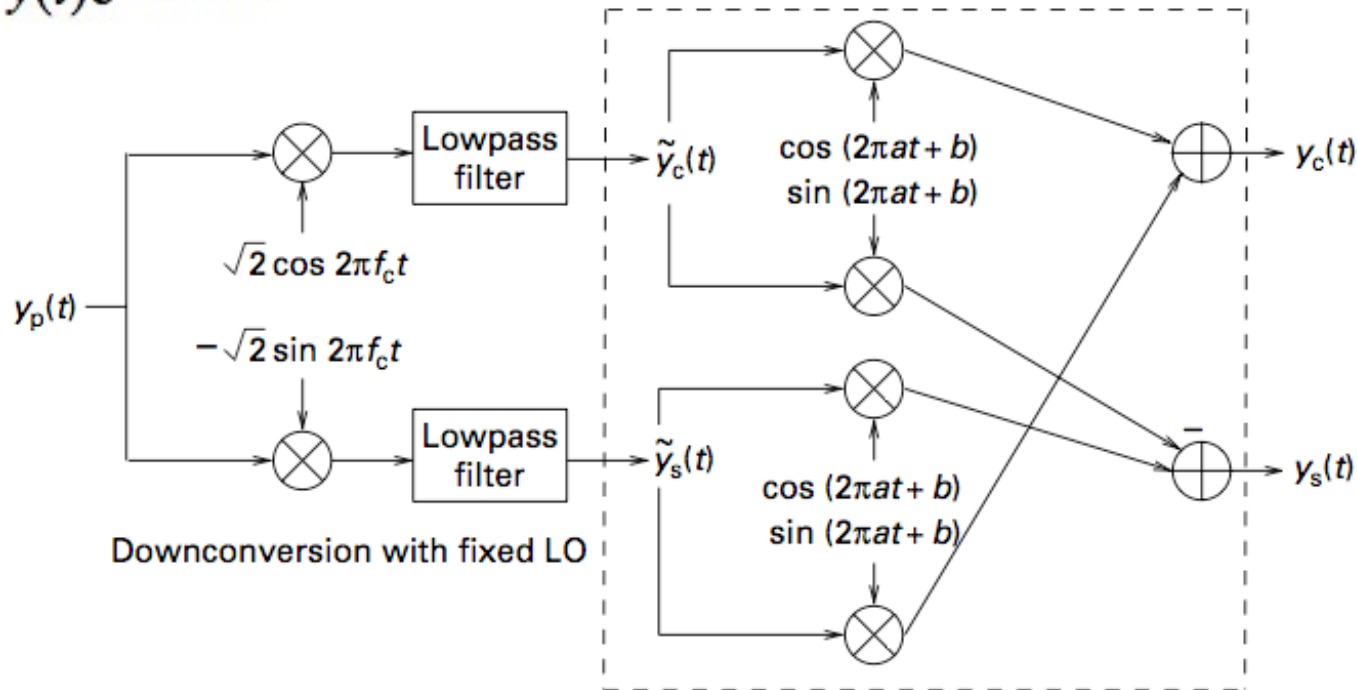
Freq/phase offsets: easy to model and undo in complex baseband

$$y_p(t) = \sqrt{2}[y_c(t) \cos(2\pi(f_c + a)t + b) - y_s(t) \sin(2\pi(f_c + a)t + b)],$$

(received passband signal, downconverted by fixed LO at receiver)

$$\tilde{y}(t) = y(t)e^{j(2\pi at + b)} \quad \text{Complex envelope with respect to downconverter}$$

$$y(t) = \tilde{y}(t)e^{-j(2\pi at + b)} \quad \text{Undo frequency and phase offset}$$



Real baseband operations for undoing frequency and phase offset

(unwrap complex-valued operations to get these)

Correlation in complex baseband

- Correlation is a fundamental operation in comm receivers
 - Match the received waveform to a template of a possibly transmitted waveform
- Coherent receiver: performs correlation assuming ideal carrier synchronization
- Noncoherent receiver: performs correlation assuming carrier phase is unknown, but is constant over the duration of the received waveform
- Complex baseband makes it easy to see the resulting correlator structures

Coherent and noncoherent receivers in complex baseband

Coherent receiver just takes passband inner product

$$\langle y_p, s_p \rangle = \text{Re}(\langle y, s \rangle) = \langle y_c, s_c \rangle + \langle y_s, s_s \rangle$$

In complex baseband,
correlate I with I, Q with Q

Coherent receiver does not work without phase sync:

$$\begin{aligned} \langle y_p, s_p \rangle &= \text{Re}(\langle Ae^{j\theta} s, s \rangle) \quad (\text{plus noise}) \\ &= A \cos \theta ||s||^2 \quad (\text{plus noise}). \end{aligned}$$

Can get completely wiped
when $\theta = \pi/2$, for example

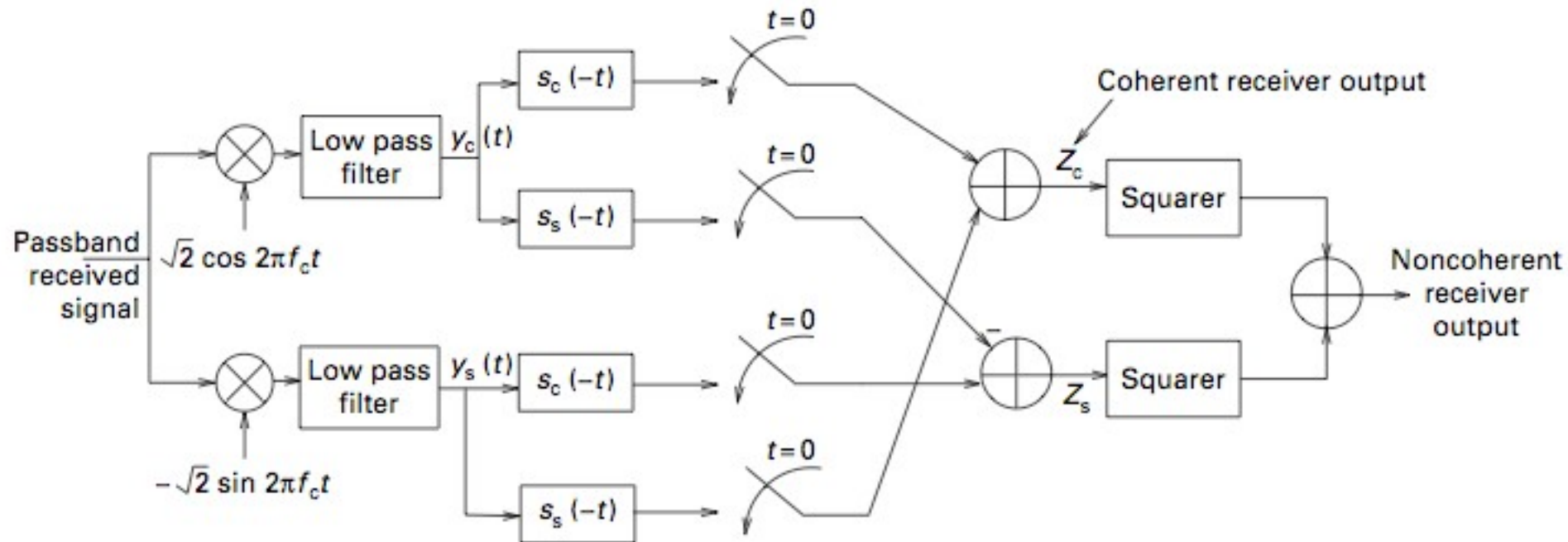
Can get rid of phase dependence by taking magnitude instead of real part:

$$|\langle y, s \rangle| = |\langle Ae^{j\theta} s, s \rangle \quad (\text{plus noise})| \approx A ||s||^2 \quad (\text{ignoring noise})$$

Noncoherent receiver involves four real inner products, including I-Q cross terms

$$|\langle y, s \rangle|^2 = (\langle y_c, s_c \rangle + \langle y_s, s_s \rangle)^2 + (\langle y_s, s_c \rangle - \langle y_c, s_s \rangle)^2$$

Coherent and noncoherent receiver operations



Bandwidth revisited

- Occupancy of positive frequencies in a physical system
 - Real baseband: one-sided
 - Passband: one-sided
 - Complex baseband: two-sided bandwidth corresponds to one-sided bandwidth of corresponding passband system
- Baseband and passband definitions do not require exact containment within a finite frequency band
- Fractional energy/power containment bandwidth

Modulation Degrees of Freedom
Why Linear Modulation is a Good Idea

Modulation degrees of freedom, I

- Ideal passband bandlimited channel of bandwidth W
- Corresponding complex baseband channel
 - Nyquist sampling theorem says that any signal falling in this band can be represented by W **complex-valued** samples per second
 - WT_o complex dimensions, or $2WT_o$ real dimensions over an observation interval of length T_o
 - Linear modulation with sinc pulse uses all available degrees of freedom (interpolation formula)
 - Bandwidth efficiency for a modulation scheme
 - Signal space description of modulation formats

Modulation degrees of freedom, II

- Physical signals and channels are analog, and live in an infinite-dimensional space
- But constraints on time and bandwidth limit us to a finite-dimensional subspace.
- The dimension of this subspace (may not be very precisely characterized) equals the **modulation degrees of freedom**

**Consider bandlimited passband channel of bandwidth W
Maps to complex baseband channel over $[-W/2, W/2]$**

Shannon's sampling theorem applied to the complex baseband channel

$$s(t) = \sum_{n=-\infty}^{\infty} s\left(\frac{n}{W}\right) p\left(t - \frac{n}{W}\right) \quad p(t) = \text{sinc}(Wt)$$

Standard interpretation: Bandlimited signal can be reconstructed from samples using sinc interpolation

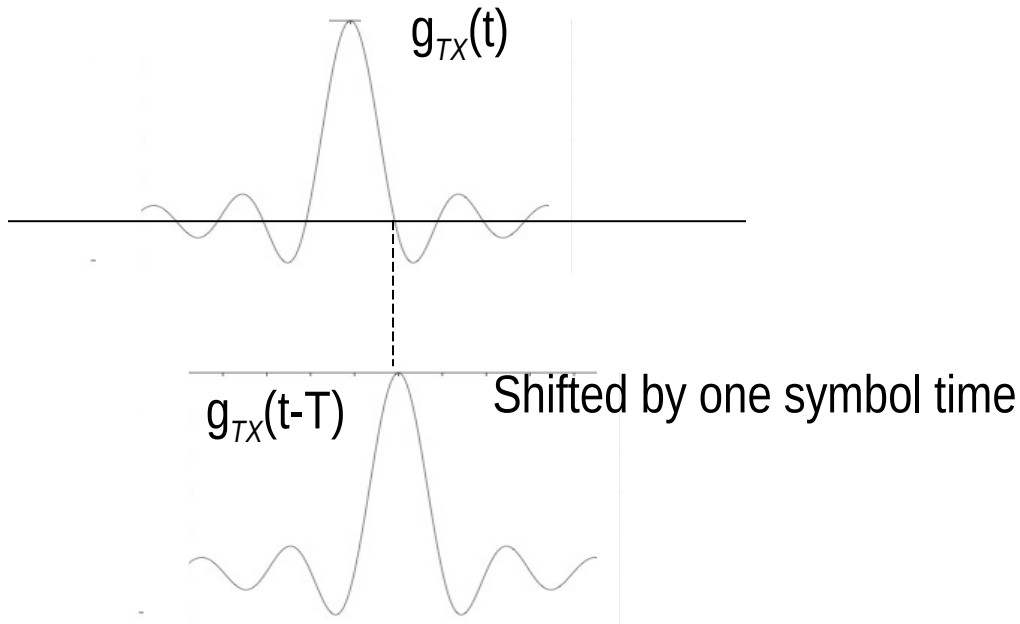
Our interpretation: Samples are symbols sent by linearly modulating sinc pulse.
Linear modulation can exploit all the available degrees of freedom in a bandlimited channel.

Modulation: what we know so far

- Ideal passband bandlimited channel of bandwidth W
- Corresponding complex baseband channel
 - Nyquist sampling theorem says that any signal falling in this band can be represented by W **complex-valued** samples per second
 - WT_o complex dimensions, or $2WT_o$ real dimensions over an observation interval of length T_o
 - Linear modulation with sinc pulse uses all available degrees of freedom (interpolation formula)
- Modulation design can be restricted to a finite-dimensional *signal space*
- Can define bandwidth efficiency of a linear modulation scheme with symbols from M -ary constellation

$$\eta_B = \frac{\log_2 M}{D} \quad (D = \text{number of degrees of freedom} = WT_o)$$

Why the sinc pulse does not work



At the peaks, only one sinc pulse contributes to overall output ==> no ISI

But for off-peak sample, we get contributions from all symbols, with contributions from "far-away" symbols decaying as $1/(\text{distance from sampling time})$

In the worst-case, the signs of these symbols (think of +1 and -1 symbols for now) conspire to add up constructively.

The sum looks like the sum of $\{1/n\}$ (the details are a bit more messy, but let's not worry about it), and grows as $\log(N)$, where N is the number of symbols.

Sum blowing up implies unbounded peak power.

Same reasoning implies that ISI can blow up if we sample slightly off-peak

So what kind of pulse should we use?

Can't we just truncate the sinc? Well, $\log N$ can get pretty bad for large N . And small N (aggressive truncation) means frequency spreading outside the given band. A controlled increase in bandwidth is far more desirable.

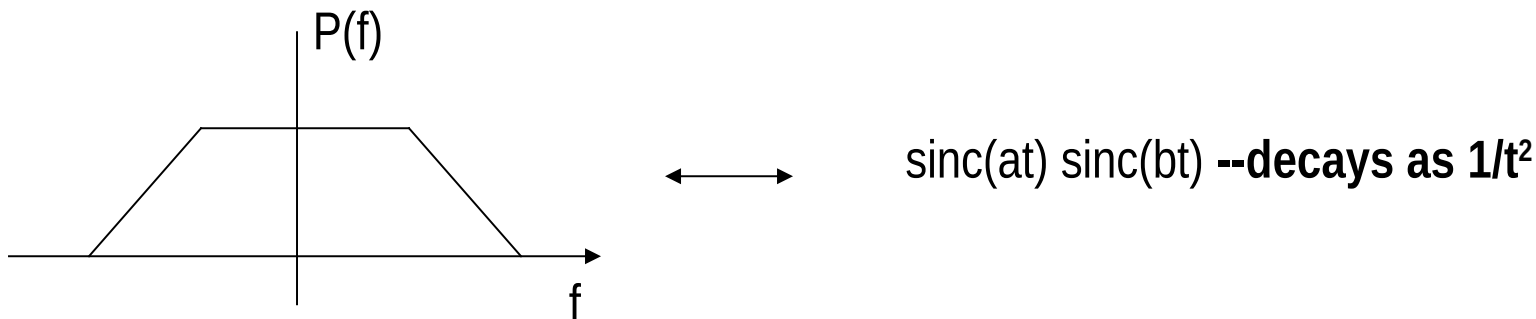
Choose the pulse to control the size of sum for peak power/ISI.

We can tightly approximate such sums by integrals for smooth functions:

$$\sum_{n=1}^N f(n) \quad \text{by} \quad \int_1^N f(t) dt$$

The integral blows up for $1/t$ decay, but does not blow up for $1/t^a$ decay, $a > 1$.

For example, $a = 2$ would lead to a convergent integral/sum regardless of how big N is. The trapezoidal frequency response below therefore works.



How should we systematically choose the pulse to conserve bandwidth but control peak power and ISI?

Modulation: what we need to figure out

- We have seen that linear modulation with sinc pulse has its problems
 - Sharp cutoff in frequency domain leads to slow $(1/t)$ decay in time domain
 - Unbounded peak power, unbounded intersymbol interference when there is sampling offset
- Need to use pulses with gentler frequency domain decay, hence faster time domain decay
 - How should we choose the modulating pulse?
 - How should we choose the symbols to be sent?
- Are there good modulation strategies that are not linear?
- Let us start with linear modulation with a general transmit pulse

Linear Modulation

Linear Modulation

- We restrict attention to baseband, without loss of generality
 - Real baseband for physical baseband channels
 - Complex baseband for physical passband channels
- Consider first examples of linear modulation without worrying about optimizing pulse choice
 - Baseband line codes
 - Two-dimensional constellations for passband modulation
- Bandwidth of linearly modulated signal
 - Depends on bandwidth of modulating pulse (not surprising)
 - Motivates using pulses of small bandwidth (while keeping peak power and ISI under control)
- Nyquist criterion for ISI avoidance
 - One possible way of choosing the pulse, so as to avoid ISI under ideal conditions
 - Nyquist and square root Nyquist pulses

Linear modulation: unified treatment for baseband and passband channels

Baseband transmitted signal $u(t) = \sum_n b[n]g_{TX}(t - nT)$

$\{b[n]\}$ Transmitted symbols

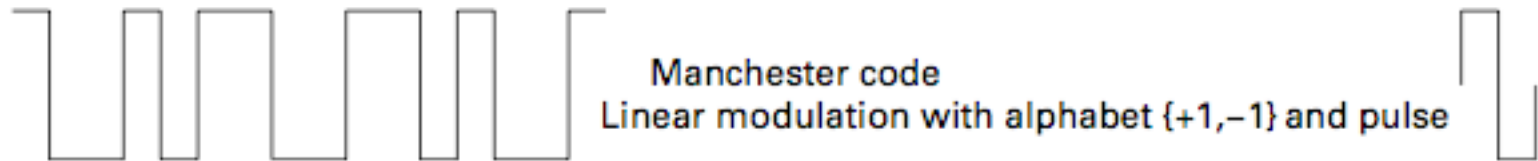
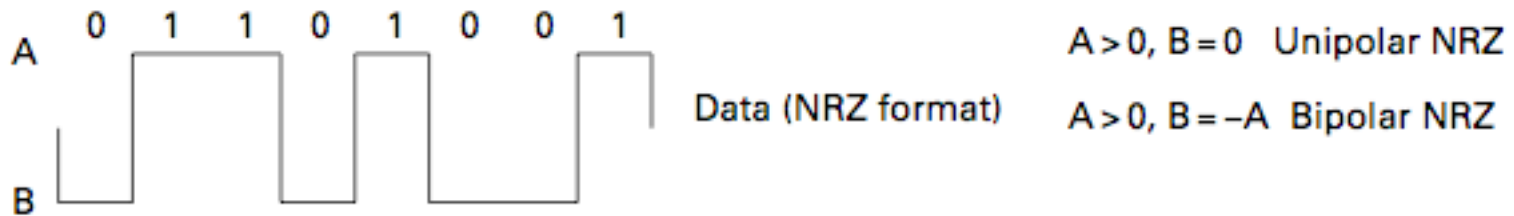
$g_{TX}(t)$ Modulating pulse (baseband)

$1/T$ Symbol rate

Physical baseband channel: $u(t)$ is the physical signal sent over the channel

Passband channel: Physical signal sent over the channel is $\text{Re}(u(t)e^{j2\pi f_c t})$

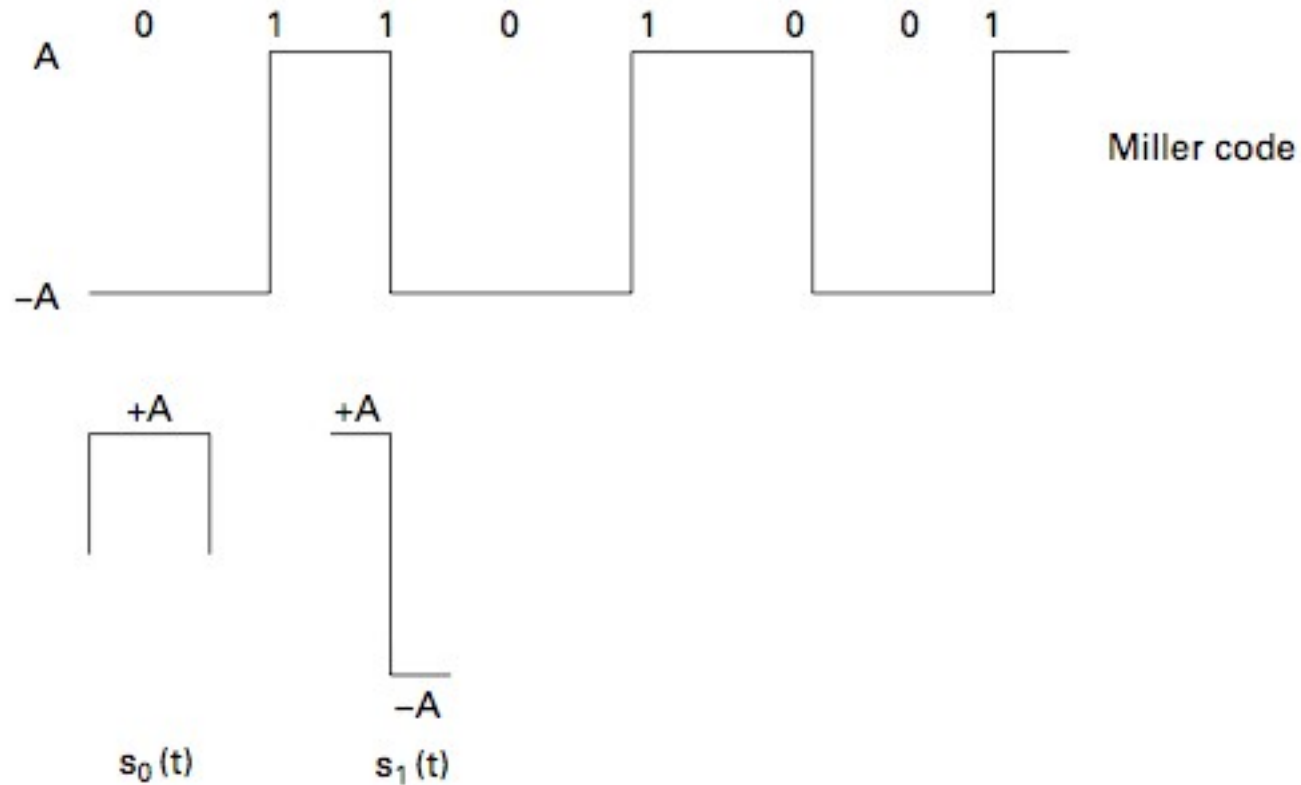
Baseband Line Codes



Linear modulation with rectangular pulses; often used for wired communication over real baseband channels

Miller code

A simple example of nonlinear modulation



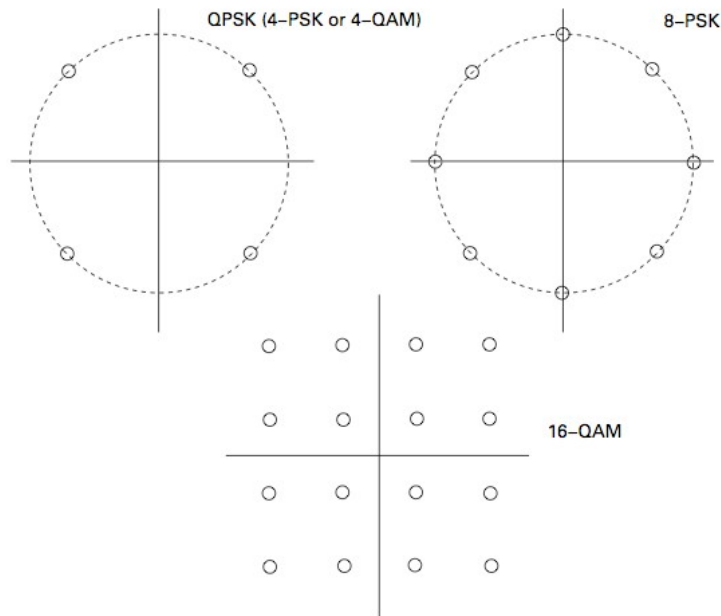
Miller code tries to minimize transitions by using memory.
Two pulses rather than one, unlike linear modulation.

Two-dimensional modulation for passband channels

Baseband signal $u(t) = \sum_n b[n]g_{TX}(t - nT)$ is the complex envelope of the physical passband signal that is sent

We work exclusively in complex baseband, so we call $u(t)$ the transmitted signal

Transmitted symbols $\{b[n]\}$ can now take complex values, typically from a fixed constellation



Some example constellations

PSK: phase shift keying

QAM: quadrature amplitude modulation

Bandwidth of linearly modulated signals

- Model as random process
 - We give an outline of the chain of reasoning
 - **Read the book for details**
- Compute the power spectral density (PSD)
- Compute bandwidth from PSD using your favorite definition of bandwidth
- Fractional power containment bandwidth is often the most useful
 - Quantifies spillage outside an allocated frequency band, for example.
 - Rigorous spec can be used to control co-channel interference

Two-dimensional modulation, contd.

Assume that the transmit pulse $g_{TX}(t)$ is real-valued (for simplicity of discussion).

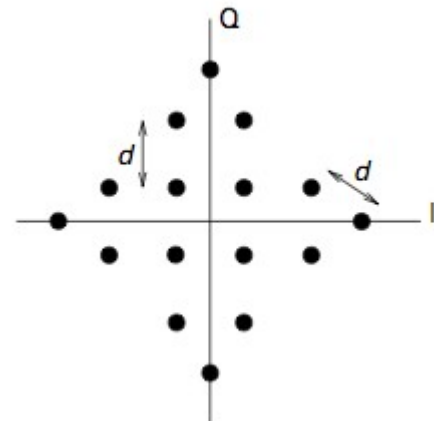
I and Q components of complex baseband transmitted signal are given by

$$u_c(t) = \sum_n \text{Re}(b[n])g_{TX}(t - nT), \quad u_s(t) = \sum_n \text{Im}(b[n])g_{TX}(t - nT).$$

In **rectangular QAM**, we choose $\text{Re}(b[n])$ and $\text{Im}(b[n])$ independently from the same real-valued constellation. For example, from $\{-1, +1\}$ for 4-QAM, and $\{-3, -1, +1, +3\}$ for 16-QAM

Amplitude and phase of complex baseband transmitted signal over n th symbol governed by $|b[n]|$ and $\arg(b[n])$. **PSK** corresponds to keeping $|b[n]|$ constant and choosing $\arg(b[n])$ from a finite set of possibilities.

There are many possible two-dimensional constellations that cannot be classified as either rectangular QAM or PSK.



Bandwidth of linear modulated signals, I

Brief detour on modeling using random processes

- Transmitted symbols modeled as random, so transmitted signal is a random process
- Power spectral density (PSD) and autocorrelation functions can be defined as empirical averages over a single sample path
 - Corresponds to a particular stream of transmitted symbols
- Bandwidth can be defined as fractional power containment bandwidth (exactly as we did for energy containment bandwidth), or as 3 dB bandwidth, etc.
 - All we need to know is the PSD
- But empirical PSD for one sample path not much good if it does not apply to other sample paths
 - What if bandwidth is much higher for some other transmitted sequence?
- We therefore typically design transmitted symbol sequences so as to get **ergodicity** (in second order statistics)
 - Impose enough variation within each transmitted sequence (e.g., by pseudorandom scrambling) that time averages along a sample path equal statistical averages across sample path for mean (usually zero DC value) and PSD/autocorrelation
- Now we can design based on statistical models

Bandwidth of linear modulated signals, II

Stationarity and cyclostationarity

- Stationary means “statistics do not change under time shifts”
- Wide sense stationary (WSS) means “second order stats do not change under time shifts”
- Cyclostationary (with respect to period T) means “statistics do not change under time shifts that are integer multiples of T ”
- Wide sense cyclostationary (with respect to period T) means “second order stats do not change under time shifts that are integer multiples of T ”
- Can compute autocorrelation and PSD as a Fourier transform pair for WSS processes
- If symbol sequence (wide sense) stationary, then linearly modulated signal is (wide sense) cyclostationary with respect to the symbol time T
 - Since shift by T in the transmitted signal is equivalent to shifting the symbol sequence
- By fuzzing up the time axis, we can “stationarize” a cyclostationary process
 - Introduce a random delay that is uniform over $[0, T]$, and is independent of everything else (of the symbol sequence, in our case)
 - The autocorrelation function and PSD for this stationarized process is exactly the same as what we would get on empirically averaging over a sample path, assuming ergodicity
- Now we can use statistical averages to compute the PSD, and hence the bandwidth

Bandwidth of linearly modulated signals, III

The core result

For a complex baseband linearly modulated signal $u(t) = \sum_{n=-\infty}^{\infty} b[n]g_{TX}(t - nT)$

with symbols $\{b[n]\}$ zero mean, uncorrelated,

the Power Spectral Density is given by

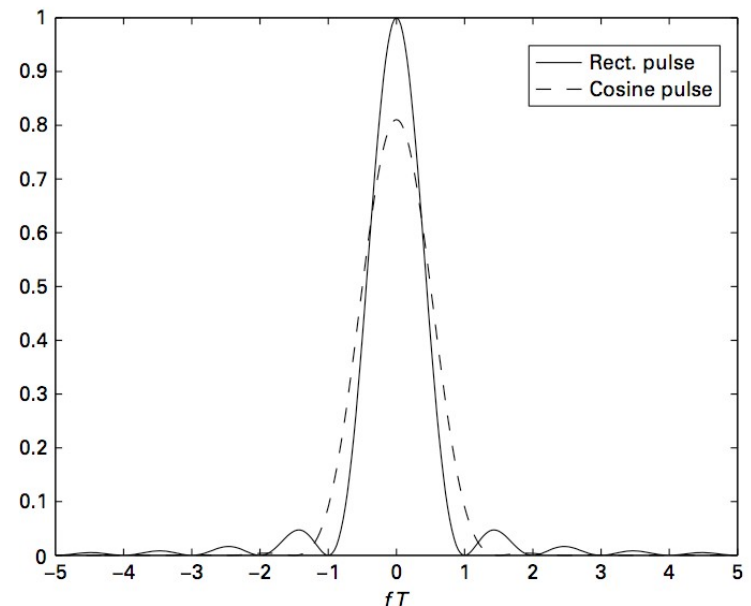
$$S_u(f) = \frac{\mathbb{E}[|b[n]|^2]}{T} |G_{TX}(f)|^2$$

- PSD of $u(t)$ is a scalar multiple of the energy spectral density of the transmit pulse.
- Fractional power containment bandwidth of $u(t)$ = Fractional energy containment bandwidth of the transmit pulse g_{TX}
- We are therefore highly motivated to reduce the bandwidth of the transmit pulse
- Aside: PSD for correlated symbols derived in Problem 2.22. Correlations can be designed to shape spectrum.

Bandwidth of linearly modulated signals, IV

Examples

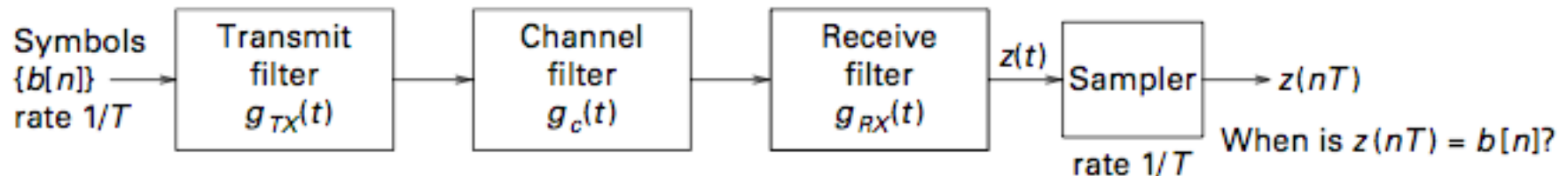
- Preferable to talk about normalized bandwidth
 - Replace f by fT (or set $T=1$, without loss of generality)
- Rectangular timelimited pulse
 - Sinc-squared spectrum has poor power containment
- Cosine timelimited pulse (used in MSK)
 - Smoother roll-off in time means better frequency containment
- But bandlimited pulses would be even better (next up--how to choose them using Nyquist criterion)



The Nyquist Criterion for ISI Avoidance
A Framework for Bandwidth-Efficient Design

Nyquist criterion for ISI avoidance

When is a noiseless linearly modulated system ISI-free?



Noiseless signal at output of receive filter
$$z(t) = \sum_n b[n]x(t - nT)$$

Effective pulse (cascade of transmit, channel and receive filters):

$$x(t) = (g_{TX} * g_C * g_{RX})(t)$$

Time domain criterion for ISI avoidance is obvious:

$$z(nT) = b[n] \quad \text{for all } n \iff x(mT) = \delta_{m0} = \begin{cases} 1, & m = 0, \\ 0, & m \neq 0. \end{cases}$$

(Effective pulse should have exactly one nonzero sample at symbol rate)

We are more interested in the implications for bandwidth occupancy...

Nyquist criterion in the frequency domain

$$z(nT) = b[n] \quad \text{for all } n \quad \longleftrightarrow \quad x(mT) = \delta_{m0} = \begin{cases} 1, & m = 0, \\ 0, & m \neq 0. \end{cases}$$

$$\longleftrightarrow \quad \boxed{1/T \sum_{k=-\infty}^{\infty} X\left(f + \frac{k}{T}\right) = 1 \quad \text{for all } f.}$$

Aliased versions of the frequency domain pulse must add up to a constant.

Any pulse satisfying this condition is said to be **Nyquist at rate $1/T$**

$$\text{Proof: Show that } B(f) = 1/T \sum_{k=-\infty}^{\infty} X\left(f + \frac{k}{T}\right) \longleftrightarrow \begin{matrix} \{x(-mT)\} \\ \text{(Fourier series)} \end{matrix}$$

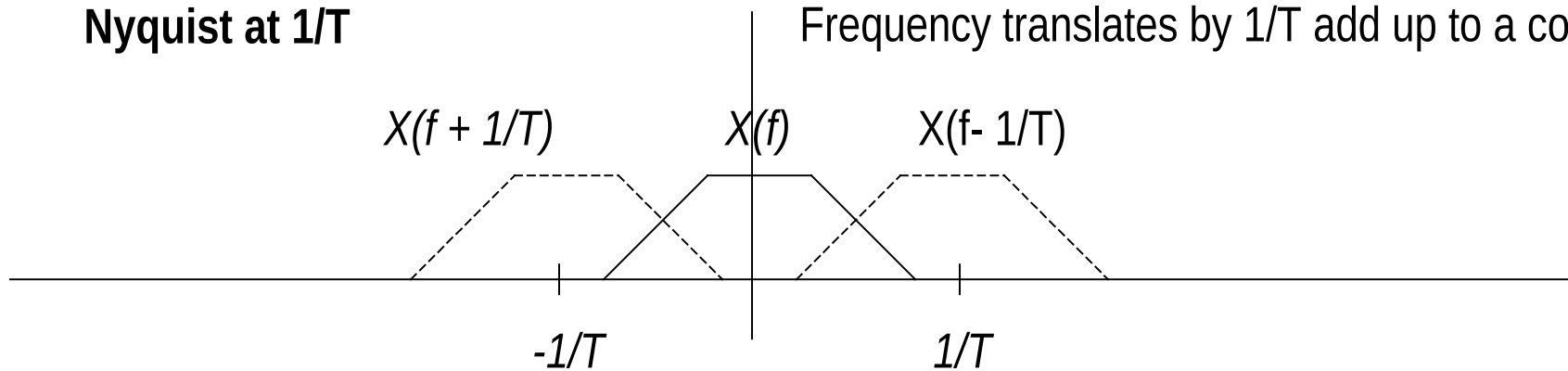
(periodic in frequency domain)

Hence $\{x(mT)\}$ is a discrete impulse if and only if $B(f)$ is constant

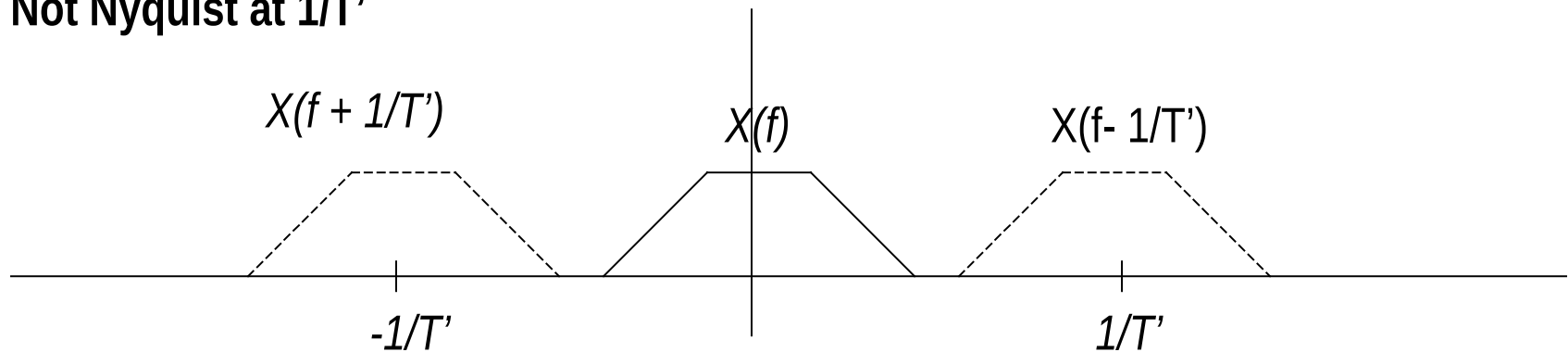
Nyquist criterion illustrated

Nyquist at $1/T$

Frequency translates by $1/T$ add up to a constant



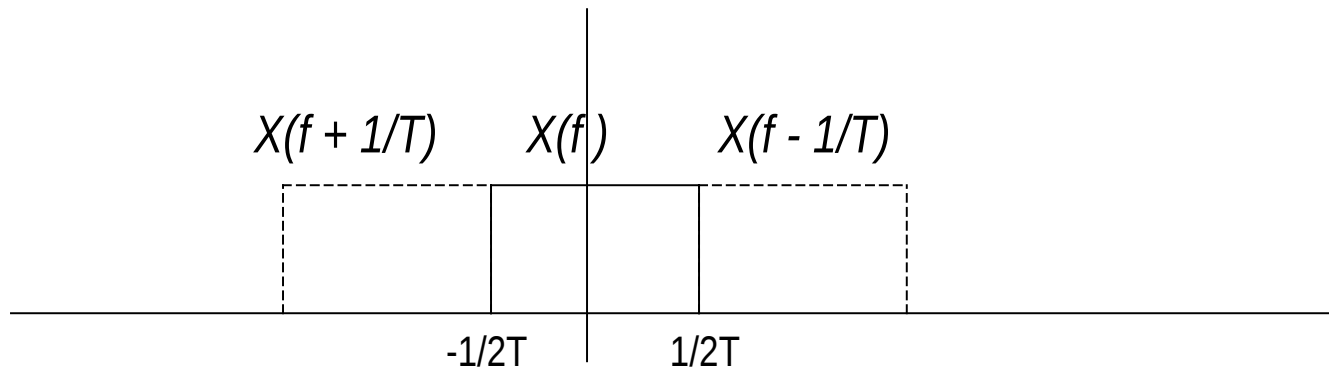
Not Nyquist at $1/T'$



Frequency translates by $1/T'$ don't add up to a constant

Minimum bandwidth Nyquist pulse

Minimum bandwidth Nyquist pulse is the sinc $x(t) = \text{sinc}(t/T)$

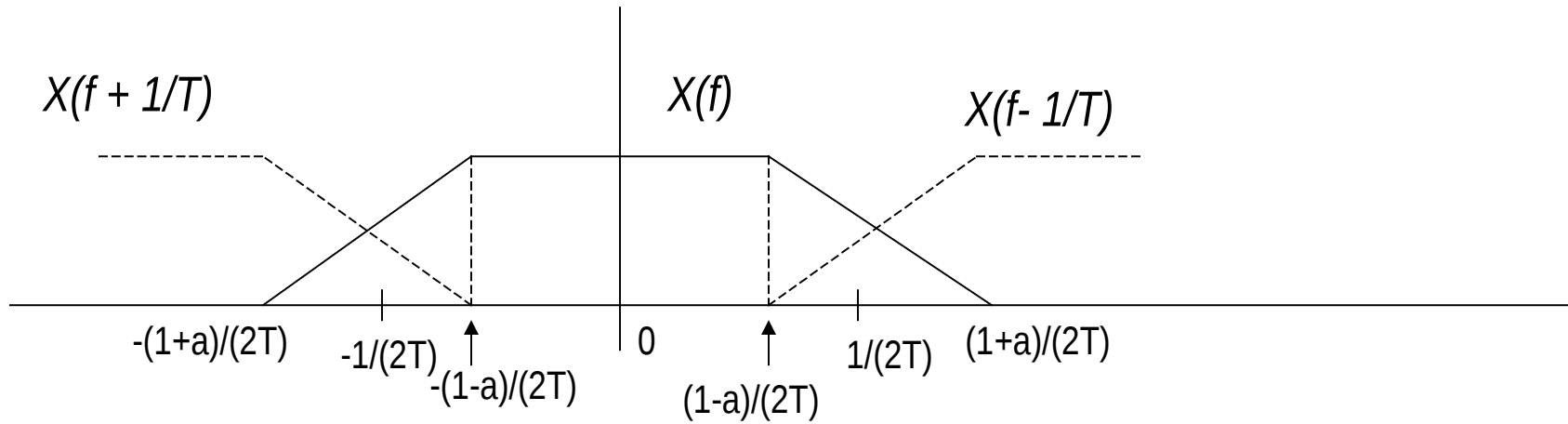


Translates at $1/T$ just touch each other, hence must be flat over band to add up to constant

But we know that sinc decays too slowly.

For faster decay in time, must go beyond the minimum bandwidth.

Design of bandlimited Nyquist pulses and excess bandwidth



Slower roll-off in frequency gives faster roll-off in time

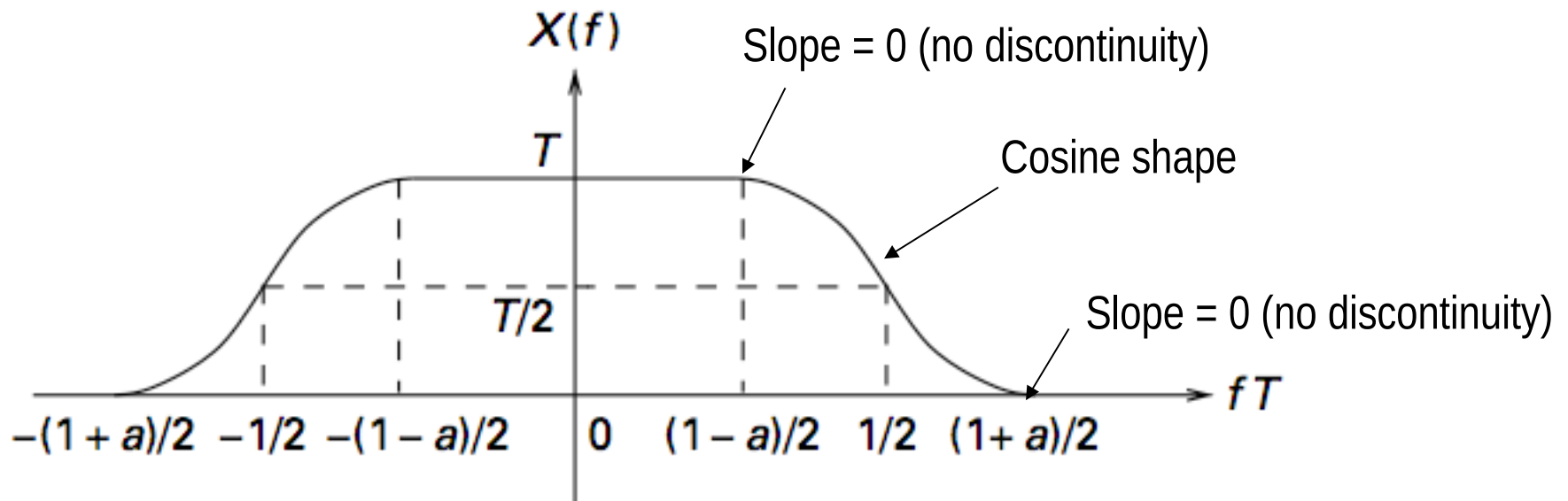
Product of two sincs means $1/t^2$ decay in time domain (good enough for peak power and ISI with timing mismatch to be bounded)

**Fractional excess bandwidth a often expressed as percentage
(e.g., $a=0.5$ corresponds to 50% excess bandwidth)**

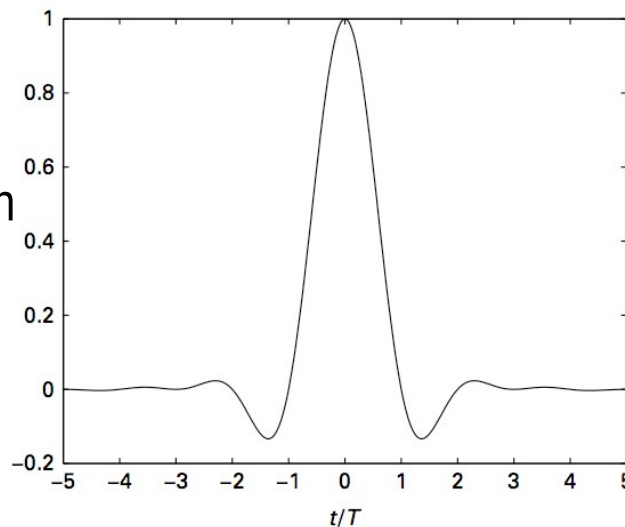
Trapezoidal frequency pulse has slope changes which translate to slower time decay

We can speed up decay in time if we make the roll-off in frequency more gentle

The Raised Cosine Pulse



Time domain pulse for
50% excess bandwidth



Time domain raised cosine pulse
has $1/t^3$ decay

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

Bandwidth efficiency for linear modulation

$$\eta_B = \log_2 M \text{ bit/symbol.}$$

Bandwidth efficiency for linear modulation
with M point constellation

$$B_{\min} = \frac{R_b}{\eta_B}$$

Minimum bandwidth needed for information rate of
 R_b bits/second

We do not count excess bandwidth when defining bandwidth efficiency
Just scale up minimum bandwidth by $(1+a)$ where a =excess bandwidth

Nyquist signaling

- Design the cascade of transmit and receive filters to be Nyquist at symbol rate
 - Channel outside our control
- Two approaches
 - Transmit filter Nyquist, receive filter wideband
 - Transmit and receive filters square root (in frequency domain) Nyquist
- Square root raised cosine (SRRC) a widely used pulse

Square root Nyquist pulses

Defn: $P(f)$ square root Nyquist at rate $1/T$ if $Q(f)=|P(f)|^2$ is Nyquist at rate $1/T$

$$Q(f) = |P(f)|^2 = P(f)P^*(f) \leftrightarrow (p * p_{MF})(t) = q(t)$$

Recall $p_{MF}(t) = p^*(-t)$

so that $q(t_0) = (p * p_{MF})(t_0) = \int p(t)p_{MF}(t_0 - t)dt = \int p(t)p^*(t - t_0)dt$

q is called the autocorrelation function of p , and is Nyquist if $q(mT) = \delta_{m0}$

Thus, p is square root Nyquist if it is uncorrelated with itself when shifted by integer multiples of T

$$\int p(t)p^*(t - mT)dt = \delta_{m0}$$

Example: Any pulse timelimited to duration T is square root Nyquist at $1/T$

Building on Linear Modulation

Linear modulation as a building block, I

Build complex waveforms using linear modulation of a *chip waveform* by a *chip sequence*
 Chip waveform chosen to be square root Nyquist at the *chip rate*

$\psi(t)$ is square root Nyquist at rate $1/T_c$.

$\longleftrightarrow Q(f) = |\Psi(f)|^2 = \Psi(f)\Psi^*(f)$ is Nyquist at rate $1/T_c$

$\longleftrightarrow \{\psi(t - kT_c)\}$ are orthonormal

Use these as a basis for constructing signals to be used for communication

M-ary signal set $s_i(t) = \sum_{k=0}^{N-1} s_i[k]\psi(t - kT_c), \quad i = 1, \dots, M$

Code vectors $\mathbf{s}_i = (s_i[0], \dots, s_i[N-1])$

Sequence of chips used for linear modulation of chip waveform

Can design code vectors to have desired properties (inner products/distances)
 Continuous-time signals inherit these properties

$$\langle s_i, s_j \rangle = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} s_i[k] s_j^*[l] \int \psi(t - kT_c) \psi^*(t - lT_c) dt = \sum_{k=0}^{N-1} s_i[k] s_j^*[k] = \langle \mathbf{s}_i, \mathbf{s}_j \rangle$$

Linear modulation as a building block, II

We shall see that for signaling in white Gaussian noise, what matters is the inner products between the signals

Building signals with linear modulation allows us to design in discrete time for a continuous time channel

Design code vectors to have desired properties (inner products/distances)

Continuous-time signals inherit these properties

$$\langle s_i, s_j \rangle = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} s_i[k] s_j^*[l] \int \psi(t - kT_c) \psi^*(t - lT_c) dt = \sum_{k=0}^{N-1} s_i[k] s_j^*[k] = \langle \mathbf{s}_i, \mathbf{s}_j \rangle$$

Much more on signal space geometry to come in Chapter 3

Orthogonal Modulation and its Variants

Orthogonal Modulation, I

Important example of **nonlinear modulation**

Good choice when we want *power efficiency* (to be shown in Chapter 3)

First let us think about what orthogonality means...

Consider two complex baseband signals $u = u_c + ju_s$ and $v = v_c + jv_s$

And the corresponding passband signals

$$u_p(t) = \text{Re}(\sqrt{2}u(t)e^{j2\pi f_c t}) \text{ and } v_p(t) = \text{Re}(\sqrt{2}v(t)e^{j2\pi f_c t})$$

Passband inner product in terms of complex baseband inner product:

$$\langle u_p, v_p \rangle = \text{Re}(\langle u, v \rangle) = \langle u_c, v_c \rangle + \langle u_s, v_s \rangle$$

Can we simply define orthogonality as $\text{Re}(\langle u, v \rangle) = 0$.

Yes, if the system is **coherent** (receiver's LO synchronized to incoming carrier in both frequency and phase)

Orthogonal Modulation, II

What if we do not have carrier sync?

Phase may be difficult to track in highly mobile environments

May choose not to track phase in order to lower cost/complexity

Received signal corresponding to v

$$\hat{v}_p(t) = \text{Re}(\sqrt{2}v(t)e^{j(2\pi f_c t + \theta)})$$

Template corresponding to u

$$u_p(t) = \text{Re}(\sqrt{2}u(t)e^{j2\pi f_c t})$$

Want these to remain orthogonal regardless of the phase shift θ

complex envelope of \hat{v}_p with respect to f_c is $\hat{v}(t) = v(t)e^{j\theta}$

$$\langle u_p, \hat{v}_p \rangle = \text{Re}(\langle u, \hat{v} \rangle) = \text{Re}(\langle u, v \rangle e^{-j\theta}).$$

Inner product vanishes for all possible phases if and only if $\langle u, v \rangle = 0$

(set phase to 0 and 90 degrees to see this)

Orthogonal Modulation, III

Two different notions of orthogonality

Coherent systems: Carrier phase and frequency synchronized

Noncoherent systems: Carrier phase and frequency not synchronized
(but phase assumed constant over signaling duration--small enough freq offset)

$\text{Re}(\langle s_i, s_j \rangle) = 0$ **Coherent orthogonality criterion**

$\langle s_i, s_j \rangle = 0$ **Noncoherent orthogonality criterion**

Let us now apply this to Frequency Shift Keying (FSK)

Frequency Shift Keying (FSK)

Send one of M tones over a symbol duration of T

$$s_i(t) = e^{j2\pi f_i t} I_{[0, T]}, \quad i = 1, \dots, M$$

Complex envelopes of the
 M possible transmitted signals

Bit rate $(\log_2 M)/T$

Bandwidth requirement for orthogonal FSK is twice for a noncoherent system

Coherent FSK: Required tone spacing is $1/(2T)$, approximate bandwidth $M/(2T)$

Noncoherent FSK: Required tone spacing is $1/T$, approximate bandwidth M/T

(See Problem 2.25)

M -ary noncoherent orthogonal signaling requires M complex dimensions

M -ary coherent orthogonal signaling requires M real dimensions (or $M/2$ complex dimensions)

Walsh-Hadamard codes

Can design discrete time orthogonal sequences, and then use linear modulation with square root Nyquist chip waveform to get continuous-time orthogonal waveforms
Walsh-Hadamard sequences are popular for this purpose
Binary, low peak-to-average ratio

Recursive construction (2^n orthogonal sequences at stage n)

$$\mathbf{H}_n = \begin{pmatrix} \mathbf{H}_{n-1} & \mathbf{H}_{n-1} \\ \mathbf{H}_{n-1} & -\mathbf{H}_{n-1} \end{pmatrix}.$$

Start with:

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \longrightarrow H_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad \text{and so on...}$$

Linearly modulated Walsh-Hadamard codes are orthogonal for noncoherent systems
For coherent systems, can use independently selectable Hadamard codes on both I and Q

Biorthogonal signaling

For a coherent system, consider an orthogonal signaling set

For every signal s , add $-s$ to the signaling set

This gives the *biorthogonal* signaling set

Example: Let $\{s_i\}$ denote M -ary Walsh-Hadamard codes.

We can use $\{s_i\}$ for noncoherent orthogonal signaling.

We can use $\{s_i, j s_i\}$ for coherent orthogonal signaling

We can use $\{s_i, j s_i, -s_i, -j s_i\}$ for (necessarily) coherent biorthogonal signaling

Increasing bandwidth efficiency (but differences get negligible as M gets large)

Other important forms of modulation

- Differential modulation: briefly mentioned in Ch. 2, more detail in Ch. 4
 - Permits operation with carrier phase sync, assuming phase approximately constant over multiple symbols
 - Encode information in phase differences over successive symbols
- Orthogonal Frequency Division Multiplexing (OFDM): Ch. 8
 - Use digital signal processing to divide channel bandwidth into narrow subchannels
 - Send symbols separately over subcarriers
 - General approach to handling frequency selective channels