

Chapter 5

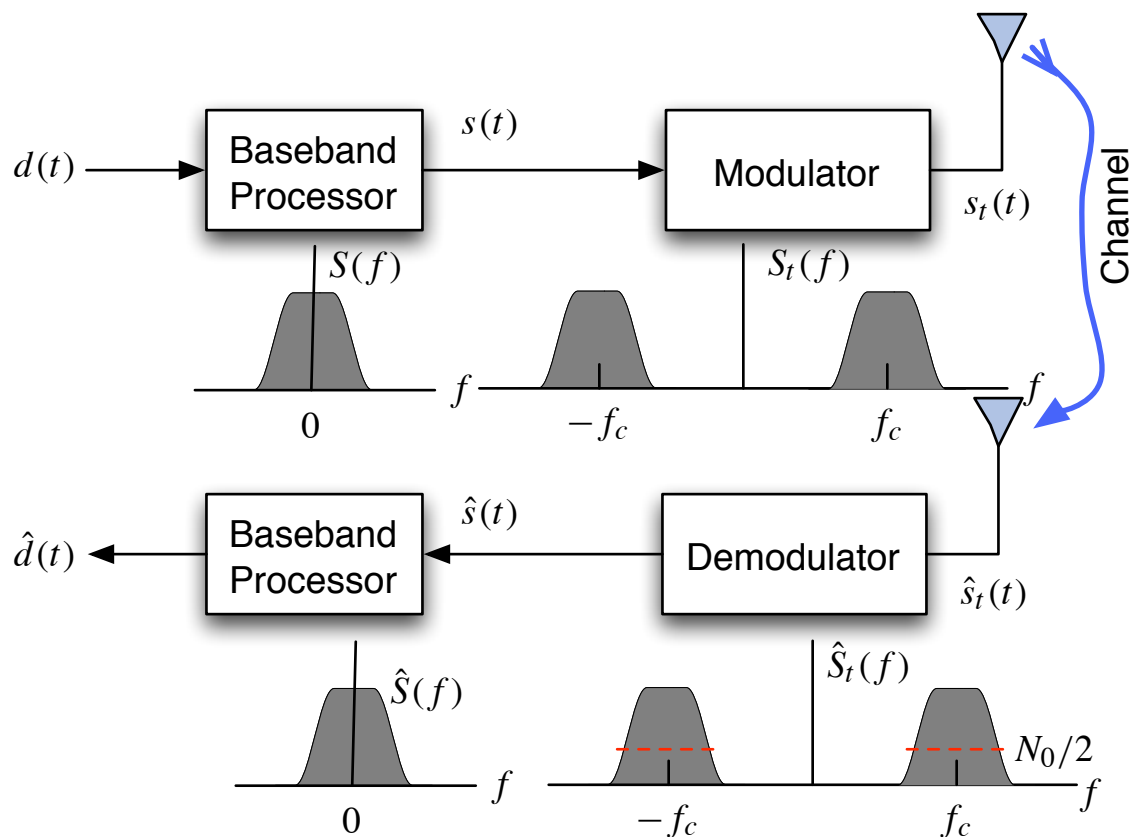
Introduction to Digital Data Transmission

Contents

5.1	Modulation	5-3
5.1.1	Modulation Classes	5-3
5.2	Linear Modulation	5-5
5.2.1	Binary Phase-Shift Keying (BPSK)	5-5
5.2.2	Quadrature-Phase-Shift Keying (QPSK) and Variations	5-9
5.3	Pulse Shaping	5-13
5.3.1	Raised Cosine Pulse	5-14
5.3.2	Square-Root Raised Cosine	5-17
5.4	Complex Baseband Representation	5-21
5.4.1	Complex Baseband Representation	5-22
5.5	Signal Space Representation	5-23
5.5.1	BPSK	5-24
5.5.2	QPSK	5-25
5.5.3	Other Two-Dimensional Schemes	5-25
5.6	Nonlinear Modulation Techniques	5-29
5.6.1	Binary Frequency-Shift Keying	5-29

5.6.2	Minimum Shift Keying	5-33
5.7	FDMA	5-41
5.7.1	Adjacent Channel Interference	5-42
5.7.2	Power Amplifier Nonlinearity	5-44
5.8	Modulation Comparison	5-47
5.8.1	Linear Channel	5-47
5.8.2	Nonlinear Channel	5-48
5.9	Channel Estimation and Tracking	5-49
5.9.1	Differential Detection	5-50
5.9.2	Pilot Transmission	5-52
5.10	Receiver Bit Error Probability Performance	5-54
5.10.1	AWGN Channel	5-55
5.10.2	Frequency Flat, Slow fading Channel	5-57
5.11	Theme Example: OFDM	5-59
5.11.1	Cyclic Prefix	5-62
5.12	Theme Example: Cordless Telephone	5-63

This chapter introduces a variety of digital modulation schemes found in modern wireless systems. The block diagram of a generic band-pass wireless link is shown below.



Generic bandpass wireless system

By its very nature, sending digital information involves a message signal that takes on only a finite number of values. At the waveform level the encoded digital message signal can be a continuous function of time, t . The data signal $d(t)$ may likely be speech that has been digitally encoded using a compression algorithm.

5.1 Modulation

The process of modulation varies some aspect of the carrier wave with respect to the modulating signal, e.g., $d(t)$. Demodulation performs the opposite operation, so as to gain recover an estimate of the information bearing signal $\hat{s}(t)$.

Three practical benefits of modulation include:

1. Shift the spectral content to an operating frequency band that is easily transmitted and received
2. The modulation operation may make the signal less vulnerable to noise and interference, e.g., frequency modulation
3. The scheme easily supports the use of a multiple-access technique

Generic Carrier Model

$$c(t) = A_c \cos(2\pi f_c t + \theta)$$

5.1.1 Modulation Classes

Modulation schemes can be compared via a number of classifications. Three such classifications are briefly explored before detailed investigations of selected schemes begins.

Linear vs Nonlinear

Two basic classes of modulation are *linear* and *nonlinear*. If the principle of superposition holds the modulation is linear.

The attributes of both linear and nonlinear modulation are explored in the remainder of this chapter.

Analog vs Digital

Modulation can also be classified along the lines of analog and digital. Here analog means the message signal, $m(t)$ is a continuous function of time, so the modulated carrier, $s_t(t)$, has attributes vary continuously over some parameter range.

With digital modulation the message signal, $d(t)$, takes on discrete values, e.g., ± 1 with the switching instants occurring at time interval T_s . The modulated carrier $s_t(t)$ will still be a function of continuous variable t , but the values the signal takes on will have a discrete nature, e.g., amplitude, frequency, or phase.

The focus in this chapter will be on digital modulation schemes.

Amplitude vs Angle

The particular attribute of the carrier signal $c(t)$ that is varied forms another basis for modulation classification. Two broad classes are amplitude and angle modulation. Angle modulation further breaks down into phase and frequency modulation.

- Amplitude modulation simply requires the carrier amplitude A_c to vary linearly with respect to $m(t)$ or $d(t)$
- With angle modulation we consider the entire argument of $\cos(\cdot)$ in $c(t)$ as the angle

$$\psi(t) = 2\pi f_c t + \theta(t)$$

of the carrier, and design a modulator so that it varies linearly with $m(t)$ or $d(t)$

- Specifically with frequency modulation, the derivative $d\psi(t)/dt$ is made to vary linearly with the message signal
- Specifically with phase modulation the carrier phase $\theta(t)$ is made to vary linearly with the message

5.2 Linear Modulation

5.2.1 Binary Phase-Shift Keying (BPSK)

To create BPSK we first construct a baseband data signal of the form

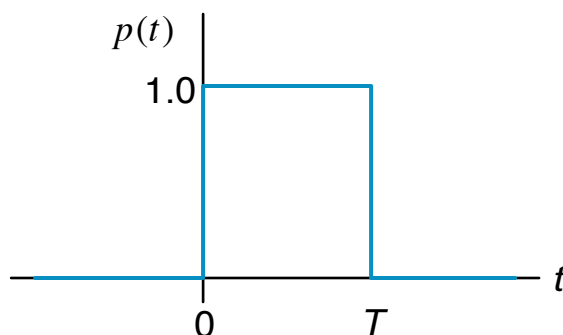
$$d(t) = \sum_k b_k p(t - kT)$$

where b_k is a bipolar bit sequence of the form

$$b_k = \begin{cases} +1, & \text{binary symbol 1} \\ -1, & \text{binary symbol 0} \end{cases}$$

- A fundamental pulse shape $p(t)$ is the rectangle shape

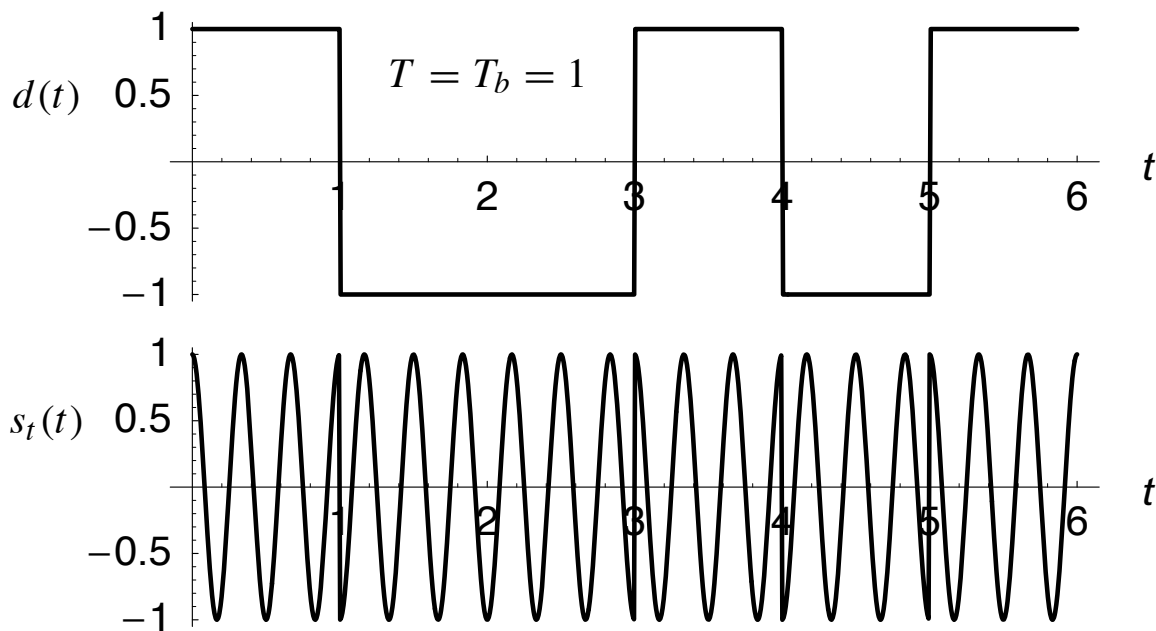
$$p(t) = \begin{cases} 1, & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$



- BPSK constitutes a form a digital phase modulation in the carrier phase is switched between $\theta(t) = 0$ and π radians depending upon the sign of b_k
- Since a phase of π radians simply changes the sign of the carrier signal, we observe that for the case of BPSK

$$s_t(t) = d(t)c(t)$$

which is of the same form as double sideband suppressed carrier (DSB-SC) modulation



BPSK using a rectangle pulse

- The power spectral density of BPSK can be shown to be of the form

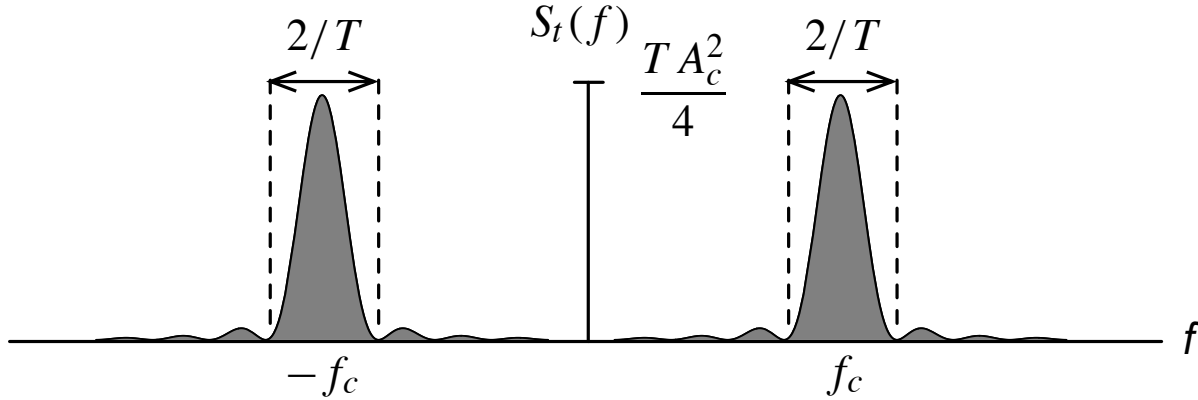
$$S_t(f) = \frac{A_c^2}{4T} [|P(f + f_c)|^2 + |P(f - f_c)|^2]$$

where $P(f) = \mathcal{F}[p(t)]$

- For the rectangular pulse shape

$$P(f) = T \frac{\sin(2\pi f T)}{2\pi f T} = T \text{sinc}(f T)$$

where $\text{sinc}(x) \equiv \sin(\pi x)/(\pi x)$



BPSK spectrum

- For a rectangular pulse shape the main lobe bandwidth, also known as the RF bandwidth is $B_{\text{RF}} = 2/T = 2R_b$, where $R_b = 1/T$ is the bit rate
- Another useful bandwidth measure is the fractional containment bandwidth, B_f , defined as

$$P_f = \frac{\int_{f_c - B_f/2}^{f_c + B_f/2} S_t(f) df}{\int_0^\infty S_t(f) df}$$

where P_f is a fraction of the total power in $s_t(t)$

Example 5.1: BPSK with Rectangle Pulse Shape B_f

- The fractional containment bandwidth of rectangle pulse shaped BPSK can be found from

$$P_f = \frac{\int_0^{B_f/2} \text{sinc}^2(fT) df}{\int_0^\infty \text{sinc}^2(fT) df}$$

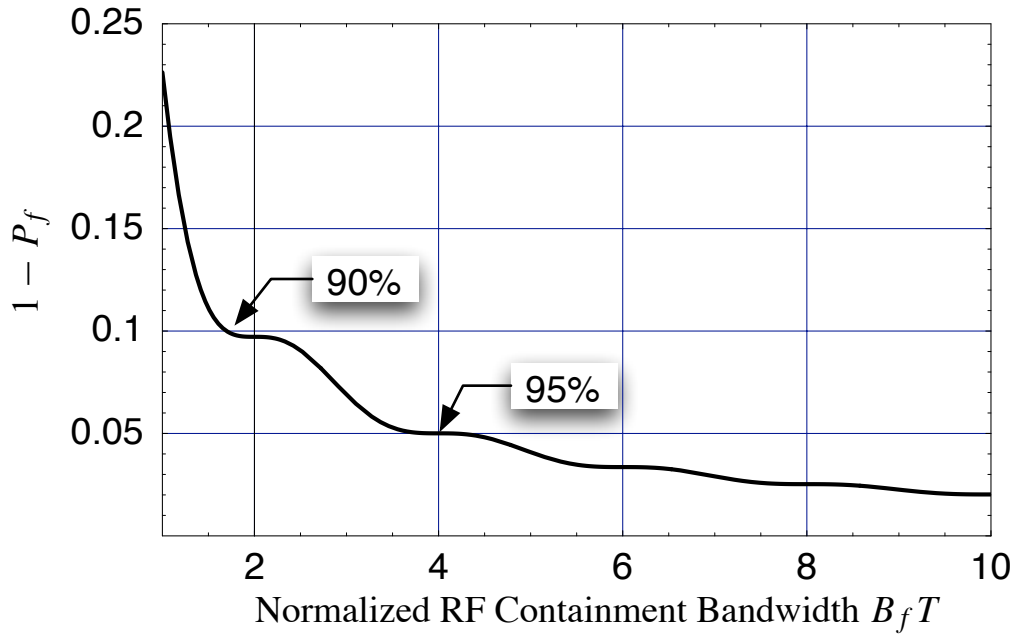
- We first work with the denominator using Parseval's theorem

$$\begin{aligned} \int_0^\infty \text{sinc}^2(fT) df &= \frac{1}{2} \int_{-\infty}^\infty \text{sinc}^2(fT) df \\ &= \frac{1}{2T} \int_{-\infty}^\infty \text{sinc}^2(x) dx = \frac{1}{2T} \end{aligned}$$

- Inserting in the numerator, and changing variables, we have

$$P_f = 2 \int_0^{B_f T/2} \text{sinc}^2(x) dx$$

- At $B_f T = 2$ the contained power is only 0.9028, while at $B_f T = 5$ the contained power has only increased to 0.9592
- The 99% containment bandwidth occurs when $B_f T = 20.57$
- Clearly pulse shaping beyond the rectangular pulse shape is needed



Rect pulse BPSK fractional out-of-band power

5.2.2 Quadriphase-Shift Keying (QPSK) and Variations

To make more effective use of the available spectrum we may choose to modulate both the sin and cos carriers (quadrature multiplexing).

- Consider a binary data stream $d(t)$ demultiplexed into two equal bit rate streams, $d_1(t)$ and $d_2(t)$ respectively

$$d(t) = \sum_k b_k p(t - kT_b)$$

$$d_1(t) = \sum_k b_{2k} p(t - kT_s)$$

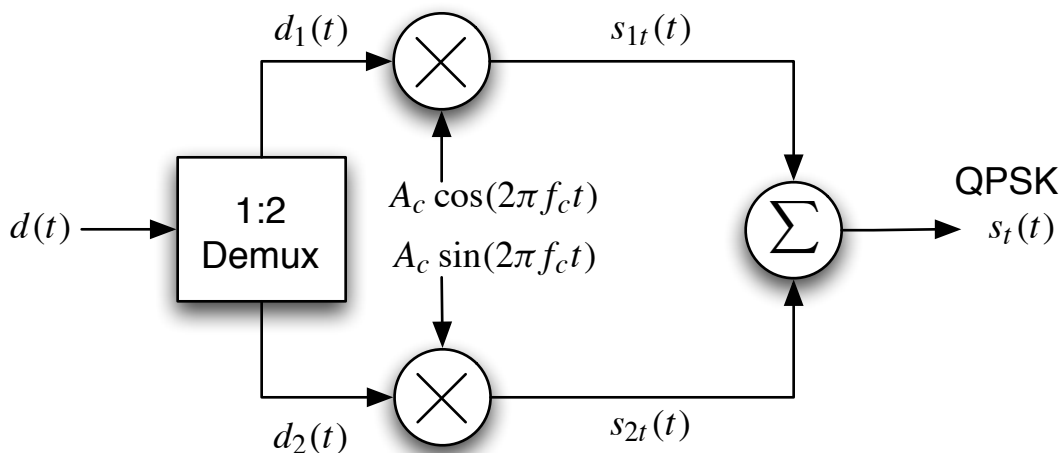
$$d_2(t) = \sum_k b_{2k+1} p(t - kT_s)$$

where $T_s = 2T_b$ is the symbol duration, which is twice the bit duration

- Note that the symbol rate is $R_s = 1/T_s = R_b/2$
- With QPSK we transmit two bits per symbol
- The data streams are applied to orthogonal, but at the same frequency, carriers

$$s_t(t) = s_1(t) + s_2(t)$$

$$s_t(t) = A_c [d_1(t) \cos(2\pi f_c t) + d_2(t) \sin(2\pi f_c t)]$$



QPSK modulator block diagram

Standard QPSK

With this form of QPSK it looks like we have two equal bit rate BPSK modulators operating in parallel.

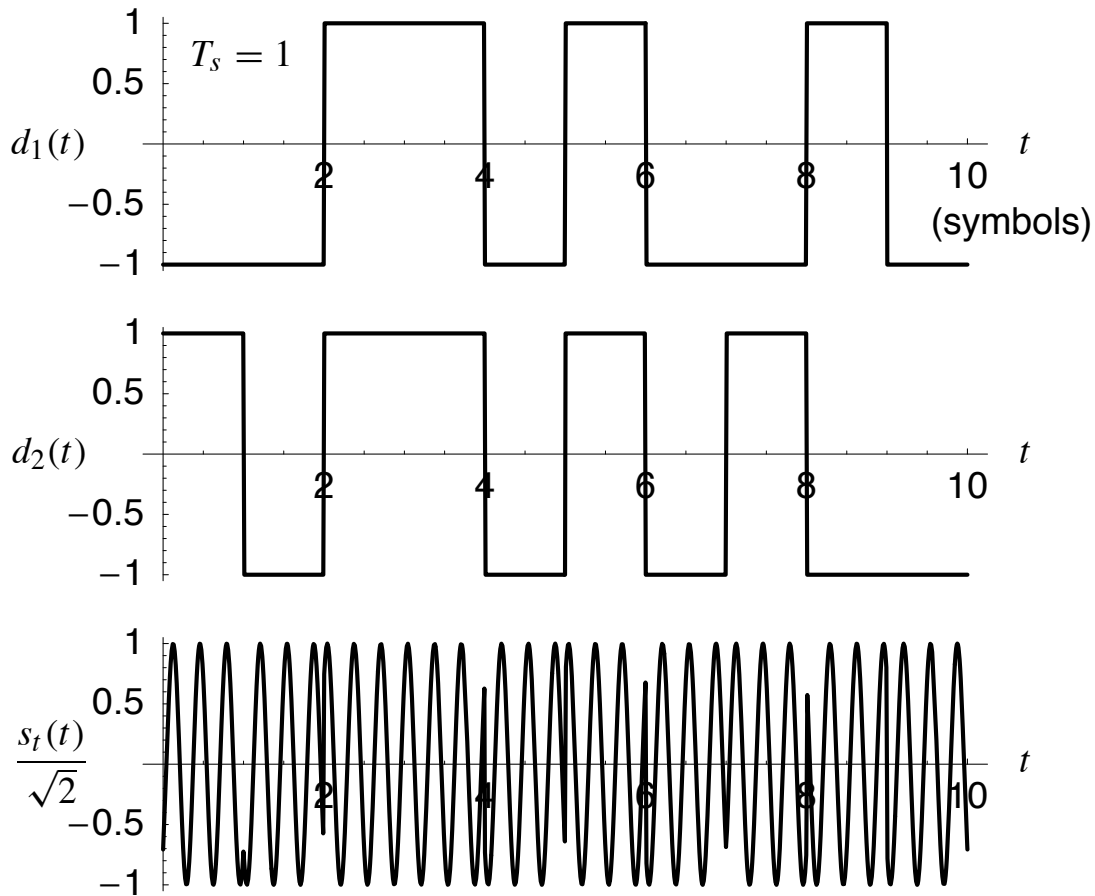
- The two carrier signals are said to be in *phase quadrature* since the sine lags the cosine by 90°

- The signal $d_1(t)$, modulating the cosine carrier, is known as the *in-phase signal* or I component, while the signal $d_2(t)$, modulating the sine carrier, is known as the *quadrature signal*
- Assuming equal bit rates and pulse shaping on the I and Q components, the spectrum of QPSK is identical to that of BPSK, that is

$$S_t(f) = \frac{A_c^2}{2T_s} [|P(f + f_c)|^2 + |P(f - f_c)|^2]$$

Note the scale factor difference from BPSK because we have two equal power carriers forming the QPSK signal.

- Rectangular pulse shaping the main lobe RF bandwidth is now $B_{\text{RF}} = 2R_s = R_b$, and bandwidth reduction of 2 compared with BPSK
- Since each carrier phase is modulated between 0° and 180° , and are in phase quadrature, composite carrier phase takes on the four values of $\{45^\circ, 135^\circ, 225^\circ, 315^\circ\}$
- A specific property of standard QPSK is that the composite carrier phase may undergo phase jumps of 0° , $\pm 90^\circ$, or $\pm 180^\circ$ every $2T_b = T_s$ seconds



QPSK using a rectangle pulse

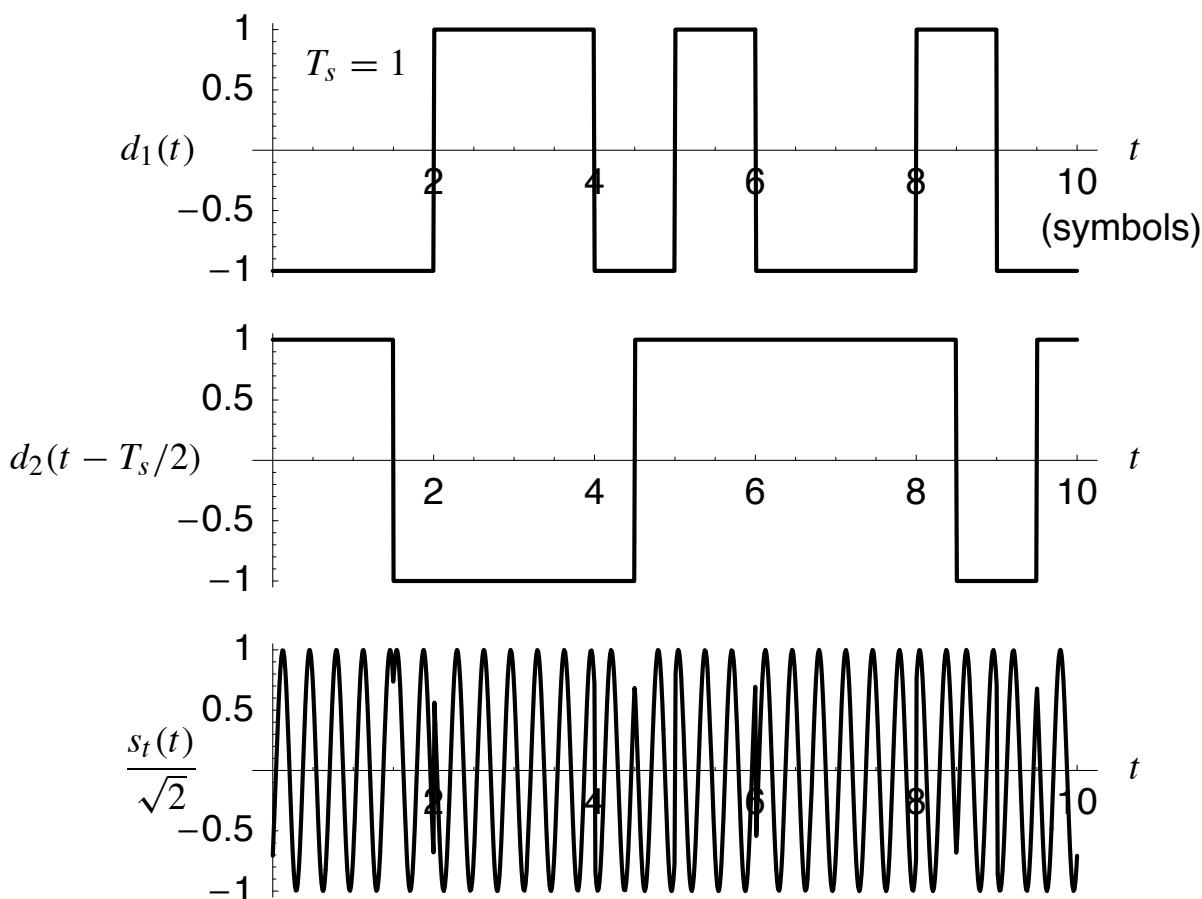
Offset QPSK (OQPSK)

The $\pm 180^\circ$ phase jumps in QPSK can be a problem when the signal is filtered and then amplified by a nonlinear power amplifier.

- OQPSK, also known as *staggered QPSK*, is formed by delaying the quadrature signal bit stream by $T_s/2 = T_b$, thus limiting phase jumps to just 0° and $\pm 90^\circ$
- The transmitted signal now takes the form

$$s_t(t) = A_c [d_1(t) \cos(2\pi f_c t) + d_2(t - T_s/2) \sin(2\pi f_c t)]$$

- The waveforms change accordingly



OQPSK using a rectangle pulse

5.3 Pulse Shaping

For both BPSK and QPSK we have seen how the rectangular pulse shape, while easy to implement, creates a wide spectral footprint in the neighborhood of the carrier frequency, f_c . We now consider the use of pulse shaping or a *premodulation filter* to better match the transmitted signal spectrum to the available channel bandwidth. We specifically desire:

1. A more compact spectrum to allow more digitally modulated carriers to occupy a frequency band allocation
2. A means to manage channel induced bandlimiting, e.g., due to multipath, which introduces *intersymbol interference* (ISI), which occurs when the energy of previous symbols interferes/overlaps with the energy of the present symbol (pulse)

5.3.1 Raised Cosine Pulse

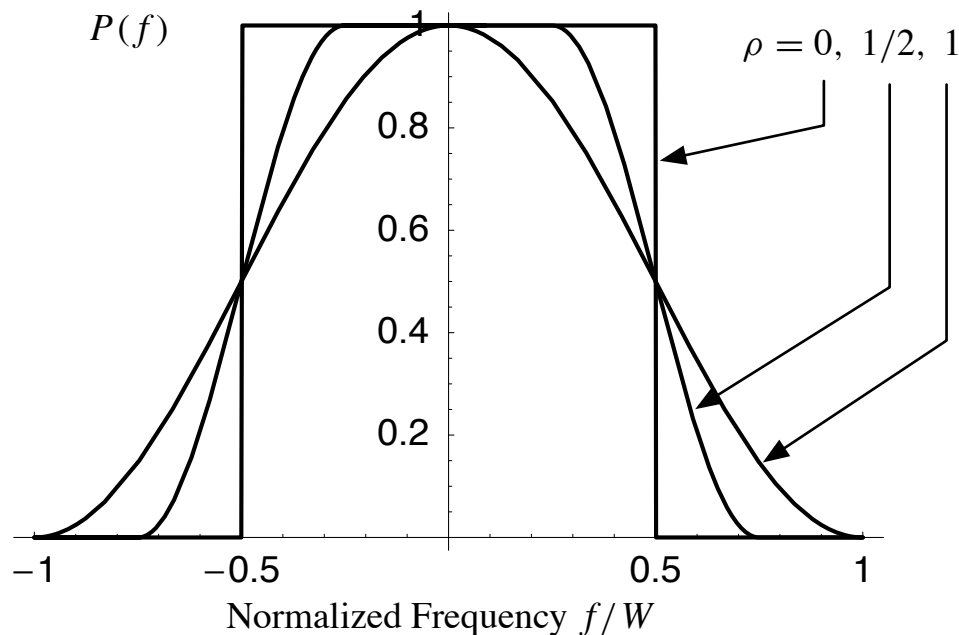
- A premodulation filter and/or pulse shaping satisfies, in part, both of the above requirements
- A popular class of pulse shaping that achieves both band limiting and ISI control is the *raised cosine* (RC) pulse
- The RC pulse has a spectrum with adjustable bandwidth in the RF (two-sided) sense running from $W = R_b$ to $2W = 2R_b$
- The pulse spectrum (Fourier transform) is defined by

$$P(f) = \begin{cases} \frac{1}{2W}, & 0 \leq |f| < f_1 \\ \frac{1}{4W} \left[1 + \cos \left(\frac{\pi}{2W\rho} (|f| - f_1) \right) \right], & f_1 \leq |f| < 2W - f_1 \\ 0, & \text{otherwise} \end{cases}$$

where f_1 sets the edge of the flat portion of the spectrum and is related to the roll-off factor ρ and W via

$$0 \leq \rho = 1 - \frac{f_1}{W} \leq 1$$

- The parameter ρ (elsewhere denoted α) controls the excess bandwidth relative to the minimum value of $W = R_b$ (one-sided $W/2$) when $\rho = 0$

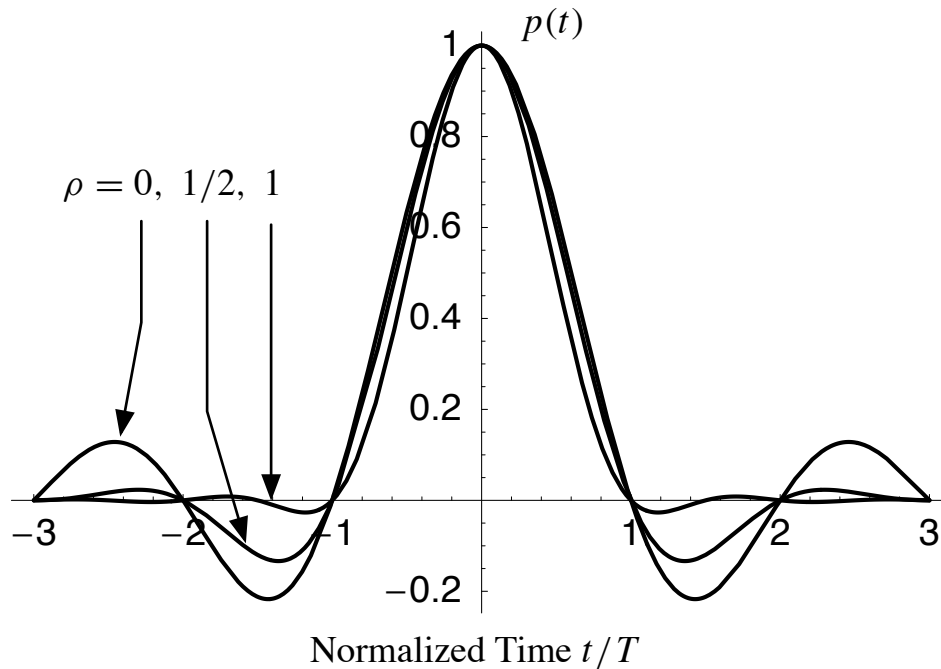


RC spectrum, $P(f)$, for various ρ values

- The RC pulse gets its name from the $1 + \cos()$ term in the spectrum definition
- The corresponding RC pulse, $p(t)$, can be obtained by inverse Fourier transforming $P(f)$

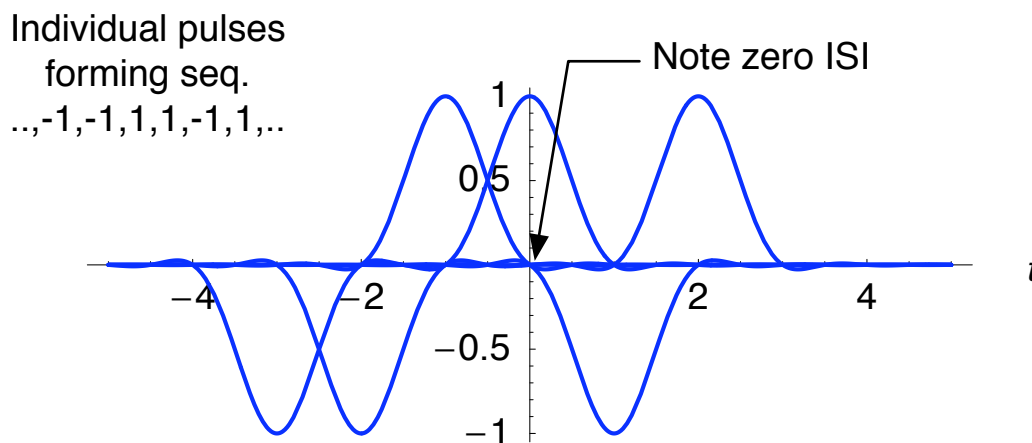
$$p(t) = \mathcal{F}^{-1}\{P(f)\} = \left(\frac{\cos(2\pi\rho Wt)}{1 - 16\rho^2 W^2 t^2} \right) \text{sinc}(2Wt)$$

- The zero ISI property of the RC pulse is that although $p(0) = 1$, $p(nT) = 0$ for $\dots, -2, -1, 1, 2, \dots$

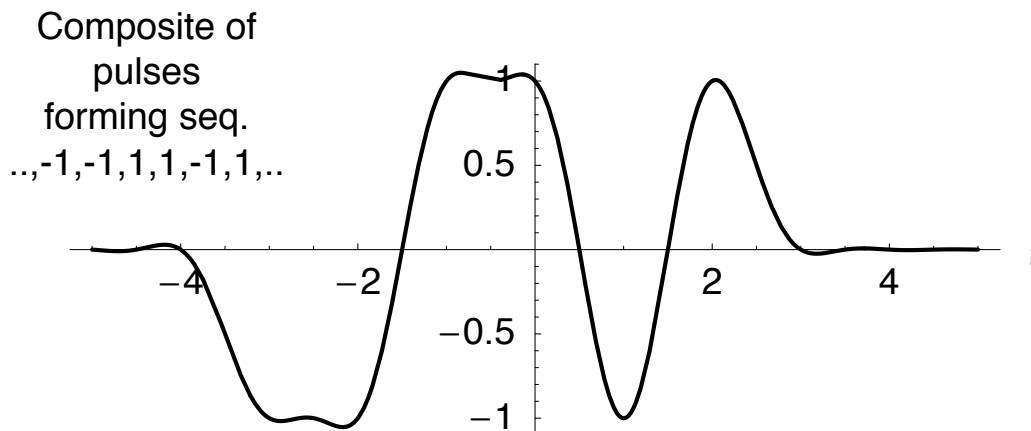
RC pulse, $p(t)$, for various ρ values

Example 5.2: RC Waveform for a Bit Pattern

Consider the waveform produced by the bit sequence $\{0, 0, 1, 1, 0, 1\}$ or in bipolar form $\{-1, -1, 1, 1, -1, 1\}$.



Waveform created with bit pattern -1,-1,1,1,-1,1

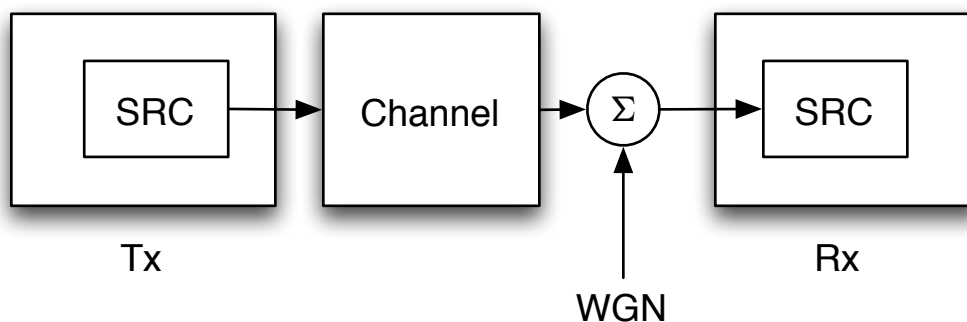


Composite waveform created with bit pattern $-1, -1, 1, 1, -1, 1$

5.3.2 Square-Root Raised Cosine

The zero ISI response holds for the RC pulse, but optimal filtering in an additive noise environment, requires that filtering/pulse shaping be distributed between the transmitter and receiver.

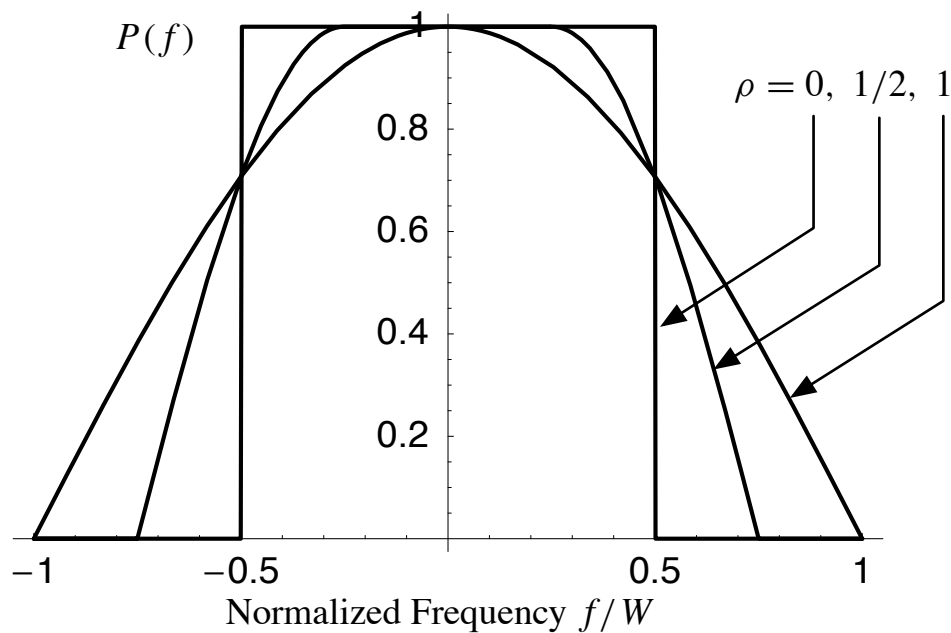
- The *root raised-cosine* or *square-root raised-cosine* (SRC) filter satisfies this requirement



SRC (also denoted RRC) used as a Tx/Rx filter pair

- The spectrum at the receiver becomes the $P^2(f)$, which for the case of the SRC is again the RC
- The SRC pulse spectrum is defined as

$$P_{\text{SRC}}(f) = \begin{cases} \frac{1}{\sqrt{2W}}, & 0 \leq |f| < f_1 \\ \frac{1}{\sqrt{2W}} \cos\left(\frac{\pi}{4W\rho}(|f| - W(1-\rho))\right), & f_1 \leq |f| < 2W - f_1 \\ 0, & \text{otherwise} \end{cases}$$



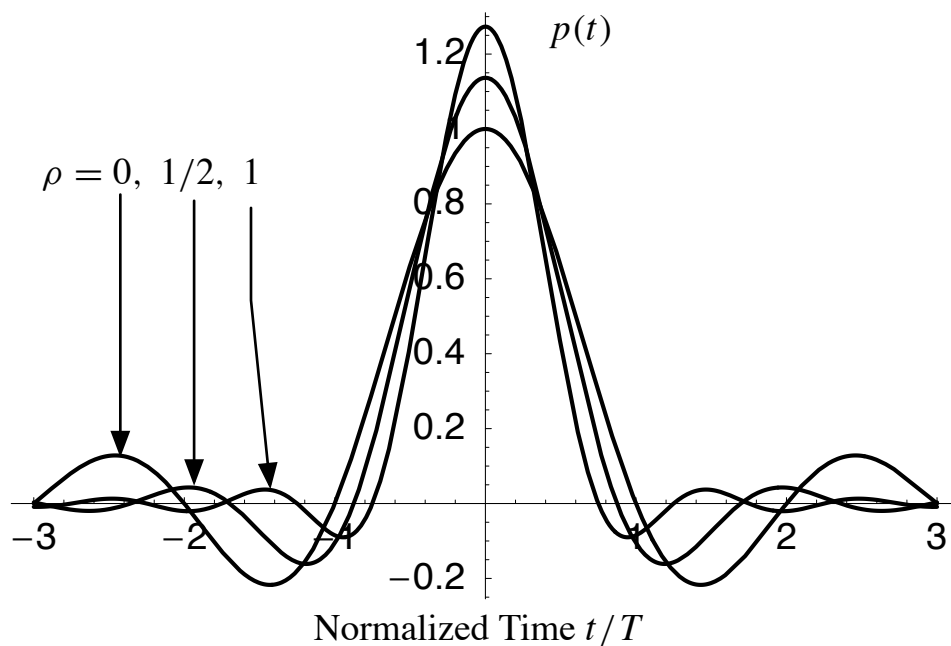
SRC spectrum, $P(f)$, for various ρ values

- The corresponding RC pulse, $p(t)$, can be obtained by inverse Fourier transforming $P(f)$

$$\begin{aligned}
 p(t) &= \mathcal{F}^{-1}\{P(f)\} \\
 &= \frac{\sqrt{2W}}{1 - 64\rho^2 W^2 t^2} \left[\frac{\sin(2\pi W(1 - \rho)t)}{2\pi W t} \right. \\
 &\quad \left. + \frac{4\rho}{\pi} \cos(2\pi W(1 + \rho)t) \right]
 \end{aligned}$$

- The SRC pulse does not have a zero ISI property as with the RC, but it does have an orthogonality condition

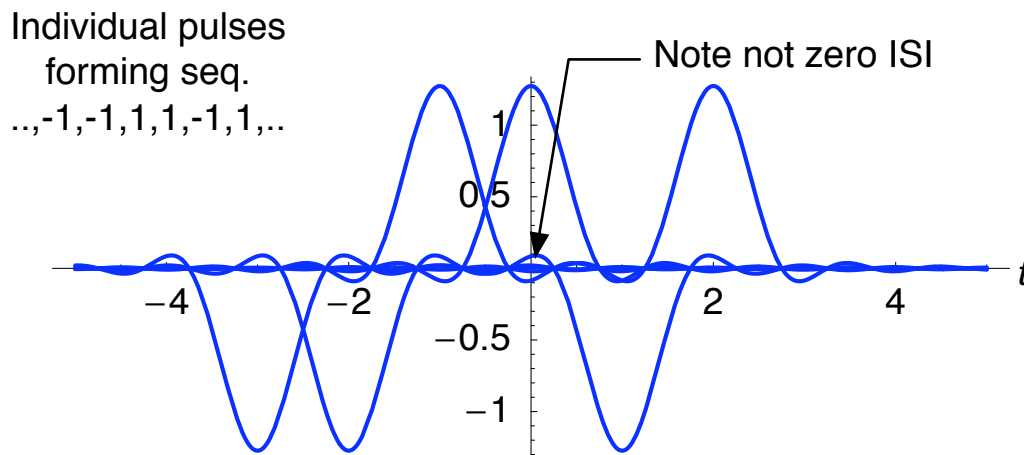
$$\int_{-\infty}^{\infty} p(t)p(t - nT) dt = 0 \text{ for } n = \pm 1, \pm 2, \dots$$



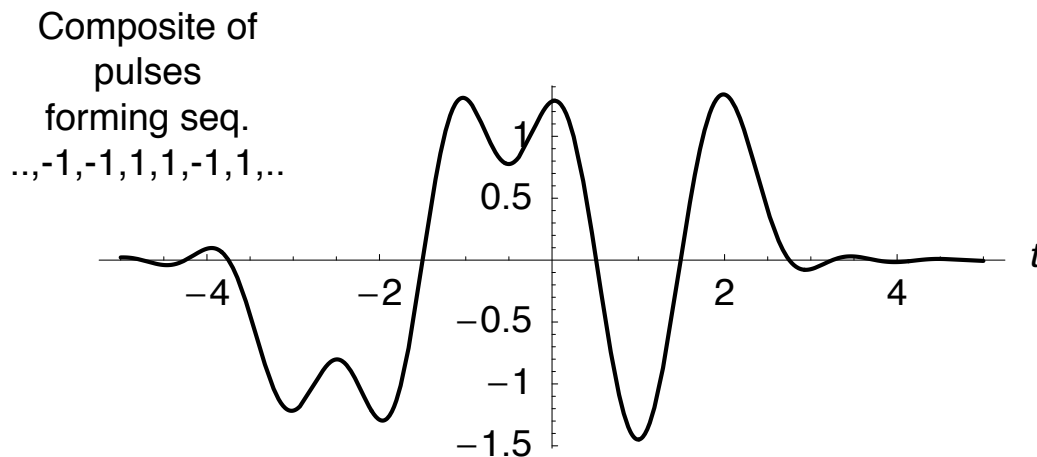
SRC pulse, $p(t)$, for various ρ values

Example 5.3: SRC Waveform for a Bit Pattern

Consider the waveform produced by the bit sequence $\{0, 0, 1, 1, 0, 1\}$ or in bipolar form $\{-1, -1, 1, 1, -1, 1\}$.



Waveform created with bit pattern $-1, -1, 1, 1, -1, 1$



Composite waveform created with bit pattern $-1, -1, 1, 1, -1, 1$

- Notice that the zero ISI condition is not met until the above waveform is passed through an SRC filter in the receiver

- Also note that the equivalent transmit waveform (figure immediately above) requires a greater dynamic range than the corresponding RC waveform

5.4 Complex Baseband Representation

As we continue to study various digital modulation schemes, we are motivated to consider the complex envelope representation. To establish the equivalence of the complex envelope form we start with the so-called *band-pass signal canonical form*:

$$s(t) = s_I(t) \cos(2\pi f_c t) - s_Q(t) \sin(2\pi f_c t)$$

- Here the in-phase modulating signal is denoted $s_I(t)$ and the quadrature modulating signal is denoted $s_Q(t)$
- The complex envelope is the complex signal

$$\tilde{s}(t) = s_I(t) + js_Q(t)$$

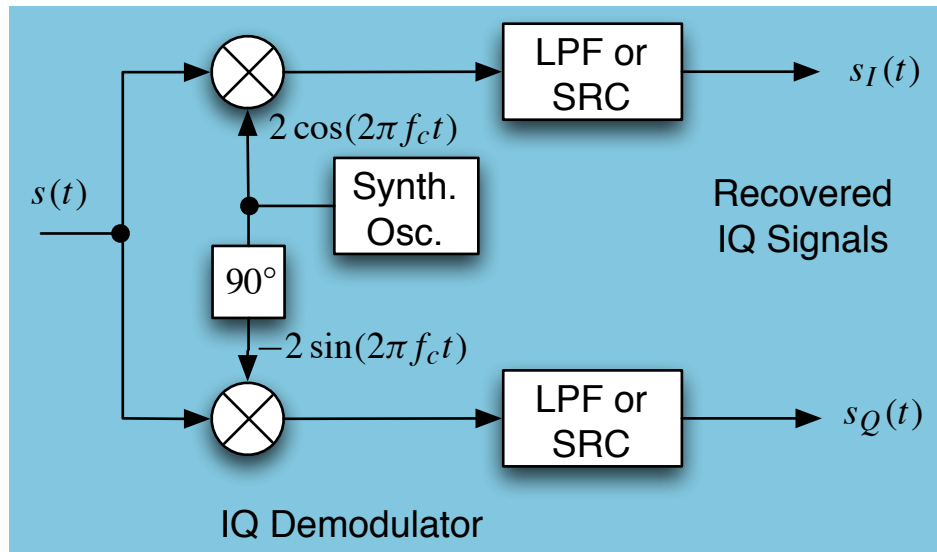
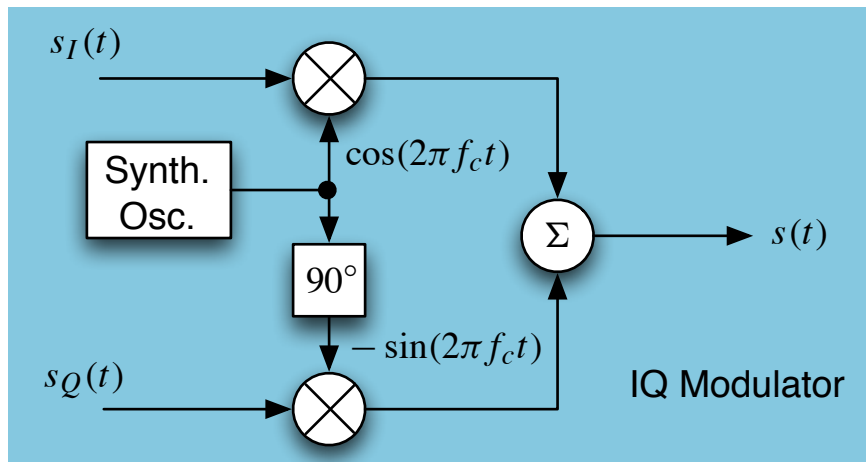
- The relationship between $s(t)$ and $\tilde{s}(t)$ is obtained by noting that

$$s(t) = \text{Re} \{ \tilde{s}(t) e^{j2\pi f_c t} \}$$

where $\exp(j2\pi f_c t) = \cos(2\pi f_c t) + j \sin(2\pi f_c t)$

- When dealing with the complex envelope the carrier frequency is effectively suppressed
- The complex envelope (*complex baseband*) form is also lends itself to simplified computer simulation and actual hardware implementations in ASIC/FPGA/general purpose DSP

- The complex envelope representation suggests the following IQ modulator/demodulator structures:



IQ Mod/Demod

5.4.1 Complex Baseband Representation

We can carry the complex envelope idea further by modeling band-pass filtering in terms of a complex baseband impulse response.

- The impulse response of a bandpass filter can be written as

$$h(t) = \text{Re} \left\{ \tilde{h}(t) e^{j2\pi f_c t} \right\}$$

where

$$\tilde{h}(t) = h_I(t) + jh_Q(t)$$

- Consider the bandpass signal $x(t)$ filtered by bandpass filter $h(t)$ to produce $y(t)$

$$y(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\lambda)h(t - \lambda) d\lambda$$

- Using complex envelopes and the complex baseband impulse response, we can equivalently write

$$\tilde{y}(t) = \frac{1}{2}\tilde{x}(t) * \tilde{h}(t) = \frac{1}{2} \int_{-\infty}^{\infty} \tilde{x}(\lambda)\tilde{h}(t - \lambda) d\lambda$$

- Once $\tilde{y}(t)$ is obtained we can return to $y(t)$ via

$$y(t) = \text{Re} \{ \tilde{y}(t)e^{j2\pi f_c t} \}$$

5.5 Signal Space Representation

As modulation schemes become more complex, a *signal space* representation becomes convenient for performance analysis purposes. Two dimensional signal *constellations* have been studied extensively.

- Traditionally the coordinate system is established via energy normalized version of the in-phase and quadrature signals, denoted $\phi_1(t)$ and $\phi_2(t)$

$$\begin{aligned}\phi_1(t) &= \sqrt{\frac{2}{T}} \cos(2\pi f_c t), \quad 0 \leq t \leq T \\ \phi_2(t) &= \sqrt{\frac{2}{T}} \sin(2\pi f_c t), \quad 0 \leq t \leq T\end{aligned}$$

- Two immediate properties of this construction approach are:

1. Orthogonality –

$$\int_0^T \phi_1(t)\phi_2(t) dt = 0$$

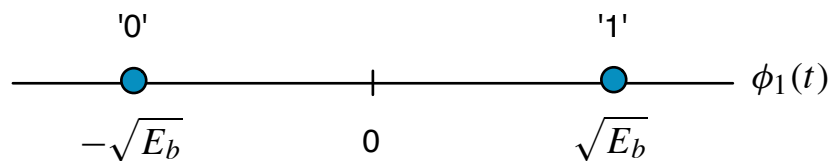
2. Unit length/unit energy –

$$\int_0^T |\phi_1(t)|^2 dt = \int_0^T |\phi_2(t)|^2 dt = 1$$

- The functions (*basis functions*) $\phi_1(t)$ and $\phi_2(t)$ constitute an *orthonormal set*
- With these two functions a variety of digital modulation schemes can be represented

5.5.1 BPSK

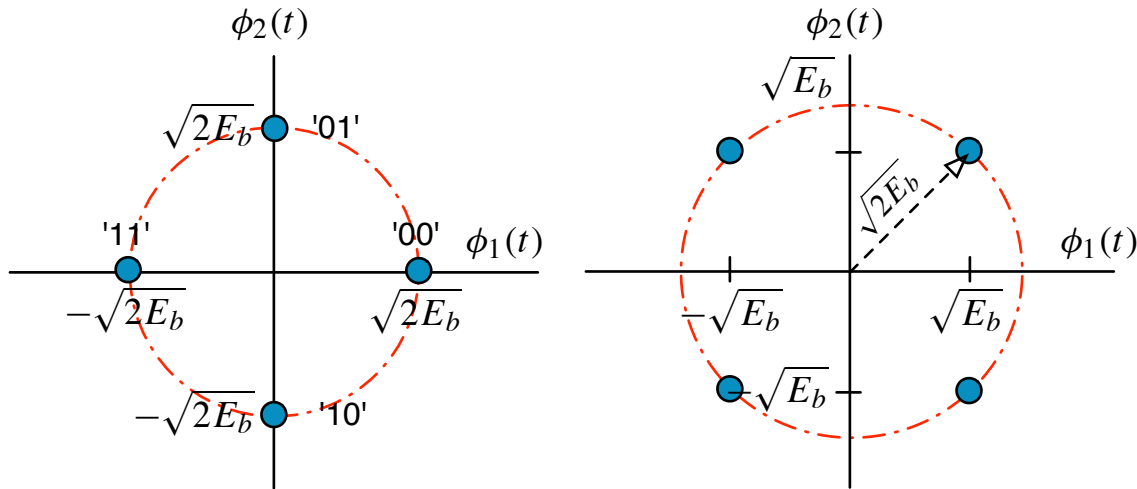
BPSK requires just a single dimension, say $\phi_1(t)$ to describe its signal constellation. Notice that in signal space the length squared has units of signal energy per bit, E_b (or symbol E_s).



BPSK 1-D signal space

5.5.2 QPSK

QPSK and its variations require two dimensions. Note $E_s = 2E_b$.



QPSK 2-D signal space: two variations

5.5.3 Other Two-Dimensional Schemes

A very large number of digital modulation schemes can be reduced to a two-dimensional signal space. Two schemes found in wireless systems are M -ary PSK (MPSK) and quadrature-amplitude modulation (QAM).

MPSK

With BPSK and QPSK defined, the logical extension is MPSK which encodes the transmission bits into one of M phase values.

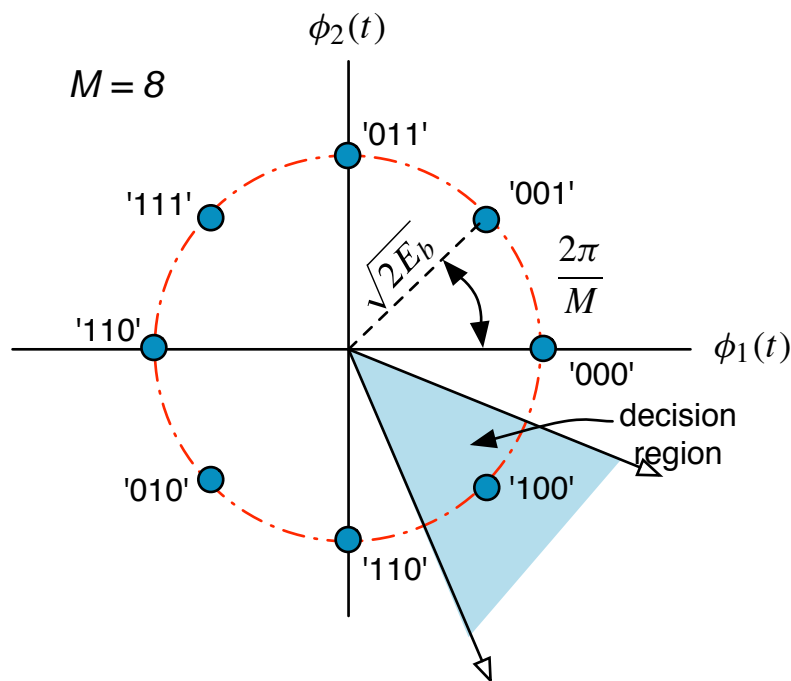
- The values that M takes on are 2, 4, 8, 16, 32, ...

- The number of bits per symbol is $\log_2(M)$, so considering the null-to-null bandwidth, MPSK has bandwidth efficiency compared to BPSK of

$$B_{\text{eff}} = \frac{R_b}{2} \log_2(M) \text{ bits/s/Hz}$$

where R_b is the serial bit rate in bits/s

- The signal space of MPSK is 2-dimensional with adjacent signal points moving closer together as M increases for a fixed E_b
- Notice that the bits are encoded using *Gray coding* which insures that two adjacent symbols differ by no more than one bit



MPSK with $M = 8$

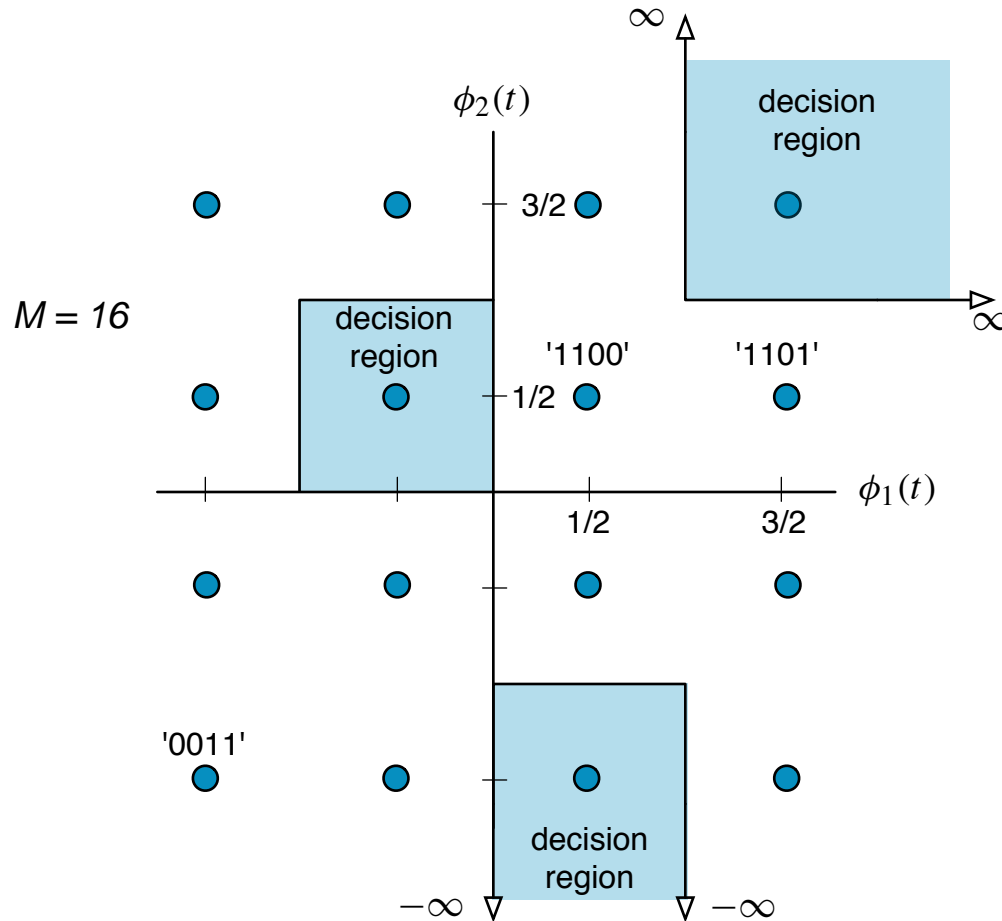
- Note that the point in signal space where a receive signal must lie in order to make a symbol decision (*decision regions*), are pie shaped in the case of MPSK

QAM

If we include amplitude modulation along with phase modulation we maintain a 2-D signal space, but now allow a much denser array of signal points. The new modulation scheme is known as *quadrature amplitude modulation* (QAM). At complex baseband the general form of QAM is

$$\tilde{s}(t) = \sum_{k=-\infty}^{\infty} a_k p(t - kT_s) + j \sum_{k=-\infty}^{\infty} b_k p(t - kT_s)$$

where a_k and b_k are the in-phase and quadrature amplitude values of the QAM symbols taking on values of $\pm 1, \pm 3, \pm 5, \dots$

QAM with $M = 16$ (16QAM)

- With QAM symbol decisions are made with different size decision regions
- The fact that information is carried in the amplitude and the phase also means that QAM is more sensitive to nonlinearities

5.6 Nonlinear Modulation Techniques

A large class of modulation schemes employ nonlinear signal processing in the modulation process. For these schemes it is convenient to represent the transmitted signal in *polar form*, that is

$$\begin{aligned} s(t) &= \{[s_I(t) + js_Q(t)]e^{j2\pi f_c t}\} \\ &= a(t) \cos[2\pi f_c t + \theta(t)] \end{aligned}$$

where

$$\begin{aligned} a(t) &= \sqrt{s_I^2(t) + s_Q^2(t)} \\ \theta(t) &= \tan^{-1} \left(\frac{s_Q(t)}{s_I(t)} \right) \end{aligned}$$

5.6.1 Binary Frequency-Shift Keying

To create a digital FM scheme we need only make the instantaneous frequency of the carrier a function of the transmission bit stream. With binary frequency-shift keying the carrier switches between two frequencies according to the transmitted bit type ‘0’ or ‘1’, e.g.,

$$s(t) = \begin{cases} \sqrt{\frac{2E_b}{T}} \cos(2\pi f_1 t + \phi), & 0 \leq t \leq T_b, \text{ 1's transmission} \\ \sqrt{\frac{2E_b}{T}} \cos(2\pi f_2 t + \phi), & 0 \leq t \leq T_b, \text{ 0's transmission} \end{cases}$$

where f_1 and f_2 are the BFSK tone frequencies and ϕ may be modeled as a random variable.

- The choice of f_1 and f_2 , or better yet $|f_2 - f_1|$ is a design parameter

- In *Sunde's FSK*¹ we choose

$$f_i = \frac{n_c + i}{T}, \quad i = 1, 2$$

with n_c a fixed integer and assume that $\phi = 0$

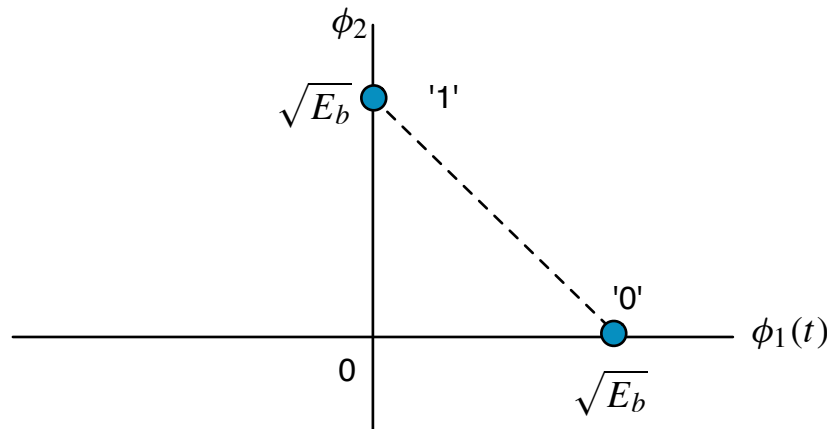
- As each successive bit is transmitted phase continuity is maintained, thus this variation of FSK is known as *continuous-phase FSK (CFSK)*
- When f_i is chosen as described above, it turns out that

$$\int_0^T \sqrt{\frac{2E_b}{T}} \cos(2\pi f_i t) \sqrt{\frac{2E_b}{T}} \cos(2\pi f_j t) dt = 0, \quad i \neq j$$

- The corresponding signal space orthonormal basis functions are

$$\phi_i(t) = \begin{cases} \sqrt{\frac{2}{T}} \cos(2\pi f_i t), & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$

- M -ary FSM (MFSK) can be created by using more orthogonal frequencies, hence the dimensionality of the signal space increases proportionately



¹Haykin and Moher p. 132.

Orthogonal BFSK signal space

BFSK Power Spectrum

The exact power spectrum of BFSK is difficult to obtain, but an approximate spectrum can be obtained using a *quasi-static* analysis. This analysis assumes the power spectrum can be formed as a superposition of the power spectrum due to on-off keying of the individual carriers f_1 and f_2 .

- Assuming a rectangular pulse shape for $p(t)$, which is also the easiest to implement in simple FSK hardware, we have the spectrum for binary on-off keying of a carrier at f_i as given by

$$S_{f_i}(f) = \frac{1}{2} \cdot \frac{A_c^2}{4T} [|P(f + f_i)|^2 + |P(f - f_i)|^2]$$

where the first 1/2 factor is the duty cycle associated with equally likely 1's and 0's and as before $P(f) = T \text{sinc}(fT)$

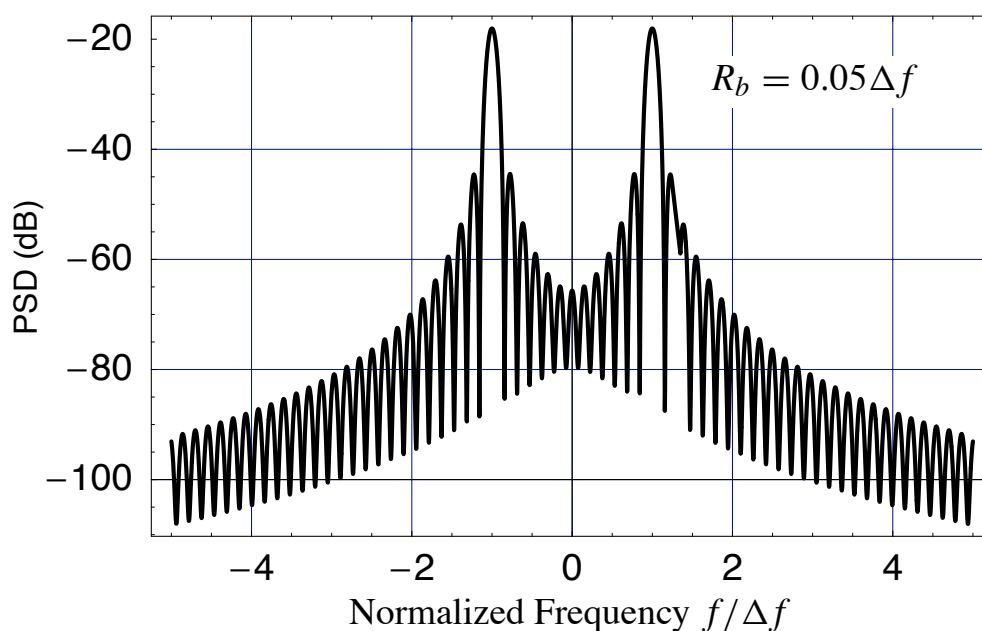
- Since we have two carrier frequencies in BFSK, superposition yields the approximate BFSK spectrum

$$S_t(f) \simeq \frac{A_c^2}{8T} [|P(f + f_1)|^2 + |P(f - f_1)|^2 + |P(f + f_2)|^2 + |P(f - f_2)|^2]$$

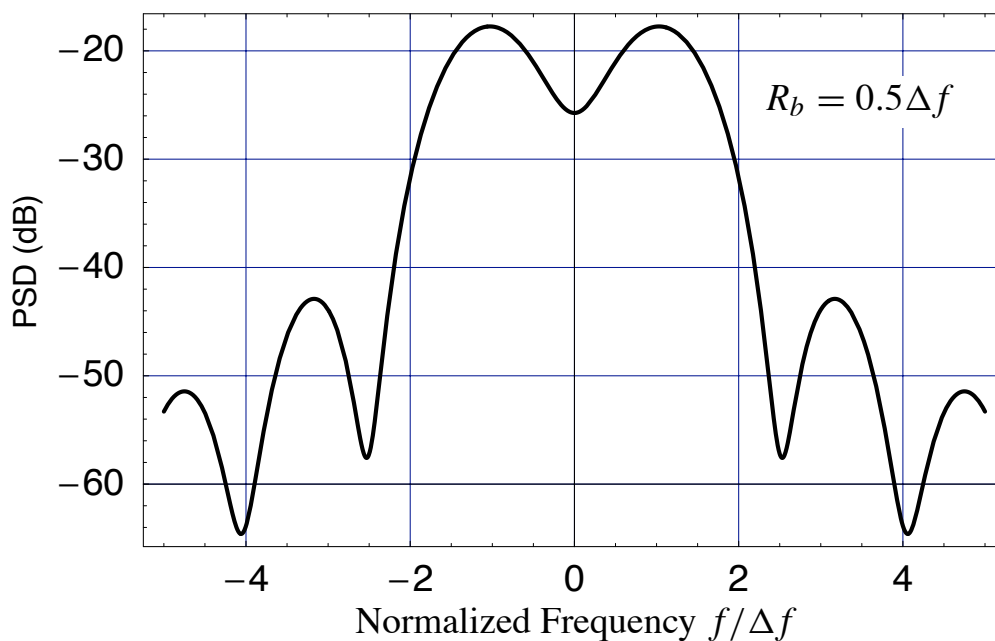
- This approximation is best when $|f_2 - f_1|$ is large compared with the bit rate R_b
- At complex baseband the power spectrum is of the form

$$S_B(f) \simeq \frac{A_c^2}{4T} [|P(f + \Delta f)|^2 + |P(f - \Delta f)|^2]$$

where $\Delta f = |f_2 - f_1|/2$



Baseband BFSK power spectrum in dB for $R_b = 0.05\Delta f$



Baseband BFSK power spectrum in dB for $R_b = 0.5\Delta f$

5.6.2 Minimum Shift Keying

- Recall that with CPFSK the carrier phase is continuous from bit-to-bit.
- Rather than using two frequencies f_1 and f_2 we may write $s(t)$ in terms of f_c and phase $\theta(t)$

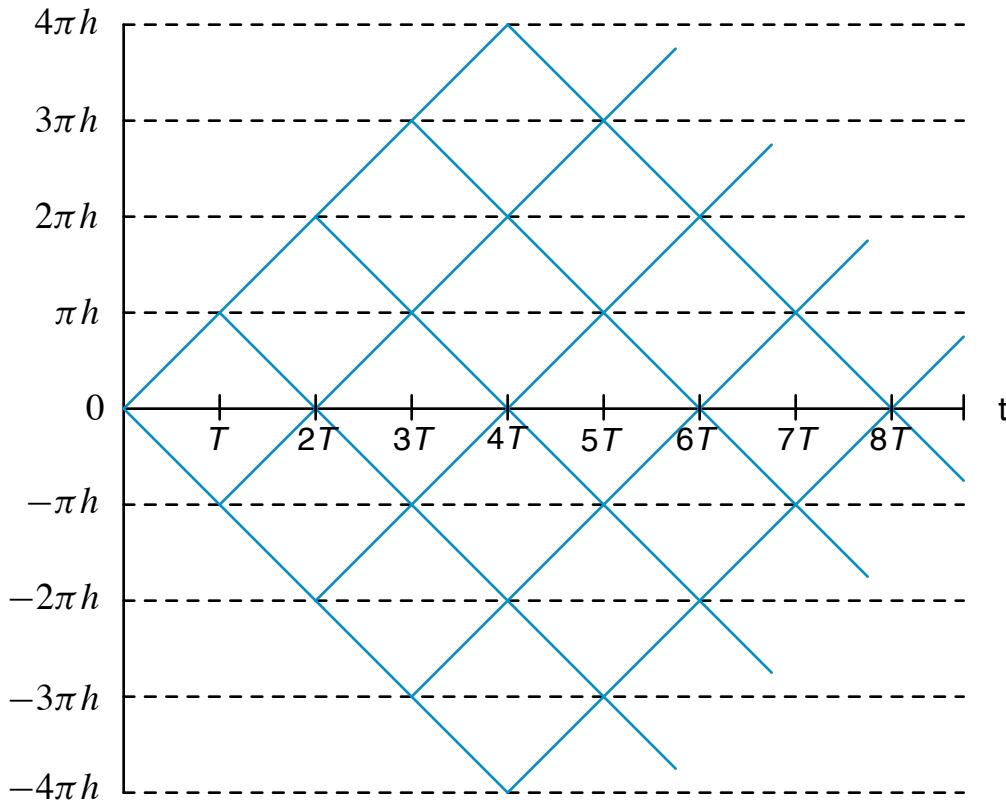
$$s(t) = \sqrt{\frac{2E_b}{T}} \cos [2\pi f_c t + \theta(t)]$$

- With CPFSK the phase linearly ramps up or down during each bit time as

$$\theta(t) = \theta(0) \pm \frac{\pi h}{T}, \quad 0 \leq t \leq T$$

where h is the CPFSK *modulation index* or *deviation ratio* and $\theta(0)$ is the accumulated excess carrier phase up to time $t = 0$

- A *phase trellis* can be constructed by noting the possible phase trajectories of $\theta(t) - \theta(0)$ for $t \geq 0$

CPFSK phase trellis for arbitrary h

- For the special case of $h = 1/2$ we have what is known as minimum-shift keying (MSK)
- The term minimum is used because when $h = 1/2$ the frequency deviation between the FSK tones is given by

$$\begin{aligned} f_1 - f_2 &= \left[f_c + \frac{h}{2T} \right] - \left[f_c - \frac{h}{2T} \right] \\ &= \frac{h}{T} = \frac{1}{2T} = \frac{R_b}{2} \end{aligned}$$

- We see that setting the frequency difference between two FSK tones to $R_b/2$ ($h = 1/2$) is the minimum tone that allows the two tones to remain orthogonal and create the minimum spectral bandwidth

- The signal space basis functions for MSK are typically expressed as

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

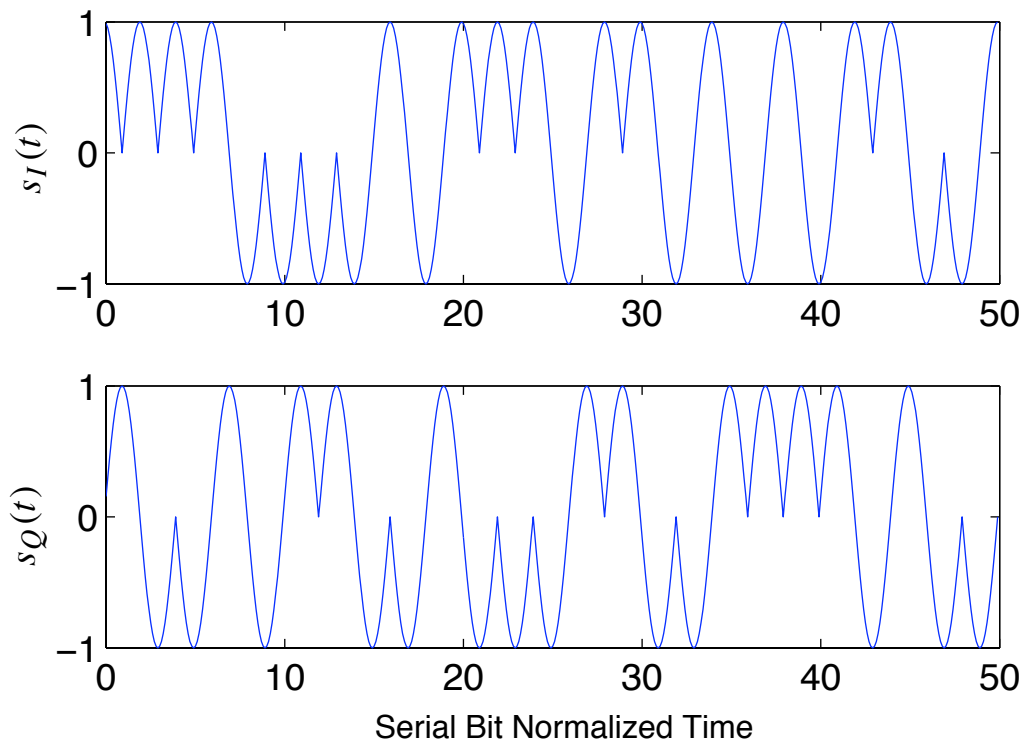
$$\phi_2(t) = \sqrt{\frac{2}{T}} \sin\left(\frac{\pi}{2T}t\right) \sin(2\pi f_c t)$$

then it can be shown that

$$s(t) = s_1\phi_1(t) + s_2\phi_2(t)$$

where s_1 and s_2 represent parallel data bits modulating I-Q carriers

- This expansion not only shows us that MSK can be generated using a standard I-Q modulator using half-sine pulse shaping
- It is also clear that the IQ phase trajectories follow the circumference of a circle traversing $\pm 90^\circ$ each serial bit period T



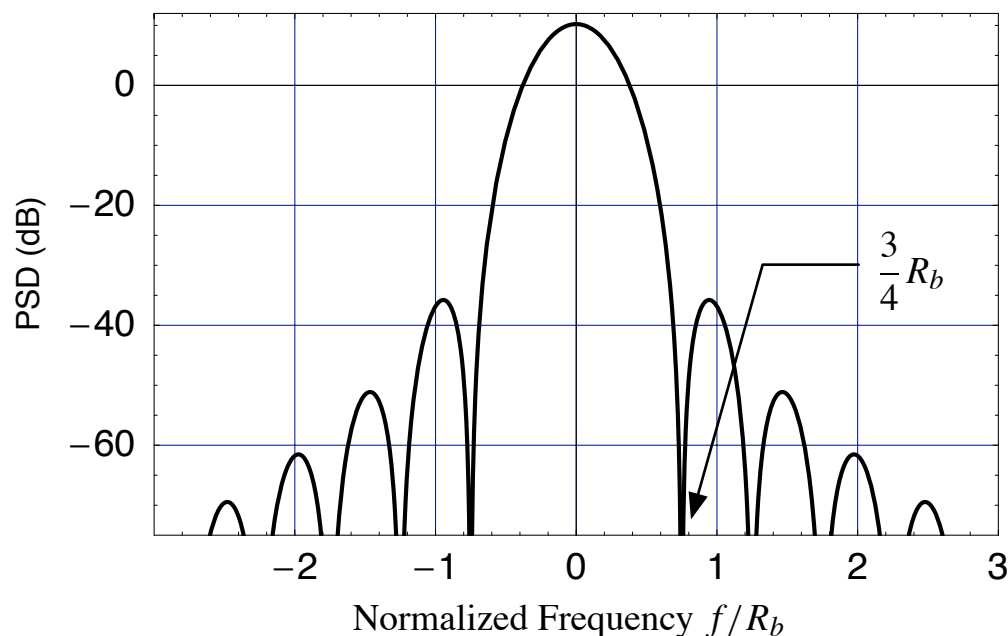
MSK complex baseband IQ waveforms

MSK Power Spectrum

The complex baseband power spectral density of MSK can be shown to be

$$S_B(f) = \frac{32E_b}{\pi^2} \left[\frac{\cos(2\pi T f)}{16T^2 f^2 - 1} \right]^2$$

where T_b is the serial bit duration as opposed to the parallel symbol duration which is $2T_b$.



MSK baseband power spectrum

MSK Properties

- Constant envelope
- Relatively narrow bandwidth
 - Wide main lobe than QPSK
 - Faster sidelobe rolloff rate than QPSK
- Coherent receiver performance equivalent to QPSK (more on this later)

Gaussian MSK

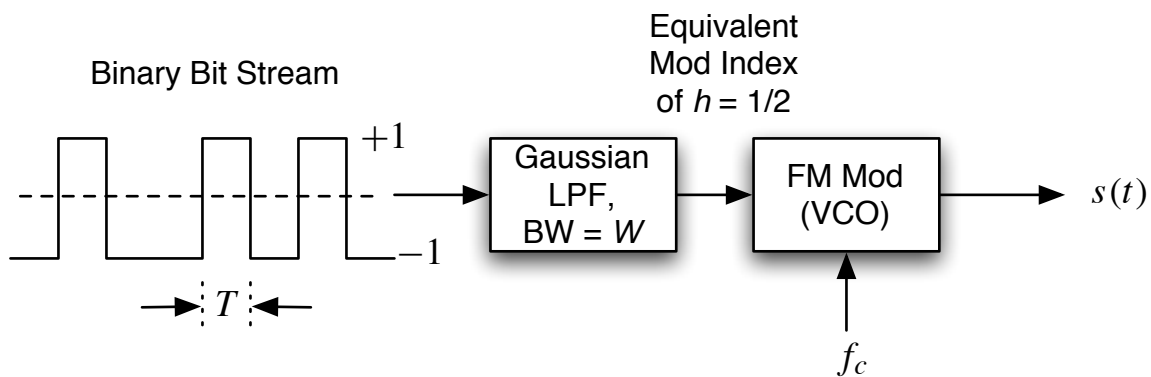
The spectrum of MSK is more compact than rectangular pulse shape BPSK/QPSK, but is still not suitable for multiple access communications.

- Another feature of MSK is that it can be generated by direct frequency modulation using a voltage controlled oscillator (VCO)
- In particular it is possible to apply prefiltering to a rectangular pulse shaped binary message stream
- With Gaussian MSK (GMSK) this shaping filter has a Gaussian impulse response and a Gaussian frequency response

$$h(t) = \sqrt{\frac{2\pi}{\ln 2}} W \exp \left[-\frac{2\pi^2}{\ln 2} W^2 t^2 \right]$$

$$H(f) = \exp \left[-\frac{\ln 2}{2} \left(\frac{f}{W} \right)^2 \right]$$

where W is the filter 3 dB bandwidth



GMSK modulator

- When this premodulation shaping filter is driven a rectangular pulse of duration T , the effective frequency pulse shaping

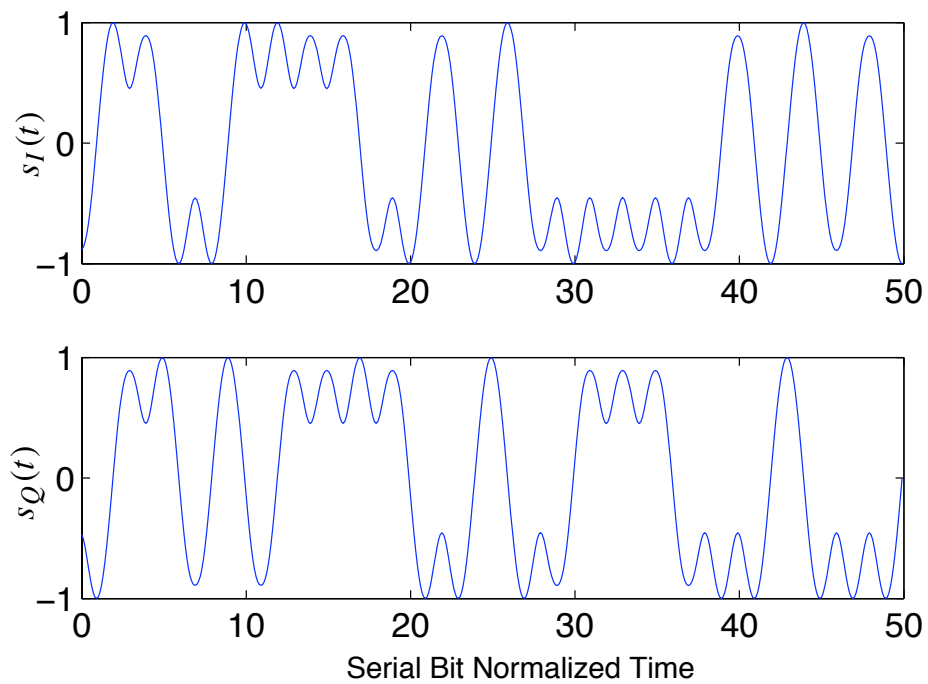
driving the FM modulator is of the form

$$g(t) = \frac{1}{2} \left\{ \operatorname{erfc} \left[\pi \sqrt{\frac{2\pi}{\ln 2}} W T \left(\frac{t}{T} - \frac{1}{2} \right) \right] - \operatorname{erfc} \left[\pi \sqrt{\frac{2\pi}{\ln 2}} W T \left(\frac{t}{T} + \frac{1}{2} \right) \right] \right\}$$

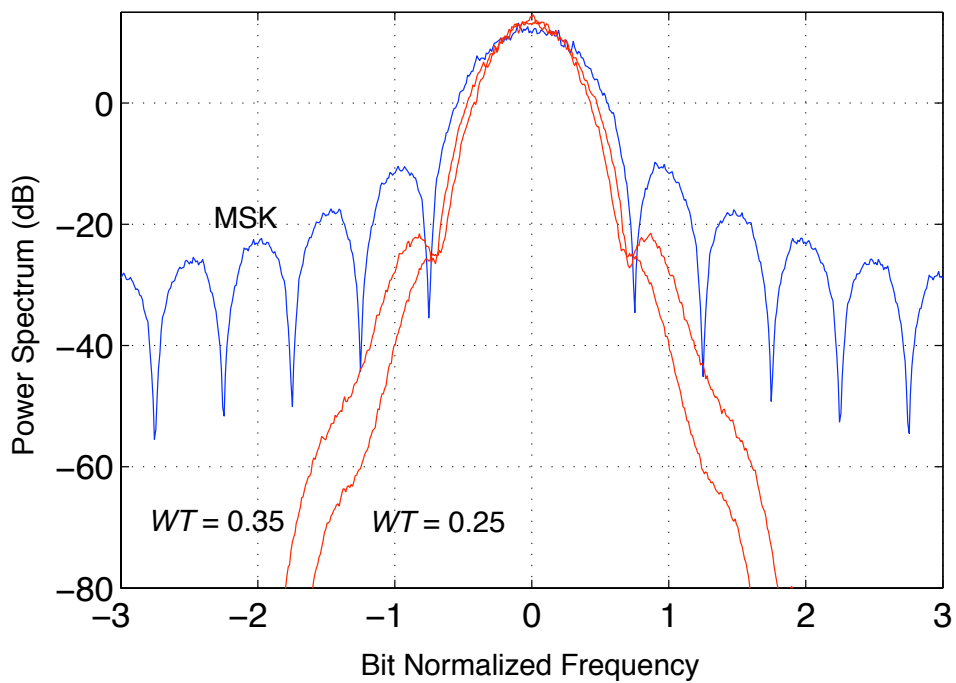
where $\operatorname{erfc}(x)$ is the complementary error function defined as

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-z^2) dz$$

- The design parameter controlling the spectral bandwidth of GMSK is the time-bandwidth product WT (more commonly BT)
- Note that $WT \rightarrow \infty$ (no filtering) results in MSK
- There is no closed-form expression for the power spectral density of GMSK, so we have to resort to simulation



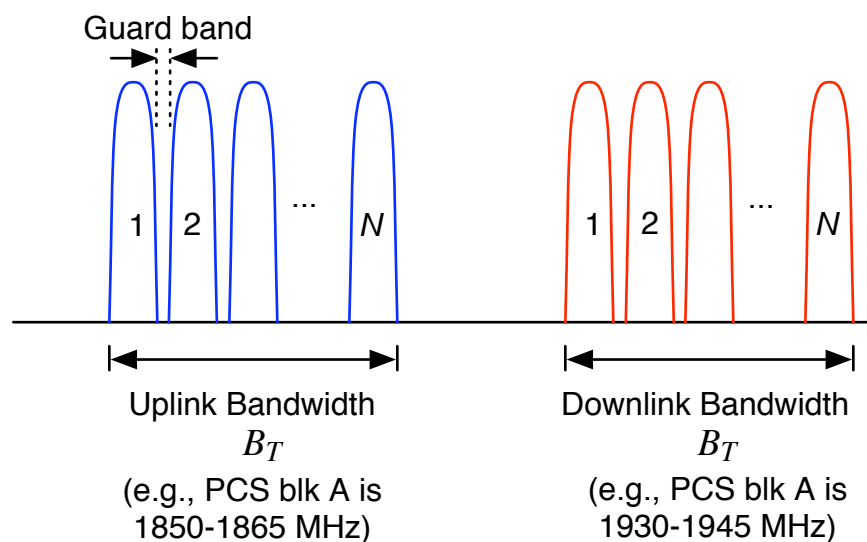
GMSK complex baseband IQ waveforms m



GMSK baseband power spectrum

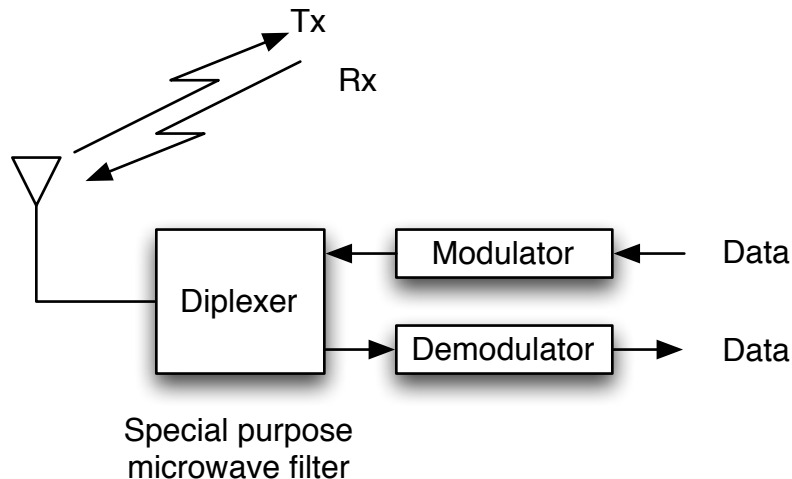
5.7 FDMA

Frequency division multiple access (FDMA) was briefly described in Chapter 1 as means to allow multiple users to communicate simultaneously. The *frequency division duplex* (FDD) scheme used with the PCS bands is shown below.



FDD/FDMA channel allocations as used in the PCS band

- Guard bands are provided as *buffer zones*, so the power radiated outside the assigned band stays between 60–80 dB below the in-band signal
- Common terminology is to refer to the base station to mobile link as the *forward-link* or downlink; while the mobile to base station link is referred to as the *reverse-link* or uplink
- When FDD is employed uplink and down link communications can occur simultaneously, meaning that the radio hardware needs a *diplexer* to separate the two signal paths



Diplexer used in an FDD enabled radio

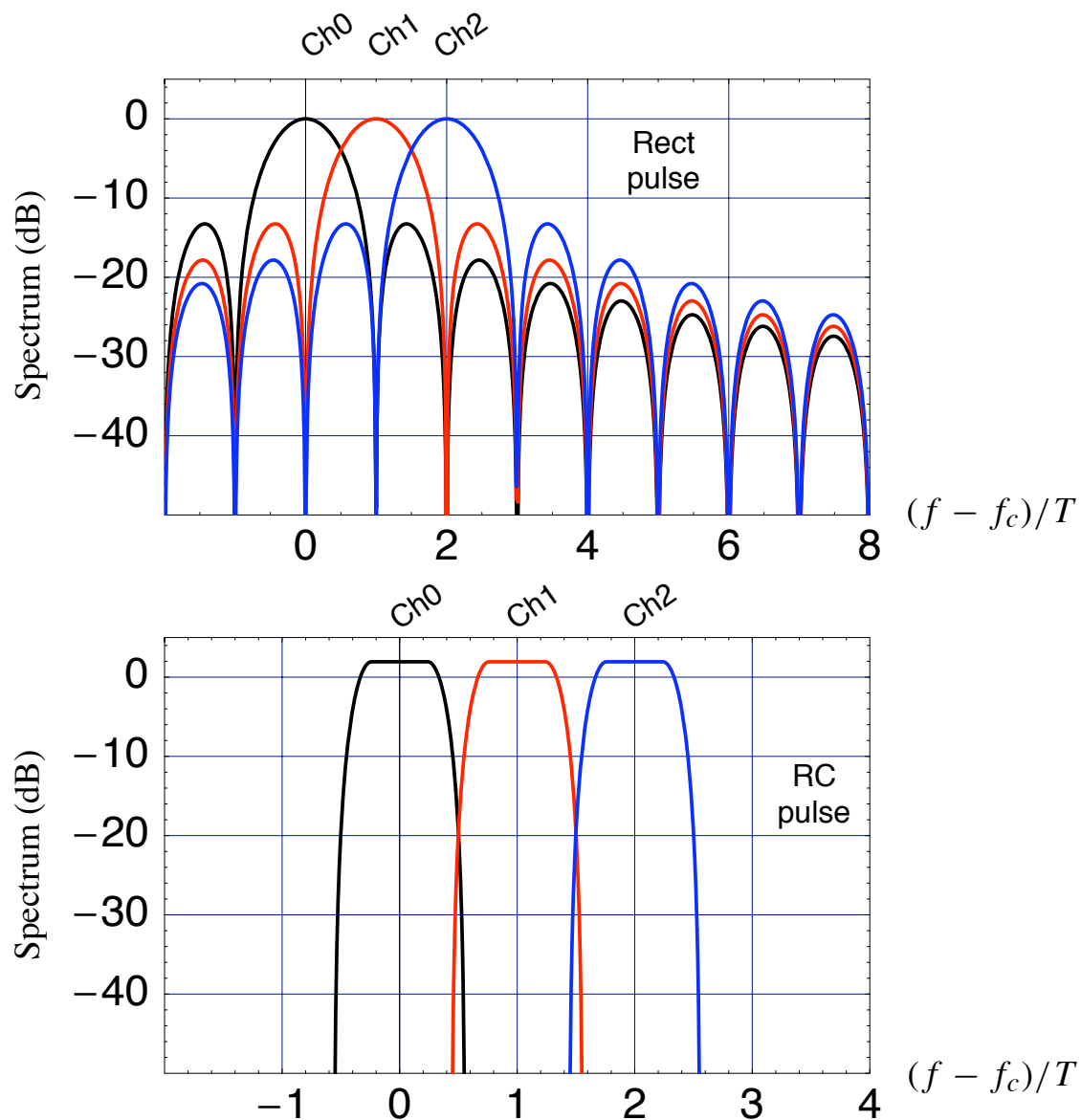
- One downside of FDMA is that the base station must have a dedicated transmitter and receiver for each carrier frequency in the downlink and uplink bandwidths, respectively
- Since the bandwidth assigned to each channel is narrow, the fading encountered tends to be flat
- Doppler spreading is still an issue however

5.7.1 Adjacent Channel Interference

One serious threat to FDMA is adjacent channel interference, that is the performance degradation resulting from spectral energy from adjacent channels leaking into the desired channel.

- Earlier we talked about bandwidth efficiency of PSK modulation, in bits/s/Hz
- A higher spectral efficiency means that more channels can be packed into bandwidth B_T

- Adjacent channel interference (ACI) is particularly a problem for non-bandlimited signals, such as rectangular pulse shaped PSK

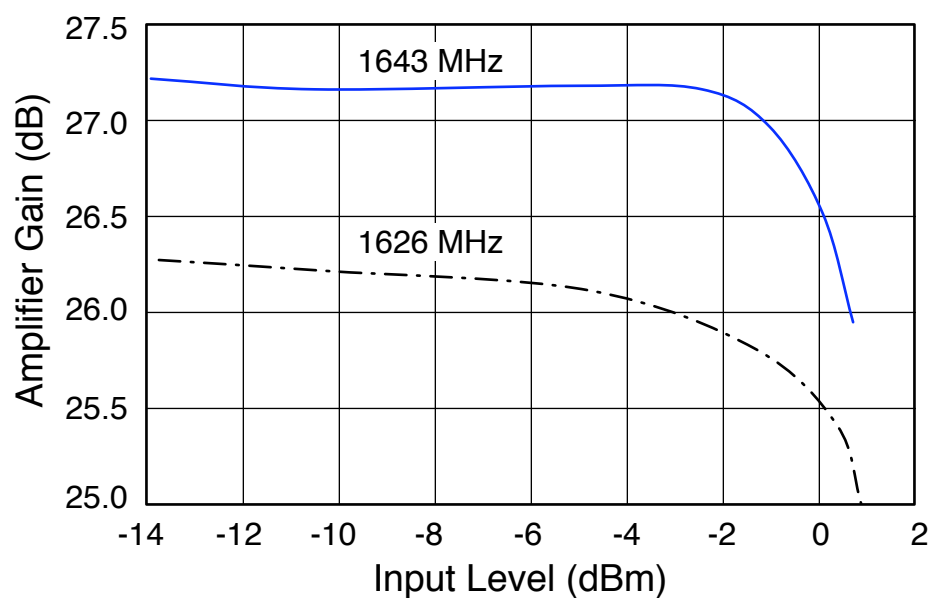


ACI comparison for Rect and RC signals

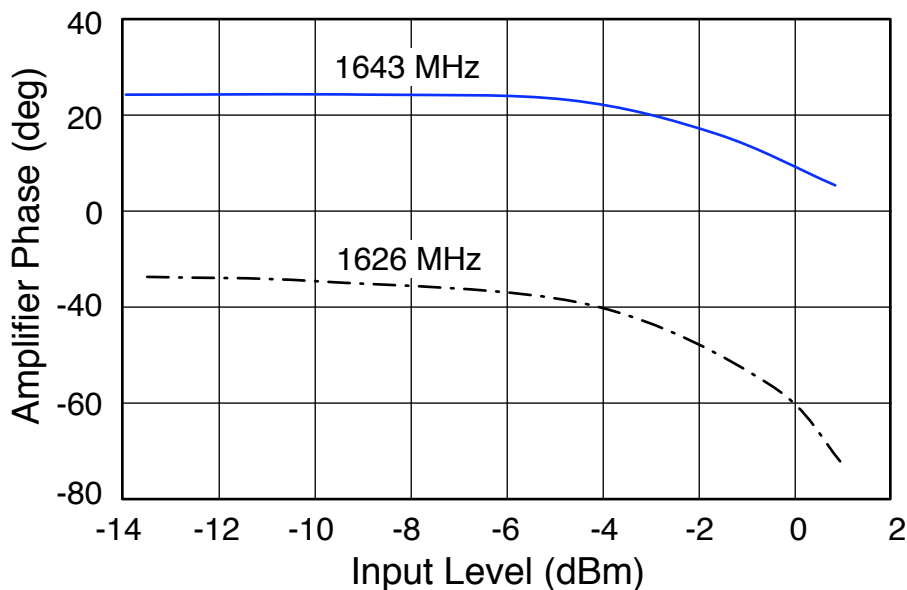
5.7.2 Power Amplifier Nonlinearity

ACI as just discussed, limits how close signals can be spaced. RF power amplifier nonlinearity also factors into ACI as spectral regrowth and intermodulation may occur.

- In portable electronics the efficiency of the RF power amplifier factors into battery lifetime
- Amplifiers that are more power efficient also tend to be more nonlinear
- There are means to linearize nonlinear amplifiers, and this is an active area of research and development topic
- A power amplifier will have both a non-constant gain and phase versus the input power level, that is also a function of frequency



Amplitude characteristics of a nonlinear amplifier



Phase characteristics of a nonlinear amplifier

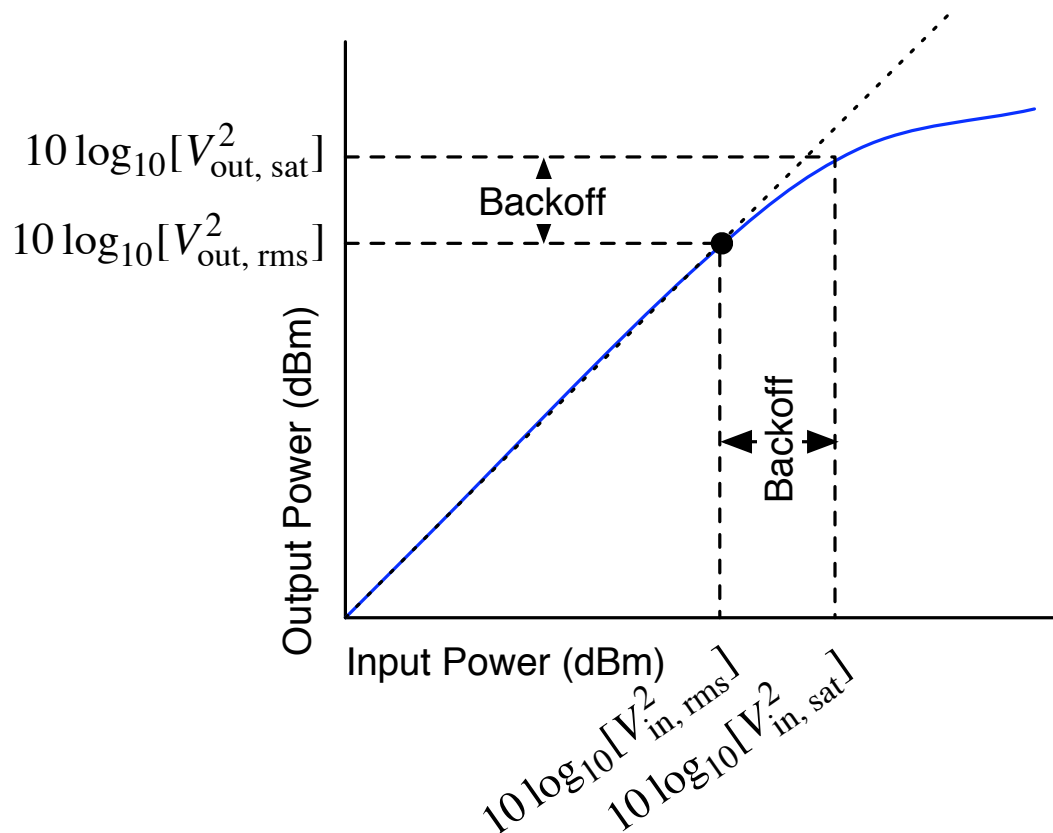
- In the above sample plots we see that when the input kept below about -8 dBm, both the amplitude and the phase remain constant
- For larger input levels we have both *AM-to-AM* and *AM-to-PM* distortion taking place
- By operating the amplifier below a particular input level, distortion can be minimized
- One measure of amplifier saturation is when the amplifier gain is reduced by 1 dB, the so-called *1dB compression point*
- If we define the input and output 1 dB compression points at $V_{in, sat}$ and $V_{out, sat}$ respectively, we can define the amplifier *in-*

put back-off and *output back-off* respectively as

$$\text{Input back-off} = 10 \log_{10} \left(\frac{V_{\text{in, rms}}}{V_{\text{in, sat}}} \right)^2$$

$$\text{Output back-off} = 10 \log_{10} \left(\frac{V_{\text{out, rms}}}{V_{\text{out, sat}}} \right)^2$$

- As we get closer to the 1dB compression point more distortion is introduced



Idealized amplifier AM-to-AM characteristic

5.8 Modulation Comparison

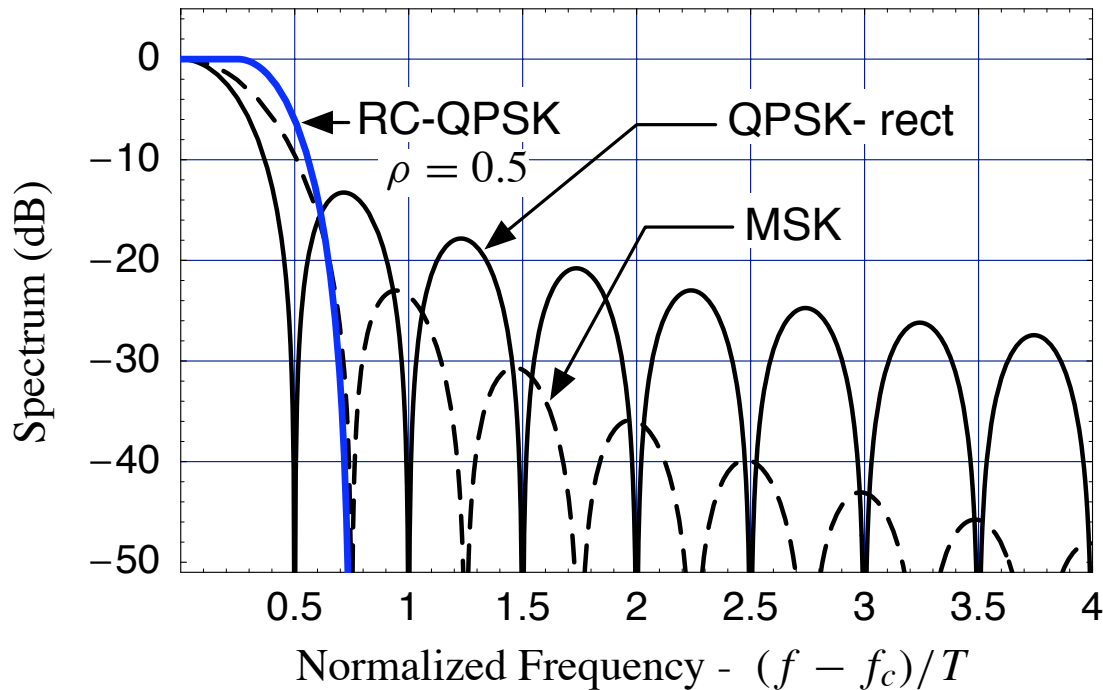
Modulation schemes can be compared on the basis of spectral efficiency and power efficiency. Spectral efficiency as we know is measures in bits/s/Hz. The channel context is important however.

- Three relevant factors include:
 1. Pulse shaping (rect, RC, SRC, Gaussian, etc.)
 2. Other filtering (Image rejection and amplifier spurious responses)
 3. Presence of nonlinearities (power amplifiers operating close to saturation)

5.8.1 Linear Channel

For a linear channel we are just interested in the ideal transmitted spectra, that is with only linear filtering at most.

- We have considered both linear and nonlinear modulation schemes
- Rectangular pulse shaping is known to be inefficient due to the slow sidelobe roll-off rate
- MSK, with the half-sine pulse shape is better, but still has a wide bandwidth
- GMSK is much better than MSK for spectral efficiency, but power efficiency is not as good
- QPSK with RC or SRC shaping is very efficient, and is good choice for many applications



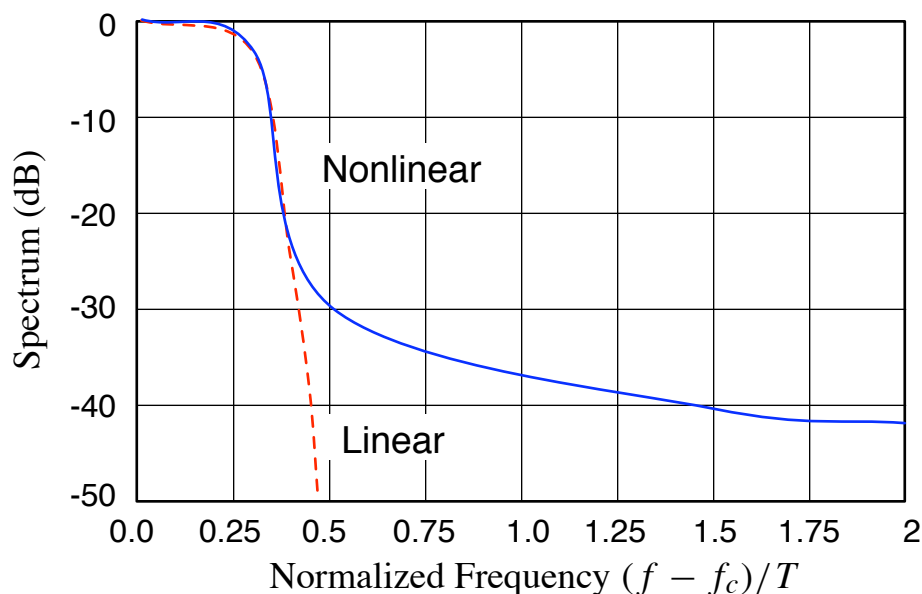
Comparison spectra: QPSK, MSK, RC-QPSK with $\rho = 0.5$

5.8.2 Nonlinear Channel

A nonlinear transmit power amplifier can effectively destroy the spectral bandlimiting achieved through linear filtering. Spectral sidelobes tend to *regrow* depending upon how close the amplifier is to saturation.

- Constant envelope modulation schemes such as QPSK with rectangular pulse shaping, MSK, and GMSK, are in theory unaffected by a nonlinear amplifier
- The sidelobe level can be kept 40 to 50 dB below the main lobe level
- When RC and SRC pulse shaping is employed envelope variations are introduced in QPSK, OQPSK, and $\pi/4$ -DQPSK

- With a nonlinear amplifier near saturation, the sidelobe level can come up to within 30 dB of the main lobe, followed by a more gradual spectral rolloff



Approximate OQPSK spectrum with 1 dB backoff

5.9 Channel Estimation and Tracking

- PSK based modulation requires a phase reference at the receiver to properly recover the message bits
- When slow fading is present recall that the channel introduces phase variations that are small relative to the modulation induced phase shifts
- There several receiver design options when it comes to phase tracking
 1. Attempt to coherently track the channel induced phase variations using a carrier phase recovery algorithm

2. Implement differential detection
 3. Implement *pilot symbol assisted modulation* (PSAM)
- Coherent detection may not always be possible, or may be too complex to implement in a low-cost receiver design

5.9.1 Differential Detection

- Differential detection relies on the fact that the carrier phase changes little from one symbol to the next, so that the previous symbol can be used to demodulate the present symbol
- Consider a transmitted signal of the form

$$s(t) = A \operatorname{Re} \{ d(t) e^{j2\pi f_c t} \}$$

where A is the carrier amplitude, f_c is the carrier frequency, and $d(t)$ is the data modulation

- At the receiver the carrier frequency will not be known exactly, so in complex baseband form we receive

$$\tilde{x}(t) = A' d(t) e^{j(2\pi \Delta f t + \phi)} + \tilde{w}(t)$$

where A' is the received signal amplitude, δf is the residual frequency error (instabilities and Doppler), and $\tilde{w}(t)$ is complex white Gaussian channel noise

- We further assume that

$$d(t) = \sum_k b_k p(t - kT)$$

where b_k are data symbols (possibly complex) and $p(t)$ is the transmitted pulse shape

- The received signal is first matched filtered to obtain improved SNR

$$y_k = \int_{nT}^{(n+1)T} p^*(t - kT) \tilde{x}(t) dt$$

- Assuming a pulse shape having unit energy, the matched filter output is approximately

$$y_k \simeq A' b_k e^{j(k2\pi\Delta f T + \phi)} + w_k$$

- We assume that the data stream was differentially decoded at the transmitter as

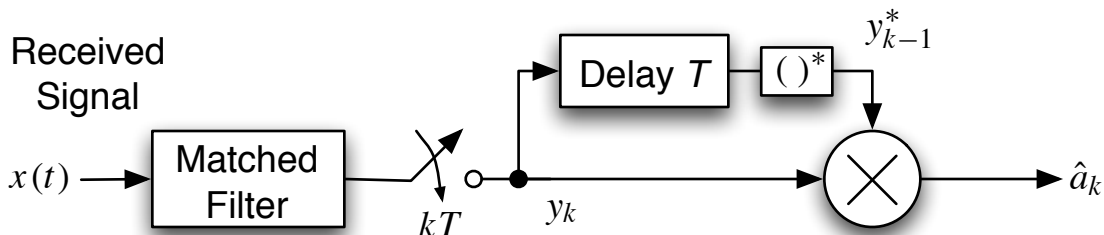
$$b_k = a_k b_{k-1}$$

where a_k is the original input data stream

- The differential detector performs a delay-and-multiply operation as follows:

$$\begin{aligned} \hat{a}_k &= y_k y_{k-1}^* \\ &= [A' b_k e^{j(k2\pi\Delta f T + \phi)} + w_k] \cdot \\ &\quad [A' b_{k-1} e^{j((k-1)2\pi\Delta f T + \phi)} + w_{k-1}]^* \\ &= b_k b_{k-1}^* e^{j2\pi\Delta f T} + \eta_k \\ &\approx a_k \eta_k \end{aligned}$$

where in the last line we have assumed that $\exp(j2\pi\Delta f T) \simeq 1$



Differential detection based receiver

- So long as the information is differentially encoded, this phase difference demodulation will work
- There is a performance penalty over fully coherent demodulation, since we form the product $y_k y_{k-1}^*$ to make symbol decisions

5.9.2 Pilot Transmission

PSAM is more complex than differential detection, but not only can the carrier phase be tracked, the channel state can also be estimated. For this to function *known pilot symbols* are inserted in the transmit data stream at regular intervals.

- Assuming the channel exhibits a fading pattern in complex baseband of the form $\tilde{\alpha}(t)$, we can write the received signal as

$$\tilde{x}(t) = \tilde{\alpha}(t)d(t) + \tilde{w}(t)$$

- Assuming Nyquist pulse shaping samples of the matched output are of the form

$$y_k = \alpha_k b_k + w_k$$

– We assume that the fading is constant over the pulse length

- Suppose the pilot symbols are known at the receiver at times $k = Ki$, where K is the pilot symbol spacing and i is an index variable

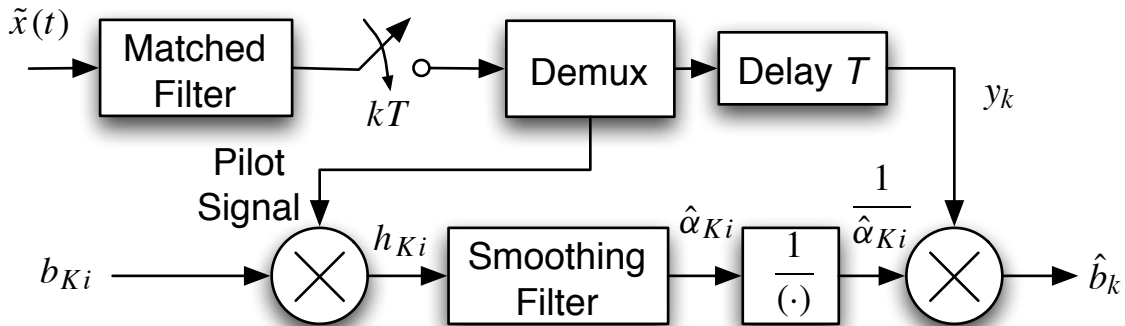
- By correlating the received pilot symbols against the known transmitted pilot symbols we can recover approximate values of the unknown fading channel

$$\begin{aligned}
 k_{Ki} &= b_{Ki}^* y_{Ki} \\
 &= \alpha_{Ki} |b_{Ki}|^2 + b_{Ki}^* w_{Ki} \\
 &= \alpha_{Ki} + w_{Ki}
 \end{aligned}$$

- Here we have assumed that the data is BPSK, i.e., $|b_{Ki}|^2 = 1$
- To obtain a better estimate of α_{Ki} a *linear minimum mean-squared error* (LMMSE) estimate is formed using multiple values of $h_{K(i+m)}$ in the neighborhood of Ki , i.e.,

$$\hat{\alpha}_{Ki} = \sum_{m=-L}^L a_m h_{K(i+m)}$$

- The Haykin text explains in detail how the LMMSE coefficient set $\{a_m\}$ is obtained
- A practical implementation is shown below



PSAM receiver for tracking channel variations

- With system fading as well as residual frequency and phase errors can be tracked
- An obvious downside is overhead needed to transmit the pilot symbols

5.10 Receiver Bit Error Probability Performance

In this section we consider the bit error rate (BER) or bit error probability (BEP) of some of the modulation schemes discussed thus far. A simple additive white Gaussian noise (AWGN) channel is considered first followed by a frequency flat, slow fading channel.

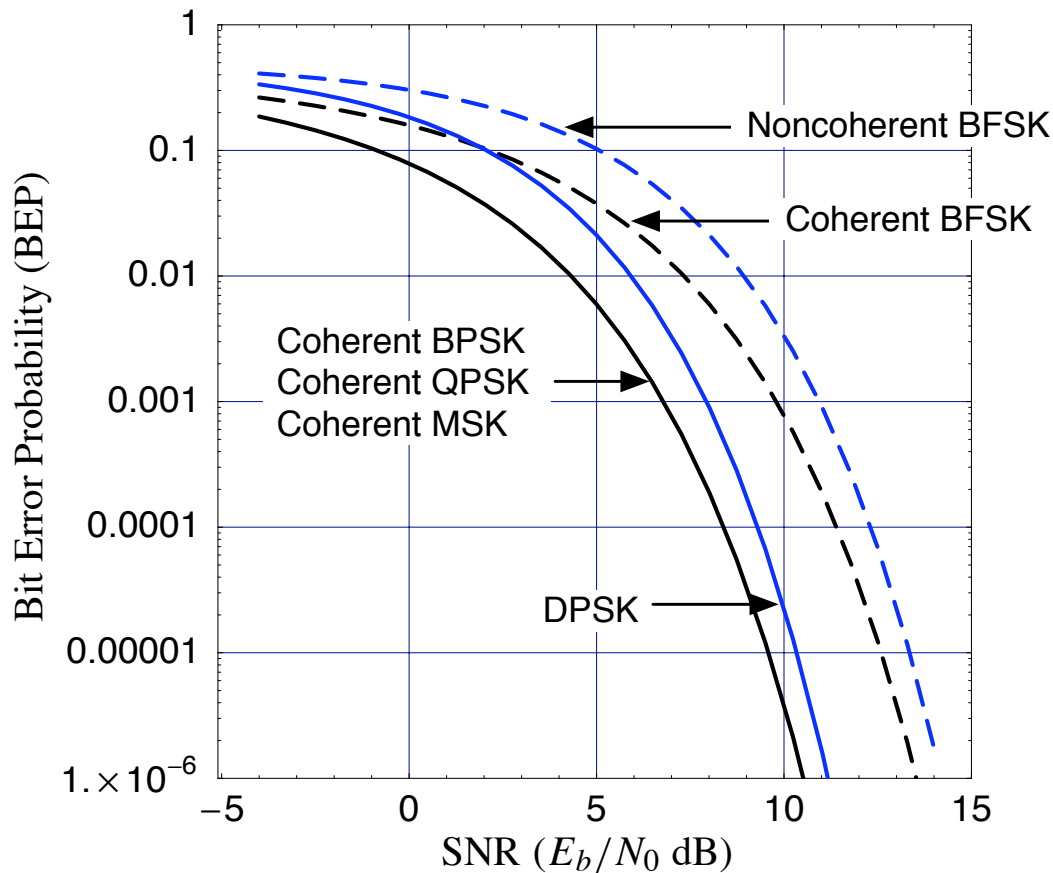
Bit error probability formulas

Signaling Scheme	AWGN	Slow Rayleigh
(a) Coh. BPSK, QPSK, MSK	$\frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$	$\frac{1}{2}\left[1 - \sqrt{\frac{\gamma_0}{1+\gamma_0}}\right]$
(b) Coh. BFSK	$\frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{2N_0}}\right)$	$\frac{1}{2}\left[1 - \sqrt{\frac{\gamma_0}{2+\gamma_0}}\right]$
(c) Binary DPSK	$\frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right)$	$\frac{1}{2(1+\gamma_0)}$
(d) NC BFSK	$\frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right)$	$\frac{1}{2+\gamma_0}$

where E_b is the energy transmitted per bit, N_0 is the one-sided WGN noise power spectral density, and γ_0 is the mean received value of E_b/N_0 under Rayleigh fading.

5.10.1 AWGN Channel

- The above AWGN expressions are exact for the indicated modulation schemes under coherent detection, differentially coherent, and noncoherent detection
 - Coherent detection means a locally generated carrier is obtained via synchronization of both residual phase and frequency; this increases receiver complexity and may not always be practical in severe fading environments
 - Differentially coherent detection means that the previous symbol is used to demodulate the present, without the need for a true locally generated carrier
 - Noncoherent means that a scheme such as frequency discriminator is employed, again a separate locally generated carrier is not obtained
- Exact closed-form BEP solutions are not available for GMSK, MPSK with $M > 4$, QAM, and MFSK. Bound can be calculated and simulations can of course be performed



BEP for an AWGN channel

- Note that in all cases the error probability decreases monotonically with increasing E_b/N_0 ; *waterfall curves*
- Coherent BPSK/QPSK/MSK has the best performance
- Coherent BFSK is 3 dB inferior to BPSK
- DPSK is less than 3 dB inferior to BPSK, with the value approaching 1 dB for high E_b/N_0
- Noncoherent BFSK is within 1 dB coherent BFSK at high E_b/N_0

GMSK

GMSK is more practical than MSK, yet its performance cannot be calculated exactly. A simple approximation for the BEP of GMSK is

$$P_e = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{\beta E_b}{2N_0}} \right)$$

where β is a constant depending upon the time bandwidth product WT .

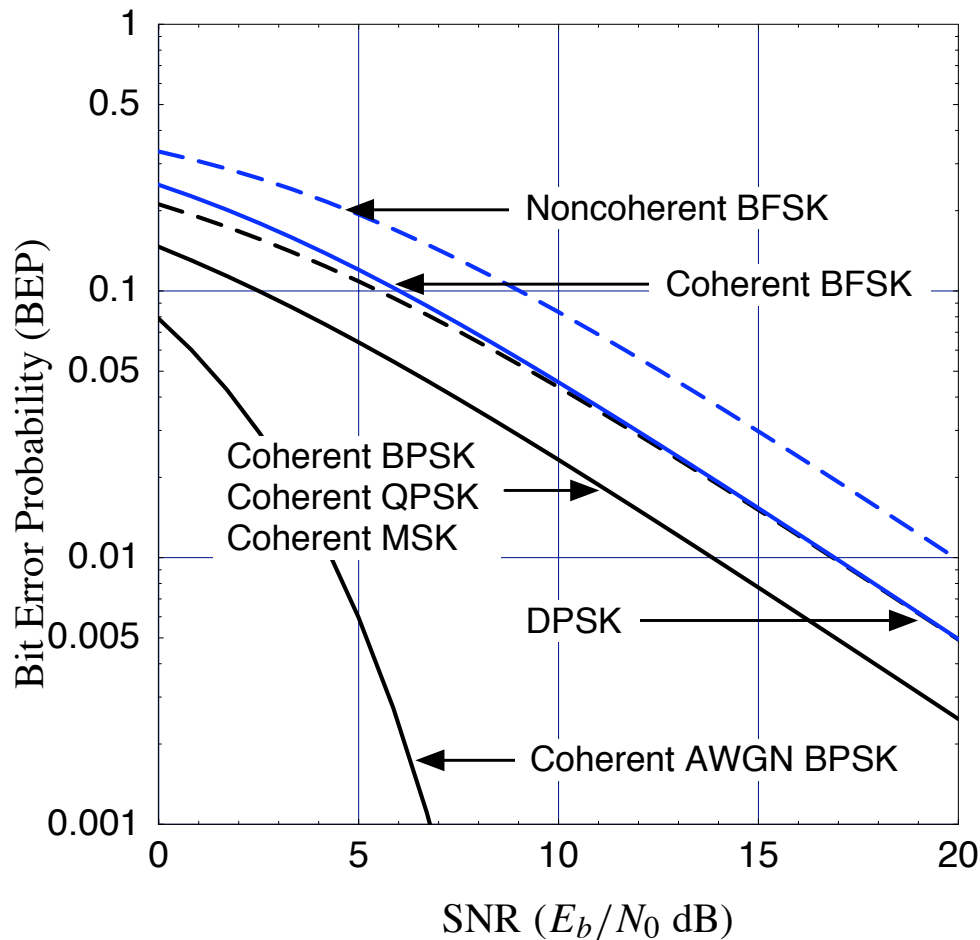
- Note that when $\beta = 2$ we have coherent MSK
- For $WT = 0.3$ the degradation is 0.46 dB which corresponds to $\beta/2 = 0.9$

5.10.2 Frequency Flat, Slow fading Channel

When fading is present we characterize the BEP performance in terms of the mean value of received E_b/N_0 given by

$$\gamma_0 = \frac{E_b}{N_0} E[\alpha^2]$$

where γ is Rayleigh random variable characterizing the envelope fading statistics.



BEP for a Rayleigh flat fading channel

- The BEP expression in the above table are radically different from the AWGN case
- The exponential-law expressions become algebraic making the waterfall much more gradual compared with the AWGN case
- A significantly larger mean SNR is required to obtain the same BEP, in fact up to 20 dB the BEP has not dropped below 10^{-3}
- The Rayleigh fading channel poses a serious challenge; more will be said about this in later chapters

5.11 Theme Example: OFDM

A modulation scheme that is becoming popular for fixed wireless data services is *orthogonal frequency division multiplexing* (OFDM). In particular the WLAN standard IEEE 802.11a uses an OFDM scheme to achieve a serial bit rate of up to 54 Mbps. The RF transmission bandwidth of 802.11a is constrained to be less than 20 MHz. One subrate of the standard sends user information at 36 MB/s, but forward error correction and coding increases this up to 48 Mb/s. This section will discuss some of the details of this rate standard.

- To pack 48 Mb/s in a 20 MHz wide RF channel requires better than 2 bits/s/Hz of bandwidth efficiency
- In this case the M -ary scheme 16-QAM is employed since it offers 4 bits/symbol
- Normally with 16-QAM, each transmitted complex baseband symbol is of the form

$$\tilde{s}(t) = b_k p(t - kT_d), \quad (k - 1)T_d \leq t < kT_d$$

where here $p(t)$ is a the pulse shape and T_d is the symbol period

- In 802.11a rather than sending the information over a single carrier, 48 orthogonal carriers are used to send $\{b_k\}$ demultiplexed into 48 parallel streams
- The symbol duration of each of these streams is $T = 48T_d$ and a pulse shape denoted, $g(t)$, applied to each stream is 48 times

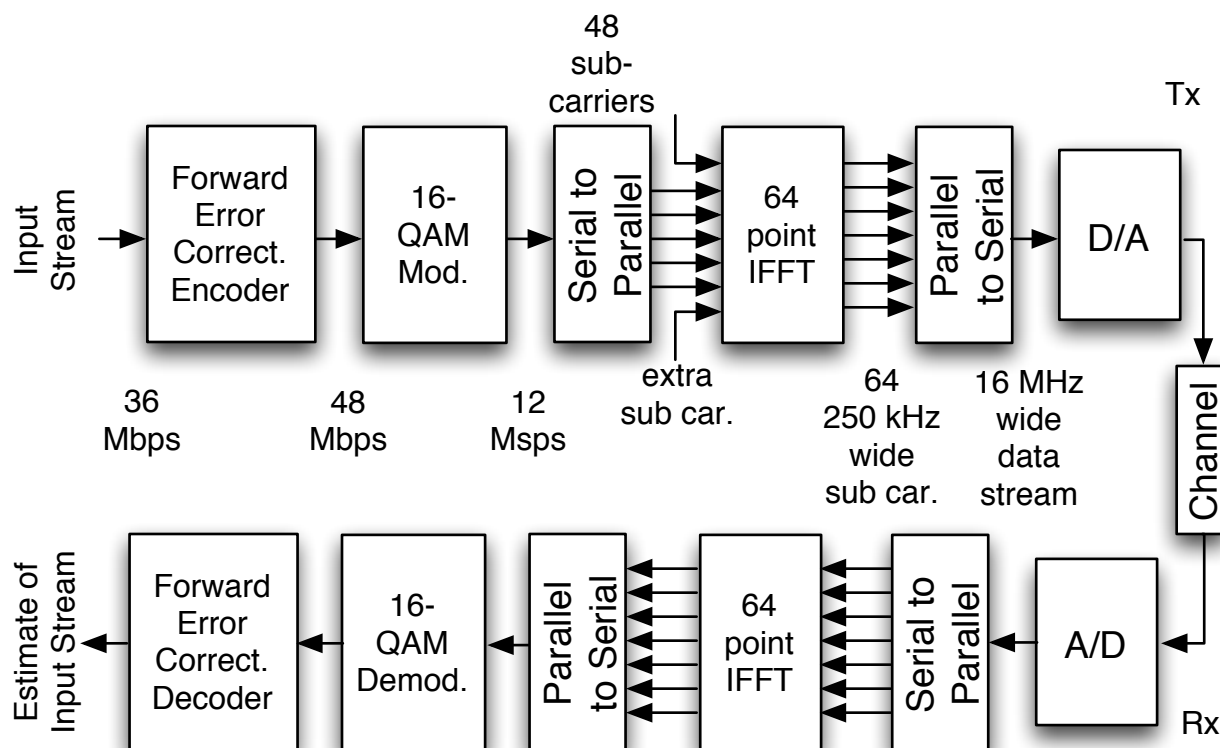
linger than $p(t)$

$$\tilde{s}_i(t) = b_{k,i} g(t - kT) e^{j2\pi f_i t}, \quad (k-1)T \leq t < kT, \quad i = 1, 2, \dots, 48$$

- The complete complex baseband transmitted signal is of the form

$$\tilde{s}(t) = \sum_{i=1}^{48} \tilde{s}_i(t)$$

- The orthogonal carriers are established through the use of the inverse fast Fourier transform (IFFT) (*formally the inverse discrete Fourier transform (IDFT)*)
- There is a corresponding *discrete Fourier transform (DFT)* (the fast version is the FFT), so both the forward and reverse operations can be performed
- For the FFT/IFFT to be fast and efficient, the size or number of points used in the computation is often a power of two
- Since 48 is not a power of two, a 64-point IFFT is used in practice
- The additional subcarriers are used for synchronization and tracking operations; unused subcarriers can be set to zero
- The transceiver block diagram for the 48 Mb/s system is shown below



IEEE 802.11a OFDM transceiver block diagram

- Subcarrier orthogonality is established by choosing

$$f_i = \frac{1}{T} \quad i = 1, 2, \dots, 48$$

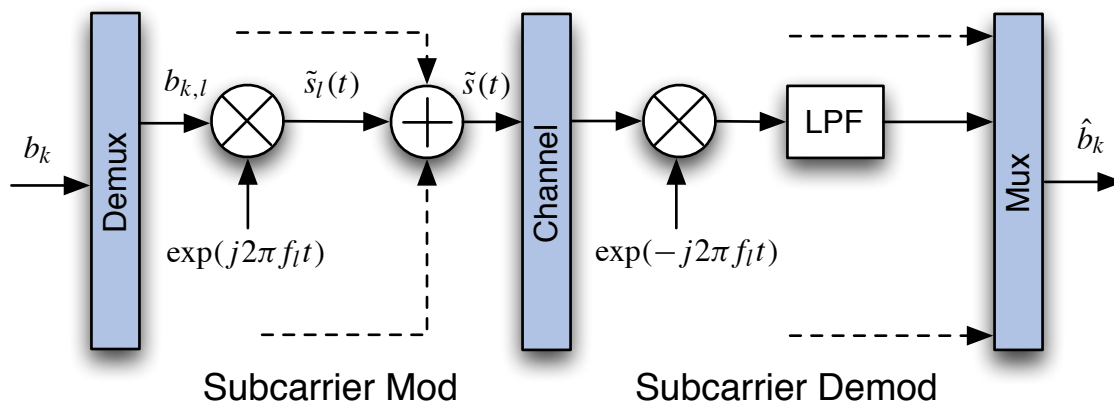
- Each subcarrier is effectively sampled at M samples per symbol of length T , thus the discrete-time version of $\tilde{s}(t)$ becomes

$$\tilde{s}(m) = \sum_{i=0}^{M-1} b_n e^{j2\pi mn/M}, \quad m = 0, 1, \dots, M-1$$

- Digital signal processing, particularly VLSI implementation makes all of this possible

5.11.1 Cyclic Prefix

- A special feature of OFDM is its ability to overcome ISI by the use of a *cyclic prefix*
- The cyclic prefix extends the duration of each OFDM symbol by a *guard time* corresponding to the maximum expected multipath delay
- By including the cyclic prefix the overall transmission bandwidth grows since a fraction of each symbol now contains a guard interval
- The receiver must remove the guard interval before the FFT processing

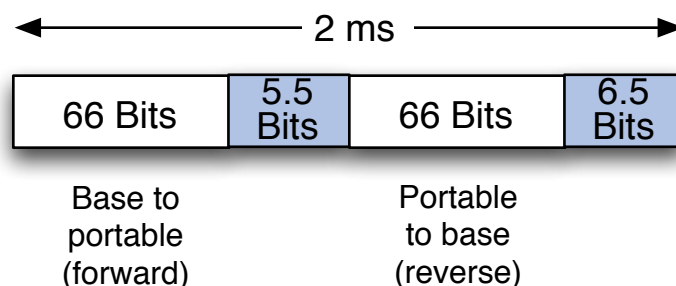


Simplified view of the OFDM modulation and demodulation process

5.12 Theme Example: Cordless Telephone

In Europe a cordless telephone standard known as CT-2 exists. 40 channels are defined over a 4 MHz bandwidth from 864.15 to 868.15 MHz. The multiple access scheme is FDMA using *time-division duplexing* (TDD) as opposed to FDD which was discussed earlier. Each channel occupies 100 kHz of bandwidth.

- With TDD the same frequency/channel is used for both transmit and receive
- In the case of CT-2 a form of GMSK, known as GFSK (Gaussian FSK), is used to provide a bandwidth efficient modulation
- The particular form of GFSK is M -ary so multiple bits are sent per symbol
- Speech is digitized using *adaptive pulse code modulation* (ADPCM) to a rate of 32 kbps
- Considering the TDD nature the overall bit rate is 72 kbps, which includes overhead for guard intervals between transmit and receive frames



CT-2 TDD frame structure

- The GFSK modulation scheme simplifies the transceiver design, and in particular allows for noncoherent reception
- *Power control*, a means to minimize interference between users, is implemented in CT-2 via a simple two level scheme
 - A power level of 5 mW is used as a default, unless the received signal power exceeds a particular level
 - When the signal is above threshold return path signaling tells the transmitter to reduce its power by about 15 dB

.