

ECE350

Peter Driessen

Fall 2020

Contents

1 IQ Signals	7
1.1 Carrier waves, amplitude and phase	7
1.1.1 Carrier wave	8
1.1.2 Modulated carrier wave with time-varying amplitude and phase	11
1.1.3 Complex envelope	14
1.1.4 Summary	19
1.1.5 Frequency shifting	20
1.1.6 Analytic signals	24
1.2 Block diagrams for real passband and complex baseband signals	25
1.2.1 Radio transmitter (modulator)	26
1.2.2 Radio receiver (demodulator)	28
1.2.3 Summary	31
1.2.4 Review	32
1.2.5 Messages	33
1.3 Analog IQ receiver	35
1.3.1 Overview and block diagram	35
1.3.2 USRP daughterboard receiver operation	36
1.3.3 Receiver operation in real notation	37
1.3.4 Receiver operation in complex notation	39
1.3.5 Another approach using digital IQ receiver	40
1.4 Spectrum Analyzers, analog and digital	40

<i>CONTENTS</i>	2
1.5 Spectrogram or spectral waterfall	43
1.5.1 Example spectrograms	43
1.6 IQ imbalance	48
1.7 Radio tuning, selecting a particular signal	49
2 Amplitude modulation (AM) transmitters	53
2.1 Amplitude modulation with full carrier	54
2.1.1 AM waveform with single tone message	55
2.1.2 AM spectrum	56
2.1.3 Power in carrier wave and sidebands	59
2.1.4 Overmodulation	60
2.1.5 AM signal with a general message	62
2.1.6 Digital Messages Transmitted Using AM	64
2.1.7 Generating an AM signal digitally in software	64
2.1.8 Analog-only method to generate an AM signal	65
2.2 Double Sideband Modulation	68
2.2.1 Double sideband suppressed carrier waveform	69
2.2.2 DSB spectrum	69
2.3 Single sideband suppressed carrier and the Hilbert transform	71
2.3.1 Derivation of SSB-SC using Hilbert transform and analytic signals	72
2.3.2 Hilbert transform definition and properties	73
2.3.3 SSB-SC signal derived from pre-envelope	74
2.3.4 Single tone modulation for SSB	75
2.3.5 SSB-SC in time and frequency domain	76
2.3.6 SSB-SC transmitter 1 using Hilbert transform	76
2.3.7 SSB-SC transmitter 2, avoids Hilbert transform	77
2.4 Review of AM, DSB, SSB transmitters in IQ format	80
2.4.1 Amplitude Modulation:	80
2.4.2 DSB-SC	80

<i>CONTENTS</i>	3
2.4.3 SSB-SC	81
3 Amplitude modulation (AM) receivers	83
3.1 Analog AM receiver	83
3.2 Digital software receivers	84
3.2.1 Digital software AM receiver in complex notation	84
3.2.2 Tuning in (selecting) one particular AM signal with the USRP	86
3.3 DSB-SC receiver	87
3.3.1 DSB receiver with frequency and phase offset	89
3.3.2 Review of I-Q receivers	90
3.3.3 Review of AM, DSB, SSB receivers in IQ format	90
3.4 Mapping from IQ receiver output to message signal	92
3.4.1 AM receiver	93
3.4.2 DSB-SC receiver with frequency error correction	93
3.4.3 SSB-SC receiver using message-bandwidth low pass filters	95
3.4.4 SSB-SC receiver using Weaver Demodulator description	97
3.4.5 SSB-SC Weaver demodulator operation in complex notation	98
3.4.6 SSB-SC Weaver demodulator operation in real notation	100
4 Super-heterodyne Receiver	102
5 Frequency Modulation	105
5.1 Overview	105
5.2 FM with General Message	106
5.3 FM with Sinusoidal Message	107
5.3.1 Narrowband FM with Sinusoidal Message	108
5.4 Power Spectrum of an FM Signal – Bessel functions	109
5.4.1 Observations about FM spectrum	112
5.4.2 Wideband FM large modulation index	112
5.4.3 Narrowband FM, small modulation index	113

5.4.4	113
5.5	Frequency modulators (transmitters)	114
5.6	Frequency Demodulation	116
5.6.1	Intuition 1: Differentiation	116
5.6.2	Intuition 2: Linear Amplitude vs. Frequency Characteristic	117
5.6.3	Intuition 3: Zero Crossing Counter	118
5.6.4	Intuition 4: Phase Locked Loop	119
5.7	Digital FM Demodulator	120
5.7.1	$I(t)$, $Q(t)$ as Real Signals	120
5.7.2	$I(t) + jQ(t)$ as a Complex Signal	120
5.7.3	Demodulator with frequency offset	121
5.8	Pre-emphasis and de-emphasis in FM	121
5.8.1	Pre- and de-emphasis circuits	121
5.8.2	De-emphasis with digital filter	123
5.9	Frequency and Phase Modulation review	127
5.9.1	Redundancy	129
5.9.2	NBFM Transmission	132
5.9.3	WBFM Transmission	133
5.9.4	Approximate Bandwidth of FM signal	134
5.9.5	FM Demodulation:	135
6	Baseband data transmission	139
6.1	Digital messages	139
6.2	Data source	140
6.3	Differential encoding and decoding	140
6.3.1	Avoiding polarity inversion	140
6.3.2	Differential encoding	141
6.3.3	Differential decoding	142
6.4	Data transmission system with pulse shaping	143

6.4.1	Representing data as square pulses in a sampled system	144
6.4.2	Pulse shaping to eliminate the sharp edges of a square pulse	144
6.4.3	Representing data sequence with Impulses instead of square pulses	149
6.4.4	Raised Cosine filter in frequency domain	151
6.4.5	Square Root Raised Cosine filter	153
6.4.6	Summary of pulse shaping methods	154
6.5	Eye diagrams	156
7	Phase shift keying	159
7.1	BPSK Transmitter	160
7.1.1	Real BPSK signal	163
7.1.2	Complex BPSK signal	164
7.1.3	BPSK at radio frequency (RF)	165
7.2	BPSK Receiver	166
7.2.1	Real BPSK signal receiver	166
7.2.2	Complex BPSK signal receiver	168
7.2.3	BPSK receiver at RF	168
7.3	QPSK Transmitter	169
7.3.1	Real QPSK signal	170
7.3.2	Complex QPSK signal	172
7.3.3	QPSK at radio frequency (RF)	173
7.4	QPSK Receiver	175
7.4.1	Real QPSK signal	175
7.4.2	Complex QPSK signal	177
7.4.3	QPSK signal at radio frequency (RF)	177
7.5	QPSK with amplitude, frequency and phase errors in the receiver	179
7.5.1	QPSK signal at particular time instant	179
7.5.2	QPSK receiver local oscillator with frequency error	180
7.5.3	QSPK receiver local oscillator with amplitude and phase errors	182

7.6	Differential BPSK	183
7.6.1	DBPSK with square pulse shape	183
7.6.2	D-BPSK with RC2 time domain pulse shape	185
7.6.3	DPSK with RRC frequency domain pulse	188
7.7	QPSK variants	188
7.7.1	Offset QPSK	188
7.7.2	$\pi/4$ QPSK	189
8	Frequency shift keying	191
8.1	FSK intuition	191
8.2	FSK complex envelope	192
8.2.1	FM equations	192
8.2.2	FSK phase ramps	193
8.2.3	FSK complex envelope	194
8.3	MSK and offset QPSK	195
A	Useful formulas	196

Chapter 1

IQ Signals

1.1 Carrier waves, amplitude and phase

Learning objectives

In this section, we learn the mathematics of a general communications signal waveform comprising a carrier wave whose amplitude and phase is modified (modulated) in step with a message signal. In this section, we describe how the communications signal may be written as the sum of two complex-valued waveforms resulting in a real-valued waveform.

Introduction

Telecommunications is defined by Webster's dictionary as "the science and technology of communications at a distance by electronic transmission".

A communications signal carries a message over a distance from point A to point B using a carrier wave.

Radio (or optical) waves are used to carry a message over a distance determined by the link budget. The radio waves travel through a channel which may be free space (wireless) or via transmission lines (e.g. waveguide, optical fiber, microstrip). The radio wave (called a carrier wave) $c(t)$ with a defined frequency is "modulated" (modified) by the message signal $m(t)$. In other words the amplitude and/or phase of the carrier wave is modified so as to include the information stored within the message.

The modified (modulated) carrier wave is called a communications signal $s(t)$ and has 3 main attributes: frequency, amplitude and phase.

A communications system transmits a message $m(t)$ to by first modulating the carrier wave $c(t)$ with the message $m(t)$ to create the communications signal $s(t)$ and transmits $s(t)$ as a radio (or optical) wave. $s(t)$ is degraded by noise and interference on the channel. The system receives the noisy and distorted $s(t)$ and extracts the message $m(t)$ by a process called demodulation.

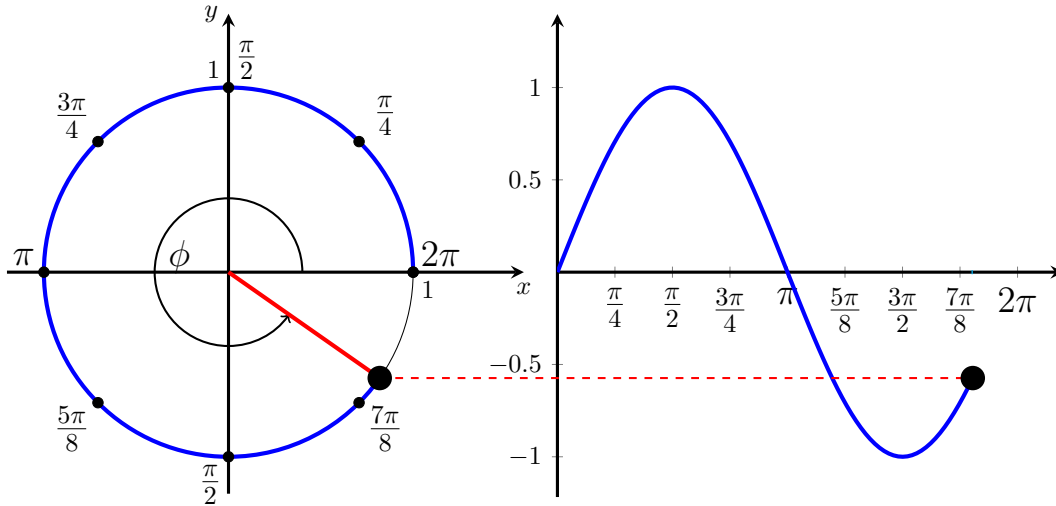


Figure 1.1: A cosine wave being created by projecting a moving point on a unit circle.

We wish to find a mathematical way to describe $s(t)$. The mathematical description of $s(t)$ and in particular the information contained in the amplitude and phase can take several forms. We choose to describe $s(t)$ using complex signals based on Euler's equation $e^{j\theta} = \cos\theta + j\sin\theta$. The reason for this choice is that complex signals are used in software-defined radios in which many of the functions of a communications systems such as frequency-shifting, filtering, modulation and demodulation are complex functions (complex arithmetic) implemented in software.

1.1.1 Carrier wave

We begin by considering the carrier wave, which is a "pure" sinewave, such as would be emitted by a laser. It contains one and only one frequency and has constant amplitude and phase that do not change with time. We will call this pure sine wave $c(t)$.

We can describe the carrier wave mathematically as $c(t) = \cos 2\pi f_c t$ or $c(t) = \sin 2\pi f_c t$ where f_c is the frequency of the carrier wave.

The carrier wave is a real (not complex) waveform that can be represented as a real function of time and as a time-varying analog (real) voltage and plotted on an oscilloscope and connected to a transmission line (e.g. coaxial cable, fiber, waveguide, microstrip) to carry it from one point to another, or connected to an antenna to radiate it into space.

A cosine or sine wave is obtained from the projection of a point moving to trace out a unit circle.

Note that we get the same cosine wave regardless of the direction of rotation, but we get the opposite polarity sine wave when we reverse the direction of rotation.

Using Euler's equation $e^{j\theta} = \cos\theta + j\sin\theta$ and also $\cos\theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$ we can write $c(t) = e^{j2\pi f_c t} + e^{-j2\pi f_c t}$. Thus $c(t)$ can be represented mathematically as the sum of two complex exponentials,

one positive and one negative. There is a scaling factor of 2 that is not important for the present discussion.

Euler's equation introduces the idea of complex numbers ($z = e^{j\theta} = \cos\theta + j\sin\theta$) that can be written in polar or rectangular form. We also introduce complex functions of time ($z(t) = e^{j2\pi f_c t} = \cos 2\pi f_c t + j\sin 2\pi f_c t$) written in polar or rectangular form.

Complex numbers are used in many contexts including communications signals and digital signal processing (transfer function of digital filters).

More generally we can write a complex number z as

$$\begin{aligned} z &= ae^{j\theta} \\ &= a \cos\theta + ja \sin\theta \\ &= x + jy \end{aligned} \tag{1.1}$$

where a is the magnitude, θ is the phase, x is the real part, y is the imaginary part. We write a complex function of time $z(t)$ as

$$\begin{aligned} z(t) &= a(t)e^{j\theta(t)} \\ &= a(t) \cos\theta(t) + ja(t) \sin\theta(t) \\ &= x(t) + jy(t) \end{aligned} \tag{1.2}$$

where $a(t)$ is the magnitude, $\theta(t)$ is the phase, $x(t)$ is the real part, $y(t)$ is the imaginary part.

How do we plot a complex function of time? One method is to plot both the amplitude and phase as two separate 2-D plots, or plot the real part and the imaginary part as two separate 2-D plots.

Another method is to visualize the positive complex exponential $z(t) = e^{j2\pi f_c t}$ in 3 dimensions: real axis (real), imaginary axis (imag) and time. A 3-D plot can be made in Matlab using the command `plot3(t, x, y)`. This 3-D curve is called a *helix*. $z(t) = e^{j2\pi f_c t}$ is a complex waveform that cannot be viewed on an 2-D oscilloscope screen as a real voltage versus time, and cannot be carried over a wire or transmission line or radiated into space via an antenna. However complex functions of time $z(t)$ are very useful for representing complex communications signals in software.

The end view of the helix (projection on the real-imag plane) shows the circle. The bottom (or top) view (projection on the real-time plane) of the helix shows the real part of the helix (a cosine wave) and the side view (projection on the imag-time plane) of the helix shows the imaginary part of the helix (a sine wave). Either the real or the imaginary parts are real waveforms that can be viewed on an oscilloscope. The end view is what would be seen on an x-y oscilloscope with the real and imaginary signals connected to the x and y inputs of the oscilloscope respectively. Since the cosine and sine wave are 90 degrees out of phase, the x-y plot is a circle.

The helix represents the carrier wave with constant amplitude and phase. In the next section we see what happens when the amplitude and phase are not constant.

In this figure, the helix arises from a counter-clockwise rotation as time increases. The positive complex exponential $e^{j2\pi f_c t}$ is defined to have a positive frequency.

We can visualize the negative complex exponential $e^{-j2\pi f_c t}$ as a helix in the opposite direction. The helix arises from a clockwise rotation as time increases. The negative complex exponential $e^{-j2\pi f_c t}$ is defined to have a negative frequency.

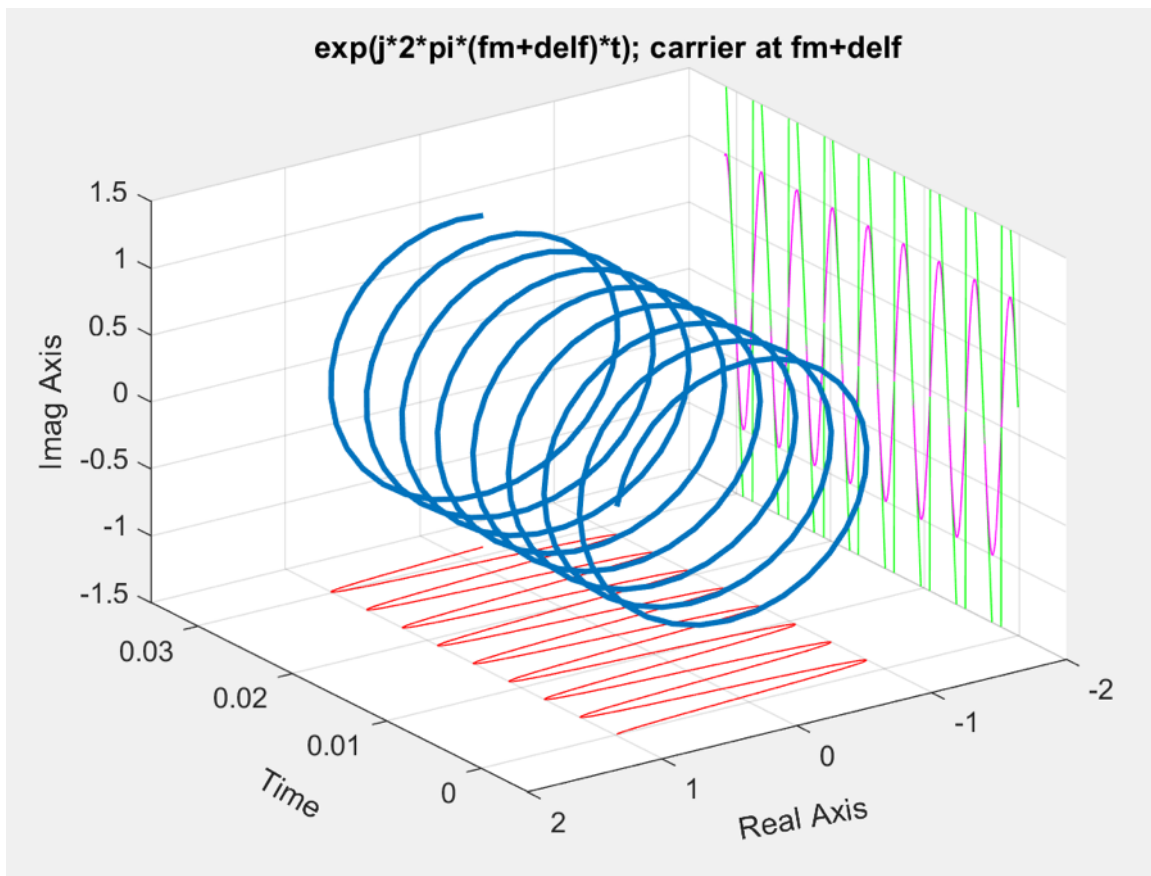


Figure 1.2: Helix with constant frequency, amplitude and phase.

The sum of the two complex exponentials $c(t) = \frac{e^{j2\pi f_c t} + e^{-j2\pi f_c t}}{2}$ (helixes in opposite directions) is a cosine wave on the real-time plane, and zero in the imag-time plane.

We can also write $c(t) = [c_+(t) + c_-(t)]/2$ where $c_+(t) = e^{j2\pi f_c t}$ and $c_-(t) = e^{-j2\pi f_c t}$. This notation shows explicitly that the real carrier wave $c(t) = \cos(2\pi f_c t)$ is made up of a positive frequency component plus a negative frequency component.

We can also write

$$c_+(t) = e^{j2\pi f_c t} = \cos 2\pi f_c t + j \sin 2\pi f_c t \quad (1.3)$$

$$c_-(t) = e^{-j2\pi f_c t} = \cos 2\pi f_c t - j \sin 2\pi f_c t \quad (1.4)$$

$$c(t) = [c_+(t) + c_-(t)]/2 = \cos 2\pi f_c t \quad (1.5)$$

It is mathematically equivalent to write the real carrier wave as

$$c(t) = \text{Re}\{c_+(t)\} = \text{Re}\{e^{-j2\pi f_c t}\} = \cos 2\pi f_c t. \quad (1.6)$$

This notation is commonly used, but the disadvantage is that this notation does not clearly show how the carrier wave is made up of positive and negative frequency components. Obtaining the real carrier wave by taking the real part implies that something (the imaginary part) is deleted. It is more clear and transparent to obtain the real carrier wave by adding the positive and negative frequency components.

This completes the discussion of the carrier wave.

1.1.2 Modulated carrier wave with time-varying amplitude and phase

The communications signal $s(t)$ is a modified (or modulated) carrier wave and has 3 main attributes: frequency, amplitude and phase. We wish to find a mathematical way to describe $s(t)$.

We will again use Euler's equation $e^{j\theta} = \cos\theta + j\sin\theta$ and also $\cos\theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$ as a starting point.

Recall the carrier wave with positive and negative frequencies is written $c(t) = e^{j2\pi f_c t} + e^{-j2\pi f_c t}$. The carrier wave may also be written in a more general form

$$\begin{aligned} c(t) &= \frac{a(t)e^{j\theta(t)} + a(t)e^{-j\theta(t)}}{2} \\ &= a(t) \cos \theta(t) \end{aligned} \quad (1.7)$$

which is valid only if $a(t) = A_c = \text{constant}$ and $\theta(t) = 2\pi f_c t$.

To write the communications signal $s(t)$ in which the amplitude, frequency and/or phase are modified to carry a message, we modify the carrier wave by changing its amplitude and phase.

$$\begin{aligned} s(t) &= \frac{a(t)e^{j\theta(t)} + a(t)e^{-j\theta(t)}}{2} \\ &= a(t) \cos \theta(t) \end{aligned} \quad (1.8)$$

where $a(t)$ represents the amplitude of the signal after modulation and $\theta(t)$ is the phase of the carrier wave. The message is contained within $a(t)$ and $\theta(t)$ in a manner to be described later. The message may be analog or digital. Note that if there is no message, then $a(t) = A_c = \text{constant}$ and $\theta(t) = 2\pi f_c t$.

Often the general form of a radio signal (or any communications signal) is written as the real part of a positive complex exponential

$$s(t) = \text{Re}\{a(t)e^{j\theta(t)}\} = a(t)\cos\theta(t) \quad (1.9)$$

This is mathematically identical to taking the projection of the helix in the figure above.

This method of writing $s(t)$ by taking the real part of a complex signal appears in many textbooks and articles, but is potentially confusing, since it appears that something (the imaginary part) is being removed from the positive complex exponential and then the mystery is what happened to that imaginary part and why we don't need it anymore. Instead, it may be better to think of adding the negative complex exponential to the positive complex exponential (adding the two helices) to get the same result, and noticing that the imaginary part is cancelled out by the addition.

Exercise: write this out in detail.

Answer:

$$\begin{aligned} s(t) &= \frac{a(t)e^{j\theta(t)} + a(t)e^{-j\theta(t)}}{2} \\ &= \frac{a(t)\cos\theta(t)}{2} + j\frac{a(t)\sin\theta(t)}{2} + \frac{a(t)\cos\theta(t)}{2} - j\frac{a(t)\sin\theta(t)}{2} \\ &= a(t)\cos\theta(t) \end{aligned} \quad (1.10)$$

Note the scaling factor of 2 appears here. This factor may be important in some circumstances where we need to know the signal amplitude and not so important in others where the signal amplitude would be changed by gain or attenuation stages in the system.

Note that the instantaneous frequency $f_i(T)$ of the signal is related to the phase $\theta(t)$ via

$$f_i(T) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} \quad (1.11)$$

For the positive exponential, the frequency is positive and visa versa.

If $d\theta = 2\pi$ and $dt = 1$ second, then the frequency is one cycle per second or 1 Hz. Thus 2π radians in one second is one cycle per second or 1 Hz. One cycle per second is one cycle of a *cos* or *sin* wave per second.

In the special case where there is no modulation applied to the carrier signal (ie: no message sent), then we say the carrier wave is "unmodulated" and we write $s(t) = c(t)$ is a carrier wave oscillating at a frequency of f_c and scaled by a constant carrier amplitude coefficient $a(t) = A_c$.

In this special case, the amplitude $a(t) = A_c$ is a constant, and the phase $\theta(t)$ increases linearly with time such that $\theta(t) = 2\pi f_c t$.

The instantaneous frequency is exactly the carrier frequency (exercise)

$$f_i(T) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} = \frac{1}{2\pi} \frac{d2\pi f_c t}{dt} = \frac{1}{2\pi} 2\pi f_c = f_c$$

as expected.

In this case, we write $a(t) = A_c$ (constant independent of time), $\theta(t) = 2\pi f_c t$ (linear increase of phase with time, 2π radians every $1/f_c$ seconds).

$$\text{Therefore } c(t) = A_c \cos(2\pi f_c t) = A_c \cos(\omega t).$$

As mentioned in the previous section, this signal is called the *carrier wave* $c(t)$ and can also be visualized as a phasor with angular frequency $\omega_c = 2\pi f_c$ and with period $T = 1/f_c$.

In general, when a message is sent and the carrier wave is modulated, the amplitude $a(t)$ of the carrier wave may be time varying, and the phase of the carrier $\theta(t)$ may have a time varying phase component $\phi(t)$ that is added to the linear phase $2\pi f_c t$, thus

$$\begin{aligned} s(t) &= \text{Re}\{a(t)e^{j\theta(t)}\} = a(t)\cos\theta(t) \\ \theta(t) &= 2\pi f_c t + \phi(t) \\ s(t) &= a(t)\cos(2\pi f_c t + \phi(t)) \end{aligned} \tag{1.12}$$

The radio signal $s(t)$ is a cosine wave at frequency f_c with time-varying amplitude $a(t)$ and phase $\phi(t)$. It is useful to write the radio signal as the real part of a complex waveform or (equivalently) as the sum of a positive frequency complex waveform $s_+(t)$ and a negative frequency complex waveform $s_-(t)$.

$$\begin{aligned} s(t) &= \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\} \\ &= \frac{a(t)e^{j\phi(t)}e^{j2\pi f_c t}}{2} + \frac{a(t)e^{-j\phi(t)}e^{-j2\pi f_c t}}{2} \\ &= \frac{s_+(t) + s_-(t)}{2} \end{aligned} \tag{1.13}$$

where

$$\begin{aligned} s_+(t) &= \frac{a(t)}{2} e^{j[2\pi f_c t + \phi(t)]} \\ &= \frac{a(t)}{2} \cos[2\pi f_c t + \phi(t)] + j \frac{a(t)}{2} \sin[2\pi f_c t + \phi(t)] \end{aligned} \tag{1.14}$$

$$\begin{aligned} s_-(t) &= \frac{a(t)}{2} e^{-j[2\pi f_c t + \phi(t)]} \\ &= \frac{a(t)}{2} \cos[2\pi f_c t + \phi(t)] - j \frac{a(t)}{2} \sin[2\pi f_c t + \phi(t)] \end{aligned} \tag{1.15}$$

When we add the positive and negative frequency signals, the imaginary parts cancel out to yield the real signal

$$s(t) = \frac{s_+(t) + s_-(t)}{2} = a(t)\cos(2\pi f_c t + \phi(t)) \quad (1.16)$$

In the equations above, the complex exponentials are written in polar form showing amplitude and phase. The radio signal may also be written

$$s(t) = \text{Re}\{s_+(t)\} = \text{Re}\{s_-(t)\}, \quad (1.17)$$

where the positive frequency signal $s_+(t) = s_a(t)$ is also called the *analytic signal*.

There is also a negative frequency analytic signal $s_-(t)$ but this definition is rarely used. However, the best way to think is that the real signal

$$s(t) = [s_+(t) + s_-(t)]/2 \quad (1.18)$$

is the sum of the positive and negative frequency signal, even though it is mathematically equivalent to $s(t) = \text{Re}\{s_+(t)\}$.

The analytic representation of a sinusoid is obtained by expressing it in terms of complex exponentials, discarding the negative frequency component, and doubling the positive frequency component. Thus if

$$\begin{aligned} s(t) &= a(t)\cos\theta(t) \\ &= a(t)e^{j\theta(t)}/2 + a(t)e^{-j\theta(t)}/2 \end{aligned} \quad (1.19)$$

then $s_+(t) = a(t)e^{j\theta(t)}$ where $\theta(t) = 2\pi f_c t + \phi(t)$.

1.1.3 Complex envelope

We define the *complex envelope* of the communications signal to be $\tilde{s}(t) = a(t)e^{j\phi(t)}$. The complex conjugate of the complex envelope is $a(t)e^{-j\phi(t)}$.

The communications signal is then written

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \\ &= a(t)\cos[2\pi f_c t + \phi(t)] \\ &= \frac{s_+(t) + s_-(t)}{2} \\ &= \frac{a(t)}{2}e^{j\phi(t)}e^{j2\pi f_c t} + \frac{a(t)}{2}e^{-j\phi(t)}e^{-j2\pi f_c t} \\ s_+(t) &= \tilde{s}(t)e^{j2\pi f_c t} \\ s_-(t) &= \tilde{s}^*(t)e^{-j2\pi f_c t} \end{aligned} \quad (1.20)$$

For the negative frequency component $s_-(t)$ we use the complex conjugate of the complex envelope $a(t)e^{-j\phi(t)}$.

To encode a message $m(t)$ on the carrier wave we vary $a(t)$ and/or $\phi(t)$ in step with the message $m(t)$. Thus $a(t)$, $\phi(t)$ are specified as a function of the message $m(t)$. These functions will be specified later when we discuss specific modulation types.

We can write the radio signal in a form where the complex envelope is separated into its real and imaginary parts $\tilde{s}(t) = a(t)e^{j\phi(t)} = I(t) + jQ(t)$

Exercise: Write $a(t)$ and $\phi(t)$ as a function of $I(t), Q(t)$.

Answer: Using the identity

$$\cos(A + B) = \cos A \cos B - \sin A \sin B$$

the general radio signal

$$s(t) = a(t)\cos(2\pi f_c t + \phi(t)) = \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\}$$

may also be written

$$\begin{aligned} s(t) &= a(t)\cos 2\pi f_c t \cos \phi(t) - a(t)\sin 2\pi f_c t \sin \phi(t) \\ &= I(t)\cos 2\pi f_c t - Q(t)\sin 2\pi f_c t \end{aligned} \quad (1.21)$$

where

$$\begin{aligned} I(t) &= a(t)\cos \phi(t) = \text{Re}\{a(t)e^{j\phi(t)}\} \\ Q(t) &= a(t)\sin \phi(t) = \text{Im}\{a(t)e^{j\phi(t)}\} \end{aligned} \quad (1.22)$$

are called the “in-phase” and “quadrature” components, respectively, and thus

$$I(t) + jQ(t) = a(t)e^{j\phi(t)}.$$

Thus we can write

$$\begin{aligned} s(t) &= \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\} \\ &= \text{Re}\{[I(t) + jQ(t)][\cos 2\pi f_c t + j\sin 2\pi f_c t]\} \\ &= I(t)\cos 2\pi f_c t - Q(t)\sin 2\pi f_c t \\ &= a(t)\cos(2\pi f_c t + \phi(t)) \end{aligned} \quad (1.23)$$

The general radio signal $s(t)$ must be a real signal that we can view on an oscilloscope.

We can write $s(t)$ as the real part of a complex signal or the sum of a positive complex exponential and a negative complex exponential.

From the above equation, we see that $s(t)$ can be described as either

1. a cosine wave with amplitude $a(t) \geq 0$ and phase $\phi(t)$, or
2. the sum of a cosine wave with amplitude $I(t)$ and a sine wave with amplitude $Q(t)$ where $I(t), Q(t)$ can be greater or less than zero.

In both cases, the message is a two-dimensional (complex) signal represented using either $a(t), \phi(t)$ (polar form) or $I(t), Q(t)$ (rectangular form). In the figure, $M(t) = a(t)$.

$I(t), Q(t)$ are functions of the message signal(s), where the exact function depends on the modulation type. A simple example is to consider the message signal to be a stereo (2-channel) music signal written as $m_L(t), m_R(t)$ and choose

$$m_L(t) = I(t), m_R(t) = Q(t)$$

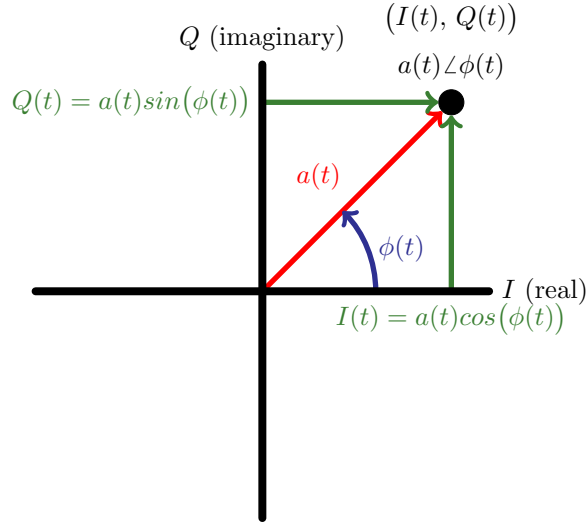


Figure 1.3: Polar and rectangular coordinates

$$I(t) + jQ(t) = a(t)e^{j\phi(t)} = \tilde{s}(t) = m_L(t) + jm_R(t)$$

is the so-called *complex baseband* signal that is a function of the message. We previously also called this the complex envelope. For a stereo music signal, the complex baseband signal is $m_L(t) + jm_R(t)$. The radio signal $s(t)$ is the so-called *real passband* signal that contains the message modulated onto the carrier wave at frequency f_c .

The complex baseband signal represents the time varying information represented by $a(t), \phi(t)$ which is transported by a carrier wave from one point to another. The complex baseband signal $\tilde{s}(t) = I(t) + jQ(t) = a(t)e^{j\phi(t)}$ can also be represented as a 3-D waveform with projection $I(t)$ onto the real axis and projection $Q(t)$ onto the imaginary axis

For example, if the amplitude is constant and the phase has only two possible values $-\pi/4$ and $-3\pi/4$ to represent a 101010 ... data sequence, then the complex baseband signal appears like a bipolar square wave, and the projections $I(t), Q(t)$ are also square waves. In a future chapter, we will recognize this waveform as Phase Shift Keying (PSK).

In another example, if the amplitude is constant and the phase is varied in a sinusoidal manner $\phi(t) = \beta \sin 2\pi f_m t$ at some frequency f_m to represent a 101010 ... data sequence (as will be seen in a later chapter), then the 3-D waveform appears as in the figure below. In this example, the helix rotates in one direction, slows down and then rotates in the opposite direction, following the sinusoidal variation in phase. The amount of rotation in each direction depends on β . The projections $I(t), Q(t)$ are periodic patterns which is not easy to interpret without also seeing the 3-D helix. In a future chapter, we will recognize this waveform as Frequency Shift Keying (FSK).

In the special case where I, Q are both constants then $s(t)$ is the sum of a cos wave and a sin wave with different amplitudes, which is a cosine wave with constant amplitude and phase. The figure below shows the radio frequency (RF) signal

$$s(t) = I \cos 2\pi f_c t - Q \sin 2\pi f_c t = a \cos 2\pi f_c t + \phi$$

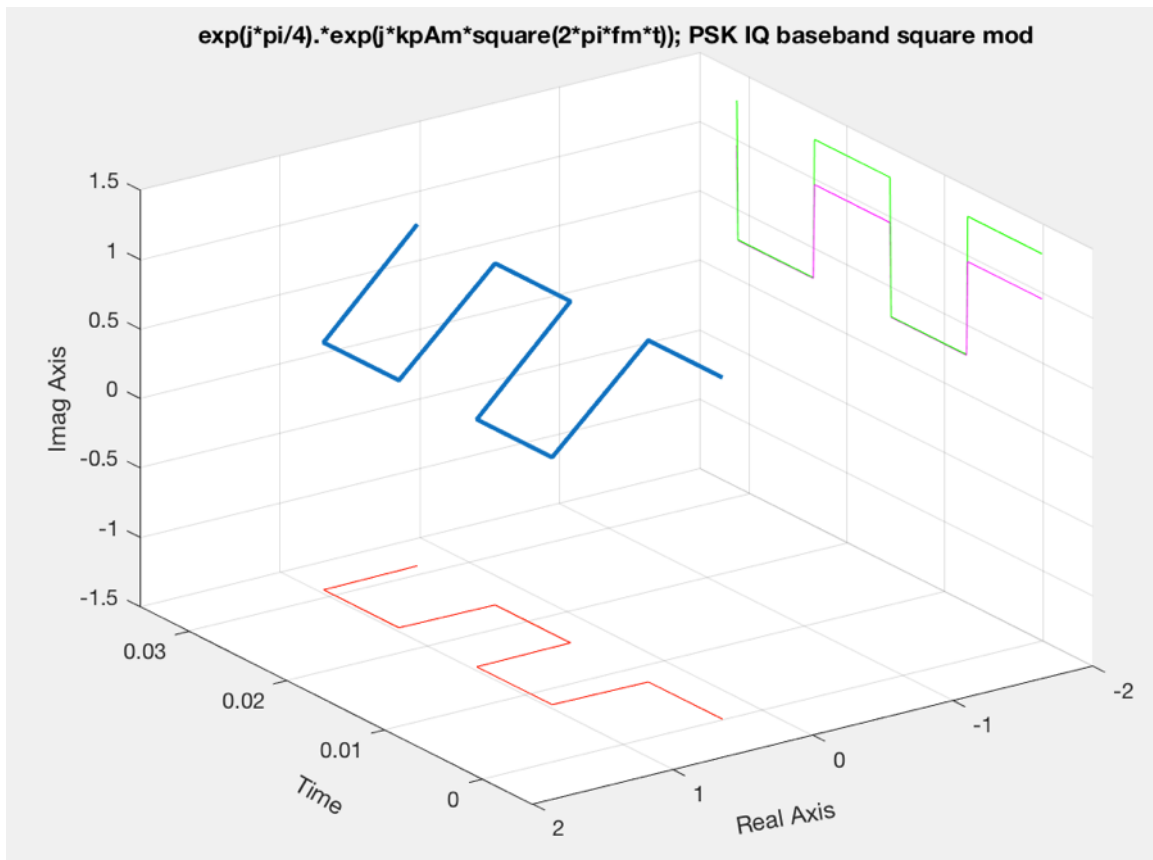


Figure 1.4: complex baseband phase shift keying with constant $a(t)$ and two values for $\phi(t)$

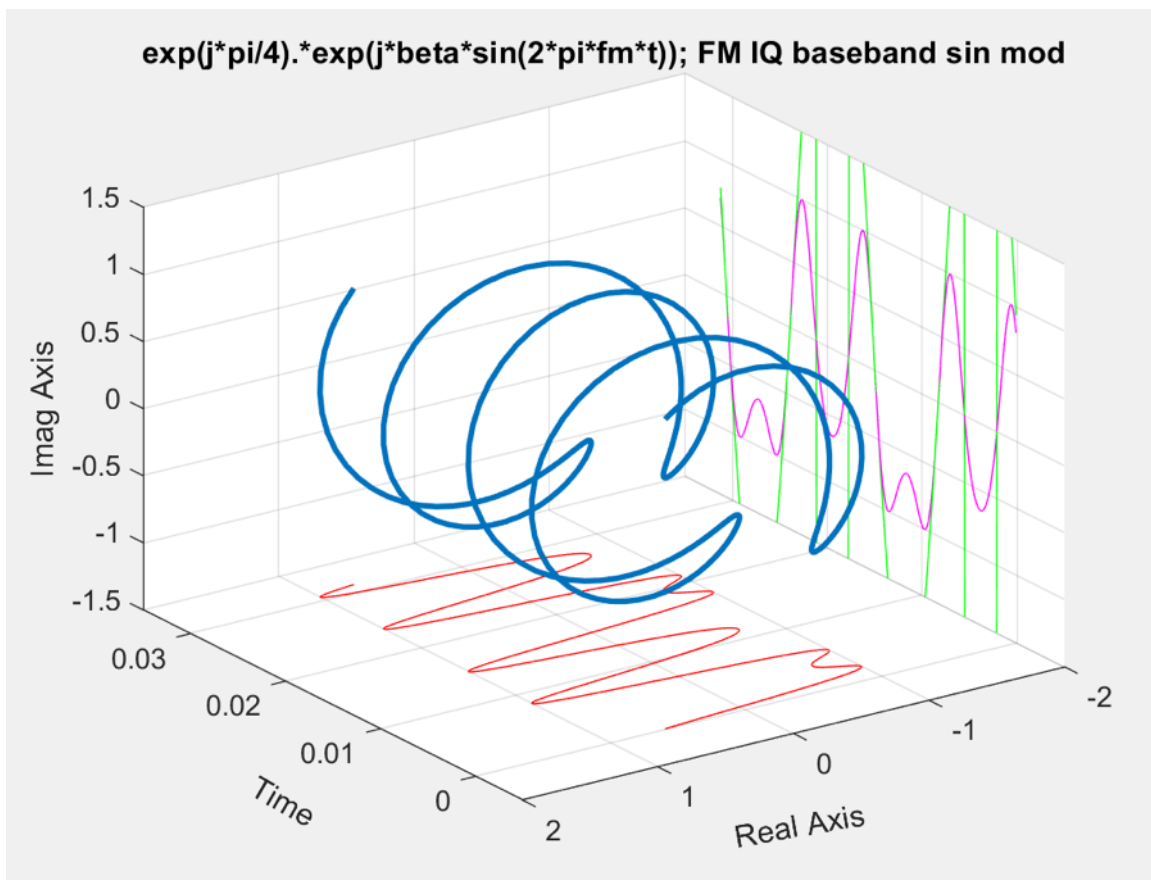
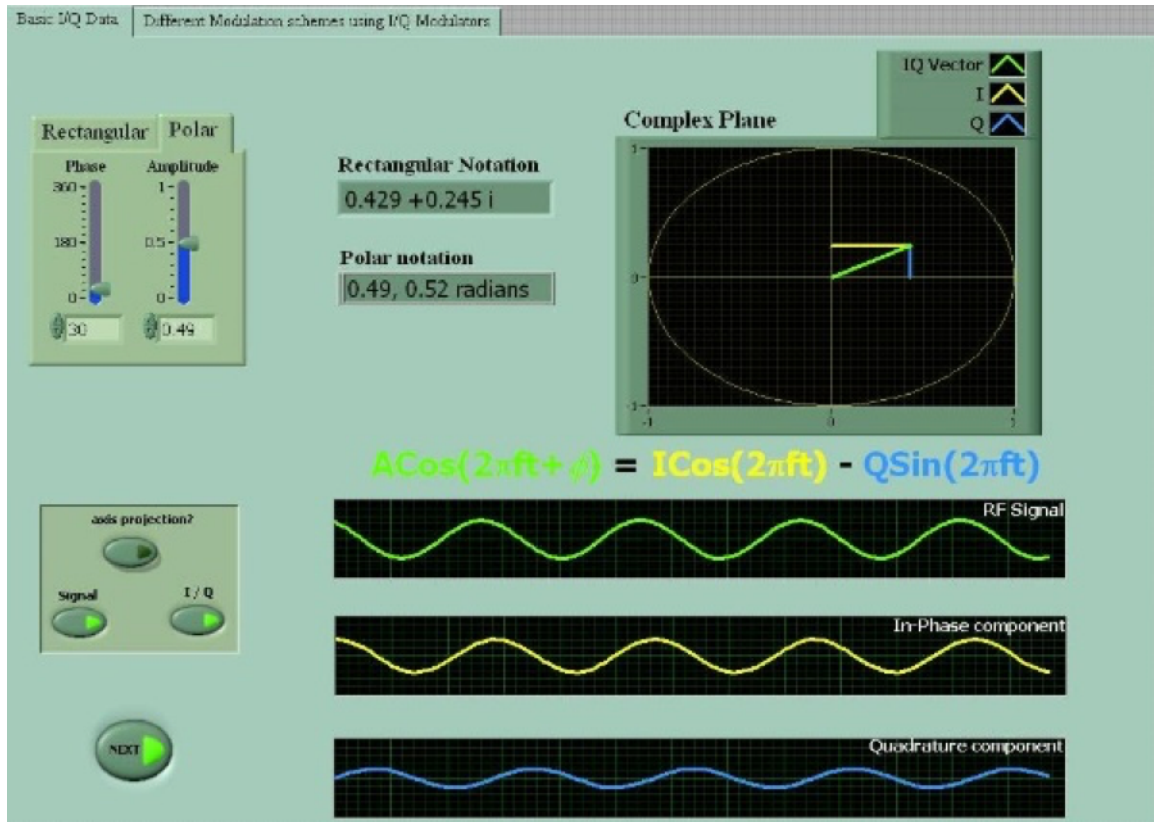


Figure 1.5: Complex baseband frequency-shift keying with constant $a(t)$ and sinusoidal $\phi(t)$

Figure 1.6: IQ projections for I, Q, a, ϕ constant independent of time

where both a, ϕ are functions of I, Q . The in-phase component $I \cos 2\pi f_c t$ and the quadrature component $Q \sin \pi f_c t$ are also cosine waves at frequency f_c with constant amplitude and phase.

Exercise: write an expression for both a, ϕ as a function of I, Q .

1.1.4 Summary

A radio (or other communications) signal may be made up of a cosine wave at a carrier frequency f_c with time varying amplitude and phase. The signal may be written as a *real passband* signal

$$s(t) = a(t) \cos(2\pi f_c t + \phi(t))$$

where the message is contained in $a(t), \phi(t)$. The signal may also be written in *complex baseband* form as $\tilde{s}(t) = I(t) + jQ(t) = a(t)e^{j\phi(t)}$ and viewed in the complex plane. The complex baseband signal contains the amplitude and phase only (i.e. the message information) and does not explicitly include the carrier frequency f_c . The real passband signal may be obtained from the complex baseband signal as

$$s(t) = \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} = a(t) \cos[2\pi f_c t + \phi(t)]$$

The real passband signal is equivalent to the sum of the positive and negative complex passband

signals

$$\begin{aligned}
 s(t) &= \frac{s_+(t) + s_-(t)}{2} \\
 &= \frac{a(t)}{2} e^{j\phi(t)} e^{j2\pi f_c t} + \frac{a(t)}{2} e^{-j\phi(t)} e^{-j2\pi f_c t} \\
 s_+(t) &= \tilde{s}(t) e^{j2\pi f_c t} \\
 s_-(t) &= \tilde{s}^*(t) e^{-j2\pi f_c t}
 \end{aligned} \tag{1.24}$$

$s_+(t)$ is also called the analytic signal.

A description of message types (analog and digital) and how they are contained in $a(t), \phi(t)$ is explained in detail a later section. A brief example: consider a digital message consisting of alternating 101010 bits. We can choose a modulation scheme as follows: assign $\phi(t) = 0$ during a 0 bit and $\phi(t) = \pi$ during a 1 bit and keep $a(t)$ constant for both 1 and 0 bits. We could also make different choices.

Exercise: sketch a graph of the passband signal waveform with this example modulation scheme.

Exercise: invent another (different) modulation scheme where we assign $a(t), \phi(t)$ as a function of the message, and sketch the passband signal waveform.

Exercise: invent a modulation scheme where we assign $I(t), Q(t)$ as a function of the message, and sketch the passband signal waveform.

1.1.5 Frequency shifting

From the definition of a (real) communications signal with information contained in $a(t), \phi(t)$,

$$\begin{aligned}
 s(t) &= \frac{s_+(t) + s_-(t)}{2} \\
 &= \frac{a(t)}{2} e^{j\phi(t)} e^{j2\pi f_c t} + \frac{a(t)}{2} e^{-j\phi(t)} e^{-j2\pi f_c t} \\
 &= a(t) \cos[2\pi f_c t + \phi(t)]
 \end{aligned} \tag{1.25}$$

we observe that it consists of the complex envelope multiplied by a positive complex exponential plus the complex conjugate of the complex envelope multiplied by a negative complex exponential. We now consider communications signals in the frequency domain.

First recall the Fourier transform. Fourier analysis shows that any complex waveform can be resolved into sinusoidal waveforms of a fundamental frequency and a number of harmonic frequencies. The spectrum analyzer effectively performs the Fourier integral:

$$\begin{aligned}
 S(f) &= \int_{t=-\infty}^{\infty} s(t) e^{-j2\pi f t} dt \\
 &= \int_{t=-\infty}^{\infty} s(t) \cos(2\pi f t) dt - j \int_{t=-\infty}^{\infty} s(t) \sin(2\pi f t) dt
 \end{aligned} \tag{1.26}$$

The integral finds the frequency components in $s(t)$ by correlating $s(t)$ with cosine and sine waves

at each frequency f . For a particular frequency $f = f_c$

$$s(t) = \cos 2\pi f_c t \quad (1.27)$$

and

$$\begin{aligned} S(f) &= \int_{t=-\infty}^{\infty} \cos(2\pi f_c t) \cos(2\pi f t) dt \\ S(f = f_c) &= \int_{t=-\infty}^{\infty} 0.5[1 + \cos(4\pi f_c t)] dt \\ &= 0.5 \int_{t=-\infty}^{\infty} dt + 0.5 \int_{t=-\infty}^{\infty} \cos(4\pi f_c t) dt \end{aligned} \quad (1.28)$$

The cosine waves are in phase for all time, the product of the two cosine waves contains DC, thus the integral integrates DC over all time, resulting in infinity (delta functions) at $f = \pm f_c$. For all other frequencies $f \neq f_c$ the cosine waves drift in and out of phase over time, there is no DC component, and the integral is zero. Thus for $s(t) = \cos(2\pi f_c t)$ we can write:

$$\begin{aligned} S(f) &= \int_{t=-\infty}^{\infty} \cos(2\pi f_c t) \cos(2\pi f t) dt \\ &= [\delta(f - f_c) + \delta(f + f_c)]/2 \end{aligned} \quad (1.29)$$

Exercise: repeat the above calculation for $s(t) = \cos(2\pi f_c t + \phi)$ where $\phi = \pi/4 = \text{constant}$.

The Fourier transform has frequency-shifting properties that will be used often in everything that follows. The properties are

$$\begin{aligned} s(t) &\Leftrightarrow S(f) \\ s(t)e^{j2\pi f_1 t} &\Leftrightarrow S(f - f_1) \\ s(t)e^{-j2\pi f_1 t} &\Leftrightarrow S(f + f_1) \end{aligned}$$

To apply these properties, recall the definitions

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \\ &= a(t) \cos[2\pi f_c t + \phi(t)] \\ &= s_+(t) + s_-(t) \\ &= a(t)e^{j\phi(t)}e^{j2\pi f_c t} + a(t)e^{-j\phi(t)}e^{-j2\pi f_c t} \end{aligned}$$

$$\begin{aligned} s_+(t) &= \tilde{s}(t)e^{j2\pi f_c t} \\ s_-(t) &= \tilde{s}^*(t)e^{-j2\pi f_c t} \end{aligned}$$

$$\begin{aligned} 2s(t) &= s_+(t) + s_-(t) \\ &= a(t)e^{j\phi(t)}e^{j2\pi f_c t} + a(t)e^{-j\phi(t)}e^{-j2\pi f_c t} \\ &= 2a(t) \cos[2\pi f_c t + \phi(t)] \\ s(t) &= \tilde{s}(t)e^{j2\pi f_c t} + \tilde{s}^*(t)e^{-j2\pi f_c t} \end{aligned}$$

Transforming these definitions into the frequency domain using the shifting properties shown below

$$\begin{aligned} 2S(f) &= S_+(f) + S_-(f) \\ &= \tilde{S}(f - f_c) + \tilde{S}^*(-(f + f_c)) \end{aligned}$$

Thus we see that the real communications signal is made up of the complex baseband signal (containing the message information) shifted by the carrier frequency. The shape of the spectrum of the complex baseband signal $\tilde{S}(f)$ is the same as the shape of the spectrum of the positive frequency signal $S_+(f)$, and the shape of the spectrum of the conjugate of the complex baseband signal $\tilde{S}^*(f)$ is the same as the shape of the spectrum of the negative frequency signal $S_-(f)$.

The Fourier transform of a real signal $s(t)$ will always be symmetrical around zero Hz, i.e. the shape of the Fourier transform (the spectrum) in the negative frequency range will be the mirror image of the Fourier transform (spectrum) in the positive frequency range.

The Fourier transform of a complex signal such as $s_+(t)$ or $\tilde{s}(t)$ will have an asymmetrical spectrum, i.e the spectrum at negative frequencies will not be the mirror image of the spectrum at positive frequencies. For $s_+(t)$ the spectrum is zero at negative frequencies but not zero at positive frequencies.

A review of how the shifting properties arise is given below, remembering that

$$\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & x \neq 0 \end{cases} \quad (1.30)$$

$$\begin{aligned}
S(f) &= \int_{t=-\infty}^{\infty} s(t)e^{-j2\pi ft} dt \\
s(t) &\Leftrightarrow S(f) \\
\delta(t) &\Leftrightarrow 1 \\
1 &\Leftrightarrow \delta(f)
\end{aligned}$$

$$\begin{aligned}
e^{j2\pi f_1 t} &\Leftrightarrow \delta(f - f_1) \\
e^{-j2\pi f_1 t} &\Leftrightarrow \delta(f + f_1)
\end{aligned}$$

Proof

$$\begin{aligned}
S(f) &= \int_{t=-\infty}^{\infty} s(t)e^{-j2\pi ft} dt \\
&= \int_{t=-\infty}^{\infty} e^{-j2\pi f_1 t} e^{-j2\pi f_1 t} s(t) dt \\
&= \int_{t=-\infty}^{\infty} e^{-j2\pi(f+f_1)t} s(t) dt = \delta(f + f_1)
\end{aligned}$$

$$\begin{aligned}
s(t)e^{j2\pi f_1 t} &\Leftrightarrow S(f) \otimes \delta(f - f_1) = S(f - f_1) \\
s(t)e^{-j2\pi f_1 t} &\Leftrightarrow S(f) \otimes \delta(f + f_1) = S(f + f_1)
\end{aligned}$$

Proof

$$\begin{aligned}
S(f) \otimes H(f) &= \int_{\alpha=-\infty}^{\infty} S(\alpha)H(f - \alpha)d\alpha \\
S(f) \otimes \delta(f) &= \int_{\alpha=-\infty}^{\infty} S(\alpha)\delta(f - \alpha)d\alpha \\
&= S(f) \\
S(f) \otimes \delta(f - f_1) &= \int_{\alpha=-\infty}^{\infty} S(\alpha)\delta(f - f_1 - \alpha)d\alpha \\
&= S(f - f_1)
\end{aligned}$$

$$\begin{aligned}
s_+(t) &\Leftrightarrow S_+(f) \\
s_+^*(t) &\Leftrightarrow S_+^*(-f)
\end{aligned}$$

Proof for specific example

$$\begin{aligned}
s_+(t) &= ae^{\phi} e^{j2\pi f_c t} \\
S_+(f) &= ae^{\phi} \delta(f - f_c) \\
s_+^*(t) &= ae^{-\phi} e^{-j2\pi f_c t} \\
&= s_-(t) \\
S_+^*(-f) &= ae^{-\phi} \delta(-f - f_c) \\
&= S_-(f)
\end{aligned}$$

then

$$\begin{aligned}
s_+^*(t) &\Leftrightarrow S_+^*(-f) \\
ae^{-\phi} e^{-j2\pi f_c t} &\Leftrightarrow ae^{-\phi} \delta(-f - f_c) \\
&\Leftrightarrow ae^{-\phi} \delta(f + f_c)
\end{aligned}$$

since

$$\delta(+f + f_c) = \delta(-f - f_c) = \infty, \quad f = -f_c \quad (1.31)$$

For a software radio system where the signal is digitized (sampled) by an analog-to-digital converter (ADC) at a sampling rate $f_s = 1/T_s$, the Fourier transform is done using the Discrete Fourier Transform (DFT) algorithm on a block of N samples. The DFT is most often implemented using the Fast Fourier transform (FFT) algorithm.

Recall the discrete Fourier transform is defined as a Fourier transform operating on a sampled periodic signal with period N samples. The time domain signal $s(t)$ is sampled at a sampling rate $f_s = 1/T_s$ to make samples $s_n = s(t = nT_s)$, where T_s is the time resolution. After the discrete Fourier transform is done on a block of N samples $s_n, 0 \leq n \leq N-1$ to obtain the frequency domain signal $S(f)$, the result is a block of N samples $S_k, 0 \leq k \leq N-1$, where $S_k = S(f = kf_0)$ is the frequency resolution $f_0 = f_s/N$. The DFT is written as follows:

$$\begin{aligned}
 S(f = kf_0) &= S_k \\
 &= \int_{t=0}^{T_0=nT_s} s(t)e^{-j2\pi ft} dt \Big|_{t=nT_s, f=kf_0} \\
 &= \sum_{n=0}^{N-1} s(nT_s)e^{-j2\pi kf_0 nT_s} \\
 &= \sum_{n=0}^{N-1} s_n e^{-j2\pi kn/N}
 \end{aligned} \tag{1.32}$$

where $T_s f_0 = 1/N$ and N is the number of samples in both the time domain and frequency domain. This calculation may be done in Matlab $S_k = fft(s_n)$ where the blocks of N samples s_n, S_k are Matlab vectors.

For a software radio system receiving real passband signals in the 2.4 GHz range (e.g. WiFi) with sampling rate of 100 MHz, the “complex baseband” signal may contain up to communications signals in a 100 MHz bandwidth, spaced by 5 MHz. Similarly for a software radio system receiving real passband signals in the 88–108 MHz FM broadcast range with sampling rate of 20 MHz, the “complex baseband” signal will contain many FM broadcast signals in the 20 MHz bandwidth from 88–108 MHz.

1.1.6 Analytic signals

Recall the definitions

$$\begin{aligned}
 2s(t) &= s_+(t) + s_-(t) \\
 &= a(t)e^{j\phi(t)}e^{j2\pi f_c t} + a(t)e^{-j\phi(t)}e^{-j2\pi f_c t} \\
 &= 2a(t)\cos[2\pi f_c t + \phi(t)] \\
 s_+(t) &= \tilde{s}(t)e^{j2\pi f_c t} \\
 s_-(t) &= \tilde{s}^*(t)e^{-j2\pi f_c t} \\
 s(t) &= \tilde{s}(t)e^{j2\pi f_c t} + \tilde{s}^*(t)e^{-j2\pi f_c t} \\
 s_-(t) &= s_+^*(t)
 \end{aligned}$$

Transforming these definitions into the frequency domain

$$2S(f) = S_+(f) + S_-(f) = \tilde{S}(f - f_c) + \tilde{S}^*(-(f + f_c))$$

Note that

$$S_-(f) = S_+^*(-f)$$

and for a real-valued signal $s(t)$

$$S(-f) = S(f)^*$$

The analytic signal $S_a(f)$ is the same as $S_+(f)$ as defined below

$$\begin{aligned} S_a(f) &= 2S(f), f > 0 \text{ and } 0, f < 0 \\ S_a(f) &= S_+(f) \\ &= 2u(f)S(f) \\ &= S(f) + \text{sgn}(f)S(f) \end{aligned} \tag{1.33}$$

where $u(f) = 1 + \text{sgn}(f)$ is the Heaviside step function

$$u(f) = \begin{cases} 1 & f > 0 \\ 0 & f < 0 \end{cases} \tag{1.34}$$

and $\text{sgn}(f)$ is the sign function

$$\text{sgn}(f) = \begin{cases} 1 & f > 0 \\ -1 & f < 0 \end{cases} \tag{1.35}$$

In a future section 2.3, we will see how to represent $S_a(f) = S_+(f)$ in the time domain using the idea of a Hilbert transform.

Exercise: write the frequency domain definition of the analytic signal for the special case where $a(t) = 1, \phi(t) = 0$.

1.2 Block diagrams for real passband and complex baseband signals

Learning objectives

In this section, we show block diagrams that implement the equations from the previous section.

Overview of this section 1.2

The implementation of the equations in section 1.1 may be done in software or hardware.

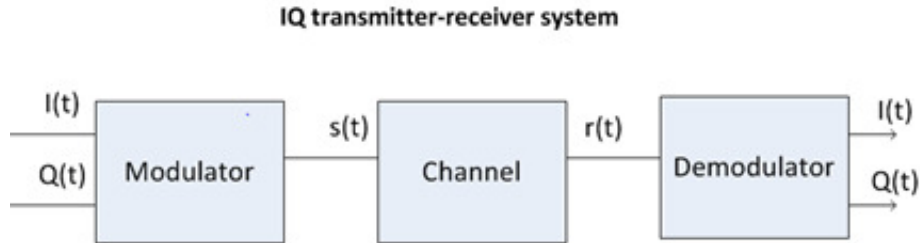
In the case of software, the complex signal equations are implemented directly in some language such as C++ and Python (for GNURadio) or in Matlab, Mathematica or similar. The signals must be represented by discrete-time samples, sampled at some sampling rate $f_s = 1/T_s$ such that $s(t)|_{t=nT_s} = s(t = nT_s) = s_n$ for real samples or $\tilde{s}(t)|_{t=nT_s} = \tilde{s}(t = nT_s) = \tilde{s}_n$ for complex samples. The samples are stored in digital memory as a vector and manipulated in software.

In the case of hardware, complex signals cannot be implemented directly from the equations since a complex signal cannot be represented by a real current that flows through a wire. The complex signal equations must be converted to real signals (voltage) that can make currents flow on wires.

Real signals can be implemented in either hardware or software.

The equations are used to implement a communications system that includes a

- transmitter (modulator) that modulates the information-bearing signal (message) onto a carrier to create a communications signal
- channel that may be a real analog thing such as optical fiber, copper wire, transmission line or antenna that radiates the signal into free space (wireless) or a simulated channel that is implemented in software. A simulated channel may be complex or real. The channel has a path loss L_0 that attenuates the communications signal as a result of the distance between transmitter and receiver. The channel may also include a time-varying filter that represents distortion caused by the physical nature of the channel.
- receiver (demodulator) that extracts the message from the communications signal



In what follows, we show block diagrams of the transmitter and receiver.

1.2.1 Radio transmitter (modulator)

Complex signal implementation - complex passband

The radio transmitter (modulator) may be described in complex form with a complex multiplication of the complex baseband message $I(t) + jQ(t)$ with $\cos 2\pi f_c t + j \sin 2\pi f_c t$ to yield a complex signal

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi(t)}e^{j2\pi f_c t} \\ &= [I(t) + jQ(t)][\cos 2\pi f_c t + j \sin 2\pi f_c t] \end{aligned} \quad (1.36)$$

The complex message $I(t), Q(t)$ is multiplied by the complex carrier wave $e^{j2\pi f_c t} = \cos 2\pi f_c t + j \sin 2\pi f_c t$ to create a complex passband signal $s_+(t)$

The complex signal implementation diagram in fig. 1.7 assumes the blocks are processing sampled digital signals in software.

If the system uses a real channel, then the radio signal $s(t)$ is the real part of this complex multiplication

$$s(t) = I(t) \cos 2\pi f_c t - Q(t) \sin 2\pi f_c t \quad (1.37)$$

If we need a real signal to transmit over a wire to an antennas, then we need a digital-to-analog converter (DAC) block after the Re block.

If we are building a computer simulation of the radio transmitter then the Re block and DAC may be omitted, and the output is the complex passband signal $s_+(t)$

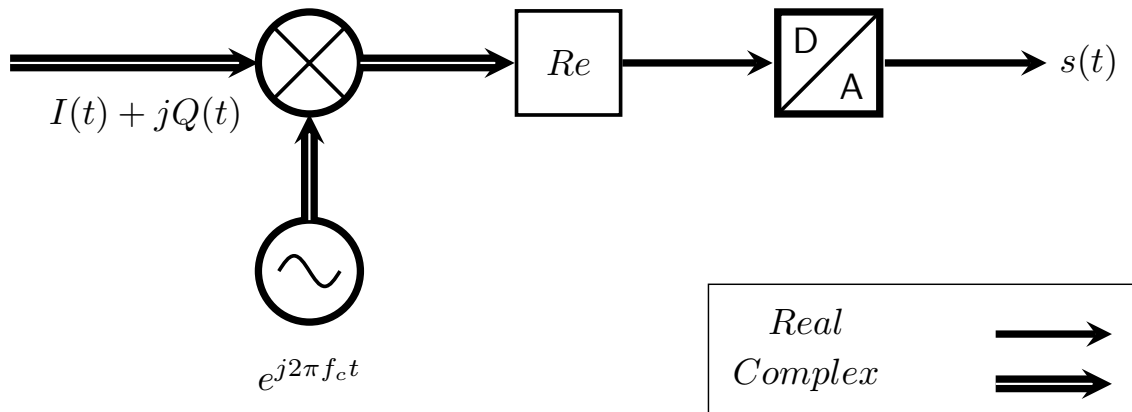


Figure 1.7: Complex implementation transmitter block diagram

Real signal implementation - real passband

In fig. 1.8 below, we show the following blocks

1. the D/A block is a digital to analog converter,
2. the *mixer* is a multiplier,
3. the *oscillators* output a cos wave for the I stream and a sin wave for the Q stream,
4. the *combiner* is an adder.

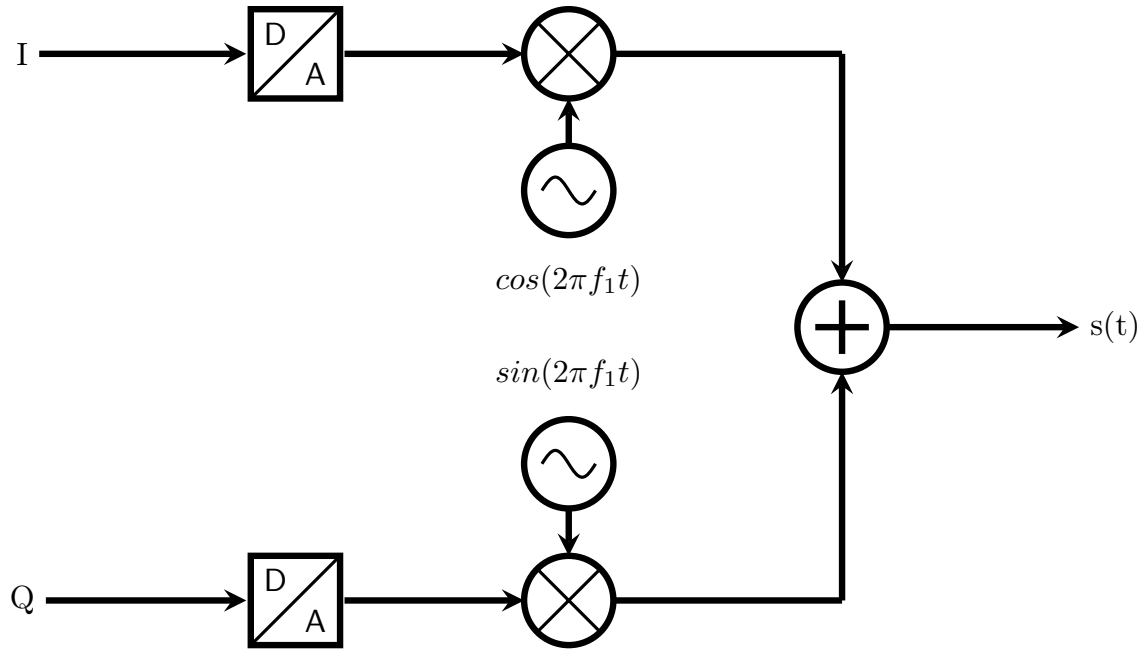


Figure 1.8: Real implementation transmitter block diagram

With the blocks defined as above, the figure shows a radio transmitter (or modulator) that produces the radio waveform $s(t) = I(t) \cos 2\pi f_c t - Q(t) \sin 2\pi f_c t$ from the message signals $I(t), Q(t)$, where in this example, $f_c = f_1$.

In the figure, the blocks process analog signals. However, if the D/As are omitted, the remaining blocks process digital signals, using software such as GNURadio or Matlab Simulink.

1.2.2 Radio receiver (demodulator)

If the entire communication system including channel is implemented by computer simulation, then the complex passband signal $s_+(t)$ is attenuated by the channel represented by a scaling factor $1/L_0$.

If the channel is real (wired or wireless), then the real passband signal $s(t)$ is transmitted over a distance via some real channel (wired or wireless), attenuated by the path loss L_0 and picked up by a receiver in the form $r(t) = s(t)/L_0$.

The receiver's task is to recover the message signals $I(t), Q(t)$ from the signal $r(t)$. This can be done using the receivers shown below.

Complex signal implementation

The radio receiver (demodulator) may be described in complex form with a complex multiplication of the complex passband signal $r_+(t) = L_0^{-1} s_+(t) = a(t) e^{j\phi(t)} e^{j2\pi f_c t} / L_0$.

We write the received signal in complex passband form

$$\begin{aligned} r_+(t) &= L_0^{-1} s_+(t) \\ &= L_0^{-1} a(t) e^{j\phi(t)} e^{j2\pi f_c t} \\ &= L_0^{-1} [I(t) + jQ(t)] [\cos 2\pi f_c t + j \sin 2\pi f_c t] \end{aligned} \quad (1.38)$$

The receiver demodulates the complex radio signal by multiplying $r_+(t)$ by the complex local oscillator $e^{-j2\pi f_c t} = \cos 2\pi f_c t - j \sin 2\pi f_c t$ to yield

$$\begin{aligned} r_+(t) e^{-j2\pi f_c t} &= [L_0^{-1} a(t) e^{j\phi(t)} e^{j2\pi f_c t}] e^{-j2\pi f_c t} \\ &= L_0^{-1} a(t) e^{j\phi(t)} \\ &= L_0^{-1} [I(t) + jQ(t)] \end{aligned} \quad (1.39)$$

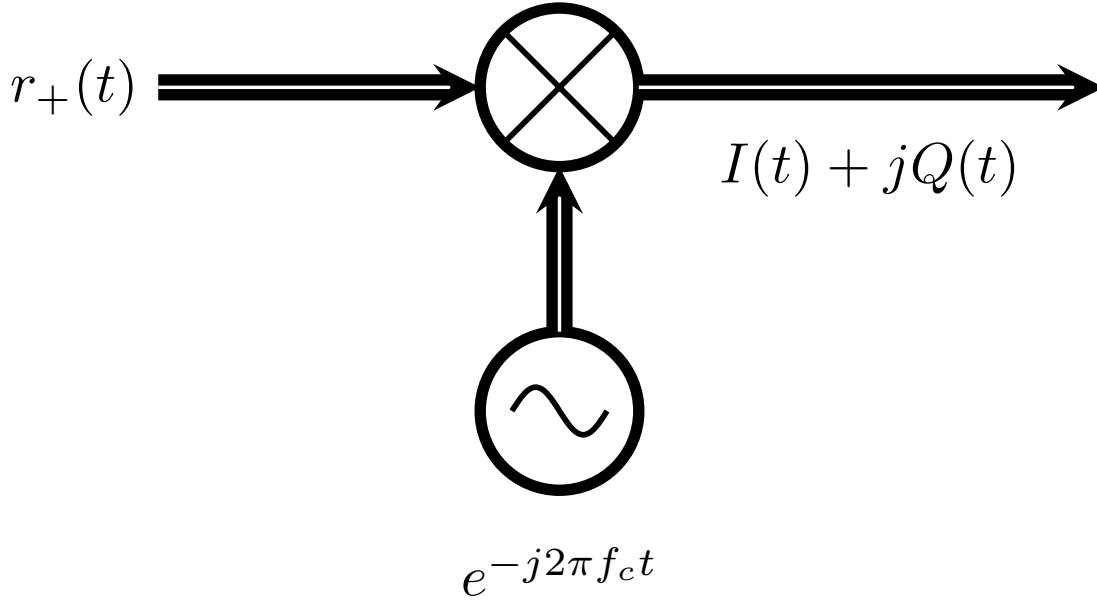


Figure 1.9: Complex baseband receiver block diagram

Real signal implementation

Recall the real passband signal

$$\begin{aligned} s(t) &= \text{Re}\{s_+(t)\} \\ &= \text{Re}\{a(t) e^{j\phi(t)} e^{j2\pi f_c t}\} \\ &= a(t) \cos[2\pi f_c t + \phi(t)] \end{aligned}$$

which is the real part of the complex passband signal.

The block diagram in fig. 1.10 is the dual of the one in fig. 1.8. In this receiver, the input signal and all processing is analog, and $I(t), Q(t)$ are digitized by an analog-to-digital converter (A/D or ADC).

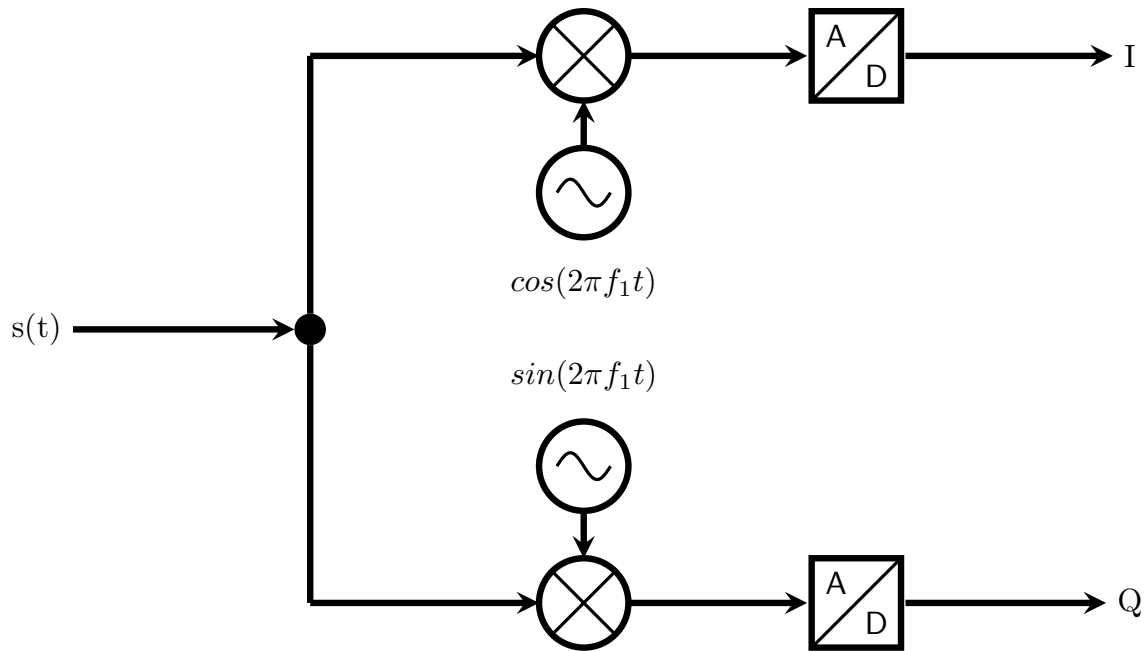
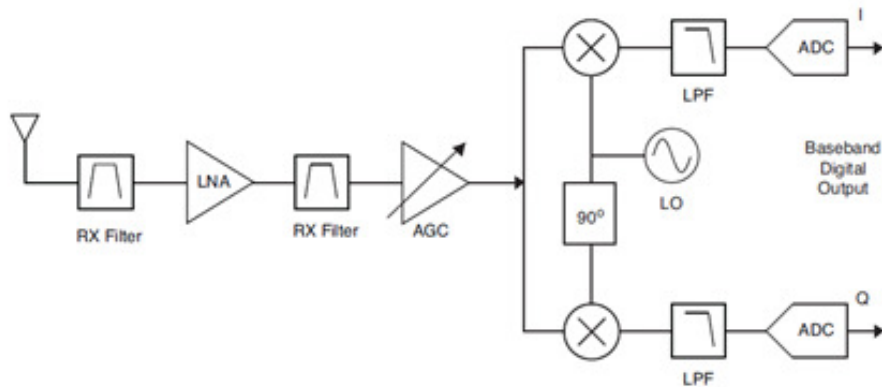


Figure 1.10: Radio receiver block diagram

In Figure 1.10, the D/A blocks should be A/D or ADC.

A more complete drawing of the receiver adds some practical components



In this figure, we have added: RX filters needed to filter out undesired signals on nearby frequencies, a Low Noise Amplifier (LNA) to amplify $r(t)$ that is typically in the microvolt range to a level in the volt range suitable for ADC, Automatic Gain Control (AGC) to adjust the gain to compensate for variations in the level $r(t)$ and low pass filters (LPF) before the ADC.

To see how this receiver works, we calculate the signals $x(t), y(t)$ at the two ADC inputs, and find that they are equal to $I(t), Q(t)$

Exercise: prove this. Several trigonometric identities will be needed:

$$\begin{aligned}\cos \alpha \cos \beta &= [\cos(\alpha - \beta) + \cos(\alpha + \beta)]/2 \\ \sin \alpha \cos \beta &= [\sin(\alpha - \beta) + \sin(\alpha + \beta)]/2 \\ \sin \alpha \sin \beta &= [\cos(\alpha - \beta) - \cos(\alpha + \beta)]/2\end{aligned}$$

Note that double frequency terms (cosine waves at $2f_c$) arise from the calculations and are filtered out by the low pass filter (LPF).

1.2.3 Summary

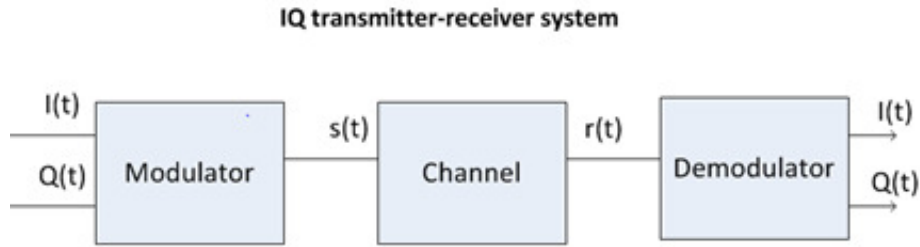
We have created a communication system with message signals $I(t), Q(t)$ that are modulated onto a carrier wave at frequency f_c to create the radio signal

$$s(t) = I(t) \cos 2\pi f_c t - Q(t) \sin 2\pi f_c t.$$

$s(t)$ travels over a distance via a channel with path loss L_o .

The receiver picks up the real passband signal $r(t) = s(t)/L_o$ and recovers the messages $I(t), Q(t)$.

When the communications system is simulated in software, we can implement it in complex passband $r_+(t) = s_+(t)/L_o$



Any signal $s(t)$ can be written as a carrier wave at frequency f_c with time-varying amplitude and phase, i.e.

$$\begin{aligned}s(t) &= a(t) \cos[2\pi f_c t + \phi(t)] \\ &= \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\} \\ &= \text{Re}\{[I(t) + jQ(t)][\cos 2\pi f_c t + j \sin 2\pi f_c t]\} \\ &= I(t) \cos 2\pi f_c t - Q(t) \sin 2\pi f_c t \\ &= a(t) \cos[2\pi f_c t + \phi(t)]\end{aligned}\tag{1.40}$$

where

$$\begin{aligned}I(t) &= a(t) \cos \phi(t) = \text{Re}\{a(t)e^{j\phi(t)}\} \\ Q(t) &= a(t) \sin \phi(t) = \text{Im}\{a(t)e^{j\phi(t)}\}\end{aligned}\tag{1.41}$$

and

$$\tilde{s}(t) = I(t) + jQ(t) = a(t)e^{j\phi(t)}\tag{1.42}$$

is called the complex envelope of the signal and also called the complex baseband signal. The complex envelope contains two real waveforms $I(t), Q(t)$ that contain the information or message.

The real passband signal $s(t)$ is obtained by multiplying the complex envelope $\tilde{s}(t)$ with the complex positive frequency carrier wave

$$c_+(t) = e^{j2\pi f_c t} = \cos 2\pi f_c t + j \sin 2\pi f_c t \quad (1.43)$$

to yield

$$s_+(t) = \tilde{s}(t)e^{j2\pi f_c t} \quad (1.44)$$

and taking the real part to yield

$$s(t) = \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \quad (1.45)$$

Exercise: obtain the real passband signal from $\tilde{s}(t)$ and $c_+(t)$ by adding the negative frequency carrier wave

$$c_-(t) = e^{-j2\pi f_c t} = \cos 2\pi f_c t - j \sin 2\pi f_c t$$

without taking the real part of anything.

1.2.4 Review

Here we review the general I-Q receiver configuration that may be implemented for all types of signals using slightly different notation. Every type of signal consists of a standard configuration coupled with a “map” that is specific to the transmission type.

$$\begin{aligned} s(t) &= a(t) \cos[2\pi f_c t + \psi(t)] \\ &= a(t)[\cos(2\pi f_c t) \cos \psi(t) - \sin(2\pi f_c t) \sin \psi(t)] \\ &= a(t) \cos \psi(t) \cos 2\pi f_c t - a(t) \sin \psi(t) \sin 2\pi f_c t \\ &= v_i(t) \cos 2\pi f_c t - v_q(t) \sin 2\pi f_c t \end{aligned} \quad (1.46)$$

where $a(t)$ is the amplitude and $\psi(t)$ is the phase. The coefficients of the carrier waves in the transmitter are referred to as $v_i(t)$ and $v_q(t)$ respectively, where

$$\begin{aligned} v_i(t) &= a(t) \cos \psi(t) \\ v_q(t) &= a(t) \sin \psi(t) \end{aligned} \quad (1.47)$$

These signals are formed by the map that converts the message $m(t)$ into $v_i(t), v_q(t)$ and used to create $s(t)$ at the IQ transmitter.

$$s(t) = v_i(t) \cos(2\pi f_c t) - v_q(t) \sin(2\pi f_c t) \quad (1.48)$$

The exact form of the map depends on the particular modulation scheme chosen, as will be discussed in later chapters.

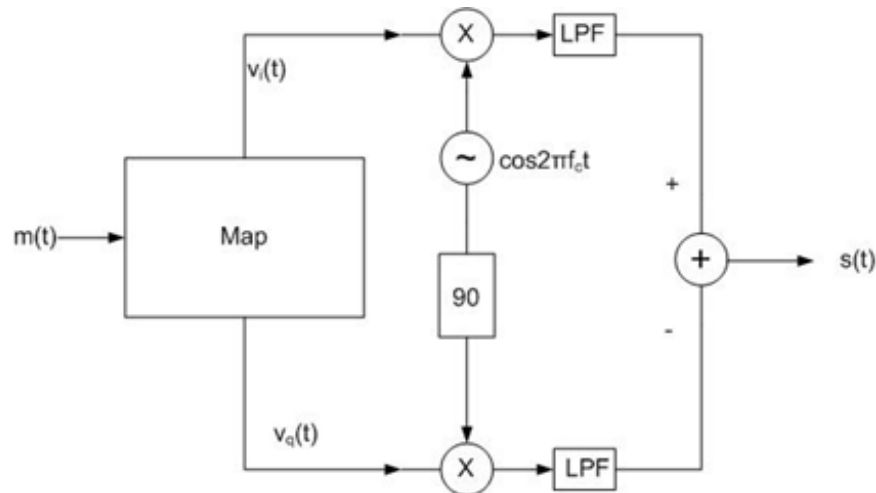


Figure 1.11: General IQ Transmitter

The IQ Receiver multiplies the incoming signal $s(t)$ by two versions of the carrier wave functions: $\cos(2\pi f_c t)$ and the 90 degree phase-shifted version $\sin(2\pi f_c t)$.

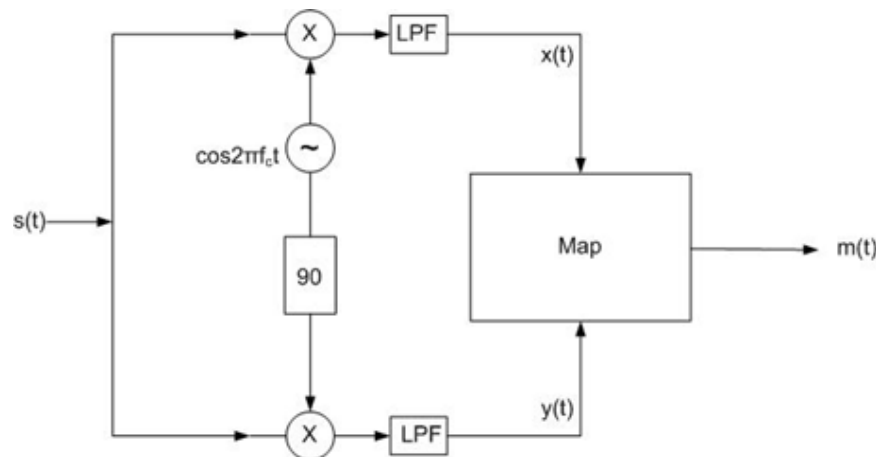


Figure 1.12: General IQ Receiver

Again, the form of the map that converts $v_i(t), v_q(t)$ to the message $m(t)$ depends on the particular modulation scheme chosen, as will be discussed in later chapters.

1.2.5 Messages

When we study modulation, we often choose a simple analog message $m(t) = A_m \cos(2\pi f_m t)$, where f_m is the modulation/message frequency and is usually on the order of Hz or kHz. We can also choose a simple digital message with an 10101010... alternating bit sequence at a bit rate $2f_m$. For this digital message, the 1 bits are the positive half of a cosine wave at frequency f_m and the 0 bits

are the negative half of the same cosine wave. This message is called a single frequency or single tone message. In practice, we wish to transmit a more interesting message that contains more than a single tone at a single frequency.

Analog message (voice or music)

A general analog message (voice or music) would be represented by a sum of cosine waves with different amplitudes and phases A_i, ψ_i at each frequency f_i ,

$$m(t) = \sum_i A_i(t) \cos(2\pi f_i t + \psi_i(t)). \quad (1.49)$$

We often assume that $m(t)$ is divided into frames of length T in the range 5-20 milliseconds. The frames may simply follow one after the other without overlap. In each frame the values of A_i, f_i, ϕ_i will in general be different.

Frames may also be windowed and overlapped. In the figure, we show each frame of length T as being N samples long, where N is typically a power of 2 in the range 256 to 2048 for audio signals sampled at an audio rate of say $N/T = 48\,000$ Hz. The figure also shows the window shape (raised cosine) and the overlap ($N/2$ samples).

Exercise: what is the length in milliseconds of an $N=256$ sample frame?

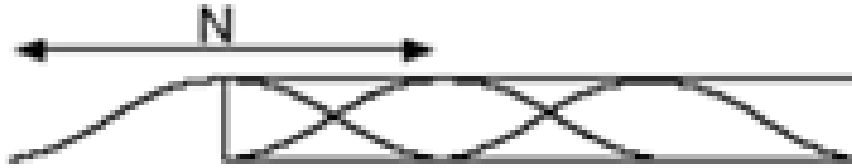


Figure 1.13: Frame length, N

During each frame k at time $t = kT$, we assume $A_i(t) = A_{i,k}, \psi_i(t) = \psi_{i,k}$ are constant, so we can write

$$m(t = kT) = \sum_i A_{i,k} \cos(2\pi f_i t + \psi_{i,k}). \quad (1.50)$$

At the discrete times $t = kT$, the message is the sum of cosine waves with frequencies f_i amplitudes $A_{i,k}$ and phases $\psi_{i,k}$ that are different for each frame k . For each frame k , the amplitudes $A_{i,k}$ and phases $\psi_{i,k}$ at the frequencies f_i can be obtained from the short-time Fourier transform of $m(t)$.

Digital message, a sequence of bits

An analog message can also represent a digital symbol sequence a_k by writing

$$m(t) = \sum_k a_k p(t - kT) \quad (1.51)$$

where $p(t)$ is a pulse that spans a finite time period, a_k may be binary symbol ± 1 (to represent binary 1 or 0) or multilevel (e.g. $\pm 1, \pm 3$ to represent 00, 01, 10, 11) and T is the symbol time.



We can have two such sequences:

$$\begin{aligned} m_I(t) &= \sum_k a_k p(t - kT) = I(t) \\ m_Q(t) &= \sum_k b_k p(t - kT) = Q(t) \end{aligned} \quad (1.52)$$

Thus we can write the complex baseband signal as

$$I(t) + jQ(t) = \sum_k (a_k + jb_k) p(t - kT) \quad (1.53)$$

where

$$c_k = a_k + jb_k = r_k e^{j\phi_k} \quad (1.54)$$

is a complex data symbol. If both a_k, b_k are binary, then c_k represents 2 bits of information at each time $t = kT$. If both a_k, b_k are multilevel $\pm 1, \pm 3$, then c_k represents 4 bits of information at each time $t = kT$.

The pulse shape $p(t)$ can have shapes other than rectangular, and can be longer than T . Other pulse shapes are described in a later chapter.

1.3 Analog IQ receiver

Learning objectives

- To describe the operation of a practical software defined radio.
- To describe the implementation of a receiver that receives real analog signals and processes them both in the analog and digital domain.
- To describe the connection between the block diagram of a software defined radio and the mathematical description of communications signals

1.3.1 Overview and block diagram

In this section, we apply the I-Q signal theory above to a particular example:

A block diagram for the Ettus USRP N210 radios used in the ELEC 350 labs is shown in fig. 1.14. For this receiver, there is an analog daughterboard that is chosen for the particular frequency range of interest (red and green lines) and a digital motherboard.

the daughterboard receives signals with carrier frequency in the range 50–2200 MHz and generates analog I-Q outputs that are in the frequency range below 50 MHz and are sampled at 100 MHz.

In fig. 1.14:

- The red blocks and connections represent real passband signals
- the green blocks and connections represent the real and imaginary parts of complex baseband signals
- The blue blocks and connections represent digital processing in the software defined part of the radio receiver.

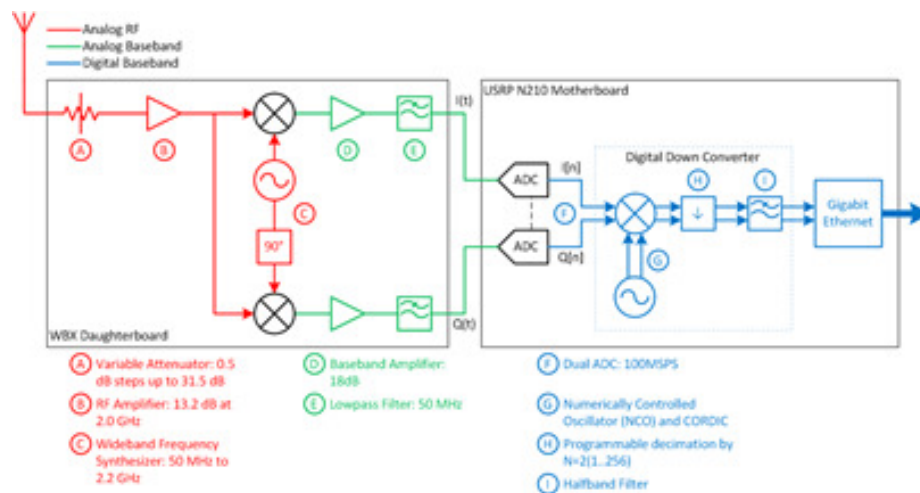


Figure 1.14: Ettus N210 block diagram (www.ettus.com)

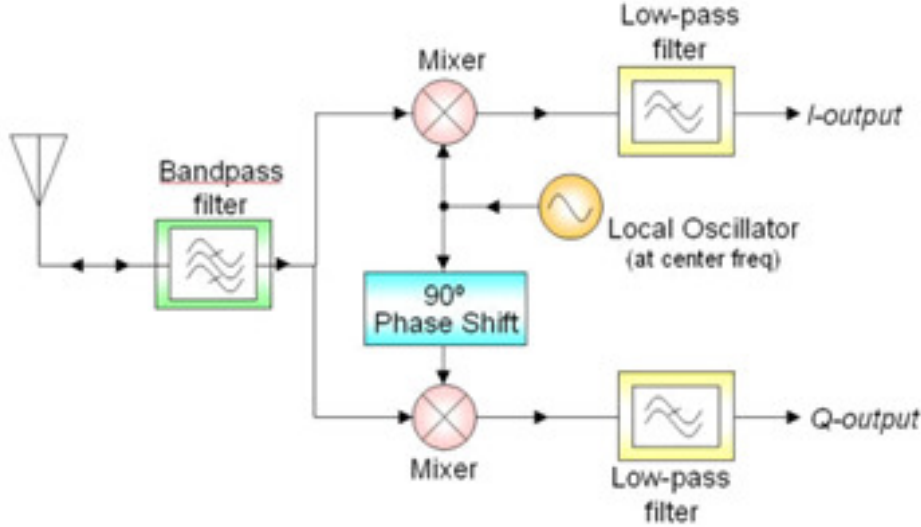
Many software defined radios that operate at frequencies above about 100 MHz have an analog I-Q stage similar to this example ahead of the analog-to-digital converter sampling at 100 MHz. The red oscillator (c) operates at a frequency f_{LO} called the local oscillator frequency, which is general is close to but not the same as the carrier frequency f_c .

In the next section, we describe the operation of the daughterboard (red and green blocks and connections).

1.3.2 USRP daughterboard receiver operation

The USRP daughterboard is designed to receive radio frequency (RF) signals at frequencies f_c in the range $f_{LO} \pm f_s/2$ MHz, where f_{LO} is the local oscillator (LO) frequency. In this example, we set f_{LO} to about 100 MHz and f_s is the sampling rate of the USRP main board ADCs, set to 100 Msps.

The USRP daughterboard operates by generating two local oscillator (LO) signals at f_{LO} and mixing (multiplying) it with a desired radio frequency (RF) signal at f_c (picked up by the antenna or fed in by a signal generator) to yield a signal at the difference frequency $f_b = f_c - f_{LO}$.



The USRP daughterboard has two local oscillators operating 90 degrees out of phase, $\cos 2\pi f_{LO}t$ and $-\sin 2\pi f_{LO}t$ and two mixers. Thus there are two receiver outputs that we call $I(t)$ and $Q(t)$ (see figure above)

The desired RF signal that we wish to receive is written $r(t) = a(t) \cos[2\pi f_c t + \phi(t)]$.

We assume $a(t) = 1$, $\phi(t) = 0$ for the moment, so the desired RF signal is simply an unmodulated carrier wave $r(t) = \cos 2\pi f_c t$.

1.3.3 Receiver operation in real notation

In what follows, given this RF signal input we will calculate the two receiver outputs $I(t)$ and $Q(t)$.

To do this, we will use some trigonometric identities

$$\cos \alpha \cos \beta = [\cos(\alpha - \beta) + \cos(\alpha + \beta)]/2$$

$$\sin \alpha \cos \beta = [\sin(\alpha - \beta) + \sin(\alpha + \beta)]/2$$

$$\sin \alpha \sin \beta = [\cos(\alpha - \beta) - \cos(\alpha + \beta)]/2$$

One of the local oscillator signals is written $\cos 2\pi f_{LO}t$.

The cos mixer multiplies this LO signal with the RF input signal to obtain

$$\cos 2\pi f_c t \cdot \cos 2\pi f_{LO}t. \quad (1.55)$$

Using one of the trigonometric identities, we can write

$$\cos 2\pi f_{LO}t \cdot \cos 2\pi f_c t = 0.5 \cos 2\pi(f_c + f_{LO})t + 0.5 \cos 2\pi(f_c - f_{LO})t \quad (1.56)$$

Thus multiplying two sine (or cosine) waves at frequencies f_{LO} and f_c results in two new sine waves: one at the sum frequency $f_c + f_{LO}$ and one at the difference frequency $f_c - f_{LO}$.

The signal at the sum frequency $f_c + f_{LO}$ is filtered out by analog low pass filters in the USRP daughterboard.

The signal at the difference frequency $f_c - f_{LO}$ is written

$$I(t) = 0.5 \cos 2\pi(f_c - f_{LO})t = 0.5 \cos 2\pi f_b t. \quad (1.57)$$

$I(t)$ can be sampled by the ADC, provided that f_c is close enough to f_{LO} , i.e. the difference is less than half the sampling rate, $|f_c - f_{LO}| < f_s/2$ or $f_{LO} - f_s/2 < f_c < f_{LO} + f_s/2$

The difference frequency $f_b = f_c - f_{LO}$, where $|f_b| < f_s/2$.

The USRP daughterboard has two local oscillators operating 90 degrees out of phase, $\cos 2\pi f_{LO}t$ and $-\sin 2\pi f_{LO}t$ and two mixers. Thus there are two outputs that we call $I(t)$ and $Q(t)$ (see figure above) where $I(t) = 0.5 \cos 2\pi(f_c - f_{LO})t$ was calculated above.

The second local oscillator signal is written $-\sin 2\pi f_{LO}t$. The sin mixer multiplies this LO signal with the RF input signal to obtain

$$\cos 2\pi f_c t \cdot (-\sin 2\pi f_{LO}t). \quad (1.58)$$

We apply a trigonometric identity with $\alpha = 2\pi f_{LO}t$, $\beta = 2\pi f_c t$ and write

$$-\sin 2\pi f_{LO}t \cdot \cos 2\pi f_c t = -0.5 \sin 2\pi(f_c + f_{LO})t + 0.5 \sin 2\pi(f_c - f_{LO})t \quad (1.59)$$

The signal at the difference frequency is written

$$Q(t) = 0.5 \sin 2\pi(f_c - f_{LO})t = 0.5 \sin 2\pi f_b t. \quad (1.60)$$

If the receiver outputs $I(t) = \cos 2\pi f_b t$ and $Q(t) = \sin 2\pi f_b t$ are displayed on a $x-y$ scope, then a circle is displayed. If $f_b < 5$ Hz or so, then the dot on the scope can be seen tracing out the circle. We can write these two receiver output signals $I(t)$ and $Q(t)$ that have time varying amplitude and phase $a(t)$, $\phi(t)$, where in this case

$$\begin{aligned} a(t) &= \sqrt{I^2(t) + Q^2(t)} \\ &= 1 \\ \phi(t) &= \arctan \frac{Q(t)}{I(t)} \\ &= \arctan \frac{\sin 2\pi f_b t}{\cos 2\pi f_b t} \\ &= \arctan(\tan 2\pi f_b t) \\ &= 2\pi f_b t \\ \tilde{r}(t) &= a(t)e^{j\phi(t)} \\ &= e^{j2\pi f_b t} \end{aligned} \quad (1.61)$$

The last equation shows how $I(t)$ and $Q(t)$ can also be represented as one complex signal $\tilde{r}(t) = I(t) + jQ(t) = a(t)e^{j\phi(t)}$ as shown in the next section.

1.3.4 Receiver operation in complex notation

We can calculate the USRP daughterboard receiver output using complex signals as follows.

Recall that the desired RF signal that we wish to receive is written $r(t) = a(t) \cos[2\pi f_c t + \phi(t)]$ where for the moment we set the path loss $L_0 = 1$.

We assume $a(t) = 1, \phi(t) = 0$ for the purpose of this section, so the desired RF signal is simply an unmodulated carrier wave $\cos 2\pi f_c t$. In complex notation we write

$$r(t) = \text{Re}\{r_+(t)\} = \text{Re}\{e^{j2\pi f_c t}\} = \cos 2\pi f_c t. \quad (1.62)$$

We can also write $r(t)$ in terms of the positive and negative frequency components.

$$r(t) = r_+(t)/2 + r_-(t)/2 \quad (1.63)$$

In complex notation, we consider only the positive frequency signal input to the receiver. The complex passband unmodulated RF signal $r_+(t) = e^{j2\pi f_c t}$ is multiplied by the complex local oscillator

$$e^{-j2\pi f_{LO} t} = \cos 2\pi f_{LO} t - j \sin 2\pi f_{LO} t \quad (1.64)$$

to yield

$$r_+(t)e^{-j2\pi f_{LO} t} = e^{j2\pi f_c t} e^{-j2\pi f_{LO} t} = e^{j2\pi f_b t} = I(t) + jQ(t) = \tilde{r}(t) \quad (1.65)$$

where the received complex baseband signal is

$$\tilde{r}(t) = I(t) + jQ(t) = e^{j2\pi f_b t}. \quad (1.66)$$

The diagram in fig. 1.15 below using complex signals performs the same function as the previous diagram above using real signals. Note that the complex signal diagram does not use the low pass filters.

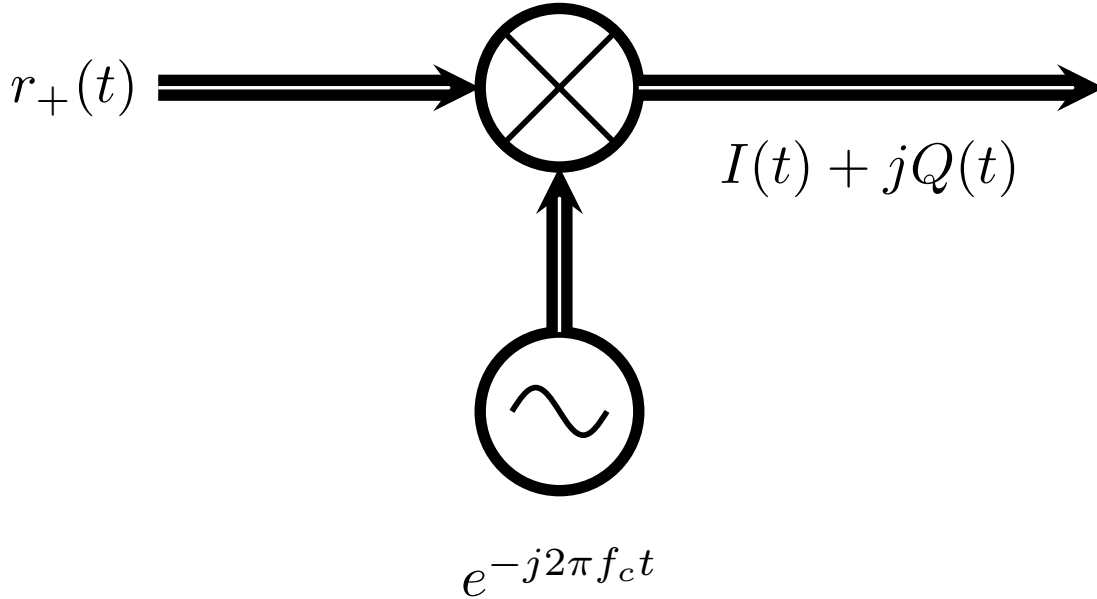


Figure 1.15: Complex signal receiver block diagram, USRP daughterboard analog IQ receiver in complex notation, in this diagram $f_c = f_{LO}$

In summary,

$$\begin{aligned}
 r_+(t) &= e^{j2\pi f_c t} \\
 I(t) &= \cos 2\pi f_b t \\
 Q(t) &= \sin 2\pi f_b t \\
 \tilde{r}(t) &= I(t) + jQ(t) \\
 &= \cos 2\pi f_b t + j \sin 2\pi f_b t \\
 &= e^{j2\pi f_b t}
 \end{aligned}$$

This is the same result as above, apart from a factor 0.5 arising from the complex notation.

If $I(t)$ and $Q(t)$ are displayed on a x - y scope, a circle is displayed. If $f_b < 5$ Hz or so, then the dot on the scope can be seen tracing out the circle.

Exercise: Repeat the above analysis in both real and complex notation when the RF input signal is a general signal $a(t) \cos[2\pi f_c t + \phi(t)]$ to find $I(t)$ and $Q(t)$ as a function of $f_b, a(t), \phi(t)$.

Exercise: Repeat the complex notation analysis for a real input signal $r(t) = \cos 2\pi f_c t$. Hint: write $r(t)$ as the sum of two complex exponentials. Some filters may be needed.

1.3.5 Another approach using digital IQ receiver

Figure 1.16 shows an alternative receiver structure where the analog stage has only one cosine oscillator followed by a low pass filter. In this structure, the single analog oscillator shifts the received bandwidth down by f_{LO} so that it is centered at 0 Hz. The signals $I(t)$, $Q(t)$ are generated after the analog-to-digital converter using exactly the same math as shown in section 1.3.3 and 1.3.4.

In the next section, we consider IQ signals in the frequency domain.

1.4 Spectrum Analyzers, analog and digital

While an oscilloscope is used to display electrical signals in an amplitude versus time format, a spectrum analyzer provides an amplitude squared (power) versus frequency format. fig. 1.17 shows the two representations for the same signal.

The Fourier analysis introduced in section 1.1.5 shows that any complex waveform can be resolved into sinusoidal waveforms of a *fundamental frequency* and a number of *harmonic frequencies*. The spectrum analyzer effectively performs the Fourier integral:

$$\begin{aligned}
 S(f) &= \int_{t=-\infty}^{\infty} s(t) e^{-j2\pi f t} dt \\
 &= \int_{t=-\infty}^{\infty} s(t) \cos(2\pi f t) dt - j \int_{t=-\infty}^{\infty} s(t) \sin(2\pi f t) dt
 \end{aligned} \tag{1.67}$$

The integral finds the frequency components in $s(t)$ by correlating $s(t)$ with cosine and sine waves at each frequency f . For a particular frequency $f = f_c$, $s(t) = \cos(2\pi f_c t)$ and

$$\begin{aligned}
 S(f_c) &= \int_{-\infty}^{\infty} \cos(2\pi f_c t) \cos(2\pi f_c t) dt \\
 &= \int_{-\infty}^{\infty} 0.5[1 + \cos(4\pi f_c t)] dt
 \end{aligned} \tag{1.68}$$

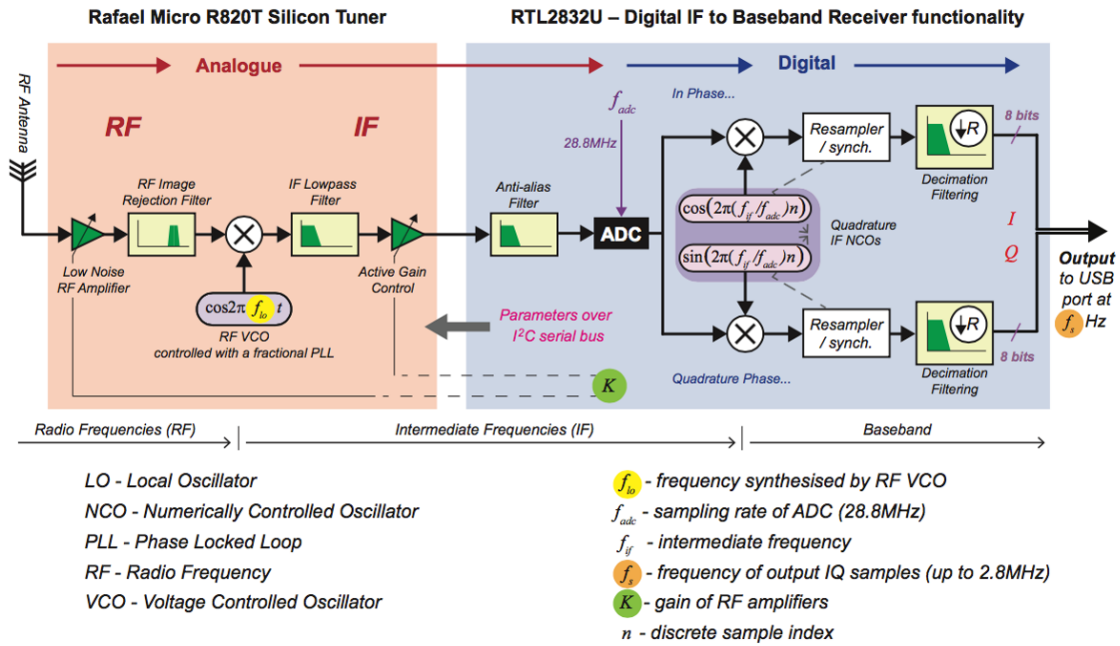


Figure 1.16: RTL-SDR block diagram (<https://www.rtl-sdr.com/about-rtl-sdr/>)

The cosine waves are in phase for all time, the product of the two cosine waves contains DC, thus the integral integrates DC over all time, resulting in infinity (delta functions) at $f = f_c$. For all other frequencies $f \neq f_c$ the cosine waves drift in and out of phase over time, there is no DC component, and the integral is zero. Thus for $s(t) = \cos(2\pi f_c t)$ we can write:

$$\begin{aligned}
 S(f) &= \int_{-\infty}^{\infty} \cos(2\pi f_c t) \cos(2\pi f t) dt \\
 &= [\delta(f - f_c) + \delta(f + f_c)]/2
 \end{aligned}
 \tag{1.69}$$

For a digital spectrum analyzer where the signal is digitized (sampled) by an analog-to-digital converter (ADC), the Fourier transform is done using the Fast Fourier Transform (FFT) algorithm, as will be discussed in more detail in the next section.

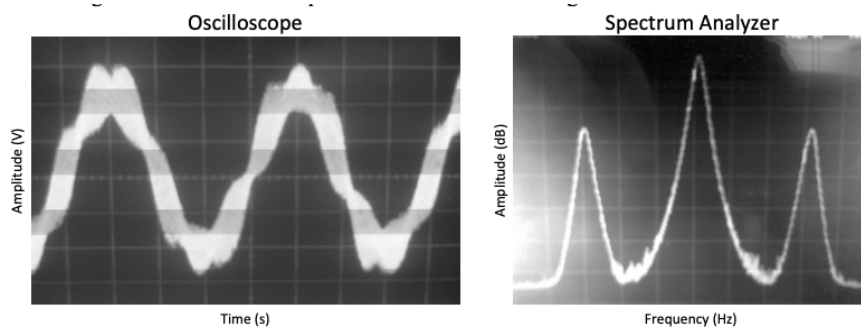


Figure 1.17: Oscilloscope and Spectrum Analyzer Displays

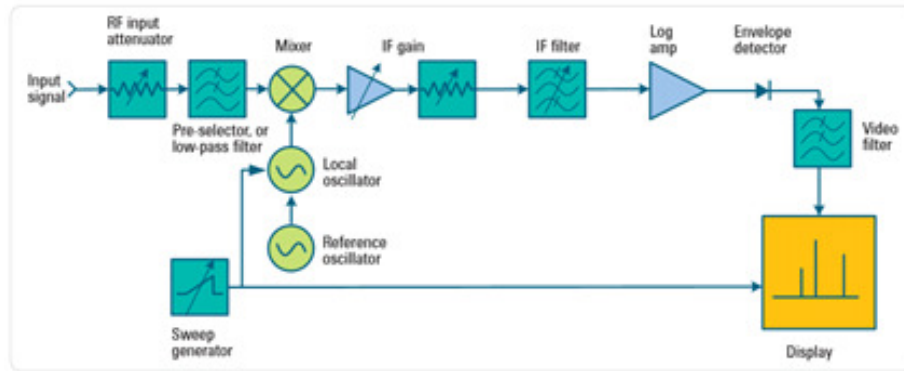


Figure 1.18: Block diagram of spectrum analyzer (<http://cp.literature.agilent.com/litweb/pdf/5952-0292.pdf>)

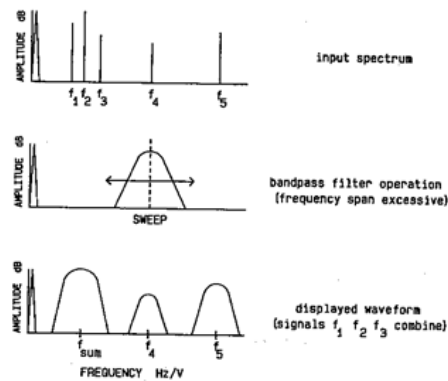


Figure 1.19: Effect of Span on Spectrum Analyzer Display

There is a practical upper limit to the sampling rate of an ADC. Thus analog spectrum analyzers are used for high frequencies. For an analog spectrum analyzer, frequency scanning is accomplished by electronically tuning a bandpass filter network across the desired frequency range. In practice, a sweep generator along with a fixed frequency bandpass filter, rather than tuning the bandpass filter, as shown in fig. 1.18 (taken from an Agilent app note).

The amplitudes of all the signals in the bandpass area (at each point during the scan) are measured to provide the amplitude versus frequency display.

It is important to note that in order to detect two signals which are narrowly spaced, the *frequency span* of the bandpass filter must be set appropriately. This concept is shown in fig. 1.19. The frequencies f_1 , f_2 and f_3 in the input spectrum are summed into a single peak on the displayed waveform due to the excessive frequency span. Note that the DC reference shown in this figure is a characteristic of the spectrum analyzer. It is always present, regardless of whether there is a DC bias on the input signal.

Unlike an oscilloscope, the amplitude scale on a spectrum analyzer can be set to a logarithmic scale (for power measurements) as well as a linear scale (for voltage measurements). The logarithmic

scale is represented in dBm, or decibels with reference to a milliwatt.

$$P_{(dBm)} = 10 \log \left(\frac{P_{in}}{1 \text{ mW}} \right) \quad (1.70)$$

It can then be shown that

$$P_{(dBm)} = 10 \log \left(\frac{V^2}{Z} \cdot \frac{1}{1 \text{ mW}} \right) \quad (1.71)$$

where V is the RMS voltage.

If the load is a standard 50Ω input,

$$\begin{aligned} P_{(dBm)} &= 20 \log V + 10 \log \frac{1}{50 \times 10^{-3}} \\ &= 20 \log V + 13 \end{aligned}$$

So the relationship between input power and input voltage is

$$P_{(dBm)} = 20 \log V + 13 \quad (1.72)$$

This is a handy formula worth remembering.

On the logarithmic dB scale, the power of signals is always positive. It is the magnitude of the signals above (in reference to) the noise level that is of interest. The analog spectrum analyzer shows only the square magnitude (or power) of each Fourier component. A digital spectrum analyzer can show phase as well as amplitude. The block diagram of the digital spectrum analyzer is the same as the general I-Q receiver, followed by an ADC (analog-to-digital converter), FFT processor and display.

1.5 Spectrogram or spectral waterfall

1.5.1 Example spectrograms

Radio signals (and other types of one dimensional signals) can be displayed with a *spectrogram* that represents the signal simultaneously in the time and frequency domains. A spectrogram has 3 dimensions: time, frequency and amplitude.

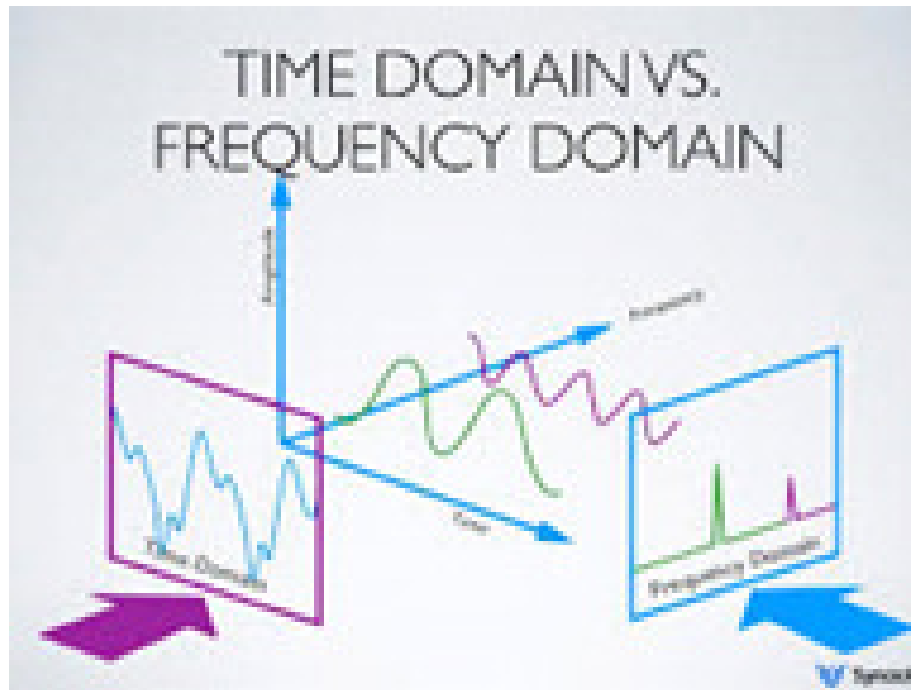
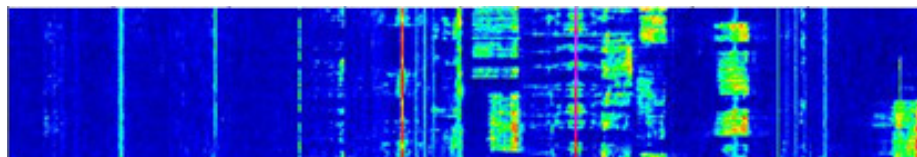


Figure 1.20: Time vs Frequency domain

Often the amplitude is represented by colour, just like a topographical map. The amplitude is colour coded using a colour map, where (for example) blue is low amplitude, green is medium amplitude and red is high amplitude. A *waterfall* display is a spectrogram where the time axis is moving, showing the most recent signal first.

Figure 1.21 shows an example snapshot in time of a spectrogram or moving spectral waterfall display. The x axis is frequency, the y axis is time (most recent at the top, time increasing downwards). On the left of the display, the blue line represents a sine wave at a fixed frequency. The green blocks represent radio signal modulated with information, and thus occupy a finite non-zero bandwidth. Note that the radio signals on a given frequency start and stop at different times.

Figure 1.21: Waterfall spectrogram with frequency on x axis, time on y axis, amplitude color coded to display z axis.

Another example of a waterfall spectrogram is shown in fig. 1.22. In this example, all the signals are morse code, i.e. sine waves switched on and off. The amplitude is grey scale coded to display the z axis amplitude, a tunable filter shown in green is used to select desired signal.

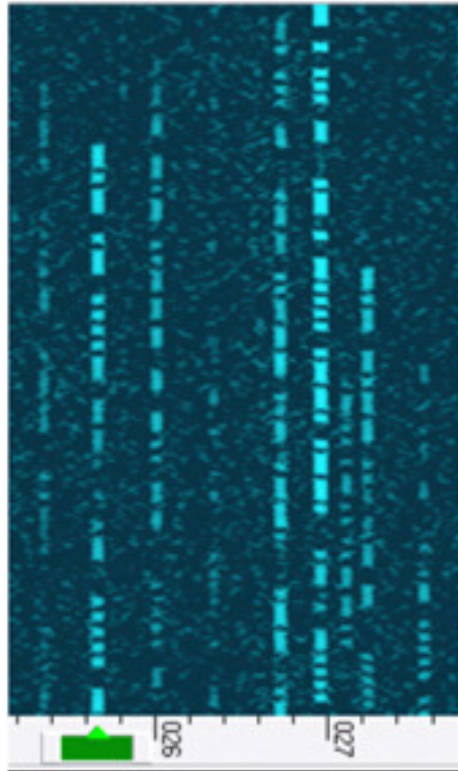


Figure 1.22: Waterfall spectrum of morse code

An example of a combination spectrum and spectral waterfall is shown in fig. 1.23. The spectrum is the output of a digital spectrum analyzer and is valid in real time. The waterfall shows the history of the spectrum with frequency on the horizontal axis, time on the vertical axis and amplitude represented by color. In fig. 1.23, there are two signals present.

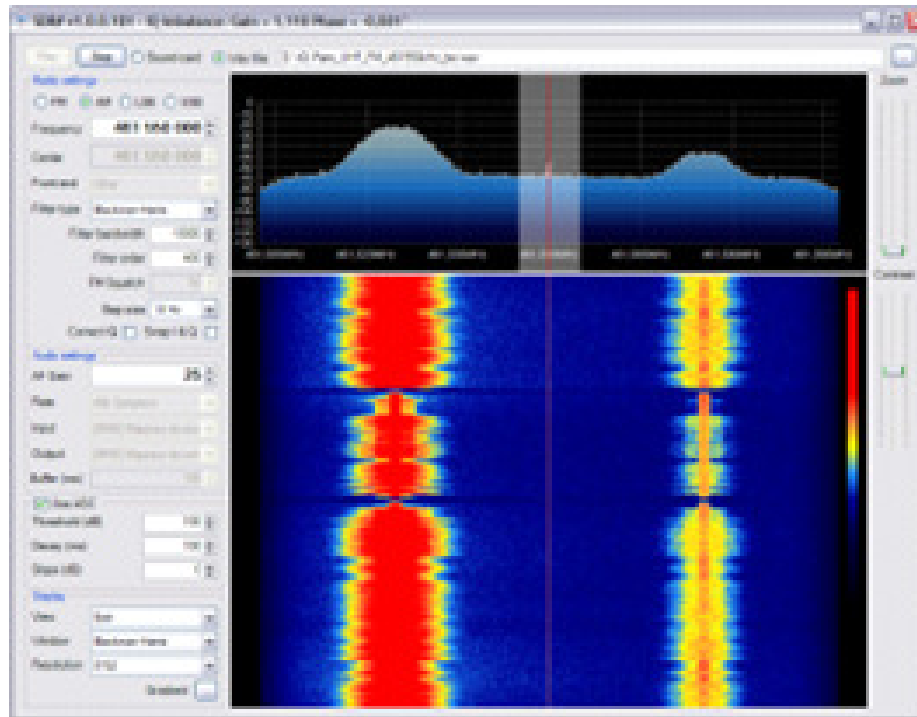


Figure 1.23: Combined spectrum and spectral waterfall

Live real-time spectral waterfalls showing radio signals in real time are available at these links: <http://www.websdr.org/> and <http://kiwisdr.com/public/>

Algorithm to generate a spectrogram

The waterfall is drawn as follows:

1. Segment the complex baseband time domain signal $\tilde{s}(t)$ sampled at a given rate f_s samples per second into frames of a fixed number of samples N , where N is typically a power of two such as 1024. Each frame is of length $T_0 = N/f_s$ seconds.
2. Take the short-time Fourier transform of a frame $\tilde{S}(f) = FFT\{\tilde{s}(t)\}$ to yield a frame of N complex samples representing the amplitude and phase versus frequency. The frequency resolution will be $f_0 = f_s/N$ Hz.
3. Draw one horizontal line at the top of the spectrogram with the color of each sample chosen to represent the amplitude. For example, blue for low amplitude and red for high amplitude and all colors in between.
4. Push all other lines in the spectrogram down by one time unit (one frame time).
5. Repeat from Step 2 for the next segment (frame).

The spectrum for a particular frame is found by computing $\tilde{S}(f) = FFT\{\tilde{s}(t)\}$ for the samples of $\tilde{s}(t)$ in that frame. The spectrum extends from $-f_s/2$ to $f_s/2$. Since $\tilde{s}(t)$ is complex, the spectrum

$\tilde{S}(f)$ is not symmetrical around zero, thus the spectrum will have a different shape for positive and negative frequencies.

The frequency axis zero point is often labeled f_{LO} in the waterfall plot instead of zero, since an RF signal at exactly $f_c = f_{LO}$ will be mixed by the IQ receiver down to $f_c - f_{LO} = f_b = 0$ Hz. In this way, the frequency axis displays signals at their actual RF frequencies within a bandwidth f_s from $f_{LO} - f_s/2$ to $f_{LO} + f_s/2$. Thus the maximum bandwidth of a complex signal sampled at f_s is f_s (not $f_s/2$ as for a real signal).

For a real signal sampled at f_s the bandwidth is $f_s/2$ and the spectrum is symmetrical about zero frequency. The spectrum shape at negative frequencies from $-f_s/2$ to 0 will be the mirror image of the spectrum shape from 0 to $f_s/2$.

The above algorithm is normally refined using frames that are windowed and overlapped (in the same way as the message frames in section 1.2.5). In the figure, we show each frame of length T_0 as being N samples long, where N is typically a power of 2 in the range 256 to 2048 for audio signals sampled at an rate equal to the desired bandwidth of the spectrogram display. The figure also shows the window shape (raised cosine) and the overlap ($N/2$ samples).

Exercise: what is the length in milliseconds of an $N = 256$ sample frame?

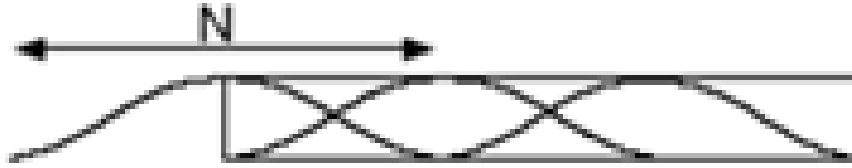


Figure 1.24: Frame length, N

Live waterfall displays with tunable filter and audio output can be found at <http://www.websdr.org/> and <http://kiwisdr.com/public/>

Exercise: write the algorithm for creating a spectrogram waterfall showing the signal samples of $\tilde{s}(t)$ and $\tilde{S}(f) = FFT\{\tilde{s}(t)\}$ in discrete-time notation where $\tilde{s}_n = \tilde{s}(t = nT_s)$, $T_s = 1/f_s$, $\tilde{S}_k = \tilde{S}(f = kf_0)$. Recall the discrete Fourier transform is defined as a Fourier transform operating on a sampled periodic signal:

$$\begin{aligned} \tilde{S}_k &= \int_{t=0}^{T_0=nT_s} \tilde{s}(t) e^{-j2\pi ft} dt \Big|_{t=nT_s, f=kf_0} \\ &= \sum_{n=0}^{N-1} \tilde{s}_n e^{-j2\pi k f_0 n T_s} \\ &= \sum_{n=0}^{N-1} \tilde{s}_n e^{-j2\pi nk/N} \end{aligned}$$

where $T_s f_0 = 1/N$.

1.6 IQ imbalance

In practice, the two local oscillators of the USRP daughterboard analog IQ receiver are such that $Q(t)$ is not exactly the same amplitude as $I(t)$ and not exactly 90 degrees out of phase with $I(t)$. This is called IQ imbalance. The practical result is that in the spectrum analyzer or waterfall view, the signals at positive frequencies will have weak images at the corresponding negative frequencies, and visa versa. These images are not really there, they are an artifact of the IQ imbalance.

If there is IQ imbalance present, then the amplitude and phase of $I(t)$ is shifted relative to what it should be by a complex factor $(1 + \alpha)e^{j\theta}$ for small values of α, θ . We write the complex output $\tilde{r}'(t)$ of the imperfect receiver with IQ imbalance as

$$\tilde{r}'(t) = I'(t) + jQ'(t) \quad (1.73)$$

$$= (1 + \alpha)e^{j\theta}I(t) + jQ(t) \quad (1.74)$$

$$I'(t) = (1 + \alpha)\cos\theta I(t) \quad (1.75)$$

$$Q'(t) = Q(t) + (1 + \alpha)\sin\theta I(t) \quad (1.76)$$

For small theta, this matches the approximation

$$\begin{aligned} Re\{out\} &= Re\{in\}(1 + \text{Magnitude}), \\ Im\{out\} &= Im\{in\} + Re\{in\} \times \text{Phase} \end{aligned} \quad (1.77)$$

where we define

$$\begin{aligned} Re\{out\} &= I'(t), \\ Re\{in\} &= I(t), \\ Im\{out\} &= Q'(t), \\ Im\{in\} &= Q(t), \\ \text{Magnitude} &= \alpha, \\ \text{Phase} &= \theta. \end{aligned}$$

In the frequency domain, IQ imbalance appears in the form of an image, i.e. every signal at frequency f_b will have a mirror image at $-f_b$. For a signal at f_b , the desired complex baseband signal is $\tilde{r}(t) = \tilde{s}(t) = e^{j2\pi f_b t}$, but with IQ imbalance, the actual observed complex baseband signal is

$$\tilde{r}'(t) = \mu e^{j2\pi f_b t} + \nu e^{-j2\pi f_b t} \quad (1.78)$$

with complex constants μ, ν , thus explicitly showing the desired signal at f_b and the mirror image at $-f_b$. The complex constants μ, ν are functions of α, θ . When $\alpha = \theta = 0, \nu = 0$.

In general, we can write the received complex signal obtained at the output of an imperfect IQ receiver as

$$\tilde{r}'(t) = \mu\tilde{r}(t) + \nu\tilde{r}^*(t) \quad (1.79)$$

where $\tilde{r}'(t)$ is the receiver output, $\tilde{r}(t) = \tilde{s}(t)$ is the desired complex signal, and $\tilde{r}^*(t)$ is the image. The image rejection is $20 \log |\nu/\mu|$ dB.

The mirror image can be nulled out by multiplying $Q'(t)$ by the inverse of this complex factor $(1 + \alpha)e^{j\theta}$, i.e. $\frac{1}{1+\alpha}e^{-j\theta} = \beta e^{j\psi}$. The values of the magnitude β and phase ψ can be found manually (with hardware or software controls) or by an adaptive algorithm.

Exercise: Verify the expressions for the imperfect receiver outputs $I'(t)$ and $Q'(t)$ given in section 4 above.

Exercise: Verify the approximate expressions for IQ imbalance correction given in Section 4 above.

1.7 Radio tuning, selecting a particular signal

The USRP receiver (section 1.3, figure fig. 1.14) is designed to receive radio frequency (RF) signals at any frequency f_c in the range $f_{LO} \pm f_s/2$ MHz, where f_{LO} is the local oscillator (LO) frequency. In this example, consider the complex signal at the output of the USRP (that is connected to the computer via Ethernet and processed in software such as GNURadio).

The GNURadio block diagram software includes a so-called ‘‘USRP source block’’ that controls the analog oscillator frequency shown in red on the USRP daughter board as well as the NCO frequency on the USRP motherboard shown in blue. The value of f_{LO} set in the USRP source block determines the frequency of both oscillators f_{red} and f_{blue} such that $f_{LO} = f_{red} + f_{blue}$. When selecting the frequency f_{LO} in the USRP source block, the values of f_{red} and f_{blue} are set automatically by the USRP hardware, so that we don’t need to think about them.

The USRP source block has a complex signal output and has the LO frequency and sampling rate as parameters. This complex signal can be live from the USRP source block output or a IQ file in 2 channel WAV format that was recorded previously. Note that this LO frequency in the GNURadio software is not the same as the LO frequency in the USRP daughterboard mentioned in an earlier section.

In this example, we choose $f_{LO} = 102$ MHz and $f_s = 2$ MHz, so that the frequency range on the spectrum or waterfall is 101–103 MHz, which is part of the FM broadcast band. Selecting a signal (radio station) at a particular frequency f_c is called ‘‘tuning’’ the radio. In this example we want to tune in the radio station at 101.9 MHz.

The USRP source block operates by generating two local oscillator signals at f_{LO} and mixing (multiplying) it with a desired radio frequency (RF) carrier wave $r_+(t) = e^{j2\pi f_c t}$ at f_c to yield a complex baseband signal $\tilde{r}(t) = I(t) + jQ(t)$ at the difference frequency $f_b = f_c - f_{LO}$, where we write

$$\begin{aligned} r_+(t)e^{-j2\pi f_{LO}t} &= e^{j2\pi f_c t} e^{-j2\pi f_{LO}t} \\ &= e^{j2\pi f_b t} \\ &= \cos 2\pi f_b t + j \sin 2\pi f_b t \\ &= I(t) + jQ(t) \end{aligned} \tag{1.80}$$

$I(t)$ and $Q(t)$ are processed correctly, provided that f_c is close enough to f_{LO} , i.e. the difference is less than half the sampling rate, $|f_c - f_{LO}| < f_s/2$ or $f_{LO} - f_s/2 < f_c < f_{LO} + f_s/2$. The difference frequency $f_b = f_c - f_{LO}$, where $|f_b| < f_s/2$.

The USRP source block function is to shift a 2 MHz wide slice of spectrum from 101–103 MHz centered at $f_{LO} = 102$ MHz down to -1 – 1 MHz (positive and negative frequencies centered around 0 Hz). The complex baseband signal $\tilde{r}_1(t) = e^{j2\pi f_b t}$ can represent positive and negative frequencies, since f_b can be positive or negative and $|f_b| < 1$ MHz. The 1 MHz slice of spectrum may contain many different signals at various frequencies within the 1 MHz range (recall waterfall plots mentioned

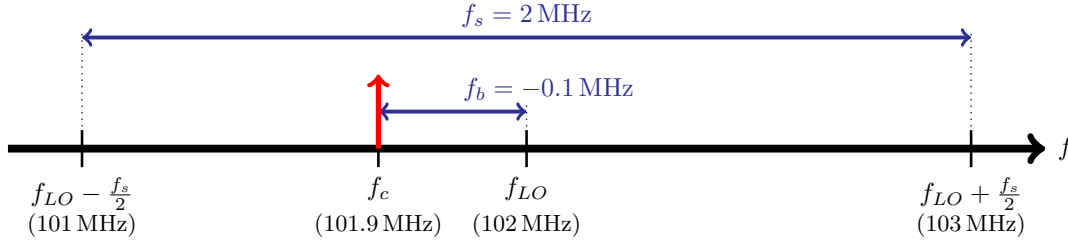


Figure 1.25: Frequency tuning example in passband (frequency axis not to scale).

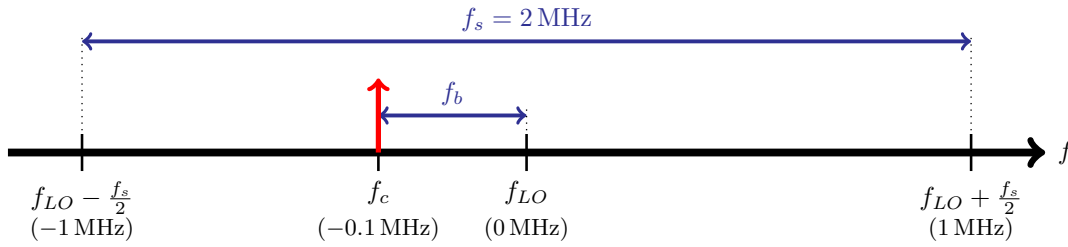


Figure 1.26: Frequency tuning example in baseband (frequency axis not to scale).

earlier). For this case, FM radio stations use carrier frequencies in 0.2 MHz steps. In the frequency range 101-103 MHz, there are 10 stations at 101.1, 101.3, 101.5 etc. up to 102.9 MHz. In this example, we wish to tune in the station at 101.9 MHz.

If the RF carrier wave is turned on and off to transmit information in e.g. Morse code, then we wish to listen to (or digitally decode) the complex baseband signal $\tilde{r}_1(t) = \cos 2\pi f_b t + j \sin 2\pi f_b t$ and no other signals. If f_b is outside the audio range we want to listen to, or if f_b is not the frequency expected at the digital decoder input, then we multiply $\tilde{r}_1(t)$ by a complex exponential $e^{-j2\pi f_d t}$ at frequency f_d to obtain another complex baseband signal

$$\begin{aligned} \tilde{r}_2(t) &= e^{j2\pi f_b t} e^{-2\pi f_d t} \\ &= e^{j2\pi(f_b - f_d)t} \\ &= e^{2\pi f_E t} \end{aligned} \tag{1.81}$$

at frequency $f_E = f_b - f_d$, where f_E is chosen to be the frequency for listening (or for the decoder).

In effect, we have shifted the spectrum twice. We first shifted f_c by f_{LO} to obtain f_b and then shifted f_b by f_d to get the exact frequency f_E we want to listen to (for Morse code) or for a digital decoder.

This idea of two successive spectrum shifts will be seen again later when we discuss the Weaver demodulator.

List of frequencies used above

- f_s : sampling rate of USRP source block
- f_{LO} : local oscillator frequency
- f_c : carrier frequency of desired signal at radio frequency near 10 MHz (passband)
- $(f_{LO} - f_s/2) < f_c < (f_{LO} + f_s/2)$: passband frequency range
- $f_b = f_c - f_{LO}$: desired signal obtained by converting to baseband
- $(-f_s/2) < f_b < (f_s/2)$: baseband frequency range (centered at 0 Hz)
- Conversion (spectrum shifting) is done by complex multiply:

$$\begin{aligned}
 \tilde{r}_1(t) &= r_+(t)e^{-j2\pi f_{LO}t} \\
 &= e^{j2\pi f_c t} e^{-j2\pi f_{LO}t} \\
 &= e^{j2\pi f_b t} \\
 &= \cos 2\pi f_b t + j \sin 2\pi f_b t \\
 &= I(t) + jQ(t)
 \end{aligned}$$

- f_E : desired frequency for listening or for decoder (could be zero or not zero)
- $f_d = f_b - f_E$: digital local oscillator in baseband used for tuning into desired signal
- f_d : is controlled by tuning knob or mouse
- The second spectrum shift is done by complex multiply:

$$\begin{aligned}
 \tilde{r}_2(t) &= e^{j2\pi f_b t} e^{-2\pi f_d t} \\
 &= e^{j2\pi (f_b - f_d)t} \\
 &= e^{2\pi f_E t}
 \end{aligned}$$

More generally, if the RF signal contains information encoded in its amplitude and phase, then the RF signal $r_+(t) = a(t)e^{j\phi(t)}e^{j2\pi f_c t}$ is multiplied by the complex local oscillator $e^{-j2\pi f_{LO}t} = \cos 2\pi f_{LO}t - j \sin 2\pi f_{LO}t$ to yield

$$\begin{aligned}
 \tilde{r}_1(t) &= r_+(t)e^{-j2\pi f_{LO}t} \\
 &= [a(t)e^{j\phi(t)}e^{j2\pi f_c t}]e^{-j2\pi f_{LO}t} \\
 &= a(t)e^{j\phi(t)}e^{j2\pi f_b t} \\
 &= I(t) + jQ(t)
 \end{aligned} \tag{1.82}$$

where the received complex baseband signal is

$$\begin{aligned}
 \tilde{r}_1(t) &= I(t) + jQ(t) \\
 &= a(t) \cos \phi(t) \cos 2\pi f_b t + ja(t) \sin \phi(t) \sin 2\pi f_b t
 \end{aligned} \tag{1.83}$$

If we want to receive the information contained in $a(t), \phi(t)$ then we multiply $\tilde{r}_1(t)$ by a complex exponential $e^{-j2\pi f_b t}$ at exactly $-f_b$ to obtain

$$\begin{aligned}
 \tilde{r}_2(t) &= \tilde{r}_1(t)e^{-j2\pi f_b t} \\
 &= a(t)e^{j\phi(t)}e^{-j2\pi f_b t} \\
 &= a(t)e^{j\phi(t)}
 \end{aligned} \tag{1.84}$$

centered at $f_E = 0$, followed by a low pass filter to filter out any other signals. We have shifted the spectrum twice, once by f_{LO} using analog circuits and a second time by f_b using software to receive the desired signal. In this case, the decoder uses $f_E = 0$

Exercise: Explain the operation of a transmitter that has a selectable carrier frequency. Sketch a block diagram and the signal spectrum. The transmitter uses the same frequency shifting operations as the receiver, but in reverse order.

Exercise: For an complex baseband signal $\tilde{s}(t) = I(t) + jQ(t) = a(t)e^{j\phi(t)}$, find an expression for $a(t)$ as a function of $I(t), Q(t)$

Exercise: Given a USRP daughterboard receiver with $f_{LO} = 142$ MHz and an incoming RF signal $f_c = 142.17$ MHz, find an expression for

1. the two real output signals from the receiver $I(t), Q(t)$, and
2. the complex output signal $I(t) + jQ(t)$ written in polar form $a(t)e^{j\phi(t)}$

Chapter 2

Amplitude modulation (AM) transmitters

In this chapter we apply the mathematics of complex signals from Chapter 1 to a particular class of signals.

In Chapter 1 we wrote a general radio signal $s(t) = a(t) \cos[2\pi f_c t + \phi(t)]$ and showed that we can use $a(t)$ and/or $\phi(t)$ or some combination thereof to represent the message $m(t)$.

In this chapter, we consider the class of signals for which the amplitude $a(t)$ varies with time but the phase $\phi(t)$ is kept constant with time. This class of signals is called *Amplitude Modulation (AM)* signals. In particular, we use the message $m(t)$ to modulate the amplitude $a(t)$ and leave the phase $\phi(t)$ constant. This method of modulation is called *amplitude modulation (AM)*.

We study AM signals and their waveforms ¹ in the time and frequency domains with messages $m(t)$ that are sine waves (single tone), general analog and digital (section 1.25). We also study how to implement a signal generator for AM signals using software as well as analog hardware.

Learning objectives

- Math expression for AM signal with sine wave message in time and frequency domain.
- Concept of sidebands in AM signal spectrum
- Power in carrier and sidebands of AM signals
- AM spectrum with general analog message
- AM signals with digital message
- Generating AM signals using software
- Generating AM signals using analog hardware

¹The waveform of a signal $s(t)$ is the shape of a graph of $s(t)$. The spectral shape of $S(f)$ is the shape of a graph of $S(f)$.

2.1 Amplitude modulation with full carrier

We consider AM signals $s(t)$ where the message $m(t)$ to modulate the amplitude $a(t)$ is varied in step with the message $m(t)$ and the phase $\phi(t)$ is left constant.

AM waves typically use carrier frequencies on the order of MHz, whereas the message frequency is typically on the order of kHz.

Amplitude modulation is used for

- broadcasting (AM radio), using carrier frequencies in the medium wave band (540–1700 kHz), the long wave band 153–279 kHz (in Europe only) and the short wave bands (3–30 MHz).
- aircraft navigational aids (190–535 kHz and 108–118 MHz)
- aircraft voice communications (118–137 MHz).

For AM, we set the amplitude

$$a(t) = A_c[1 + k_a m(t)] \quad (2.1)$$

with $|k_a m(t)| < 1$ or $[1 + k_a m(t)] > 0$ and $f_c \gg w$, where w is the bandwidth of $m(t)$.

Note that $a(t)$ is a DC term 1 plus the message $m(t)$ scaled by a gain factor k_a . We choose k_a such that $a(t) > 0$ for all values of $m(t)$. We also assume that the bandwidth of $m(t)$ is much less than the carrier frequency f_c used.

For AM, we assume $\phi(t)$ is a constant or zero. Thus the transmitted AM signal is of the form:

$$\begin{aligned} s(t) &= a(t) \cos(2\pi f_c t + \phi) \\ &= A_c[1 + k_a m(t)] \cos(2\pi f_c t + \phi) \\ \phi(t) &= \phi = \text{constant} \end{aligned} \quad (2.2)$$

ϕ is set to zero for convenience, but could be any constant value independent of time.

From the mathematics we can see that we have modulated the amplitude $a(t)$ of the signal $s(t)$ in step with the message $m(t)$, so that $s(t)$ is an AM signal.

For the special case where there is no message to send, $m(t) = 0$ and $s(t) = A_c \cos(2\pi f_c t)$ which is simply the carrier wave. In practice, $m(t)$ could be an analog or digital message as described in section 1.2.5.

For a single tone message, $m(t) = A_m \cos(2\pi f_m t)$, the resulting AM wave is given by the following:

$$s(t) = A_c[1 + k_a A_m \cos(2\pi f_m t)] \cos(2\pi f_c t) \quad (2.3)$$

Recall that an analog message (voice or music) is time varying in the practical case and will normally change from one frame to the next. (Recall that frames for audio are typically 5–23 ms). The AM signal above with $m(t) = A_m \cos(2\pi f_m t)$ is valid for one frame in which the message contains only one frequency f_m during that frame. In the next frame, the message could be the same or could be different.

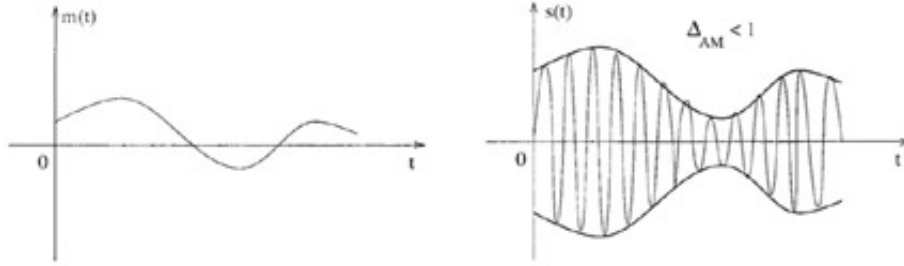


Figure 2.1: AM modulated signal

2.1.1 AM waveform with single tone message

For our initial analysis of $s(t)$, we will assume the message $m(t)$ is of constant frequency f_m for all frames. The AM signal $s(t)$ appears in the time domain to be the carrier wave with an envelope² that replicates the shape of the modulating tone.

$$\begin{aligned}
 m(t) &= A_m \cos 2\pi f_m t \\
 &\text{where } f_m = \text{message frequency} \\
 a(t) &= A_c(1 + k_a m(t)) \\
 &\text{where } \phi(t) = 0 \\
 s(t) &= A_c(1 + k_a m(t)) \cos 2\pi f_c t \\
 s(t) &= A_c(1 + k_a A_m \cos 2\pi f_m t) \cos 2\pi f_c t \\
 s(t) &= A_c(1 + \mu \cos 2\pi f_m t) \cos 2\pi f_c t \\
 s(t) &= A_c \cos 2\pi f_c t + A_c \mu \cos 2\pi f_m t \cos 2\pi f_c t \\
 \mu &= k_a A_m < 1 \text{ is called the modulation index}
 \end{aligned}$$

We use the character $\mu = A_m k_a$ for simplification and to serve as the modulation index. The modulation index μ can be given as a percentage where 1 = 100%.

Using the complex I-Q notation and assuming $\phi(t) = \phi = \text{constant}$ not necessarily zero

$$\begin{aligned}
 s(t) &= a(t) \cos(2\pi f_c t + \phi(t)) \\
 &= \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\}
 \end{aligned} \tag{2.4}$$

²the envelope here is the real part of the complex envelope

Thus for AM

$$\begin{aligned}
s(t) &= \text{Re}\{A_c[1 + k_a m(t)]e^{j2\pi f_c t} e^{j\phi}\} \\
&= \text{Re}\{A_c[1 + \mu \cos 2\pi f_m t]e^{j2\pi f_c t} e^{j\phi}\} \\
&= A_c[1 + \mu \cos 2\pi f_m t] \cos[2\pi f_c t + \phi] \\
a(t) &= A_c[1 + \mu \cos 2\pi f_m t] \\
\phi(t) &= \phi \\
\tilde{s}(t) &= a(t)e^{j\phi} = A_c[1 + \mu \cos 2\pi f_m t]e^{j\phi} \\
&= I(t) + jQ(t) \\
I(t) &= a(t) \cos \phi = A_c[1 + \mu \cos 2\pi f_m t] \cos \phi \\
Q(t) &= a(t) \sin \phi = A_c[1 + \mu \cos 2\pi f_m t] \sin \phi
\end{aligned}$$

When $\phi = 0$

$$\begin{aligned}
s(t) &= \text{Re}\{A_c[1 + k_a m(t)]e^{j2\pi f_c t}\} \\
&= \text{Re}\{A_c[1 + \mu \cos 2\pi f_m t]e^{j2\pi f_c t}\} \\
&= A_c[1 + \mu \cos 2\pi f_m t] \cos[2\pi f_c t] \\
a(t) &= A_c[1 + \mu \cos 2\pi f_m t] \\
\phi(t) &= 0 \\
\tilde{s}(t) &= a(t) = A_c[1 + \mu \cos 2\pi f_m t] \\
&= I(t) + jQ(t) \\
I(t) &= a(t) = A_c[1 + \mu \cos 2\pi f_m t] \\
Q(t) &= 0
\end{aligned}$$

We will use the version of $s(t)$ with $\phi = 0$ in the subsequent sections.

2.1.2 AM spectrum

What does the AM wave look like in the frequency domain?

We first consider the special case where the message $m(t) = A_m \cos(2\pi f_m t)$ with $A_m = 1$, and take the Fourier transform of the message

$$\begin{aligned}
m(t) &\xrightarrow{\text{Fourier}} M(f) \\
A_m \cos(2\pi f_m t) &\xrightarrow{\text{Fourier}} \frac{1}{2} A_m [\delta(f - f_m) + \delta(f + f_m)]
\end{aligned} \tag{2.5}$$

This result is obtained from the Fourier transform properties for a cos wave.

$$\begin{aligned}
m(t) &\leftrightarrow M(f) \\
m(t) &= \cos(2\pi f_m t) \\
&= (e^{j2\pi f_m t} + e^{-j2\pi f_m t})/2 \\
e^{j2\pi f_m t} &\leftrightarrow \delta(f - f_m) \text{ positive frequencies} \\
e^{-j2\pi f_m t} &\leftrightarrow \delta(f + f_m) \text{ negative frequencies} \\
M(f) &= \frac{1}{2} [\delta(f - f_m) + \delta(f + f_m)]
\end{aligned}$$

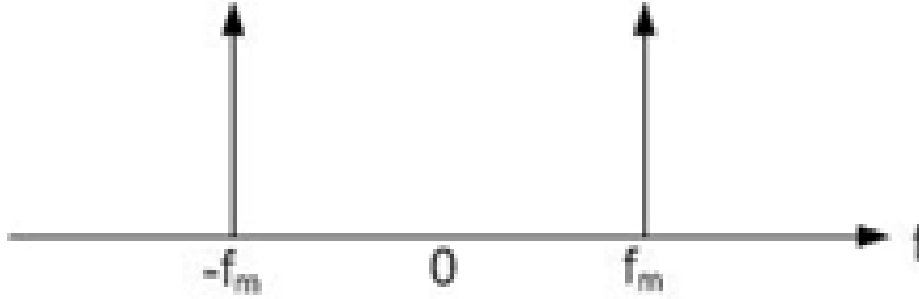


Figure 2.2: Frequency Spectrum of a Single Tone

Note that the cos wave contains both positive and negative frequencies arising from the exponentials rotating clockwise and counterclockwise respectively as time increases.

The spectrum for any real signal is symmetrical about zero frequency. The spectrum of the message $M(f)$ can be drawn as shown in fig. 2.2, with both positive and negative frequencies. Note that the delta functions are drawn as vertical lines with arrows that suggest they go to infinity. The delta function $\delta(f - f_m)$ is zero for all frequencies $f \neq f_m$. For the frequency $f = f_m$, the delta function $\delta(f - f_m)$ has infinite height, zero width and area 1, so that we can write the area under the curve $\int_{-\infty}^{\infty} \delta(f - f_m) = 1$.

The spectrum of the carrier wave is written

$$c(t) \leftrightarrow C(f)$$

$$c(t) = \cos(2\pi f_c t)$$

$$= (e^{j2\pi f_c t} + e^{-j2\pi f_c t})/2$$

$$e^{j2\pi f_c t} \leftrightarrow \delta(f - f_c) \text{ positive frequencies} \quad (2.6)$$

$$e^{-j2\pi f_c t} \leftrightarrow \delta(f + f_c) \text{ negative frequencies} \quad (2.7)$$

$$C(f) = \frac{1}{2}[\delta(f - f_c) + \delta(f + f_c)] \quad (2.8)$$

$C(f)$ appears as two delta function spikes, one at f_c and one at $-f_c$.

To find the spectrum of the AM signal

$$\begin{aligned} s(t) &= A_c(1 + \mu \cos 2\pi f_m t) \cos 2\pi f_c t \\ &= A_c \cos 2\pi f_c t + A_c \mu \cos 2\pi f_m t \cos 2\pi f_c t \end{aligned} \quad (2.9)$$

we take the Fourier transform $s(t) \xrightarrow{\text{Fourier}} S(f)$ and obtain the result shown in fig. 2.3.

We will discuss the result for $S(f)$ first, and then do the mathematics.

Observe that the spectrum of an AM signal contains both positive and negative frequencies. The AM signal spectrum is the sum of two components: the spectrum of the carrier wave $c(t)$ at $\pm f_c$ plus the spectrum of the message ($M(f)$) shifted both up and down by f_c . There is a very important principle at work here: the frequency shifting properties in Section 1.1.5. Notice that in the expression

$$s(t) = A_c \cos 2\pi f_c t + A_c \mu \cos 2\pi f_m t \cos 2\pi f_c t$$

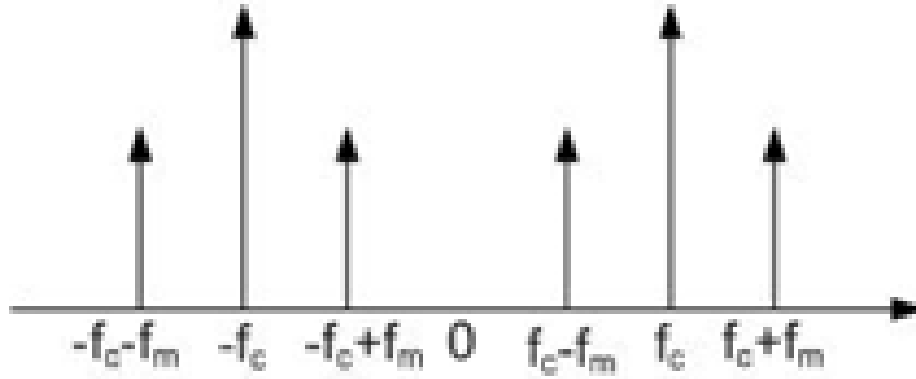


Figure 2.3: Frequency Spectrum of AM Wave

the message signal $m(t)$ is multiplied by the real carrier wave $c(t)$. Whenever a message signal is multiplied by a real carrier wave containing both positive and negative frequencies, the spectrum of the message signal is shifted both up and down.

This result for the AM signal spectrum can be demonstrated in two different ways. The first method is to do it algebraically by simplifying $s(t)$ in the time domain and then converting the resulting collection of individual sinusoidal terms to their frequency domain representation. Note: The trigonometric identity $\cos \alpha \cos \beta = \frac{1}{2}[\cos(\alpha + \beta) + \cos(\alpha - \beta)]$ is used to produce line three from line two below:

$$\begin{aligned}
 s(t) &= A_c[1 + \mu \cos(2\pi f_m t)] \cos(2\pi f_c t) \\
 &= A_c \cos(2\pi f_c t) + A_c \mu \cos(2\pi f_m t) \cos(2\pi f_c t) \\
 &= A_c \cos(2\pi f_c t) + (A_c \mu / 2) \cos(2\pi [f_c + f_m] t) + (A_c \mu / 2) \cos(2\pi [f_c - f_m] t) \quad (2.10)
 \end{aligned}$$

The second and third cosine terms represent the so-called *sidebands*³ as seen in the frequency domain. *The frequency of the sidebands relative to the carrier frequency holds the useful information describing the message.*

The AM signal spectrum $S(f)$ may be written by taking the Fourier transform of each of the 3 terms in the expression for $s(t)$ above to obtain

$$\begin{aligned}
 S(f) &= \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] \\
 &\quad + \frac{A_c \mu}{4} [\delta(f - f_c - f_m) + \delta(f + f_c + f_m)] \\
 &\quad + \frac{A_c \mu}{4} (\delta(f - f_c + f_m) + \delta(f + f_c - f_m)) \quad (2.11)
 \end{aligned}$$

Exercise: Obtain $s(f)$ by writing $s(t)$ for an AM signal as $s(t) = [s_+(t) + s_-(t)]/2$, taking the Fourier transform of the positive and negative frequency terms separately and adding them together to get the final result for $S(f)$

Exercise: show that $\mu = \frac{A_{cMAX} - A_{cMIN}}{A_{cMAX} + A_{cMIN}} = A_m k_a$, where A_{cMAX} and A_{cMIN} are respectively the

³Here is yet another term for a kind of *band*, to go with the terms baseband and passband

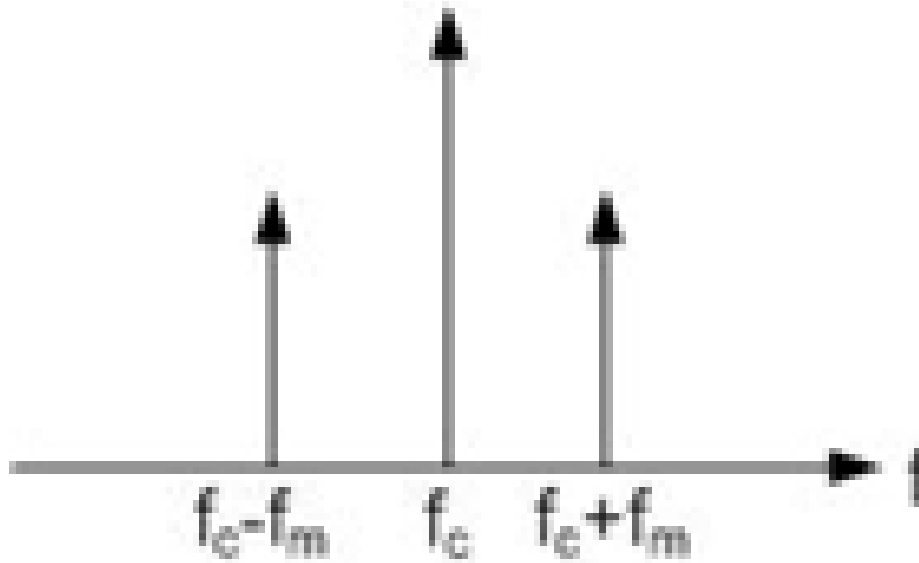


Figure 2.4: Spectral Representation of AM signal showing positive frequencies only

highest and lowest positive amplitude obtained by $s(t)$. Hint: $A_{cMAX} = A_c(1 + \mu)$; $A_{cMIN} = A_c(1 - \mu)$, $\frac{A_{c\max}}{A_{c\min}} = \frac{A_c(1+\mu)}{A_c(1-\mu)}$.

Observe that even though we seem to only modify the amplitude of the carrier wave at f_c sidebands are observed at $f_c + f_m$ and $f_c - f_m$ in the frequency domain representation. How could there be frequencies at $f_c + f_m$ and $f_c - f_m$ when all we are doing is varying the amplitude of the carrier at f_c without changing its frequency?

The reason is that to see the sidebands at $f_c + f_m$ and $f_c - f_m$ we must observe the signal over a sufficiently long time (at least one cycle of the message $m(t)$).⁴ If we observe the signal for a short time over a small fraction of a cycle of $m(t)$, then we see only the carrier wave at frequency f_c with amplitude $a(t) = A_c[1 + \mu \cos 2\pi f_m t]$ at time t and we don't see any sidebands.

The second method of showing the result is done using Fourier transform properties. Note that a Fourier transform is taken over a long time including at least one cycle of the message $m(t)$, so that we expect to see the sidebands. The result can also be shown using a Fourier series, where the signal is assumed to be periodic with a period equal to one cycle of the message.

2.1.3 Power in carrier wave and sidebands

The power of a signal is a quantity of interest and one may wish to calculate the power of the carrier or sideband components of the signal specifically. Using the time domain expression for $s(t)$, the power is proportional to the square of the coefficient of each cosine term. Using the frequency domain expression $S(f)$, the power is proportional to the square of the coefficient of each δ -function term.

⁴If we have one complete cycle of $m(t)$ then we have a periodic signal and can take the Fourier series (or Fourier transform) to observe the sidebands.

The power versus frequency curve is displayed on a spectrum analyzer, typically in dBm, or decibel relative to one milliwatt at 50 Ω . For this case, $P(\text{dBm}) = 20 \log V + 13$ where V is rms voltage, as was shown in Section 1.4.

In what follows we assume 1 Ω rather than 50 Ω for convenience as is typically done for communication signals mathematics. The power into one ohm $[s(t)]^2 = |S(f)|^2$ without the need for scaling factors to account for impedance.

Consider an AM signal with single tone modulation $m(t) = A_m \cos 2\pi f_m t$:

$$s(t) = A_c [1 + \mu \cos(2\pi f_m t)] \cos(2\pi f_c t)$$

which, in the frequency domain is

$$s(t) = A_c \cos(2\pi f_c t) + (A_c \mu / 2) \cos(2\pi [f_c + f_m] t) + (A_c \mu / 2) \cos(2\pi [f_c - f_m] t) \quad (2.12)$$

- Power at carrier at f_c : $A_c^2/2$
- Power of sideband at $f_c + f_m$: $(1/2)(A_c \mu / 2)^2 = (1/8)A_c^2 \mu^2$
- Power of sideband at $f_c - f_m$: $(1/2)(A_c \mu / 2)^2 = (1/8)A_c^2 \mu^2$
- Carrier power: $2 \left(\frac{A_c}{2}\right)^2 = \frac{A_c^2}{2}$
 - factor of 2 because positive and negative frequencies
- Upper sideband power: $2 \left(\frac{A_c \mu}{4}\right)^2 = \frac{A_c^2 \mu^2}{8}$
- Lower sideband power: $2 \left(\frac{A_c \mu}{4}\right)^2 = \frac{A_c^2 \mu^2}{8}$
- $\frac{USB+LSB \text{ Power}}{\text{Total Power}} = \frac{\frac{A_c^2 \mu^2}{8} + \frac{A_c^2 \mu^2}{8}}{\frac{A_c^2}{2} + \frac{A_c^2 \mu^2}{8} + \frac{A_c^2 \mu^2}{8}} = \frac{\mu^2}{2 + \mu^2}$

Exercise: If the power at $f_c = 0$ dBm, what is the power at $f_c + f_m$?

Answer: Given: $A_c^2/2 = 1$ mW; $(1/8)A_c^2 \mu^2 = (1 \text{ mW})\mu^2/4$. If $\mu = 1$ then the power at $f_c + f_m$ is $(1 \text{ mW})/4 = 0.25$ mW. A drop in power by half is -3 dB, thus one quarter of the power is -6 dB.

In the special case where $\mu = 1$ (see fig. 2.5), we have 100% modulation (minimum envelope value is zero).

2.1.4 Overmodulation

Given the AM signal

$$s(t) = A_c [1 + \mu \cos(2\pi f_m t)] \cos(2\pi f_c t)$$

, with modulation index $\mu > 1$ (greater than 100%) the real envelope, $a(t) = [1 + \mu \cos(2\pi f_m t)]$ becomes less than zero and no longer looks like the message being sent. This case is called “overmodulation”. The phase of the carrier wave is shifted by 180 degrees when $[1 + \mu \cos(2\pi f_m t)]$ is less than zero.

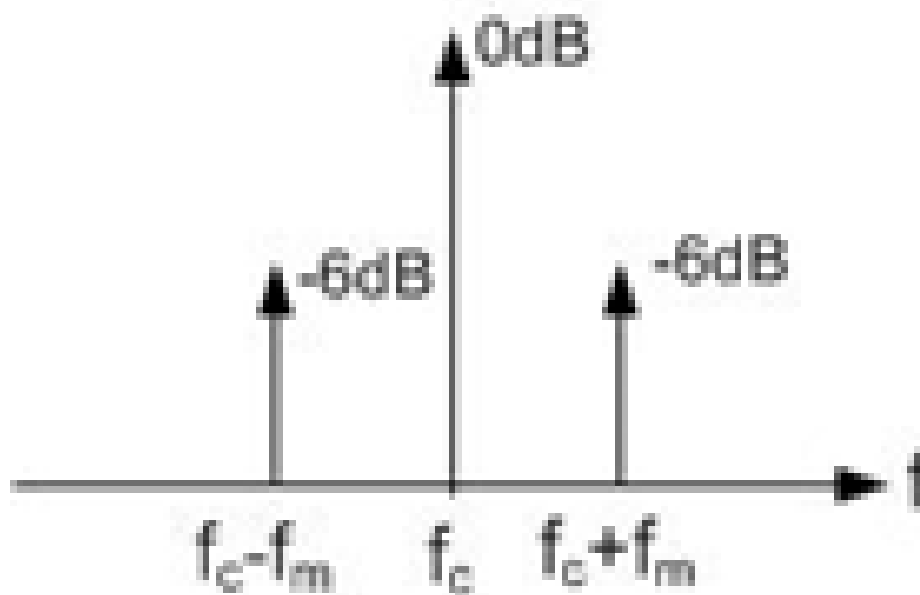


Figure 2.5: AM Spectrum: Sidebands with modulation index 1

In the frequency domain, the sum of the amplitudes (not the power) of the sidebands exceeds the amplitude of the carrier.

Exercise: What happens to the complex envelope $\tilde{s}(t)$ when overmodulation occurs? Write an expression for $\tilde{s}(t)$.

Answer:

$$\begin{aligned}
 \tilde{s}(t) &= a(t)e^{j\phi} = A_c(1 + \mu\cos 2\pi f_m t) \\
 &= I(t) + jQ(t) \\
 I(t) &= A_c(1 + \mu\cos 2\pi f_m t) \text{ may be } < 0 \text{ when } \mu > 1 \\
 Q(t) &= 0 \\
 |\tilde{s}(t)| &= |a(t)e^{j\phi}| \\
 &= |I(t) + jQ(t)| \\
 &= \sqrt{I^2(t) + Q^2(t)} \\
 &= |A_c[1 + \mu\cos 2\pi f_m t]| \\
 &\geq 0
 \end{aligned}$$

The real envelope $|a(t)|$ is the rectified version of $I(t)$.

Exercise: Write an expression for $\tilde{S}(f)$ and $s(f)$ for the overmodulation case with $\mu = 2$.

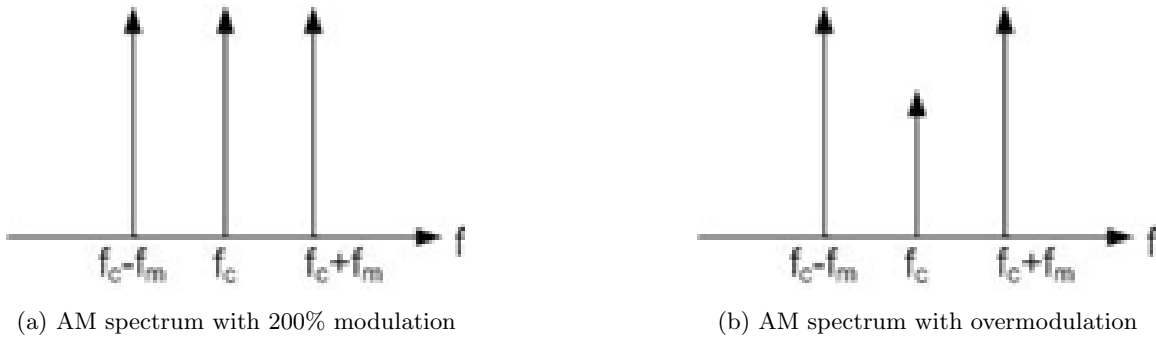

 Figure 2.6: AM waveforms for various Δ values


Figure 2.7: AM spectra with various modulation levels

2.1.5 AM signal with a general message

We can write the AM signal in time and frequency domain for a general message $m(t)$ as

$$s(t) = A_c[1 + k_a m(t)] \cos 2\pi f_c t \quad (2.13)$$

$$= A_c[1 + k_a m(t)] \frac{(e^{j2\pi f_c t} + e^{-j2\pi f_c t})}{2} \quad (2.14)$$

$$= \frac{A_c}{2} e^{j2\pi f_c t} + \frac{A_c}{2} e^{-j2\pi f_c t} + \frac{A_c k_a}{2} m(t) e^{j2\pi f_c t} + \frac{A_c k_a}{2} m(t) e^{-j2\pi f_c t} \quad (2.15)$$

To find the frequency domain expression $S(f)$ use the Fourier transform properties

$$\begin{aligned} m(t) &\leftrightarrow M(f) \\ e^{j2\pi f_c t} &\leftrightarrow \delta(f - f_c) \\ e^{-j2\pi f_c t} &\leftrightarrow \delta(f + f_c) \\ \exp(j2\pi f_c t)m(t) &\leftrightarrow M(f - f_c) \\ \exp(-j2\pi f_c t)m(t) &\leftrightarrow M(f + f_c) \end{aligned}$$

The result is

$$S(f) = \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] + \frac{A_c k_a}{2} [M(f - f_c) + M(f + f_c)] \quad (2.16)$$

This expression shows both positive and negative frequencies. We often draw a picture of the positive frequencies only, and we can write

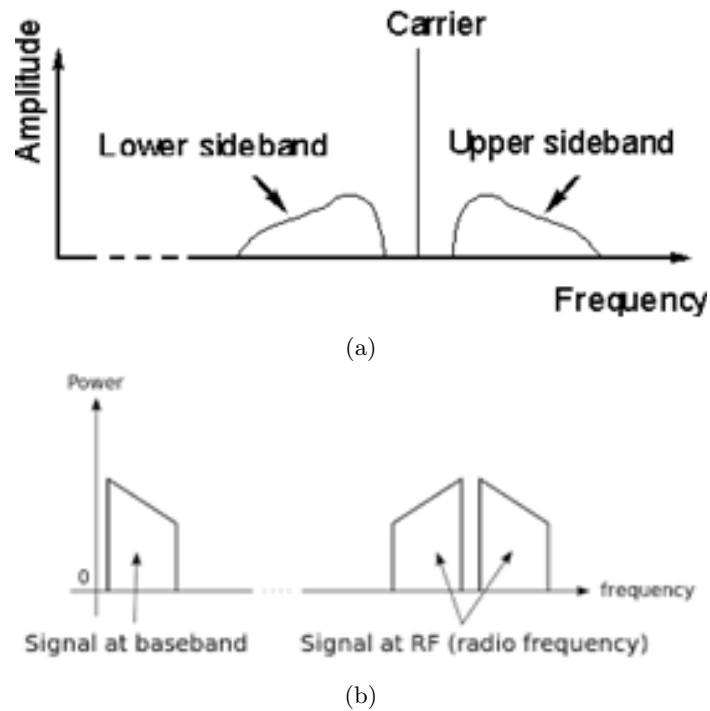


Figure 2.8: Upper and lower side bands as well as base band spectra all showing only positive frequencies.

$$S(f > 0) = \frac{A_c}{2} \delta(f - f_c) + \frac{A_c k_a}{2} M(f - f_c)$$

The modulation spectrum $M(f)$ is shifted up so it is centered around f_c as in fig. 2.8.

In fig. 2.9, both positive and negative frequencies are shown. In this figure we write $\omega = 2\pi f$, $\omega_c = 2\pi f_c$.

Observe that the AM signal spectrum is the sum of two components: the spectrum of the carrier wave at $\pm f_c$ plus the spectrum of the message shifted both up and down by f_c . Observe that because the message spectrum contains both positive and negative frequencies, the AM signal spectrum contains components both above and below the carrier frequency.

In general, $M(f)$ will change shape with each frame. For a particular frame, and for illustration purposes, we draw it as it would be for a voice signal: an asymmetrical shape that is zero for $f < 300$ Hz, peaks near $f = 1000$ Hz, and is zero for $f > 2700$ Hz. (Think of the green line in `sndpeek` <https://soundlab.cs.princeton.edu/software/sndpeek/>). The negative frequencies will be the mirror image of this shape.

Exercise: Obtain $s(f)$ by writing $s(t)$ for an AM signal as $s(t) = [s_+(t) + s_-(t)]/2$, taking the Fourier transform of the positive and negative frequency terms separately and adding them together to get the final result for $S(f)$.

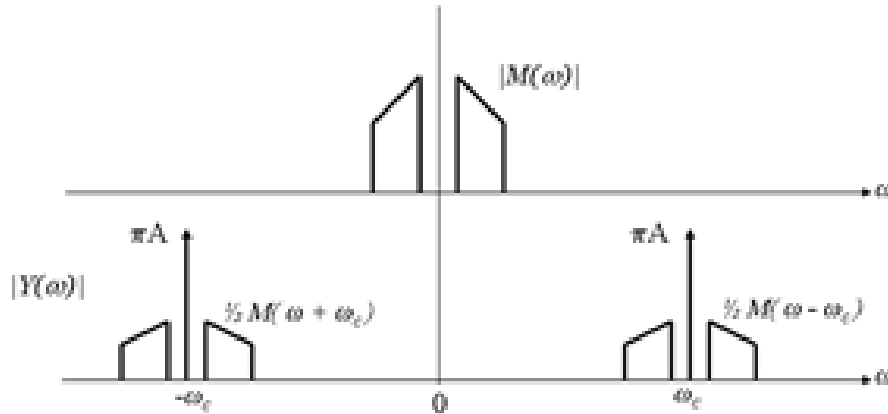


Figure 2.9: Spectrum of upper and lower sidebands showing negative frequencies

2.1.6 Digital Messages Transmitted Using AM

Digital messages can be sent by using a binary modulating signal; the amplitude of the carrier is multiplied by a high voltage (logic one) and a low voltage (logic zero). A modulated wave with the capacity to represent a binary bit stream is the result. This method is known as amplitude shift keying (ASK). The most common case is where logic 1 is a positive voltage, say +1 volts and a logic 0 is a negative voltage say -1 volts. In general we can write $m(t) = \pm A_m$. Thus we can write the digital AM signal

$$\begin{aligned} s(t) &= A_c[1 + k_a m(t)] \cos(2\pi f_c t) \\ &= A_c[1 \pm k_a A_m] \cos 2\pi f_c t \end{aligned} \tag{2.17}$$

If we choose A_m such that $k_a A_m = 1$, then

$$\begin{aligned} s(t) &= A_c[1 \pm k_a A_m] \cos 2\pi f_c t \\ &= A_c[1 \pm 1] \cos 2\pi f_c t \\ &= 2A_c \cos 2\pi f_c t \text{ for logic 1} \\ &= 0 \text{ for logic 0} \end{aligned}$$

depending on whether a logic 1 or a logic 0 was sent. In this case the amplitude is shifted from $2A_c$ to 0, and it is called on-off keying (OOK) and can be used for Morse code transmission.

2.1.7 Generating an AM signal digitally in software

Next we look at how to build transmitters and receivers. We have a message that we wish to send and we need a way to send and receive it.

One approach is to use digital means, for example by programming the USRP using GNU Radio. The USRP includes the DAC and analog IQ mixer, filters and amplifiers connected to an antenna.

The transmitted AM signal is of the form:

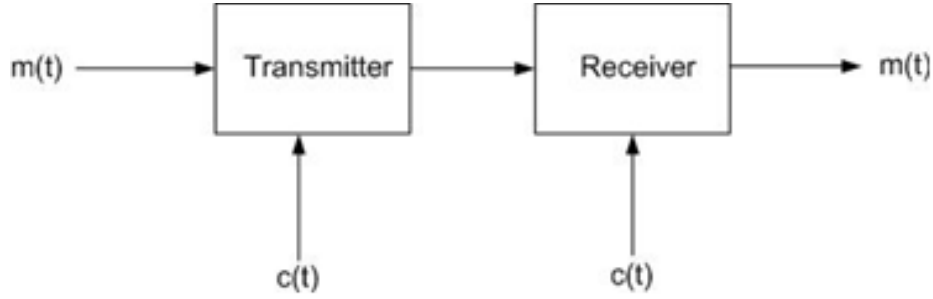
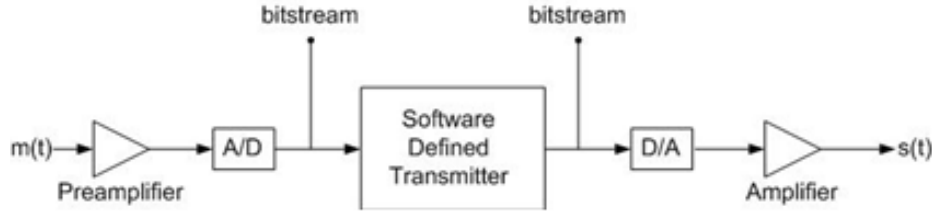
Figure 2.10: Conceptual modulation and demodulation of the Signal $m(t)$ 

Figure 2.11: Digital transmitter configuration

$$\begin{aligned} s(t) &= a(t) \cos(2\pi f_c t + \phi) \\ &= A_c [1 + k_a m(t)] \cos(2\pi f_c t + \phi) \end{aligned} \quad (2.18)$$

where ϕ is a constant phase. We can write the AM signal $s(t)$ as the real part of a complex signal:

$$\begin{aligned} s(t) &= \text{Re}[a(t)e^{j\phi}e^{j2\pi f_c t}] \\ &= \text{Re}[\tilde{s}(t)e^{j2\pi f_c t}] \end{aligned}$$

where the complex envelope:

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi} \\ &= a(t) \cos \phi + ja(t) \sin \phi \\ &= i(t) + jq(t) \end{aligned}$$

Usually we choose $\phi = 0$, so that $i(t) = a(t) = 1 + k_a m(t)$, $q(t) = 0$

Given the message $m(t)$, the digital AM transmitter simply implements the equations $i(t) = 1 + k_a m(t)$

2.1.8 Analog-only method to generate an AM signal

In the days before software radio, only analog circuit-theory based methods were available to generate an AM signal. The mathematics tells us that we want to multiply the message signal $m(t)$ with the carrier wave $c(t)$, in particular: $s(t) = [1 + k_a m(t)]c(t)$. So we need to build an analog multiplier.

One purely analog method uses a non-linear circuit with output

$$v_2(t) = \sum_i a_i [v_1(t)]^i = a_1 v_1(t) + a_2 v_1^2(t) + \dots$$

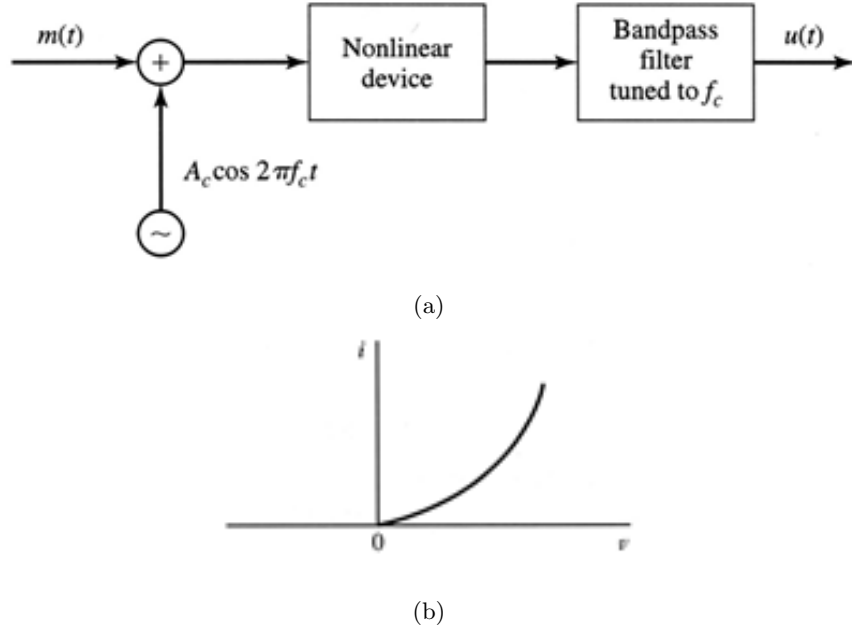


Figure 2.12: Fundamental Analog Transmission Method

Any non-linearity $v_2(t) = \sum_i a_i [v_i(t)]^i$ will cause multiplication where $i \geq 2$.

This non-linearity could be a diode operating with a small input voltage. The non-linearity includes a square term where the input is multiplied by itself.

As shown by the identity, $\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha + \beta) + \cos(\alpha - \beta)]$ and also in the previous section, the result of multiplying two signals is sinusoidal components located at the sum and the difference of carrier and message frequencies.

We model the non-linearity as the first two terms of the Taylor series expansion

$$v_2(t) = \sum_{i=1,2} a_i [v_i(t)]^i \quad (2.19)$$

where the input signal $v_1(t) = m(t) + c(t)$ is the sum (not the product) of the message and the carrier. Thus we write

$$\begin{aligned} v_2(t) &= a_1 v_1(t) + a_2 [v_1(t)]^2 \\ v_1(t) &= m(t) + c(t) \\ v_i(t) &= m(t) + A_c \cos(2\pi f_c t) \\ &\vdots \\ v_2(t) &= a_1 [m(t) + A_c \cos(2\pi f_c t)] + a_2 [m(t) + A_c \cos(2\pi f_c t)]^2 \\ v_2(t) &= a_1 m(t) + a_1 A_c \cos(2\pi f_c t) + a_2 m^2(t) + 2a_2 m(t) A_c \cos(2\pi f_c t) + a_2 A_c^2 \cos^2(2\pi f_c t) \end{aligned}$$

Using the trigonometric identity, $\cos^2(\alpha) = (1/2)[1 + \cos(2\alpha)]$, we arrive at

$$a_2 A_c^2 \cos^2(2\pi f_c t) = (1/2) a_2 A_c^2 [1 + \cos(4\pi f_c t)] \quad (2.20)$$

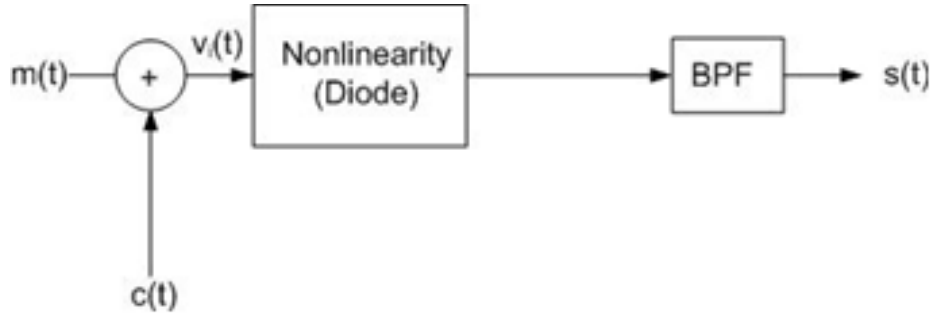


Figure 2.13: Adding Message and Carrier Prior to Non-linearity

Sorting by frequency we get the following simplified expression:

$$\begin{aligned}
 v_2(t) = & a_1 m(t) + a_2 m^2(t) + (1/2)a_2 A_c^2 \\
 & + \cos(2\pi f_c t)[a_1 A_c + 2a_2 A_c m(t)] \\
 & + \cos(4\pi f_c t)[(1/2)a_2 A_c^2]^2
 \end{aligned} \tag{2.21}$$

We now apply a band-pass filter centred at f_c , to get the desired modulator output signal

$$s(t) = a_1 A_c \cos 2\pi f_c t [1 + (2a_2/a_2)m(t)]$$

Thus we have created an AM wave $s(t)$ by adding $m(t) + c(t)$ and then using a non-linearity and band pass filter. For this AM wave $\mu = 2a_2/a_1$.

Exercise: write a frequency domain expression for $v_2(t)$.

Answer:

$$\begin{aligned}
 V_2(f) = & a_1 \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] \\
 & + 2a_2 A_c [M(f - f_c) + M(f + f_c)] \\
 & + a_1 M(f) + a_2 FT[m^2(t)] \\
 & + \frac{a_2 A_c^2}{2} [\delta(f) + \delta(f - 2f_c) + \delta(f + 2f_c)]
 \end{aligned}$$

Another type of non-linearity is diode switching.

$$v_2(t) = \begin{cases} v_1(t), & c(t) > 0 \\ 0, & c(t) < 0 \end{cases}$$

In this case, the message waveform plus the carrier wave signal is half-wave rectified and the result is band-pass filtered at the carrier wave frequency.

Exercise: Show that the filter output contains an AM wave. Hint:

$$\begin{aligned}
 v_1(t) &= c(t) + m(t) \\
 v_2(t) &= [c(t) + m(t)]g_p(t)
 \end{aligned}$$

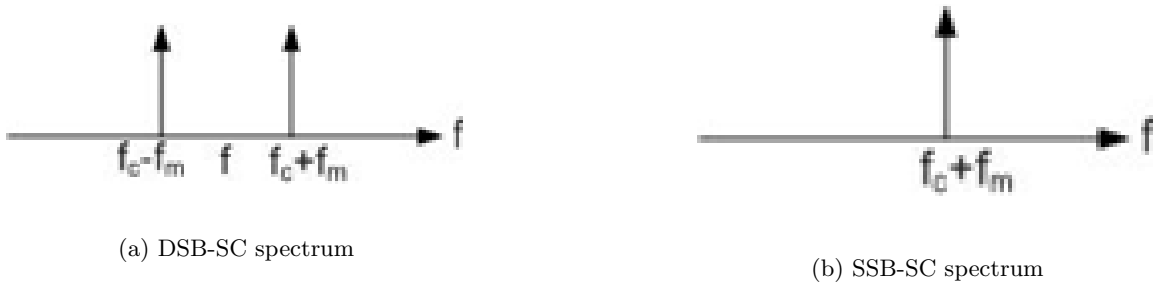


Figure 2.14: Spectra for double and single side band signals

where $g_p(t)$ is a uni-polar square wave at f_c .

Answer: Find the Fourier series for the square wave

$$\begin{aligned} v_2(t) &= [c(t) + m(t)] \left[\frac{1}{2} + \frac{2}{\pi} \cos 2\pi f_c t + \dots \right] \\ &= \frac{A_c}{2} \left[1 + \frac{4}{\pi A_c} m(t) \right] \cos 2\pi f_c t + \dots \end{aligned}$$

which is an AM wave with $k_a = \frac{4}{\pi A_c}$ plus components at frequencies removed from f_c

2.2 Double Sideband Modulation

Carrier waves require power to transmit but they do not in themselves contain any information about the message. Sometimes we want to transmit an AM wave without its carrier wave (this lessens the power requirements for transmission and thus may be a more economical course of action for certain applications). Filtering out (suppressing) the carrier frequency band leads to a double sideband - suppressed carrier signal (DSB-SC) and further filtering out one of either the upper sideband (USB) or lower sideband (LSB) will produce what is called a single sideband - suppressed carrier signal (SSB-SC).

Yet another type is vestigial sideband modulation (VSB) which can be used practically for transmission of television signals. In essence VSB = SSB + Carrier.

Modulation Type	Advantages	Disadvantages
AM	Easily demodulated	High power requirements due to carrier
DSB-SC	Lower power requirements	Carrier location required. Redundant SB information
SSB-SC	Less bandwidth and power requirements	Don't know where carrier is
VSB	Less bandwidth and power requirements. Lack of redundancy	

Table 2.1: Amplitude based modulation schemes

2.2.1 Double sideband suppressed carrier waveform

For a double sideband signal we know that the only difference from the AM case is the lack of carrier term. Thus with a message $m(t) = A_m \cos 2\pi f_m t$ and a carrier wave $c(t) = A_c \cos 2\pi f_c t$, the DSB-SC signal is represented by

$$\begin{aligned} s(t) &= m(t)c(t) \\ &= A_c A_m \cos(2\pi f_m t) \cos(2\pi f_c t) \\ &= \frac{A_c A_m}{2} [\cos(2\pi[f_c - f_m]t) + \cos(2\pi[f_c + f_m]t)] \end{aligned} \quad (2.22)$$

where we let the constant $A_c A_m$ represent any scaling that has taken place by way of amplification or any hardware effects. We have used the trigonometric identity

$$\cos \alpha \cos \beta = \frac{1}{2} [\cos(\alpha - \beta) + \cos(\alpha + \beta)]$$

For DSB-SC with message $m(t) = A_m \cos 2\pi f_m t$, $I(t) = A_c \cos(2\pi f_m t)$ and $Q(t) = 0$. A DSB-SC modulator is simply a multiplier that multiplies the message signal $m(t)$ with the carrier $c(t)$ as illustrated in fig. 2.15.

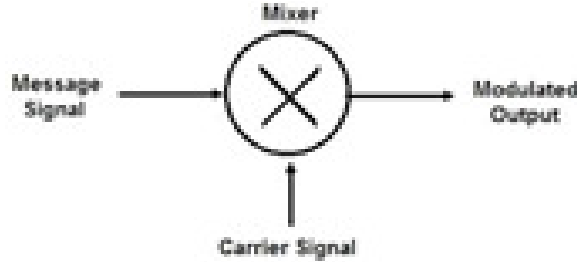


Figure 2.15: DSB-SC modulator block diagram

Figure 2.16 shows the waveforms ($m(t)$, $c(t)$, $s(t)$ respectively) for DSB-SC. Note the phase reversal of the carrier wave when the message signal is less than zero.

DSB may be view as an extreme case of AM overmodulation

$$\begin{aligned} s(t) &= A_c [1 + \mu \cos(2\pi f_m t)] \cos(2\pi f_c t) \\ &= A_c \cos(2\pi f_c t) + A_c \mu \cos(2\pi f_m t) \cos(2\pi f_c t) \end{aligned}$$

with $\mu \gg 1$ and $A_c \ll 1$ and μA_c held constant. In this case, the power in the sidebands is much greater than the power in the carrier. In the limit $\mu \rightarrow \infty$ and $A_c \rightarrow 0$ and $\mu A_c = 1$, the overmodulated AM signal becomes a DSB signal.

2.2.2 DSB spectrum

The spectrum of the single tone message $m(t) = A_m \cos 2\pi f_m t$ is shown in fig. 2.18.

The spectrum of a DSB signal with carrier $c(t) = A_c \cos 2\pi f_c t$ and message $m(t) = A_m \cos 2\pi f_m t$ is two peaks at $f_c - f_m$ and $f_c + f_m$ (and no power at f_c) as shown in fig. 2.19.

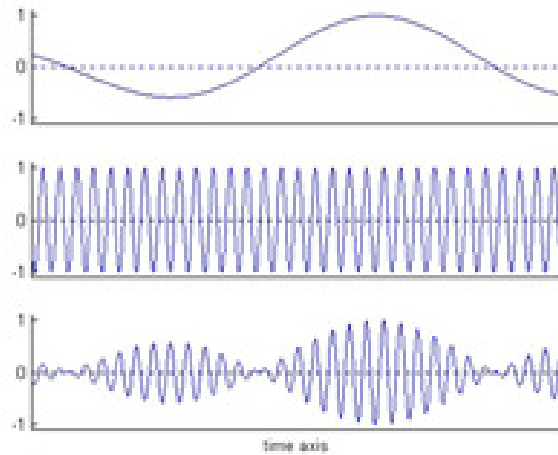


Figure 2.16: DSB-SC waveforms

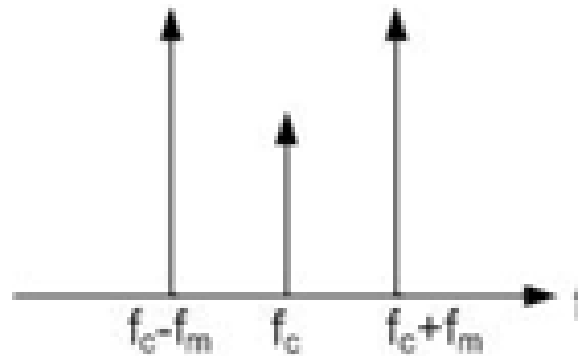


Figure 2.17: DSB as limiting case of overmodulated AM with weak but non-zero carrier

The DSB signal spectrum will also include negative frequencies with peaks at $-(f_c - f_m)$ and $-(f_c + f_m)$ (not shown in fig. 2.19).

Notice the general principle at work here: multiplying two cos waves together yields the sum and difference frequencies, as can be seen from the trigonometric identity $\cos \alpha \cos \beta = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$.

There is another related principle at work here. Notice that in the expression

$$s(t) = m(t)c(t) = A_c A_m \cos 2\pi f_m t \cos 2\pi f_c t$$

the message signal is multiplied by the carrier wave. Whenever a message signal is multiplied by a carrier wave, the spectrum of the message signal is shifted both up and down by the carrier frequency.

Observe that because the message spectrum contains both positive and negative frequencies, the signal spectrum contains components both above and below the (suppressed) carrier frequency.

Exercise: for a DSB signal $s(t)$, write an expression for $S(f)$

Exercise: for a DSB signal, find the power in each sideband.

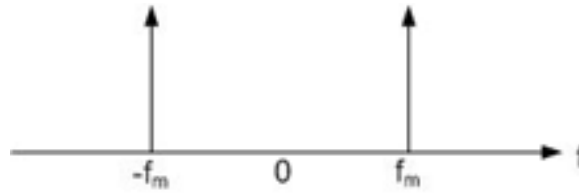


Figure 2.18: Spectrum of single tone message

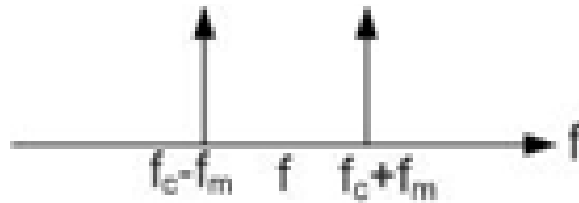


Figure 2.19: DSB spectrum showing both peaks

Exercise: for a DSB-SC signal, find the complex envelope

Exercise: write the spectrum of a DSB signal with a single tone message using delta functions.

Exercise: write the spectrum of a DSB signal with a general message $m(t) \leftrightarrow M(f)$

2.3 Single sideband suppressed carrier and the Hilbert transform

SSB type modulation may be viewed as double sideband with one of the sidebands removed.

Ironically, we will see that to *remove* the sideband, we *add* an additional term to the expression for DSB-SC.

The information content in all variations of AM is in the sidebands. The advantage of SSB is that all the available transmit power is used for one sideband only, whereas for DSB the available



(a) DSB-SC Spectrum

(b) SSB-SC Spectrum

Figure 2.20: Caption

transmit power is divided between two sidebands and for AM the available power is divided between the carrier wave as well as the two sidebands. Thus SSB is more power-efficient than DSB or AM.

SSB is commonly used for long distance over-the-horizon voice communications on the shortwave bands 1.8 - 30 MHz that can propagate worldwide via the ionosphere. SSB may also be used on other frequencies where power efficiency is important. Radio services using SSB include marine mobile, aeronautical mobile and Amateur radio.

2.3.1 Derivation of SSB-SC using Hilbert transform and analytic signals

We can start the derivation of SSB-SC signal mathematics by considering the message spectrum $M(f)$. Note that $M(f)$ has two sidebands on either side of DC or zero frequency. Thus $M(f)$ contains both positive and negative frequencies. For any $m(t)$ that is real-valued with Fourier transform $M(f)$, we find that $M(-f) = M^*(f)$ where $*$ denotes complex conjugate (same amplitude, negative of the phase).

We can create an SSB-SC waveform in the same way as we created DSB-SC, except that we add one new step:

1. eliminate one of the sidebands (say the negative frequency sideband) in the message spectrum to create a new message with asymmetrical spectrum,
2. multiply the new message by the carrier wave as was done for DSB-SC.

The negative frequency sideband is eliminated by multiplying $M(f)$ by the Heaviside step function $2u(f) = 1 + \text{sgn}(f) = 2$ where

$$u(f) = \begin{cases} 1 & f > 0 \\ 0 & f < 0 \end{cases} \quad (2.23)$$

This is exactly the same approach taken at the end of section 1.1 (equation (1.33)). The new message can be written as analytic signal

$$M_a(f) = M_+(f) = 2u(f)M(f) = [1 + \text{sgn}(f)]M(f) = M(f) + \text{sgn}(f)M(f)$$

and has an asymmetrical spectrum with only positive frequencies. The new message in the time domain is

$$\begin{aligned} m_+(t) &= IFT\{M_a(f)\} \\ &= IFT\{M(f)\} + IFT\{\text{sgn}(f)M(f)\} \\ &= m(t) + IFT\{\text{sgn}(f)\} \otimes IFT\{M(f)\} \\ &= m(t) + \frac{j}{\pi t} \otimes m(t) \\ &= m(t) + j\hat{m}(t) \end{aligned}$$

where \otimes is the convolution operator and we have used the Fourier transform pair $\frac{j}{\pi t} \leftrightarrow \text{sgn}(f)$. $\hat{m}(t)$ is the so-called Hilbert transform of $m(t)$ defined below. Recall that multiplication in the frequency domain corresponds to convolution in the time domain.

The new message

$$\begin{aligned} m_+(t) &= m_a(t) \\ &= m(t) + j\hat{m}(t) \end{aligned} \quad (2.24)$$

with only positive frequencies is an *analytic signal* also called the *pre-envelope*.

2.3.2 Hilbert transform definition and properties

From the above section $\hat{m}(t) = m(t) \otimes (1/\pi t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{m(\tau)}{t-\tau} d\tau$ is called the Hilbert Transform of $m(t)$.

The Hilbert transform is a special kind of non-causal filter with impulse response $1/\pi t$. It turns out that the Hilbert transform shifts each sinusoidal component of $m(t)$ by 90 degrees.

For $m(t) = \cos 2\pi f_m t$, $\hat{m}(t) = \cos(2\pi f_m t - \pi/2) = \sin 2\pi f_m t$ for any and all f_m . The proof is in showing that

$$\begin{aligned} \cos 2\pi f_m t \otimes (1/\pi t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos 2\pi f_m \tau}{t-\tau} d\tau \\ &= \sin 2\pi f_m t \end{aligned}$$

The proof is easier in the frequency domain

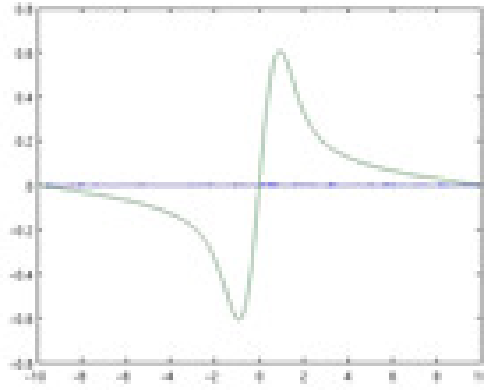
$$\begin{aligned} m(t) &\leftrightarrow M(f) \\ \hat{m}(t) &\leftrightarrow -j \operatorname{sgn}(f) M(f) \\ m_+(t) &= m(t) + j\hat{m}(t) \\ &\leftrightarrow M(f) + \operatorname{sgn}(f) M(f) \\ &\leftrightarrow M_+(f) \end{aligned}$$

Thus the Hilbert transformer is an LTI system with transfer function $H(f) = -j \operatorname{sgn} f$ and impulse response $h(t) = \frac{1}{\pi t}$. The transfer function shows that every positive frequency is multiplied by $-j$ (and every negative frequency is multiplied by $+j$), thus shifting the phase by 90 degrees. For example, for a single tone message

$$\begin{aligned} m(t) &= \cos 2\pi f_m t \\ &= \frac{1}{2} [e^{j2\pi f_m t} + e^{-j2\pi f_m t}] \end{aligned}$$

recall

$$\begin{aligned} e^{j2\pi f_m t} &\leftrightarrow \delta(f - f_m) \\ M(f) &= \frac{1}{2} [\delta(f - f_m) + \delta(f + f_m)] \\ \hat{M}(f) &= -j \operatorname{sgn}(f) M(f) \\ &= \frac{1}{2} [-j\delta(f - f_m) + j\delta(f + f_m)] \\ \hat{M}(f) &= \frac{1}{2j} [\delta(f - f_m) - \delta(f + f_m)] \end{aligned}$$

Figure 2.21: Impulse response of $H(f)$.

thus

$$\begin{aligned}\hat{m}(t) &= \frac{1}{2j} [e^{j2\pi f_m t} - e^{-j2\pi f_m t}] \\ &= \sin 2\pi f_m t\end{aligned}$$

Figure 2.21 shows an approximation of the impulse response. In theory $h(t)$ flips from minus infinity to infinity at $t = 0$.

An analytic signal (or pre-envelope) is defined as $m_+(t) = m(t) + j\hat{m}(t)$, where $\hat{m}(t) = m(t) \otimes (1/\pi t)$ is the Hilbert transform of $m(t)$. The pre-envelope contains only positive frequencies. We will use this pre-envelope to create an SSB-SC signal with only the upper sideband, also called simply an *upper sideband (USB)* signal.

The reasoning above can be repeated to define a pre-envelope with only negative frequencies $m_-(t) = m(t) - j\hat{m}(t)$ which is used to create a *lower sideband (LSB)* signal.

2.3.3 SSB-SC signal derived from pre-envelope

An SSB-SC signal is created by taking the analytic signal or pre-envelope as the message signal, shifting it up to the carrier frequency, and taking the real part. Recall from the Fourier transform properties that multiplying any signal by a complex exponential at f_c shifts the spectrum of that signal by f_c . For a baseband message signal $M(f)$ centered at $f = 0$, the spectrum is shifted so that it is centered at f_c .

$$\begin{aligned}m(t) &\leftrightarrow M(f) \\ m(t)e^{j2\pi f_c t} &\leftrightarrow M(f - f_c)\end{aligned}$$

For DSB-SC, $s(t) = m(t)c(t)$. For SSB-SC with upper sideband only

$$\begin{aligned} s(t) &= \text{Re}\{[m(t) + j\hat{m}(t)]e^{j2\pi f_c t}\} \\ &= \text{Re}\{[m(t) + j\hat{m}(t)][\cos 2\pi f_c t + j \sin 2\pi f_c t]\} \\ &= m(t) \cos 2\pi f_c t - \hat{m}(t) \sin 2\pi f_c t \end{aligned}$$

In general, an SSB-SC signal with a general message can be written

$$s(t) = m(t) \cos 2\pi f_c t \mp \hat{m}(t) \sin 2\pi f_c t$$

where the minus sign is for upper sideband and the plus sign for lower sideband.

If $m(t) = \cos(2\pi ft)$, then $\hat{m}(t) = \sin(2\pi ft)$ where we denote the Hilbert transform by the hat \wedge symbol, and

$$\begin{aligned} s(t) &= \cos 2\pi f_m t \cos 2\pi f_c t \mp \sin 2\pi f_m t \sin 2\pi f_c t \\ &= \cos 2\pi(f_c \pm f_m)t \end{aligned} \tag{2.25}$$

which is the single sideband at $f_c \pm f_m$ (plus for USB, minus for LSB).

Thus for USB, $I(t) = m(t)$ and $Q(t) = \hat{m}(t)$. For a single tone message $m(t) = A_m \cos 2\pi f_m t$, $I(t) = \cos(2\pi f_m t)$ and $Q(t) = \sin(2\pi f_m t)$. The constant A_c accounts for the signal power.

A single sideband USB signal $s(t) = m(t) \cos 2\pi f_c t - \hat{m}(t) \sin 2\pi f_c t$ can be written as

$$\begin{aligned} s(t) &= \text{Re}\{a(t)e^{j\phi}e^{j2\pi f_c t}\} \\ &= \text{Re}\{[m(t) + j\hat{m}(t)]e^{j2\pi f_c t}\} \end{aligned}$$

Thus the complex envelope of an SSB-SC signal $a(t)e^{j\phi} = m(t) + j\hat{m}(t)$ is an analytic signal created from the (real) message $m(t)$.

2.3.4 Single tone modulation for SSB

$$\begin{aligned} m(t) &= A_m \cos 2\pi f_m t \\ \hat{m}(t) &= A_m \sin 2\pi f_m t \end{aligned}$$

$$\begin{aligned} s(t) &= \frac{A_c}{2} [m(t) \cos 2\pi f_c t \mp \hat{m}(t) \sin 2\pi f_c t] \\ &= \frac{A_c A_m}{2} [\cos 2\pi f_m t \cos 2\pi f_c t \mp \sin 2\pi f_m t \sin 2\pi f_c t] \\ &= \frac{A_c A_m}{2} \cos[2\pi(f_c \pm f_m)t] \end{aligned}$$

using $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$.

$s(t)$ is a single tone at frequency $f_c \pm f_m$.

2.3.5 SSB-SC in time and frequency domain

$$\begin{aligned} m(t) &\Leftrightarrow M(f) \\ s(t) &\Leftrightarrow S(f) \\ &\Leftrightarrow M(f - f_c) \end{aligned}$$

for $f > f_c$ where f_c is a positive frequency.

Proof

$$s(t) = \frac{A_c}{2} [m(t) \cos 2\pi f_c t \mp \hat{m}(t) \sin 2\pi f_c t] \begin{cases} -, & \text{USB} \\ +, & \text{LSB} \end{cases}$$

$$\hat{M}(f) = M(f)H(f)$$

$$= -j \operatorname{sgn}(f)M(f)$$

$$S(f) = \frac{A_c}{2} [M(f) \otimes \frac{1}{2} \{\delta(f - f_c) + \delta(f + f_c)\}]$$

$$\mp \frac{A_c}{2} (-j \operatorname{sgn}(f))M(f) \otimes \frac{1}{2j} \{\delta(f - f_c) - \delta(f + f_c)\}]$$

$$S(f) = \frac{A_c}{4} [M(f) \otimes \delta(f - f_c) + M(f) \otimes \delta(f + f_c)$$

$$\pm \operatorname{sgn}(f)M(f) \otimes \delta(f - f_c)$$

$$\mp \operatorname{sgn}(f)M(f) \otimes \delta(f + f_c)]$$

$$S(f) = \frac{A_c}{4} [M(f - f_c) + M(f + f_c)$$

$$\pm \operatorname{sgn}(f - f_c)M(f - f_c) \mp \operatorname{sgn}(f + f_c)M(f + f_c)]$$

$$S(f) = \frac{A_c}{2} M(f - f_c)$$

where $f - f_c > 0$, $f > f_c + \frac{A_c}{2}M(f + f_c)$, $f + f_c > 0$, $f < -f_c$.

2.3.6 SSB-SC transmitter 1 using Hilbert transform

To create an SSB-SC signal at f_c using the SSB-SC equations above, we can:

1. Take the Hilbert transform $\hat{m}(t) = m(t) \otimes (1/\pi t)$ of the message.
2. Create the analytic signal $\tilde{m}(t) = m(t) + j\hat{m}(t)$ for USB or $\tilde{m}(t) = m(t) - j\hat{m}(t)$ for LSB
3. Upconvert it to the desired carrier frequency by multiplying by $e^{j2\pi f_c t}$
4. Take the real part.

In GNU Radio, the upconversion to a radio frequency (RF) wave at f_c (steps 3 and 4) is the function of the *USRP Sink* block. Thus an SSB signal is generated by a USRP sink block (standard IQ transmitter) with inputs $i(t) = m(t)$ and $q(t) = \hat{m}(t)$ for USB.

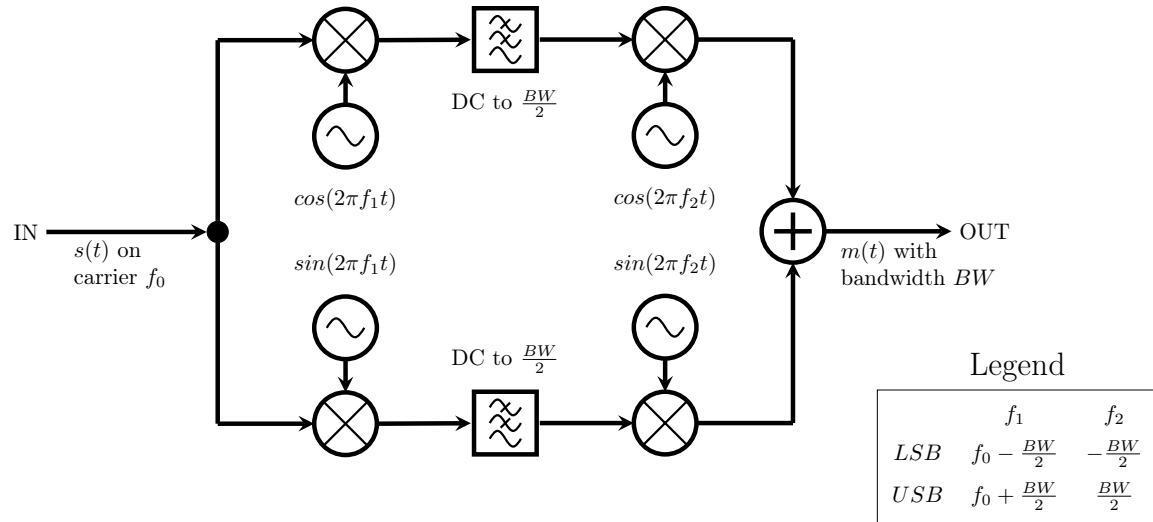


Figure 2.22: Real block diagram for Weavers modulator.

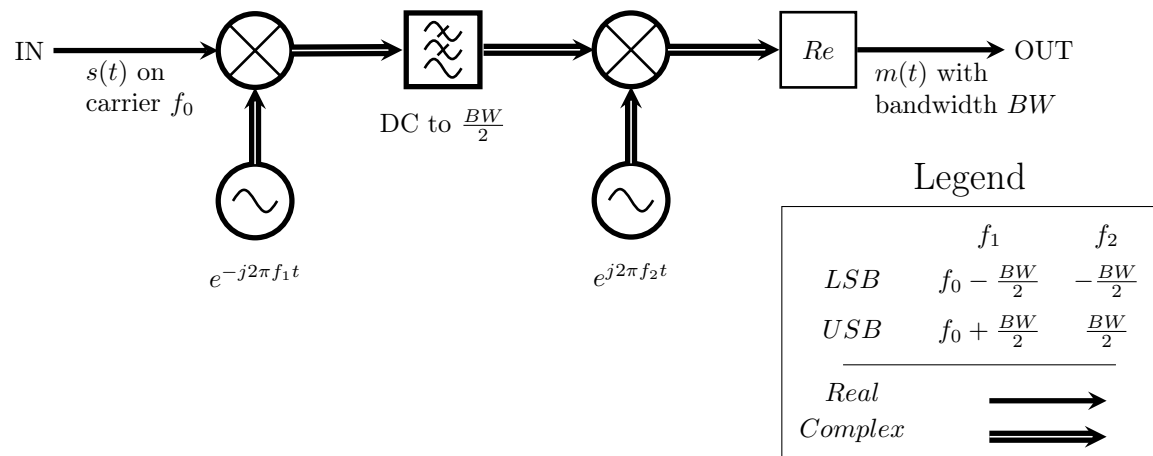


Figure 2.23: Complex block diagram for Weavers modulator.

2.3.7 SSB-SC transmitter 2, avoids Hilbert transform

The Hilbert transform can be inconvenient to compute. An SSB-SC signal can be created without using the Hilbert transform with the so-called Weaver modulator. The block diagram of this modulator in real form is shown in fig. 2.22. The complex form is shown in fig. 2.23.

We will now do an exercise where the signs of the exponentials are chosen for USB.

LO1 operating at f_1 is chosen to be in the centre of the band of the message $m(t)$, e.g. $f_1 = 1500$ Hz for a voice bandwidth message 300–2700 Hz.

LO2 operates at $f_c + f_1$.

In the figure below, we observe how the multiplying the message by the complex exponentials

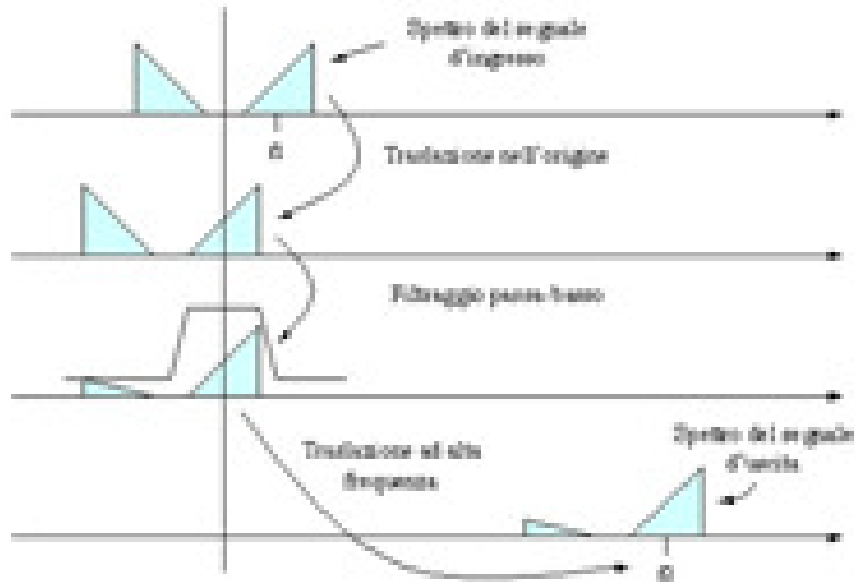


Figure 2.24: Stages of a USB Weaver's modulator

shift the spectrum of the message.

The operations to create the USB signal are as follows:

1. The real message spectrum $m(t)$ is first shifted down by f_1 thus making it a complex signal $\tilde{s}_1(t)$ with an asymmetrical spectrum.
2. $\tilde{s}_1(t)$ is low pass filtered to remove the lower sideband resulting in $\tilde{s}_2(t)$ which is now a complex signal
3. $\tilde{s}_2(t)$ is shifted up by $f_c + f_1$ resulting in another complex signal related to $s(t)$ with positive frequencies only.
4. The real part of this signal is selected, or the real and imaginary parts of this signal are added together, resulting in a real signal $s(t)$ with positive and negative frequency components.
5. The net result of the two frequency conversions is that the zero frequency component of $m(t) \leftrightarrow M(f)$ has been shifted up to the suppressed carrier frequency f_c and the components of $m(t) \leftrightarrow M(f)$ above zero frequency (i.e. the message content) are shifted to be above f_c , i.e. $M(f) \rightarrow M(f - f_c)$

The operations to create the USB signal are written mathematically as follows. At the input there is the message, $m(t)$. After the complex multiply with $e^{-j2\pi f_1 t}$ we have

$$\begin{aligned}\tilde{s}_1(t) &= m(t)e^{-j2\pi f_1 t} \\ &= m(t) \cos 2\pi f_1 t - jm(t) \sin 2\pi f_1 t\end{aligned}$$

and after the LPF filter the result, \tilde{s}_2 is the analytic signal shifted by f_1 . This does not arise from

algebra, but is clear from the frequency diagram.

$$\begin{aligned}\tilde{s}_2(t) &= [m(t) + j\hat{m}(t)]e^{-j2\pi f_1 t} \\ &= m(t) \cos 2\pi f_1 t - j\hat{m}(t) \sin 2\pi f_1 t + j\hat{m}(t) \cos 2\pi f_1 t + \hat{m}(t) \sin 2\pi f_1 t\end{aligned}$$

After the mixer we have

$$\begin{aligned}\tilde{s}_3(t) &= \tilde{s}_2(t)e^{j2\pi(f_c+f_1)t} \\ &= [m(t) + j\hat{m}(t)]e^{-j2\pi f_1 t}e^{j2\pi(f_c+f_1)t} \\ &= [m(t) + j\hat{m}(t)]e^{j2\pi f_c t} \\ &= m(t) \cos 2\pi f_c t - \hat{m}(t) \sin 2\pi f_c t + j[\hat{m}(t) \cos 2\pi f_c t + m(t) \sin 2\pi f_c t] \\ &= s(t)\end{aligned}$$

as the SSB-SC output USB. We can choose i or q .

Exercise: mathematically show low pass filtering $\tilde{s}_1(t)$ to get $\tilde{s}_2(t)$. Hint: write $m(t) = [m_+(t) + m_-(t)]/2$

Exercise: repeat the figure and mathematics for LSB

It is also possible to make a real USB signal by adding the real and imaginary branch outputs together, adding $i + q$ not $i + jq$. If i and q are added

$$\begin{aligned}s(t) &= m(t) \cos 2\pi f_c t - \hat{m}(t) \sin 2\pi f_c t + \hat{m}(t) \cos 2\pi f_c t + m(t) \sin 2\pi f_c t \\ &= m(t) [\cos 2\pi f_c t + \sin 2\pi f_c t] + \hat{m}(t) [\cos 2\pi f_c t - \sin 2\pi f_c t] \\ &= m(t) \cos(2\pi f_c t - \pi/4)/\sqrt{2} + \hat{m}(t) \cos(2\pi f_c t + \pi/4)/\sqrt{2} \\ &= m(t) \cos(2\pi f_c t - \pi/4)/\sqrt{2} + \hat{m}(t) \cos(2\pi f_c t - \pi/4 + \pi/2)/\sqrt{2} \\ &= m(t) \cos(2\pi f_c t - \pi/4)/\sqrt{2} - \hat{m}(t) \sin(2\pi f_c t - \pi/4)/\sqrt{2}\end{aligned}$$

thus

$$\begin{aligned}s(t) &= \text{Re}\{\tilde{s}_3(t)e^{-j\pi/4}\} \\ &= \text{Re}\{[m(t) + j\hat{m}(t)]e^{j(2\pi f_c t - \pi/4)}/\sqrt{2}\}\end{aligned}$$

using

$$\begin{aligned}\cos(\alpha + \pi/4) &= \cos \alpha \cos(\pi/4) - \sin \alpha \sin(\pi/4) \\ &= \sqrt{2}[\cos \alpha - \sin \alpha] \\ \cos(\alpha - \pi/4) &= \cos \alpha \cos(\pi/4) + \sin \alpha \sin(\pi/4) \\ &= \sqrt{2}[\cos \alpha + \sin \alpha] \\ \cos(\alpha + \pi/2) &= -\sin \alpha\end{aligned}$$

This completes the discussion of the operations in an SSB-SC transmitter that creates an SSB-SC signal $s(t) = m(t) \cos 2\pi f_c t \mp \hat{m}(t) \sin 2\pi f_c t$ with message $m(t)$, where the upper sign is for USB and the lower sign is for LSB.

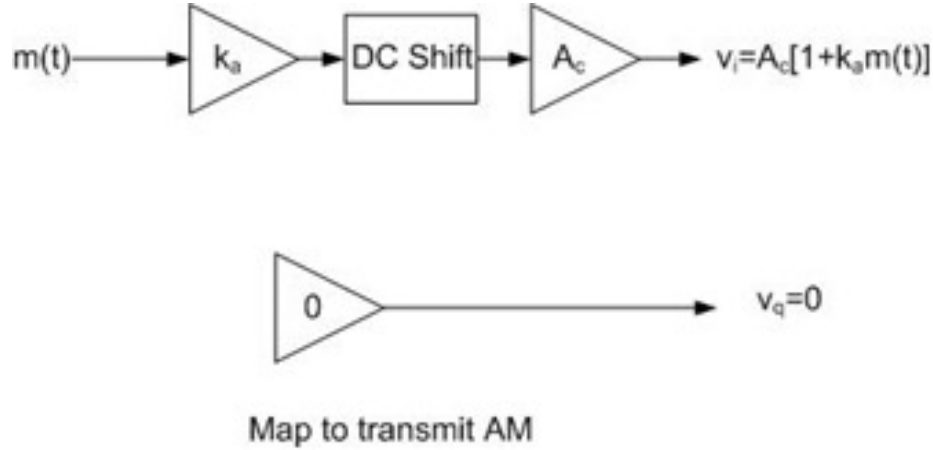


Figure 2.25: Map to transmit AM

2.4 Review of AM, DSB, SSB transmitters in IQ format

In preparation for discussing receivers, we review the AM, DSB, SSB signals in IQ format. Each transmission method (AM, DSB-SC, SSB, etc) requires its own unique map that serves to form $v_i(t)$ and $v_q(t)$ from the message $m(t)$.

2.4.1 Amplitude Modulation:

We already know that $s(t) = A_c[1 + k_a m(t)] \cos(2\pi f_c t)$, and for the special case where $m(t) = \cos 2\pi f_m t$ then $s(t) = A_c[1 + k_a A_m \cos(2\pi f_m t)] \cos(2\pi f_c t)$. Since in general $s(t) = v_i(t) \cos(2\pi f_c t) - v_q(t) \sin(2\pi f_c t)$, to form $v_i(t)$ and $v_q(t)$ in the IQ transmitter/receiver we see that $v_i(t) = A_c[1 + k_a A_m \cos(2\pi f_m t)]$ and $v_q(t) = 0$. Thus our map must produce such functions.

Recall that $v_i(t) = a(t) \cos \psi(t)$ and $v_q(t) = a(t) \sin \psi(t)$, thus $a(t) = A_c[1 + k_a m(t)] = A_c[1 + k_a A_m \cos(2\pi f_m t)]$ and $\psi(t) = 0$.

We could also choose $\psi(t) \neq 0$ and write

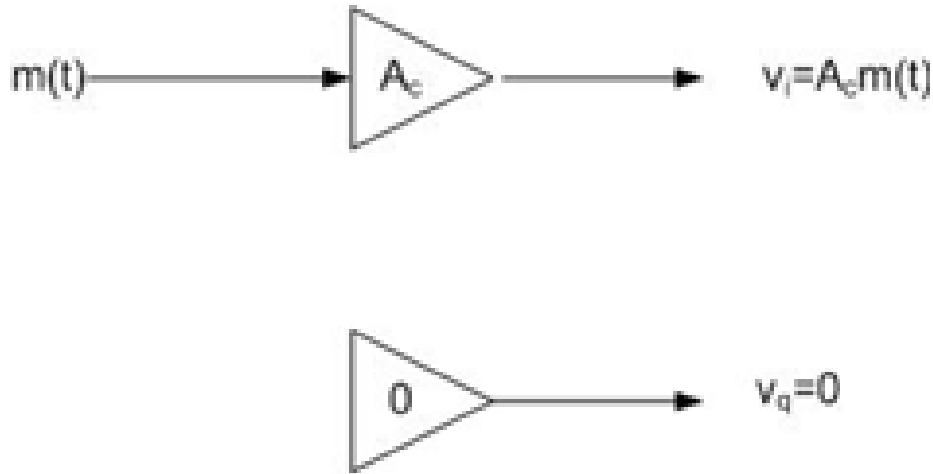
$$\begin{aligned} v_i(t) &= a(t) \cos \psi(t) = A_c[1 + k_a m(t)] \cos \psi(t) \\ v_q(t) &= a(t) \sin \psi(t) = A_c[1 + k_a m(t)] \sin \psi(t) \end{aligned}$$

The value of the phase $\psi(t)$ does not affect the message contained within $a(t)$.

2.4.2 DSB-SC

For a double sideband signal we know that the only difference from the AM case is the lack of carrier term. Thus with a message $m(t) = A_m \cos 2\pi f_m t$ the DSB-SC signal is represented by

$$\begin{aligned} s(t) &= A_c A_m \cos(2\pi f_m t) \cos(2\pi f_c t) \\ &= \frac{A_c A_m}{2} [\cos(2\pi[f_c - f_m]t) + \cos(2\pi[f_c + f_m]t)] \end{aligned}$$



Map to transmit DSB-SC

Figure 2.26: Map to transmit DSB-SC

where we let the constant $A_c A_m$ represent any scaling that has taken place by way of amplification or any hardware effects. For DSB-SC, $v_i(t) = A_c \cos(2\pi f_m t)$ and $v_q(t) = 0$.

2.4.3 SSB-SC

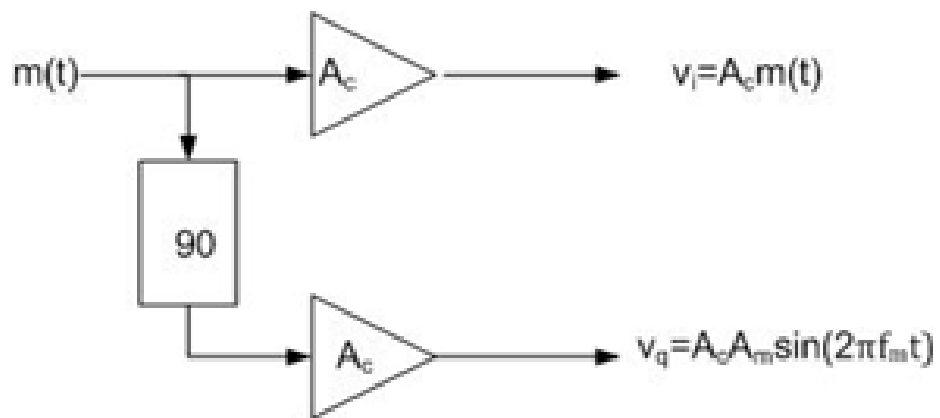
SSB type modulation might be understood as double sideband with one of the sidebands removed. It would follow that adding something to the double sideband map would be adequate to generate single sideband signal. We require a transform which when taken of a certain signal, would negate that signal if the two were added together. We may use, for this purpose, the Hilbert Transform that shifts all frequencies by 90 degrees.

In general $s(t) = m(t) \cos 2\pi f_c t \mp \hat{m}(t) \sin 2\pi f_c t$ where the minus sign is for upper sideband and the plus sign for lower sideband

If $m(t) = \cos(2\pi f t)$, then $\hat{m}(t) = \sin(2\pi f t)$ where we denote the Hilbert transform by the hat \wedge symbol, and

$$\begin{aligned} s(t) &= \cos 2\pi f_m t \cos 2\pi f_c t \mp \sin 2\pi f_m t \sin 2\pi f_c t \\ &= \cos 2\pi(f_c \pm f_m)t \end{aligned}$$

Thus the map for SSB is $v_i(t) = m(t)$ and $v_q(t) = \hat{m}(t)$. For a single tone message $m(t) = A_m \cos 2\pi f_m t$, $v_i(t) = \cos(2\pi f_m t)$ and $v_q(t) = \sin(2\pi f_m t)$. The constant A_c accounts for the signal power.



Map to transmit SSB-SC

Figure 2.27: Map to transmit SSB-SC

Chapter 3

Amplitude modulation (AM) receivers

The purpose of a receiver is to extract or recover the message $m(t)$ from the modulated wave $s(t)$. There exist many different kinds of receiver structures and designs. Digital and analog methods are used, each having a specific structure that is unique to the method of modulation used. (We shall focus here on AM, DSB, and SSB).

3.1 Analog AM receiver

Analog demodulation of the received AM wave (shown in eq. (3.1) scaled by the path loss L_0) can be achieved via an envelope detector made from analog circuit components. The envelope represents the modulating message wave. The envelope detector extracts the message through the use of a rectifying diode and a RC lowpass filter with cutoff frequency chosen so that the high frequency carrier term at f_c is removed while the message with lower frequency f_m plus a DC term remains. A series capacitor is used to remove the DC.

$$\begin{aligned} r(t) &= \frac{s(t)}{L_0} \\ &= \frac{a(t)}{L_0} \cos(2\pi f_c t + \phi) \\ &= \frac{A_c}{L_0} [1 + k_a m(t)] \cos(2\pi f_c t + \phi) \end{aligned} \tag{3.1}$$

Mathematically speaking, the analog envelope detector finds the magnitude of the complex envelope of $s(t)$. Figure 3.1 shows the waveform as it exists at different points in the circuit.

Exercise: do the mathematics to show that if $s(t)$ is the input to the envelope detector circuit with rectifier and low pass filter $s(t)$ then the output of the circuit is the magnitude of the complex envelope of $s(t)$.

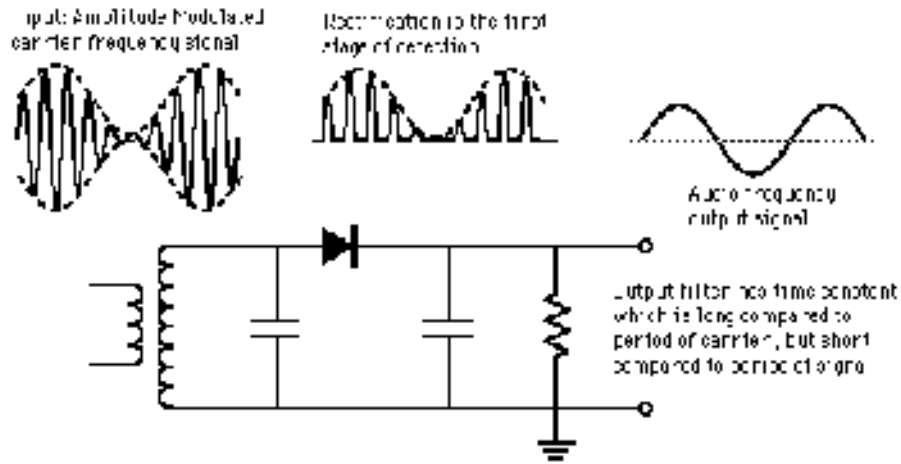


Figure 3.1: AM envelope detector with waveforms shown at each stage.

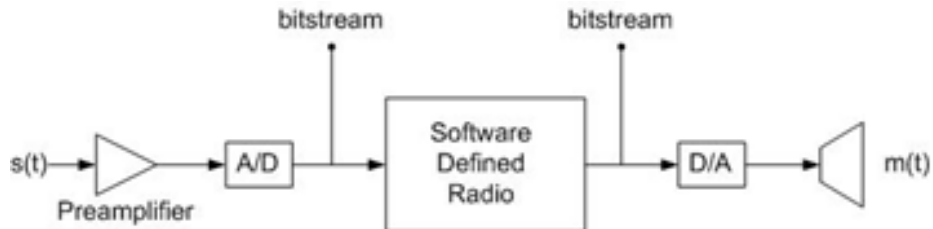


Figure 3.2: Demodulation principles block diagram

Exercise: write expressions for the signal at the input to the diode, at the output of the diode before the low pass filter, and at the output of the low pass filter.

3.2 Digital software receivers

Figure 3.2 is a representation of the main principles involved in demodulating an AM signal using a digital architecture.

3.2.1 Digital software AM receiver in complex notation

The transmitted AM signal is of the form:

$$\begin{aligned}
 s(t) &= a(t) \cos(2\pi f_c t + \phi) \\
 &= A_c [1 + k_a m(t)] \cos(2\pi f_c t + \phi)
 \end{aligned}$$

where ϕ_t is a constant phase.

The received AM signal $r(t) = s(t)/L_0$ will in general have a different phase ϕ_r so that

$$\begin{aligned} r(t) &= a(t) \cos(2\pi f_c t + \phi_r) \\ &= A_c [1 + k_a m(t)] \cos(2\pi f_c t + \phi_r) \end{aligned}$$

We have assumed $L_0 = 1$ for convenience so that $r(t) = s(t)$ and write $\phi = \phi_r$. In practice, $L_0 \simeq 10^{-4}$ to 10^{-6} so that the received signal is only a few microvolts. The preamplifier stage with voltage gain up to 10^6 is needed to restore $r(t)$ to a level that is suitable for the ADC (typically a few volts).

The AM receiver recovers $m(t)$ from $s(t)$. One method is to recover the magnitude of the complex envelope $a(t) = 1 + k_a m(t)$ and subtract the DC component to obtain $m(t)$.

To show how this is done in software with the USRP and GNU Radio Companion (GRC), recall that the USRP source block has a complex output with real and imaginary components $i(t)$ and $q(t)$.

In what follows we assume $L_0 = 1$ so that $r(t) = s(t)$. We can write the AM signal $s(t)$ as the real part of a complex signal:

$$\begin{aligned} s(t) &= \text{Re}[a(t)e^{j\phi}e^{j2\pi f_c t}] \\ &= \text{Re}[\tilde{s}(t)e^{j2\pi f_c t}] \end{aligned}$$

where the complex envelope

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi} \\ &= a(t) \cos \phi + ja(t) \sin \phi \\ &= i(t) + jq(t) \end{aligned}$$

Thus the USRP source block with frequency set to f_c will have outputs:

$$\begin{aligned} &i(t)a(t) \cos \phi \\ q(t) &= a(t) \sin \phi \end{aligned}$$

To obtain $a(t)$, we take the magnitude of the complex envelope $\tilde{s}(t)$. Thus we can write:

$$\begin{aligned} |\tilde{s}(t)| &= |i(t) + jq(t)| \\ &= |a(t) \cos \phi + ja(t) \sin \phi| \\ &= a(t) |\cos \phi + j \sin \phi| \\ &= a(t) \sqrt{\cos^2 \phi + \sin^2 \phi} \\ &= a(t) \end{aligned}$$

This shows that we can recover $a(t) = 1 + k_a m(t)$ regardless of the value of ϕ . The GRC *Complex to Magnitude* block allows us to obtain the magnitude of the complex envelope by performing the function $a(t) = |i(t) + jq(t)|$.

If there is frequency offset, then $\phi = 2\pi\Delta f t$, but as we have just seen, $|\tilde{s}(t)| = a(t)$ is not affected by the value of ϕ and thus not affected by any time variation in ϕ caused by a frequency offset Δf , provided that $s(t)$ is within the filter bandwidth of the receiver.

Exercise: write an expression for $s(t)$ and $\tilde{s}(t)$ with a frequency offset.

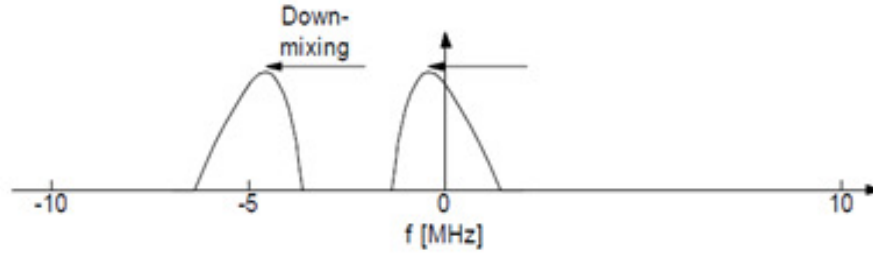


Figure 3.3: Downmixing

3.2.2 Tuning in (selecting) one particular AM signal with the USRP

The USRP multiplies the real valued radio frequency signal $s(t)$ by $e^{-j2\pi f_c t}$ to generate $i(t) + jq(t)$. This process is called *complex downmixing* and is equivalent to the standard IQ receiver shown in fig. 1.9.

Recall the AM spectrum

$$S(f) = \frac{A_c}{2} [\delta(f - f_c) + \delta(f + f_c)] + \frac{A_c k_a}{2} [M(f - f_c) + M(f + f_c)] \quad (3.2)$$

The spectrum $S(f)$ of the real radio frequency (RF) signal $s(t)$ is symmetric about zero. The complex downmixing shifts the frequency of $S(f)$ down by f_c .

$$\begin{aligned} s(t) &\leftrightarrow S(f) \\ e^{-j2\pi f_c t} s(t) &\leftrightarrow S(f + f_c) \end{aligned}$$

Exercise: write an expression for $S(f + f_c)$, and show that one of the terms needs to be filtered out.

Answer: The shifted AM spectrum is

$$S(f + f_c) = \frac{A_c}{2} [\delta(f) + \delta(f + 2f_c)] + \frac{A_c k_a}{2} [M(f) + M(f + 2f_c)] \quad (3.3)$$

After complex downmixing, the resulting signal is complex and the frequency spectrum $S(f + f_c)$ is no longer symmetric about zero. The terms at $-2f_c$ are filtered out, leaving only the message plus a DC component.

The complex signal output from the USRP source block $i(t) + jq(t)$ is bandlimited to the sampling rate of the USRP source block. The USRP source block output can be recorded to a file and used again at a later time. This file source will have the same sampling rate and bandwidth as the USRP sink block used to record it.

With a sampling rate of 256 kHz and complex samples, the bandwidth will be 256 kHz (because the complex signal spectrum is not symmetric and does not have redundant mirror-image positive and negative frequencies).

The AM aircraft band 108–137 MHz is divided into 25 kHz channels. With a file source sampled at 256 kHz, there can be as many as 10 different AM aircraft signals.

The AM aircraft signal with carrier frequency $f_c = f_d$ (f_d is set in the USRP source block) will appear at 0 Hz after the downconversion (at the USRP source output). Other signals at carrier frequencies $f_c \pm n f_0$ kHz will appear at multiples of $f_0 = 25$ Hz away from 0 Hz.

We need to create a filter to select the one signal we want (the one with carrier frequency f_c that now appears at 0 Hz). A low pass filter with 12.5 kHz cutoff frequency will do the job, since all the other signals are centered at frequencies at least 25 kHz away from 0 Hz.

To “tune in” (receive) one of the other signals, we can shift the spectrum of the USRP source output by $n f_0$ kHz by multiplying the complex signal $i(t) + jq(t)$ by $e^{-j2\pi n f_0 t} = \cos 2\pi n f_0 t - j \sin 2\pi n f_0 t$, so that the signal that first appeared at $n f_0$ Hz now appears at 0 Hz.

Exact tuning is not required, since a frequency offset Δf results in a changing phase $\phi(t) = 2\pi \Delta f t$ that does not affect the AM demodulation process.

3.3 DSB-SC receiver

Recall this time that a DSB wave is comprised of only the two (redundant) message frequency bands (the carrier has been filtered out before transmission):

$$\begin{aligned} s(t) &= m(t) c(t) \\ &= m(t) \cos(2\pi f_c t) \end{aligned} \tag{3.4}$$

The modulated double sideband signal has no carrier frequency and therefore exhibits a periodic phase reversal which makes the simple envelope detector described previously inadequate for demodulation (the rectifying diode will leave us with a rectified version of the envelope: received envelope = $|m(t)|$).

How to properly receive $m(t)$ from $s(t)$ for a DSB-SC signal? One general method may be understood by observing the result of multiplying the modulated wave $s(t)$ by the carrier frequency, $c(t)$ (fig. 3.4). Let $r(t)$ represent the received signal within the receiver structure.

$$\begin{aligned} r(t) &= s(t)c(t) \\ &= m(t)c(t)c(t) \\ &= m(t) \cos^2(2\pi f_c t) \\ &= m(t) \left[\frac{1}{2}(1 + \cos 4\pi f_c t) \right] \\ &= \frac{1}{2}m(t) + \frac{1}{2} \cos 4\pi f_c t \end{aligned}$$

Thus we have created a signal comprised of a high frequency term (a DSB signal and an amplitude scaled version of our message. All that must be done to extract $m(t)$ from the signal is to apply a lowpass filter to the signal: $r_{LP}(t) = \frac{1}{2}m(t)$ (fig. 3.5).

This method, however effective, requires the party operating the receiver to know the exact carrier frequency, f_c and the phase of the carrier wave. Both instances of the carrier wave (that used here at the receiver and that which is used initially in the modulation process at the transmitter) must both have exactly the same phase and frequency.

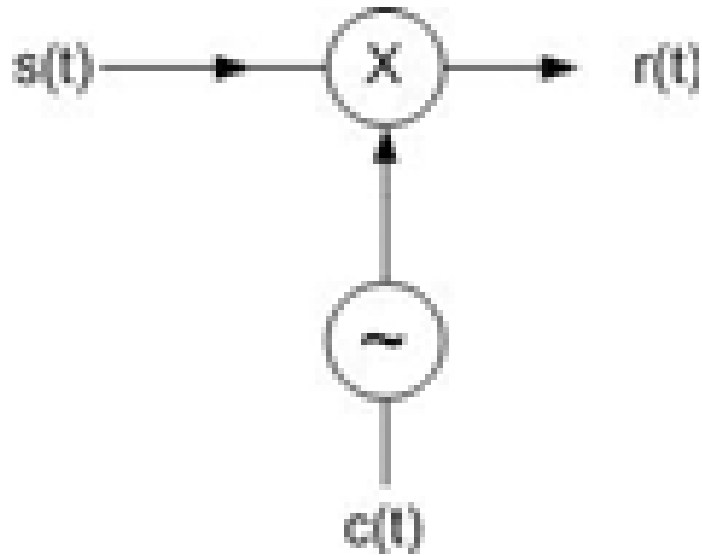
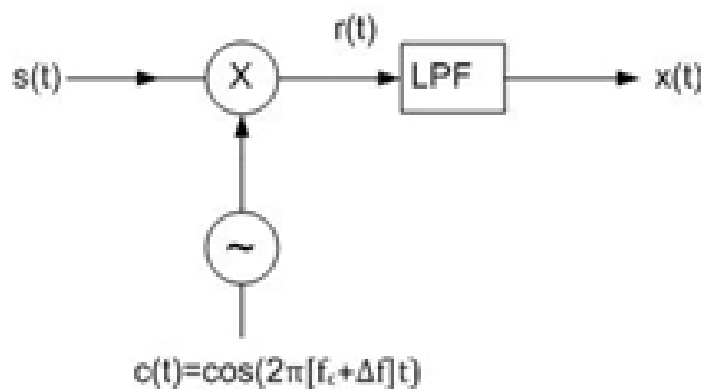


Figure 3.4: DSB receiver first step

Figure 3.5: Production of Signal $x(t)$ with LPF

3.3.1 DSB receiver with frequency and phase offset

When the frequency and phase are not exactly the same, the error terms (frequency error Δf and phase shift ϕ) may be represented by $\psi = 2\pi\Delta ft + \phi$. Using the appropriate trig identities, we now may write the received signal before the low pass filter

$$\begin{aligned}
 r(t) &= m(t) \cos(2\pi f_c t) \cos[2\pi f_c t + 2\pi(\Delta f)t + \phi] \\
 &= m(t) \cos(2\pi f_c t) \cos[2\pi f_c t + \psi] \\
 &= m(t) \cos(2\pi f_c t) [\cos(2\pi f_c t) \cos \psi - \sin(2\pi f_c t) \sin \psi] \\
 &= m(t) \cos^2(2\pi f_c t) \cos \psi - m(t) \cos(2\pi f_c t) \sin(2\pi f_c t) \sin \psi \\
 &= \frac{m(t)}{2} [1 + \cos(4\pi f_c t)] \cos \psi - \frac{m(t)}{2} \sin(4\pi f_c t) \sin \psi
 \end{aligned}$$

Thus a more general resultant signal is attained when we take into account the possibility of frequency error and phase shift in the local oscillator that provides the receiver's version of $c(t)$. After lowpass filtering this signal, we achieve the following result:

$$r_{LP}(t) = \frac{m(t)}{2} \cos(\psi) \quad (3.5)$$

So if $\psi = 0$ then we get the ideal case where $\cos \psi = 1$ and the output of the LPF is a scaled version of the message. If, however $\psi \neq 0$ then get above LPF output where the message is multiplied by the cosine of the error term.

Various values of ψ lead to different effects on the extracted message shown in table 3.1.

Specific details of ψ	Effect on $r_{LP}(t)$
$\psi = 2\pi\Delta ft + \phi$	$r_{LP}(t) = \frac{m(t)}{2} \cos(2\pi\Delta ft + \phi)$ Message will be scaled and will be distorted
$\Delta f = 0$ $\psi = \phi = \text{constant}$	$r(t) = k \frac{m(t)}{2}$ where $k = \cos \phi$ is a scaling factor Message can be recovered
$\phi = 0, \psi = 2\pi\Delta ft$	$r(t) = \frac{m(t)}{2} \cos(2\pi\Delta ft)$ Message will include distortion
$\psi = 0$	$r_{LP}(t) = \frac{m(t)}{2}$ Exact message is readily recoverable

Table 3.1: Effect of various ψ values on $r_{LP}(t)$

The distorted signals can be interpreted by considering the frequency error cosine term $\cos 2\pi\Delta ft$ to be a sort of low frequency “carrier”, where typically $\Delta f = 100$ Hz.

Assuming $m(t) = \cos 2\pi f_m t$, with typical $f_m = 1000$ Hz the receiver output

$$\begin{aligned}
 r(t) &= m(t) \cos 2\pi\Delta ft = \cos 2\pi f_m t \cos 2\pi\Delta ft \\
 &= 0.5 \cos 2\pi(f_m + \Delta f)t + 0.5 \cos 2\pi(f_m - \Delta f)t
 \end{aligned}$$

has outputs with frequencies at $f_m + \Delta f$ and $f_m - \Delta f$, so the message at frequency f_m now appears in two places, at $f_m - \Delta f$ and $\Delta f + f_m$. For our example, with frequency error 100 Hz, the message at 1000 Hz now appears at 900 and 1100 Hz.

We may extract the true message frequency only if we know the exact carrier frequency. Typical error ratios compared to carrier frequency ($\frac{\Delta f}{f_c}$) are as follows: a factor of 10^{-4} for cheap equipment, 10^{-6} for a good crystal oscillator, 10^{-11} for a Rubidium oscillator, and approaching 10^{-13} with GPS discipline (control).

Later we will see a method whereby a DSB-SC receiver can find the frequency error and eliminate it.

3.3.2 Review of I-Q receivers

Here we review the general I-Q receiver configuration that may be implemented for all types of signals. The receiver for every type (for AM, DSB-SC, SSB) consists of a standard configuration coupled with a “map” that is specific to the transmission type.

$$\begin{aligned} s(t) &= a(t) \cos[2\pi f_c t + \psi(t)] \\ &= a(t) [\cos(2\pi f_c t) \cos \psi(t) - \sin(2\pi f_c t) \sin \psi(t)] \\ &= a(t) \cos \psi(t) \cos 2\pi f_c t - a(t) \sin \psi(t) \sin 2\pi f_c t \\ &= v_i(t) \cos 2\pi f_c t - v_q(t) \sin 2\pi f_c t \end{aligned}$$

where $a(t)$ is the amplitude and $\psi(t)$ is the phase. The coefficients of the carrier waves in the transmitter are referred to as $v_i(t)$ and $v_q(t)$ respectively, where

$$\begin{aligned} v_i(t) &= a(t) \cos \psi(t) \\ v_q(t) &= a(t) \sin \psi(t) \end{aligned}$$

These signals are formed by the map and used to create $s(t)$ at the IQ Transmitter.

$$s(t) = v_i(t) \cos(2\pi f_c t) - v_q(t) \sin(2\pi f_c t)$$

The IQ Receiver multiplies the incoming signal $s(t)$ by two versions of the carrier wave functions: $\cos(2\pi f_c t)$ and the 90 degree phase-shifted version $\sin(2\pi f_c t)$.

Recall the DSB-SC receiver (fig. 3.4) uses only one version of the carrier wave and thus has a single cosine oscillator.

The general I-Q receiver has two oscillators, \cos and \sin as shown in fig. 3.7.

3.3.3 Review of AM, DSB, SSB receivers in IQ format

The receiver must determine a best guess of the information stored in the signals $v_i(t)$ and $v_q(t)$. The generic IQ receiver architecture produces $x(t)$ and $y(t)$ by multiplying the received signal $s(t)$ by local carriers $\cos 2\pi f_c$ and $\sin 2\pi f_c t$. The algebra describing the production of these signals assuming no frequency error appears below. We will use the identities

$$\begin{aligned} \cos \alpha \cos \beta &= [\cos(\alpha - \beta) + \cos(\alpha + \beta)]/2 \\ \cos^2 \alpha &= [1 + \cos(2\alpha)]/2 \\ \sin \alpha \sin \beta &= [\cos(\alpha - \beta) - \cos(\alpha + \beta)]/2 \\ \sin \alpha \cos \beta &= [\sin(\alpha - \beta) + \sin(\alpha + \beta)]/2 \end{aligned}$$

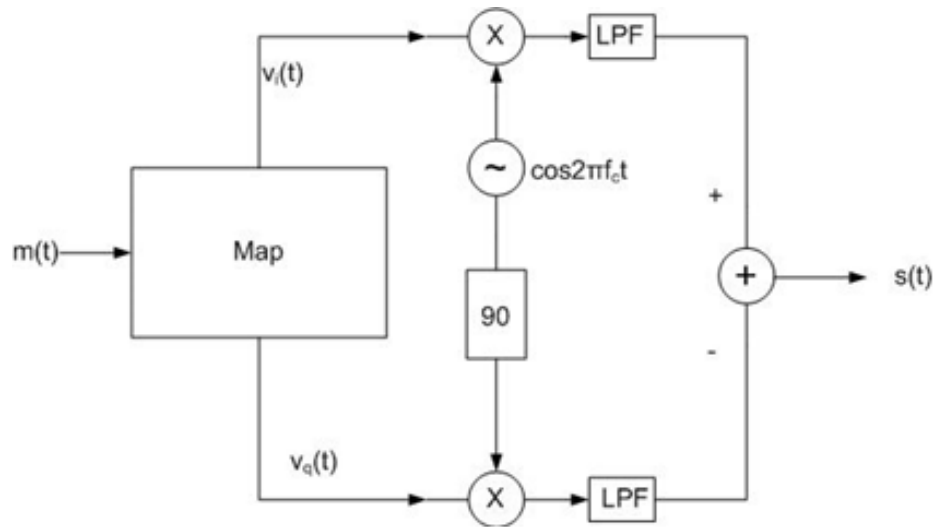


Figure 3.6: General IQ Transmitter

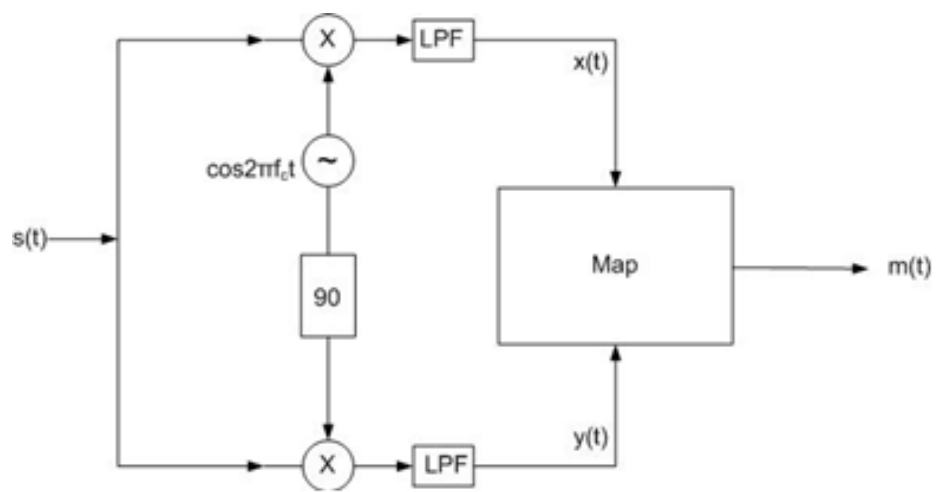


Figure 3.7: General IQ Receiver

For $x(t)$, before the LPF we have,

$$\begin{aligned}
 s(t) \cos 2\pi f_c t &= [v_i(t) \cos(2\pi f_c t) - v_q(t) \sin(2\pi f_c t)] \cos(2\pi f_c t) \\
 &= v_i(t) \cos^2(2\pi f_c t) - v_q(t) \sin(2\pi f_c t) \cos(2\pi f_c t) \\
 &= \frac{v_i(t)}{2} [1 + \cos(2\pi[2f_c]t)] - \frac{v_q(t)}{2} \sin(2\pi[2f_c]t) \\
 &= \frac{v_i(t)}{2} + \frac{v_i(t)}{2} \cos(2\pi[2f_c]t) - \frac{v_q(t)}{2} \sin(2\pi[2f_c]t)
 \end{aligned}$$

Recall $v_i(t)$ and $v_q(t)$ are formed from the message frequency only, so of their frequency components are at or near f_m . The expression $v_i(t) \cos[2\pi(2f_c)t]$ results in two terms at the sum and difference frequencies ($2f_c \pm f_m$), both of which are near double the carrier frequency. The same can be said of $v_q(t) \sin[2\pi(2f_c)t]$. Thus after the LPF is applied to $s(t) \cos 2\pi f_c t$ the only remaining term will be $x(t) = \frac{v_i(t)}{2}$.

Similarly for $y(t)$, before the LPF we have,

$$\begin{aligned}
 s(t) \sin 2\pi f_c t &= [v_i(t) \cos(2\pi f_c t) - v_q(t) \sin(2\pi f_c t)] \sin(2\pi f_c t) \\
 &= v_i(t) \sin(2\pi f_c t) \cos(2\pi f_c t) - v_q(t) \sin^2(2\pi f_c t) \\
 &= \frac{v_i(t)}{2} \sin(2\pi[2f_c]t) - \frac{v_q(t)}{2} + \frac{v_q(t)}{2} \cos(2\pi[2f_c]t)
 \end{aligned}$$

And after the LPF we end up with $y(t) = \frac{v_q(t)}{2}$.

We often write that the IQ receiver output $x(t) = v_i$ and $y(t) = v_q(t)$, thus neglecting the scaling factor $\frac{1}{2}$, since this scaling factor in real hardware will depend on the gains of the amplifiers.

The receiver will still work in the case where $\psi = 2\pi\Delta f t + \phi \neq 0$ as long as the oscillator that produces the signals $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$ is corrected such that its frequency mirrors the error (a voltage controlled oscillator, VCO, outputs at frequency $2\pi(f_c + \Delta f)t + \phi$ so that the apparent carrier frequency received is the same as that which is used in demodulation).

When we include the possibility of frequency error so to represent the most general case, the incoming signal can be represented by $s(t) = v_i(t) \cos(2\pi[f_c + \Delta f]t + \phi) - v_q(t) \sin(2\pi[f_c + \Delta f]t + \phi)$.

3.4 Mapping from IQ receiver output to message signal

Each modulation method (AM, DSB-SC, SSB, etc) requires its own unique map that serves to extract the message $m(t)$ from $x(t)$ and $y(t)$. The maps for receiving $m(t)$ from these signals for different transmission types are outlined below.

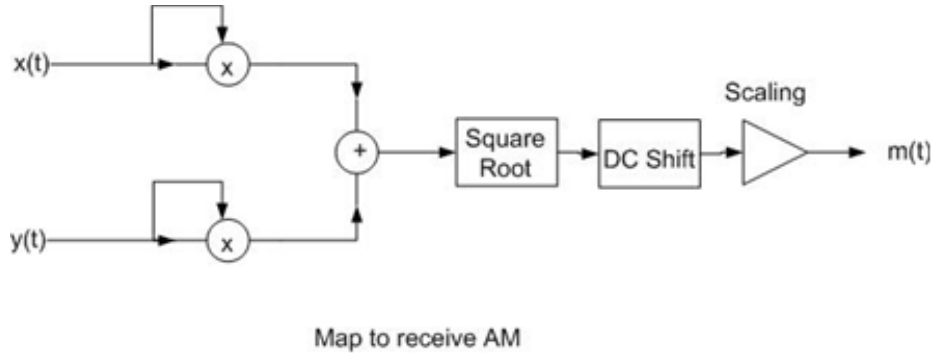


Figure 3.8: Map to receive AM

3.4.1 AM receiver

Recall for AM

$$\begin{aligned}
 x(t) &= v_i(t) \\
 &= a(t) \cos \psi(t) \\
 &= A_c [1 + k_a m(t)] \cos \psi(t) \\
 y(t) &= v_q(t) \\
 &= a(t) \sin \psi(t) \\
 &= A_c [1 + k_a m(t)] \sin \psi(t)
 \end{aligned}$$

The AM receiver map extracts $m(t)$ by first obtaining $a(t) = A_c [1 + k_a m(t)]$, removing the DC component and scaling the signal by $1/k_a$. $a(t)$ is obtained from

$$a(t) = \sqrt{x^2(t) + y^2(t)}$$

since

$$\begin{aligned}
 \sqrt{x^2(t) + y^2(t)} &= \sqrt{a^2(t) \cos^2 \psi(t) + a^2(t) \sin^2 \psi(t)} \\
 &= a(t)
 \end{aligned}$$

regardless of the value of $\psi(t) = 2\pi\Delta f t + \phi$. Thus an AM receiver will function even with a frequency error Δf

3.4.2 DSB-SC receiver with frequency error correction

Recall for DSB-SC, $v_i(t) = m(t)$, $v_q(t) = 0$. For the case where $m(t) = A_c \cos 2\pi f_c t$, $v_i(t) = A_c \cos(2\pi f_m t)$ and $v_q(t) = 0$.

The demodulating carrier frequency must be controlled to correct for the possibility of frequency error Δf such that $\psi = 2\pi\Delta f t$. With a frequency error Δf , the incoming transmitted wave is of the form:

$$s(t) = v_i(t) \cos(2\pi[f_c + \Delta f]t) - v_q(t) \sin(2\pi[f_c + \Delta f]t).$$

For DSB-SC with a single tone message $m(t) = \cos(2\pi f_m t)$ we have $v_i(t) = A_c \cos(2\pi f_m t)$ and $v_q(t) = 0$. Thus

$$\begin{aligned} s(t) &= v_i(t) \cos(2\pi f_c t + 2\pi \Delta f t) \\ &= A_c \cos 2\pi f_m t \cos 2\pi (f_c + \Delta f) t \\ &= A_c \cos 2\pi [(f_c + \Delta f) + f_m] t + \cos 2\pi [(f_c + \Delta f) - f_m] t \end{aligned}$$

To find $x(t)$: before the LPF,

$$\begin{aligned} s(t) \cos 2\pi f_c t &= A_c \cos(2\pi f_m t) \cos 2\pi (f_c + \Delta f) t \cos(2\pi f_c t) \\ &= \frac{A_c}{2} [\cos 2\pi [(f_c + \Delta f) + f_m] t + \cos 2\pi [(f_c + \Delta f) - f_m] t] \cos 2\pi f_c t \\ &= \frac{A_c}{4} [\cos(2\pi [(2f_c + \Delta f) + f_m] t) + \cos(2\pi [f_m + \Delta f] t) \\ &\quad + \cos(2\pi [f_m - (\Delta f)] t) + \cos(2\pi [(2f_c + \Delta f) - f_m] t)] \end{aligned}$$

while after the LPF we have,

$$x(t) = \frac{A_c}{4} [\cos(2\pi [f_m + \Delta f] t) + \cos(2\pi [f_m - \Delta f] t)]$$

For a general message $m(t)$, we find $x(t) = \frac{1}{2} m(t) \cos 2\pi \Delta f t$.

Similarly for $y(t)$: before the LPF,

$$\begin{aligned} s(t) \sin 2\pi f_c t &= A_c \cos 2\pi f_m t \cos 2\pi (f_c + \Delta f) t \sin 2\pi f_c t \\ &= \frac{A_c}{2} [\cos(2\pi [f_m + (f_c + \Delta f)] t) + \cos(2\pi [f_m - (f_c + \Delta f)] t)] \sin(2\pi f_c t) \\ &= \frac{A_c}{4} [\sin(2\pi [f_m + (2f_c + \Delta f)] t) - \sin(2\pi [f_m + (\Delta f)] t) \\ &\quad + \sin(2\pi [f_m - (\Delta f)] t) - \sin(2\pi [f_m - (2f_c + \Delta f)] t)] \end{aligned}$$

while after the LPF we have,

$$y(t) = \frac{A_c}{4} [-\sin(2\pi [f_m + (\Delta f)] t) + \sin(2\pi [f_m - (\Delta f)] t)]$$

For a general message $m(t)$, we find $y(t) = \frac{1}{2} m(t) \sin 2\pi \Delta f t$.

The received signals are distorted by the frequency offset. The DSB-SC receiver does not inherently know the frequency error but it manipulates the signals $x(t)$ and $y(t)$ to acquire it. We multiply $x(t)$ and $y(t)$ together and produce the following control signal for the VCO.

Using $\sin \alpha \cos \alpha = (1/2) \sin 2\alpha$

$$\begin{aligned}
 e(t) &= x(t)y(t) \\
 &= \frac{A_c^2}{16} [-\sin(2\pi[f_m + (\Delta f)]t) + \sin(2\pi[f_m - (\Delta f)]t)] \\
 &\quad \times [\cos(2\pi[f_m + (\Delta f)]t) + \cos(2\pi[f_m - (\Delta f)]t)] \\
 &= \frac{A_c^2}{32} [-\sin(4\pi[f_m + \Delta f]t) - \sin(4\pi f_m t) - \sin(4\pi \Delta f t) \\
 &\quad + \sin(4\pi f_m t) + \sin(-4\pi \Delta f t) + \sin(4\pi[f_m - \Delta f]t)] \\
 &= \frac{A_c^2}{32} [-2 \sin(4\pi \Delta f t)] \\
 &= -\frac{A_c^2}{16} \sin(4\pi \Delta f t)
 \end{aligned}$$

For a general message $m(t)$

$$\begin{aligned}
 e(t) &= x(t)y(t) \\
 &= \frac{1}{4} m^2(t) \cos 2\pi \Delta f t \sin 2\pi \Delta f t \\
 &= \frac{1}{8} m^2(t) \sin 4\pi \Delta f t
 \end{aligned}$$

Typical values of f_m are 300–3000 Hz for voice, whereas a typical value for Δf is 100 Hz or less. $e(t)$ is low pass filtered with a cutoff frequency less than 300 Hz so that all frequency components f_m of $m(t)$ are filtered out. Thus after the low pass filter, $e_{LPF}(t) = \frac{1}{8} \langle m^2(t) \rangle \sin 4\pi \Delta f t$ where $\langle m^2(t) \rangle$ is a DC value representing the average power in $m(t)$. Thus even without knowing the error frequency Δf , we can produce a sinusoid, directly dependant on Δf . This allows us to control the frequency for the local oscillator at the receiver to replicate the incoming carrier frequency. In the special case where

$$\begin{aligned}
 m(t) &= A_c \cos 2\pi f_m t \\
 m^2(t) &= \frac{A_c^2}{2} + \frac{A_c^2}{2} \cos 4\pi f_m t \\
 \langle m^2(t) \rangle &= \frac{A_c^2}{2} \\
 e_{LPF}(t) &= \frac{1}{8} \langle m^2(t) \rangle \sin 4\pi \Delta f t \\
 &= \frac{A_c^2}{16} \sin 4\pi \Delta f t
 \end{aligned}$$

The above analysis can be repeated using a general message $m(t)$ and a general offset $\psi = 2\pi \Delta f t + \phi \neq 0$.

3.4.3 SSB-SC receiver using message-bandwidth low pass filters

We consider two methods of receiving an SSB-SC signal using low pass filters with cutoff frequency equal to the bandwidth B of the message signal. For a single tone message, we choose $f_m = B$.

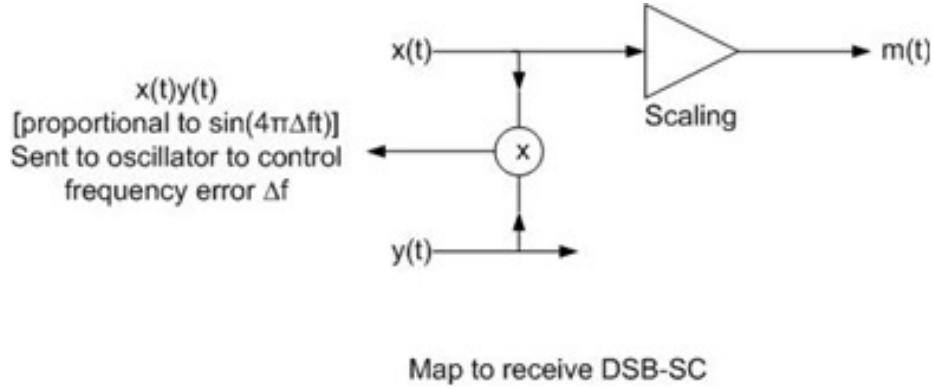


Figure 3.9: Map to receive DSB-SC

In Method 1, we multiply the real SSB signal by $\cos 2\pi f_c t$ and low pass filter to get $m(t)$. This method is used in analog receivers. To demodulate SSB-SC, the receiver is required to know the exact carrier frequency for proper demodulation.

Exercise: show the mathematical steps to obtain $m(t)$ from $s(t)$ using this method.

In method 2, we use an IQ receiver. For an SSB-SC signal $s(t) = m(t) \cos 2\pi f_c t \mp m(t) \sin 2\pi f_c t$ with $-$ for USB and $+$ for LSB. Considering the single tone message $m(t) = A_c \cos 2\pi f_m t$, for

- LSB we write $s(t) = A_c \cos(2\pi[f_c - f_m]t)$,
- USB we write $s(t) = A_c \cos(2\pi[f_c + f_m]t)$.

For USB, the IQ receiver output

$$\begin{aligned} x(t) &= A_c \cos 2\pi(f_c + f_m)t \cos 2\pi f_c t \\ &= \frac{A_c}{2} [\cos 2\pi(2f_c + f_m)t + \cos 2\pi f_m t] \\ y(t) &= A_c \cos 2\pi(f_c + f_m)t \sin 2\pi f_c t \\ &= \frac{A_c}{2} [\sin 2\pi(2f_c + f_m)t - \sin 2\pi f_m t] \end{aligned}$$

comes

$$x(t) = \frac{A_c}{2} \cos 2\pi f_m t$$

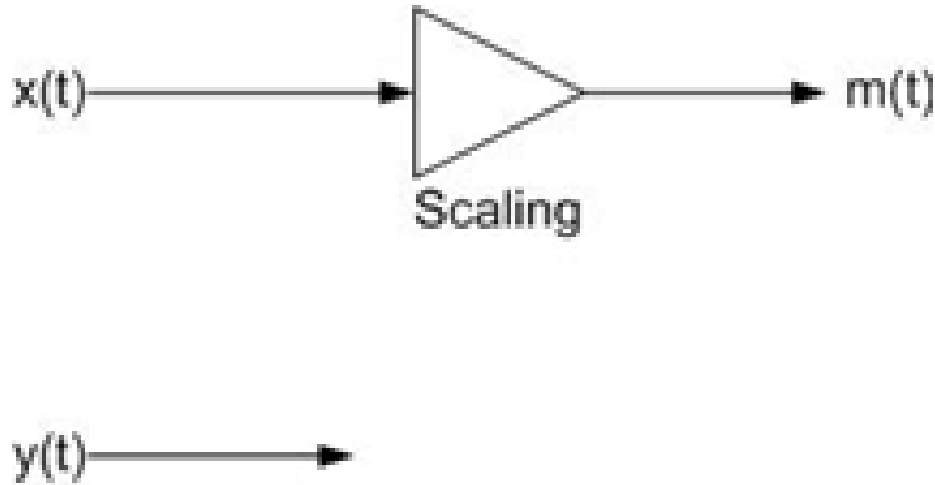
and

$$y(t) = \frac{A_c}{2} \sin 2\pi f_m t$$

after the LPF output. In both cases, an LPF leaves us with the message (cosine term) and a phase shifted version (sine term). We simply choose the signal in $x(t)$ and further demodulation only serves to scale the amplitude of the message.

Exercise: repeat this calculation with a small frequency offset Δf .

The problem with this method is that if there is another signal present where the other (opposite) sideband would have been if we were using DSB or AM instead of SSB, then that signal will also be demodulated and will interfere with the desired message.



Map to receive SSB-SC

Figure 3.10: Map to receive SSB-SC

Exercise: show that this is true for the USB signal $s_{USB}(t) = A_c \cos(2\pi[f_c + f_m]t)$ with an interfering signal $s_{int}(t) = A_c \cos(2\pi[f_c - f_m]t)$.

3.4.4 SSB-SC receiver using Weaver Demodulator description

A third method of receiving SSB-SC uses the Weaver demodulator which is analogous to the Weaver modulator discussed earlier.

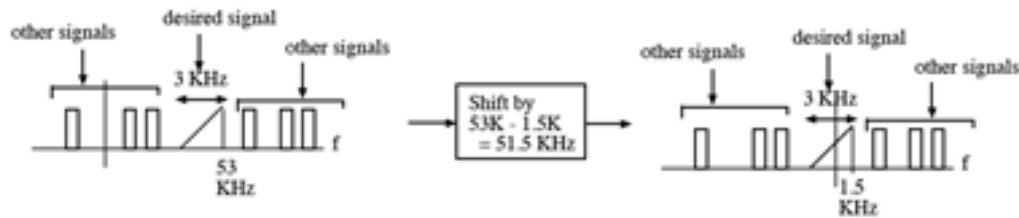
The Weaver demodulator uses a low pass filter with cutoff at $B/2$ (half the message bandwidth) compared to the full message bandwidth low pass filter B used in the previous section above.

The Weaver demodulator operates in several steps, shown below in complex notation for the USB case. We assume the message occupies a bandwidth of 0–3000 Hz.

$$s(t) \rightarrow \otimes \Rightarrow LPF(f_1) \Rightarrow \otimes \Rightarrow m(t)$$

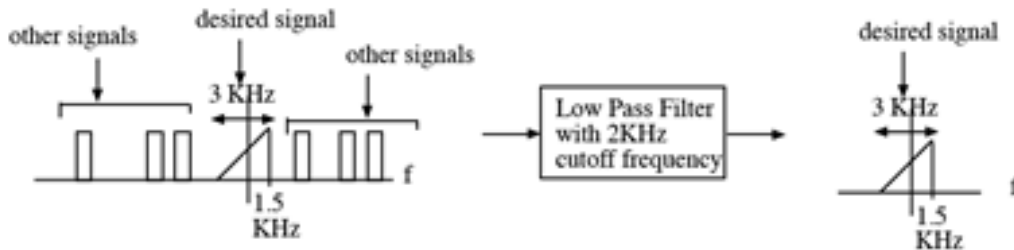
$$\begin{array}{ccc} \uparrow \tilde{s}_1(t) & \tilde{s}_2(t) & \uparrow \tilde{s}_3(t) \\ e^{-j2\pi(f_0+f_1)t} & & e^{j2\pi f_1 t} \end{array}$$

1. downconvert the real USB signal $s(t) = m(t) \cos 2\pi f_c t - m(t) \sin 2\pi f_c t$ to $\tilde{s}_1(t)$ using a complex local oscillator $e^{j2\pi(f_c+f_1)t}$ with a frequency offset $f_1 = 1.5$ kHz relative to $f_0 = f_c = 53$ kHz as shown in the figure below. We choose $f_1 = B/2$ to be a frequency in the approximate middle of the message bandwidth, e.g. 1500 Hz for a 0–3000 Hz voice signal.



2.

3. Low pass filter the result to eliminate adjacent undesired signals. The resulting signal $\tilde{s}_2(t)$ is a complex signal with asymmetric spectrum in the frequency range $-B/2$ to $B/2$ or -1.5 – 1.5 kHz.



4.

5. Do a frequency shift of $\tilde{s}_2(t)$ with a complex local oscillator $e^{j2\pi f_1 t}$ where $f_1 = B/2 = 1.5$ kHz as defined above, resulting in $\tilde{s}_3(t)$ in the frequency range 0 to B or 0 – 3000 Hz. $\tilde{s}_3(t)$ is an analytic signal with positive frequencies only.
6. Take the real part of $\tilde{s}_3(t)$ to obtain the real signal $m(t)$ with both positive and negative frequencies that can be listened to or decoded correctly.
7. The net result of the two frequency shifts and LPF in the Weaver demodulator is that the suppressed carrier frequency in $s(t)$ is shifted to zero frequency, the message is recovered and all other unwanted signals are filtered out.

The problem with the interfering signal mentioned in the previous section does not arise for the Weaver demodulator because the LPF bandwidth is $B/2$ instead of B .

The Weaver demodulator also works for Morse code or other digital signals that are processed by ear or other decoder that operates on a real (not complex) signal. In this case, the bandwidth B may be set to an appropriate value (e.g. 50 Hz for Morse code) and the frequency offset f_1 is set to a tone (pitch) that is pleasant to the ear (e.g. 400 Hz).

The operations of a Weaver demodulator can also be shown in real notation as shown in fig. 3.11.

3.4.5 SSB-SC Weaver demodulator operation in complex notation

We consider two cases for the input signal, real and complex

The mathematics below is for a real input signal that arises when the signal is taken directly from antenna and preamp and (perhaps) an analog real downconverter stage (cos oscillator only).

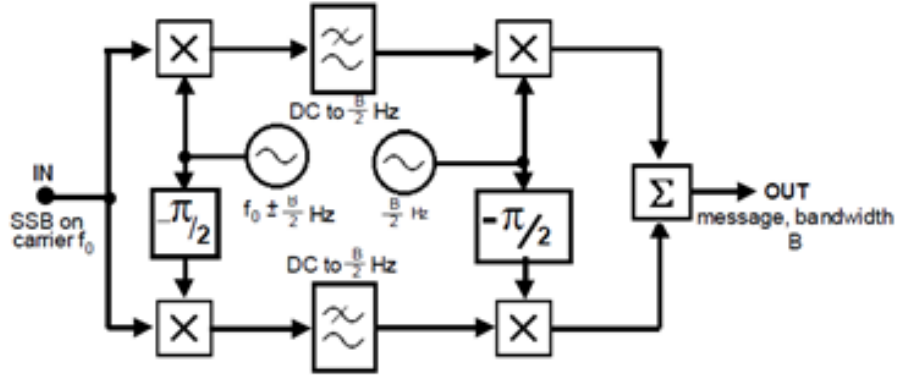


Figure 3.11: Real SSB receiver

At the input we have

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \\ &= m(t)\cos 2\pi f_c t - m(t)\sin 2\pi f_c t \end{aligned}$$

where $\tilde{s}(t) = m(t) + jm(t)$. After the first mixer we have

$$\begin{aligned} \tilde{s}_1(t) &= [m(t)\cos 2\pi f_c t - m(t)\sin 2\pi f_c t]e^{-j2\pi(f_c+f_1)t} \\ &= 0.5[m(t)e^{j2\pi f_c t} + m(t)e^{-j2\pi f_c t} + jm(t)e^{j2\pi f_c t} - jm(t)e^{-j2\pi f_c t}]e^{-j2\pi(f_c+f_1)t} \\ &= 0.5[m(t)e^{-2\pi f_1 t} + m(t)e^{-j4\pi(f_c+f_1)t} + jm(t)e^{-j2\pi f_1 t} - jm(t)e^{-j4\pi(f_c+f_1)t}]. \end{aligned}$$

After the LPF the analytic signal must be shifted

$$\begin{aligned} \tilde{s}_2(t) &= m(t)e^{-2\pi f_1 t} + jm(t)e^{-j2\pi f_1 t} \\ &= [m(t) + jm(t)]e^{-j2\pi f_1 t} \end{aligned}$$

and after the second mixer we have

$$\begin{aligned} \tilde{s}_3(t) &= \tilde{s}_2(t)e^{j2\pi f_1 t} \\ \tilde{s}(t) &= m(t) + jm(t) \\ m(t) &= \text{Re}\{\tilde{s}(t)\} \end{aligned}$$

Exercise: mathematically show low pass filtering $\tilde{s}_1(t)$ to get $\tilde{s}_2(t)$. Hint: write $m(t) = [m_+(t) + m_-(t)]/2$.

Exercise: repeat the figure and mathematics for LSB.

The mathematics below is for a complex input that arises when a complex downconversion (e.g. cos and sin oscillators in an analog mixer with two outputs I and Q) has taken place ahead of the USB receiver. There is a complex input

$$\tilde{s}(t)e^{j2\pi f_c t} = m(t)\cos 2\pi f_c t - m(t)\sin 2\pi f_c t + j[m(t)\cos 2\pi f_c t + m(t)\sin 2\pi f_c t]$$

where $\tilde{s}(t) = m(t) + jm(t)$. After the first mixer we have

$$\begin{aligned} \tilde{s}_1(t) &= \tilde{s}(t)e^{j2\pi f_c t}e^{-j2\pi(f_c+f_1)t} \\ &= \tilde{s}(t)e^{-j2\pi f_1 t}. \end{aligned}$$

After the LPF eliminats other unwanted signals $\tilde{s}_2(t) = \tilde{s}_1(t)$. After the second mixer we have

$$\begin{aligned}\tilde{s}_3(t) &= \tilde{s}_2(t)e^{j2\pi f_1 t} \\ &= \tilde{s}(t)e^{-j2\pi f_1 t}e^{j2\pi f_1 t} \\ &= \tilde{s}(t) \\ &= m(t) + jm(t) \\ m(t) &= \text{Re}\{\tilde{s}(t)\}\end{aligned}$$

3.4.6 SSB-SC Weaver demodulator operation in real notation

The real version of the Weaver demodulator can be built entirely in analog, or entirely in digital. It may also be split, with the first cos/sin mixer in analog (as done in the USRP daughterboard) and the second mixer digital (implemented in GNURadio software).

In fig. 3.11, for USB, the first oscillator is $f_0 + B/2$.

The mathematics for the Weaver demodulator in real notation is done in the example below, assuming that the message is the sum of cos waves at 500 Hz and 2000 Hz. This example should give an appreciation of the value of complex signals to simply both the concepts and the algebra.

Recall

$$\begin{aligned}\cos \alpha \cos \beta &= [\cos(\alpha - \beta) + \cos(\alpha + \beta)]/2 \\ \sin \alpha \cos \beta &= [\sin(\alpha - \beta) + \sin(\alpha + \beta)]/2 \\ \sin \alpha \sin \beta &= [\cos(\alpha - \beta) - \cos(\alpha + \beta)]/2 \\ \cos \alpha \sin \beta &= [-\sin(\alpha - \beta) + \sin(\alpha + \beta)]/2 \\ \cos(\alpha + \beta) &= \cos \alpha \cos \beta - \sin \alpha \sin \beta \\ \sin(\alpha + \beta) &= \sin \alpha \cos \beta + \cos \alpha \sin \beta\end{aligned}$$

The message is

$$\begin{aligned}m(t) &= \cos 2\pi f_2 t + \cos 2\pi f_3 t \\ &= \cos 2\pi 500 t + \cos 2\pi 2000 t\end{aligned}$$

For a real SSB signal (upper sideband)

$$\begin{aligned}s(t) &= m(t) \cos 2\pi f_c t - m(t) \sin 2\pi f_c t \\ &= [\cos 2\pi f_2 t + \cos 2\pi f_3 t] \cos 2\pi f_c t - [\sin 2\pi f_2 t + \sin 2\pi f_3 t] \sin 2\pi f_c t \\ &= \cos 2\pi f_2 t \cos 2\pi f_c t - \sin 2\pi f_2 t \sin 2\pi f_c t + \cos 2\pi f_3 t \cos 2\pi f_c t - \sin 2\pi f_3 t \sin 2\pi f_c t \\ &= \cos 2\pi(f_2 + f_c)t + \cos 2\pi(f_3 + f_c)t\end{aligned}$$

Writing $f_0 = f_c$ yields

$$s(t) = \cos 2\pi(f_2 + f_0)t + \cos 2\pi(f_3 + f_0)t.$$

The first oscillator is at $f_0 + B_2$ for upper sideband. The upper branch after the first mixer and low

pass filter but before the second mixer has $B_2 = B/2 = f_1$ and

$$\begin{aligned}
 s(t) \cos 2\pi(f_0 + B_2)t &= 0.5[\cos 2\pi(f_2 + f_0)t + \cos 2\pi(f_3 + f_0)t] \cos 2\pi(f_0 + B_2)t \\
 &= 0.5[\cos 2\pi(f_2 - B_2)t + \cos 2\pi(f_3 - B_2)t] \\
 &= 0.5[\cos 2\pi(500 - 1500)t + \cos 2\pi(2000 - 1500)t] \\
 &= 0.5[\cos 2\pi(-1000)t + \cos 2\pi(500)t]
 \end{aligned}$$

The upper branch after the second mixer is described by

$$\begin{aligned}
 &0.5[\cos 2\pi(f_2 - B_2)t + \cos 2\pi(f_3 - B_2)t] \cos 2\pi B_2 t \\
 &= 0.25[\cos 2\pi f_2 t + \cos 2\pi(f_2 - 2B_2)t + \cos 2\pi f_3 t + \cos 2\pi(f_3 - 2B_2)t] \\
 &= 0.25[\cos 2\pi(500)t + \cos 2\pi(500 - 3000)t + \cos 2\pi(2000)t + \cos 2\pi(2000 - 3000)t] \\
 &= 0.25[\cos 2\pi(500)t + \cos 2\pi(-2500)t + \cos 2\pi(2000)t + \cos 2\pi(-1000)t]
 \end{aligned}$$

The lower branch after the first mixer and low pass filter but before second mixer is described by

$$\begin{aligned}
 s(t)[- \sin 2\pi(f_0 + B_2)t] &= - [\cos 2\pi(f_2 + f_0)t + \cos 2\pi(f_3 + f_0)t] \sin 2\pi(f_0 + B_2)t \\
 &= - 0.5[\cos 2\pi(f_2 + f_0)t \sin 2\pi(f_0 + B_2)t \\
 &\quad - \cos 2\pi(f_3 + f_0)t \sin 2\pi(f_0 + B_2)t] \\
 &= 0.5[\sin 2\pi(f_2 - B_2)t + \sin 2\pi(f_3 - B_2)t] \\
 &= 0.5[\sin 2\pi(500 - 1500)t + \sin 2\pi(2000 - 1500)t] \\
 &= 0.5[\sin 2\pi(-1000)t + \sin 2\pi(500)t]
 \end{aligned}$$

The lower branch after second mixer is described by

$$\begin{aligned}
 &0.5[\sin 2\pi(f_2 - B_2)t + \sin 2\pi(f_3 - B_2)t][- \sin 2\pi B_2 t] \\
 &= 0.25[- \cos 2\pi(f_2 - 2B_2)t + \cos 2\pi f_2 t - \cos 2\pi(f_3 - 2B_2)t + \cos 2\pi f_3 t] \\
 &= 0.25[- \cos 2\pi(500 - 3000)t + \cos 2\pi(500)t - \cos 2\pi(2000 - 3000)t + \cos 2\pi(2000)t] \\
 &= 0.25[- \cos 2\pi(-2500)t + \cos 2\pi(500)t - \cos 2\pi(-1000)t + \cos 2\pi(2000)t]
 \end{aligned}$$

Then the sum of the second mixer outputs is

$$\begin{aligned}
 0.5[\cos 2\pi f_2 t + \cos 2\pi f_3 t] &= 0.5[\cos 2\pi(500)t + \cos 2\pi(2000)t] \\
 &= 0.5m(t)
 \end{aligned}$$

Chapter 4

Super-heterodyne Receiver

Before a transmission signal is inputted into an IQ receiver, it is often advantageous to apply amplification as the typical amplitude range at this point is on the order of microvolts. A far more practical signal level to work with is one on the order of Volts and such a gain is usually achieved through a series of op amps. Such a large gain requirement around a single carrier frequency can raise practical issues such as the occurrence of feedback due to the high nature of the frequencies needing to be amplified.

A super-heterodyne (or colloquially: superhet) receiver offers a means of overcoming such issues by shifting the carrier (usually down in magnitude) and effectively splitting the gain requirements over multiple frequencies: the original carrier and what is referred to as the intermediate frequency (f_{IF}). This action helps to prevent feedback which may lead to inaccuracies in the demodulated message.

Another advantage of using the superhet design is that it allows the receiver to shift any carrier frequency to an industry standard. This allows the components of the standardized IQ receiver to be mass produced at a very low cost. Amplifiers that support higher frequencies tend to have a narrower bandwidth as well so by shifting the carrier we also remove this limitation. The architecture of the super-heterodyne is shown in fig. 4.1.

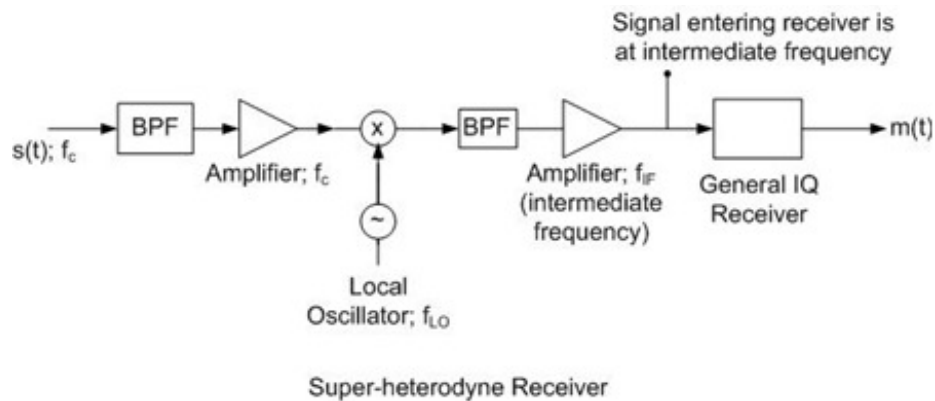


Figure 4.1: Super-heterodyne Receiver

The receiver is tuned to receive different carrier frequencies by changing the frequency of the so-called Local Oscillator (as distinct from the Remote oscillator in the transmitter some distance away from the receiver). The tuning knob on a radio receiver controls the frequency of the Local Oscillator. A common intermediate frequency used for AM receivers is 455 kHz. Once f_{IF} is selected, we may decide how to tune the local oscillator so to produce such a carrier:

$$f_{LO} = f_c \pm f_{IF}$$

With the local oscillator tuned to this frequency we get the following pair: $f_c + (f_c \pm f_{IF}) = 2f_c \pm f_{IF}$ and $f_c - (f_c \pm f_{IF}) = \mp f_{IF}$. We can easily bandpass filter the high frequency term out meaning we have successfully shifted our carrier to a desired intermediate frequency.

An issue with the super-heterodyne receiver which is important to note is the existence of an “image” frequency (f_{IM}). If a signal exists at this frequency and proper filtering is not implemented then it will be shifted to the intermediate frequency, along with the signal at the desired f_c , and will cause direct interference.

If $f_{LO} = f_c \pm f_{IF}$ then $f_{IM} = f_{LO} \pm f_{IF}$. It is often best to consider specific numbers rather than formulas with plus/minus signs. For example, consider $f_{RF} = f_c = 1070$ kHz and $f_{IF} = 455$ kHz.

Then we can choose $f_{LO} = 1070 - 455 = 615$ kHz, so that the incoming signal at 1070 kHz mixes with the LO at 615 kHz to yield an output at the 455 kHz IF. Thus by setting the LO to 615 kHz we have “tuned in” the signal at 1070 kHz.

The image frequency will be $615 - 455 = 160$ kHz, since an incoming (image) signal at 160 kHz can also mix with the same LO at 615 kHz to yield an output at the 455 kHz IF.

We can also choose $f_{LO} = 1070 + 455 = 1525$ kHz, so that the incoming signal at 1070 kHz mixes with the LO at 1525 kHz to yield an output at the 455 kHz IF. Thus by setting the LO to 1525 kHz we have “tuned in” the signal at 1070 kHz.

In this case, the image frequency will be $1525 + 455 = 1980$ kHz, since an incoming (image) signal at 1980 kHz can also mix with the same LO at 1525 kHz to yield an output at the 455 kHz IF.

The superheterodyne principle can also be applied to complex signals. For example, in the USRP, the daughterboard is a complex local oscillator that downconverts a slice of radio frequencies centered around f_{LO} to a zero frequency IF complex baseband signal (with asymmetrical spectrum). This signal is sampled by the main USRP board.

Chapter 5

Frequency Modulation

5.1 Overview

In this chapter, we describe angle modulation, which includes both frequency and phase modulation. In both cases, amplitude is kept constant.

The phase angle of the carrier wave varies with the message signal.

For phase modulation, the phase angle is linearly related to the message

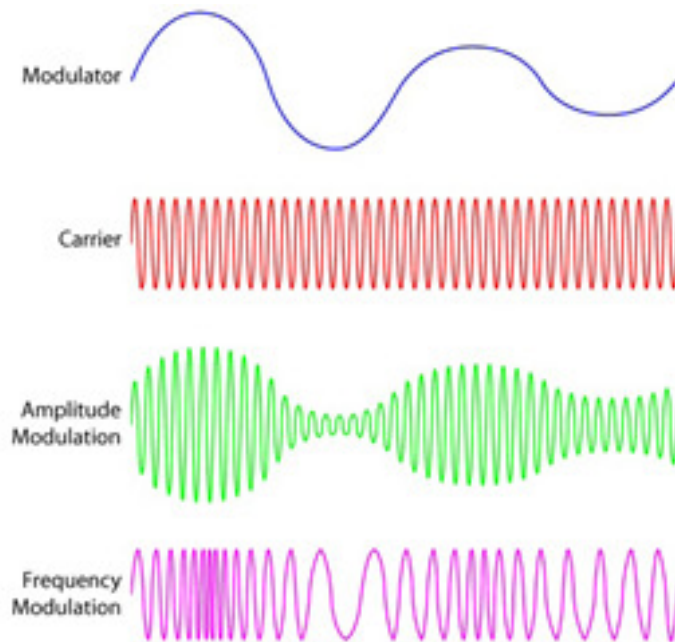
For frequency modulation (FM), the derivative of the phase angle is linearly related to the message.

FM and PM signals have a constant amplitude or *constant envelope*, thus enabling the use of power-efficient nonlinear power amplifiers.

FM signals are more resistant to noise than AM, but at the cost of larger bandwidth.

FM is widely used for radio broadcasting, analog TV audio and public safety two-way radios.

The figure below compares AM and FM



5.2 FM with General Message

To derive the equation for an FM wave, we have 3 starting points:

1. Recall that idea of AM is that the instantaneous amplitude $a(t)$ of a carrier wave is varied linearly with the baseband message signal $m(t)$ around a constant bias value A_c so that $a(t) = A_c + A_c k_a m(t)$
1. By analogy, the idea of frequency modulation is that the instantaneous frequency $f_i(t)$ of a carrier wave is varied linearly with the baseband message signal $m(t)$ around a constant value f_c . Thus for a general message $m(t)$ the idea of FM is:

$$f_i(t) = f_c + k_f m(t)$$

where the constant k_f represents the frequency sensitivity of the modulator, expressed in hertz per volt.

1. Recall that a general signal is written:

$$s(t) = a(t) \cos \theta(t) = a(t) \cos [2\pi f_c t + \phi(t)]$$

where $\theta(t) = 2\pi f_c t + \phi(t)$ has a linear variation at rate f_c and a time varying part $\phi(t)$

1. Recall the instantaneous frequency of a signal is:

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} = f_c + \frac{1}{2\pi} \frac{d\phi(t)}{dt}$$

Combining these 3 starting points, we can write:

$$\begin{aligned}\theta(t) &= 2\pi \int_0^t f_i(\alpha) d\alpha = 2\pi \int_0^t [f_c + k_f m(\alpha)] d\alpha \\ &= 2\pi f_c t + k_f \int_0^t m(\alpha) d\alpha \\ s(t) &= A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right]\end{aligned}$$

The FM signal $s(t)$ can be written in standard IQ format:

$$\begin{aligned}s(t) &= i(t) \cos 2\pi f_c t - q(t) \sin 2\pi f_c t \\ s(t) &= \operatorname{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\} = \\ &= \operatorname{Re}\{[I(t) + jQ(t)][\cos 2\pi f_c t + j\sin 2\pi f_c t]\} \\ &= I(t) \cos 2\pi f_c t - Q(t) \sin 2\pi f_c t \\ &= a(t) \cos[2\pi f_c t + \phi(t)]\end{aligned}$$

Thus for FM:

$$\begin{aligned}a(t) &= A_c \\ \phi(t) &= 2\pi k_f \int_0^t m(\alpha) d\alpha \\ I(t) &= A_c \cos 2\pi k_f \int_0^t m(\alpha) d\alpha \\ Q(t) &= A_c \sin 2\pi k_f \int_0^t m(\alpha) d\alpha\end{aligned}$$

5.3 FM with Sinusoidal Message

Now consider a sinusoidal modulating wave defined by

$$m(t) = A_m \cos(2\pi f_m t)$$

We re-derive the equation for the FM wave with the same 3 starting points.

1. The idea of FM is that the instantaneous frequency of the resulting FM wave equals:

$$\begin{aligned}f_i(t) &= f_c + k_f A_m \cos(2\pi f_m t) \\ &= f_c + \Delta f \cos(2\pi f_m t)\end{aligned}$$

where $\Delta f = k_f A_m$.

Thus the message causes the instantaneous frequency to vary above and below the carrier frequency f_c , from $f_c - \Delta f$ to $f_c + \Delta f$. The quantity Δf is called the frequency deviation, since the instantaneous frequency deviates from the carrier by that amount.

1. A general signal:

$$s(t) = a(t) \cos \theta(t)$$

1. The instantaneous frequency of a signal is:

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt}$$

Combining these 3 starting points, we can write:

$$\begin{aligned}\theta(t) &= 2\pi f_c t + \frac{\Delta f}{f_m} \sin(2\pi f_m t) \\ &= 2\pi f_c t + \beta \sin(2\pi f_m t)\end{aligned}$$

Where $\beta = \frac{k_f A_m}{f_m} = \frac{\Delta f}{f_m}$ is called the modulation index of the FM wave.

Thus the FM wave itself is given In terms of β by:

$$s(t) = A_c \cos [2\pi f_c t + \beta \sin(2\pi f_m t)]$$

If β is small we have narrowband FM (NBFM) and if β is large (compared to one radian) we have wideband FM.

Thus for a sinusoidal modulating wave $m(t) = A_m \cos(2\pi f_m t)$ the FM wave can be written:

$$s(t) = I(t) \cos 2\pi f_c t - Q(t) \sin 2\pi f_c t = a(t) \cos[2\pi f_c t + \phi(t)]$$

where:

$$\begin{aligned}a(t) &= A_c \\ \phi(t) &= \beta \sin 2\pi f_m t \\ i(t) &= A_c \cos \beta \sin 2\pi f_m t \\ q(t) &= A_c \sin \beta \sin 2\pi f_m t\end{aligned}$$

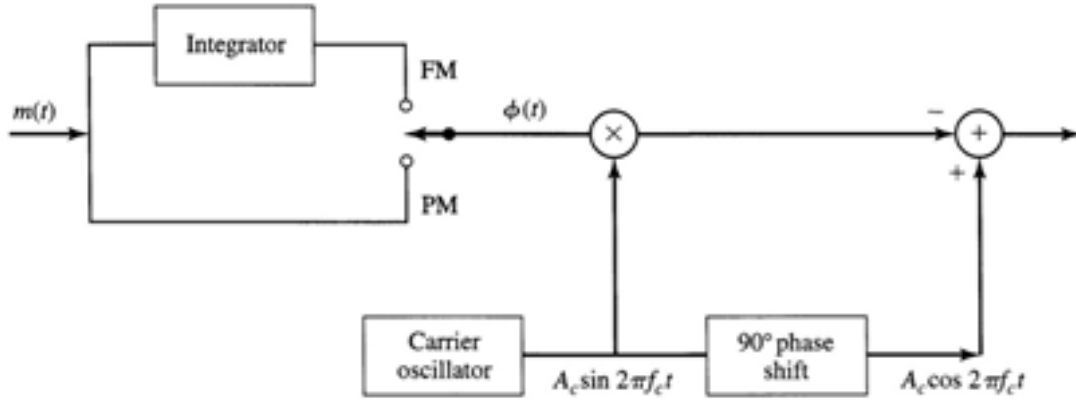
5.3.1 Narrowband FM with Sinusoidal Message

When β is small compared to one radian, i.e. $\beta \ll 1$, the FM wave may be approximated using $\cos x \approx 1, \sin x \approx x$ for $x \ll 1$ to obtain

$$s(t) \simeq A_c \cos(2\pi f_c t) - \beta A_c \sin(2\pi f_c t) \sin(2\pi f_m t)$$

Thus for $\beta \ll 1$

$$\begin{aligned}i(t) &= A_c \\ q(t) &= \beta A_c \sin 2\pi f_m t\end{aligned}$$



Narrow-

band Frequency and Phase Modulator

Exercise: find the complex envelope $\tilde{s}(t)$ for a NBFM signal. Recall that in general

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi(t)} \\ s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \\ a(t) &= \sqrt{i^2(t) + q^2(t)} \\ \phi(t) &= \arctan \frac{q(t)}{i(t)} \end{aligned}$$

Compare the NBFM waveform equation with the AM equation, note the similarity and differences.

$$s(t) = A_c \cos 2\pi f_c t + A_c \mu \cos 2\pi f_m t \cos 2\pi f_c t$$

Observe that NBFM signal wave requires essentially the same transmission bandwidth (i.e., $2f_m$) as an AM wave.

5.4 Power Spectrum of an FM Signal – Bessel functions

Consider the FM signal with a single tone message $m(t) = A_m \cos(2\pi f_m t)$ so that

$$\begin{aligned} s(t) &= A_c \cos[2\pi f_c t + \beta \sin(2\pi f_m t)] \\ &= \text{Re}\{A_c e^{j(2\pi f_c t + \beta \sin(2\pi f_m t))}\} = \text{Re}\{A_c e^{j\beta \sin 2\pi f_m t} e^{j2\pi f_c t}\} \end{aligned}$$

In what follows, we will evaluate the FM complex envelope and find that it is given in term of a special function called a Bessel function. Once we have defined this function, we continue with finding the FM power spectrum and bandwidth.

The FM complex envelope $a(t)e^{j\phi(t)} = A_c e^{j\beta \sin 2\pi f_m t}$ is periodic with period $1/f_m$ so we can write the complex envelope as a complex Fourier series with index n

$$A_c e^{j\beta \sin 2\pi f_m t} = A_c \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n f_m t}$$

The Fourier coefficients are given by the integral

$$c_n = f_m \int_{t=0}^{1/f_m} e^{j\beta \sin 2\pi f_m t} e^{-j2\pi n f_m t} dt \triangleq J_n(\beta)$$

This integral depends on two parameters n and β , but cannot be evaluated in terms of elementary functions, so it is given a name $J_n(\beta)$ called the Bessel function of the first kind and order n .

By a change of variable $u = 2\pi f_m t$ we can write

$$J_n(\beta) = \frac{1}{2\pi} \int_{u=0}^{2\pi} e^{j(\beta \sin u - nu)} du$$

The function $J_n(\beta)$, the Bessel function of the first kind of order n , is plotted in the figure below for values of n from 0 to 4.



$J_n(\beta)$ looks like a damped cosine wave for $n = 0$ and damped sine waves for $n \geq 1$

The Bessel function has many properties and identities (just like cos and sin)

$$J_n \beta = (-1)^n J_{-n}(\beta)$$

$$\sum_{n=-\infty}^{\infty} J_n^2(\beta) = 1$$

Also, for small values of β , we have:

$$J_0(\beta) \simeq 1, \quad J_1(\beta) \simeq \frac{\beta}{2}$$

and:

$$J_n(\beta) \simeq 0, \quad n > 1$$

Now that we have introduced the Bessel function, we continue to find the FM power spectrum and bandwidth.

The FM complex envelope can be written

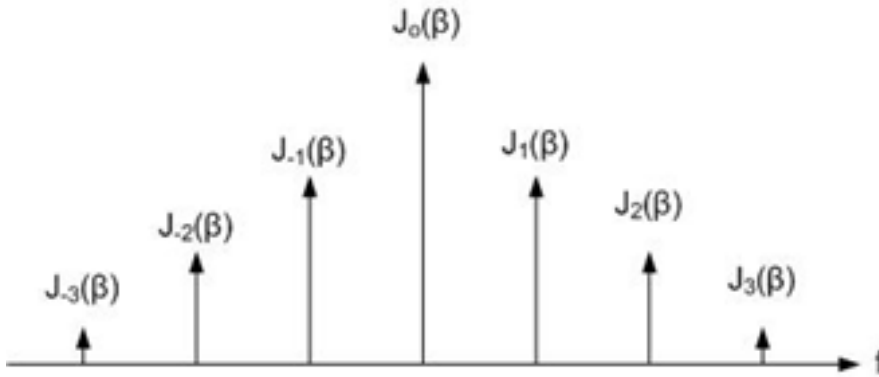
$$a(t)e^{j\phi(t)} = A_c e^{j\beta \sin 2\pi f_m t} = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi n f_m t}$$

Thus the FM signal is written

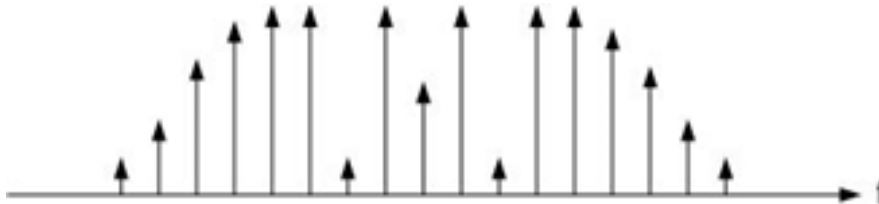
$$\begin{aligned} s(t) &= \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\} = \text{Re}\{A_c e^{j\beta \sin 2\pi f_m t} e^{j2\pi f_c t}\} \\ &= \text{Re}\{A_c [\sum_{n=-\infty}^{\infty} J_n(\beta) e^{j2\pi n f_m t}] e^{j2\pi f_c t}\} \\ &= A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos [2\pi(f_c + n f_m)t] \end{aligned}$$

The discrete spectrum of the FM wave is obtained as

$$S(f) = \frac{A_c}{2} \sum_{n=-\infty}^{\infty} J_n(\beta) [\delta(f - f_c - n f_m) + \delta(f + f_c + n f_m)]$$



Bessel Function value $[J_n(\beta)]$ defining the FM spectrum depends upon β



Example of FM spectrum with relatively large β

As for the case of AM, we may ask why it is that for a sinusoidal message we see discrete sidebands at $f_c \pm n f_m$, since the instantaneous frequency given by

$f_i(t) = f_c + \Delta f \cos(2\pi f_m t) = f_c + \beta f_m \cos(2\pi f_m t)$ can be any continuous value and is not restricted to a discrete set of values.

If we were to view the frequency domain representation of the FM signal and the observation (time) interval is far less than the message period $1/f_m$, the spectrum consists of one impulse or “spike” at $f_i(t)$ that moves around in frequency in a continuous fashion as time progresses. If the observation interval is greater than $1/f_m$, then the spectrum is discrete with sidebands spaced at intervals of f_m . The derivation of the FM spectrum above uses the complex Fourier series and thus assumes an observation interval of one message period.

5.4.1 Observations about FM spectrum

We make the following observations about the FM spectrum as observed over an interval greater than the message period $1/f_m$

1. the spectrum includes components at frequencies $f_c \pm n f_m$ that are integer multiples of the message frequency above and below the carrier frequency.
2. The number of significant peaks increases as β increases.
3. As β increases, more and more peaks become significant, but some peaks become smaller, all in accordance with the Bessel function curve.
4. Thus the shape (‘envelope’) of the spectrum depends on β in a complicated way.
5. It is not obvious from the spectrum plot, but for any value of β , and thus for any spectrum shape, the powers in the all of the peaks added together adds up to 1. This is because the average power of an FM wave developed across a 1 ohm resistor is given by:

$P = \frac{A_c^2}{2} \cdot \sum_{n=-\infty}^{\infty} J_n^2(\beta) = \frac{A_c^2}{2}$ independent of the values of β , since all the Bessel function power spikes in the power spectrum add up to 1.

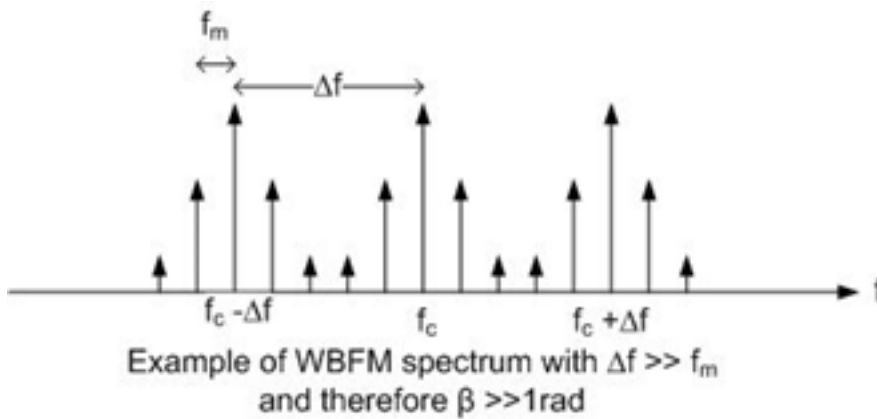
The FM spectrum depends upon the modulation index β as mentioned above ; this means it is a function of Δf and f_m .

5.4.2 Wideband FM large modulation index

Beta is large (WBFM): $\beta = \frac{\Delta f}{f_m} = 10$; $\Delta f = 10f_m$;

$$f_{max} = f_c + \Delta f; f_{min} = f_c - \Delta f$$

Wideband Frequency Modulation: $\Delta f \ll f_m$; $\beta \gg 1$

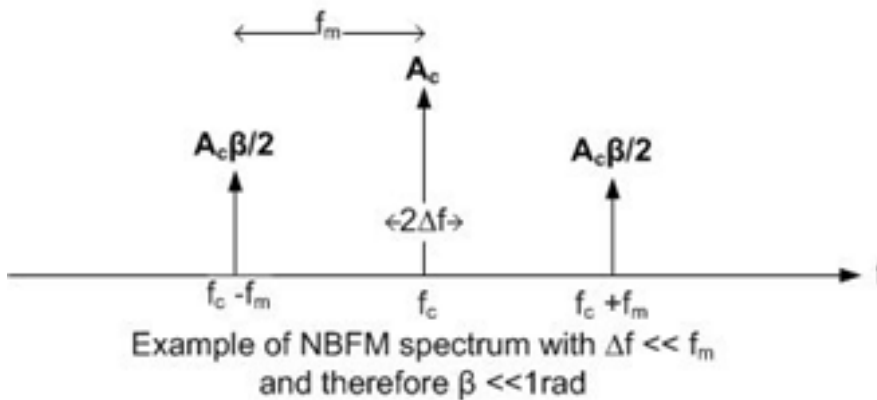


5.4.3 Narrowband FM, small modulation index

Beta is small (NBFM); $\beta = \frac{\Delta f}{f_m} = 0.1$; $\Delta f = 0.1 f_m$

$$f_i(t) = f_c + k_f m(t) = f_c + \Delta f \cos(2\pi f_m t)$$

Narrowband Frequency Modulation: $\Delta f \ll f_m$; $\beta \ll 1$



5.4.4

Effective bandwidth of FM – Carson’s rule

For practical purposes, the bandwidth of the FM wave corresponds to the bandwidth containing 98% of the signal power.

The effective bandwidth of the FM signal is approximately given by Carson's formula:

$$B = 2(1 + \beta)f_m = 2\Delta f + 2f_m = 2(\Delta f + f_m) = 2(k_f A_m + f_m)$$

Carson's formula is intuitively reasonable: the bandwidth must be at least twice the frequency deviation since the instantaneous frequency changes from $f_c - \Delta f$ to $f_c + \Delta f$.

The bandwidth must also be twice the modulating frequency.

For a general message $m(t) = \sum_i A_i(t) \cos(2\pi f_i t + \psi_i(t))$ containing many cos waves, we can estimate the FM bandwidth by considering the highest frequency (i.e. the message bandwidth W) and highest amplitude A in the message. For this case, we replace $f_m \rightarrow W$ and $A_m \rightarrow A$ and find

$$B = 2k_f A + 2W = 2(1 + D)W$$

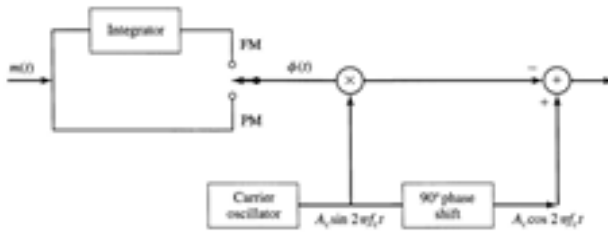
$$DW = k_f A$$

Thus the deviation ratio $D = \frac{k_f A}{W}$ has the same role as the modulation index $\beta = \frac{k_f A_m}{f_m}$ in determining the bandwidth of an FM signal.

5.5 Frequency modulators (transmitters)

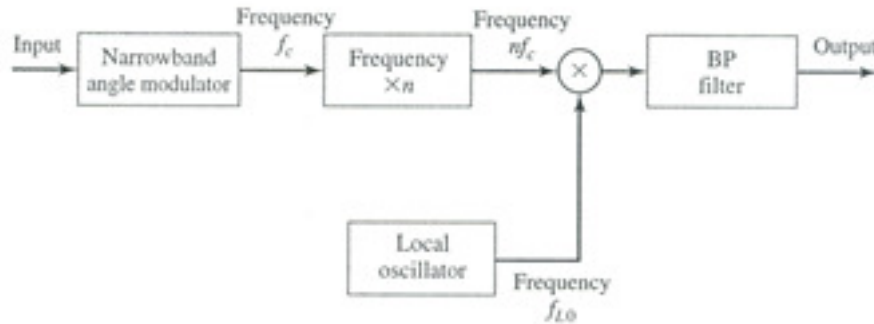
A common analog method is so-called indirect FM modulation.

The process starts with a narrowband modulator as shown above, and repeated below, that implements $s(t) \simeq A_c \cos(2\pi f_c t) - \beta A_c \sin(2\pi f_c t) \sin(2\pi f_m t)$



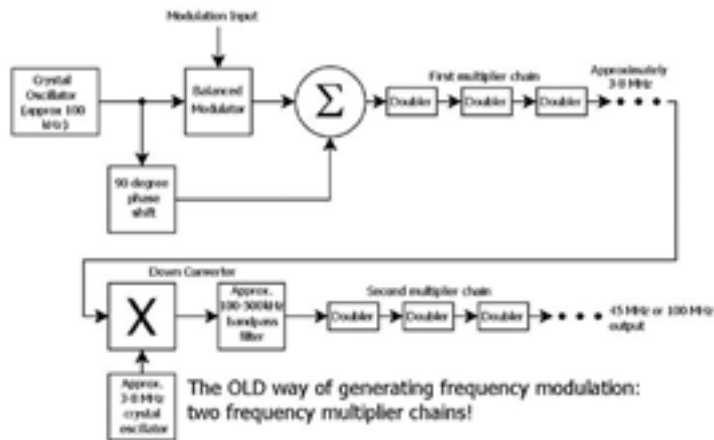
Narrowband angle modulator

The next step is to multiply the signal by itself several times, thus increasing both the carrier frequency and the modulation index, as shown in the figure and mathematics below



$$\begin{aligned}
 s(t) &= A_c \cos(2\pi f_c t + \beta \sin 2\pi f_m t) \\
 s^2(t) &= A_c^2 \cos^2(2\pi f_c t + \beta \sin 2\pi f_m t) \\
 &= \text{DC} + A_c^2 \cos(4\pi f_c t + 2\beta \sin 2\pi f_m t) \\
 s^4(t) &= \text{DC} + A_c^4 \cos(8\pi f_c t + 4\beta \sin 2\pi f_m t)
 \end{aligned}$$

We observe that both the deviation and modulation index increase when the signal is multiplied by itself. In practice the multiplication may be many times, as illustrated in the example below. A downconversion stage may be required if the desired carrier frequency is low and the desired deviation (β) is large.



In this figure, the message waveform is integrated and then phase modulated in accordance with the FM equation

$$s(t) = A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right]$$

A digital method of building an FM transmitter simply implements the FM equation in software.

5.6 Frequency Demodulation

Frequency demodulation extracts the original message wave from the frequency-modulated wave. We describe two basic devices, the analog-like frequency discriminator and a digital FM demodulator.

How to extract the message from an FM signal? We consider four intuitive approaches.

5.6.1 Intuition 1: Differentiation

We know that from the idea of FM $f_i(t) = f_c + k_f m(t) = f_c + \frac{1}{2\pi} \frac{d\phi(t)}{dt}$ so that $m(t) = \frac{1}{2\pi k_f} \frac{d\phi(t)}{dt}$

Since $s(t) = a(t)\cos[2\pi f_c t + \phi(t)]$ with $a(t) = A_c$ (constant) we see that if we differentiate $s(t)$ we will get a term in $\frac{ds(t)}{dt}$ that looks like $\frac{d\phi(t)}{dt}$

A demodulator based on this intuition is in the figure below.

$$s(t) \rightarrow \text{LIMITER} \rightarrow \text{BPF} \rightarrow \text{DIFFERENTIATOR} \rightarrow \\ \text{ENVELOPE DETECTOR} \rightarrow \text{DC BLOCK} \rightarrow m(t)$$

We differentiate $s(t)$ as a first step to extract the message $m(t)$. Assuming A_c is constant we have:

$$s(t) = A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right] \\ \frac{ds(t)}{dt} = -A_c [2\pi f_c + 2\pi k_f m(t)] \sin \left[2\pi f_c t + 2\pi k_f \int_0^t m(t) dt \right]$$

This expression is in the form of an AM signal

$$2\pi f_c A_c [1 + k_a m(t)] \cos[2\pi f_c t + \phi(t)] = a(t) \cos[2\pi f_c t + \phi(t)]$$

with envelope $2\pi A_c f_c \left[1 + \frac{k_f}{f_c} m(t) \right]$ with $k_a = k_f/f_c$ and phase $2\pi k_f \int_0^t m(\alpha) d\alpha - \pi/2$

The resulting AM signal can be demodulated by an envelope detector to obtain DC plus the message. The envelope detector ignores the phase.

Note that if the FM signal has $a(t) \neq A_c$ is not constant (e.g. due to channel fading or noise), then differentiation will not work, since for this case

$$\frac{ds(t)}{dt} = -a(t) [2\pi f_c + 2\pi k_f m(t)] \sin \left[2\pi f_c t + 2\pi k_f \int_0^t m(t) dt \right] \\ + \frac{da(t)}{dt} \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right]$$

To make $a(t) = A_c$ is constant, we can use a limiter (e.g. with back to back diodes) ahead of the discriminator. The limiter limits (clips) the input signal so that it is of constant amplitude. The

output of the limiter looks like a square wave with changing frequency, and will contain harmonics at odd multiples of f_c . The limiter must be followed by a bandpass filter at f_c to restore the signal $s(t) = A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right]$

A block diagram of this FM demodulator is

$$s(t) \rightarrow \text{LIMITER} \rightarrow \text{BPF} \rightarrow \text{DISCRIMINATOR} \rightarrow \\ \text{ENVELOPE DETECTOR} \rightarrow \text{DC BLOCK} \rightarrow m(t)$$

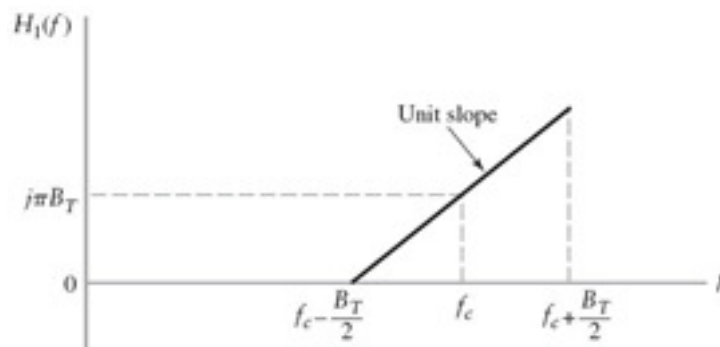
5.6.2 Intuition 2: Linear Amplitude vs. Frequency Characteristic

We know that from the idea of FM $f_i(t) = f_c + k_f m(t)$. If we have a circuit that has output amplitude that increases linearly with frequency, then the circuit output amplitude will vary in step with the message.

One such circuit is the ideal differentiator, with a transfer function given by $H(f) = j2\pi f$.

Proof: If $s(t) \rightarrow S(f)$ then $ds(t)/dt \rightarrow j2\pi f S(f)$

The transfer function acts as a frequency to voltage converter, and is illustrated below, centered at the carrier frequency



The slope is such that the transfer function changes by $j2\pi B_T$ over a bandwidth B_T centered at f_c

The action of an ideal differentiator (Figure 3(a)) can be approximated by any device whose magnitude transfer function is reasonably linear, within the range of frequencies of interest.

In Figure 3(b) an RL circuit approximation to a differentiator is used followed by an envelope detector. The RL circuit is a *high pass filter*.

A bandpass version of this circuit is shown in Figure 3(c). These discriminators are known as slope detectors. A more linear response can be obtained by taking the difference between two

bandpass magnitude responses, as is done by the balanced discriminator shown in Figure 3(d).

In all cases (a)-(d), a limiter and bandpass filter is required ahead of the discriminator

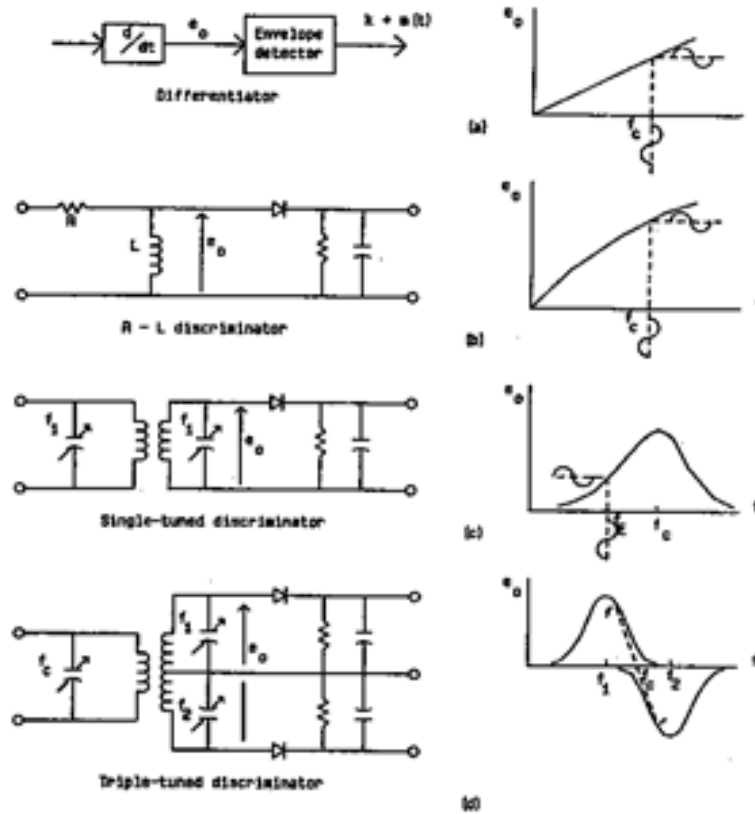
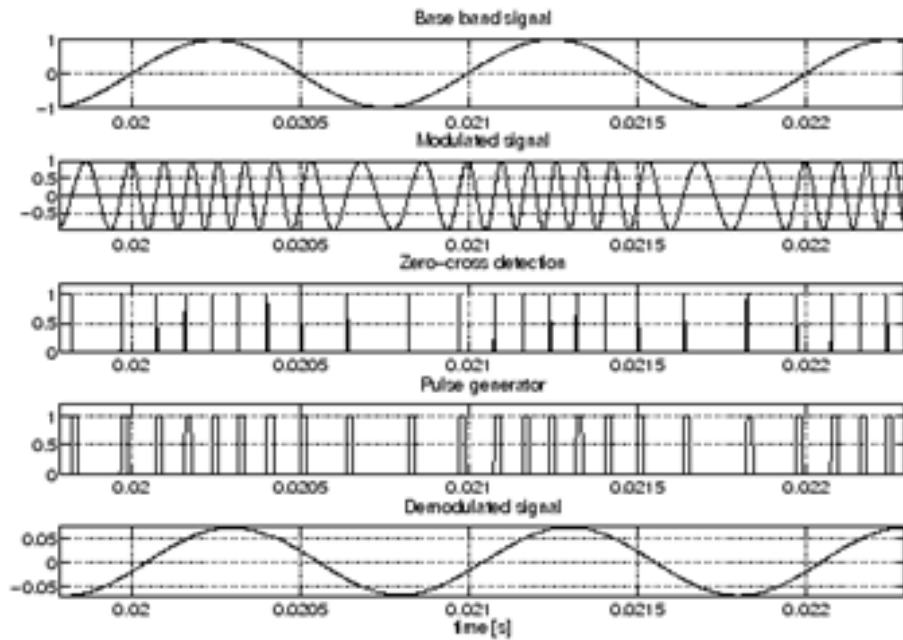


Figure 3 FM Detectors

5.6.3 Intuition 3: Zero Crossing Counter

The message information is contained in the time (location) of the zero crossings, and the amplitude can be ignored, as shown in the figure below.



5.6.4 Intuition 4: Phase Locked Loop

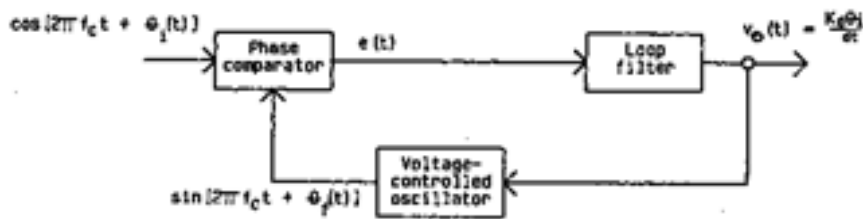


Figure 4 Phase-Locked Loop

This circuit uses ideas from the Costas loop receiver used for DSB. The circuit comprises a phase comparator (multiplier), low pass (loop) filter and voltage controlled oscillator (equivalent to an FM modulator).

With signal input $s(t) = \cos[2\pi f_c t + \theta_i(t)]$ and FM modulator (VCO) output $\sin[2\pi f_c t + \theta_f(t)]$, the phase comparator (multiplier) output signal $e(t) = k_c [\theta_i(t) - \theta_f(t)]$ plus double frequency terms. With high gain in the loop filter, $e(t) \approx 0$ and the VCO output frequency is the same as the input signal frequency. Thus the VCO input must be the same as the message. The loop filter output voltage is proportional to the instantaneous frequency of the input, and FM demodulation is achieved.

5.7 Digital FM Demodulator

A digital FM demodulator starts with the I and Q outputs of a general IQ receiver. Recall for an FM signal:

$$s(t) = A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right]$$

$$I(t) = A_c \cos 2\pi k_f \int_0^t m(\alpha) d\alpha$$

$$Q(t) = A_c \sin 2\pi k_f \int_0^t m(\alpha) d\alpha$$

To extract $m(t)$ from $I(t), Q(t)$ we show two methods that can be implemented in software.

5.7.1 $I(t), Q(t)$ as Real Signals

Formula: $m(t) = \frac{d}{dt} \arctan \left\{ \frac{Q(t)}{I(t)} \right\}$

Block diagram: $I(t), Q(t) \rightarrow \text{DIVIDE} \rightarrow \text{ARCTAN} \rightarrow d/dt \rightarrow m(t)$

Proof:

$$\begin{aligned} \frac{d}{dt} \arctan \left\{ \frac{Q(t)}{I(t)} \right\} &= \frac{d}{dt} \arctan \frac{A_c \sin 2\pi k_f \int_0^t m(\alpha) d\alpha}{A_c \cos 2\pi k_f \int_0^t m(\alpha) d\alpha} \\ &= \frac{d}{dt} \arctan \left\{ \tan 2\pi k_f \int_0^t m(\alpha) d\alpha \right\} = \\ &= \frac{d}{dt} 2\pi k_f \int_0^t m(\alpha) d\alpha = 2\pi k_f m(t) \end{aligned}$$

This method is not good in practice because when $I(t)$ is small, the division by a small number will cause numerical problems.

5.7.2 $I(t) + jQ(t)$ as a Complex Signal

$$s(t) = \text{Re} \{ a(t) e^{j\phi(t)} e^{j2\pi f_c t} \} = \text{Re} \{ [I(t) + jQ(t)] e^{j2\pi f_c t} \} = \text{Re} \{ \tilde{s}(t) e^{j2\pi f_c t} \}$$

Formula:

$$m(t) = \arg[\tilde{s}(t-1)\tilde{s}^*(t)]$$

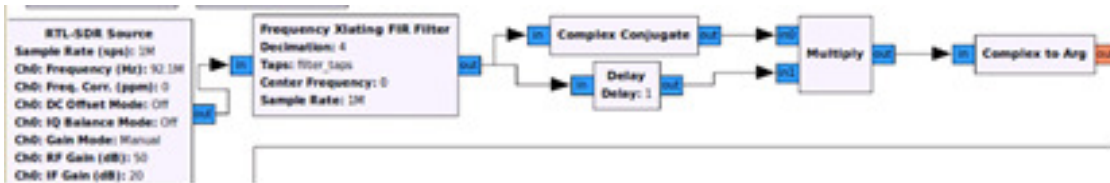
Where:

$$(t-1) \rightarrow z^{-1}$$

represents one sample delay

Block diagram:

$$\tilde{s}(t) \rightarrow \tilde{s}(t-1), \tilde{s}^*(t) \rightarrow \text{MULTIPLY} \rightarrow \text{ARG}$$



Proof:

$$\begin{aligned} \arg[\tilde{s}(t-1)\tilde{s}^*(t)] &= \arg[a(t-1)e^{j\phi(t-1)}a(t)e^{-j\phi(t)}] \\ &= \phi(t-1) - \phi(t) \approx \frac{d\phi}{dt} = 2\pi k_f m(t) \end{aligned}$$

This method of FM demodulation is commonly used in software defined radios.

Alternate Formula:

$$m(t) = \arg[\tilde{s}(t)\tilde{s}^*(t-1)]$$

Proof:

$$\begin{aligned} \arg[\tilde{s}(t)\tilde{s}^*(t-1)] &= \arg[a(t)e^{j\phi(t)}a(t-1)e^{-j\phi(t-1)}] \\ &= \phi(t) - \phi(t-1) \approx \frac{d\phi}{dt} = 2\pi k_f m(t) \end{aligned}$$

5.7.3 Demodulator with frequency offset

In the event of a frequency offset, the complex baseband signal will include a rotating exponential representing the frequency offset

$$\tilde{s}(t) = a(t)e^{j\phi(t)}e^{j(2\pi f_b t + \psi)}$$

In this case,

$$\begin{aligned} \arg[\tilde{s}(t)\tilde{s}^*(t-1)] &= \arg[a(t)e^{j\phi(t)}e^{j(2\pi f_b t + \psi)}a(t-1)e^{-j\phi(t-1)}e^{-j(2\pi f_b(t-1) + \psi)}] \\ &= \phi(t) - \phi(t-1) + 2\pi f_b t - 2\pi f_b(t-1) \approx \frac{d\phi}{dt} - 2\pi f_b = 2\pi(k_f m(t) - f_b) \end{aligned}$$

Thus the frequency offset results in a DC offset added to (or subtracted from) the message. If the frequency offset is large, then the DC offset may be larger than the peak value of the message. Since the message is represented digitally, the DC offset may approach or exceed the maximum (or minimum) value that can be represented in which case the message may be clipped.

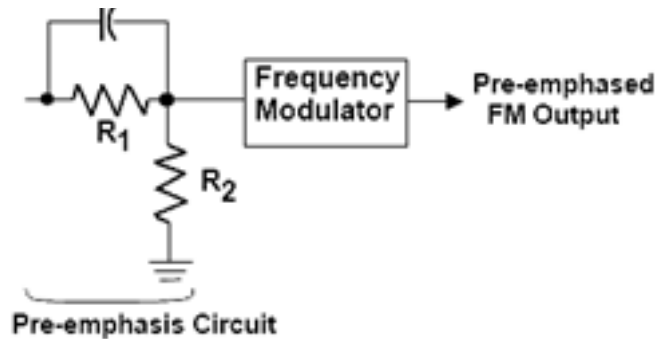
5.8 Pre-emphasis and de-emphasis in FM

5.8.1 Pre- and de-emphasis circuits

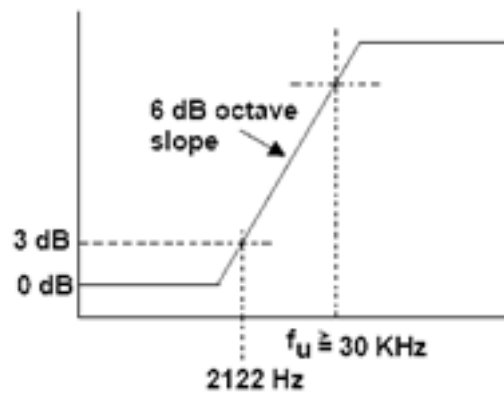
All versions of the FM demodulator are essentially a differentiator or high pass filter whose output will increase linearly with the input frequency. Thus for a message consisting of two cos waves at f_1 and $f_2 > f_1$ of equal amplitude, then at the FM demodulator output, the wave at f_2 will have a larger amplitude than the wave at f_1 .

For $f_2 = 2f_1$, the amplitude of f_2 will be twice the amplitude at f_1 , and the power will be 4 times higher (6 dB). The noise (with components at all frequencies) will also have a high pass characteristic with larger amplitude at higher frequencies.

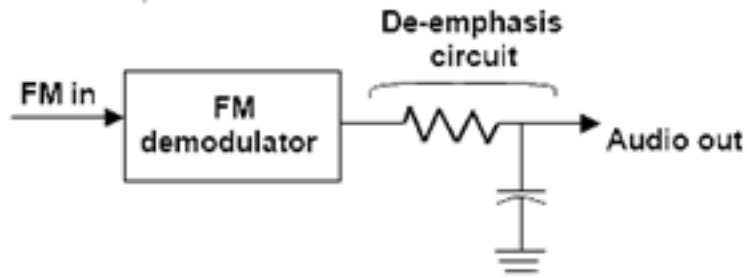
To compensate for this effect, the transmitted message signal is filtered with a so-called pre-emphasis filter to also increase the amplitude of the message at high frequencies. Thus both the message signal and the noise will have increasing amplitude at higher frequencies. The pre-emphasis filter is a single pole filter with 6 dB per octave (or 20 dB per decade) slope.



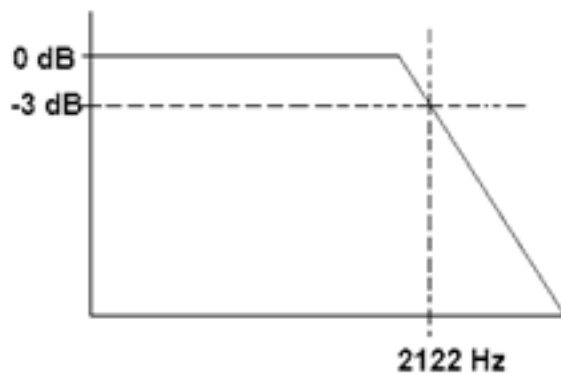
(a) Pre-emphasis Circuit



The FM receiver contains a de-emphasis filter to decrease the amplitude at high frequencies so that the message is reconstructed correctly. For a message with cos waves at f_1 and $f_2 > f_1$ of equal amplitude, then at the FM demodulator output after the de-emphasis filter, the wave at f_2 will have the same amplitude as the wave at f_1 .



(c) De-emphasis circuit



(d) De-emphasis Curve

The de-emphasis filter can be implemented with an analog RC circuit as shown in the figure.

For Broadcast FM in North America, the RC time constant is set to 75 usec, corresponding to a cutoff frequency of $\frac{1}{2\pi RC} = \frac{1}{2\pi 75 \cdot 10^{-6}} = 2122$ Hz.

The impulse response of the RC circuit is $h(t) = e^{-t/RC}$ so that the signal decays to $1/e = 36.8$ % of its initial value in time $t = RC$

5.8.2 De-emphasis with digital filter

This filter can be implemented digitally with a single pole IIR filter with difference equation

$$y[n] - (1 - \alpha)y[n - 1] = \alpha x[n] \text{ or } y_n = (1 - \alpha)y_{n-1} + \alpha x_n$$

with transfer function derived as follows:

$$Y(z) - z^{-1}(1 - \alpha)Y(z) = \alpha X(z)$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\alpha}{1 - (1 - \alpha)z^{-1}} = \frac{\alpha z}{z - (1 - \alpha)}$$

This filter has a single pole at $1 - \alpha$ and a single zero at 0.

The impulse response may be found by inverse z-transform or iteration to find

$$h[n] = \alpha(1 - \alpha)^n u[n]$$

This filter is implemented in GNURadio GRC as a single pole IIR filter block with parameter α , see

http://gnuradio.org/doc/sphinx-3.7.0/filter/filter_blk.html#gnuradio.filter.single_pole_iir_filter_cc

The Matlab representation is $y = \text{filter}(b, a, x)$ for a general filter, where we define the vectors $a = [1 \ a_1 \ a_2]$, $b = [1 \ b_1 \ b_2]$, see www.mathworks.com/help/matlab/ref/filter.html

for the filter with difference equation

$$y[n] + a_1 y[n - 1] + a_2 y[n - 2] = b_0 x[n] + b_1 x[n - 1] + b_2 x[n - 2]$$

and transfer function

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

For the filter considered here $y[n] - (1 - \alpha)y[n - 1] = \alpha x[n]$

Thus $a = [a_1 \ a_2] = [1 \ -(1 - \alpha)]$ and $b = [b_0] = [\alpha]$.

The filter characteristics (including frequency response, phase response, phase delay, group delay, pole-zero plot, impulse response) can be viewed using the Matlab command **fvtool(b,a)**, see www.mathworks.com/help/signal/ref/fvtool.html

For example, if $\alpha = 0.1$ then set the values in **fvtool** to be $b = 0.1$ and $a = [1 \ -0.9]$

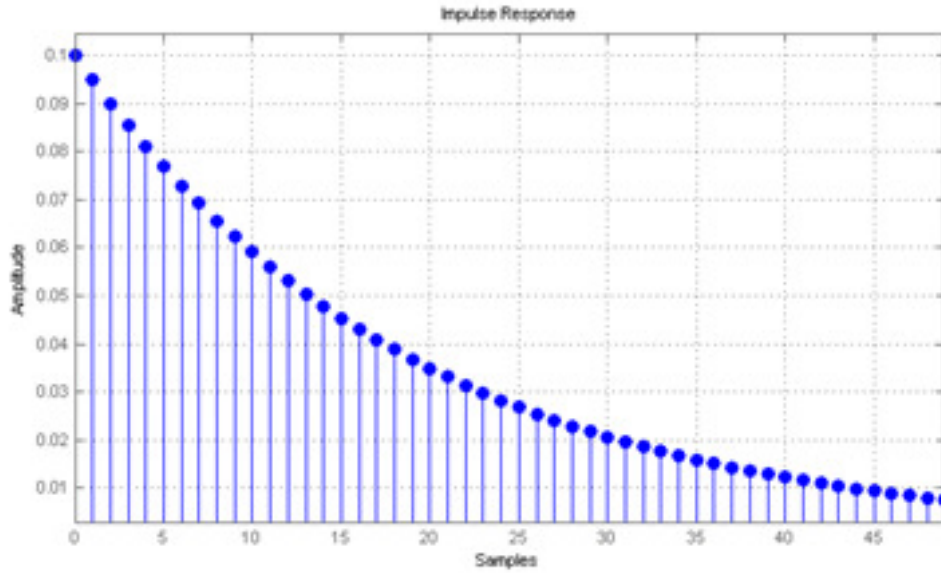
The impulse response of the digital filter is $h[n] = \alpha(1 - \alpha)^n u[n]$ and thus is an exponential decay similar to that of the analog RC filter.

The signal decays to $1/e = 36.8\%$ of its initial value in d samples such that

$$\frac{h[n = d]}{h[n = 0]} = \frac{h_d}{h_0} = \frac{\alpha(1 - \alpha)^d}{\alpha} = (1 - \alpha)^d = e^{-1} \text{ or } 1 - \alpha = e^{-1/d}$$

Given the sampling rate of the filter f_s and the desired time constant $t = RC$ we set the number of samples d needed for the filter output to decay to $1/e = 36.8\%$ of its initial value during the time from $t = 0$ to $t = RC$ to be $d/f_s = RC$ or $d = RC f_s$

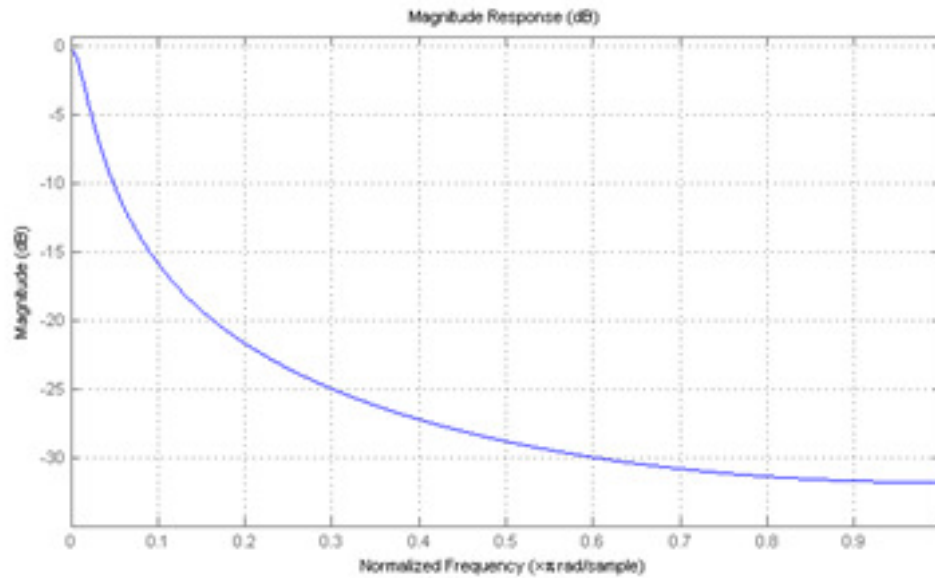
For example, if we assume a sampling rate of 250 KHz and time constant of 75 usec, we set $d = RCf_s = 75 \cdot 10^{-6} \cdot 0.25 \cdot 10^6 = 19$ samples.



The value of α is found via $1 - \alpha = e^{-1/d}$. In this example $1 - \alpha = e^{-1/19} = 0.9487$

An expression for the frequency response is obtained by choosing $z = e^{j2\pi f/f_s}$ around the unit circle in the z -plane, so that

$$H(f) = H(z)|_{z=e^{j2\pi f/f_s}} = \frac{\alpha}{1 - (1 - \alpha)z^{-1}}|_{z=e^{j2\pi f/f_s}} = \frac{\alpha}{1 - (1 - \alpha)e^{-j2\pi f/f_s}} = \frac{1}{\alpha^{-1} - (\alpha^{-1} - 1)e^{-j2\pi f/f_s}}$$



The gain at $f = 0$ or $z = 1$ is $H(z = 1) = \frac{\alpha}{1-(1-\alpha)} = \frac{\alpha}{\alpha} = 1$

The gain at $f = f_s/2$ or $z = e^{j\pi} = -1$ is $H(z = -1) = \frac{\alpha}{1+(1-\alpha)} = \frac{\alpha}{2+\alpha} = \frac{1}{1+2/\alpha}$

The gain is reduced by 3dB at a cutoff frequency f_C such that

$$1/d = 2\pi f_C/f_s \text{ or } (1 - \alpha) = e^{-1/d} = e^{-2\pi f_C/f_s}.$$

The gain at f_C such that $(1-\alpha) = e^{-2\pi f_C/f_s}$ is $H(f = f_C) = \frac{\alpha}{1-(1-\alpha)e^{-j2\pi f/f_s}} = \frac{1-e^{-2\pi f_C/f_s}}{1-e^{-2\pi f_C/f_s}e^{-j2\pi f_C/f_s}}$

Here f_C/f_s is normalized to the sampling rate so that $0 \leq f_C/f_s \leq 0.5$

Recall for the analog filter the cutoff frequency $\frac{1}{2\pi RC} = \frac{1}{2\pi 75 \cdot 10^{-6}} = 2122\text{Hz}$.

We want the gain for the digital filter at 2122 Hz to be 3 dB down.

In this case, assuming a sampling rate of 250 KHz, the normalized frequency is $\frac{f_C}{f_s} = \frac{2122}{250000} = 0.00848$

Substituting these values into the 3dB cutoff frequency result $(1 - \alpha) = e^{-2\pi f_C/f_s}$ we find

$e^{-2\pi f_C/f_s} = e^{-0.00848} = 0.9481 = (1-\alpha)$ consistent with the value of α found via $1-\alpha = e^{-1/d}$ above, and also consistent with the frequency (magnitude) response plot.

Thus a digital de-emphasis filter with time constant 75 usec, cutoff frequency 2122 Hz and sampling rate $f_s = 250000$ Hz may be built using a single pole IIR filter with parameter $\alpha = 1 - 0.948 = 0.052$

5.9 Frequency and Phase Modulation review

Frequency and phase modulation (FM and PM) are two modulation types that encode information within a carrier wave just as in amplitude modulation but instead, the message is represented by varying frequency as opposed to varying amplitude. We start once again with the general signal $s(t)$ formed by multiplying the amplitude function $a(t)$ by a carrier wave $c(t) = \cos[2\pi f_c t + \psi(t)]$.

$$s(t) = a(t) \cos[2\pi f_c t + \psi(t)]$$

Recall that $\psi(t) = 2\pi \Delta f t + \varphi(t)$. Where $\varphi(t)$ represents phase shift but in contrast to AM, it is a function of time. Additionally we begin by neglecting any frequency error Δf leaving us with $\psi(t) = \varphi(t)$ and therefore $s(t) = a(t) \cos[2\pi f_c t + \varphi(t)]$. We let theta represent the argument of the carrier:

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

$$s(t) = a(t) \cos\theta(t)$$

In PM, we use phase shift term, $\varphi(t)$ to encode our message within the frequency of the modulated wave and leave our amplitude function constant.

$$\phi(t) = k_p m(t)$$

$$a(t) = A_c$$

Thus phase modulation involves a modulated wave of form $s(t) = A_c \cos[2\pi f_c t + k_p m(t)]$. An important parameter to which we will refer for both FM and PM, is the signal's instantaneous frequency. This function is denoted as the derivative of the argument:

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt} \theta(t)$$

$$\text{PM: } f_i(t) = \frac{1}{2\pi} \frac{d}{dt} \theta(t) = \frac{1}{2\pi} \frac{d}{dt} [2\pi f_c t - k_p m(t)] = f_c + \frac{1}{2\pi} k_p \frac{d}{dt} m(t)$$

A question that may strike at this point is, "How would we then represent frequency modulation?"

We simply choose $\varphi(t)$ so that $f_i(t) = m(t)$. This means that since $s(t) = A_c \cos[2\pi f_c t + \varphi(t)]$, we let $\varphi(t) = 2\pi \int_{-\infty}^t k_f m(\alpha) d\alpha$. This produces the following transmitted signal with its particular intermediate frequency:

$$\text{FM: } s(t) = A_c \cos[2\pi f_c t + 2\pi k_f \int_{-\infty}^t m(\alpha) d\alpha]$$

$$f_i(t) = \frac{1}{2\pi} \frac{d}{dt} \theta(t) = f_c + k_f m(t)$$

Recall that for AM the function representing the modulated wave's amplitude at any given time is $a(t) = A_c [1 + k_a m(t)] = A_c + A_c [k_a m(t)]$. Analogously for FM $f_i(t) = f_c + k_f m(t)$.

Now that we have the correct form, we can write a complete expression for the FM signal.

$$\begin{aligned}\theta(t) &= 2\pi \int_{-\infty}^t f_i(\alpha) d\alpha \\ \theta(t) &= 2\pi \int_{-\infty}^t [f_c + k_f m(\alpha)] d\alpha \\ \theta(t) &= 2\pi f_c t + 2\pi \int_{-\infty}^t k_f m(\alpha) d\alpha\end{aligned}$$

$s(t) = a(t) \cos[2\pi f_c t + \varphi(t)]$; where $a(t) = A_c$ constant

So we have $\varphi(t) = 2\pi \int_{-\infty}^t k_f m(\alpha) d\alpha$

We will assume that the message $m(t) = A_m \cos(2\pi f_m t)$ meaning that the instantaneous frequency becomes $f_i(t) = f_c + k_f A_m \cos(2\pi f_m t)$. Therefore...

$$\begin{aligned}\varphi(t) &= 2\pi \int_{-\infty}^t k_f A_m \cos(2\pi f_m \alpha) d\alpha \\ \varphi(t) &= 2\pi k_f A_m \frac{\sin(2\pi f_m \alpha)}{2\pi f_m} \Big|_{-\infty}^t \\ \varphi(t) &= \frac{k_f A_m}{f_m} \sin(2\pi f_m t)\end{aligned}$$

A real message is formed by a succession of frequencies representing information. The frequency deviation from f_c is denoted by Δf . This value represents how far on the spectrum the instantaneous frequency is away from the carrier frequency and is equal to $k_f A_m$.

$\varphi(t) = \frac{\Delta f}{f_m} \sin(2\pi f_m t)$; $\beta = \frac{\Delta f}{f_m}$ is called the modulation index

$$s(t) = A_c \cos[2\pi f_c t + \beta \sin 2\pi f_m t]$$

When the modulation index is large compared to one radian, we call the modulation type wideband FM (WBFM). When it is small it is referred to as narrowband FM (NBFM).

Recall that μ , the modulation index for AM, specifies the difference between the maximum and minimum amplitude of the modulated wave just as β specifies the maximum frequency deviation in FM.

AM: $s(t) = [A_c + A_c \mu \cos 2\pi f_m t] \cos 2\pi f_c t$; $\mu = A_m k_a$

$$s(t) = A_c \cos(2\pi f_c t) + \frac{A_c \mu}{2} \cos[2\pi(f_c + f_m)t] + \frac{A_c \mu}{2} \cos[2\pi(f_c - f_m)t]$$

To get the amplitude spectrum $S(f)$, we need to take the Fourier transform of the signal. If the signal $s(t)$ is periodic, we can write its Fourier series but we need $\frac{f_c}{f_m}$ to be an integer.

$$s(t) = \sum_{-\infty}^{+\infty} J_n(\beta) \cos 2\pi(f_c + n f_m)t$$

The discrete spectrum is, $S(f) = \frac{A_c}{2} \sum_{-\infty}^{+\infty} J_n(\beta) [\delta(f - f_c - n f_m) + \delta(f + f_c + n f_m)]$

Note this spectrum's symmetry as the Bessel Function, $J_n(\beta) = (-1)^n J_{-n}(\beta)$. Also note that the sidebands that are spaced at $nf_m, n = \pm(1, 2, 3, \dots)$. The values of the Bessel Function are decreasing in magnitude such that $\sum J_n^2(\beta) = 1$.

The Bessel function J_n depends upon n and β like a damped sine wave. Using the signal's Beta value, and the integer n , we can find the value of $J_n(\beta)$.

Once again we draw comparisons to the form of an AM wave. When presented in a similar fashion, it is plain to see that both modulation types are analogues of each other:

$$\text{AM: } a(t) = A_c + A_c \mu \cos(2\pi f_m t)$$

$$s(t) = A_c \sum_{n=-1}^1 k_n(\mu) \cos[2\pi(f_c + nf_m)t]; \mu = A_m k_a$$

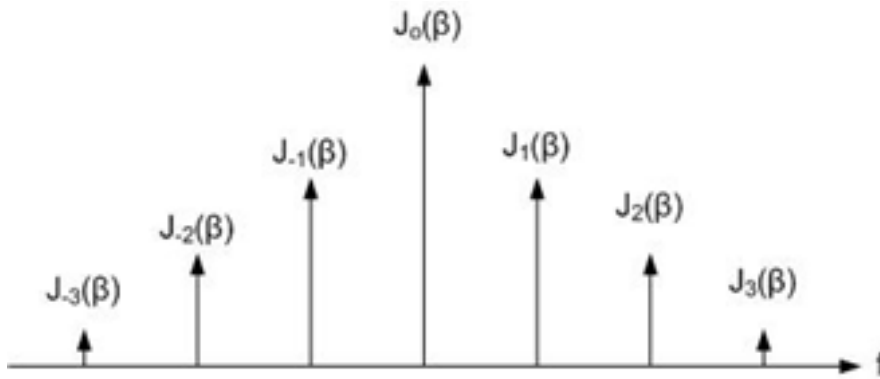
$$k_0 = 1, k_{\pm 1} = \frac{\mu}{2}; \text{ for } n \neq 0, \pm 1, k_n = 0$$

$$\text{FM: } f_i(t) = f_c + \Delta f \cos 2\pi f_m t$$

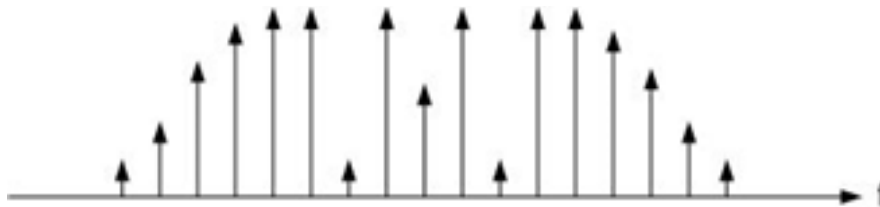
$$s(t) = A_c \sum J_n(\beta) \cos[2\pi(f_c + \mu f_m)t]; \beta = \frac{\Delta f}{f_m}$$

5.9.1 Redundancy

Thus far we have discussed in detail, single sideband (suppressed carrier), double sideband (suppressed carrier), standard AM, and FM. The first of these has no redundancy as only one sideband exists holding the message information (this lack of redundancy occurs at the expense of ease of demodulation). DSB-SC includes two copies of the message, allowing the carrier to be located with an IQ receiver; AM signals are not extremely power efficient but they can be received with a simple analog envelope detector; FM waves have many sidebands and hence are highly redundant. One should not look on redundancy in the FM wave as wasteful however. It means that the FM wave can be received even with interference. In FM, one can properly demodulate a signal provided its signal to noise ratio (S/N) is above the "capture threshold", meaning that inherent noise is not above a level that renders the information indiscernible.



Bessel Function value $[J_n(\beta)]$ defining the FM spectrum depends upon β



Example of FM spectrum with relatively large β

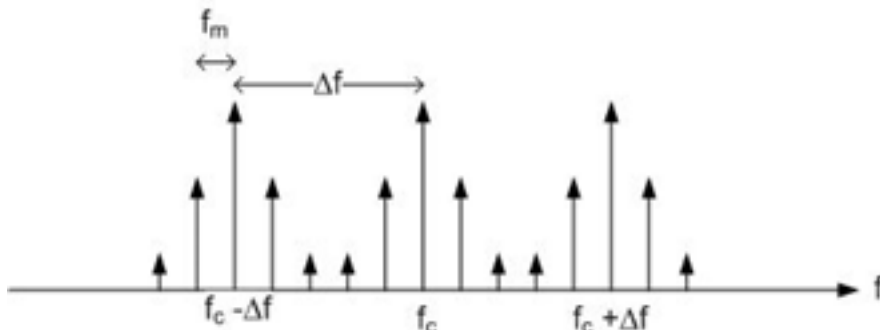
The FM spectrum depends upon the modulation index β ; this means it is a function of Δf and f_m .

Consider these examples:

1. Beta is large (WBFM): $\beta = \frac{\Delta f}{f_m} = 10$; $\Delta f = 10f_m$;

$$f_{max} = f_c + \Delta f; f_{min} = f_c - \Delta f$$

Wideband Frequency Modulation: $\Delta f \ll f_m$; $\beta \gg 1$

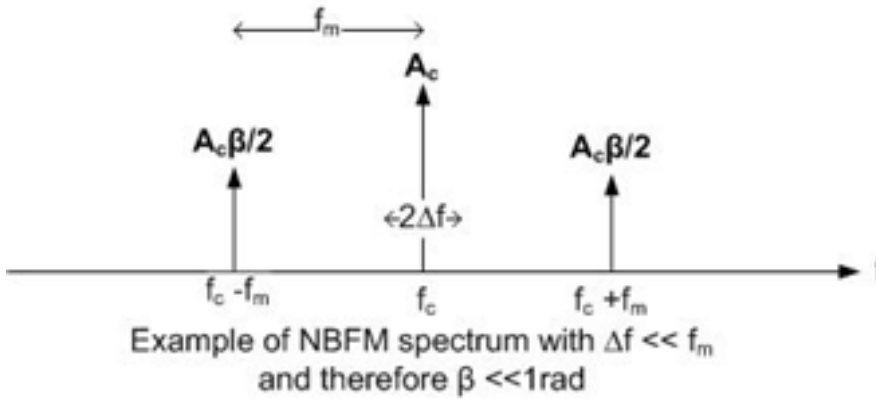


Example of WBFM spectrum with $\Delta f \gg f_m$ and therefore $\beta \gg 1$ rad

- Beta is small (NBFM); $\beta = \frac{\Delta f}{f_m} = 0.1$; $\Delta f = 0.1f_m$

$$f_i(t) = f_c + k_f m(t) = f_c + \Delta f \cos(2\pi f_m t)$$

Narrowband Frequency Modulation: $\Delta f \ll f_m$; $\beta \ll 1$



If we were to view the frequency domain representation of the FM signal and the observation (time) interval is far less than the message period $T_m = \frac{1}{f_m}$, the spectrum consists of one impulse or “spike” at $f_i(t)$. If the observation interval is much greater than $\frac{1}{f_m}$, then the spectrum is discrete with sidebands spaced at intervals of f_m .

A useful and highly practical exercise is to simplify the FM signal expression, $s(t)$ for narrowband FM (when $\beta \ll 1$) and interpret what it means:

In general, $s(t) = A_c \cos[2\pi f_c t + \beta \sin(2\pi f_m t)]$

Now we apply the trigonometric identity $\cos(\alpha + \beta) = \frac{1}{2}[\cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta)]$

So we get, $s(t) = \frac{A_c}{2}[\cos(2\pi f_c t) \cos[\beta \sin(2\pi f_m t)] - \sin(2\pi f_c t) \sin[\beta \sin(2\pi f_m t)]]$

Now use narrowband approximation, $\beta \ll 1$ and recall that for small x , $\sin x \approx x$, $\cos x \approx 1$

Thus: $s(t) = A_c \cos(2\pi f_c t) (1) - A_c \sin(2\pi f_c t) [\beta \sin(2\pi f_m t)]$

So for NBFM: $s(t) = A_c \cos(2\pi f_c t) - A_c \sin(2\pi f_c t) [\beta \sin(2\pi f_m t)]$

Recall the general modulated signal of any type must be of the form:

$$s(t) = v_i(t) \cos(2\pi f_c t) + v_q(t) \sin(2\pi f_c t) \text{ or } s(t) = x(t) \cos(2\pi f_c t) + y(t) \sin(2\pi f_c t)$$

As is apparent from the above NBFM approximation, $x(t) = A_c$ and $y(t) = A_c \beta \sin(2\pi f_m t)$. In general for frequency modulation, $x(t) = \cos(\beta \sin[2\pi f_m t]) = \cos(2\pi k_f \int_0^t m(\alpha) d\alpha)$ and $y(t) = \sin(\beta \sin[2\pi f_m t]) = \sin(2\pi k_f \int_0^t m(\alpha) d\alpha)$.

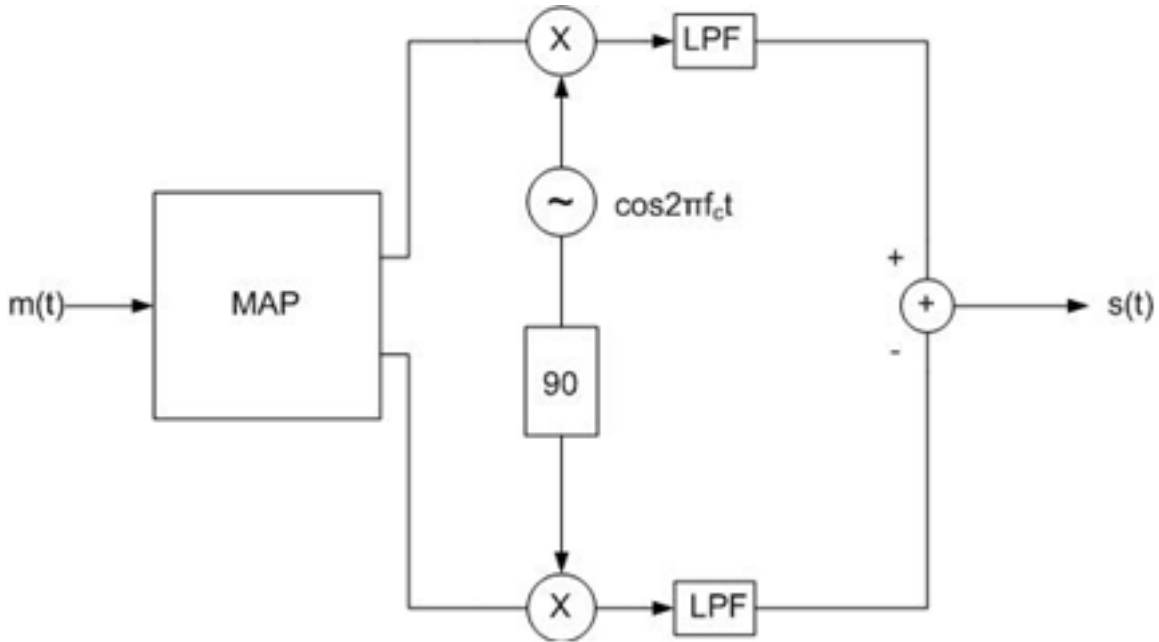
We may seem to be constantly bringing up similarities to AM, but due to NBFM comprising two sidebands and a carrier (as displayed below) we can think of this wave as being an AM/FM hybrid of sorts.

Given the identity, $\sin(\alpha)\sin(\beta) = \frac{1}{2}\cos(\alpha - \beta) - \frac{1}{2}\cos(\alpha + \beta)$

We have...

FM: $s(t) = A_c \cos(2\pi f_c t) - \frac{A_c}{2} \beta \cos[2\pi(f_c - f_m)t] + \frac{A_c}{2} \beta \cos[2\pi(f_c + f_m)t]$

AM: $s(t) = A_c \cos(2\pi f_c t) + \frac{A_c}{2} \mu \cos[2\pi(f_c + f_m)t] + \frac{A_c}{2} \mu \cos[2\pi(f_c - f_m)t]$



General IQ Transmitter

5.9.2 NBFM Transmission

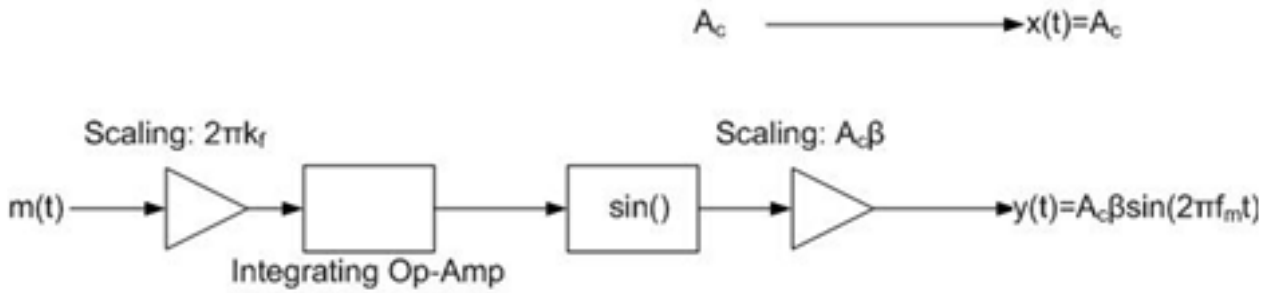
We now seek a method to build an FM transmitter for general, single-tone message $m(t)$, using analog components. We need a way to control instantaneous frequency. We may use an RLC circuit for which the resonant frequency is the following:

$$f_i(t) = \frac{1}{2\pi\sqrt{LC(t)}}; \text{for } \beta \ll 1; \beta = \frac{\Delta f}{f_c} = \frac{k_f A_m}{f_m}; k_f \ll 1$$

$$s(t) = A_c \cos(2\pi f_c t) - A_c \sin(2\pi f_c t) (2\pi k_f \int_0^t m(\alpha) d\alpha)$$

So with a voltage regulated capacitor or varactor (of capacitance $C(t)$) we can control the capacitance and thus the instantaneous frequency output of the transmitter's local oscillator.

However, the maximum variation in capacitance achievable with a varactor is not large enough to successfully produce a wideband signal (WBFM requires great frequency deviation). This approach will work well for narrowband frequency modulation but not beyond that.



Map for NBFM Transmission

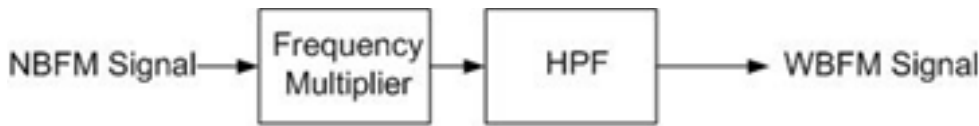
5.9.3 WBFM Transmission

An analog transmitter for WBFM uses the voltage controlled capacitor in a slightly more complicated manner in which an intermediate frequency is defined as the resonant frequency of the RLC circuit.

$$f_{IF} = \frac{1}{2\pi\sqrt{LC}}$$

The instantaneous frequency is $f_i(t) = f_{IF} + \Delta f \cos(2\pi f_m t)$

In an effort to produce a wideband signal, we may take a NBFM signal and run it through a frequency multiplier which serves to duplicate sidebands and effectively increase redundancy. The parameters f_c , Δf , k_f , f_m become nf_c , $n\Delta f$, nk_f , f_m where n serves to indicate the location of each set of sidebands on the spectrum.



WBFM production from NBFM

If we consider our carrier to be at $f_c = nf_{IF}$ in which case, our frequency deviation becomes $n\Delta f$ and our instantaneous frequency would

$$nf_{IF} + n\Delta f \cos(2\pi f_m t)$$

The first term in the above represents the carrier frequency and the second shows a wide deviation in frequency for large n .

5.9.4 Approximate Bandwidth of FM signal

To acquire the approximate bandwidth W (or B) we may use “Carson’s Rule” which states that $W = 2(\Delta f + f_m) = 2f_m \left(1 + \frac{\Delta f}{f_m}\right) = 2f_m(1 + \beta)$ where β is the modulation index (bandwidth expansion factor). For NBFM, $W \cong 2f_m$.

AM and FM Spectra:

$$\text{AM: } s(t) = A_c \sum_{n=-1}^1 k_n(\mu) \cos[2\pi(f_c + n f_m)t]$$

$$\text{FM: } s(t) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \cos[2\pi(f_c + n f_m)t]$$

$$\text{NBFM: } s(t) = A_c \sum_{n=-1}^1 J_n(\beta) \cos[2\pi(f_c + n f_m)t]$$

Observe that for AM, the summation includes the carrier and two sidebands (-1, 0, 1) as does NBFM (analogous to AM), while WBFM includes all sidebands summed over infinity (though low power make higher order terms negligible).

Stereo FM

We have two messages, $m_1(t)$ and $m_2(t)$ each representing left and right parts of the audio signal $m_L(t)$ and $m_R(t)$. We have some choices for how we will modulate the FM wave. We can combine amplitude and frequency modulation:

1. $s(t) = a(t) \cos[2\pi f_c t + \varphi(t)]; m_1(t) = a(t); m_2(t) = \varphi(t)$

This is not typically a good idea for FM since we are storing information in $a(t)$ and many FM receiver designs ignore amplitude variations as often a hard limiter is applied to unify the incoming signal’s amplitude. A second possibility is to store $m_1(t)$ on the upper sideband $m_2(t)$ on lower sideband using Hilbert Transforms of each to ensure that they each only occupy a single sideband:

We have managed to run into the same problem of storing information within the amplitude here once again. We now try the idea of sum and difference frequencies:

1. $m_m(t) = m_L(t) + m_R(t)$; a mono signal

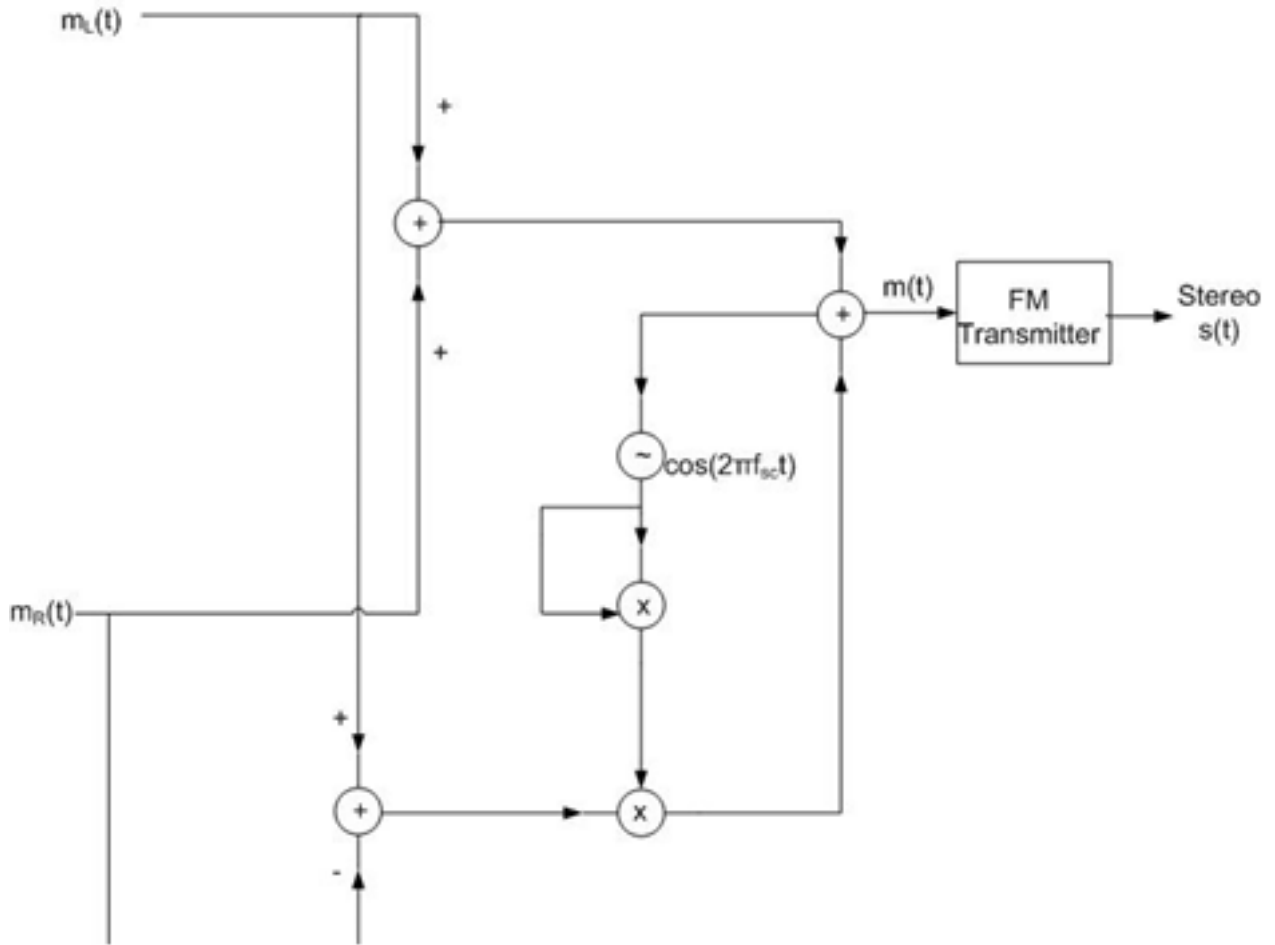
$$m_s(t) = m_L(t) - m_R(t); \text{ a difference signal}$$

Therefore we have a new form of message storing all pertinent information for stereo audio. In this method, we will modulate solely the frequency:

$$m(t) = m_L(t) + m_R(t) + \cos(2\pi f_{sc}t) + [m_L(t) - m_R(t)]\cos(4\pi f_{sc}t)$$

$m(t) = m_m(t) + [m_s(t)]\cos(4\pi f_{sc}t) + \cos(2\pi f_{sc}t)$; f_{sc} is the sub-carrier.

$$s(t) = A_c \cos\left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha\right]$$



Stereo FM Transmitter

5.9.5 FM Demodulation:

Let us begin to understand the principles of FM reception with an intuitive approach to the extraction of the argument, $\theta(t)$. Recall that the message is stored in this angle as demonstrated below.

$$s(t) = A_c \cos \theta(t)$$

$$s(t) = A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right]$$

Taking the derivative will yield the following form which is reminiscent of our AM modulated wave. A simple envelope detector or the IQ receiver may be used to acquire the envelope $[a(t)]$ when a transmission has this form $s(t) = a(t) \cos [\theta(t)]$.

$$\frac{d}{dt} s(t) = A_c \sin [\theta(t)] \frac{d\theta}{dt} = -A_c \sin [\theta(t)] [2\pi f_c + 2\pi k_f m(t)]$$

$$\frac{d\theta(t)}{dt} = 2\pi f_c \left[1 + \frac{k_f}{f_c} m(t) \right]$$

$$\frac{d}{dt}s(t) = -A_c \sin[\theta(t)] 2\pi f_c \left[1 + \frac{k_f}{f_c} m(t)\right]$$

$$a(t) = -A_c 2\pi f_c \left[1 + \frac{k_f}{f_c} m(t)\right]$$

This is the amplitude function or envelope and our signal has the form $s(t) = a(t)\sin[\theta(t)]$. Upon extracting this, we may remove the DC component and scale appropriately to get $m(t)$.



Unenlightened Approach to FM Demodulation

Practical experimentation has will expose a problem with this approach. As an FM wave's amplitude tends to vary with frequency, the envelope detector will also detect these variations. The message received will not be the one sent.

One way which the amplitude of an FM wave is forced constant is the use of a. This device "clips" any sinusoid whose amplitude is greater than a specified magnitude. The result is a uniform amplitude for all frequencies throughout a signal. A hard-limiter creates unwanted "out of band" distortion terms but these can be filtered out. Clipping a signal will introduce many harmonic frequencies (which are mathematically necessary to form the near-square appearance of a clipped waveform) but the zero crossings (time axis intercepts) will remain unchanged; clipping is the loss of amplitude, not frequency information which is okay considering we have stored no info within amplitude variations.



$s(t)$ Processed Though a Hard Limiter

We seek now, a more effective approach to this demodulation technique after limiting has taken place. The steps involved in successful FM demodulation of and FM wave are as follows.

1. Limiter- As previously discussed, the amplitude changes of the FM wave are suppressed and the wave is clipped.
1. Filter- We apply a filter (BPF) to the signal to remove the many additional harmonics that have been introduced through the limiting process. With this we have removed any present amplitude alterations which posed a problem in the demodulation technique discussed above.
1. Differentiator- The derivative of the signal must be taken. Recall that a differentiator has frequency response $H(f) = j2\pi f$.

Recall the form of the frequency response of a bandpass filter. We may make the BPF wider or narrower with a certain centre frequency by altering the circuit which represents it. If we place two BPFs with opposing polarities, each centred at offset frequencies, we get a linear portion with a constant slope. We may vary the parameters of each BPF and their respective centre frequencies in order to acquire our ideal constant slope of $H(f) = j2\pi f$. When the bandpass filter is applied to the FM signal, the output is the differentiated FM wave: $\frac{d}{dt}s(t) = -A_c \sin[\theta(t)] 2\pi f_c [1 + \frac{k_f}{f_c} m(t)]$.

With the effect of the limiter being zero amplitude deviation prior to differentiation, we now can extract an envelope (instantaneous frequency) using an envelope detector.

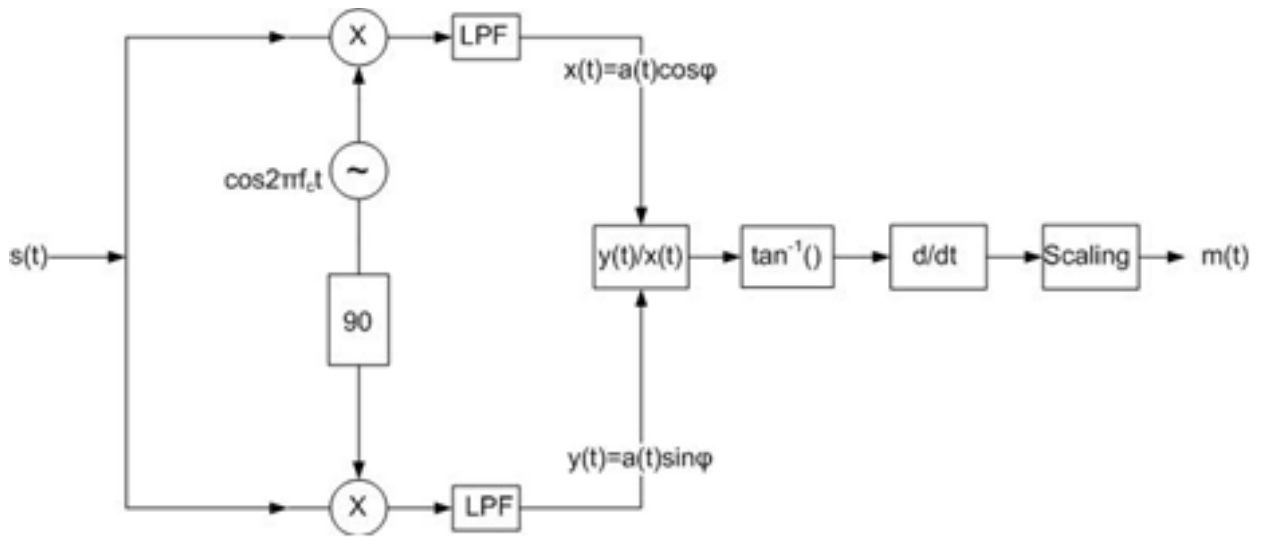
In practice the instantaneous frequency can be found using an IQ detector. We still first apply a limiter and a bandpass filter then we run the signal through the general IQ configuration:

$$\begin{aligned} s(t) &= A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right] \\ s(t) &= a(t) \cos [2\pi f_c t + \varphi(t)] \\ s(t) &= a(t) \cos [\theta(t)] \\ a(t) &= A_c \end{aligned}$$

From this, we acquire the following terms: $x(t) = a(t) \cos [\varphi(t)]$ and $y(t) = a(t) \sin [\varphi(t)]$. We divide these [keeping in mind that $\sin(a)/\cos(a) = \tan(a)$] and we are left with $\tan[\varphi(t)]$. Taking the inverse tangent of this expression yields $\varphi(t)$, which is proportional to the integral of the message. Using a differentiator (software version or the analog filter method outlined above) and finally scaling and DC-shifting as necessary we produce our message.

$$\begin{aligned} \varphi(t) &= 2\pi k_f \int_0^t m(\alpha) d\alpha \\ \frac{d}{dt} \varphi(t) &= 2\pi k_f m(t) \\ f_i(t) &= \frac{1}{2\pi} d\theta/dt = f_c + \frac{1}{2\pi} d\varphi/dt = f_c + k_f m(t) \end{aligned}$$

There are many ways to demodulate an FM wave. It is noteworthy to point out that if one is implementing a software method, simply analyzing the zero crossings will produce the instantaneous frequency for an interval, regardless of amplitude. This method works especially well if frequency shift keying is used (FSK) in which a set number of frequencies (two for binary, sixteen for hexadecimal, etc) are used to transmit a message. One can write a program to find the carrier frequency and determine when each of these frequencies are active and thus what information is encoded.



FM Receiver With Explicit "Map"

Chapter 6

Baseband data transmission

6.1 Digital messages

$m(t)$ is a digital message 11100101 as shown in the figure below.



Figure 1 Digital Message

In Figure 1, $p(t)$ is chosen to be a so-called square pulse, where we can write $p(t) = 1$ for $0 < t < T$ and 0 otherwise. We send one such pulse for each data bit, a positive pulse $p(t)$ for logic 1 and a negative pulse for logic 0. In the figure, the message is shown in the interval $0 < t < 8T$ and the data bits are 11100101.

For the period $0 < t < T$, $m(t) = A_0 p(t) = +1$ and the data is logic 1. For the next period $T < t < 2T$, $m(t) = A_1 p(t - T) = +1$, and the data is logic 1. In the example figure above, the message is shown in the interval $0 < t < 8T$ and the information bits are 11100101

We can write $m(t) = A_0 p(t) + A_1 p(t - T) + A_2 p(t - 2T) + \dots = \sum_k A_k p(t - kT)$ where the real constants $A_k = \pm 1$ depending on whether a logic 1 or a logic 0 was sent.

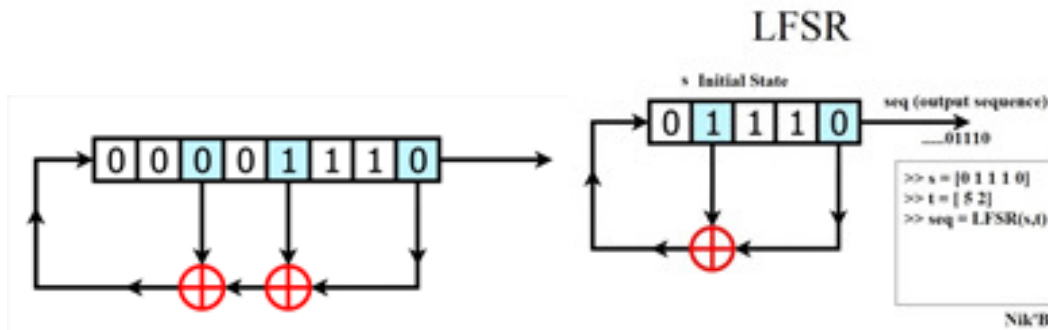
In the figure above, $A_0 = +1, A_1 = +1, A_2 = +1, A_3 = -1, A_4 = -1, A_5 = +1, A_6 = -1, A_7 = +1$

The constants may also be chosen to be $A_k = \pm 1, \pm 3$. In this 4 level system, each pulse represents one of 4 combinations of 2 bits: 00, 01, 11, 10.

6.2 Data source

The data source produces the data symbols $A_k = \pm 1$ that may be stored in memory or generated on the fly. For testing purposes, it is convenient to have a random data source, with an equal number of logic 1s and logic 0s, or an equal number of positive and negative pulses. This can be achieved with a so-called linear feedback shift register.

Two examples are shown in the figure below.



At each time step (symbol time) T , the input on the left is determined by the XOR gate output and all the bits slide along one step and the output is the bit on the right. If the feedback shift register is of length M and the feedback taps (XOR gates) are selected correctly, the output sequence will have length $2^M - 1$ before it repeats.

6.3 Differential encoding and decoding

6.3.1 Avoiding polarity inversion

A data receiver may easily invert the polarity of the received sequence. This can happen for example if the data is represented by phase shifts as will be shown in section 7 on phase shift keying. It is easy to detect changes in phase but much more difficult if not impossible to detect absolute phase.

To avoid this problem, the data sequence is differentially encoded. *Data are represented in terms of the changes between successive signal elements (bits) rather than the signal elements (bits) them-*

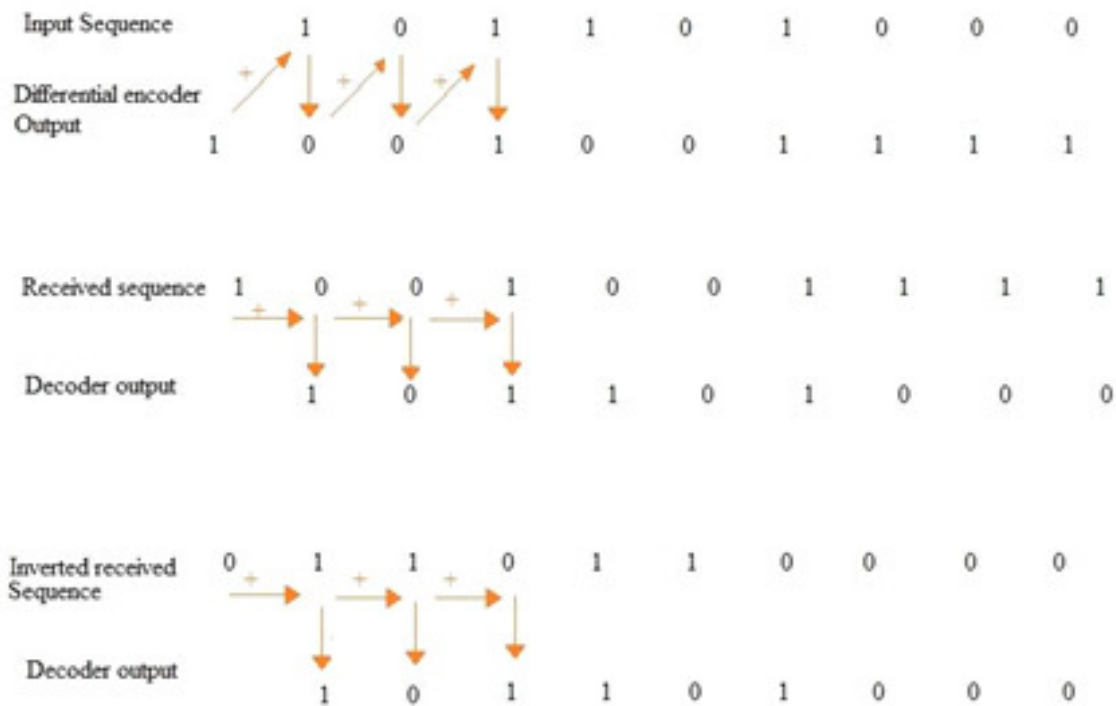
selves.

The polarity of the differentially encoded signal can be unintentionally inverted by the receiver. This inversion will not have any effect on the decoded signal waveform, as shown in the example below. The encoder output starts with a 1, but this is arbitrary, it could also be a 0. The received sequence can be inverted or not inverted. The decoder output is the same in both cases.

6.3.2 Differential encoding

The arrows show how the differential encoding works. At the beginning, we have $1+1=0$, then $0+0=0$, then $0+1=1$, etc.

The arrows also show how the differential decoding works. For the received sequence, we have $1+0=1$, then $0+0=0$, then $0+1=1$, etc. For the inverted sequence, $0+1=0$, $1+1=0$, $1+0=1$, etc. In both cases, a change in the received sequence results in the decoder outputting a “1” bit, and no change results in a “0” bit.



Working of Differential encoder and Differential Decoder.

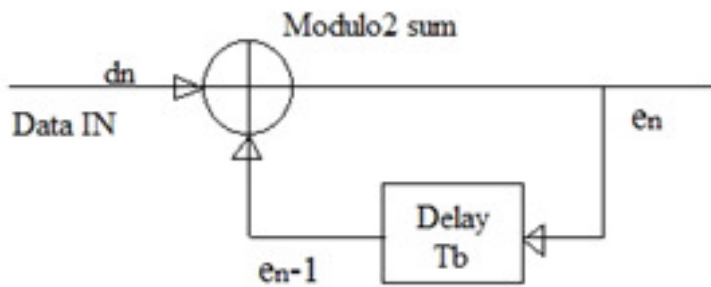
Exercise: repeat the figure above, but with the first bit in the differential encoder set to 0 instead of 1.

In some cases, the decoder output is polarity reversed by convention, so that a change in the received sequence is decoded as a “0” bit and no change as a “1” bit.

In software simulation, there is no chance that a bit sequence will be reversed. So when creating a simulation model of a Differentially Encoded sequence, we can ignore this baseband coding.

In the differential encoder, input data bits are modulo 2 SUM with the previous output bits. Modulo 2 SUM is same as EX-OR. The differential encoder equation is

$$e_n = d_n \oplus e_{n-1}$$



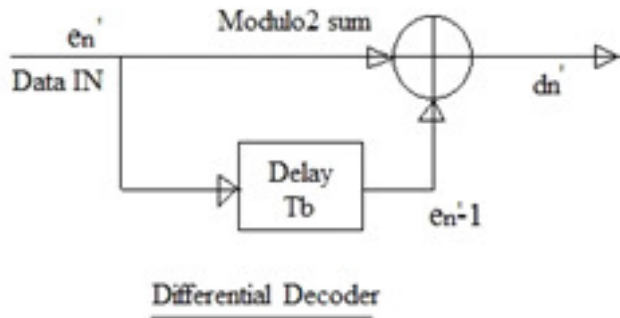
Differential Encoder

6.3.3 Differential decoding

In the **differential decoder**, the current input and a delayed version of the same is fed to the module 2 sum. This produces the output bits. The differential decoder equation is

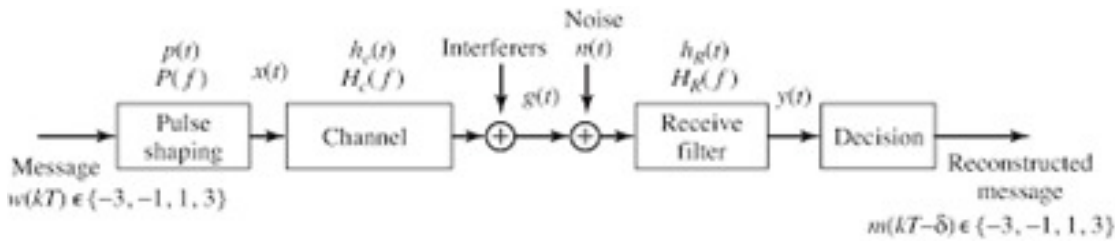
$$d'_n = e'_n \oplus e'_{n-1}$$





6.4 Data transmission system with pulse shaping

A data transmission system that sends bits from one place to another contains transmitter with pulse shaping as shown above followed by a receiver with receive filter and a decision device. The receive filter is ideally matched to the transmit pulse shape filter, but may be a simple low pass filter.



The received message waveform after filtering by both the transmitter pulse shaping filter and the receiver low pass or matched filter is a sequence of delayed pulses $m(t) = \sum_k A_k p(t - kT)$. In general the waveform $m(t)$ will have a peak amplitude that varies depending on the data sequence A_k

The decision device carries out two tasks:

1. (a) i. A. samples this waveform $m(t)$ at the time $t = kT + \delta$ of maximum amplitude for each symbol, where the value of δ depends on the pulse shape, and
 - B. decides what is the value of the symbol (1 or -1 for a binary two level system, or $\pm 1, \pm 3$ for a 4 level system).

Pulse shaping can be done in multiple ways. Here we consider three methods:

1. (a) i. A. Generate square pulses and filter with low pass filter.

- B. Generate impulses and filter with *time domain* FIR pulse-shaping filter.
- C. Generate impulses and filter with *frequency domain* FIR pulse-shaping filter.
- D. Generate impulses and filter with Gaussian pulse-shaping filter (*t and f domain*)

In the following sections, we describe how to

- represent data as square pulses in a sampled system
- filter the square pulses in time domain
- represent data sequence with impulses instead of square pulses
- filter impulses with time domain RC filter
- filter impulses with raised cosine filter shape in frequency domain
- filter impulses with root raised cosine filter in frequency domain

6.4.1 Representing data as square pulses in a sampled system

For a sampling frequency f_s and symbol time T , there are $f_s T$ samples per symbol. Each square pulse contains $f_s T$ samples of identical value.

The data symbols $A_k = \pm 1$ are represented by one sample per symbol and come from the data source at a rate $1/T$ symbols per second. To generate the square pulse, each sample of data must be repeated $f_s T$ times, so that the square pulse is represented by $f_s T$ samples running at sampling frequency f_s to make up a symbol that takes time T to transmit. We also refer to the symbol time as the symbol length. A sequence of such symbols will look like the waveform $m(t)$ shown in the figure below with $f_s T$ samples per symbol.

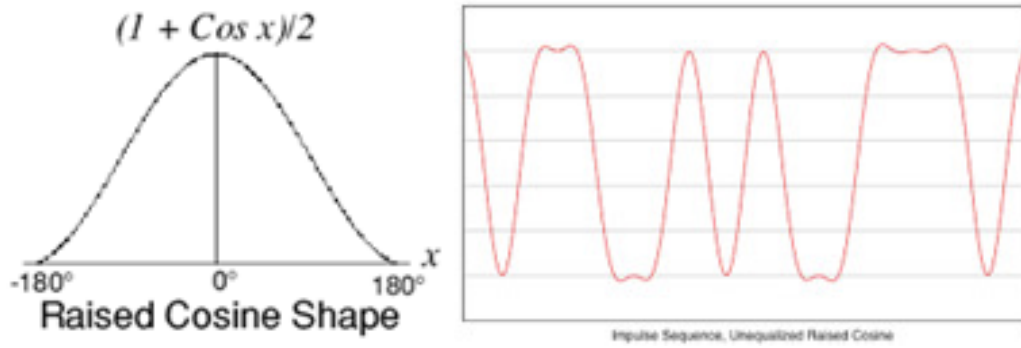


6.4.2 Pulse shaping to eliminate the sharp edges of a square pulse

The square pulse can be low pass filtered with a cutoff frequency of $1/(2T)$ since this is the highest frequency in the data stream for an alternating data sequence 101010... $A_k = 1, k$ odd, $A_k = -1, k$ even. Ideally the low pass filter has an impulse response such that with a square pulse input starting at $t = 0$ and ending at $t = T$, the output is approximately a *raised cosine* pulse in the time domain, starting at $t = 0$ and ending at $t = 2T$ as shown in the figure below.

We write the RC2 raised cosine pulse $p_2(t) = [1 - \cos(\pi t/T)]/2$ so that $p_2(t) = 0$ for $t \in \{0, 2T\}$ and $p_2(t) = 1$ for $t = T$ (since $\cos \pi = -1$).

When a data sequence is sent with these pulses, the waveform has rounded edges with minimal overshoot as shown in the figure on the right.



Raised cosine from $t = 0$ to $t = 2T$

Mathematically we can write that the sequence $m(t) = A_0p(t) + A_1p(t - T) + A_2p(t - 2T) + \dots = \sum_k A_k p(t - kT)$ using square pulses $p(t) = 1$ for $0 < t < T$ and 0 otherwise

is low pass filtered by an analog Bessel filter with impulse response $h(t)$ to yield another sequence of pulses

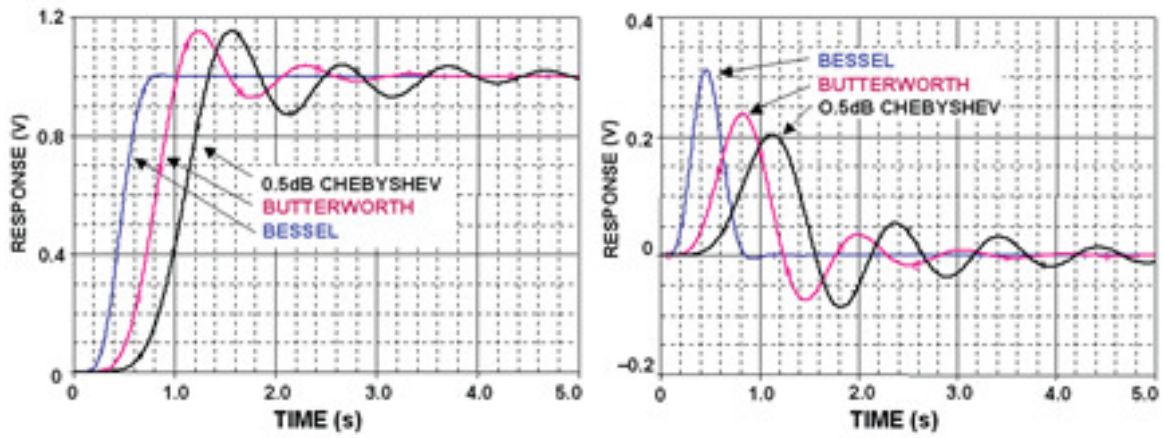
$$m(t) = A_0p_2(t) + A_1p_2(t - T) + A_2p_2(t - 2T) + \dots = \sum_k A_k p_2(t - kT)$$

where $p_2(t) = p(t) \otimes h(t)$ has an approximately raised cosine shape from $t = 0$ to $t = 2T$.

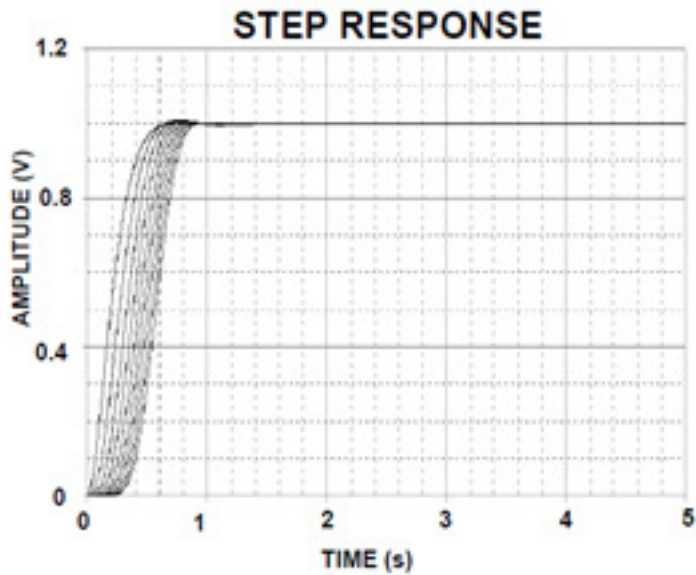
In other words, the response of $h(t)$ to a single square pulse $p(t) = 1$ from $t \leq 0 \leq T$ will be the raised cosine pulse from $t = 0$ to $t = 2T$

For an alternating sequence $A_k = 1, k$ odd, $A_k = -1, k$ even, $m(t) = \cos \pi t/T = \cos 2\pi f_m t$ where $f_m = 1/2T$ (*exercise*)

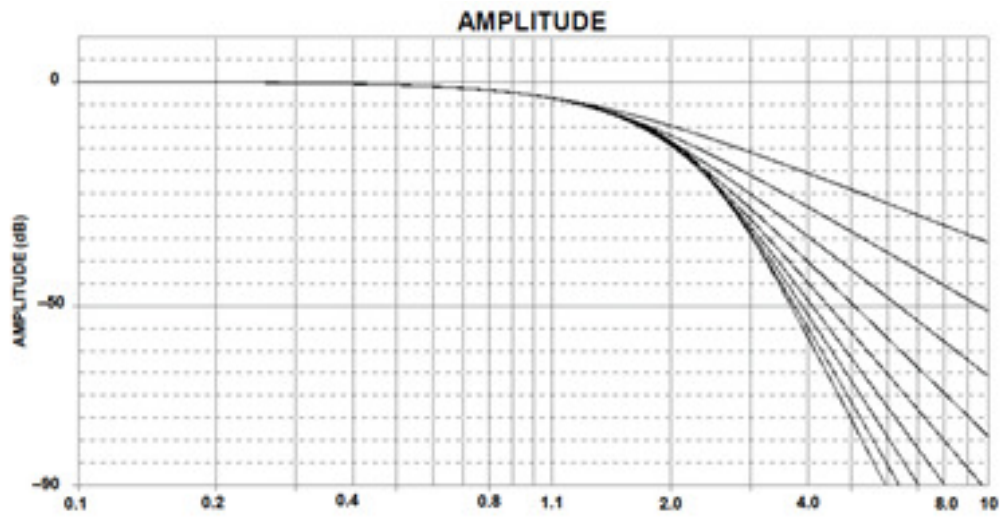
The step response (integral of the impulse response) of the Bessel filter $h(t)$ approximates half of the raised cosine shape, as shown in the figures below, starting at $p_2(t = 0) = 0$ to $p_2(t = T) = 1$.



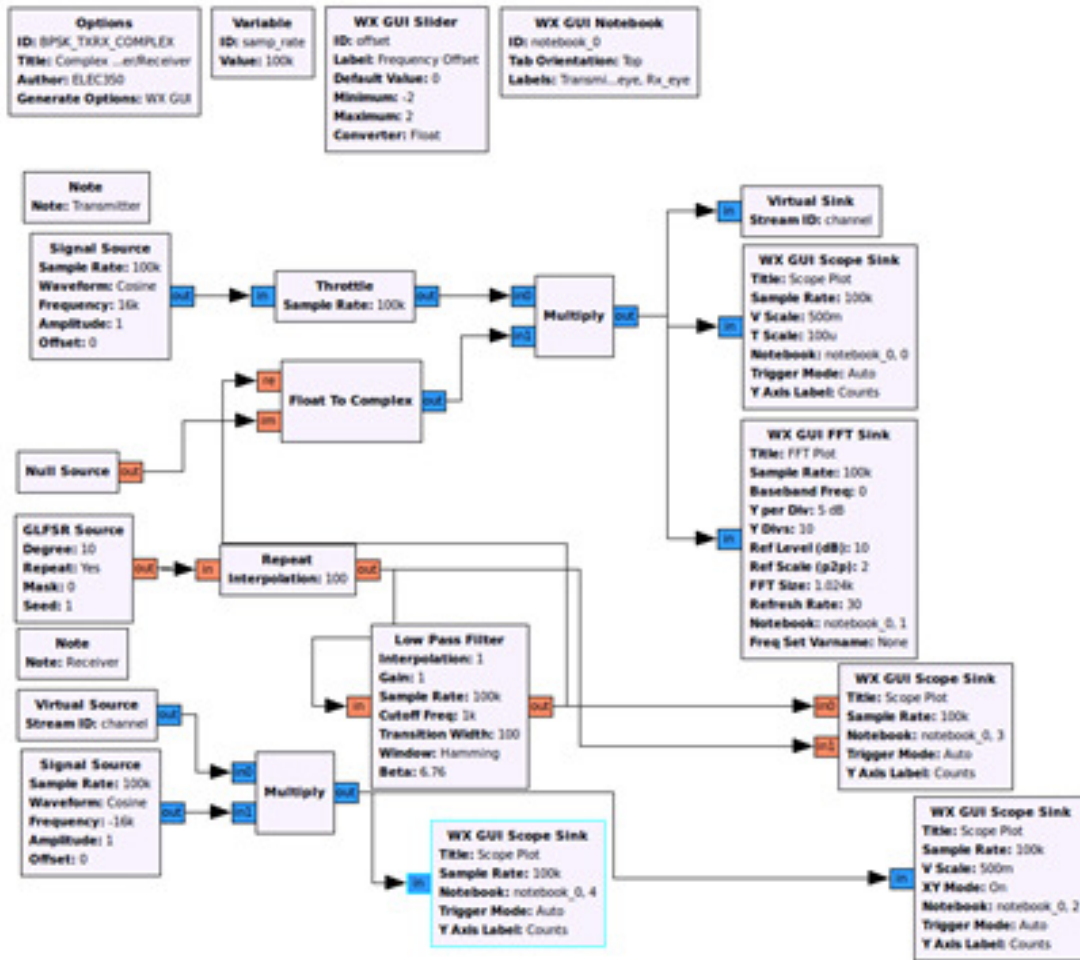
A family of Bessel step responses is shown below, depending on the number of poles.



The frequency response of this filter is shown below



Pulse shaping of a square pulse using a low pass filter is shown in the GRC flowgraph below. The sampling rate is set to $f_s = 100K$.



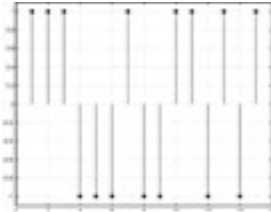
The linear feedback shift register (GLFSR) data source generates one sample for each data symbol $A_k = \pm 1$ at a rate $1/T$ symbols per second. The Repeat block interpolates to make $f_s T = 100$ samples for each data symbol, so that each square pulse contains 100 samples of identical value. The symbol rate is thus

$$\frac{1}{T} = \frac{1000 \text{ symbols}}{\text{sec}} = \frac{100K \text{ samples}}{\frac{\text{sec}}{100 \text{ samples}} \text{ symbol}} = \frac{f_s}{f_s T}$$

The low pass filter cutoff frequency is set to 1000 Hz and approximates the Bessel filter.

6.4.3 Representing data sequence with Impulses instead of square pulses

The data sequence can be represented by the sequence of impulses $m_\delta(t) = A_0\delta(t) + A_1\delta(t - T) + A_2\delta(t - 2T) + \dots = \sum_k A_k\delta(t - kT)$.

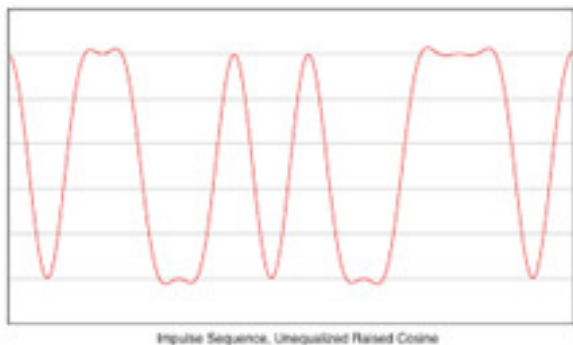


This sequence of impulses $m_\delta(t)$ as shown above is a waveform with only one sample per symbol, so that we can write $m_{\delta,k} = m_\delta(t = kT)$ and the sampling rate is $1/T$.

For transmission, this waveform is convolved with a pulse shaping filter with impulse response $p(t)$ and frequency response $P(f)$ where the convolution operates at a sample rate f_s . To represent this sequence of impulses $m_{\delta,k} = m_\delta(t = kT)$ as a waveform sampled at f_s , the impulses must be filtered by an interpolating FIR filter that yields one sample $A_k = \pm 1$ followed by $f_s T - 1$ samples set to zero, so that the $f_s T$ samples per symbol have only one non-zero value. This interpolating FIR filter will have only one non-zero tap and $f_s T - 1$ zero taps.

This waveform $m_\delta(t)$ sampled at f_s is then in turn convolved with an (regular, non-interpolating) FIR filter (operating at f_s) with impulse response $p(t)$. The output of this FIR filter yields the message waveform $m(t)$. The filter coefficients of the FIR filter will be the pulse shape $p(t)$ sampled at f_s . $p(t)$ can be a square pulse $p(t) = 1$ for $0 < t < T$ and 0 otherwise and thus will have $f_s T$ taps. However, the FIR coefficients can also represent a longer pulse, for example the raised cosine pulse $p(t) = p_2(t) = 0.5(1 - \cos(\pi t/T))$, $0 < t < 2T$, 0 otherwise and will have $2f_s T$ taps in this case. Note that the raised cosine pulse extends over 2 symbol periods, so that adjacent symbol pulses will overlap.

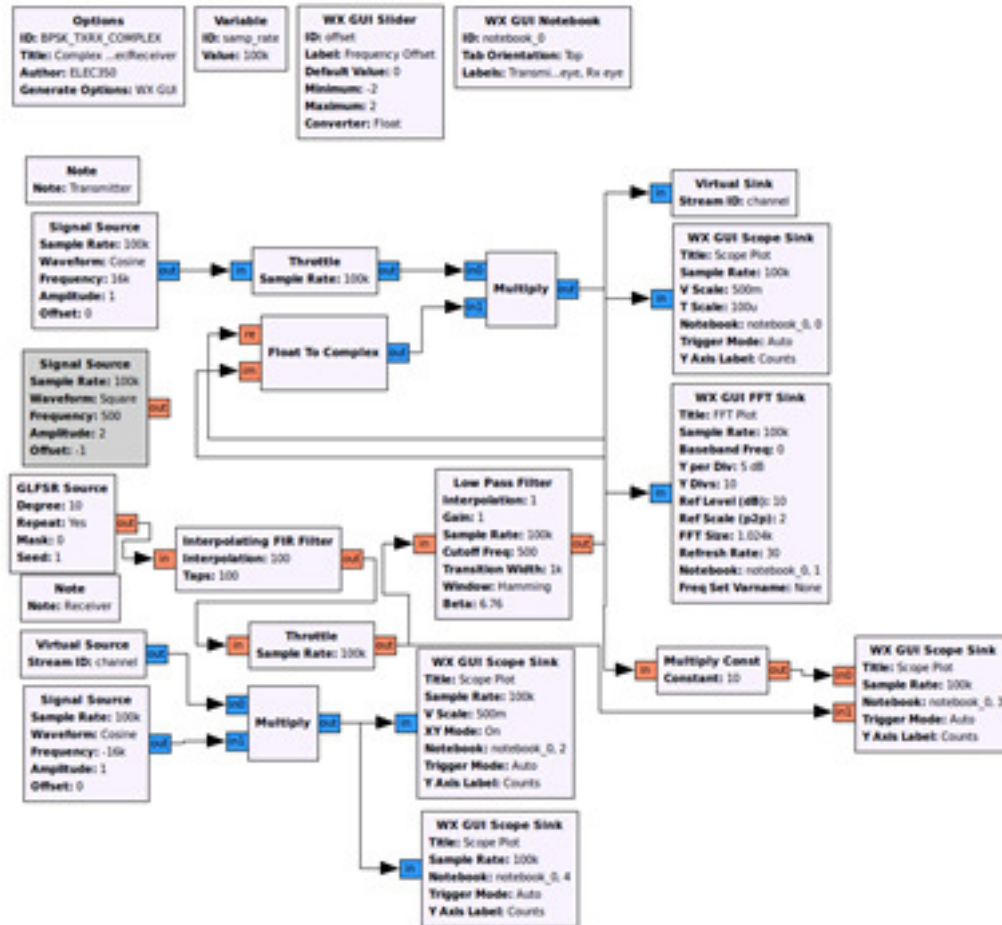
The message waveform will appear as shown before and below, sampled at $f_s T$ samples per symbol.



Thus we can generate time domain raised cosine pulses $p_2(t)$ using either a sequence of square pulses

or a sequence of impulses.

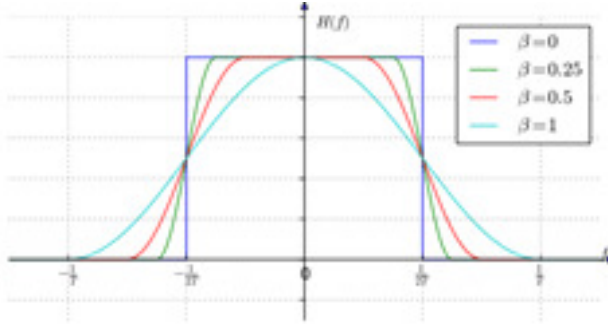
Pulse shaping of a impulses with a pulse shaping filter is shown in the GRC flowgraph below. The sample rate f_s is set to 100K, the symbol rate $1/T=1000$ so that $f_s T=100$.



The linear feedback shift register (GLFSR) data source generates one sample for each data symbol. These samples are filtered by an *interpolating FIR filter* that yields one sample $A_k = \pm 1$ followed by $f_s T - 1 = 99$ samples set to zero, so that the $f_s T = 100$ samples per symbol have only one non-zero value. This interpolating FIR filter will have only one non-zero tap and $f_s T - 1 = 99$ zero taps. The output of this filter is a sequence of impulses at a rate $1/T = 1000$ Hz and sampling rate $f_s = 100K$. This output is filtered by a low pass filter with cutoff 500 Hz and transition width 1000 Hz whose impulse response approximates the time domain RC2 pulse shape.

6.4.4 Raised Cosine filter in frequency domain

$p(t)$ can be chosen to obtain a particular shape in the time domain, as shown in the previous section. However, $p(t)$ can also be chosen to have a particular shape in the frequency domain, i.e. a frequency response $P(f) = H(f)$. One example choice is a raised cosine shape in the frequency domain $H(f) = H_{RRC}(f)$, as per the figure below.



$$H(f) = \begin{cases} T, & |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2} \left[1 + \cos \left(\frac{\pi T}{\beta} \left[|f| - \frac{1-\beta}{2T} \right] \right) \right], & \frac{1-\beta}{2T} < |f| \leq \frac{1+\beta}{2T} \\ 0, & \text{otherwise} \end{cases}$$

The frequency response can be varied with different values of the roll-off factor β and the symbol time T . The roll-off factor is a measure of the excess bandwidth δf

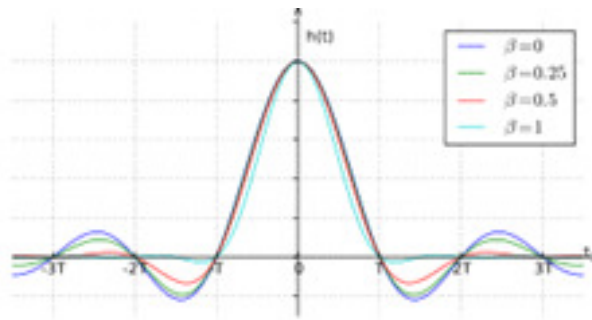
of the pulse shaping filter, the bandwidth beyond the Nyquist bandwidth $1/2T$. Thus β is defined

$$\beta = \frac{\delta f}{1/2T} = 2T\delta f$$

The impulse response $h(t)$ is a sinc shape extending for a time usually truncated to $6T$

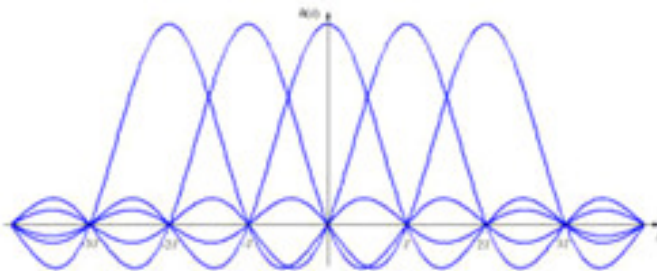
$$h(t) = \text{sinc} \left(\frac{t}{T} \right) \frac{\cos \left(\frac{\pi \beta t}{T} \right)}{1 - \frac{4\beta^2 t^2}{T^2}}$$

as shown in the figure below

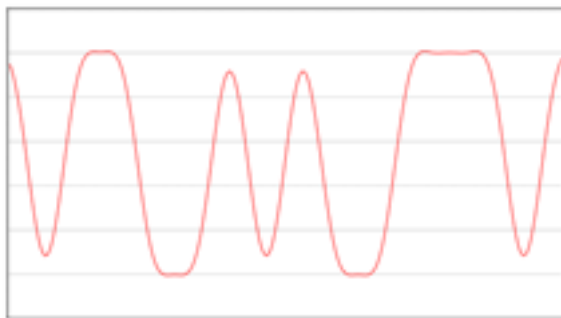


When $\beta = 1$, $H(f) = [1 + \cos(\pi fT)]T/2, |f| \leq 1/T$. Note that this equation has the same form as the time domain pulse $p_2(t) = [1 - \cos(\pi t/T)]/2$ for $t \in \{0, 2T\}$. The sign difference results from the different limits on the variables t and f .

A sequence of these sinc-shaped pulses will overlap, as shown in the figure below for a data sequence 11111.



Because of the sinc shaped pulses, the message waveform will have variable peak amplitude as shown below.



6.4.5 Square Root Raised Cosine filter

A receiver is optimum if the receiver filter matches the transmit filter. Thus we often use a Square Root Raised Cosine (RRC) pulse shaping filter $H_{RRC}(f)$ at the transmitter and another RRC filter at the receiver, so that the combination of the two filters yields a Raised Cosine (RC) pulse shape at the receiver output.

$$H_{RC}(f) = H_{RRC}^2(f)$$

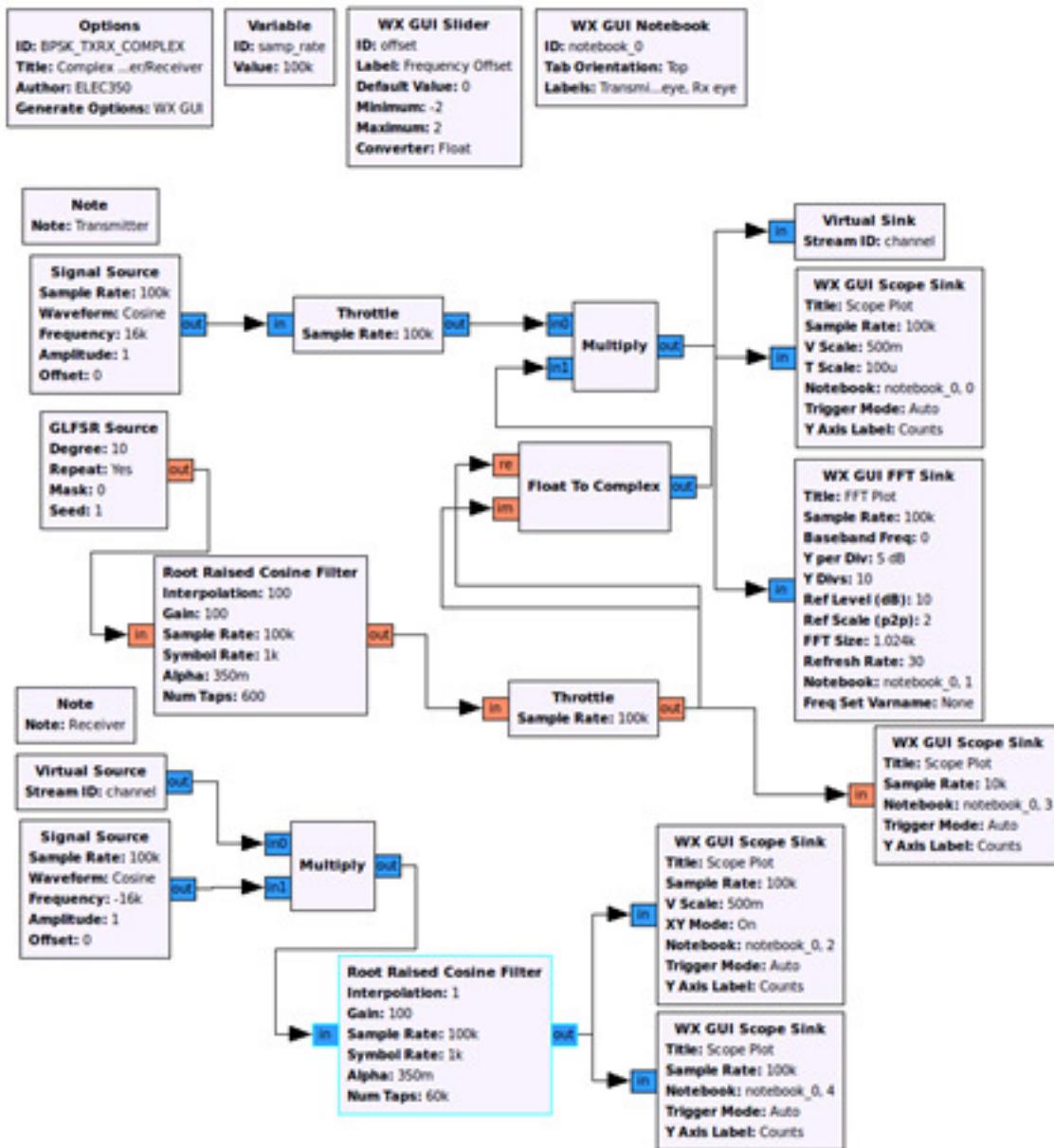
The RRC filter impulse response looks similar to the RC impulse response but with lower sidelobes. The mathematical expression for $h_{RRC}(t)$ is

$$h(t) = \begin{cases} \frac{1}{\sqrt{T_s}} \left(1 - \beta + 4\frac{\beta}{\pi}\right), & t = 0 \\ \frac{\beta}{\sqrt{2T_s}} \left[\left(1 + \frac{2}{\pi}\right) \sin\left(\frac{\pi}{4\beta}\right) + \left(1 - \frac{2}{\pi}\right) \cos\left(\frac{\pi}{4\beta}\right) \right], & t = \pm \frac{T_s}{4\beta} \\ \frac{1}{\sqrt{T_s}} \frac{\sin\left[\pi\frac{t}{T_s}(1-\beta)\right] + 4\beta\frac{t}{T_s} \cos\left[\pi\frac{t}{T_s}(1+\beta)\right]}{\pi\frac{t}{T_s} \left[1 - \left(4\beta\frac{t}{T_s}\right)^2\right]}, & \text{otherwise} \end{cases}$$

Pulse shaping of data using an RRC filter is shown in the GRC flowgraph below.

The RRC blocks does the interpolation by 100 to yield 100 samples per symbol time T.

The sample rate f_s is set to 100K. $f_s T$ is set to 100, so that the symbol rate $1/T=1000$.



6.4.6 Summary of pulse shaping methods

Pulse shaping can be done in multiple ways. Here we consider three methods:

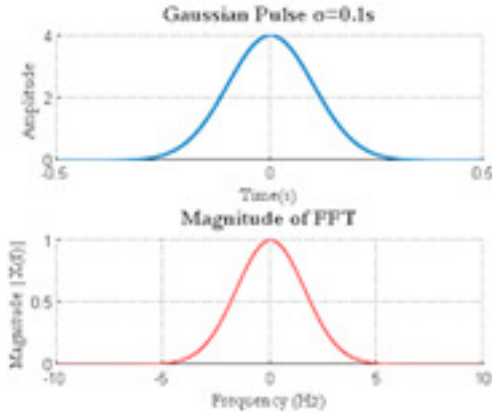
1. (a) i. A. Generate square pulses and low pass filter, commonly used in analog systems.
B. Generate impulses and low pass filter with *time domain* RC FIR filter.

In this case the RC FIR filter $p(t)$ is strictly time limited with zero value outside a time range and $P(f)$ is theoretically has sidelobes spread out at all frequencies. In practice, the sidelobe amplitude versus frequency will be negligible beyond frequencies greater than a few multiples of $1/T$. Thus $P(f)$ will not spread significantly in bandwidth beyond a few multiples of $1/T$.

1. (a) i. A. Generate impulses and filter with *frequency domain* RRC FIR filter.

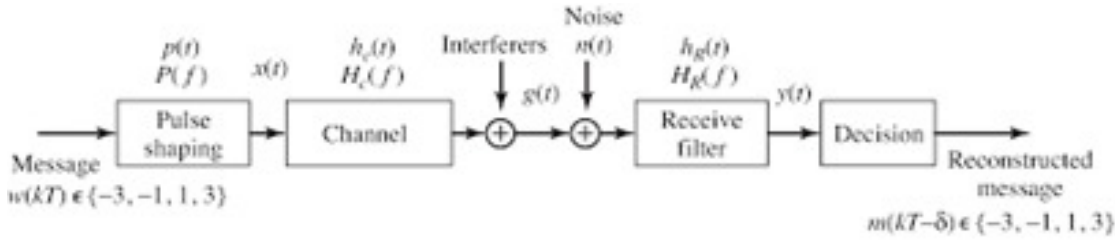
In this case $P(f)$ is strictly band limited with zero value outside a frequency range and $p(t)$ theoretically has sidelobes spread out over all time. In practice, the sidelobe amplitude versus time will be negligible beyond times greater than a few multiples of T . Thus the FIR filter length is chosen so that the sinc-shaped pulse $p(t)$ does not spread in time beyond a few multiples of T . The RRC shape is preferred over the RC shape in the frequency domain so that the receiver RRC filter can be matched to the transmit RRC filter, yielding an overall RC system (transmitter plus receiver) response.

1. (a) i. A. The middle ground between 2 and 3 above is to generate impulses and use a filter with Gaussian shaped impulse response. A Gaussian pulse shape will be Gaussian in both time and frequency domains,



The Gaussian pulse theoretically spreads over all frequencies and all time, since the Gaussian tails extend to $\pm\infty$. In practice, Gaussian pulses will be both time limited and bandwidth limited.

6.5 Eye diagrams

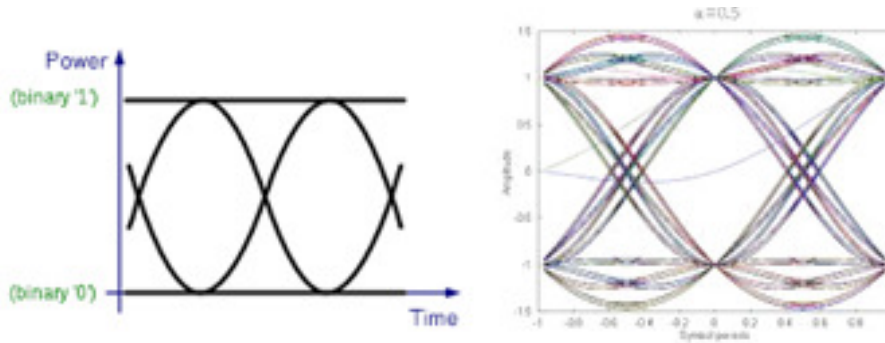


The received message waveform after filtering by both the transmitter pulse shaping filter and the receiver low pass or matched filter is a sequence of pulses that in general will have a variable peak amplitude. The message waveform may be further varied by the channel filter (causing waveform distortion), noise and interference.

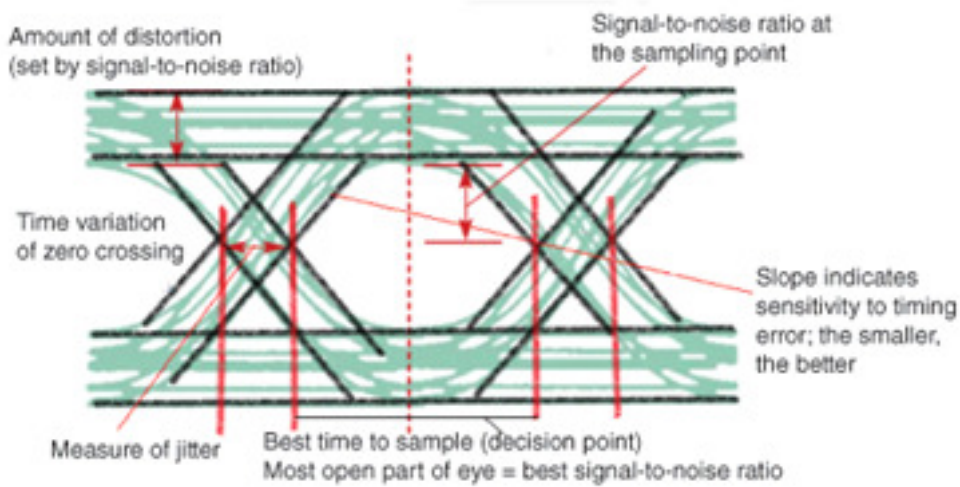
To estimate the probability of a decision error, it is useful to view the message waveform modulo the symbol time T . This can be achieved in practice by triggering the scope on the zero crossings of the message waveform. The resulting picture is a so-called *eye diagram*.

An ideal eye diagram is shown below for the two cases:

1. (a) i. A. time domain RC pulse shape on the left and
 B. the frequency domain RRC pulse shape (with sinc-shaped pulses) on the right.

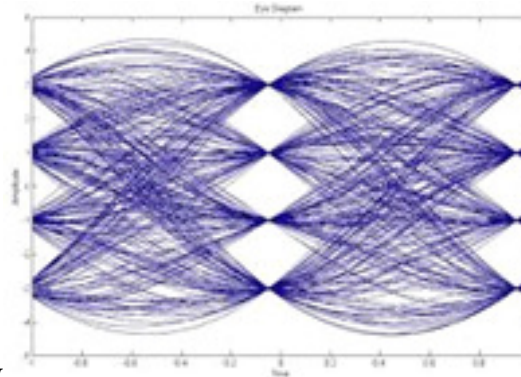


In practice, these eye diagrams are further modified by the channel filter, noise and interference, resulting in a general eye diagram as shown below.

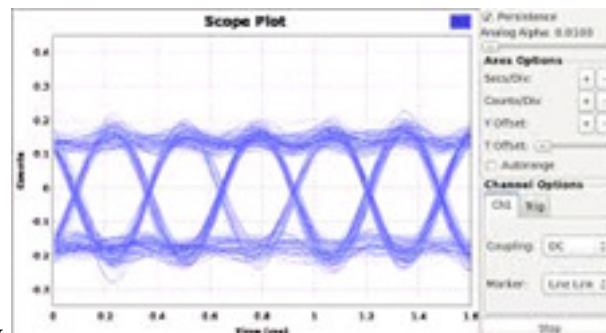


The eye must be open to achieve a low probability of decision error.

The decision device samples this waveform at the time of maximum amplitude for each symbol and then decides what is the value of the symbol (1 or -1 for a binary two level eye diagram above, or



$\pm 1, \pm 3$ for a 4 level eye diagram below



An eye diagram with some noise is shown below.

This completes the discussion of baseband data transmission, digital messages, pulse shaping and

eye diagrams.

Chapter 7

Phase shift keying

Phase shift keying is a digital form of general phase modulation, where the carrier phase is modulated in step with the message waveform. The word “keying” is historical and refers to the on-off keying done by a Morse code key acting as a switch.



For phase modulation, the complex envelope

$$\begin{aligned}\tilde{s}(t) &= a(t)e^{j\phi} = a(t)\cos\phi + ja(t)\sin\phi \\ &= i(t) + jq(t)\end{aligned}$$

has a constant amplitude $a(t) = A_c$ and time varying phase $\phi(t)$.

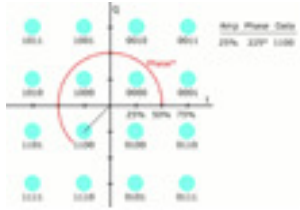
For digital phase modulation, $\phi(t)$ is selected from a finite number of discrete values at the discrete symbol times kT . The discrete values of the complex envelope at the times kT plotted in the 2-D complex plane is called a signal *constellation*.

In general, $\phi(t)$ will assume other values for times t in between the discrete symbol times kT .

Binary phase shift keying (BPSK) uses two discrete phase values, and Quadrature phase shift keying (QPSK) uses four discrete phase values. In general, n -PSK uses n discrete phase values.

In practice, as will be seen later, for BPSK and QPSK with any pulse shape other than a square pulse, both the amplitude and phase will vary with time.

PSK can be combined with amplitude shift keying (ASK) to yield amplitude-phase keying (APK). One form of APK is Quadrature Amplitude Modulation (QAM). One example is 16QAM that uses 16 discrete amplitude and phase values. The signal constellation for 16QAM is shown below. This constellation has 12 discrete phase values and 3 discrete amplitude values. Each point in the constellation represents 4 bits of information.



7.1 BPSK Transmitter

In this section, we show the formulas for a BPSK signal derived from the general complex signal notation. We also show how BPSK is a special case of DSB-SC.

A general signal

$$s(t) = \text{Re}[a(t)e^{j\phi}e^{j2\pi f_c t}] = \text{Re}[\tilde{s}(t)e^{j2\pi f_c t}],$$

where the complex envelope

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi} = a(t)\cos\phi + ja(t)\sin\phi \\ &= i(t) + jq(t) \end{aligned}$$

and $i(t), q(t)$ are obtained from the message signal.

Recall that for DSB-SC, we choose $i(t) = m(t)$

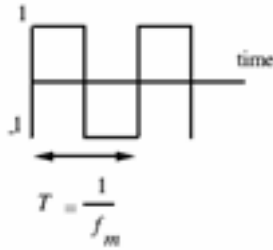
and $q(t) = 0$, so that

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} = \text{Re}\{[i(t) + jq(t)]e^{j2\pi f_c t}\} \\ &= \text{Re}\{m(t)e^{j2\pi f_c t}\} = m(t)\cos 2\pi f_c t \end{aligned}$$

For DSB-SC, $\tilde{s}(t) = m(t)$, i.e. the complex envelope is equal to the (real) message.

We can choose $m(t) = \cos 2\pi f_m t$ to represent a 10 alternating digital message during one cycle of the cos wave, representing a 1 bit when $m(t) = 1$ and a 0 bit when $m(t) = -1$. Thus $f_m = 1/2T$ and T is the symbol time.

We could also choose $m(t)$ to be a square wave with period $2T$. The frequency of the cos (or square) wave is one-half the symbol rate.

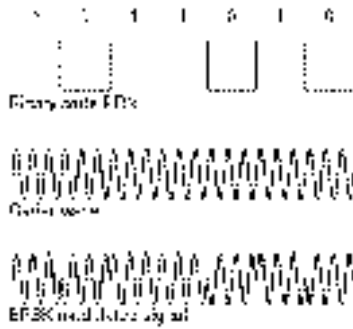


For binary phase shift keying (BPSK), we also choose $i(t) = m(t)$ and $q(t) = 0$ and we choose

$$m(t) = A_0p(t) + A_1p(t - T) + A_2p(t - 2T) + \dots = \sum_k A_k p(t - kT)$$

If $p(t)$ is chosen to be a square pulse and $A_i = +1, i$ even, $-1, i$ odd then $m(t)$ is a square wave with period $2T$.

Thus BPSK is created in the same way as DSB-SC, i.e. multiply the message by the carrier.



The BPSK signal is written

$$s(t) = \text{Re}[m(t)e^{j2\pi f_c t}] = m(t) \cos 2\pi f_c t = \sum_k A_k p(t - kT) \cos 2\pi f_c t$$

The signal constellation for BPSK consists of two points located at $A_k = \pm 1$, one at +1 and one at -1, as shown in Figure 2.

Figure 2 BPSK Constellation

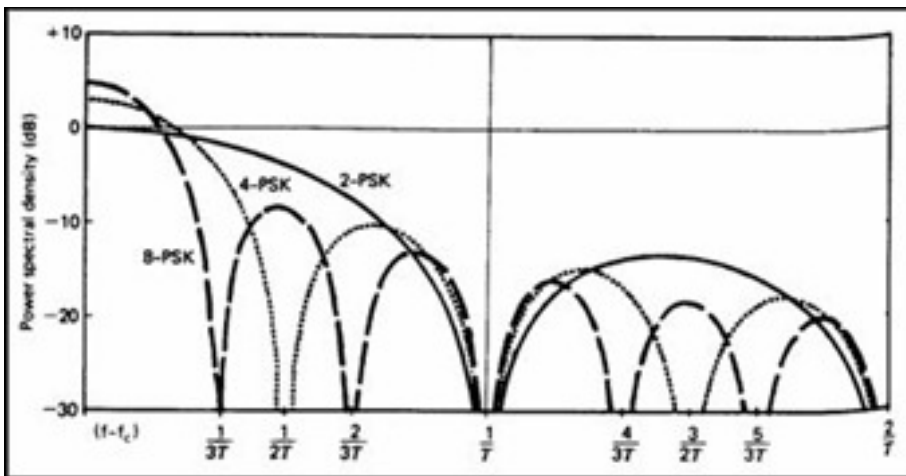
In polar form $A_k = a_k e^{j\phi_k}$, so that with $A_k = \pm 1$, $a_k = 1$ and $\phi_k = n\pi$ are the 2 possible phases of a BPSK signal.

For BPSK, $\tilde{s}(t) = m(t)$, i.e. the complex envelope is equal to the message.

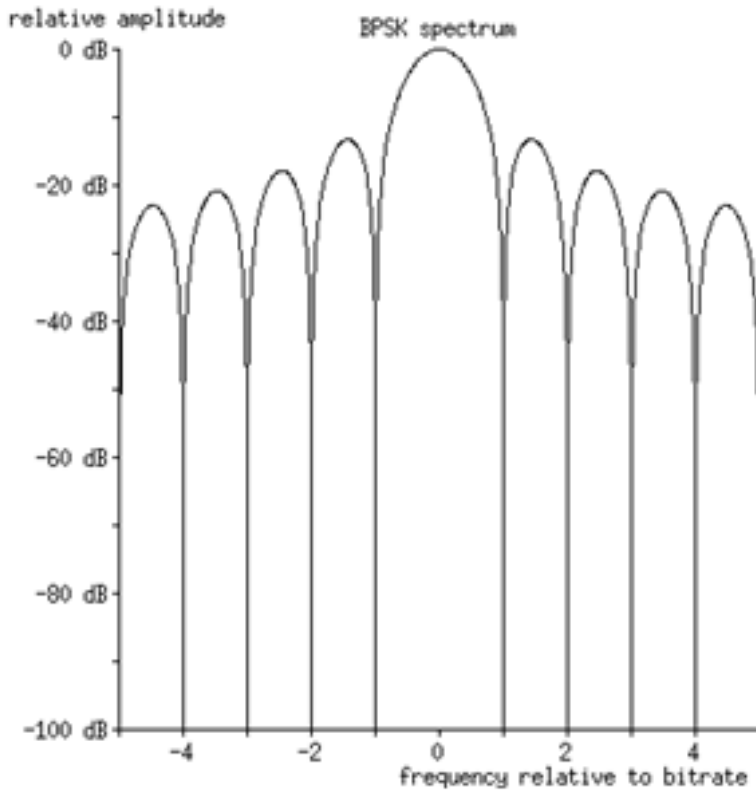
The spectrum of the BPSK signal will depend on the data sequence and pulse shape $i(t) = m(t) = \sum_k A_k p(t - kT)$.

For an alternating sequence $A_k = 1, k \text{ odd}, A_k = -1, k \text{ even}$, with time domain RC2 raised cosine pulse shape $p(t) = p_2(t) = [1 - \cos(\pi t/T)]/2$, $m(t) = \cos \pi t/T = \cos 2\pi f_m t$ where $f_m = 1/2T$, so that in this case, the BPSK spectrum is the same as for DSB-SC with single tone modulation with delta function spikes at $f_c \pm f_m = f_c \pm 1/2T$

In general, the data sequence is random and can be modelled as the output of a linear feedback shift register. The spectrum has a sinc shape as shown below for 2-PSK, where the left axis is the carrier frequency f_c and the spectrum is symmetrical about f_c . The mirror image at frequencies below f_c is not shown.



The full symmetrical spectrum is shown below on a dB scale



7.1.1 Real BPSK signal

To create a real BPSK signal in GRC at intermediate frequency f_1 using these equations, we carry out two steps:

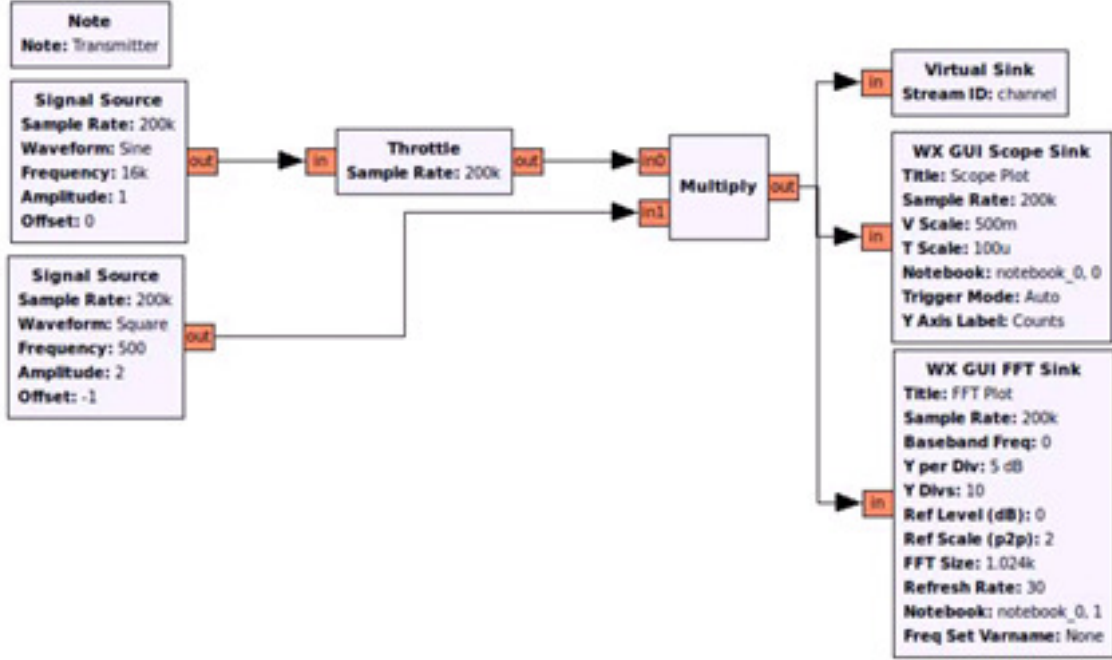
Generate the real baseband signal $i(t) = m(t) = \sum_k A_k p(t - kT)$

1. (a) i. A. Upconvert it to the desired carrier frequency by multiplying by $\cos 2\pi f_1 t = 0.5\{e^{j2\pi f_1 t} + e^{-j2\pi f_1 t}\}$ to yields $s(t) = \text{Re}[m(t)e^{j2\pi f_1 t}] = m(t) \cos 2\pi f_1 t$

$$\begin{array}{c}
 m(t) \rightarrow \otimes \rightarrow s(t) \\
 \uparrow \\
 \cos 2\pi f_1 t
 \end{array}$$

The spectrum of $s(t)$ will be centered at $\pm f_1$ and has both positive and negative frequency components.

A GRC flowgraph for real BPSK is shown below.



The real BPSK signal can also be generated as follows, recalling $i(t) = m(t)$ and $q(t) = 0$

$$i(t) \rightarrow \otimes \Rightarrow \tilde{s}_1(t) \Rightarrow Re \rightarrow i(t) \cos 2\pi f_1 t$$

$$\uparrow$$

$$e^{2\pi f_1 t}$$

The real BPSK signal may be written

$$s(t) = i(t) \cos 2\pi f_1 t$$

$$\text{for } p(t) = 1 \text{ for } 0 \leq t \leq T \text{ (square pulse)}$$

$$s(t) = a_k \cos(2\pi f_1 t + \phi_k) \text{ for } kT \leq t \leq (k+1)T$$

$$= \cos(2\pi f_1 t + n\pi) \text{ for } kT \leq t \leq (k+1)T$$

where the value of n is chosen to represent the phase at time k .

7.1.2 Complex BPSK signal

To create a complex (analytic) BPSK signal in GRC at intermediate frequency f_1 using these equations, we carry out two steps:

1. (a) i. A. Generate the real baseband signal $i(t) = m(t) = \sum_k A_k p(t - kT)$
- B. Upconvert it to the desired carrier frequency by multiplying by $e^{j2\pi f_1 t}$. The

results is a complex (analytic) signal .

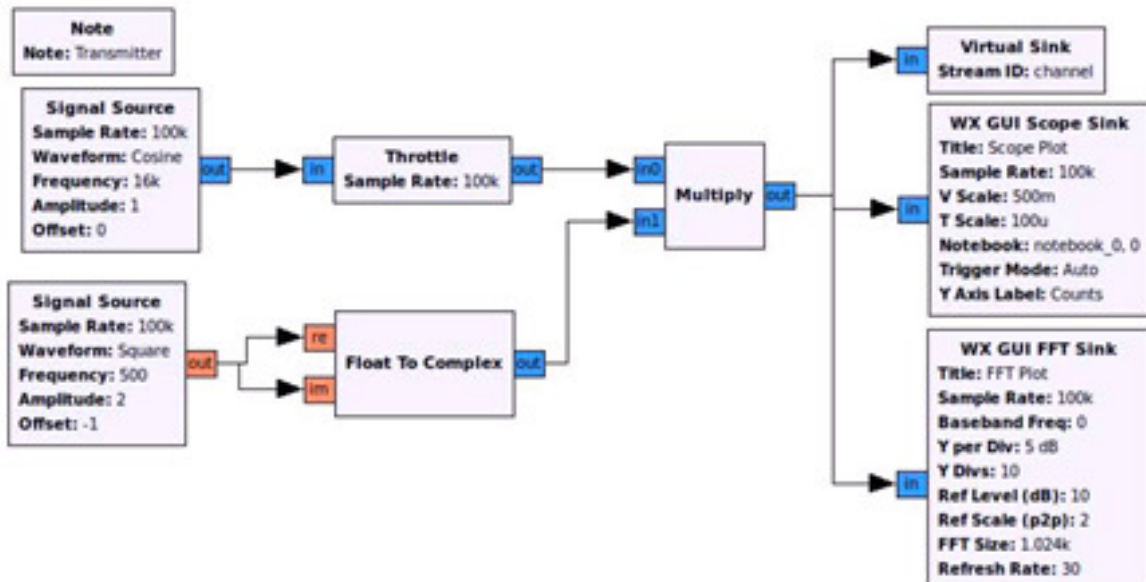
$$\begin{aligned} \tilde{s}_1(t) &= i(t)e^{j2\pi f_1 t} \\ &= i(t) \cos 2\pi f_1 t + j[i(t)\sin 2\pi f_1 t] \\ &= i_1(t) + jq_1(t) \end{aligned}$$

$$m(t) \rightarrow \otimes \begin{matrix} \uparrow \\ e^{j2\pi f_1 t} \end{matrix} \rightarrow \tilde{s}_1(t) = i_1(t) + jq_1(t) = i(t)e^{j2\pi f_1 t}$$

The spectrum of $\tilde{s}_1(t)$ will be centered at f_1 . Since the signal is analytic, there is no negative frequency component at $-f_1$.

A GRC flowgraph is shown below, where the message a square wave 1010 data sequence. In this flowgraph, the message is set to be

$m(t) + jm(t)$ but could also have been set to $m(t) + j0$ by using a null source connected to the lower input of the Float to Complex block.



7.1.3 BPSK at radio frequency (RF)

The USRP sink block with complex input $\tilde{s}_1(t)$ is used to transmit the BPSK signal near a radio frequency (RF) f_c . The USRP transmitter multiplies $\tilde{s}_1(t)$ by $e^{j2\pi f_c t}$ and takes the real part to generate a real RF signal centered at $f_1 + f_c$ and $-f_1 - f_c$.

Recall from above that $i_1(t) = i(t) \cos 2\pi f_1 t$, $q_1(t) = i(t) \sin 2\pi f_1 t$ and $i(t) = m(t) = \sum_k A_k p(t - kT)$

The real BPSK signal at RF is written

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}_1(t)e^{j2\pi f_c t}\} = \text{Re}\{[i_1(t) + jq_1(t)]e^{j2\pi f_c t}\} \\ &= i_1(t) \cos 2\pi f_c t - q_1(t) \sin 2\pi f_c t \\ &= i(t) \cos 2\pi f_1 t \cos 2\pi f_c t - i(t) \sin 2\pi f_1 t \sin 2\pi f_c t \\ &= i(t) \cos 2\pi(f_c + f_1)t \end{aligned}$$

We can get the same result using the exponentials instead of the sine and cosines

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}_1(t)e^{j2\pi f_c t}\} = \text{Re}\{i(t)e^{j2\pi f_1 t}e^{j2\pi f_c t}\} \\ &= \text{Re}\{i(t)e^{j2\pi(f_1+f_c)t}\} = i(t) \cos 2\pi(f_c + f_1)t \end{aligned}$$

BPSK at RF is generated as follows

$$i(t) \rightarrow \otimes \Rightarrow \tilde{s}_1(t) = i_1(t) + jq_1(t) \Rightarrow \otimes \Rightarrow \text{Re} \rightarrow i(t) \cos 2\pi(f_c + f_1)t$$

$$\begin{array}{ccc} & \uparrow & \uparrow \\ & e^{2\pi f_1 t} & e^{2\pi f_c t} \end{array}$$

The USRP sink block does the second complex multiply and taking the real part.

The GRC flowgraph is the same as the complex BPSK above with a USRP sink block in place of the Virtual Sink.

The real RF signal $s(t)$ is centered at $f_1 + f_c$ and $-f_1 - f_c$.

Since the signal is real, there are both positive and negative frequency components.

To transmit the BPSK signal at f_c the USRP sink block could also have used the complex input $\tilde{s}(t) = i(t) = m(t)$ with the imaginary part set to zero.

In this case the USRP transmitter multiplies $\tilde{s}(t)$ by $e^{j2\pi f_c t}$ and takes the real part to generate a real RF signal centered at f_c and $-f_c$. However, this practice is not recommended, since $\tilde{s}(t) = m(t)$ may contain DC, and the USRP does not pass DC.

7.2 BPSK Receiver

To receive the BPSK signal, the USRP source block is used to downconvert the RF signal to complex baseband $i(t) + jq(t)$

We can also simulate the BPSK receiver within GNU Radio without using the USRP source block.

The simulation is done by using the BPSK signal generated at intermediate frequency f_1

7.2.1 Real BPSK signal receiver

We first work with the real BPSK signal $s(t) = \text{Re}[m(t)e^{j2\pi f_1 t}] = m(t) \cos 2\pi f_1 t$

We use a standard IQ receiver set up for f_1 with complex output $i(t) + jq(t)$

$$s(t) \rightarrow \otimes \Rightarrow LPF \Rightarrow i(t) + jq(t)$$

$$\uparrow \tilde{s}_1(t)$$

$$e^{-j2\pi f_1 t}$$

If there is no frequency or phase offset, then $q(t) = 0$

The IQ receiver above is implemented with a complex multiplier. Low pass filters will be required in the complex case, since the input is real.

In GNURadio, the real BPSK signal $s(t)$ is one input of a float to complex block, with the second input set to zero (null source), and the resulting complex signal output (with the imaginary part set to zero) is connected to one input of a complex multiplier.

The IQ receiver can be implemented with real cosines and sines and low pass filters.

With a suitable low pass filter, the received data $i(t) + jq(t)$ at sampling times $t = kT + \tau$ is $A_k = \pm 1$ (real), where τ is adjusted to sample in the middle of the pulse $p(t)$.

τ is a timing offset to sample the data away from the transitions between different bits.

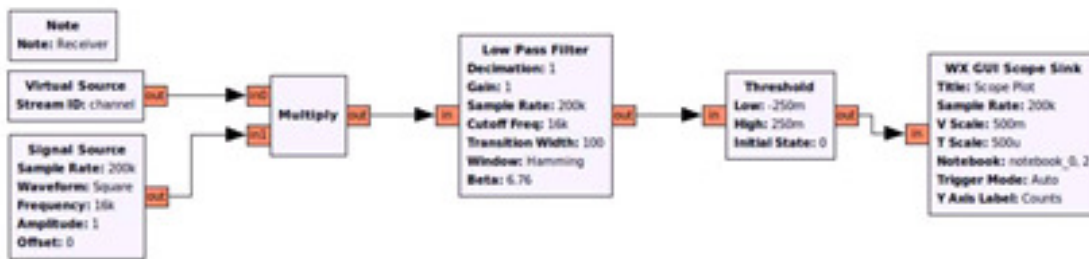
The simplest receiver assumes there is no frequency or phase offset, and can be implemented with a real local oscillator

$$s(t) \rightarrow \otimes \Rightarrow LPF \Rightarrow i(t)$$

$$\uparrow$$

$$\cos 2\pi f_1 t$$

The GRC example implementation shown below uses a square wave local oscillator, but could also have used a cosine wave. There is no reason to choose one or the other. Either will work, since the low pass filter will take out all harmonics of the square wave.



7.2.2 Complex BPSK signal receiver

We can also receive a complex BPSK signal

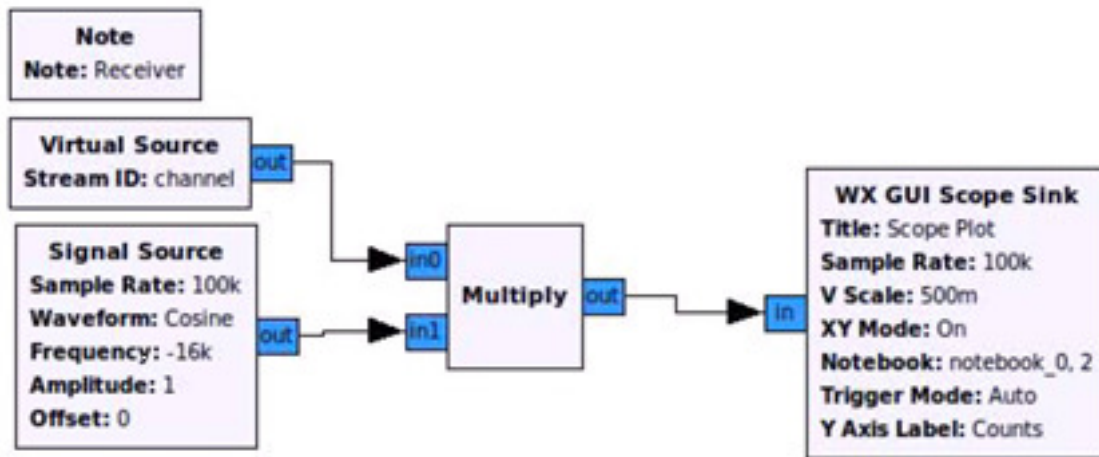
$$\begin{aligned} \tilde{s}_1(t) &= i(t)e^{j2\pi f_1 t} \\ &= i(t) \cos 2\pi f_1 t + j[i(t) \sin 2\pi f_1 t] \\ &= i_1(t) + jq_1(t) \end{aligned}$$

Again, we use a standard IQ receiver with complex local oscillator. In this case, we can use the complex IQ receiver that multiplies $\tilde{s}_1(t)$ by $e^{-j2\pi f_1 t}$ (note the minus sign for downconversion). An LPF is not needed. If there is no frequency or phase offset, then $q(t) = 0$

$$\tilde{s}_1(t) \Rightarrow \otimes \Rightarrow i(t) + jq(t)$$

$$\begin{array}{c} \uparrow \\ e^{-j2\pi f_1 t} \end{array}$$

The GRC flowgraph below uses a complex local oscillator.



7.2.3 BPSK receiver at RF

A BPSK receiver at RF using the USRP would have a GRC flowgraph as above, except that the virtual source is replaced with a USRP source.

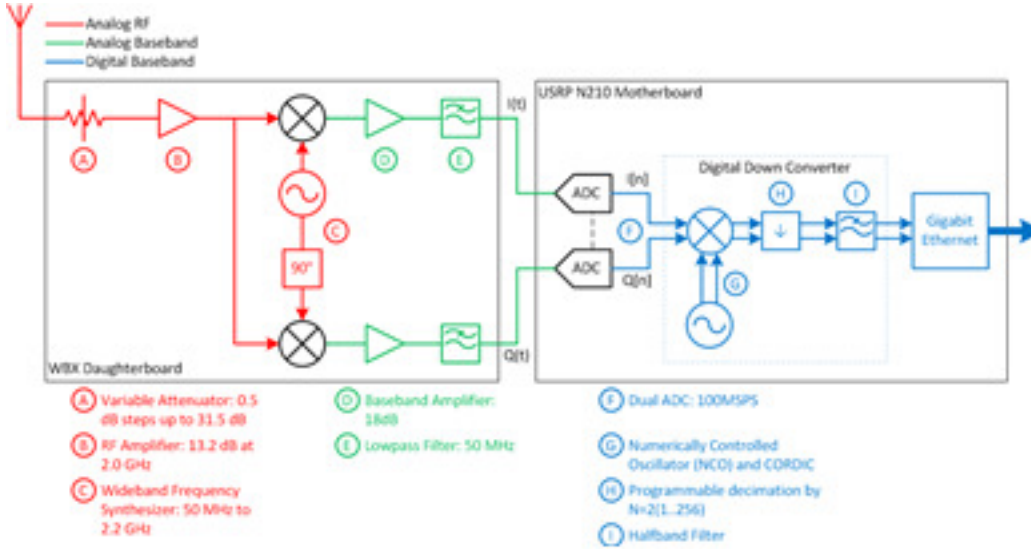
In the diagram below, the USRP source block carries out the complex downconversion with local oscillator at f_c to produce the complex baseband signal $\tilde{s}_1(t)$ that is subsequently processed by the GRC flowgraph.

$$s(t) \rightarrow \otimes \Rightarrow LPF \Rightarrow \tilde{s}_1(t) \Rightarrow \otimes \Rightarrow i(t) + jq(t)$$

$$\begin{array}{c} \uparrow \\ e^{-j2\pi f_c t} \end{array} \qquad \begin{array}{c} \uparrow \\ e^{-j2\pi f_1 t} \end{array}$$

The USRP source block actually does two complex downconversions as shown below.

The Gigabit Ethernet output contains the I and Q samples of $\tilde{s}_1(t)$



7.3 QPSK Transmitter

Recall that we can send separate messages $m_1(t), m_2(t)$ on the I and Q channels, i.e. $i(t) = m_1(t), q(t) = m_2(t)$

We can choose $m_1(t)$ and $m_2(t)$ to be separate digital messages with the same kind of waveform as shown above. When we use both the I and Q channels, the modulation is called Quadrature Phase Shift Keying (QPSK).

To write an expression for QPSK, we can use the same general form as for BPSK above, except that the constants representing the data are now complex instead of real.

A digital message can be written as a complex envelope

$$\tilde{s}(t) = \sum_k C_k p(t - kT) = C_0 p(t) + C_1 p(t - T) + C_2 p(t - 2T) + \dots$$

where $\tilde{s}(t) = i(t) + jq(t)$ is a complex message waveform,

$C_k = A_k + jB_k = a_k e^{j\phi_k}$ is the digital data, where $A_k = a_k \cos \phi_k$ and $B_k = a_k \sin \phi_k$

From the expression for $\tilde{s}(t)$, we see that for the period $0 < t < T$,

$\tilde{s}(t) = A_0 + jB_0 = a_0 e^{j\phi_0}$ is a complex constant. For the next period $T < t < 2T$, $\tilde{s}(t) = A_1 + jB_1 = a_1 e^{j\phi_1}$, etc. In general, for the time period $kT < t < (k+1)T$ $\tilde{s}(t) = A_k + jB_k = a_k e^{j\phi_k}$ is a fixed amplitude and phase.

In rectangular form, $\tilde{s}(t) = i(t) + jq(t)$, where

$$i(t) = \sum_k A_k p(t - kT) = A_0 p(t) + A_1 p(t - T) + A_2 p(t - 2T) + \dots$$

and

$$q(t) = \sum_k B_k p(t - kT) = B_0 p(t) + B_1 p(t - T) + B_2 p(t - 2T) + \dots$$

For QPSK, we choose $A_k = \pm 1$ and $B_k = \pm 1$, so that $C_k = \pm 1 \pm j$.

There are 4 possible values of C_k depending on whether a logic 00, 01, 10 or 11 was sent.

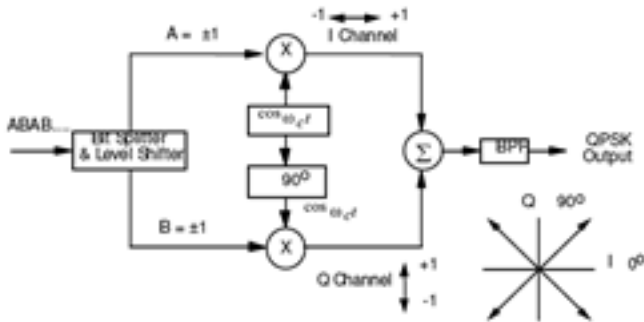
The signal constellation for QPSK is shown in Figure 3 and consists of the 4 points located at $C_k = \pm 1 \pm j$

Figure 3 QPSK Constellation

In polar form $C_k = A_k + jB_k = a_k e^{j\phi_k}$, so that with $A_k = \pm 1$ and $B_k = \pm 1$, $a_k = \sqrt{2}$ and $\phi_k = \frac{(2n-1)\pi}{4}$ are the 4 possible phases of a QPSK signal.

7.3.1 Real QPSK signal

To create a real QPSK signal in GRC at intermediate frequency f_2 using these equations, we generate the complex baseband signal $\tilde{s}(t) = i(t) + jq(t)$ and upconvert it to the desired carrier frequency with a standard (real) IQ transmitter



The result is

$$\begin{aligned}
 s(t) &= i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t \\
 \text{for } p(t) &= 1 \text{ for } 0 \leq t \leq T \text{ (square pulse)} \\
 s(t) &= a_k \cos(2\pi f_2 t + \phi_k) \text{ for } kT \leq t \leq (k+1)T \\
 &= \sqrt{2} \cos(2\pi f_2 t + \frac{(2n-1)\pi}{4}) \text{ for } kT \leq t \leq (k+1)T
 \end{aligned}$$

where the value of n is chosen to represent the phase at time k .

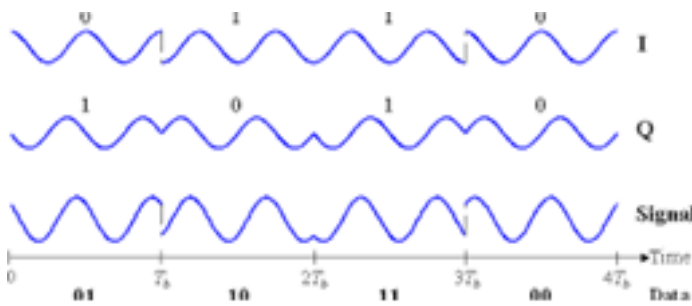
In the figure below, the top waveform is the first term

$$i(t) \cos 2\pi f_2 t$$

the middle waveform is the second term

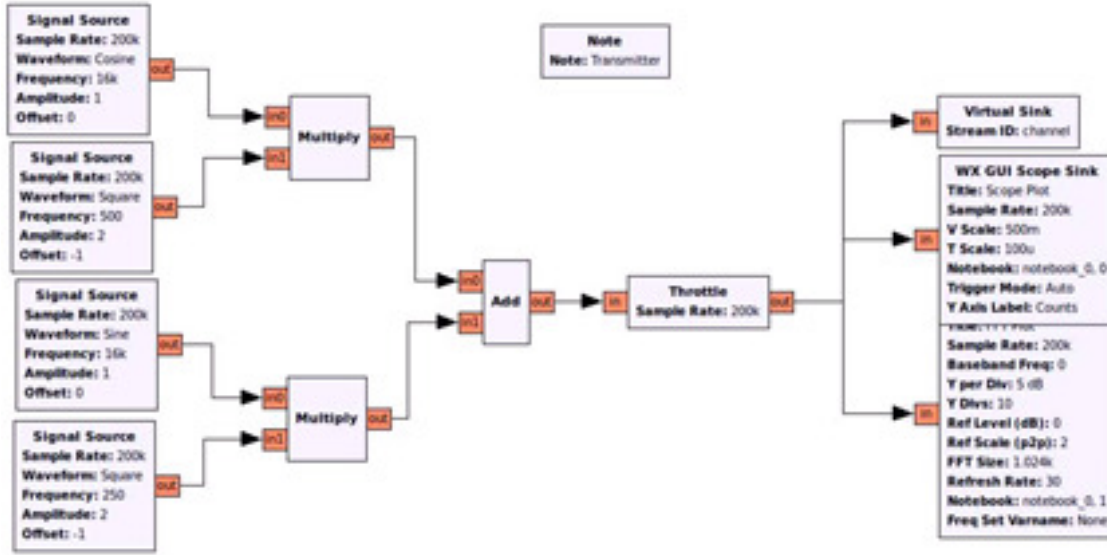
$$-q(t) \sin 2\pi f_2 t$$

and the bottom waveform is the real QPSK signal $s(t)$.



The spectrum of $s(t)$ will be centered at $\pm f_1$ and has both positive and negative frequency components.

A GRC flowgraph of the real QPSK transmitter is shown below



The result at the virtual sink output is

$$s(t) = i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t$$

The real QPSK signal can also be generated using complex notation

$$m_1(t) + jm_2(t) = \tilde{s}(t) = i(t) + jq(t) \Rightarrow \otimes \Rightarrow \tilde{s}_2(t) \Rightarrow Re \rightarrow s(t)$$

$$\uparrow$$

$$e^{j2\pi f_2 t}$$

7.3.2 Complex QPSK signal

To create an complex QPSK signal in GRC at intermediate frequency f_2 using these equations, we create the complex baseband signal $\tilde{s}(t) = i(t) + jq(t)$ and upconvert it to the desired carrier frequency by multiplying by $e^{j2\pi f_2 t}$. The result is

another complex signal

$$\begin{aligned} \tilde{s}_2(t) &= [i(t) + jq(t)]e^{j2\pi f_2 t} \\ &= i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t + j[q(t) \cos 2\pi f_2 t + i(t) \sin 2\pi f_2 t] \\ &= i_2(t) + jq_2(t) \end{aligned}$$

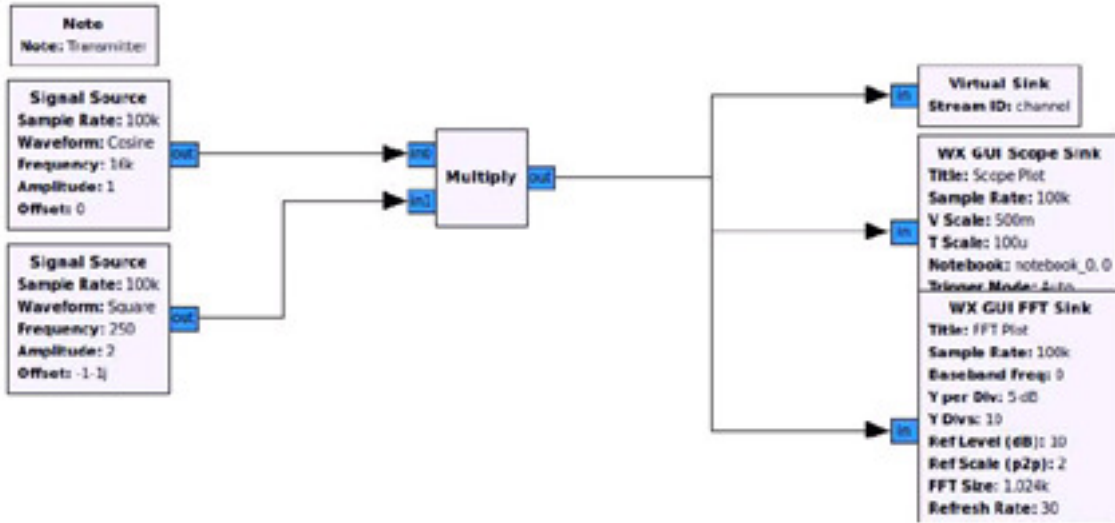
The complex QPSK signal is generated as follows

$$m_1(t) + jm_2(t) = \tilde{s}(t) = i(t) + jq(t) \Rightarrow \otimes \Rightarrow \tilde{s}_2(t)$$

$$\uparrow$$

$$e^{j2\pi f_2 t}$$

A GRC flowgraph is shown below. The complex message $m_1(t) + jm_2(t)$ is multiplied by a complex local oscillator $e^{j2\pi f_2 t}$



We can create a second QPSK signal at a different intermediate frequency f_3 in the same way to yield

$$\tilde{s}_3(t) = i_3(t) + jq_3(t)$$

We can create a composite signal consisting of several QPSK signals at f_2, f_3 etc.

7.3.3 QPSK at radio frequency (RF)

The upconversion of a QPSK signal to a radio frequency (RF) wave at f_c is the function of the USRP sink block (standard IQ transmitter) with inputs $i_2(t)$ and $q_2(t)$.

The USRP transmitter accepts a complex baseband signal $\tilde{s}_2(t)$ and multiplies it by $e^{j2\pi f_c t}$ and takes the real part to generate a real RF signal $s(t)$.

$$\tilde{s}(t) \Rightarrow \otimes \Rightarrow \tilde{s}_2(t) \Rightarrow \otimes \Rightarrow Re \rightarrow s(t)$$

$$\begin{matrix} \uparrow & & \uparrow \\ e^{j2\pi f_2 t} & & e^{j2\pi f_c t} \end{matrix}$$

$$s(t) = i(t) \cos 2\pi(f_c + f_2)t - q(t) \sin 2\pi(f_c + f_2)t$$

Recall $m_1(t) + jm_2(t) = \tilde{s}(t) = i(t) + jq(t)$ and $\tilde{s}_2(t) = i_2(t) + jq_2(t)$

The USRP sink block does the second complex multiply and takes the real part.

The GRC flowgraph is the same as the complex QPSK above with a USRP sink block in place of the Virtual Sink.

Compare this to the BPSK case, where $q(t) = 0$

$$i(t) \rightarrow \otimes \Rightarrow \tilde{s}_1(t) \Rightarrow \otimes \Rightarrow Re \rightarrow i(t) \cos 2\pi(f_c + f_1)t$$

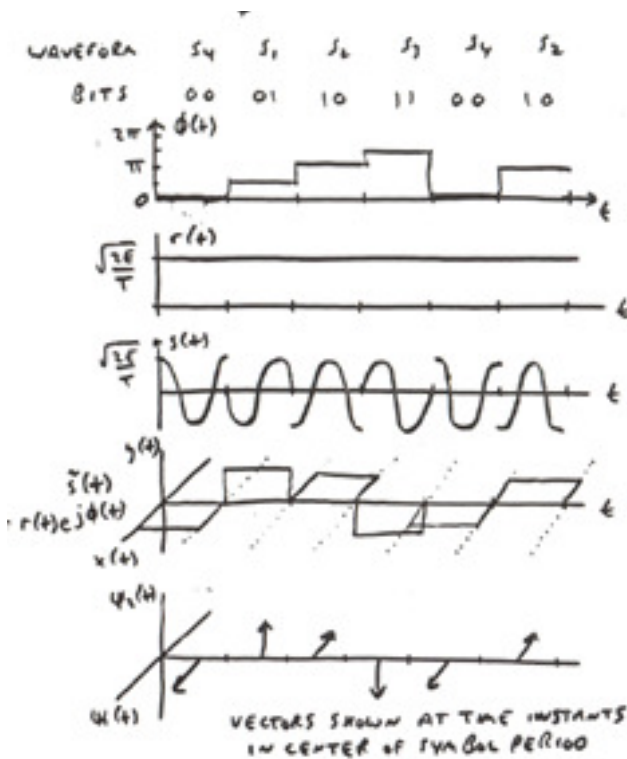
$$\begin{matrix} \uparrow & & \uparrow \\ e^{2\pi f_1 t} & & e^{2\pi f_c t} \end{matrix}$$

Similarly, the entire group of QPSK signals $\tilde{s}_2(t) + \tilde{s}_3(t)$ can be upconverted to RF by the USRP sink block.

The USRP sink block could also have used the complex input $\tilde{s}(t) = i(t) + jq(t)$

to transmit the QPSK signal at f_c . In this case the USRP transmitter multiplies $\tilde{s}(t)$ by $e^{j2\pi f_c t}$ and takes the real part to generate a real RF signal centered at f_c and $-f_c$. However, this practice is not recommended, since $\tilde{s}(t) = i(t) + jq(t)$ may contain DC, and the USRP does not pass DC.

The waveforms are represented as shown below in both complex baseband and real passband.



7.4 QPSK Receiver

To receive the QPSK signal, the USRP source block is used to downconvert the RF signal to complex baseband $i(t) + jq(t)$

We can also simulate the QPSK receiver within GNU Radio without using the USRP source block.

The simulation is done by using the QPSK signal generated at intermediate frequency f_2

7.4.1 Real QPSK signal

We can receive the real QPSK signal

$$s(t) = i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t$$

We use a standard IQ receiver set up for f_2 with complex output $i(t) + jq(t)$

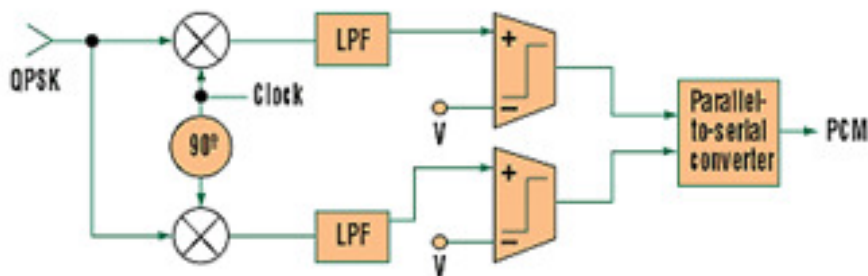
This IQ receiver can be implemented with real cosines and sines and low pass filters.

With a suitable low pass filter, the received data $i(t)$ at sampling times $t = kT +$ is $A_k = \pm 1$ and the received data $q(t)$ is $B_k = \pm 1$, where t is adjusted to sample in the middle of the pulse $p(t)$ The received data $i(t) + jq(t)$ at sampling times $t = kT +$ is $C_k = \pm 1 \pm j$

$$s(t) \rightarrow \otimes \begin{matrix} \uparrow \tilde{s}_2(t) \\ e^{-j2\pi f_2 t} \end{matrix} \Rightarrow LPF \Rightarrow i(t) + jq(t)$$

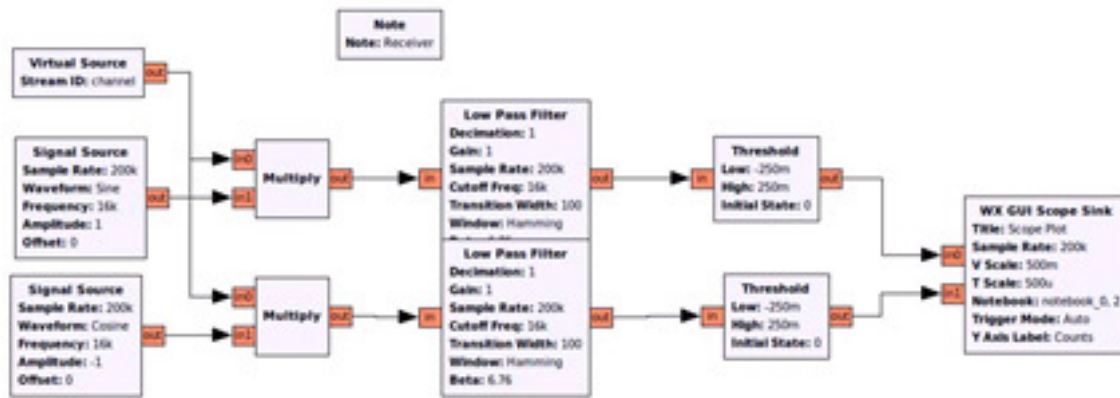
The IQ receiver above is implemented with a complex multiplier. Low pass filters will be required in the complex case, since the input is real.

The IQ receiver can be also implemented with real cosines and sines and low pass filters.



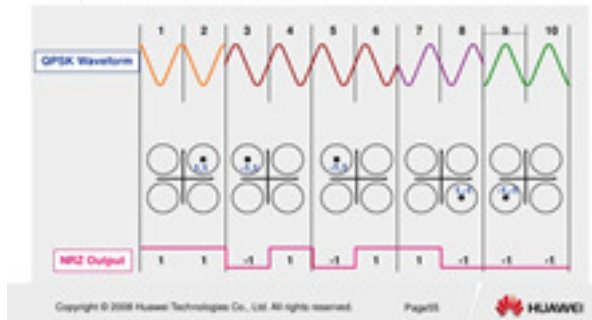
1. Typically, QPSK signals are demodulated using the approach shown here. It produces terms at double the oscillator's frequency, thus the presence of a low-pass filter.

A GRC flowgraph of the real QPSK receiver is shown below.



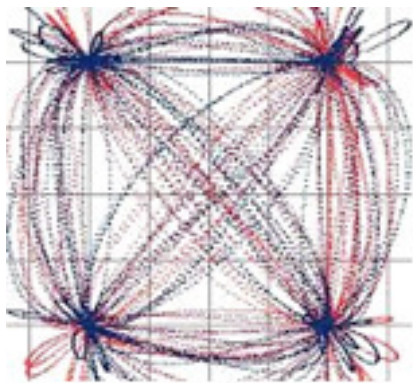
Demodulation

• QPSK Constellation Diagram



An XY scope plot will show the signal constellation. At sampling times $t = kT +$ the constellation shows $C_k = \pm 1 \pm j$

If the signal is not sampled and we display the received data $i(t) + jq(t)$ at all times t , then the signal constellation shows the paths between the 4 points $C_k = \pm 1 \pm j$



The exact trajectories (shape of the path curves) will depend on the pulse shaping filter.

7.4.2 Complex QPSK signal

We can also receive the complex QPSK signal

$$\begin{aligned}\tilde{s}_2(t) &= [i(t) + jq(t)]e^{j2\pi f_2 t} \\ &= i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t + j[q(t) \cos 2\pi f_2 t + i(t) \sin 2\pi f_2 t] \\ &= i_2(t) + jq_2(t)\end{aligned}$$

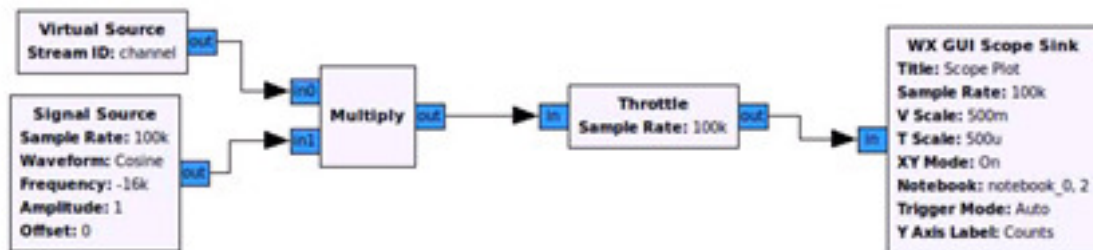
Again, we use a standard IQ receiver. In this case, we can use the complex IQ receiver that multiplies $\tilde{s}_2(t)$ by $e^{-j2\pi f_2 t}$ (note the minus sign for downconversion). An LPF is not needed.

$$\tilde{s}_2(t) \Rightarrow \otimes \Rightarrow \tilde{s}_4(t) = i(t) + jq(t)$$

$$\begin{array}{c} \uparrow \\ e^{-j2\pi f_2 t} \end{array}$$

$$\begin{aligned}\tilde{s}_4(t) &= i_4(t) + jq_4(t) = \tilde{s}_2(t)e^{-j2\pi f_2 t} = [i(t) + jq(t)]e^{j2\pi f_2 t}e^{-j2\pi f_2 t} \\ &= i(t) + jq(t)\end{aligned}$$

A GRC flowgraph is shown below. The XY scope mode is used to show the signal constellation. A practical receiver will include a low pass or RRC matched filter.



The received data $i(t) + jq(t)$ at sampling times $t = kT +$ is $C_k = \pm 1 \pm j$

7.4.3 QPSK signal at radio frequency (RF)

A BPSK receiver at RF using the USRP would have a GRC flowgraph as above, except that the virtual source is replaced with a USRP source.

The USRP receiver multiplies the real valued radio frequency signals $s(t)$ by $e^{j2\pi f_2 t}$ to generate $\tilde{s}_2(t)$

In the diagram below, the USRP source block carries out the complex downconversion with local oscillator at f_c and low pass filters to produce the complex baseband signal $\tilde{s}_2(t)$ that is subsequently processed by the GRC flowgraph.

$$s(t) \rightarrow \otimes \Rightarrow LPF \Rightarrow \tilde{s}_2(t) \Rightarrow \otimes \Rightarrow i(t) + jq(t)$$

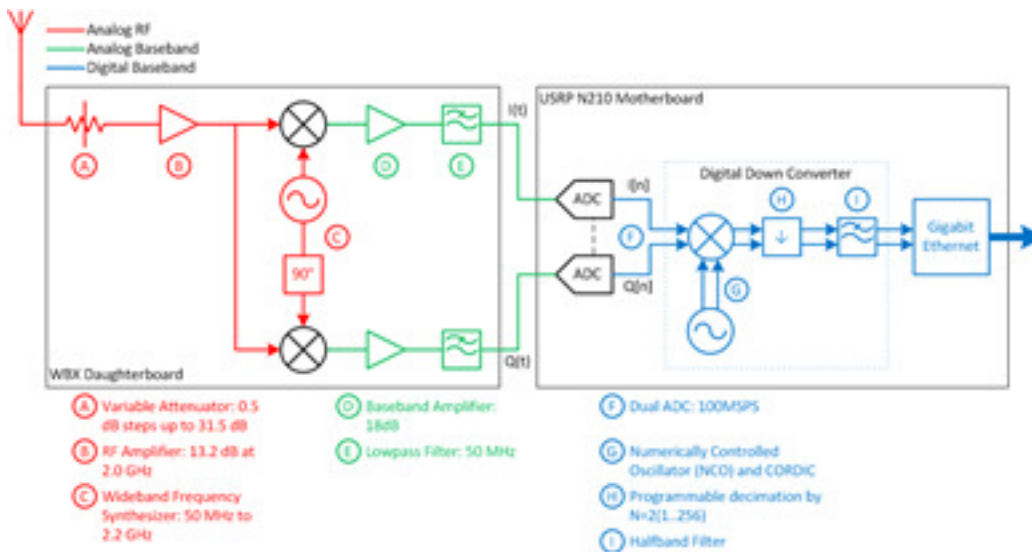
$$\uparrow \qquad \qquad \qquad \uparrow$$

$$e^{-j2\pi f_c t} \qquad \qquad \qquad e^{-j2\pi f_2 t}$$

The USRP source block has a complex output $\tilde{s}_2(t)$

The USRP source block actually does two complex downconversions as shown below.

The Gigabit Ethernet output contains the I and Q samples of $\tilde{s}_2(t)$



The complex signal output from the USRP source block $\tilde{s}_2(t)$ is bandlimited to the sampling rate of the USRP source block. The USRP source block output can be recorded to a file and used again at a later time. This file source will have the same sampling rate and bandwidth as the USRP source block used to record it.

With a sampling rate of 200 kHz and complex samples, the bandwidth will be 200 kHz (because the complex signal spectrum is not symmetric and does not have redundant mirror-image positive and negative frequencies)

We can “tune in” (receive) any of the QPSK signals in this bandwidth by shifting the spectrum of the USRP source output using GRC. We shift by f_2 Hz by multiplying the USRP source output (complex signal) $\tilde{s}_2(t)$ by $e^{-j2\pi f_2 t} = \cos 2\pi f_2 t - j \sin 2\pi f_2 t$ to produce $i(t) + jq(t)$, so that the signal that first appeared at f_2 Hz now appears at zero Hz.

With a suitable low pass filter, the received data $i(t) + jq(t)$ at sampling times $t = kT +$ is $C_k = \pm 1 \pm j$, where $p(t)$ is adjusted to sample in the middle of the pulse $p(t)$

7.5 QPSK with amplitude, frequency and phase errors in the receiver

7.5.1 QPSK signal at particular time instant

For a real QPSK signal for a particular time instant $t = kT +$ we can write

$$s(t) = A_k \cos 2\pi f_c t + B_k \sin 2\pi f_c t$$

where $A_k, B_k = \pm 1$

Proof of why we can write this:

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

From PSK theory: A digital message can be written as a complex envelope

$$\tilde{s}(t) = \sum_k C_k p(t - kT) = C_0 p(t) + C_1 p(t - T) + C_2 p(t - 2T) + \dots$$

where

$\tilde{s}(t) = i(t) + jq(t)$ is a complex message waveform,

$C_k = A_k + jB_k = a_k e^{j\phi_k}$ is the digital data, where $A_k = a_k \cos \phi_k$ and $B_k = a_k \sin \phi_k$

At sampling time $t = mT +$,

$$\begin{aligned} \tilde{s}(t = mT +) &= \sum_k C_k p((m - k)T +) \\ &= C_0 p(mT +) + C_1 p((m - 1)T +) + C_2 p((m - 2)T +) + \dots \\ &= C_{m-1} p(-T) + C_m p(0) + C_{m+1} p(+T) \\ &= C_m \end{aligned}$$

Only one term survives since $p(t) \neq 0$ only for $t \simeq$

Changing variables $m \rightarrow k$ we can write at time $t = kT +$

$$\begin{aligned} \tilde{s}(t) &= C_k = A_k + jB_k \\ &= i(t) + jq(t)|_{t=kT+} \\ i(t)|_{t=kT+} &= i(t = kT +) = A_k \\ q(t)|_{t=kT+} &= q(t = kT +) = B_k \end{aligned}$$

$$\begin{aligned}
s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \\
&= \text{Re}\{(A_k + jB_k)(\cos 2\pi f_c t + j \sin 2\pi f_c t)\} \\
&= A_k \cos 2\pi f_c t - B_k \sin 2\pi f_c t \\
&= i(t) \cos 2\pi f_c t - q(t) \sin 2\pi f_c t|_{t=kT+}
\end{aligned}$$

In polar form $\tilde{s}(t) = C_k = A_k + jB_k = a_k e^{j\phi_k}$, so that with $A_k = \pm 1$ and $B_k = \pm 1$, $a_k = \sqrt{2}$ and $\phi_k = \frac{(2n-1)\pi}{4}$ are the 4 possible phases of a QPSK signal.

7.5.2 QPSK receiver local oscillator with frequency error

In this example, we show what happens when the receiver local oscillator $I_{LO}(t) - jQ_{LO}(t) = e^{-j2\pi f_{LO} t}$ runs at $f_{LO} = f_c + \Delta f$ instead of f_c

We can consider this for the analog IQ local oscillator in the USRP. However, this is not normally done because the USRP cannot pass the DC that arises from the complex data symbols. $\tilde{s}(t) = C_k = A_k + jB_k = i(t) + jq(t)|_{t=kT+}$

Here we consider a frequency error within GRC where the complex baseband signal is available from the USRP source block.

We assume that the USRP source block output is a complex QPSK signal with message waveform $\tilde{s}(t) = i(t) + jq(t)$ and data symbols $\tilde{s}(t) = C_k = A_k + jB_k = i(t) + jq(t)|_{t=kT} = \pm 1 \pm j$

The complex baseband QPSK signal from the USRP source block output appears on a carrier frequency f_2 as shown in the complex QPSK transmitter section above

$$\begin{aligned}
\tilde{s}_2(t) &= [i(t) + jq(t)]e^{j2\pi f_2 t} \\
&= i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t + j[q(t) \cos 2\pi f_2 t + i(t) \sin 2\pi f_2 t] \\
&= i_2(t) + jq_2(t)
\end{aligned}$$

We use the complex IQ receiver that multiplies $\tilde{s}_2(t)$ by $e^{-j2\pi f_2 t}$ (note the minus sign for downconversion). An LPF is not needed. However, in this case we consider a frequency error (or tuning error) Δf in the local oscillator, so that we multiply $\tilde{s}_2(t)$ by $e^{-j2\pi(f_2 + \Delta f)t}$

$$\begin{aligned}
\tilde{s}_2(t) &\Rightarrow \otimes \Rightarrow \tilde{s}_4(t) = i_4(t) + jq_4(t) \\
&\quad \uparrow \\
&\quad e^{-j2\pi(f_2 + \Delta f)t}
\end{aligned}$$

$$\begin{aligned}
\tilde{s}_4(t) &= i_4(t) + jq_4(t) = \tilde{s}_2(t)e^{-j2\pi f_2 t} = [i(t) + jq(t)]e^{j2\pi f_2 t}e^{-j2\pi(f_2 + \Delta f)t} \\
&= [i(t) + jq(t)]e^{-j2\pi \Delta f t}
\end{aligned}$$

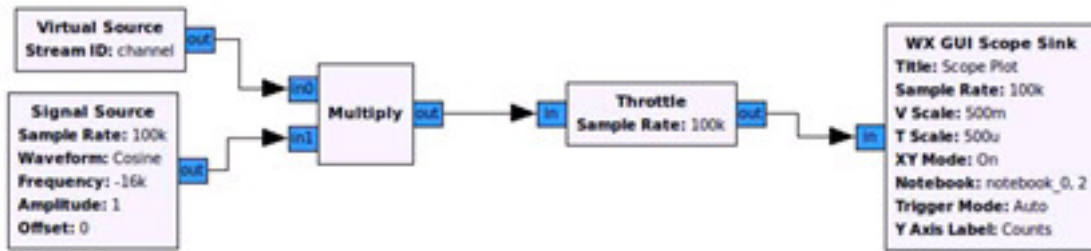
At the sampling time $t = kT +$

$$\tilde{s}_4(t) = [i(t) + jq(t)]e^{-j2\pi\Delta ft}|_{t=kT+} = [A_k + jB_k]e^{-j2\pi\Delta f(kT+)}$$

in polar form

$$\tilde{s}_4(t) = C_k e^{-j2\pi\Delta ft} = a_k e^{j\phi_k} e^{-j2\pi\Delta ft} = a_k e^{j(\phi_k - 2\pi\Delta ft)}|_{t=kT+}$$

A GRC flowgraph is shown below. In this example, the frequency is set to -16K, and is connected to a GUI slider to test the tuning error.



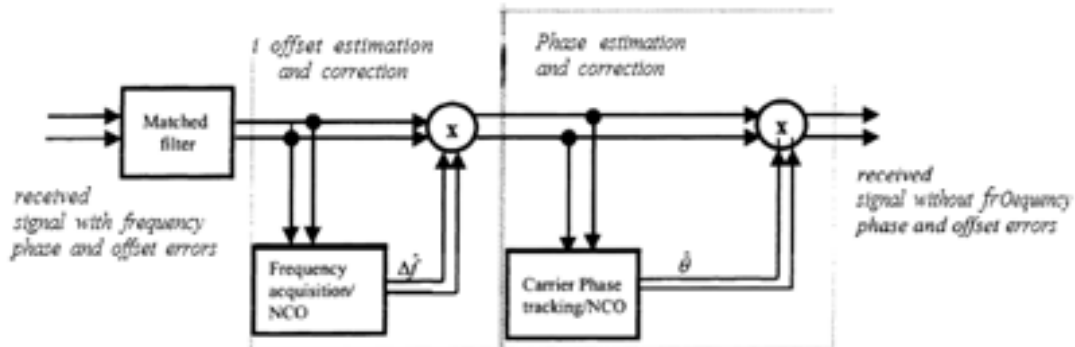
For perfect tuning with $\Delta f = 0$, the received data $\tilde{s}_4(t)$ at sampling times $t = kT +$ is $C_k = \pm 1 \pm j$

With a tuning error, the transmitted phase ϕ_k is received as $\phi_k - 2\pi\Delta ft|_{t=kT+} = \phi_k - 2\pi\Delta f(kT +)$

The phase rotates with time at a rate $2\pi\Delta f$ radians per second. This rotation can be observed by moving the GRC slider. An animated GIF would be useful here.

In order to receive the correct data symbols $C_k = \pm 1 \pm j$ it is necessary to adjust the tuning error to zero. This can be done manually with the slider. In a practical receiver, a frequency offset correction (or phase tracking) circuit is employed, also called *carrier acquisition* and *carrier tracking*.

An example of such a circuit is shown below. The operation of this circuit is beyond the scope of the present work.



7.5.3 QPSK receiver local oscillator with amplitude and phase errors

The receiver local oscillator can be written $I_{LO}(t) - jQ_{LO}(t) = e^{-j2\pi f_{LO}t}$ if it is perfect. However, for this case, assume it has the correct frequency $f_{LO} = f_c$, but has amplitude and phase errors, so that we can write

$$\begin{aligned} I_{LO}(t) &= (1 + \delta/2) \cos(2\pi f_{LO}t - \theta/2) \\ Q_{LO}(t) &= (1 - \delta/2) \sin(2\pi f_{LO}t + \theta/2) \end{aligned}$$

This type of error is likely to occur in the analog complex downmixer. It will not occur unintentionally in GRC. The effect of these errors can be modelled and tested in GRC.

Exercise:

Draw the constellation diagram of the received signals assuming that $\theta = 30^\circ$, $\delta = 0.2$ and compare to the ideal constellation with $\theta = 0^\circ$, $\delta = 0$. Show all calculations needed.

Implement this receiver in GNURadio GRC and include the flowgraph.

This problem can be solved using either real or complex notation, show the signals both ways.

Real notation, find output of I and Q branches of real QPSK receiver and plot for different values of the data $A_k = \pm 1$, $B_k = \pm 1$

$$\begin{aligned} s(t) &= A_k \cos 2\pi f_c t - B_k \sin 2\pi f_c t = i(t) \cos 2\pi f_c t - jq(t) \sin 2\pi f_c t \\ \text{at sampling time } t &= kT, i(t) = A_k, q(t) = B_k \end{aligned}$$

$$\begin{aligned} I_{LO}(t) &= (1 + \delta/2) \cos(2\pi f_{LO}t - \theta/2) \\ Q_{LO}(t) &= (1 - \delta/2) \sin(2\pi f_{LO}t + \theta/2) \end{aligned}$$

$$\begin{aligned} I(t) &= s(t)I_{LO}(t) = [A_k \cos 2\pi f_c t - B_k \sin 2\pi f_c t] \cdot (1 + \delta/2) \cos(2\pi f_{LO}t - \theta/2) \\ &= A_k(1 + \delta/2) \cos 2\pi f_c t \cos(2\pi f_{LO}t - \theta/2) - B_k(1 + \delta/2) \sin 2\pi f_c t \cos(2\pi f_{LO}t - \theta/2) \\ &= 0.5A_k(1 + \delta/2) \cos(\theta/2) - 0.5B_k(1 + \delta/2) \sin(\theta/2) \end{aligned}$$

$$\begin{aligned} Q(t) &= s(t)Q_{LO}(t) = [A_k \cos 2\pi f_c t - B_k \sin 2\pi f_c t] \cdot (1 - \delta/2) \sin(2\pi f_{LO}t + \theta/2) \\ &= A_k(1 - \delta/2) \cos 2\pi f_c t \sin(2\pi f_{LO}t + \theta/2) - B_k(1 - \delta/2) \sin 2\pi f_c t \sin(2\pi f_{LO}t + \theta/2) \\ &= 0.5A_k(1 - \delta/2) \sin(\theta/2) - 0.5B_k(1 - \delta/2) \cos(\theta/2) \end{aligned}$$

Complex notation at sampling time

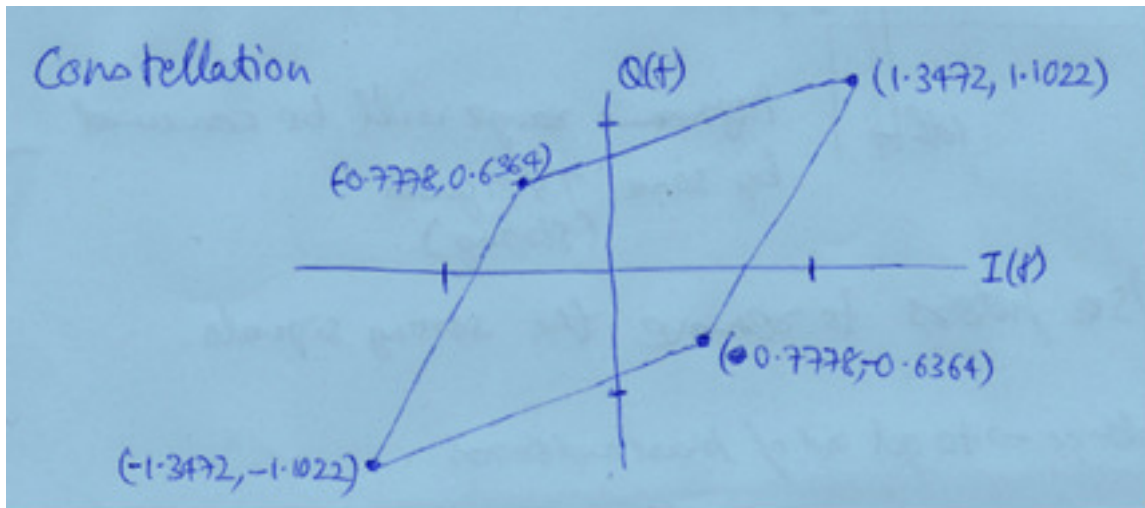
$$I_{LO}(t) + jQ_{LO}(t) = (1 + \delta/2) \cos(2\pi f_{LO}t - \theta/2) + j(1 - \delta/2) \sin(2\pi f_{LO}t + \theta/2)$$

$$\begin{aligned} s(t) &= \text{Re}\{\tilde{s}(t)e^{j2\pi f_c t}\} \\ \tilde{s}(t) &= A_k + jB_k = a_k e^{j\phi_k} \\ I(t) + jQ(t) &= \tilde{s}(t)e^{j2\pi f_c t}[I_{LO}(t) - jQ_{LO}(t)] \\ &= (A_k + jB_k)e^{j2\pi f_c t}[I_{LO}(t) - jQ_{LO}(t)] \\ &= (A_k + jB_k)e^{j2\pi f_c t}[(1 + \delta/2) \cos(2\pi f_{LO}t - \theta/2) - j(1 - \delta/2) \sin(2\pi f_{LO}t + \theta/2)] \\ &\text{fill in missing steps} \\ I(t) &= A_k(1 + \delta/2) \cos(\theta/2) - B_k(1 + \delta/2) \sin(\theta/2) \\ Q(t) &= A_k(1 - \delta/2) \sin(\theta/2) - B_k(1 - \delta/2) \cos(\theta/2) \end{aligned}$$

Substituting numbers $\theta = 30^\circ, \delta = 0.2$

A_k	B_k	$I(t)$	$Q(t)$
1	1	1.347	1.102
1	-1	0.778	-0.636
-1	1	-0.778	0.636
-1	-1	-1.347	-1.102

The result is that with amplitude and phase errors, the originally square constellation is distorted and looks like a trapezoid.



7.6 Differential BPSK

In this section we consider differential BPSK, in which the information is encoded in the phase difference between two adjacent symbols. D-BPSK (also written DPSK) has the advantage that the absolute phase (i.e. a coherent reference signal) is not required. In practice, the absolute phase is not generally available unless the transmitter and receiver are phase-locked via GPS.

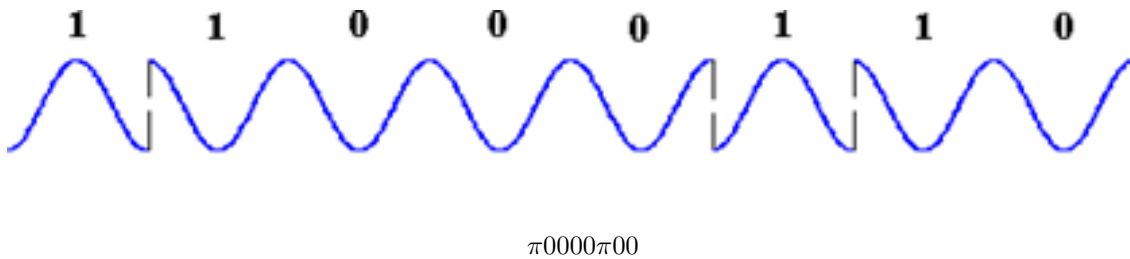
Terminology: If absolute phase is required, the receiver is *coherent*. If only phase differences are required, the receiver is *non-coherent*.

7.6.1 DBPSK with square pulse shape

A DBPSK waveform with square pulse shape is shown in the diagram below with one cycle of the carrier per bit time. This waveform is not used in practical systems because the sharp transitions will cause the bandwidth to be wider than necessary.

In this figure, the carrier wave is $\cos 2\pi f_c t + \phi$ where $\phi = 0$ or π

The phases are marked below the waveform.



From the phases $\pi 0000\pi 00$ the transmitted data is 10000100.

The message bits above the figure are the result of differentially decoding the transmitted data.

The message bits above the figure are shifted to the right by a half cycle of the carrier wave.

To see how the differential decoding works in this case, shift the message bits a half cycle to left so the bits are right on top of any change in phase.

Then the leftmost “1” bit is a reference followed by 1 = change, 0 = no change.

From the phases $\pi 0000\pi 00$ the transmitted data is 10000100.

Differential decoding of the transmitted data is the message data

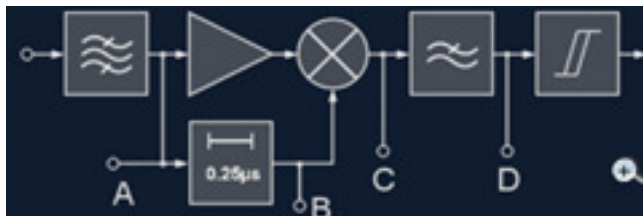
The transmitted data is the differentially encoded message data

10000100 transmitted data

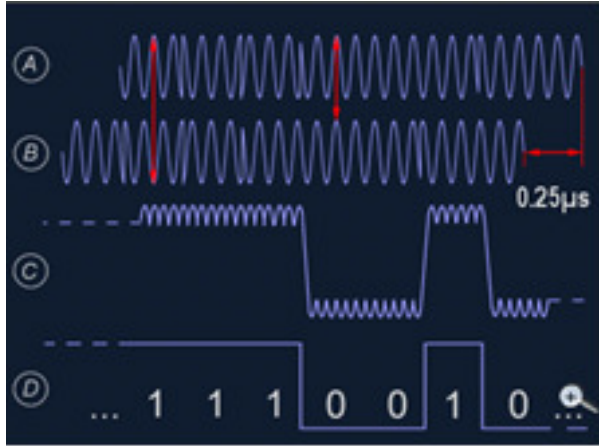
X1000110 diff decoded message data shown in the figure, where X is arbitrary.

10000100 diff encoding of diff decoded message data should yield transmitted data

In the figure below, differential decoding is illustrated with 3 cycles of carrier per symbol time. The differential decoder can be implemented with a delay and multiply operation with a delay of one symbol time. The carrier is removed by a low pass filter. In this example we have operated directly on the signal waveform, we do not need to demodulate first before doing the differential decoding.



A is the transmitted waveform, B is the delayed waveform, C is the output resulting from a delay and multiply operation, and D is the output of a decision device (slicer) that decides 1 or 0.



7.6.2 D-BPSK with RC2 time domain pulse shape

In this example, we show DBPSK waveforms with an RC2 time domain pulse shape, where the pulse spans two symbol periods.

One application of this waveform is low-rate data communication on the shortwave bands. It is used for live contacts on the amateur radio bands at a data rate of $R=31.25$ baud, corresponding to a keyboard speed of approximately 50wpm.

This mode uses a differential binary encoding where reversal of the carrier phase indicates the symbol '0' and no change in the carrier phase indicates the symbol '1'. Figure 1 shows an example of a modulated BPSK31 waveform. For a string of '1' symbols, the modulated waveform is equivalent to just transmitting the carrier. For a string of '0' symbols, the modulated waveform is equivalent to transmitting the DSB-SC signal:

$$s(t) = A \cos 2\pi f_c t \cos 2\pi f_m t = \frac{A}{2} \cos 2\pi(f_c + f_m)t + \frac{A}{2} \cos 2\pi(f_c - f_m)t \text{ where } f_m = R/2 = 15.625$$

Therefore, the bandwidth of the BPSK31 signal is only 31.25 Hz. This allows many operators to make contacts within a narrow bandwidth without interfering with each other.

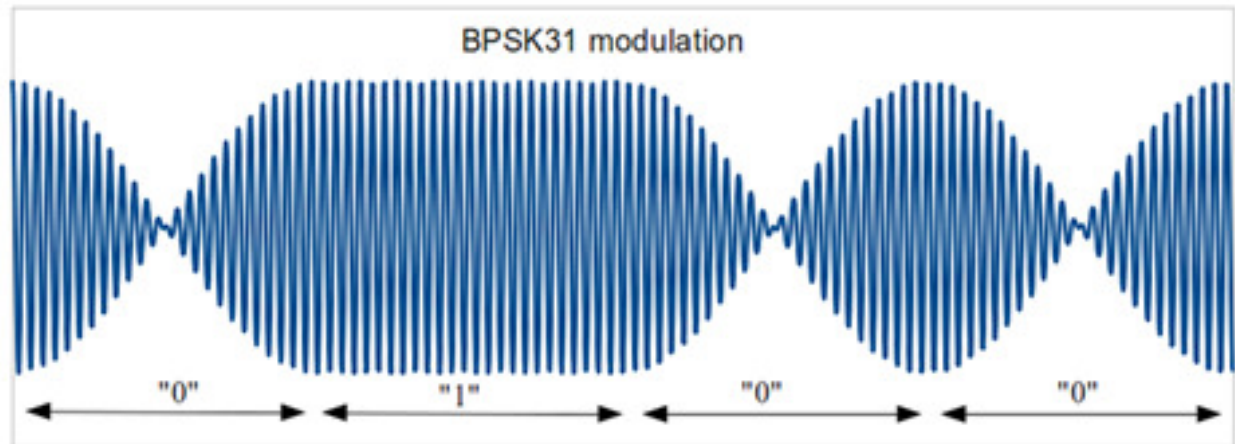


Figure 1 BPSK31 Waveform

BPSK31 uses a variable-length coding to encode the alphabet. More common characters use shorter codes while less common characters use longer codes. Additionally, each code must begin and end with a '1' symbol and may not contain two or more consecutive '0' symbols. A table of codes and additional information is at <http://aintel.bi.ehu.es/psk31theory.html>

Any string of two or more '0' symbols is ignored for the purposes of decoding the received message. This variable-length coding scheme has the following advantages:

1. Transmitting a string of '0' symbols will not result in any decoded message, but allows the PLL in the receiver to synchronize to the nominal 31.25 Hz symbol rate. Transmitting a string of '0' symbols (known as the idle signal) before the message text ensures that the receiver synchronizes to the symbol rate of the transmitted signal.
2. No additional bits are required to synchronize the start and end of characters. Since all characters start and end with '1' and the string '00' may not appear within a character, we know that the string '001' indicates the start of a character and the string '100' indicates the end of a character.

The BPSK31 mode does not use any form of error correction, although an intelligent decoder may try to make a best guess for a string of symbols that does not map to any character. There is a QPSK variant that provides error correction capabilities.

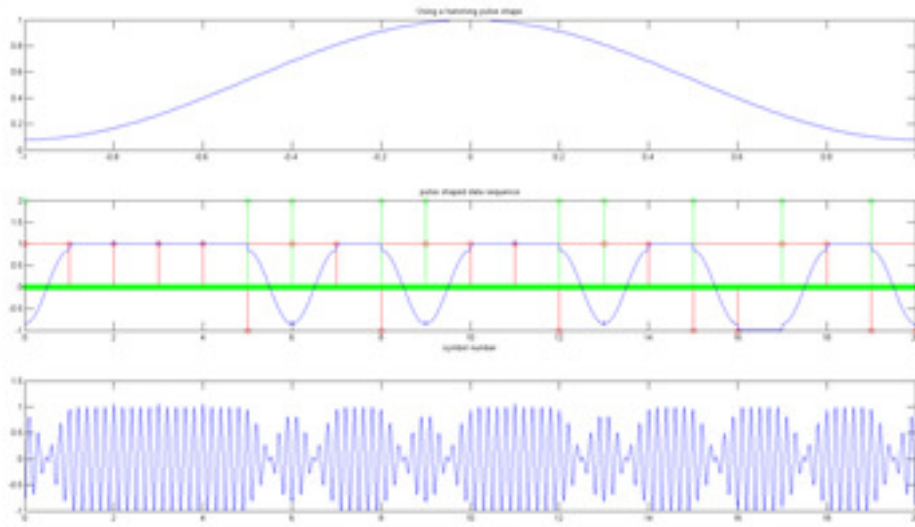
Exercise, write the BPSK31 signal in the standard BPSK format for a real BPSK signal

$$s(t) = \text{Re}[m(t)e^{j2\pi f_c t}] = m(t) \cos 2\pi f_c t = \sum_k D_k p(t - kT)$$

$D_k = A_k \otimes D_{k-1}$ is the differentially encoded message A_k

$$p_2(t) = [1 - \cos(\pi t/T)]/2 \text{ for } t \in \{0, 2T\}.$$

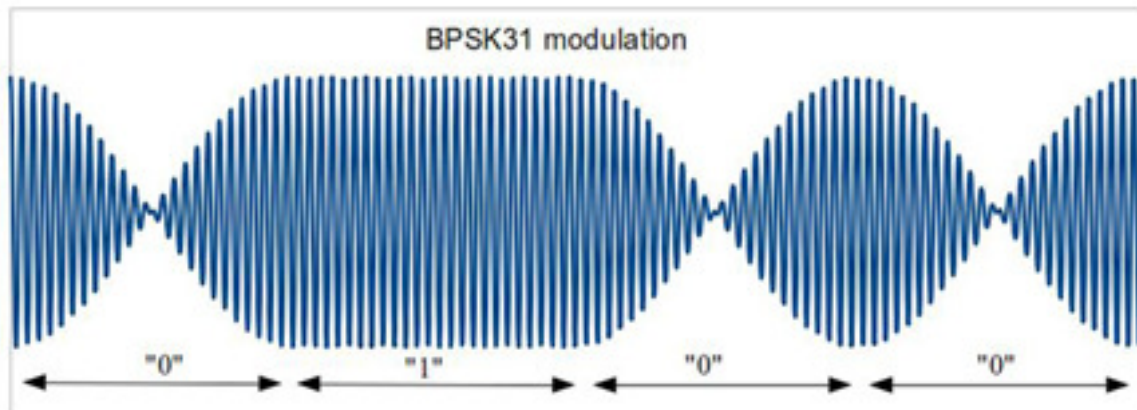
Hint: see the figure below.



The red impulses show the impulses D_k corresponding to the differentially *encoded* message bits. The green pulses are the decoder output, i.e. the differentially decoded pulses corresponding to the message bits.

The message waveform $m(t)$ is shown in blue. Each red impulse is convolved with the pulse $p(t)$ to yield an RC pulse spanning two symbol periods. The pulses overlap resulting in the blue waveform $m(t)$. Note that $m(t)$ has zero crossings one-half symbol time after a green “1” message bit.

The signal waveform with carrier is shown at the bottom. The envelope of the signal waveform crosses zero at $T/2$ after a green “1” message bit. Thus the message can be obtained by observing the zero crossings. In the figure below, the message is obtained in this way, however, decoder output is polarity reversed by convention for BPSK31, so that a change in the received sequence is decoded as a “0” bit and no change as a “1” bit.



7.6.3 DPSK with RRC frequency domain pulse

The wireless LAN standard 802.11b includes different modes for bit rates of 1,2,5.5 and 11 Mbits/sec. IN all caes the signal looks like an 11 MHz BPSK and QPSK waveform using square-root raised cosine (RRC) pulse shaping in the frequency domain with 75% excess bandwidth.

7.7 QPSK variants

In this section, we consider two variants of QPSK: offset QPSK and $\pi/4$ QPSK.

Types of QPSK



- ◆ **Conventional QPSK** has transitions **through zero** (i.e. 180° phase transition). Highly linear amplifiers required.
- ◆ **In Offset QPSK**, the phase transitions are limited to 90° , the transitions on the I and Q channels are staggered.
- ◆ **In $\pi/4$ QPSK** the set of constellation points are toggled each symbol, so transitions through zero cannot occur. This scheme produces **the lowest envelope variations**.
- ◆ **All QPSK schemes** require linear power amplifiers

25

7.7.1 Offset QPSK

The offset QPSK signal is the same as the QPSK signal, except that the quadrature component is delayed by half a symbol period.

Thus we modify the complex baseband signal for offset QPSK

$\tilde{s}(t) = i(t) + jq(t - T/2)$, where we use the expressions for $i(t), q(t)$ from QPSK, repeated here for convenience

$$i(t) = \sum_k A_k p(t - kT) = A_0 p(t) + A_1 p(t - T) + A_2 p(t - 2T) + \dots$$

and

$$q(t) = \sum_k B_k p(t - kT) = B_0 p(t) + B_1 p(t - T) + B_2 p(t - 2T) + \dots$$

For QPSK, we choose $A_k = \pm 1$ and $B_k = \pm 1$, so that $C_k = \pm 1 \pm j$.

The real offset QPSK signal is

$$s(t) = i(t) \cos 2\pi f_2 t - q(t - T/2) \sin 2\pi f_2 t$$

7.7.2 $\pi/4$ QPSK

The $\pi/2$ QPSK signal is the same as QPSK except that we code

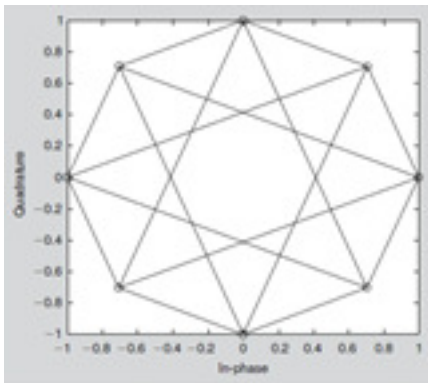
$$\begin{aligned} A_k &= A_{k-1} \cos \phi_k - B_{k-1} \sin \phi_k = \cos \delta_k \\ B_k &= A_{k-1} \sin \phi_k - B_{k-1} \cos \phi_k = \sin \delta_k \end{aligned}$$

We define the phases ϕ_k per the following table

A_k	B_k	ϕ_k
1	1	$\pi/4$
0	1	$3\pi/4$
0	0	$-3\pi/4$
1	0	$-\pi/4$

We use these expressions for A_k, B_k in the expressions for $i(t), q(t)$ from QPSK and find the real QPSK signal $s(t) = i(t) \cos 2\pi f_2 t - q(t) \sin 2\pi f_2 t$

We find the phases δ_k take on 8 possible values $n\pi/4$ for $n\{1, \dots, 8\}$. However, successive values of δ_k are never $n\pi/2$ apart, as shown in the signal constellation below.



We can prove this is true provided that

$$f_0 = \frac{2m - n}{4T} \text{ and } f_1 = \frac{2m + n}{4T} \text{ so that } f_1 - f_0 = \frac{n}{2T}$$

This relationship determines the condition for orthogonality between the two frequencies.

When $n = 1$ we have a special case of FSK called *minimum shift keying* (MSK).

For MSK, the frequency difference is half the symbol rate.

8.2 FSK complex envelope

8.2.1 FM equations

We want to write FSK in terms of the complex envelope

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

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi} = a(t)\cos\phi + ja(t)\sin\phi \\ &= i(t) + jq(t) \end{aligned}$$

We expect the FSK signal to have constant amplitude $a(t) = A_c$ and time varying phase $\phi(t)$.

To find the complex envelope for FSK, we start with the FM equations, in particular

$$f_i(t) = f_c + k_f m(t)$$

where the message is a square wave, or more precisely,

$$m(t) = \sum_k A_k p(t - kT) \text{ and } A_k = \pm 1, \begin{cases} p(t) = 1 \text{ for } 0 \leq t \leq T \\ p(t) = 0 \text{ otherwise} \end{cases}$$

The modulation sensitivity k_f is chosen so that

$$\begin{aligned} f_i(t) &= f_c + k_f A_k = f_c + \Delta f = f_0 \text{ for binary 0} \\ f_i(t) &= f_c + k_f A_k = f_c - \Delta f = f_1 \text{ for binary 1} \end{aligned}$$

The FSK signal is written in standard format

$$s(t) = a(t)\cos\theta(t) = a(t)\cos[2\pi f_c t + \phi(t)]$$

where $\theta(t) = 2\pi f_c t + \phi(t)$ has a linear variation at rate f_c and a time varying part $\phi(t)$

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} = f_c + \frac{1}{2\pi} \frac{d\phi(t)}{dt}$$

Combining these 3 starting points, we can write:

$$\begin{aligned} \theta(t) &= 2\pi \int_0^t f_i(\alpha) d\alpha = 2\pi \int_0^t [f_c + k_f m(\alpha)] d\alpha \\ &= 2\pi f_c t + k_f \int_0^t m(\alpha) d\alpha \\ s(t) &= A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t m(\alpha) d\alpha \right] \\ &= A_c \cos \left[2\pi f_c t + 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha \right] \\ &= A_c \cos [2\pi f_c t + \phi(t)] \end{aligned}$$

Thus

$$\phi(t) = 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha$$

8.2.2 FSK phase ramps

The integral

$$\phi(t) = 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha$$

can be defined over intervals $kT \leq t \leq (k+1)T$ for each pulse with data A_k

For example, if we assume $A_0 = +1$ for $0 \leq t \leq T$

$$\begin{aligned} \phi(t) &= 2\pi k_f \int_0^t A_0 \cdot 1 \cdot d\alpha = 2\pi k_f t = 2\pi \Delta f t \text{ for } 0 \leq t \leq T \\ f_i(t) &= f_c + \frac{1}{2\pi} \frac{d\phi(t)}{dt} = f_c + k_f = f_c + \Delta f = f_1 \end{aligned}$$

Observe that the phase is a ramp that starts at $\phi(t=0) = 0$ and increases linearly to $\phi(t=T) = 2\pi \Delta f T$

Recall that for orthogonal FSK, $f_1 - f_0 = \frac{n}{2T}$ and note that $f_1 - f_0 = 2\Delta f$

Thus $\Delta f = \frac{n}{4T}$, so that the smallest ($n=1$) frequency shift away from the center frequency f_c is one quarter the symbol rate, i.e.

$$\begin{aligned} f_1 &= f_c + \Delta f = f_c + 1/4T \\ f_0 &= f_c - \Delta f = f_c - 1/4T \end{aligned}$$

From the phase ramp result $\phi(t=T) = 2\pi \Delta f T$, we find

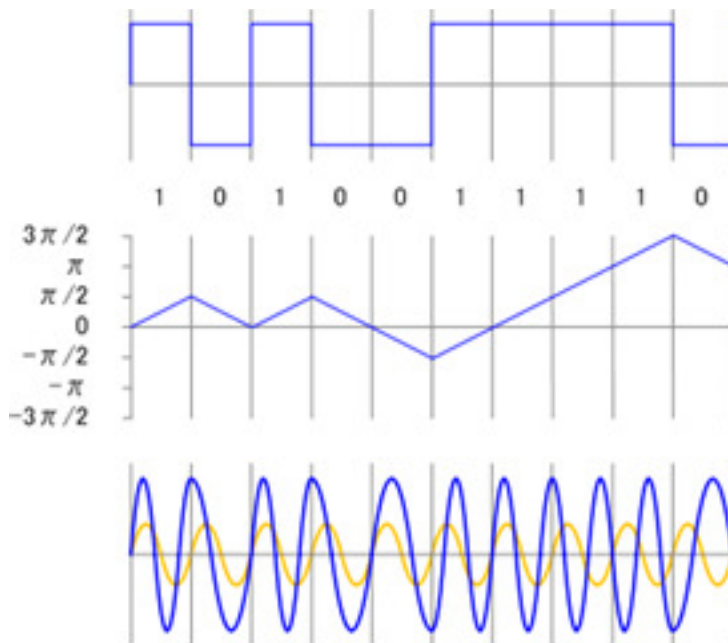
$$\phi(t=T) = 2\pi \frac{n}{4T} T = \frac{n\pi}{2}$$

If we choose the two FSK frequencies to be as close as possible and still orthogonal ($n = 1$), then the phase ramps from $\phi(t = 0) = 0$ to $\phi(t = T) = \pi/2$

Exercise: Continue the example for the next symbol. Assume $A_1 = -1$ for $T \leq t \leq 2T$, and show that the phase will ramp down from

$$\phi(t = T) = \pi/2 \text{ to } \phi(t = 2T) = 0$$

The figure below shows the phase $\phi(t)$ and the FSK signal $s(t)$ for an example bit pattern 1010011110



The phase may wander far away from zero. In most cases, we choose the data pattern to have the same number of 1s and 0s on average, so that the phase will stay close to zero on average.

8.2.3 FSK complex envelope

The FM signal $s(t)$ can be written in standard IQ format:

$$\begin{aligned} s(t) &= \text{Re}\{a(t)e^{j\phi(t)}e^{j2\pi f_c t}\} = \\ &= \text{Re}\{[I(t) + jQ(t)][\cos 2\pi f_c t + j\sin 2\pi f_c t]\} \\ &= I(t)\cos 2\pi f_c t - Q(t)\sin 2\pi f_c t \\ &= a(t)\cos[2\pi f_c t + \phi(t)] \end{aligned}$$

Thus for FSK:

$$\begin{aligned}
a(t) &= A_c \\
\phi(t) &= 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha \\
I(t) &= A_c \cos 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha \\
Q(t) &= A_c \sin 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha
\end{aligned}$$

As shown above, if we assume $A_0 = +1$ for $0 \leq t \leq T$

$$\begin{aligned}
\phi(t) &= 2\pi k_f \int_0^t \sum_k A_k p(\alpha - kT) d\alpha \\
&= 2\pi k_f \int_0^t A_0 \cdot 1 \cdot d\alpha = 2\pi k_f t = 2\pi \Delta f t \text{ for } 0 \leq t \leq T \\
I(t) &= \cos \phi(t) = \cos 2\pi \Delta f t \text{ for } 0 \leq t \leq T \\
Q(t) &= \sin \phi(t) = \sin 2\pi \Delta f t \text{ for } 0 \leq t \leq T \\
f_i(t) &= f_c + \frac{1}{2\pi} \frac{d\phi(t)}{dt} = f_c + k_f = f_c + \Delta f = f_1
\end{aligned}$$

8.3 MSK and offset QPSK

The real offset QPSK signal with a symbol rate $1/T$ is

$$s(t) = i(t) \cos 2\pi f_2 t - q(t - T/2) \sin 2\pi f_2 t$$

Assuming the data symbols at the sampling time are $i(t) = I_k = \pm 1, q(t) = Q_k = \pm 1, kT \leq t \leq (k+1)T$

Consider a variation of this QPSK signal at a slower symbol rate $1/2T$ and a cosine pulse shape over two symbol periods

$$s(t) = i(t) \cos \frac{\pi t}{2T} \cos 2\pi f_2 t - q(t - T) \sin \frac{\pi t}{2T} \sin 2\pi f_2 t$$

From trigonometric identities, we can write

$$s(t) = \cos[2\pi(f - I_k Q_k \frac{1}{4T})t + \frac{\pi}{2}(1 - I_k)], kT \leq t \leq (k+1)T$$

Thus MSK and offset QPSK with cosine pulse shaping are equivalent with

$$\begin{aligned}
f_1 &= f_c + \Delta f = f_c + 1/4T \\
f_0 &= f_c - \Delta f = f_c - 1/4T
\end{aligned}$$

Appendix A

Useful formulas

$$e^{j\theta} = \cos \theta + j \sin \theta$$

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

$$\sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j}$$

$$\begin{aligned} e^{j(\alpha+\beta)} &= \cos(\alpha + \beta) + j \sin(\alpha + \beta) \\ &= e^{j\alpha} e^{j\beta} \\ &= (\cos \alpha + j \sin \alpha) (\cos \beta + j \sin \beta) \\ &= \cos \alpha \cos \beta - \sin \alpha \sin \beta + j (\cos \alpha \sin \beta + \sin \alpha \cos \beta) \end{aligned}$$

$$\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$$

$$\sin(\alpha + \beta) = \cos \alpha \sin \beta + \sin \alpha \cos \beta$$

$$\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$$

$$\sin(\alpha - \beta) = \sin \alpha \cos \beta - \cos \alpha \sin \beta$$

$$\cos \alpha \cos \beta = \frac{\cos(\alpha - \beta) + \cos(\alpha + \beta)}{2}$$

$$\sin \alpha \cos \beta = \frac{\sin(\alpha - \beta) + \sin(\alpha + \beta)}{2}$$

$$\sin \alpha \sin \beta = \frac{\cos(\alpha - \beta) - \cos(\alpha + \beta)}{2}$$

$$\cos \alpha \sin \beta = \frac{-\sin(\alpha - \beta) + \sin(\alpha + \beta)}{2}$$

$$\begin{aligned} M(f) &= \text{FFT}\{m(t)\}, m(t) \leftrightarrow M(f) \\ &= \int_{t=-\infty}^{\infty} m(t)e^{-j2\pi ft} dt \end{aligned}$$

$$\begin{aligned} e^{j2\pi f_c t} &\leftrightarrow \delta(f - f_c) \\ m(t)e^{j2\pi f_c t} &\leftrightarrow M(f - f_c) \end{aligned}$$

$$\begin{aligned} \tilde{s}(t) &= a(t)e^{j\phi(t)} \\ s_+(t) &= \tilde{s}(t)e^{j2\pi f_c t} \\ s_-(t) &= \tilde{s}(t)^* e^{-j2\pi f_c t} \\ s(t) &= s_+(t)/2 + s_-(t)/2 \\ &= \text{Re}[s_+(t)] \\ &= \text{Re}[s_-(t)] \end{aligned}$$

$$\begin{aligned} \tilde{S}(f) &= \text{FFT}\{\tilde{s}(t)\}, \tilde{s}(t) \leftrightarrow \tilde{S}(f) \\ \tilde{s}_n &= \tilde{s}(t = nT_s), T_s = 1/f_s, \tilde{S}_k = \tilde{S}(f = kf_0) \end{aligned}$$

$$\begin{aligned} \tilde{S}_k &= \int_{t=0}^{T_0=nT_s} \tilde{s}(t)e^{-j2\pi ft} dt \Big|_{t=nT_s, f=kf_0} \\ &= \sum_{n=0}^{N-1} \tilde{s}_n e^{-j2\pi kf_0 nT_s} \\ &= \sum_{n=0}^{N-1} \tilde{s}_n e^{-j2\pi nk/N} \end{aligned}$$

$$T_s f_0 = T_s/T_0 = f_0/f_s = 1/N$$