Communication Systems

January 26, 2014

Contents

Current Status:

0.1 Introduction

This book will eventually cover a large number of topics in the field of electrical communications. The reader will also require a knowledge of Time and Frequency Domain representations, which is covered in-depth in the Signals and Systems[1] book. This book will, by necessity, touch on a number of different areas of study, and as such is more than just a text for aspiring Electrical Engineers. This book will discuss topics of analog communication schemes, computer programming, network architectures, information infrastructures,

---

1   http://en.wikibooks.org/wiki/Signals%20and%20Systems

communications circuit analysis, and many other topics. It is a large book, and varied, but it should be useful to any person interested in learning about an existing communication scheme, or in building their own. Where previous Electrical Engineering books were grounded in theory (notably the Signals and Systems[2] book), this book will contain a lot of information on current standards, and actual implementations. It will discuss how current networks and current transmission schemes work, and may even include information for the intrepid engineer to create their own versions of each.

This book is still in an early stage of development. Many topics do not yet have pages, and many of the current pages are stubs. Any help would be greatly appreciated.

---

2   http://en.wikibooks.org/wiki/Signals%20and%20Systems

1 Communications Basics

It is important to know the difference between a baseband signal, and a broad band signal. In the Fourier Domain, a baseband signal is a signal that occupies the frequency range from 0Hz up to a certain cutoff. It is called the baseband because it occupies the **base**, or the lowest range of the spectrum.

In contrast, a broadband signal is a signal which does not occupy the lowest range, but instead a higher range, 1MHz to 3MHz, for example. A wire may have only one baseband signal, but it may hold any number of broadband signals, because they can occur anywhere in the spectrum.

w: Baseband[1]

1.1 Wideband vs Narrowband

in form of frequency modulation. wideband fm has been defined as that in which the modulation index normally exceeds unity.

1.2 Frequency Spectrum

A graphical representation of the various frequency components on a given transmission medium is called a frequency spectrum.

Consider a situation where there are multiple signals which would all like to use the same wire (or medium). For instance, a telephone company wants multiple signals on the same wire at the same time. It certainly would save a great deal of space and money by doing this, not to mention time by not having to install new wires. How would they be able to do this? One simple answer is known as **Time-Division Multiplexing**.

1.3 Time Division Multiplexing

w:Time-division_multiplexing[2]

Time-Division Multiplexing (TDM) is a convenient method for combining various digital signals onto a single transmission media such as wires, fiber optics or even radio. These signals may be interleaved at the bit, byte, or some other level. The resulting pattern may

---

1   http://en.wikipedia.org/wiki/%20Baseband
2   http://en.wikipedia.org/wiki/Time-division_multiplexing

be transmitted directly, as in digital carrier systems, or passed through a modem to allow the data to pass over an analog network. Digital data is generally organized into frames for transmission and individual users assigned a time slot, during which frames may be sent. If a user requires a higher data rate than that provided by a single channel, multiple time slots can be assigned.

Digital transmission schemes in North America and Europe have developed along two slightly different paths, leading to considerable incompatibility between the networks found on the two continents.

BRA (basic rate access) is a single digitized voice channel, the basic unit of digital multiplexing.

BRA Comparison		
Characteristic	North America	Europe
Basic channel rate	64 Kbps	64 Kbps
CODEC format	μ-Law	A-Law
Binary format	Folded binary	Sign magnitude
Voice Channels per frame	24	30
Signaling	Bit robbing	Dedicated channel

Figure 1

Digital Transmission Systems[2]

Designation	User	Bit Rate [Mbps]	Line Code	Media	Repeater Spacing
T1	AT&T	1.544	AMI/B8ZS	Twisted pair	6 Kft
CEPT1	CCITT	2.048	HDB3 (B4ZS)	Twisted pair	2 Km
T1C	AT&T	3.152	Bipolar	Twisted pair	6 Kft
T148	ITT	2.37 ternary	4B3T	Twisted pair	6 Kft
9148A	GTE	3.152	1-D² duobinary	Twisted pair	6 Kft
T1D	AT&T	3.152	1+D duobinary	Twisted pair	6 Kft
T1G	AT&T	6.443	4-level	Twisted pair	6 Kft
T2	AT&T	6.312	B6ZS	Low cap twisted pair	14.8 Kft
LD-4	Canada	274.176	B3ZS	Coax	1.9 Km
T4M	AT&T	274.176	Polar	Coax binary NRZ	5.7 Kft

Figure 2

1.3.1 North American TDM

The various transmission rates are not integral numbers of the basic rate. This is because additional framing and synchronization bits are required at every multiplexing level.

DS Hierarchy			
Designation	# Voice Channels	Bit Rate [Mbps]	Comments
DS'–0	1	.064	100% duty cycle, unipolar
DS–1	24	1.544	Media: 22 AWG cable Repeater spacing: 6000 ft Signal: bipolar RZ
DS–1C	48	3.152	Media: 22 AWG cable Repeater spacing: 6000 ft Signal: bipolar RZ
DS–2	96	6.312	media: low capacitance 22 AWG Max repeater spacing: 14,800 ft Signal: bipolar RZ, B6ZS
DS–3	672	44.736	Media: microwave, fiber Signal: RZ, B3ZS
DS-3A	1344	91.04	
DS–4	4032	274.176	Media: coax, microwave Signal: bipolar NRZ

Figure 3

In North America, the basic digital channel format is known as DS-0. These are grouped into frames of 24 channels each. A concatenation of 24 channels and a start bit is called a frame. Groups of 12 frames are called multiframes or superframes. These vary the start bit to aid in synchronizing the link and add signaling bits to pass control messages.

w:Digital_Signal_1[3]

---

3   http://en.wikipedia.org/wiki/Digital_Signal_1
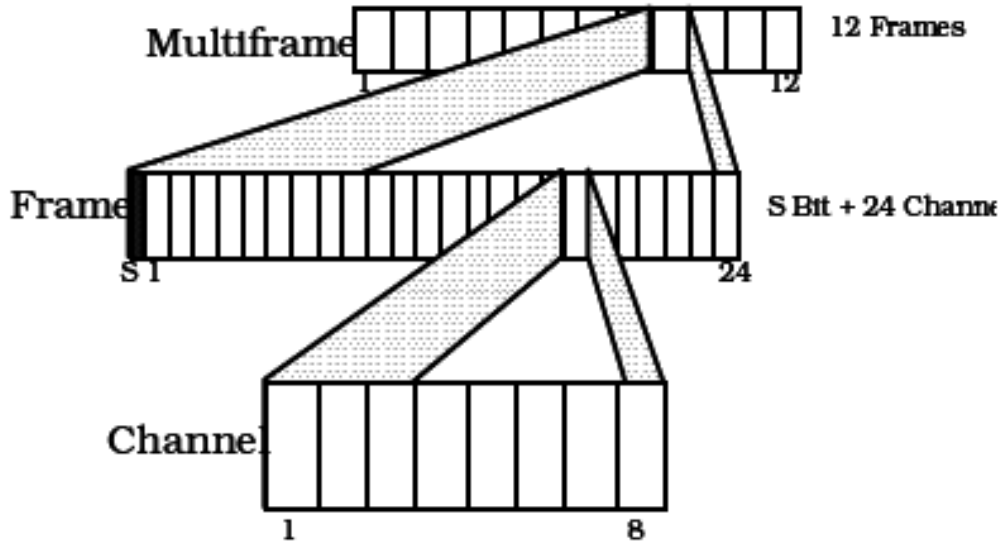
Figure 4

S Bit Synchronization

The S bit is used to identify the start of a DS-1 frame. There are 8 thousand S bits per second. They have an encoded pattern, to aid in locating channel position within the frame.

Frame	1	2	3	4	5	6	7	8	9	10	11	12
S Bit	1	0	0	0	1	1	0	1	1	1	0	0

Figure 5

This forms a regular pattern of 1 0 1 0 1 0 for the odd frames and 0 0 1 1 1 0 for the even frames. Additional synchronization information is encoded in the DS-1 frame when it is used for digital data applications, so lock is more readily acquired and maintained.

For data customers, channel 24 is reserved as a special sync byte, and bit 8 of the other channels is used to indicate if the remaining 7 bits are user data or system control information. Under such conditions, the customer has an effective channel capacity of 56 Kbps.

To meet the needs of low speed customers, an additional bit is robbed to support sub-rate multiplexer synchronization, leaving 6 x 8 Kbps = 48 Kbps available. Each DS-0 can be utilized as:

- 5 x 9.6 Kbps channels or
- 10 x 4.8 Kbps channels or
- 20 x 2.48 Kbps channels.

In the DS-2 format, 4 DS-1 links are interleaved, 12 bits at a time. An additional 136 Kbps is added for framing and control functions resulting in a total bit rate of 6.312 Mbps.

Signaling

Signaling provides control and routing information. Two bits, called the A and B bits, are taken from each channel in frames 6 and 12 in the multiframe. The A bit is the least significant bit in each channel in frame 6, and the B bit is the least significant bit in each channel in frame 12. This provides a signaling rate of 666 2/3 bps per channel.

The quality of voice transmission is not noticeably affected when 2% of the signal is robbed for signaling. For data, it may be a different story. If the data is encoded in an analog format such as FSK or PSK, then robbing bits is of no consequence, but if the data is already in digital form, then robbing bits results in unacceptable error rates. It is for this reason that in North America, a 64 Kbps clear channel cannot readily be switched through the PSTN. This means that data customers are limited to 56 Kbps clear channels. This simple condition has a profound effect on the development of new services such as ISDN. In most facilities, the A and B bits represent the status of the telephone hook switch, and correspond to the M lead on the E&M interface of the calling party.

ESF

CCITT has modified the North American digital hierarchy for the deployment of ISDN, by means of recommendation G.704. ESF consists of 24 DS-0 channels in a frame, but groups them into a 24-frame multiframe instead of the usual 12-frame multiframe.

The S bit is renamed the F bit, but only 1/4 of them are used for synchronization. This is possible because of improvements in frame search techniques and allows more signaling states to be defined.

Bit robbing is still used for signaling over an ESF link, but with the advent of ISDN, it will not be permitted. Instead, channel 24 is used to support a D channel.

F Bit Function	Frame Position	Comments
Sync pattern	4, 8, 12, 16, 20, 24	Pattern: 001011
CRC check	2, 6, 10, 14, 18, 22	Is the remainder from modulo-2 division of all the bits in the previous frame by the binary polynomial x^6+x+1
Maintenance	all odd frames	Currently not specified, but will convey: maintenance, diagnostic, and status information

Figure 6

Typical T1 CPE Application

The large telecom carriers are not the only ones who deploy high-speed TDM facilities. In many cases, heavy users of voice or data services can reduce their transmission costs by concentrating their numerous low speed lines on to a high speed facility.

There are many types of T1 multiplexers available today. Some are relatively simple devices, while others allow for channel concatenation, thus supporting a wide range of data rates. The ability to support multiple DS-0s allows for easy facilitation of such protocols as the video teleconferencing standard, Px64.
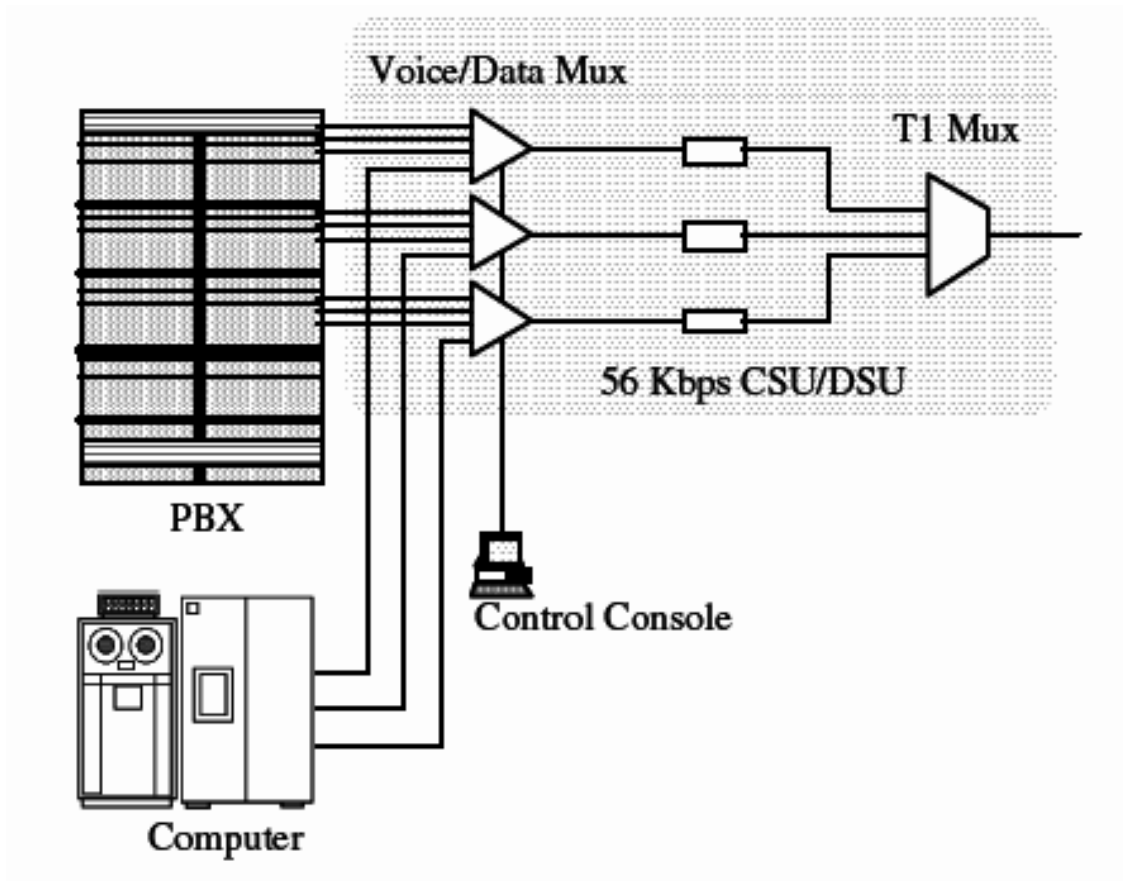
Figure 7

Multiplexers

Multiplexing units are often designated by the generic term M_{ab} where a is input DS level and b is the output DS level. Thus, an M13 multiplexer combines 28 DS–1s into a single DS–3 and an M23 multiplexer combines 7 DS–2s into a single DS–3.
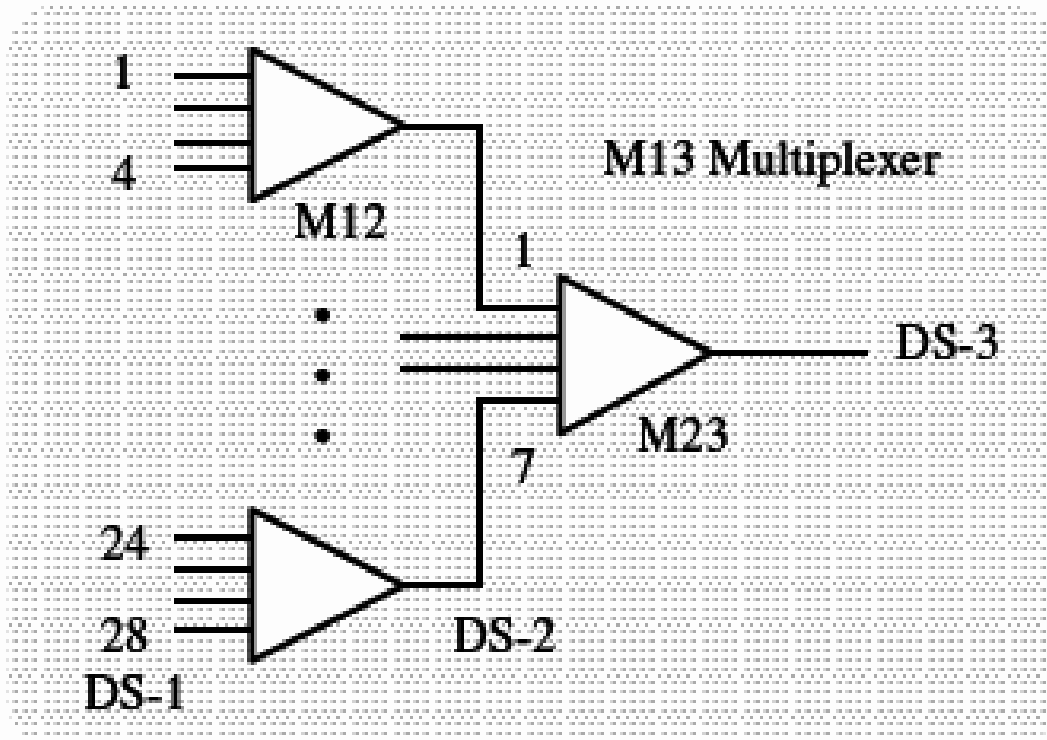
Figure 8

ZBTSI

ZBTSI (zero byte time slot interchange) is used on DS–4 links. Four DS-1 frames are loaded into a register, and renumbered 1–96. If there are any empty slots [all zeros], the first framing bit is inverted and all blank slots are relocated to the front of the frame. Channel 1 is then loaded with a 7-bit number corresponding to the original position of the first empty slot. Bit 8 used to indicate whether the following channel contains user information or another address for an empty slot.

If there is a second vacancy, bit 8 in the previous channel is set, and the empty slot address is placed in channel 2. This process continues until all empty positions are filled.

The decoding process at the receiver is done in reverse. Borrowing 1 in 4 framing bits for this system is not enough to cause loss of synchronization and provides a 64 Kbps clear channel to the end-user.

1.3.2 European TDM Carriers

European systems were developed along slightly different principles. The 64 Kbps channel is still the basic unit, but signaling is not included in each channel. Instead, common channel signaling is used. In a level 1 carrier, channels 0 and 16 are reserved for signaling

and control. This subtle difference means that European systems did not experience the toll fraud and 56 k bottlenecks common to North American systems, and they experience a much larger penetration of ISDN services.

Level	# Voice Channels	Bit Rate [Mbps]	Line Code
0	1	.064	
1	30	2.048	HDB3
2	120	8.448	HDB3
3	480	34.368	HDB3
4	1920	139.264	CMI
5	7680	565.148	

Figure 9
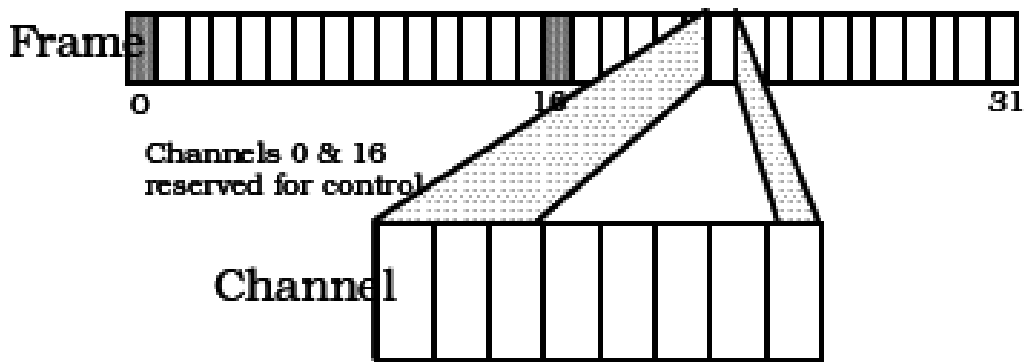
Figure 10

1.4 Zero Substitutions

In order to prevent transmission systems from loosing lock on the data stream, it is necessary to avoid long strings of zeros. One of the most effective ways of doing this is to replace the zeros with a predetermined code. This substitution must be done in such a way that the receiver can identify it and strip it off before passing the data stream to the client.

AMI provides a simple means of detecting substitutions. In the normal course of events, alternate marks are inverted. Therefor, deliberately inducing a bipolarvariation at the transmitter can alert the receiver of a substitution. However, a single violation is indistinguishable from a transmission error. Consequently, some additional condition must also occur.

There are two common methods to create a second condition:

· Create a second bipolar violation in the opposite direction, within a specified time. This has the effect of keeping the average signal level at zero.

· Count the number of marks from the last substitution to predict the next type of violation

1.4.1 B6ZS

B6ZS (binary six zero substitution) is used on T2 AMI transmission links.

Synchronization can be maintained by replacing strings of zeros with bipolar violations. Since alternate marks have alternate polarity, two consecutive pulses of the same polarity constitute a violation. Therefore, violations can be substituted for strings of zeros, and the receiver can determine where substitutions were made.

Since the last mark may have been either positive (+) or negative (-), there are two types of substitutions:

Polarity of previous mark	Substitution					
-	0	-	+	0	+	-
+	0	+	-	0	-	+

Figure 11

These substitutions force two consecutive violations. A single bit error does not create this condition.

B6ZS Example:

Original data:	0	-	0	0	0	0	0	0	0	+	0	-	+
After substitution:	0	-	0	-	+	0	+	-	0	+	0	-	+

Original data:	0	+	0	0	0	0	0	0	0	-	0	+	-
After substitution:	0	+	0	+	-	0	-	+	0	-	0	+	-

Figure 12

1.4.2 B8ZS

This scheme uses the same substitution as B6ZS.

Polarity of previous mark	Substitution							
-	0	0	0	-	+	0	+	-
+	0	0	0	+	-	0	-	+

Figure 13

1.4.3 B3ZS

B3ZS is more involved than B6ZS, and is used on DS–3 carrier systems. The substitution is not only dependent on the polarity of the last mark, but also on the number of marks since the last substitution.

Previous Mark Polarity	Number of marks since the last substitution	
	Odd	Even
-	0 0 -	+ 0 +
+	0 0 +	- 0 -

Figure 14

B3ZS Example:

Let's suppose that there were 27 marks since the last substitution:

Original data:

    - 0 0 0 0 + 0 0 0 - + 0 0 0 0 -

After Substitution:

    - 0 0 - 0 + 0 0 + - + - 0 - 0 -

Figure 15

1.4.4 HDB3

HDB3 (high density binary 3) introduces bipolar violations when four consecutive zeros occur. It can therefore also be called B4ZS. The second and thirds zeros are left unchanged,

but the fourth zero is given the same polarity as the last mark. The first zero may be modified to a one to make sure that successive violations are of alternate polarity.

Previous Mark Polarity	Number of marks since the last substitution	
	Odd	Even
-	0 0 0 -	+ 0 0 +
+	0 0 0 +	- 0 0 -

Figure 16

HDB3 is used in Europe. Violation, or V pulses are injected after three consecutive zeros. The fourth zero is given the same polarity as the last mark. In the event of long strings of zeros occurring, a succession of single polarity pulses would occur, and a dc offset would build-up.

To prevent this, the first zero in a group of 4, may be modified to a 1. This B or balancing pulse assures that successive violations are of alternate polarity.
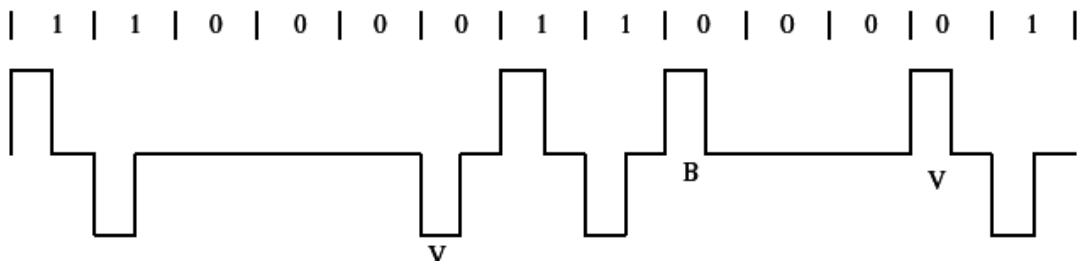
Figure 17

1.4.5 Block Code Substitution

These schemes operate on bytes rather than a bit at a time. Some transmit the signal as binary levels, but most use multi-level pulses. Some authors categorize these as line codes.

A binary block code has the designation $n\text{B}m\text{B}$, where n input bits are encoded into m output bits. The most common of these is the 3B4B code.

3B4B Coding	
Input	Output
000	- - + - or + + - +
001	- - + +
010	- + - +
011	- + + -
100	+ - - +
101	+ - + -
110	+ + - -
111	- + - - or + - + +

Figure 18

In Europe 4B3T, which encodes 4 binary bits into 3 ternary levels, has been selected as the BRA for ISDN. In North America, 2B1Q which encodes 2 binary bits into 1 quaternary level has been selected for BRA.

2B1Q Coding	
Input	Output
00	-3
01	-1
10	+1
11	+3

Figure 19

Some block codes do not generate multilevel pulses. For example, 24B1P or 24B25B simply adds a P or parity bit to a 24 bit block.

1.5 Benefits of TDM

TDM is all about cost: fewer wires and simpler receivers are used to transmit data from multiple sources to multiple destinations. TDM also uses less bandwidth than Frequency-Division Multiplexing (FDM) signals, unless the bitrate is increased, which will subsequently increase the necessary bandwidth of the transmission.

1.6 Synchronous TDM

Synchronous TDM is a system where the transmitter and the receiver both know exactly which signal is being sent. Consider the following diagram:

```
Signal A ---> |---| |A|B|C|A|B|C|   |------| ---> Signal A
Signal B ---> |TDM| -------------> |De-TDM| ---> Signal B
Signal C ---> |---|                |------| ---> Signal C
```
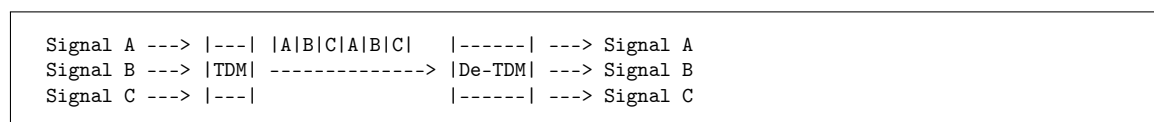
In this system, starting at time-slice 0, every third time-slice is reserved for Signal A; starting at time-slice 1, every third time-slice is reserved for Signal B; and starting at time-slice 2, every third time-slice is reserved for Signal C. In this situation, the receiver (De-TDM) needs only to switch after the signal on each time-slice is received.

The data flow of each input connection is divided into units where each input occupies one input time slot. Each input connection has a time slot alloted in the output irrespective of the fact whether it is sending data or not.

```
      A -----|A3|A2|A1|---> |---| .............|C3|B3|A3|C2|B2|A2|C1|B1|A1|
|------| --->  A
          |  |  |  |                   |          |          |
      B -----|B3|B2|B1|---> |MUX| -------------|--------|--------|---------->
|De-MUX| --->  B
          |  |  |  |                   |          |          |
      C -----|C3|C2|C1|---> |---|      |          |          |
|------| --->  C
                <-->                          <-------->
              Bit Interval                   Frame (x seconds)
```
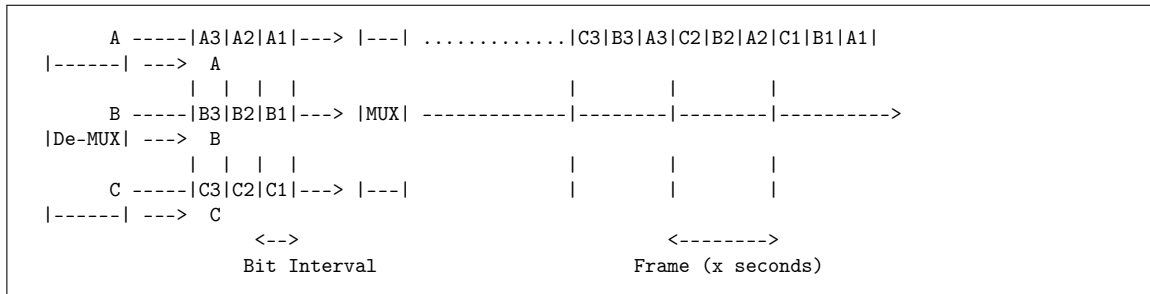
Sync TDM is inefficient when one or more input lines have no data to send. Thus, it is used with lines with high data traffic.

Sampling rate is same for all signals. Maximum sampling rate = twice the maximum frequency all the signals.

1.7 Statistical TDM

Synchronous TDM is beneficial because the receiver and transmitter can both cost very little. However, consider the most well-known network: the Internet. In the Internet, a given computer might have a data rate of 1kbps when hardly anything is happening, but might have a data rate of 100kbps when downloading a large file from a fast server. How are the time-slices divided in this instance? If every time slice is made big enough to hold 100Kbps, when the computer isn't downloading any data, all of that time and electricity will be wasted. If every time-slice is only big enough for the minimum case, the time required to download bigger files will be greatly increased.

The solution to this problem is called **Statistical TDM**, and is the solution that the Internet currently uses. In Statistical TDM, each data item, known as the payload (we used time-slices to describe these earlier), is appended with a certain amount of information about who sent it, and who is supposed to receive it (the header). The combination of a payload and a header is called a **packet**. Packets are like envelopes in the traditional "snail mail" system: Each packet contains a destination address and a return address as well as some enclosed data. Because of this, we know where each packet was sent from and where it is going.

The downside to statistical TDM is that the sender needs to be smart enough to write a header, and the receiver needs to be smart enough to read the header and (if the packet is to be forwarded,) send the packet toward its destination.

1.7.1 Link Utilization

w:Queuing_theory[4]

Statistical multiplexing attempts to maximize the use of a communication path. The study of this is often called *queuing theory.* A queue is simply a line of customers or packets waiting to be served. Under most circumstances, the arrival rate is unpredictable and therefor follows a random or Poisson distribution pattern, whereas the service time is constant.

The utilization or fraction of time actually used by a packet multiplexing system to process packets is given by:

$$\rho = \lambda\tau = \frac{\alpha NR}{M} = \frac{\alpha}{K} = \frac{\lambda}{M}$$

where

$\lambda =$ average arrival rate in bps

$\tau =$ service time in seconds

$\alpha =$ fraction of time each source is transmitting

$N =$ number of inputs sources

$R =$ source data rate in bps

$M =$ effective channel capacity in bps

$K =$ compression ratio

Figure 20

The queue length or average number of items waiting to be served is given by:

$$q = \frac{\rho^2}{2\,(1-\rho)} + \rho$$

4   http://en.wikipedia.org/wiki/Queuing_theory
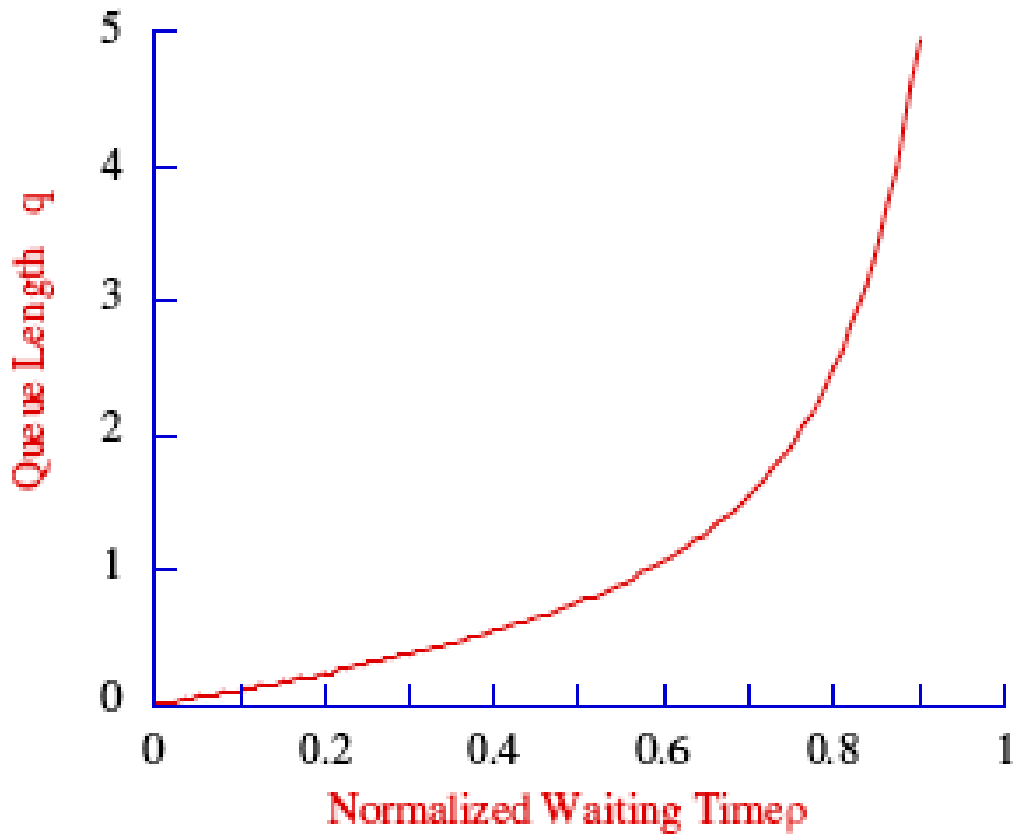
Figure 21

Example

A T1 link has been divided into a number of 9.6 Kbps channels and has a combined user data rate of 1.152 Mbps. Access to this channel is offered to 100 customers, each requiring 9.6 Kbps data 20% of the time. If the user arrival time is strictly random find the T1 link utilization.

Solution

The utilization or fraction of time used by the system to process packets is given by:

$$\rho = \frac{\alpha N R}{M} = \frac{0.2 \times 100 \times 9.6 \times 10^3}{1.152 \times 10^6} = 0.167$$

A 24 channel system dedicated to DATA, can place five 9.6 Kbps customers in each of 23 channels, for a total of 115 customers. In the above statistical link, 100 customers created an *average* utilization of 0.167 and were easily fitted, with room to spare if they transmit on the average 20% of the time. If however, the customer usage were not randomly distributed, then the above analysis would have to be modified.

This example shows the potential for statistical multiplexing. If channels were assigned on a demand basis (only when the customer had something to send), a single T1 may be able to support hundreds of low volume users.

A utilization above 0.8 is undesirable in a statistical system, since the slightest variation in customer requests for service would lead to buffer overflow. Service providers carefully monitor delay and utilization and assign customers to maximize utilization and minimize cost.

1.8 Packets

Packets will be discussed in greater detail once we start talking about digital networks (specifically the Internet). Packet headers not only contain address information, but may also include a number of different fields that will display information about the packet. Many headers contain error-checking information (checksum, Cyclic Redundancy Check) that enables the receiver to check if the packet has had any errors due to interference, such as electrical noise.

1.9 Duty Cycles

Duty cycle is defined as " the time that is effectively used to send or receive the data, expressed as a percentage of total period of time." The more the duty cycle , the more effective transmission or reception.

We can define the pulse width, τ, as being the time that a bit occupies from within its total alloted bit-time T_{b}. If we have a duty cycle of D, we can define the pulse width as:

$$\tau = DT_b$$

Where:

$$0 < \tau \leq T_b$$

The pulse width is equal to the bit time if we are using a 100% duty cycle.

1.10 Introduction

It turns out that many wires have a much higher bandwidth than is needed for the signals that they are currently carrying. Analog Telephone transmissions, for instance, require only 3 000 Hz of bandwidth to transmit human voice signals. Over short distances, however, twisted-pair telephone wire has an available bandwidth of nearly 100000 Hz!

There are several terrestrial radio based communications systems deployed today. They include:

- Cellular radio
- Mobile radio
- Digital microwave radio

Mobile radio service was first introduced in the St. Louis in 1946. This system was essentially a radio dispatching system with an operator who was able to patch the caller to the PSTN via a switchboard. Later, an improved mobile telephone system, IMTS, allowed customers to dial their own calls without the need for an operator. This in turn developed into the cellular radio networks we see today.

The long haul PSTNs and packet data networks use a wide variety of transmission media including

- Terrestrial microwave
- Satellite microwave
- Fiber optics
- Coaxial cable

In this section, we will be concerned with terrestrial microwave systems. Originally, microwave links used FDM exclusively as the access technique, but recent developments are changing analog systems to digital where TDM is more appropriate.

1.10.1 Fixed Access Assignment

Three basic methods can be used to combine customers on to fixed channel radio links:

- FDMA - (frequency division multiple access) analog or digital
- TDMA - (time division multiple access) three conversation paths are time division multiplexed in 6.7 mSec time slots on a single carrier.
- CDMA - (code division multiple access) this uses spread spectrum techniques to increase the subscriber density. The transmitter hops through a pseudo-random sequence of frequencies. The receiver is given the sequence list and is able to follow the transmitter. As more customers are added to the system, the signal to noise will gradually degrade. This is in contrast to AMPS where customers are denied access once all of the frequencies are assigned code division multiple access [digital only]
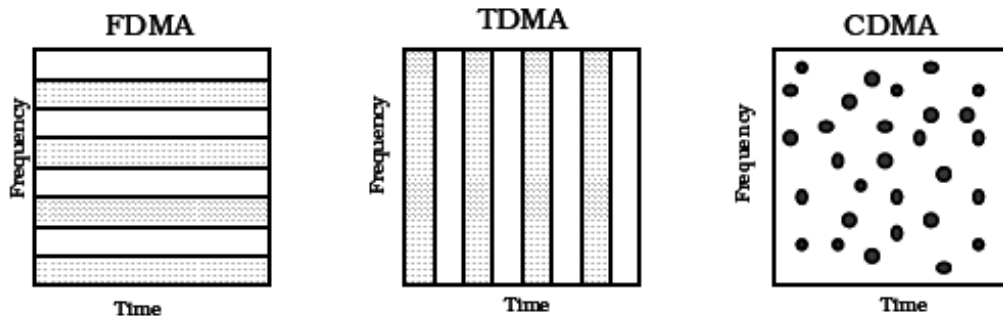
Figure 22

1.11 What is FDM?

w: Frequency-division_multiplexing[5]

Frequency Division Multiplexing (FDM) allows engineers to utilize the extra space in each wire to carry more than one signal. By frequency-shifting some signals by a certain amount, engineers can shift the spectrum of that signal up into the unused band on that wire. In this way, multiple signals can be carried on the same wire, without having to divy up time-slices as in Time-Division Multiplexing schemes.In analog transmission, signals are commonly multiplexed using frequency-division multiplexing (FDM), in which the carrier bandwidth is divided into subchannels of different frequency widths, each carrying a signal at the same time in parallel

> Broadcast radio and television **channels** are separated in the frequency spectrum using FDM. Each individual channel occupies a finite frequency range, typically some multiple of a given base frequency.

Traditional terrestrial microwave and satellite links employ FDM. Although FDM in telecommunications is being reduced, several systems will continue to use this technique, namely: broadcast & cable TV, and commercial & cellular radio.

1.11.1 Analog Carrier Systems

The standard telephony voice band [300 – 3400 Hz] is heterodyned and stacked on high frequency carriers by single sideband amplitude modulation. This is the most bandwidth efficient scheme possible.
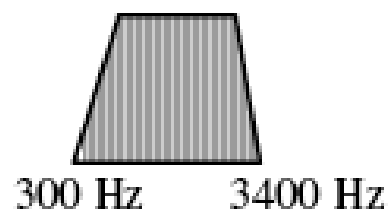
Figure 23

The analog voice channels are pre-grouped into threes and heterodyned on carriers at 12, 16, and 20 kHz. The resulting upper sidebands of four such pregroups are then heterodyned on carriers at 84, 96, 108, and 120 kHz to form a 12-channel group.

---

5   http://en.wikipedia.org/wiki/%20Frequency-division_multiplexing

Since the lower sideband is selected in the second mixing stage, the channel sequence is reversed and a frequency inversion occurs within each channel.
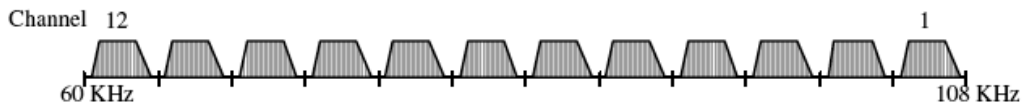
Figure 24

This process can continue until the available bandwidth on the coaxial cable or microwave link is exhausted.

Analog Carrier Systems			
# Voice Channels	Bandwidth	Spectrum	Terminology
12	48 KHz	60 – 108 KHz	Group
60	240 KHz	312 – 552 KHz	Supergroup
300	1.232 MHz	812 – 2044 KHz	CCITT Mastergroup
600	2.520 MHz	564 – 3084 KHz	Mastergroup
900	3.872 MHz	8.516 – 12.388 MHz	CCITT Supermastergroup
3600	16.984 MHz	.564 – 17.548 MHz	Jumbogroup
10800	57.442 MHz	3.124 – 60.566 MHz	

Figure 25

In the North American system, there are:

- 12 channels per group
- 5 groups per supergroup
- 10 super groups per mastergroup
- 6 master groups per jumbogroup

In the European CCITT system, there are:

- 12 channels per group
- 5 groups per supergroup
- 5 super groups per mastergroup
- 3 master groups per supermastergroup

There are other FDM schemes including:

- L600 - 600 voice channels 60–2788 kHz
- U600 - 600 voice channels 564–3084 kHz
- L3 - 1860 voice channels 312–8284 kHz, comprised of 3 mastergroups and a supergroup

- L4 - 3600 voice channels, comprised of six U600s

1.12 Benefits of FDM

FDM allows engineers to transmit multiple data streams simultaneously over the same channel, at the expense of bandwidth. To that extent, FDM provides a trade-off: faster data for less bandwidth. Also, to demultiplex an FDM signal requires a series of bandpass filters to isolate each individual signal. Bandpass filters are relatively complicated and expensive, therefore the receivers in an FDM system are generally expensive.

1.13 Examples of FDM

As an example of an FDM system, Commercial broadcast radio (AM and FM radio) simultaneously transmits multiple signals or "stations" over the airwaves. These stations each get their own frequency band to use, and a radio can be tuned to receive each different station. Another good example is cable television, which simultaneously transmits every channel, and the TV "tunes in" to which channel it wants to watch.

1.14 Orthogonal FDM

w:Orthogonal_frequency-division_multiplexing[6]

Orthogonal Frequency Division Multiplexing (OFDM) is a more modern variant of FDM that uses orthogonal sub-carriers to transmit data that does not overlap in the frequency spectrum and is able to be separated out using frequency methods. OFDM has a similar data rate to traditional FDM systems, but has a higher resilience to disruptive channel conditions such as noise and channel fading.

1.15 Voltage Controlled Oscillators (VCO)

w:Voltage-controlled_oscillator[7]

A **voltage-controlled oscillator** (VCO) is a device that outputs a sinusoid of a frequency that is a function of the input voltage. VCOs are *not* time-invariant, linear components. A complete study of how a VCO works will have to be relegated to a more detailed section based on feedback and oscillators. This page will, however, attempt to answer some of the basic questions about VCOs.

A basic VCO has input/output characteristics as such:

---

6   http://en.wikipedia.org/wiki/Orthogonal_frequency-division_multiplexing
7   http://en.wikipedia.org/wiki/Voltage-controlled_oscillator

```
v(t) ----|VCO|----> sin(a[f + v(t)]t + O)
```

VCOs are often implemented using a special type of diode called a "Varactor". Varactors, when reverse-biased, produce a small amount of capacitance that varies with the input voltage.

1.16 Phase-Locked Loops

w:Phase-locked_loop[8]

If you are talking on your cellphone, and you are walking (or driving), the phase angle of your signal is going to change, as a function of your motion, at the receiver. This is a fact of nature, and is unavoidable. The solution to this then, is to create a device which can "find" a signal of a particular frequency, negate any phase changes in the signal, and output the clean wave, phase-change free. This device is called a Phase-Locked Loop (PLL), and can be implemented using a VCO.

1.17 Purpose of VCO and PLL

VCO and PLL circuits are highly useful in modulating and demodulating systems. We will discuss the specifics of how VCO and PLL circuits are used in this manner in future chapters.

1.18 Varactors

As a matter of purely professional interest, we will discuss varactors here.

1.19 Further reading

- Clock and data recovery[9] has detailed information about designing and analyzing PLLs. (VCO)

1.20 What is an Envelope Filter?

If anybody has some images that they can upload, it would be much better then these ASCII art things.

---

8   http://en.wikipedia.org/wiki/Phase-locked_loop
9   http://en.wikipedia.org/wiki/Clock%20and%20data%20recovery

The envelope detector is a simple analog circuit that can be used to find the peaks in a quickly-changing waveform. Envelope detectors are used in a variety of devices, specifically because passing a sinusoid through an envelope detector will suppress the sinusoid.

1.21 Circuit Diagram

In essence, an envelope filter has the following diagram:

```
o------+------+------o
+      |      |      +
       \      (c)
vin    /R     |      vout
       \      |
-      |      |      -
o------+------+------o
```

Where (c) represents a capacitor, and R is a resistor. Under zero input voltage (vin = 0), the capacitor carries no charge, and the resistor carries no current. When vin is increased, the capacitor stores charge until it reaches capacity, and then the capacitor becomes an open circuit. At this point, all current in the circuit is flowing through the resistor, R. As voltage decreases, the capacitor begins to discharge it's stored energy, slowing down the state change in the circuit from high voltage to low voltage.

1.22 Positive Voltages

By inserting a diode at the beginning of this circuit, we can negate the effect of a sinusoid, dipping into negative voltage, and forcing the capacitor to discharge faster:

```
   diode
o-->|--+------+------o
+      |      |      +
       \      (c)
vin    /R     |      vout
       \      |
-      |      |      -
o------+------+------o
```

1.23 Purpose of Envelope Filters

Envelope filters help to find the outer bound of a signal that is changing in amplitude.

Envelope Filters are generally used with AM demodulation, discussed later.

(Envelope Detectors)

Modulation is a term that is going to be used very frequently in this book. So much in fact, that we could almost have renamed this book "Principals of Modulation", without having to delete too many chapters. So, the logical question arises: What exactly is modulation?

1.24 Definition

Modulation is a process of mixing a signal with a sinusoid to produce a new signal. This new signal, conceivably, will have certain benefits over an un-modulated signal, especially during transmission. If we look at a general function for a sinusoid:

$$f(t) = A\sin(\omega t + \phi)$$

we can see that this sinusoid has 3 parameters that can be altered, to affect the shape of the graph. The first term, A, is called the magnitude, or amplitude of the sinusoid. The next term, ω is known as the frequency, and the last term, ϕ is known as the phase angle. All 3 parameters can be altered to transmit data.

The sinusoidal signal that is used in the modulation is known as the **carrier signal**, or simply "the carrier". The signal that is used in modulating the carrier signal(or sinusoidal signal) is known as the "data signal" or the "message signal". It is important to notice that a simple sinusoidal carrier contains no information of its own.

In other words we can say that modulation is used because the some data signals are not always suitable for direct transmission, but the modulated signal may be more suitable.

1.25 Types of Modulation

There are 3 basic types of modulation: Amplitude modulation, Frequency modulation, and Phase modulation.

amplitude modulation

a type of modulation where the amplitude of the carrier signal is modulated (changed) in proportion to the message signal while the frequency and phase are kept constant.

frequency modulation

a type of modulation where the frequency of the carrier signal is modulated (changed) in proportion to the message signal while the amplitude and phase are kept constant.

phase modulation

a type of modulation where the phase of the carrier signal is varied accordance to the low frequency of the message signal is known as phase modulation.

1.26 Why Use Modulation?

Clearly the concept of modulation can be a little tricky, especially for the people who don't like trigonometry. Why then do we bother to use modulation at all? To answer this question, let's consider a channel that essentially acts like a bandpass filter: both the lowest frequency components and the highest frequency components are attenuated or unusable in some way, with transmission only being practical over some intermediate frequency range. If we can't send low-frequency signals, then we need to shift our signal up the frequency ladder. Modulation allows us to send a signal over a bandpass frequency range. If every signal gets its own frequency range, then we can transmit multiple signals *simultaneously* over a single channel, all using different frequency ranges.

Another reason to modulate a signal is to allow the use of a smaller antenna. A baseband (low frequency) signal would need a huge antenna because in order to be efficient, the antenna needs to be about 1/10th the length of the wavelength. Modulation shifts the baseband signal up to a much higher frequency, which has much smaller wavelengths and allows the use of a much smaller antenna.

1.27 Examples

Think about your car radio. There are more than a dozen (or so) channels on the radio at any time, each with a given frequency: 100.1MHz, 102.5MHz etc... Each channel gets a certain range (usually about 0.22MHz), and the entire station gets transmitted over that range. Modulation makes it all possible, because it allows us to send voice and music (which are essential baseband signals) over a bandpass (or "Broadband") channel.

1.28 non-sinusoidal modulation

A sine wave at one frequency can be separated from a sine wave at another frequency (or a cosine wave at the same frequency) because the two signals are "orthogonal".

There are other sets of signals, such that every signal in the set is orthogonal to every other signal in the set.

A simple orthogonal set is time multiplexed division (TDM) -- only one transmitter is active at any one time.

Other more complicated sets of orthogonal waveforms -- Walsh codes and various pseudo-noise codes such as Gold codes and maximum length sequences -- are also used in some communication systems.

The process of combining these waveforms with data signals is sometimes called "modulation", because it is so very similar to the way modulation combines sine waves are with data signals.

1.29 further reading

- Data Coding Theory/Spectrum Spreading[10]
- Wikipedia:Walsh code[11]
- Wikipedia:Gold code[12]
- Wikipedia:pseudonoise code[13]
- Wikipedia:maximum length sequence[14]

There is lots of talk nowadays about buzzwords such as "Analog" and "Digital". Certainly, engineers who are interested in creating a new communication system should understand the difference. Which is better, analog or digital? What is the difference? What are the pros and cons of each? This chapter will look at the answers to some of these questions.

1.30 What are They?

What exactly is an analog signal, and what is a digital signal?

Analog

 Analog signals are signals with continuous values. Analog signals are continuous in both time and value. Analog signals are used in many systems, although the use of analog signals has declined with the advent of cheap digital signals.

Digital

 Digital signals are discrete in time and value. Digital signals are signals that are represented by binary numbers, "1" or "0". The 1 and 0 values can correspond to different discrete voltage values, and any signal that *doesn't quite fit* into the scheme just gets rounded off.

Digital signals are sampled, quantized & encoded version of continuous time signals which they represent. In addition, some techniques also make the signal undergo encryption to make the system more tolerant to the channel.

1.31 What are the Pros and Cons?

Each paradigm has its own benefits and problems.

Analog

 Analog systems are less tolerant to noise, make good use of bandwidth, and are easy to manipulate mathematically. However, analog signals require hardware receivers and transmitters that are designed to perfectly fit the particular transmission. If you are

---

10   `http://en.wikibooks.org/wiki/Data%20Coding%20Theory%2FSpectrum%20Spreading`
11   `http://en.wikipedia.org/wiki/Walsh%20code`
12   `http://en.wikipedia.org/wiki/Gold%20code`
13   `http://en.wikipedia.org/wiki/pseudonoise%20code`
14   `http://en.wikipedia.org/wiki/maximum%20length%20sequence`

working on a new system, and you decide to change your analog signal, you need to completely change your transmitters and receivers.

Digital

Digital signals are more tolerant to noise, but digital signals can be completely corrupted in the presence of excess noise. In digital signals, noise could cause a 1 to be interpreted as a 0 and vice versa, which makes the received data different than the original data. Imagine if the army transmitted a position coordinate to a missile digitally, and a single bit was received in error? This single bit error could cause a missile to miss its target by miles. Luckily, there are systems in place to prevent this sort of scenario, such as checksums and CRCs, which tell the receiver when a bit has been corrupted and ask the transmitter to resend the data. The primary benefit of digital signals is that they can be handled by simple, standardized receivers and transmitters, and the signal can be then dealt with in software (which is comparatively cheap to change).

Discreet Digital and Analogue

Discreet data has a fixed set of possible values.

Digital data is a type of Discreet data where the fixed value can either be 1 or 0.

Analogue data can take on any real value.

1.32 Sampling and Reconstruction

The process of converting from analog data to digital data is called "sampling". The process of recreating an analog signal from a digital one is called "reconstruction". This book will not talk about either of these subjects in much depth beyond this, although other books on the topic of EE might, such as A-level Physics (Advancing Physics)/Digitisation[15].

1.33 further reading

- Electronics/Digital to Analog & Analog to Digital Converters[16]

Signals need a channel to follow, so that they can move from place to place. These **Communication Mediums**, or "channels" are things like wires and antennae that transmit the signal from one location to another. Some of the most common channels are listed below:

---

15 `http://en.wikibooks.org/wiki/A-level%20Physics%20%28Advancing%20Physics%29%2FDigitisation`

16 `http://en.wikibooks.org/wiki/Electronics%2FDigital%20to%20Analog%20%26%20Analog%20to%20Digital%20Converters`

1.34 Twisted Pair Wire

w:Twisted Pair[17] **Twisted Pair** is a transmission medium that uses two conductors that are twisted together to form a pair. The concept for the twist of the conductors is to prevent interference. Ideally, each conductor of the pair basically receives the same amount of interference, positive and negative, effectively cancelling the effect of the interference. Typically, most inside cabling has four pairs with each pair having a different twist rate. The different twist rates help to further reduce the chance of crosstalk by making the pairs appear electrically different in reference to each other. If the pairs all had the same twist rate, they would be electrically identical in reference to each other causing crosstalk, which is also referred to as capacitive coupling. Twisted pair wire is commonly used in telephone and data cables with variations of categories and twist rates.

w:Shielded Twisted Pair[18] Other variants of Twisted Pair are the **Shielded Twisted Pair** cables. The shielded types operate very similar to the non-shielded variety, except that Shielded Twisted Pair also has a layer of metal foil or mesh shielding around all the pairs or each individual pair to further shield the pairs from electromagnetic interference. Shielded twisted pair is typically deployed in situations where the cabling is subjected to higher than normal levels of interference.

1.35 Coaxial Cable

w:Coaxial Cable[19] Another common type of wire is **Coaxial Cable**. Coaxial cable (or simply, "coax") is a type of cable with a single data line, surrounded by various layers of padding and shielding. The most common coax cable, common television cable, has a layer of wire mesh surrounding the padded core, that absorbs a large amount of EM interference, and helps to ensure a relatively clean signal is transmitted and received. Coax cable has a much higher bandwidth than a twisted pair, but coax is also significantly more expensive than an equal length of twisted pair wire. Coax cable frequently has an available bandwidth in excess of hundreds of megahertz (in comparison with the hundreds of kilohertz available on twisted pair wires).

Originally, Coax cable was used as the backbone of the telephone network because a single coaxial cable could hold hundreds of simultaneous phone calls by a method known as "Frequency Division Multiplexing" (discussed in a later chapter). Recently however, Fiber Optic cables have replaced Coaxial Cable as the backbone of the telephone network because Fiber Optic channels can hold many more simultaneous phone conversations (thousands at a time), and are less susceptible to interference, crosstalk, and noise then Coaxial Cable.

---

17   `http://en.wikipedia.org/wiki/Twisted%20Pair`
18   `http://en.wikipedia.org/wiki/Shielded%20Twisted%20Pair`
19   `http://en.wikipedia.org/wiki/Coaxial%20Cable`

1.36 Fiber Optics

w:Glass Fibers[20] **Fiber Optic** cables are thin strands of glass that carry pulses of light (frequently infrared light) across long distances. Fiber Optic channels are usually immune to common RF interference, and can transmit incredibly high amounts of data very quickly. There are 2 general types of fiber optic cable: single frequency cable, and multi-frequency cable. single frequency cable carries only a single frequency of laser light, and because of this there is no self-interference on the line. Single-frequency fiber optic cables can attain incredible bandwidths of many gigahertz. Multi-Frequency fiber optics cables allow a Frequency-Division Multiplexed series of signals to each inhabit a given frequency range. However, interference between the different signals can decrease the range over which reliable data can be transmitted.

1.37 Wireless Transmission

In **wireless transmission systems**, signals are propagated as Electro-Magnetic waves through free space. Wireless signals are transmitted by a transmitter, and received by a receiver. Wireless systems are inexpensive because no wires need to be installed to transmit the signal, but wireless transmissions are susceptible not only to EM interference, but also to physical interference. A large building in a city, for instance can interfere with cell-phone reception, and a large mountain could block AM radio transmissions. Also, WiFi internet users may have noticed that their wireless internet signals don't travel through walls very well.

There are 2 types of antennas that are used in wireless communications, **isotropic**, and **directional**.

1.37.1 Isotropic

People should be familiar with isotropic antennas because they are everywhere: in your car, on your radio, etc... Isotropic antennas are omni-directional in the sense that they transmit data out equally (or nearly equally) in all directions. These antennas are excellent for systems (such as FM radio transmission) that need to transmit data to multiple receivers in multiple directions. Also, Isotropic antennas are good for systems in which the direction of the receiver, relative to the transmitter is not known (such as cellular phone systems).

1.37.2 Directional

Directional antennas focus their transmission power in a single narrow direction range. Some examples of directional antennas are satellite dishes, and wave-guides. The downfall of the directional antennas is that they need to be pointed directly at the receiver all the time to maintain transmission power. This is useful when the receiver and the transmitter are not moving (such as in communicating with a geo-synchronous satellite).

---

20   http://en.wikipedia.org/wiki/Glass%20Fibers

1.38 Receiver Design

It turns out that if we know what kind of signal to expect, we can better receive those signals. This should be intuitive, because it is hard to find something if we don't know what precisely we are looking for. How is a receiver supposed to know what is data and what is noise, if it doesnt know what data looks like?

Coherent transmissions are transmissions where the receiver knows what type of data is being sent. Coherency implies a strict timing mechanism, because even a data signal may look like noise if you look at the wrong part of it. In contrast, noncoherent receivers don't know exactly what they are looking for, and therefore noncoherent communication systems need to be far more complex (both in terms of hardware and mathematical models) to operate properly.

This section will talk about coherent receivers, first discussing the "Simple Receiver" case, and then going into theory about what the optimal case is. Once we know mathematically what an optimal receiver should be, we then discuss two actual implementations of the optimal receiver.

It should be noted that the remainder of this book will discuss optimal receivers. After all, why would a communication's engineer use anything that is less than the best?

1.39 The Simple Receiver

A simple receiver is just that: simple. A general simple receiver will consist of a low-pass filter (to remove excess high-frequency noise), and then a sampler, that will select values at certain points in the wave, and interpolate those values to form a smooth output curve. In place of a sampler (for purely analog systems), a general envelope filter can also be used, especially in AM systems. In other systems, different tricks can be used to demodulate an input signal, and acquire the data. However simple receivers, while cheap, are not the best choice for a receiver. Occcasionally they are employed because of their price, but where performance is an issue, a better alternative receiver should be used.

1.40 The Optimal Receiver

Mathematically, Engineers were able to predict the structure of the optimal receiver. Read that sentence again: Engineers are able to design, analyze, and build the best possible receiver, for any given signal. This is an important development for several reasons. First, it means that no more research should go into finding a better receiver. The best receiver has already been found, after all. Second, it means any communications system will not be hampered (much) by the receiver.

1.40.1 Derivation

here we will attempt to show how the coherent receiver is derived.

1.40.2 Matched Receiver

The matched receiver is the logical conclusion of the optimal receiver calculation. The matched receiver convolves the signal with itself, and then tests the output. Here is a diagram:

```
   s(t)----->(Convolve with r(t))----->
```

This looks simple enough, except that convolution modules are often expensive. An alternative to this approach is to use a correlation receiver.

1.40.3 Correlation Receiver

The correlation receiver is similar to the matched receiver, instead with a simple switch: The multiplication happens first, and the integration happens second.

Here is a general diagram:

```
            r(t)
             |
             v
  s(t) ----->(X)----->(Integrator)--->
```
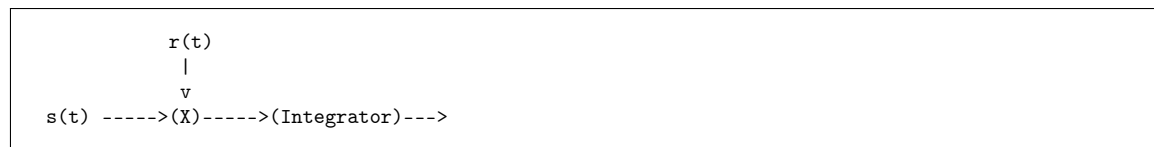
In a digital system, the integrator would then be followed by a threshold detector, while in an analog receiver, it might be followed by another detector, like an envelope detector.

1.41 Conclusion

To do the best job of receiving a signal, we need to know the form of the signal that we are sending. This should seem obvious, we can't design a receiver until after we've decided how the signal will be sent. This method poses some problems however, in that the receiver must be able to line up the received signal with the given reference signal to work the magic: If the received signal and the reference signal are out of sync with each other, either as a function of an error in phase or an error in frequency, then the optimal receiver will not work.

1.42 further reading

w:demodulation[21]

---

21   http://en.wikipedia.org/wiki/demodulation

2 Analog Modulation

2.1 Analog Modulation Overview

Let's take a look at a generalized sinewave:

$$x\left(t\right) = A\sin\left(\omega t + \theta\right)$$

It consists of three components namely; amplitude, frequency and phase. Each of which can be decomposed to provide finer detail:

$$x(t) = As(t)\sin(2\pi[f_c + kf_m(t)]t + \alpha\phi(t))$$

2.2 Types of Analog Modulation

We can see 3 parameters that can be changed in this sine wave to send information:

- $As(t)$. This term is called the "Amplitude", and changing it is called "Amplitude Modulation" (AM)
- $kf_m(t)$ This term is called the "Frequency Shift", and changing it is called "Frequency Modulation"
- $\alpha\phi(t)$. this term is called the "Phase angle", and changing it is called "Phase Modulation".
- The terms frequency and phase modulation are often combined into a more general group called "Angle Modulation".

2.3 The Breakdown

Each term consists of a coefficient (called a "scaling factor"), and a function of time that corresponds to the information that we want to send. The scaling factor out front, A, is also used as the transmission power coefficient. When a radio station wants their signal to be stronger (regardless of whether it is AM, FM, or PM), they "crank-up" the power of A, and send more power out onto the airwaves.

2.4 How we Will Cover the Material

We are going to go into separate chapters for each different type of modulation. This book will attempt to discuss some of the mathematical models and techniques used with

different modulation techniques. It will also discuss some practical information about how to construct a transmitter/receiver, and how to use each modulation technique effectively.

Amplitude modulation is one of the earliest radio modulation techniques. The receivers used to listen to AM-DSB-C are perhaps the simplest receivers of any radio modulation technique; which may be why that version of amplitude modulation is still widely used today. By the end of this module, you will know the most popular versions of amplitude modulation, some popular AM modulation circuits, and some popular AM demodulation circuits.

2.5 Amplitude Modulation

w:Amplitude_modulation[1]

Amplitude modulation (AM) occurs when the amplitude of a carrier wave is modulated, to correspond to a source signal. In AM, we have an equation that looks like this:

$$F_{signal}(t) = A(t)\sin(\omega t)$$

We can also see that the phase of this wave is irrelevant, and does not change (so we don't even include it in the equation).

ⓘ Information

AM Radio uses AM modulation

AM Double-Sideband (AM-DSB for short) can be broken into two different, distinct types: Carrier, and Suppressed Carrier varieties (AM-DSB-C and AM-DSB-SC, for short, respectively). This page will talk about both varieties, and will discuss the similarities and differences of each.

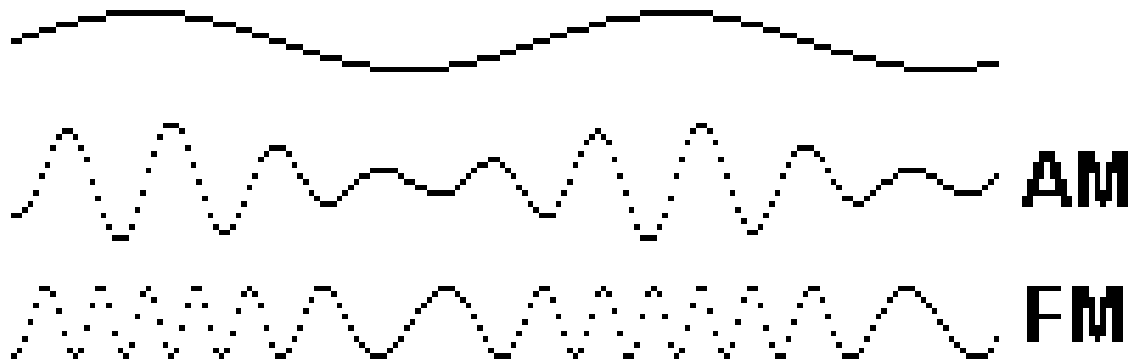1   http://en.wikipedia.org/wiki/Amplitude_modulation

Figure 26

2.5.1 Characteristics

Modulation Index

Amplitude modulation requires a high frequency constant carrier and a low frequency modulation signal.

A sine wave carrier is of the form $e_c = E_c \sin(\omega_c t)$

A sine wave modulation signal is of the form $e_m = E_m \sin(\omega_m t)$

Notice that the amplitude of the high frequency carrier takes on the shape of the lower frequency modulation signal, forming what is called a modulation envelope.
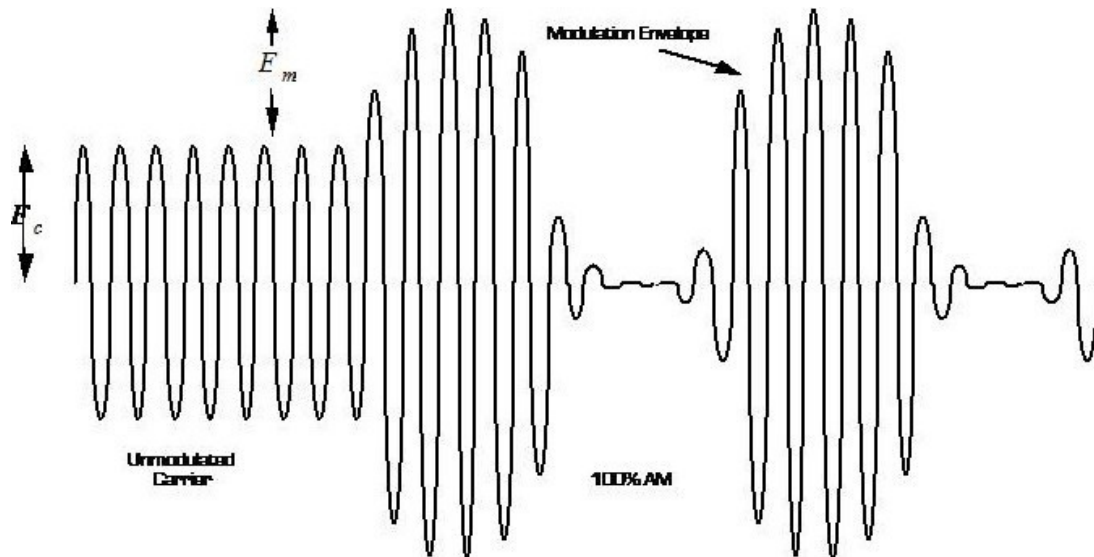
Figure 27

The modulation index is defined as the ratio of the modulation signal amplitude to the carrier amplitude.

$m_{am} = \frac{E_m}{E_c}$ where $0 \leq m_{am} \leq 1$

The overall signal can be described by:

$$e_{am} = (E_c + E_m \sin(\omega_m t)) \sin(\omega_c t)$$

More commonly, the carrier amplitude is normalized to one and the am equation is written as:

$$e_{am} = (1 + m_{am} \sin(\omega_m t)) \sin(\omega_c t)$$

In most literature this expression is simply written as:

$$e = (1 + m \sin \omega_m t) \sin \omega_c t$$

If the modulation index is zero ($m_{am} = 0$) the signal is simply a constant amplitude carrier.

If the modulation index is 1 ($m_{am} = 1$), the resultant waveform has maximum or 100% amplitude modulation.
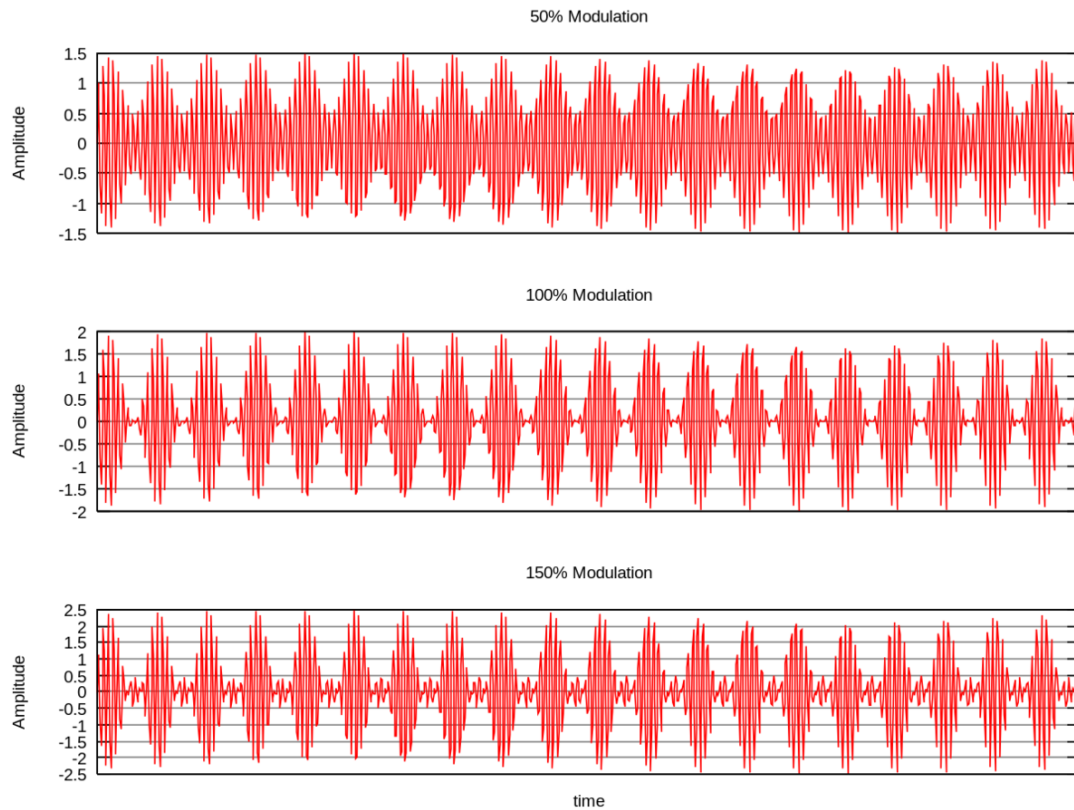
Figure 28

Sidebands

Expanding the normalized AM equation:

$$e = (1 + m\sin\omega_m t)\sin\omega_c t$$

we obtain:

$$e = \sin\omega_c t + \tfrac{m}{2}\cos(\omega_c - \omega_m)t - \tfrac{m}{2}\cos(\omega_c + \omega_m)t$$

where:

$\sin\omega_c t$ represents the carrier

$\tfrac{m}{2}\cos(\omega_c - \omega_m)t$ represents the lower sideband

$\tfrac{m}{2}\cos(\omega_c + \omega_m)t$ represents the upper sideband

The sidebands are centered on the carrier frequency. They are the sum and difference frequencies of the carrier and modulation signals. In the above example, they are just single frequencies, but normally the baseband modulation signal is a range of frequencies and hence two bands are formed.

2.5.2 AM Modulator

The standard amplitude modulation equation is:

$$e_{am} = (1 + m \sin \omega_m t) \sin \omega_c t$$

From this we notice that AM involves a multiplication process. There are several ways to perform this function electronically. The simplest method uses a switch.

Switching Modulators

Switching modulators can be placed into two categories: unipolar and bipolar.

Bipolar Switching Modulator

The bipolar switch is the easiest to visualize. Note that an AM waveform appears to consist of a low frequency dc signal whose polarity is reversing at a carrier rate.
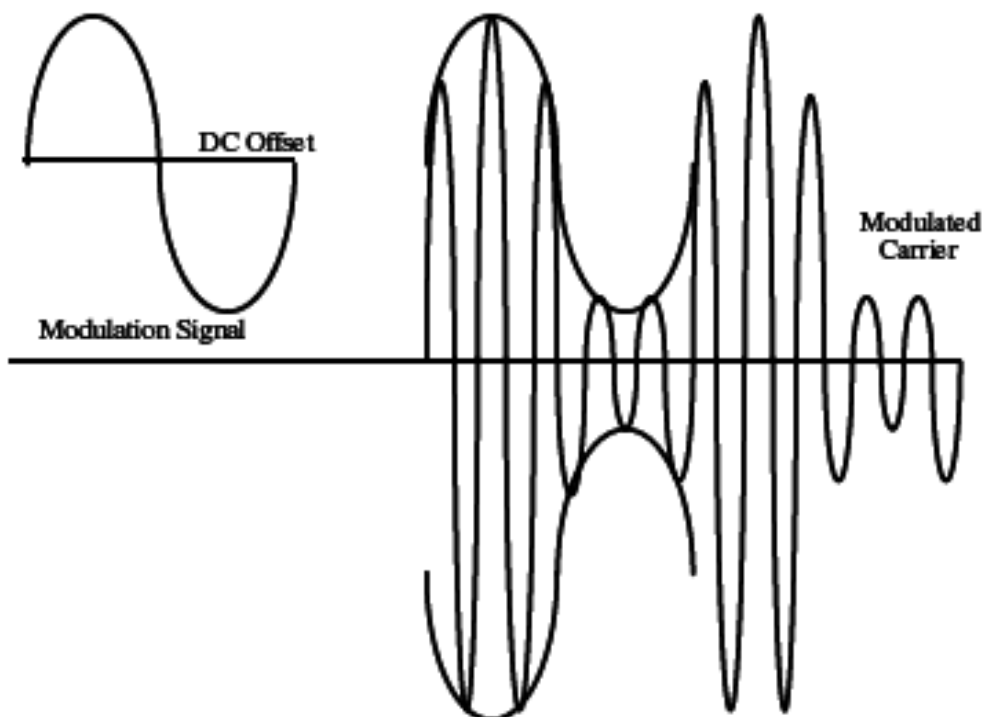

Figure 29

The AM signal can be created by multiplying a dc modulation signal by ± 1.
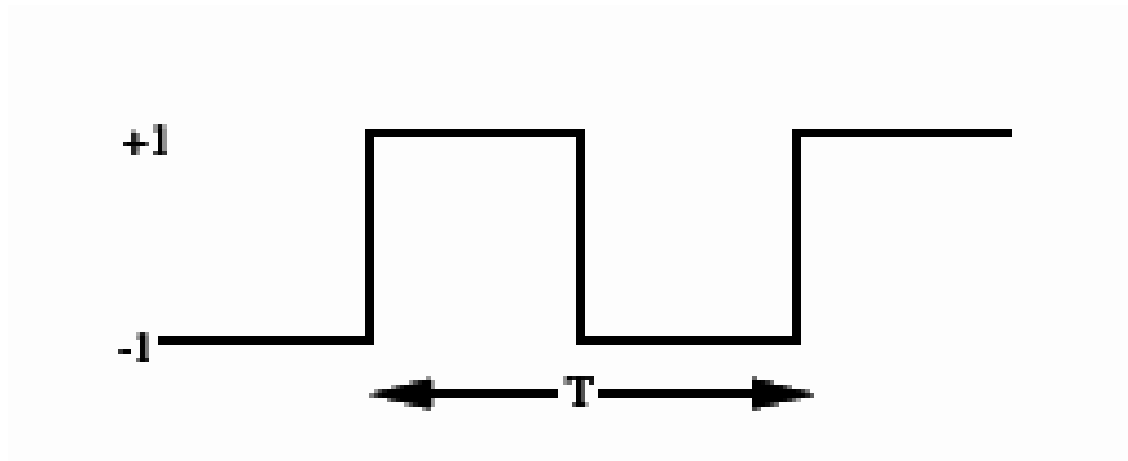
Figure 30

However, since the square wave contains lots of harmonics, the resulting multiplication will contain lots of extraneous frequencies. Mathematically, the spectrum of the square wave signal (given by the Fourier Transform) is of the form:

$$F\{f(t)\} = \sum_{n=1}^{\infty} \frac{4}{n\pi} \sin\left(\frac{n\pi}{2}\right) \cos\left(\frac{n2\pi t}{T}\right)$$

This seems complicated but, if the square wave switching function has a 50% duty cycle, this simplifies to:

$$F\{f(t)\} = \frac{4}{\pi} \sum_{n=1,3,5...}^{\infty} \frac{1}{n} \cos\left(\frac{n2\pi t}{T}\right)$$

This tells us that the square wave is actually composed of a series of cosines (phase shifted sines) at odd multiples of the fundamental switching frequency. Therefore, using this signal to multiply the baseband signal results in AM signals being generated at each of the odd harmonics of the switching (carrier) frequencies. Since the amplitude of the harmonics decreases rapidly, this technique is practical for only the first few harmonics, and produces an enormous amount of unwanted signals (noise).
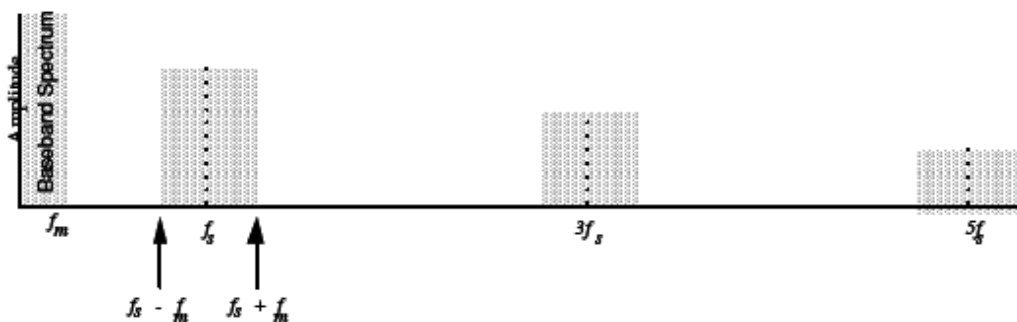

Figure 31

A band pass filter can be used to select any one of the AM signals. The number of different output frequencies can be significantly reduced if the multiplier accepts sinewaves at the carrier input.

w:Double-sideband suppressed-carrier transmission[2]

Removing the DC component from the input eliminates the carrier signal and creates DSBSC modulation.

Physically this is done by reversing the signal leads:

Reverse at the carrier rate
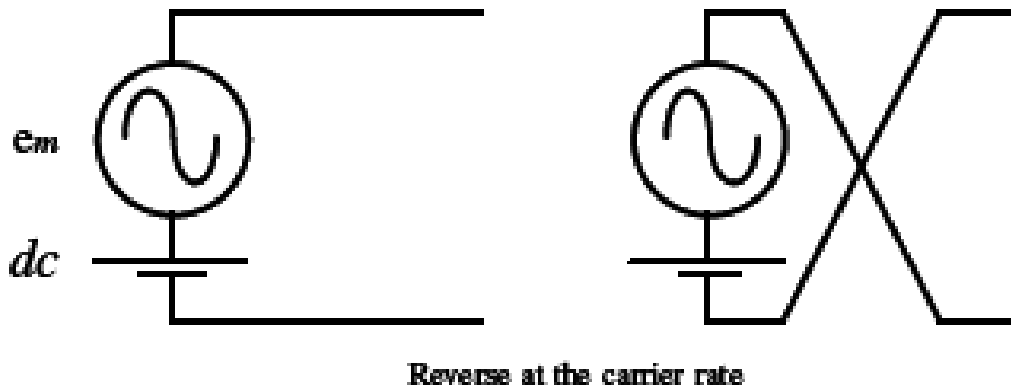
Figure 32

The process of reversing the polarity of a signal is easily accomplished by placing two switch pairs in the output of a differential amplifier. The Mc1496 Balanced Modulator[3] is an example of such a device.

2   http://en.wikipedia.org/wiki/Double-sideband%20suppressed-carrier%20transmission
3   http://www.onsemi.com/pub_link/Collateral/MC1496-D.PDF
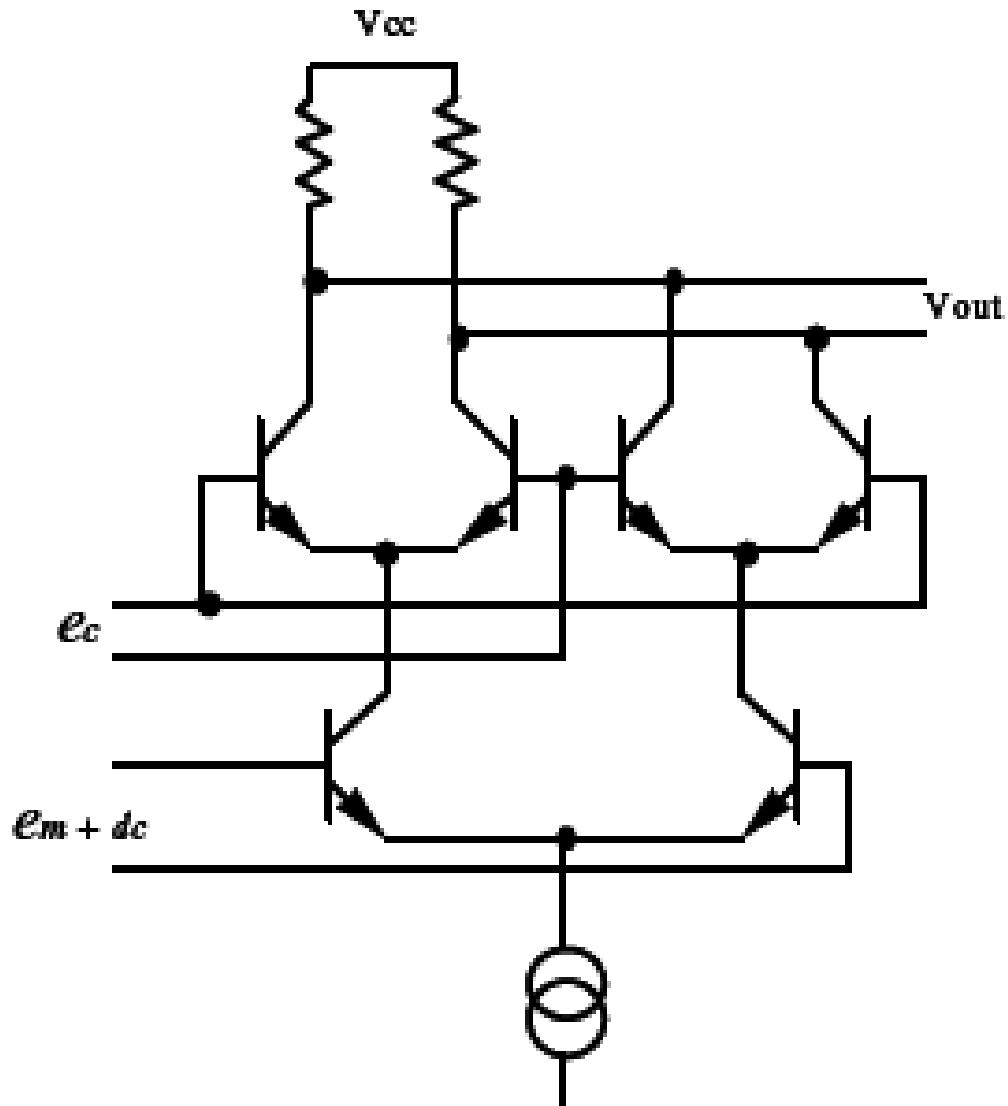
Figure 33

Unipolar Switching Modulator

As previously mentioned, an AM signal can be created by multiplying a dc modulation signal by 0 & 1.
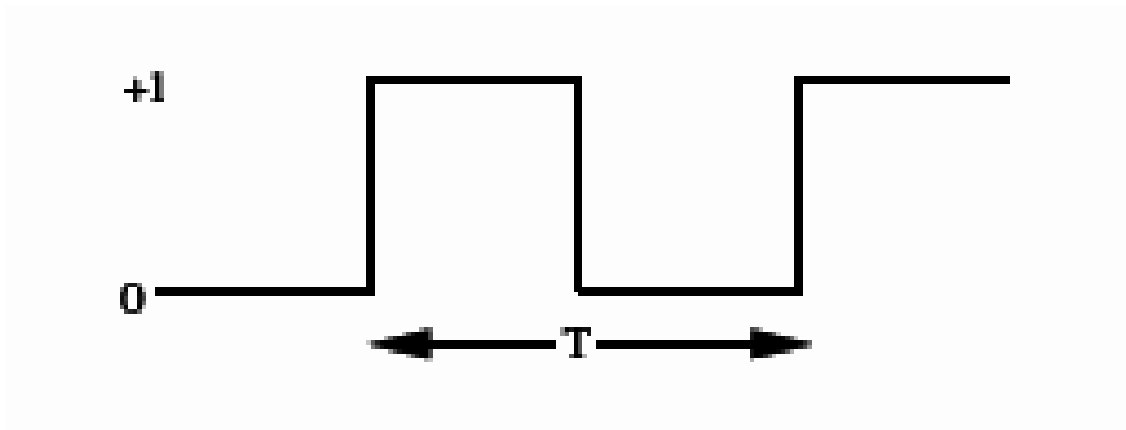
Figure 34

The spectrum of this signal is defined by:

$$F\{f(t)\} = \frac{1}{2} + \sum_{n=1}^{\infty} \frac{2}{n\pi} \sin\left(\frac{n\pi}{2}\right) \cos\left(\frac{n2\pi t}{T}\right)$$

Physically this is done by turning the modulation signal on and off at the carrier rate:

Gate the modulation at the carrier rate
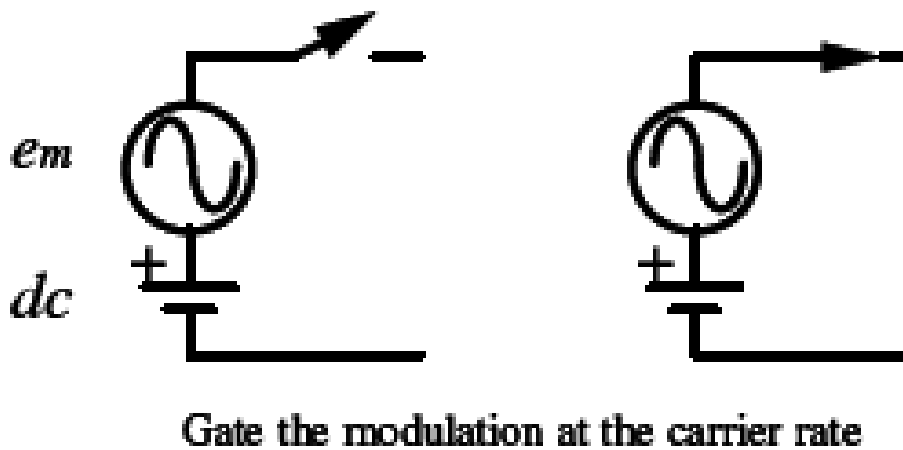
Figure 35

A high amplitude carrier can be used to turn a diode on and off. A dc bias is placed on the modulation signal to make certain that only the carrier (not the modulation signal) can reverse bias the diode.

Figure 36

Before the diode

After the diode
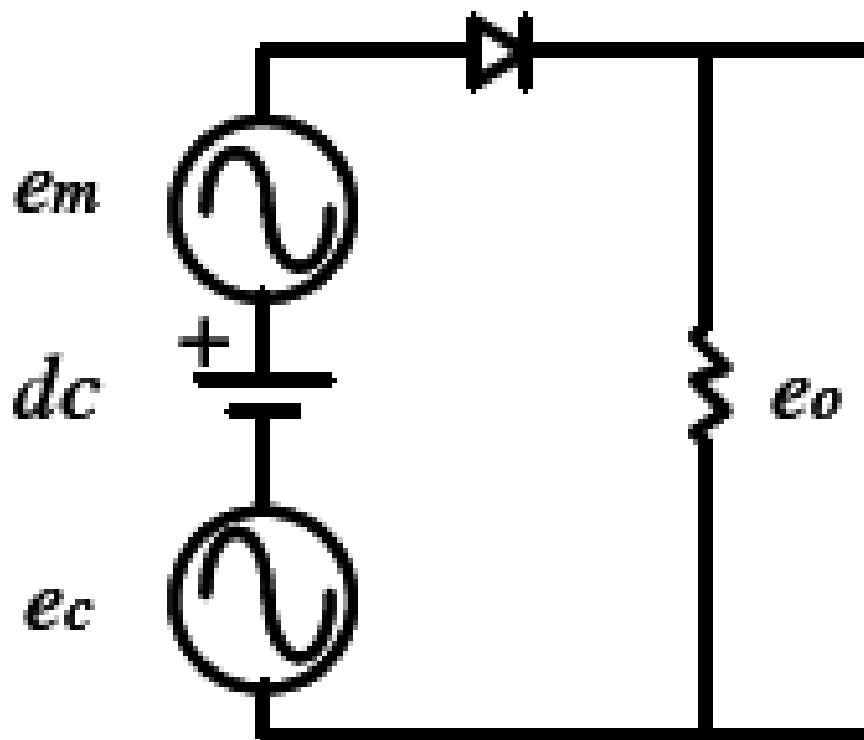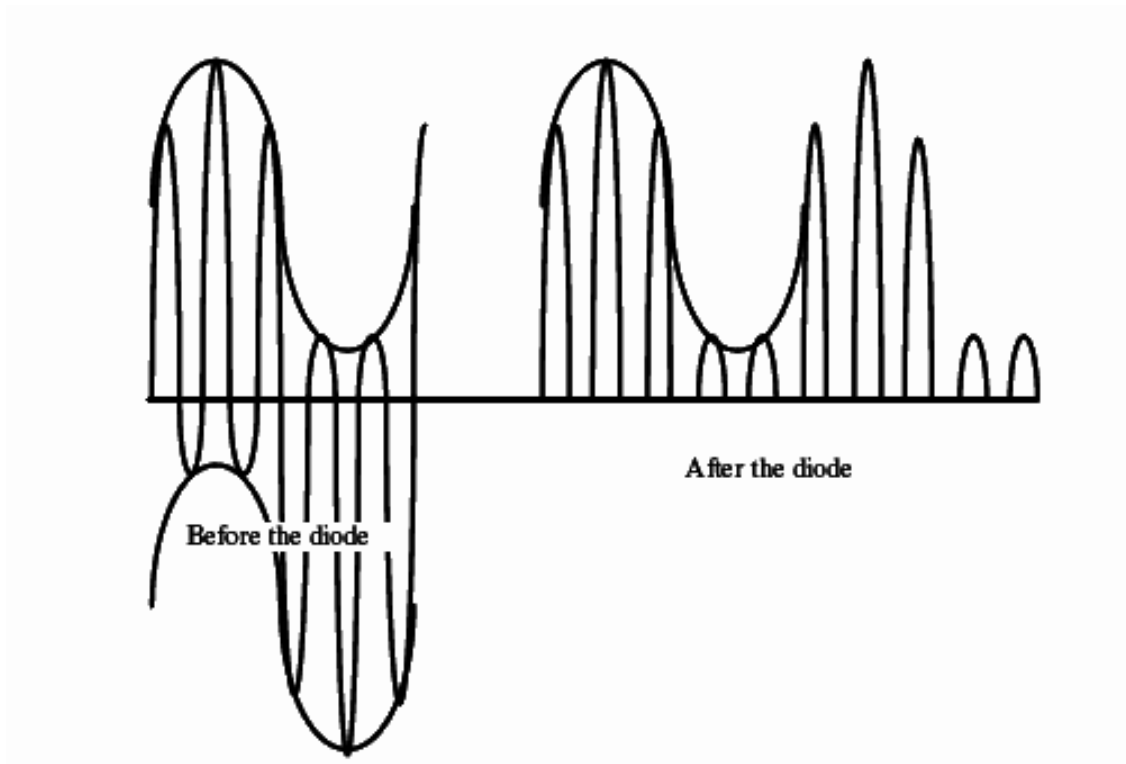
Figure 37

It may not seem obvious, but the output of this circuit contains a series of AM signals. A bandpass filter is needed to extract the desired one. Normally it is the 1st or 3rd harmonic of the fundamental. (The 1st harmonic is the fundamental.)

Collector Modulator

The diode switching modulator is incapable of producing high power signals since it is a passive device. A transistor can be used to overcome this limitation. A collector modulator is used for high level modulation.
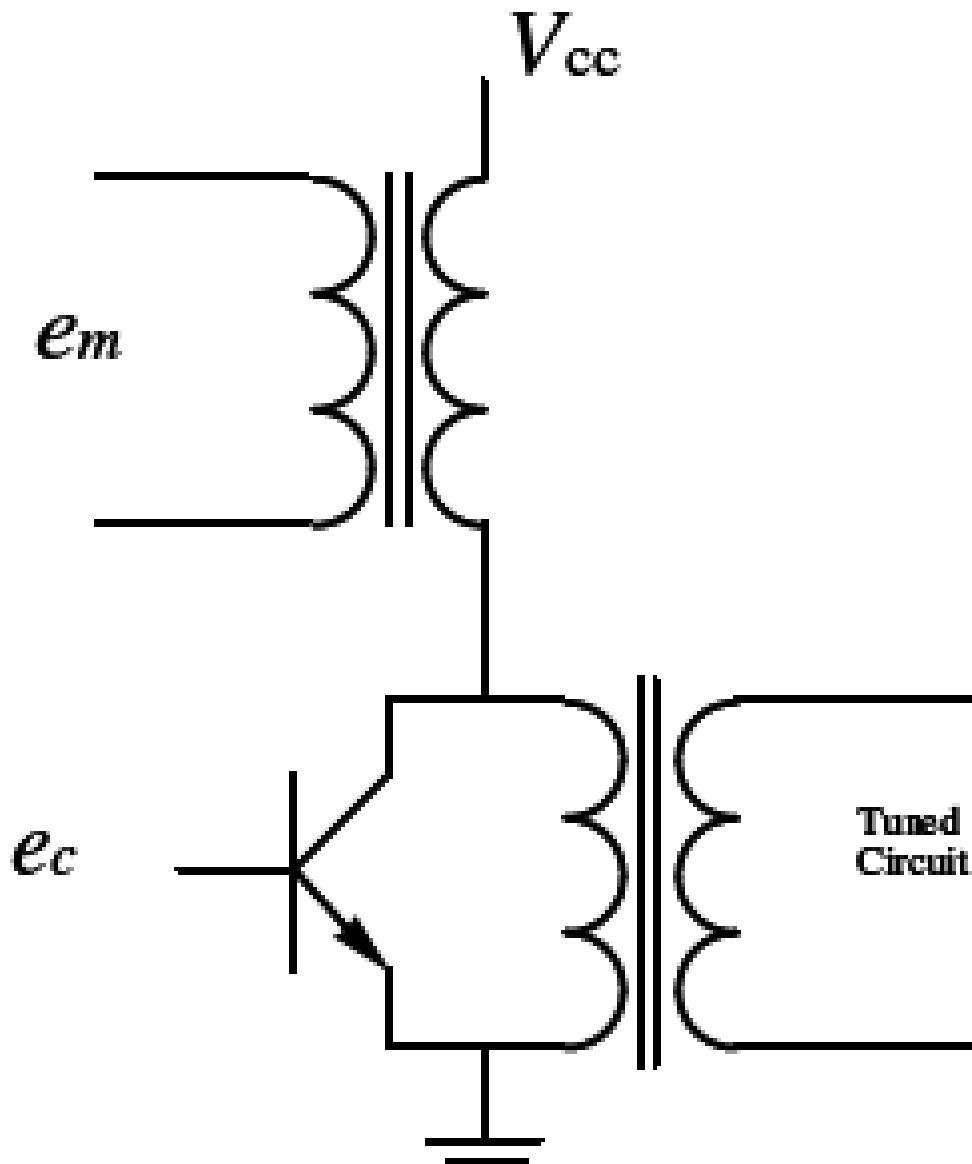
Figure 38

Square Law Modulator

The voltage-current relationship of a diode is nonlinear near the knee and is of the form:

$$i(t) = av(t) + bv^2(t)$$

The coefficient a and b are constants associated with the particular diode.
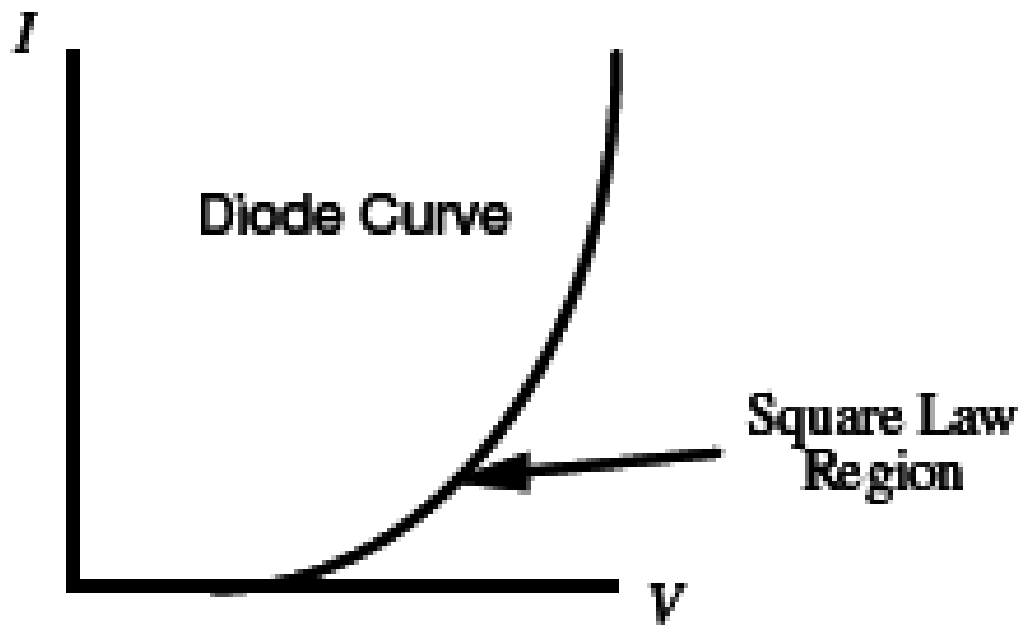
Figure 39

Amplitude modulation occurs if the diode is kept in the square law region when signals combine.
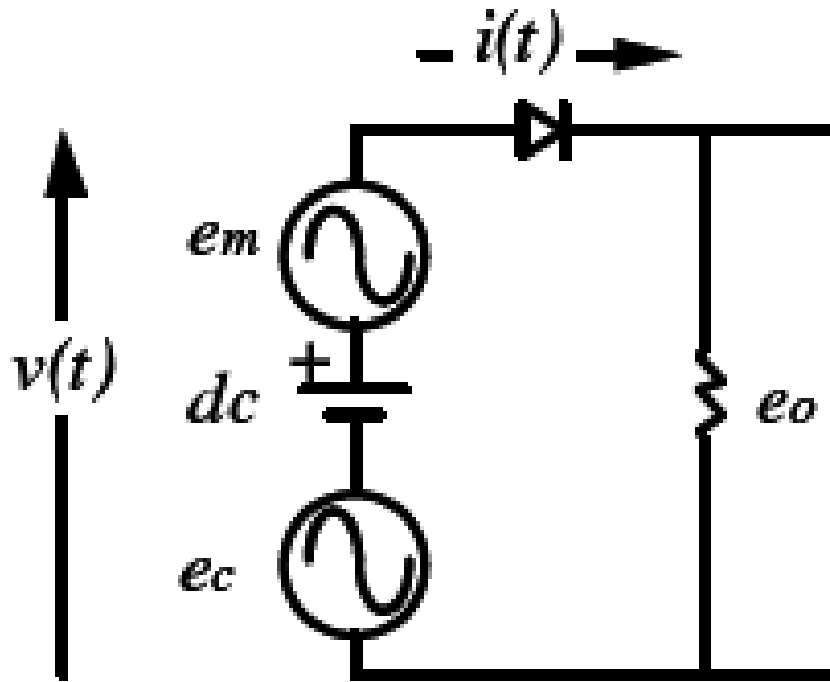
Figure 40

Let the injected signals be of the form:

$$k = \text{dcbias}$$

$$e_m = E_m \sin\omega_m t = \text{modulationsignal}$$

$$e_c = E_c \sin\omega_c t = \text{carriersignal}$$

The voltage applied across the diode and resistor is given by:

$$v(t) = k + e_m + e_c$$

The current in the diode and hence in the resistor is given by:

$$i(t) = a(k + e_m + e_c) + b(k + e_m + e_c)^2$$

Which expands to:

$$i(t) = \underbrace{k(a\_bk)}_{\text{dc}} + \underbrace{(a+2bk)e_m}_{\text{originalmodulatingsignal}} + \underbrace{(a+2bk)e_c}_{\text{carrier}} + \underbrace{2be_m e_c}_{\text{2sidebands}} + \underbrace{be_m^2}_{\text{2xmodulatingfrequency}} + \underbrace{be_c^2}_{\text{2xcarrierfrequency}}$$

Modulation Index Measurement

It is sometimes difficult to determine the modulation index, particularly for complex signals. However, it is relatively easy to determine it by critical observation. There are two practical methods to derive the modulation index. 1)By representing a AM wave as it is in time domain.(using maxima - minima terms.) 2)By Trapezoidal method.
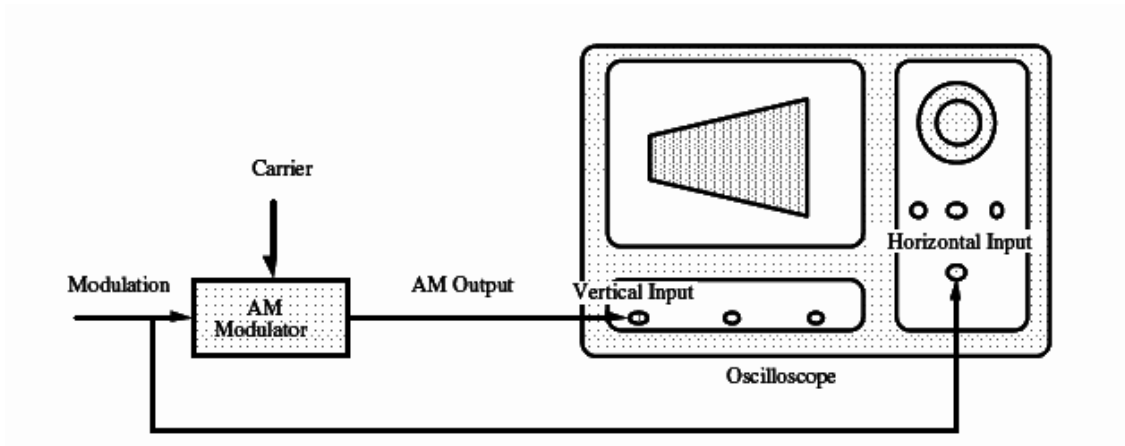
Figure 41

The trapezoidal oscilloscope display can be used to determine the modulation index.
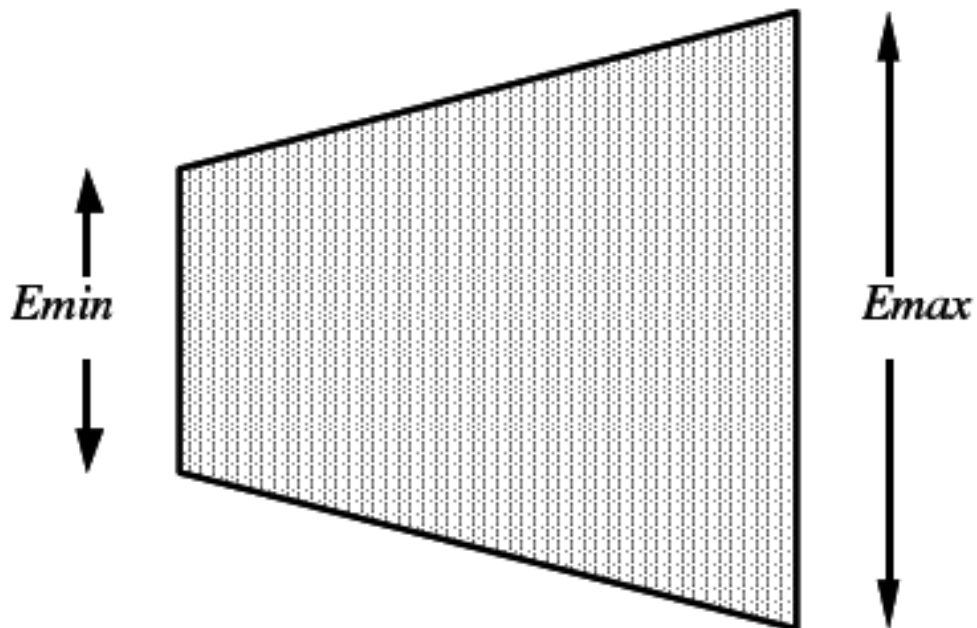
Figure 42

AM modulation index: $m = \frac{E_{\max} - E_{\min}}{E_{\max} + E_{\min}}$

The trapezoidal display makes it possible to quickly recognize certain types of problems, which would reduce the AM signal quality.
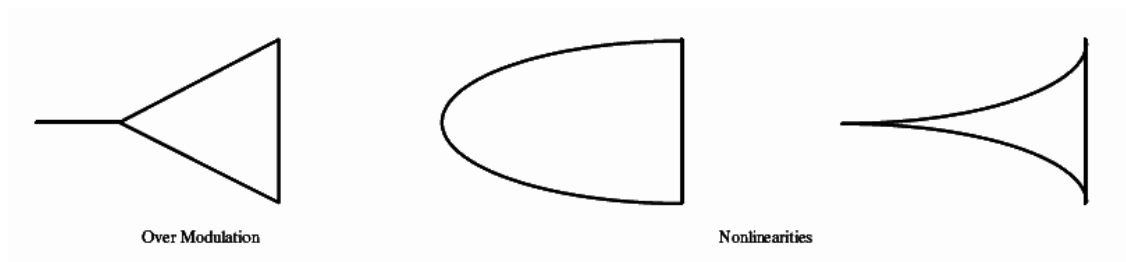
Figure 43

The highest authorized carrier power for AM broadcast in the US is 50 kilowatts, although directional stations are permitted 52.65 kilowatts to compensate for losses in the phasing system. The ERP can be much higher

C-QUAM

The basic idea behind the C-Quam modulator is actually quite simple. The output stage is an ordinary AM modulator however; the carrier signal has been replaced by an amplitude limited vector modulator. Therefore, the limiter output is really a phase-modulated signal.
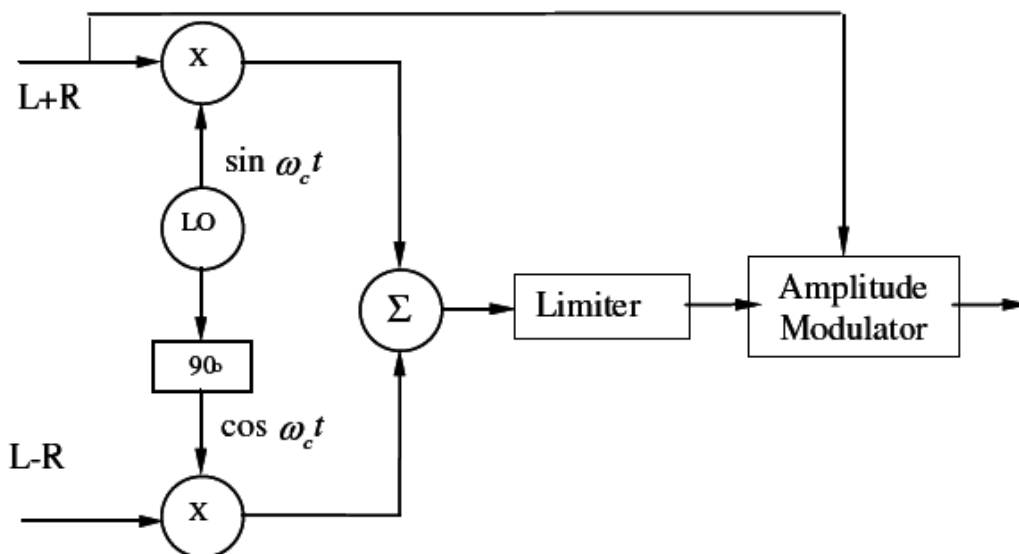
Figure 44

A standard AM receiver will detect the amplitude variations as L+R. A stereo receiver will also detect the phase variations and to extract L-R. It will then process these signals to separate the left and right channels.

To enable the stereo decoder, a 25 Hz pilot tone is added to the L-R channel.

2.5.3 AM Receivers

The most common receivers in use today are the super heterodyne type. They consist of:

- Antenna
- RF amplifier
- Local Oscillator and Mixer
- IF Section
- Detector and Amplifier

The need for these subsystems can be seen when one considers the much simpler and inadequate TRF or tuned radio frequency amplifier.

TRF Amplifier

It is possible to design an RF amplifier to accept only a narrow range of frequencies, such as one radio station on the AM band.
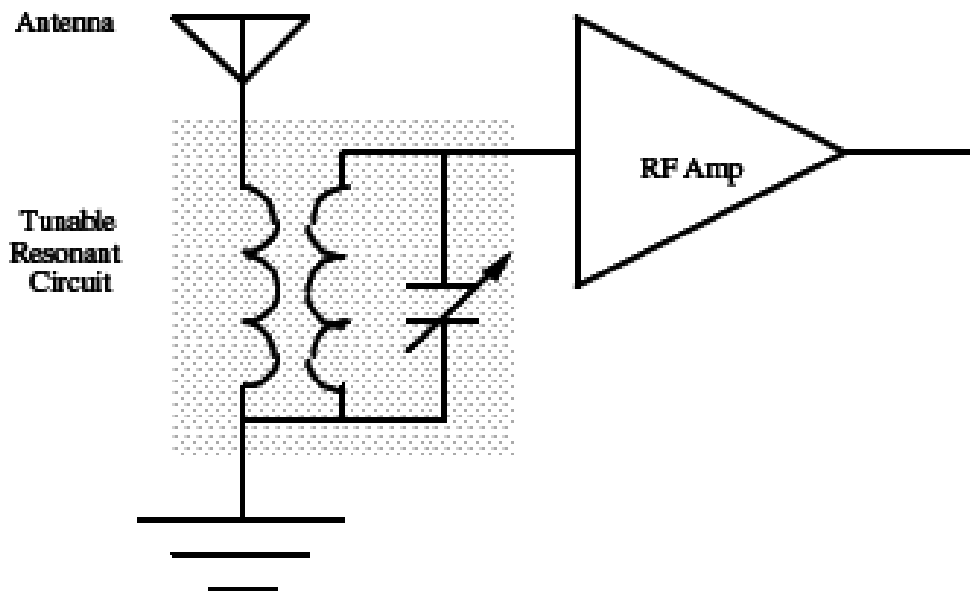
Figure 45

By adjusting the center frequency of the tuned circuit, all other input signals can be excluded.
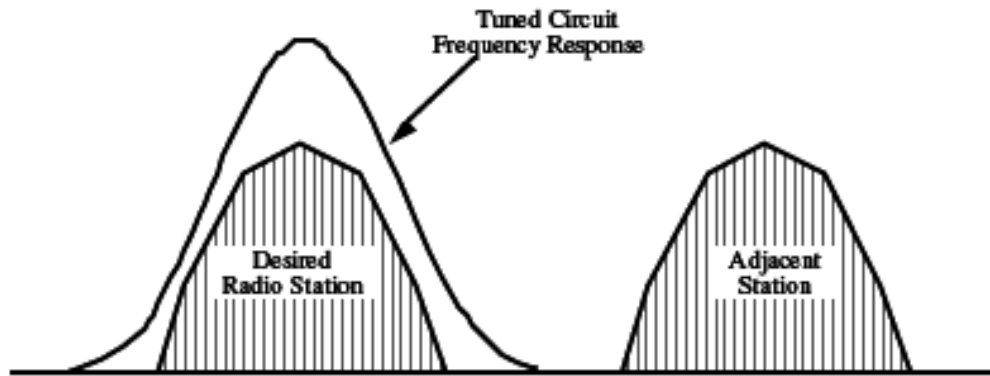
Figure 46

The AM band ranges from about 500 kHz to 1600 kHz. Each station requires 10 kHz of this spectrum, although the baseband signal is only 5 kHz.

Recall that for a tuned circuit: $Q = \frac{f_c}{B}$. The center or resonant frequency in an RLC network is most often adjusted by varying the capacitor value. However, the Q remains approximately constant as the center frequency is adjusted. This suggests that as the bandwidth varies as the circuit is tuned.

For example, the Q required at the lower end of the AM band to select only one radio station would be approximately:

$$Q = \frac{f_c}{B} = \frac{500kHz}{10kHz} = 50$$

As the tuned circuit is adjusted to the higher end of the AM band, the resulting bandwidth is:

$$B = \frac{f_c}{Q} = \frac{1600kHz}{50} = 30kHz$$

A bandwidth this high could conceivably pass three adjacent stations, thus making meaningful reception impossible.

To prevent this, the incoming RF signal is heterodyned to a fixed IF or intermediate frequency and passed through a constant bandwidth circuit.

Superheterodyne Receiver
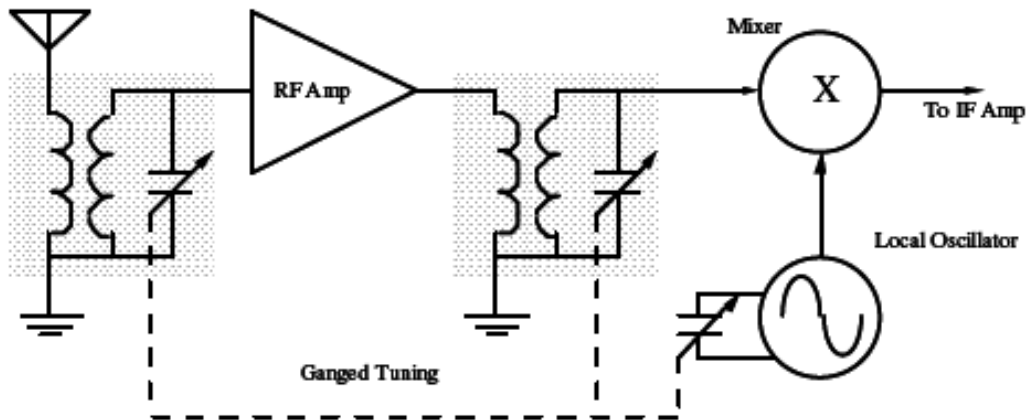
Figure 47

The RF amplifier boosts the RF signal into the mixer. It has broad tuning and amplifies not just one RF station, but many of them simultaneously. It also amplifies any input noise and even contributes some of its own.

The other mixer input is a high frequency sine wave created by a local oscillator. In AM receivers, it is always 455 kHz above the desired station carrier frequency. An ideal mixer will combine the incoming carrier with the local oscillator to create sum and difference frequencies. .

A real mixer combines two signals and creates a host of new frequencies:

- A dc level

- The original two frequencies

- The sum and difference of the two input frequencies

- Harmonics of the two input frequencies

- Sums and differences of all of the harmonics

Since the RF amplifier passes several radio stations at once, the mixer output can be very complex. However, the only signal of real interest is the difference between the desired station carrier frequency and the local oscillator frequency. This difference frequency, also called the IF (intermediate frequency) will alway be 455 kHz. By passing this through a 10 kHz BPF (band pass filter) centered at 455 kHz, the bulk of the unwanted signals can be eliminated.

Local Oscillator Frequency

Since the mixer generates sum and difference frequencies, it is possible to generate the 455 kHz IF signal if the local oscillator is either above or below the IF. The inevitable question is which is preferable.

Case I The local Oscillator is above the IF. This would require that the oscillator tune from (500 + 455) kHz to (1600 + 455) kHz or approximately 1 to 2 MHz. It is normally the capacitor in a tuned RLC circuit, which is varied to adjust the center frequency while the inductor is left fixed.

Since $f_c = \frac{1}{2\pi\sqrt{LC}}$

solving for C we obtain $C = \frac{1}{L(2\pi f_c)^2}$

When the tuning frequency is a maximum, the tuning capacitor is a minimum and vice versa. Since we know the range of frequencies to be created, we can deduce the range of capacitance required.

$$\frac{C_{\max}}{C_{\min}} = \frac{L\left(2\pi f_{\max}\right)^2}{L\left(2\pi f_{\min}\right)^2} = \left(\frac{2MHz}{1MHz}\right)^2 = 4$$

Making a capacitor with a 4:1 value change is well within the realm of possibility.

Case II The local Oscillator is below the IF. This would require that the oscillator tune from (500 - 455) kHz to (1600 - 455) kHz or approximately 45 kHz to 1145 kHz, in which case:

$$\frac{C_{\max}}{C_{\min}} = \left(\frac{1145kHz}{45kHz}\right)^2 \approx 648$$

It is not practical to make a tunable capacitor with this type of range. Therefore the local oscillator in a standard AM receiver is above the radio band.

Image Frequency

Just as there are two oscillator frequencies, which can create the same IF, two different station frequencies can create the IF. The undesired station frequency is known as the image frequency.
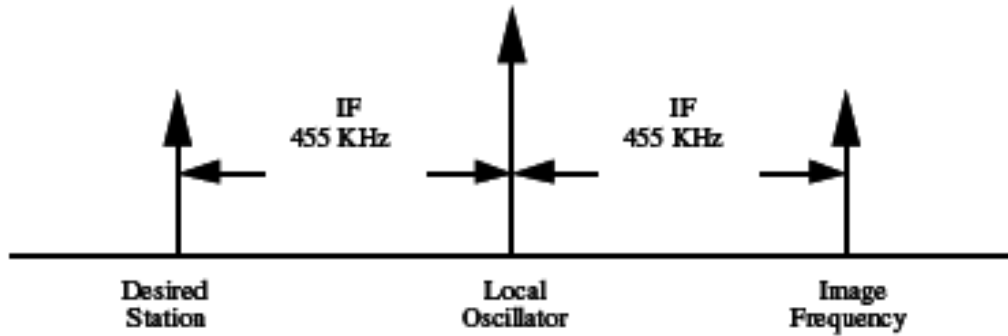
$$f_{image} = f_s + 2f_{IF} \text{ if } f_o > f_s$$

$$f_{image} = f_s - 2f_{IF} \text{ if } f_s > f_o$$

Figure 48

If any circuit in the radio front end exhibits non-linearities, there is a possibility that other combinations may create the intermediate frequency. Once the image frequency is in the mixer, there is no way to remove it since it is now heterodyned into the same IF band as the desired station.

2.6 AM Demodulation

AM Detection

There are two basic types of AM detection, coherent and non-coherent. Of these two, the non-coherent is the simpler method.

- Non-coherent detection does not rely on regenerating the carrier signal. The information or modulation envelope can be removed or detected by a diode followed by an audio filter.
- Coherent detection relies on regenerating the carrier and mixing it with the AM signal. This creates sum and difference frequencies. The difference frequency corresponds to the original modulation signal.

Both of these detection techniques have certain drawbacks. Consequently, most radio receivers use a combination of both.

Envelope Detector
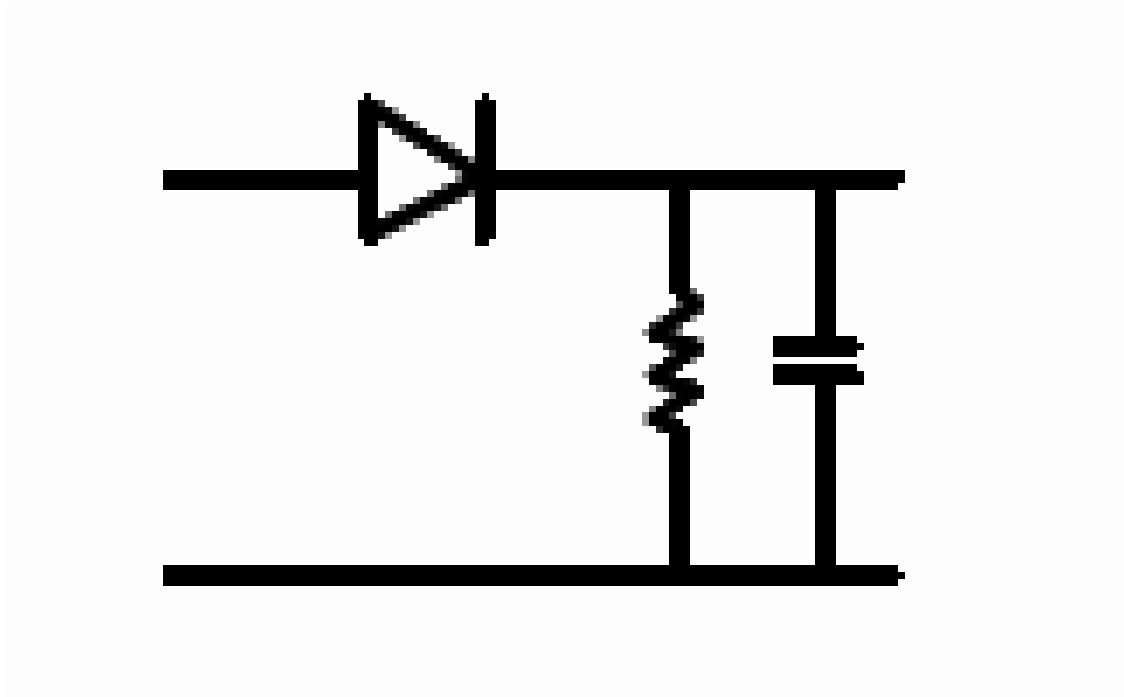
Figure 49

When trying to demodulate an AM signal, it seems like good sense that only the amplitude of the signal needs to be examined. By only examining the amplitude of the signal at any given time, we can remove the carrier signal from our considerations, and we can examine the original signal. Luckily, we have a tool in our toolbox that we can use to examine the amplitude of a signal: The Envelope Detector.

An envelope detector is simply a half wave rectifier followed by a low pass filter. In the case of commercial AM radio receivers, the detector is placed after the IF section. The carrier at this point is 455 kHz while the maximum envelope frequency is only 5 kHz. Since the ripple component is nearly 100 times the frequency of the highest baseband signal and does not pass through any subsequent audio amplifiers.

An AM signal where the carrier frequency is only 10 times the envelope frequency would have considerable ripple:
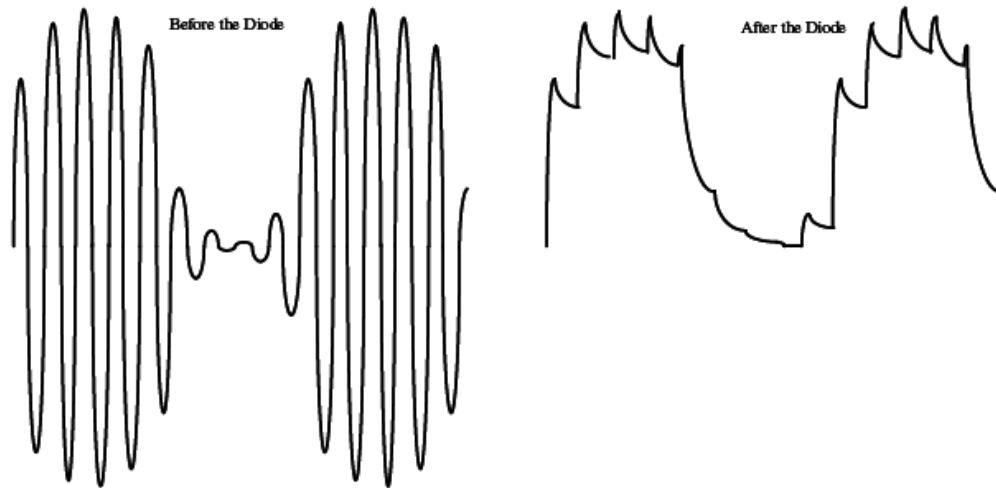
Figure 50

Synchronous Detector

In a synchronous or coherent detector, the incoming AM signal is mixed with the original carrier frequency.
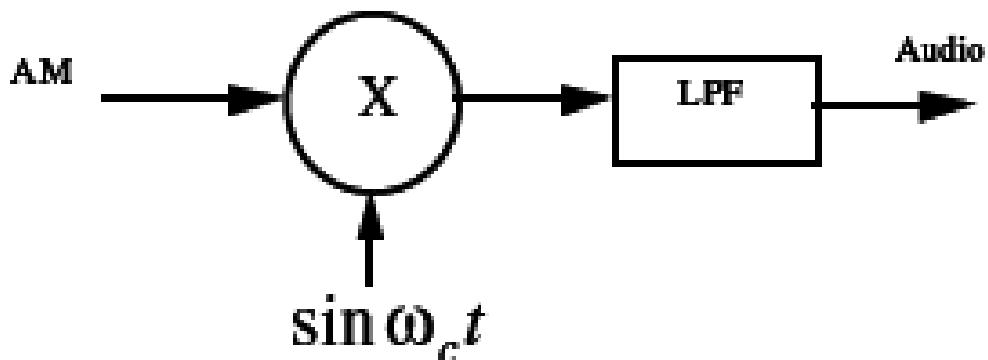
Figure 51

If you think this looks suspiciously like a mixer, you are absolutely right! A synchronous detector is one where the difference frequency between the two inputs is zero Hz. Of in other words, the two input frequencies are the same. Let's check the math.

Recall that the AM input is mathematically defined by:

$$e_{am} = \underbrace{\sin \omega_c t}_{\text{Carrier}} + \underbrace{\frac{m}{2} \sin \left(\omega_c - \omega_m \right) t}_{\text{LowerSideband}} - \underbrace{\frac{m}{2} \sin \left(\omega_c + \omega_m \right) t}_{\text{UpperSideband}}$$

At the multiplier output, we obtain:

$$\text{mixerout} = e_{am} \times \sin\omega_c t = \underbrace{-\frac{m}{2}\sin\omega_m t}_{\text{Original Modulation Signal}} \underbrace{-\frac{1}{2}\sin 2\omega_c t - \frac{m}{4}\sin\left(2\omega_c - \omega_m\right)t + \frac{m}{4}\sin\left(2\omega_c + \omega_m\right)t}_{\text{AM signal centered at 2 times the carrier frequency}}$$

The high frequency component can be filtered off leaving only the original modulation signal.

This technique has one serious drawback. The problem is how to create the exact carrier frequency. If the frequency is not exact, the entire baseband signal will be shifted by the difference. A shift of only 50 Hz will make the human voice unrecognizable. It is possible to use a PLL (phase locked loop), but making one tunable for the entire AM band is not trivial.

As a result, most radio receivers use an oscillator to create a fixed intermediate frequency. This is then followed by an envelope detector or a fixed frequency PLL.

Squaring Detector

The squaring detector is also a synchronous or coherent detector. It avoids the problem of having to recreate the carrier by simply squaring the input signal. It essentially uses the AM signal itself as a sort of wideband carrier.
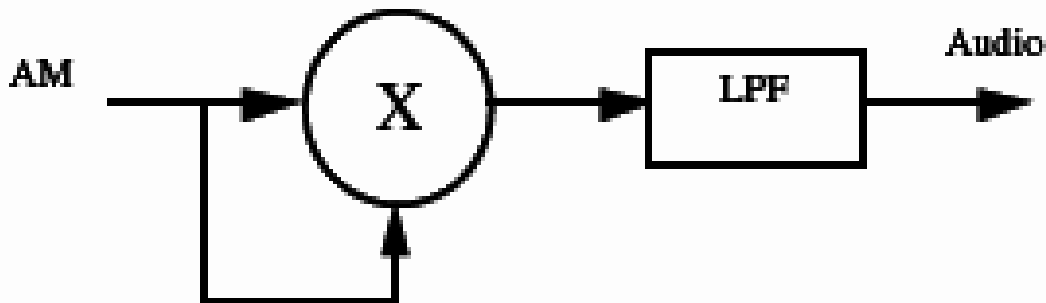
Figure 52

The output of the multiplier is the square of the input AM signal:

$$(e_{am})^2 = \left(\sin\omega_c t + \frac{m}{2}\cos\left(\omega_c - \omega_m\right)t - \frac{m}{2}\cos\left(\omega_c + \omega_m\right)t\right)^2$$

Since the input is being multiplied by the $\sin\omega_c t$ component, one of the resulting difference terms is the original modulation signal. The principle difficulty with this approach is trying to create a linear, high frequency multiplier.

2.7 AM-DSBSC

AM-DSB-SC is characterized by the following transmission equation:

$$v(t) = As(t)\cos(2\pi f_c t)$$

It is important to notice that s(t) can contain a negative value. AM-DSB-SC requires a coherent receiver, because the modulation data can go negative, and therefore the receiver needs to know that the signal is negative (and not just phase shifted). AM-DSB-SC systems are very susceptible to frequency shifting and phase shifting on the receiving end. In this equation, A is the transmission amplitude.

Double side band suppressed carrier modulation is simply AM without the broadcast carrier. Recall that the AM signal is defined by:

$$e_{am} = (1 + m\sin\omega_m t)\sin\omega_c t = \sin\omega_c t + \frac{m_2}{2}\cos\left(\omega_c - \omega_m\right)t - \frac{m_2}{2}\cos\left(\omega_c + \omega_m\right)t$$

The carrier term in the spectrum can be eliminated by removing the dc offset from the modulating signal:

$$e_{DSBSC} = m\sin\omega_m t\sin\omega_c t = \frac{m_2}{2}\cos\left(\omega_c - \omega_m\right)t - \frac{m_2}{2}\cos\left(\omega_c + \omega_m\right)t$$

2.7.1 Double Balanced Ring Modulator

One of the circuits which is capable of creating DSBSC is the double balance ring modulator.
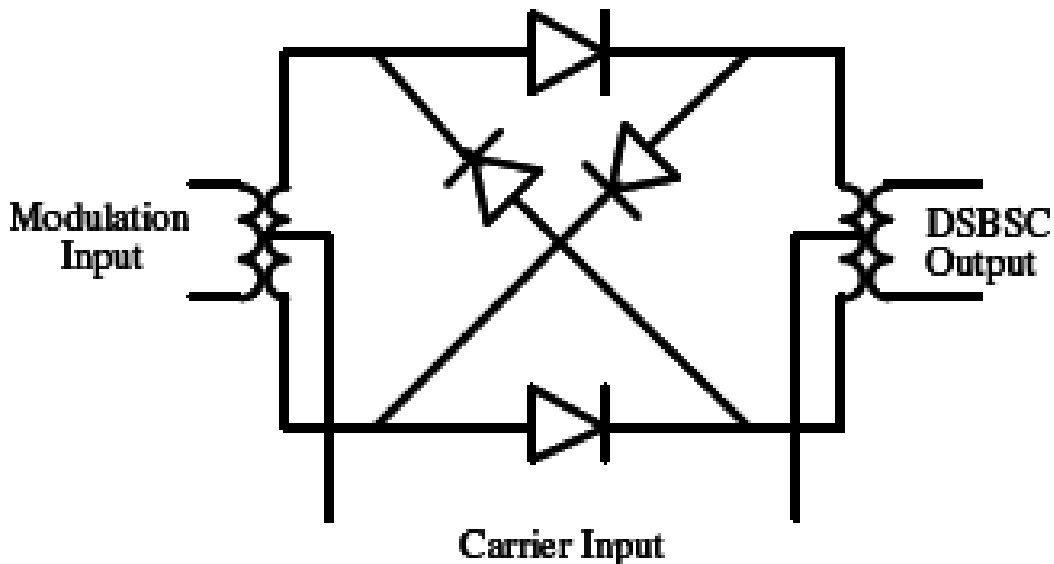
Figure 53

If the carrier is large enough to cause the diodes to switch states, then the circuit acts like a diode switching modulator:
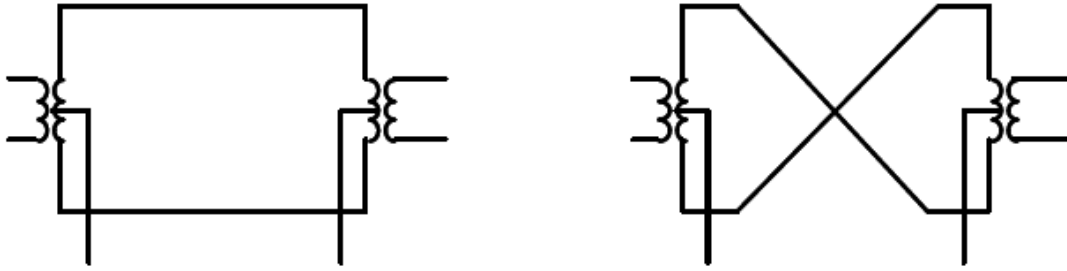
Figure 54

The modulation signal is inverted at the carrier rate. This is essentially multiplication by ± 1. Since the transformers cannot pass dc, there is no term which when multiplied can create an output carrier. Since the diodes will switch equally well on either cycle, the modulation signal is effectively being multiplied by a 50% duty cycle square wave creating numerous DSBSC signals, each centered at an odd multiple of the carrier frequency. Bandpass filters are used to extract the frequency of interest.

Some IC balanced modulators use this technique, but use transistors instead of diodes to perform the switching.

2.7.2 Push Pull Square Law Balanced Modulator
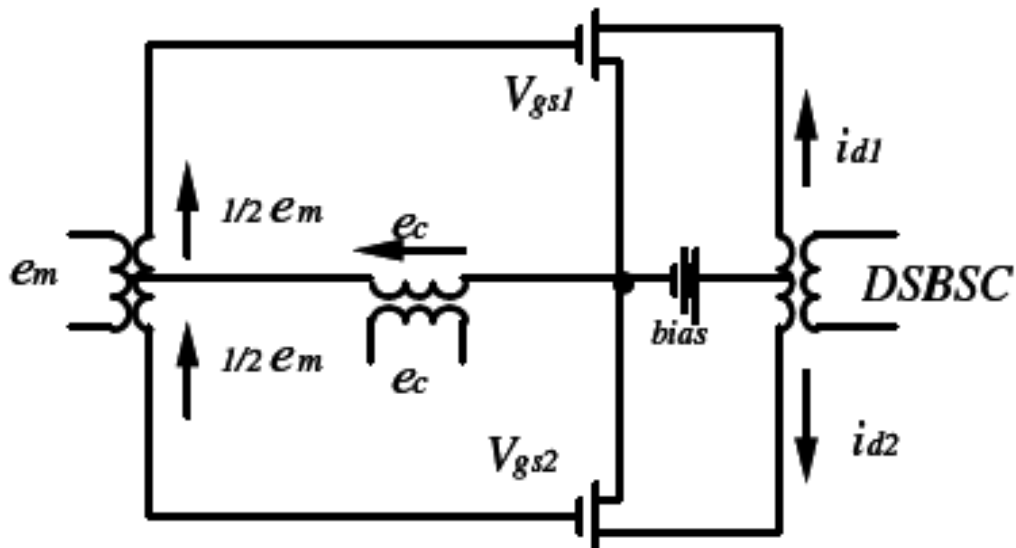
Figure 55

This circuit uses the same principles as the diode square law modulator. Since dc cannot pass through the transformer, it would be expected that there would be no output signal at the carrier frequency.

The drain current vs. gate-source voltage is of the form:

$$i_d = i_0 + av_{gs} + v_{gs}^2$$

The net drain current in the output transformer is given by:

$$i_{net} = i_{d1} - i_{d2} = i_0 + av_{gs1} + bv_{gs1}^2 - \left(i_0 + av_{gs2} + bv_{gs2}^2\right)$$

$$i_{net} = a\left(v_{gs1} - v_{gs2}\right) + b\left(v_{gs1} + v_{gs2}\right)\left(v_{gs1} - v_{gs2}\right)$$

By applying KVL around the gate loops we obtain:

$$v_{gs1} = \frac{e_m}{2} + e_c \qquad v_{gs2} = -\frac{e_m}{2} + e_c$$

Putting it all together we obtain:

$$i_{net} = a\left(\frac{e_m}{2} + e_c + \frac{e_m}{2} - e_c\right) + b\left(\frac{e_m}{2} + e_c - \frac{e_m}{2} + e_c\right)\left(\frac{e_m}{2} + e_c + \frac{e_m}{2} - e_c\right)$$

$$i_{net} = ae_m + 2be_ce_m$$

From this we note that the first term is the originating modulation signal and can easily be filtered off by a high pass filter. The second term is of the form:

$$\sin\omega_m t \sin\omega_c t = \frac{1}{2}\cos\left(\omega_c - \omega_m\right)t - \frac{1}{2}\cos\left(\omega_c + \omega_m\right)t$$

which is AM DSBSC.

2.8 AM-DSB-C

In contrast to AM-DSB-SC is AM-DSB-C, which is categorized by the following equation:

$$v(t) = A[s(t) + c]\cos(2\pi f_c t)$$

Where c is a positive term representing the carrier. If the term $[s(t) + c]$ is always non-negative, we can receive the AM-DSB-C signal non-coherently, using a simple envelope detector to remove the cosine term. The +c term is simply a constant DC signal and can be removed by using a blocking capacitor.

It is important to note that in AM-DSB-C systems, a large amount of power is wasted in the transmission sending a "boosted" carrier frequency. since the carrier contains no information, it is considered to be wasted energy. The advantage to this method is that it

greatly simplifies the receiver design, since there is no need to generate a coherent carrier signal at the receiver. For this reason, this is the transmission method used in conventional AM radio.

AM-DSB-SC and AM-DSB-C both suffer in terms of bandwidth from the fact that they both send two identical (but reversed) frequency "lobes", or bands. These bands (the upper band and the lower band) are exactly mirror images of each other, and therefore contain identical information. Why can't we just cut one of them out, and save some bandwidth? The answer is that we can cut out one of the bands, but it isn't always a good idea. The technique of cutting out one of the sidebands is called **Amplitude Modulation Single-Side-Band** (AM-SSB). AM-SSB has a number of problems, but also some good aspects. A compromise between AM-SSB and the two AM-DSB methods is called **Amplitude Modulation Vestigial-Side-Band** (AM-VSB), which uses less bandwidth then the AM-DSB methods, but more than the AM-SSB.

2.8.1 Transmitter

A typical AM-DSB-C transmitter looks like this:

```
                c     cos(...)
                |        |
  Signal ---->(+)---->(X)----> AM-DSB-C
```

which is a little more complicated than an AM-DSB-SC transmitter.

2.8.2 Receiver

An AM-DSB-C receiver is very simple:

```
   AM-DSB-C ---->|Envelope Filter|---->|Capacitor|----> Signal
```
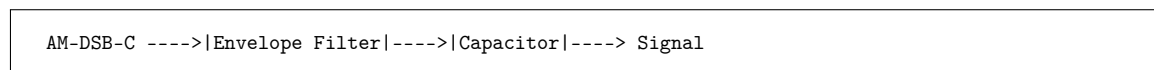
The capacitor blocks the DC component, and effectively removes the +c term.

2.9 AM-SSB

To send an AM-SSB signal, we need to remove one of the sidebands from an AM-DSB signal. This means that we need to pass the AM-DSB signal through a filter, to remove one of the sidebands. The filter, however, needs to be a very high order filter, because we need to have a very aggressive roll-off. One sideband needs to pass the filter almost completely unchanged, and the other sideband needs to be stopped completely at the filter.

To demodulate an AM-SSB signal, we need to perform the following steps:

1. Low-pass filter, to remove noise
2. Modulate the signal *again* by the carrier frequency
3. Pass through another filter, to remove high-frequency components

4. Amplify the signal, because the previous steps have attenuated it significantly.

AM-SSB is most efficient in terms of bandwidth, but there is a significant added cost involved in terms of more complicated hardware to send and receive this signal. For this reason, AM-SSB is rarely seen as being cost effective.

Single sideband is a form of AM with the carrier and one sideband removed. In normal AM broadcast, the transmitter is rated in terms of the carrier power. SSB transmitters attempt to eliminate the carrier and one of the sidebands. Therefore, transmitters are rated in PEP (peak envelope power).

$$PEP = \frac{(\text{peakenvelopevoltage})^2}{2R_L}$$

With normal voice signals, an SSB transmitter outputs 1/4 to 1/3 PEP.

There are numerous variations of SSB:

- SSB - Single sideband - amateur radio
- SSSC - Single sideband suppressed carrier - a small pilot carrier is transmitted
- ISB - Independent sideband - two separate sidebands with a suppressed carrier. Used in radio telephony.
- VSB - Vestigial sideband - a partial sideband. Used in broadcast TV.
- ACSSB - Amplitude companded SSB

There are several advantages of using SSB:

- More efficient spectrum utilization

- Less subject to selective fading

- More power can be placed in the intelligence signal

- 10 to 12 dB noise reduction due to bandwidth limiting

2.9.1 Filter Method

The simplest way to create SSB is to generate DSBSC and then use a bandpass filter to extract one of the sidebands.
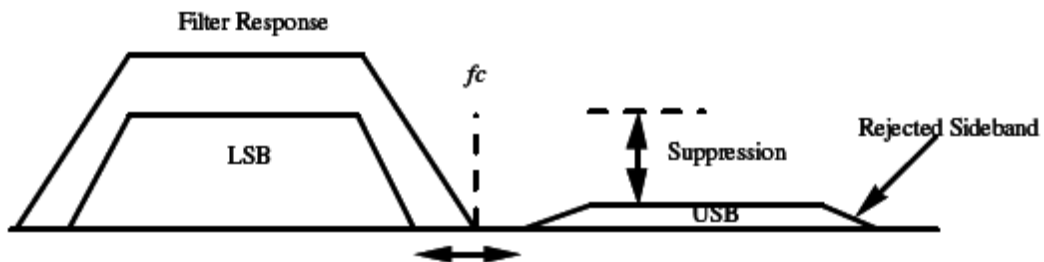
Figure 56

This technique can be used at relatively low carrier frequencies. At high frequencies, the Q of the filter becomes unacceptably high. The required Q necessary to filter off one of the sidebands can be approximated by:

$$Q \approx \frac{f_c \sqrt{S}}{4 \Delta f}$$

where:

$$f_c = \mathrm{carrier frequency}$$

$$\Delta f = \mathrm{sideband separation}$$

$$S = \mathrm{sideband suppression (not in dB)}$$

Several types of filters are used to suppress unwanted sidebands:

- LC - Maximum Q = 200
- Ceramic - Maximum Q = 2000
- Mechanical - Maximum Q = 10,000
- Crystal - Maximum Q = 50,000

In order to reduce the demands placed upon the filter, a double heterodyne technique can be used.
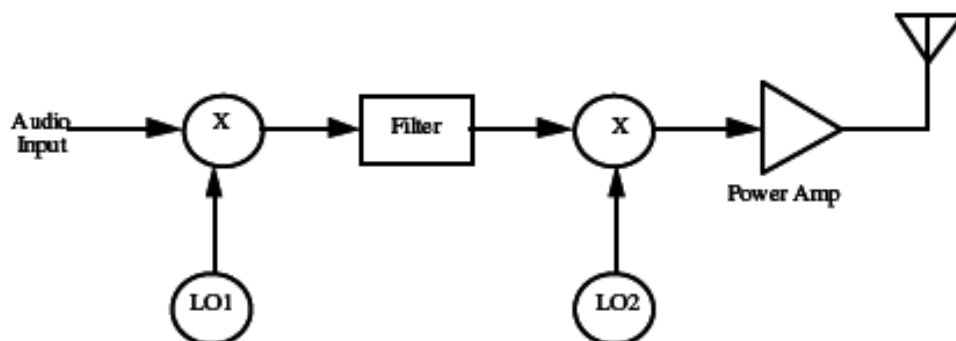
Figure 57

The first local oscillator has a relatively low frequency thus enabling the removal of one of the sidebands produced by the first mixer. The signal is then heterodyned a second time, creating another pair of sidebands. However, this time they are separated by a sufficiently large gap that one can be removed by the band limited power amplifier or antenna matching network.

Example

Observe the spectral distribution under the following conditions:

- Audio baseband = 100 HZ to 5 KHz

- LO1 = 100 kHz

- LO2 = 50 MHz

The spectral output of the first mixer is:
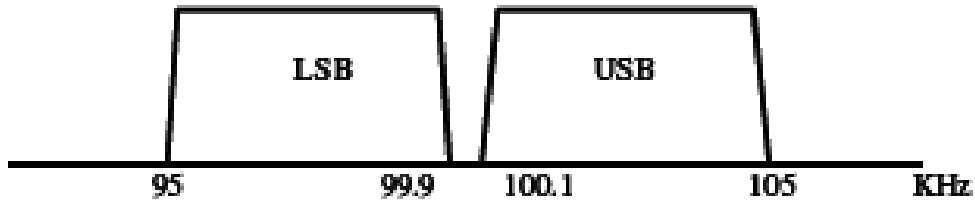
Figure 58

If the desired sideband suppression is 80 dB, the Q required to filter off one of the sidebands is approximately:

$$S = \log^{-1}\left(\frac{80}{20}\right) = 10^4$$

$$Q \approx \frac{f_c\sqrt{S}}{4\Delta f} = \frac{100 \times 10^3\sqrt{10^4}}{4 \times 200} = 12500$$

It is evident that a crystal filter would be needed to remove the unwanted sideband.

After the filter, only one sideband is left. In this example, we'll retain the USB. The spectrum after the second mixer is:
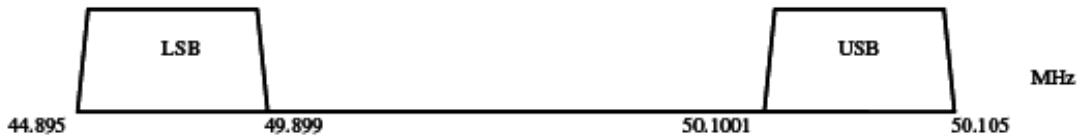
Figure 59

The Q required to suppress one of the side bands by 80 dB is approximately:

$$Q \approx \frac{f_c\sqrt{S}}{4\Delta f} = \frac{50 \times 10^6\sqrt{10^4}}{4 \times 200.2 \times 10^3} = 6244$$

Thus, we note that the required Q drops in half.

This SSB filter technique is used in radiotelephone applications.

2.9.2 Phase Shift Method
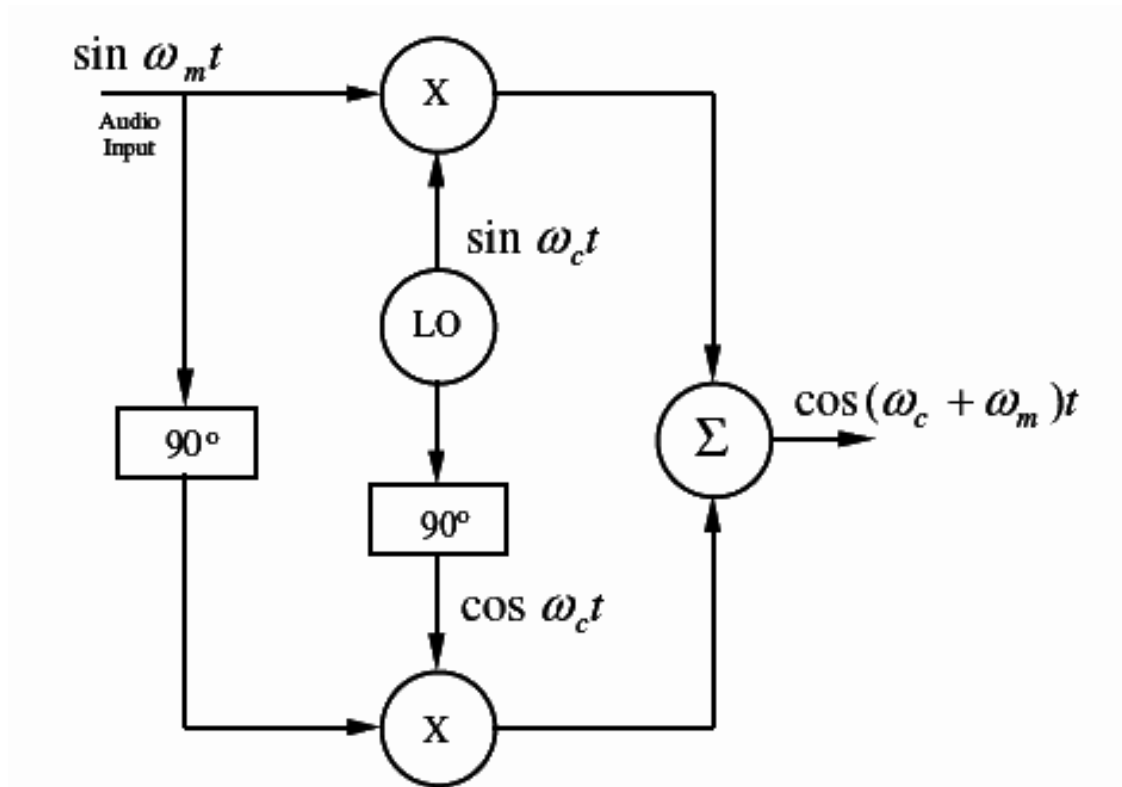
Figure 60

The output from the top mixer is given by:

$$\sin\omega_m t \sin\omega_c t = \frac{1}{2}\cos\left(\omega_c - \omega_m\right)t - \frac{1}{2}\cos\left(\omega_c + \omega_m\right)t$$

The output from the bottom mixer is given by:

$$\cos\omega_m t \cos\omega_c t = \frac{1}{2}\cos\left(\omega_c - \omega_m\right)t + \frac{1}{2}\cos\left(\omega_c + \omega_m\right)t$$

The output of the sumer is:

$$\cos\left(\omega_c - \omega_m\right)t$$

which corresponds to the lower sideband.

The major difficulty with this technique is the need to provide a constant 90° phase shift over the entire input audio band. To overcome this obstacle, the Weaver or third method uses an audio sub carrier, which is phase shifted.

2.9.3 Weaver Method

The Weaver or 'third' method places the baseband signal on a low frequency quadrature carrier.
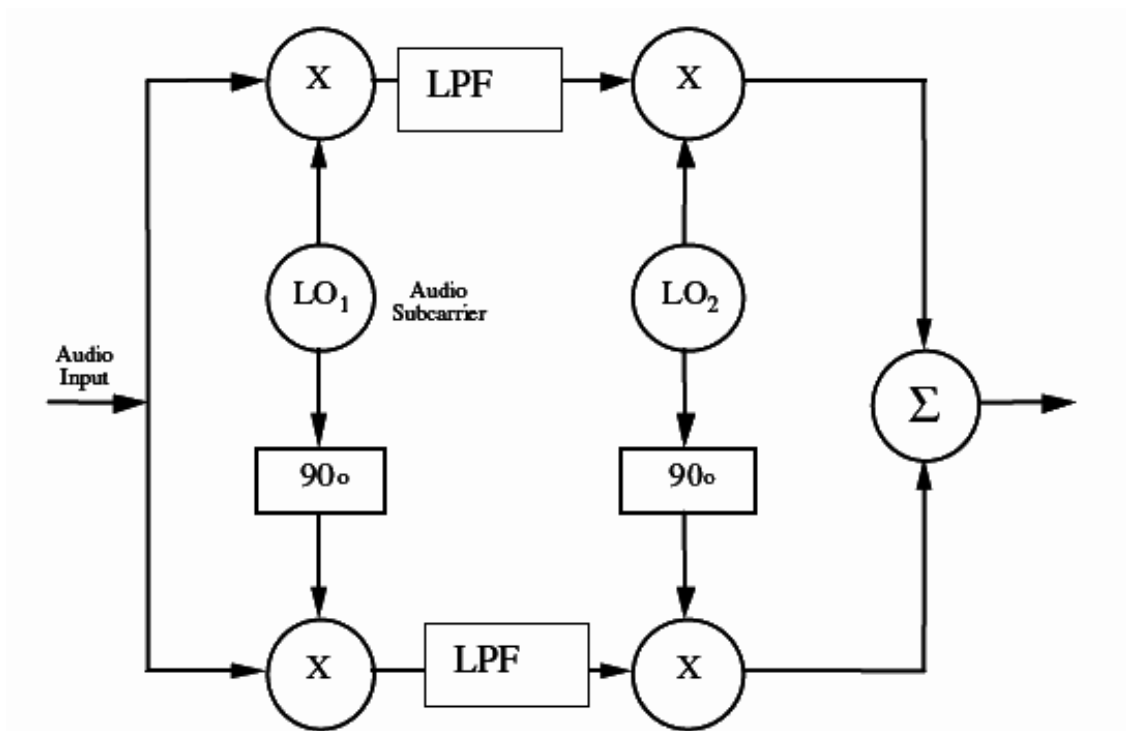
Figure 61

This has the advantage of not requiring a broadband phase shifter however; the use of four mixers makes it awkward and seldom used.

AM Modulation Waveforms

Time Domain

Frequency Domain

AM

LSB   Carrier   USB

DSBSC

LSB   USB

SSBSC

LSB

or

USB
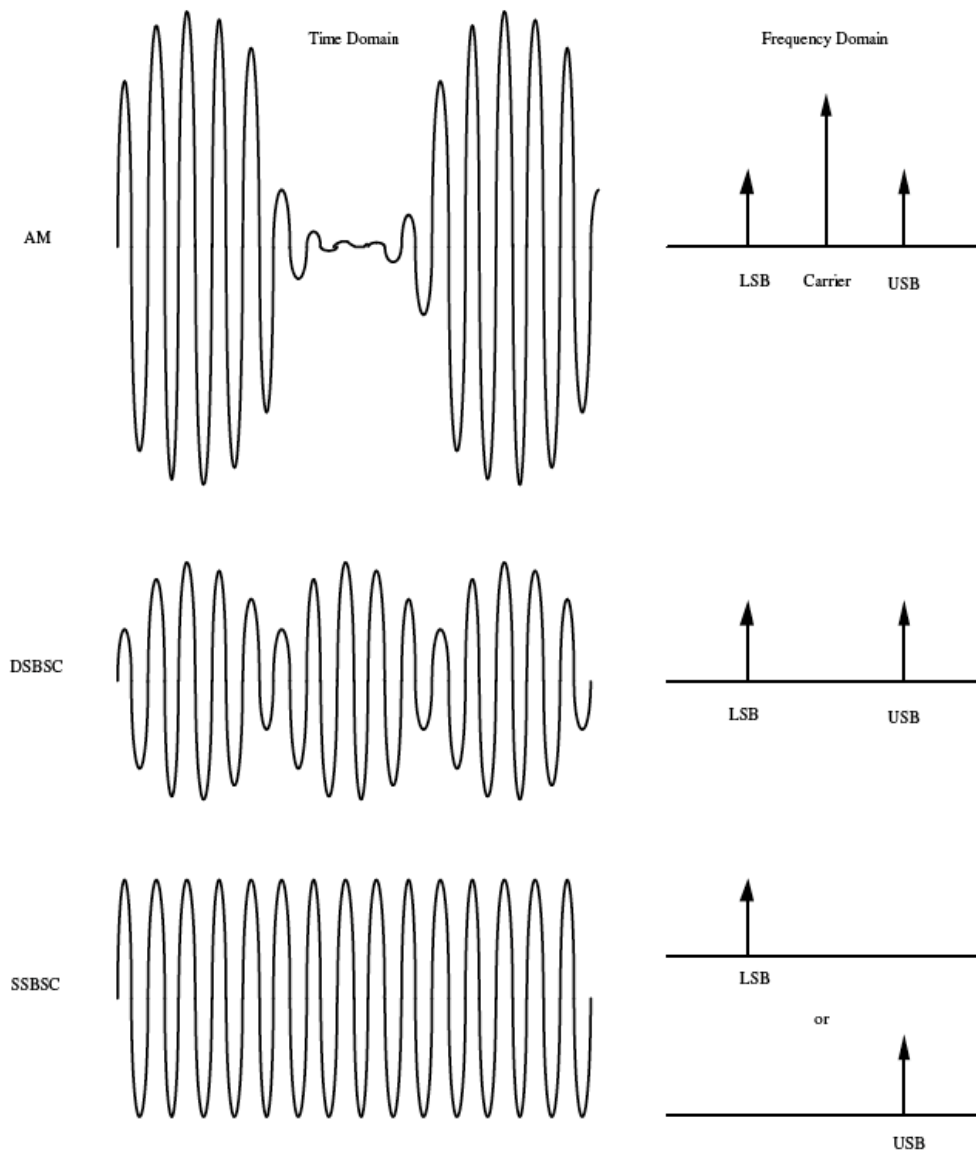
Figure 62

2.9.4 SSB Transmitter

AM-SSB transmitters are a little more complicated:

```
          cos(...)
             |
  Signal ---->(X)---->|Low-Pass Filter|----> AM-SSB
```
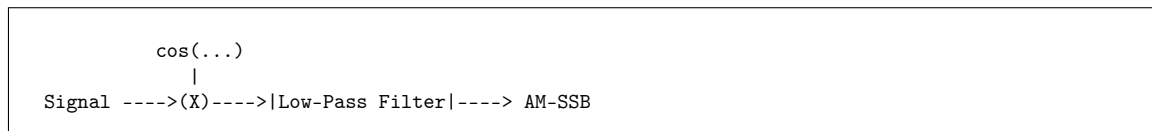
The filter must be a very high order, for reasons explained in that chapter.

2.9.5 SSB Receiver

An AM-SSB receiver is a little bit complicated as well:

```
            cos(...)
               |
  AM-SSB ---->(X)---->|Low-Pass Filter|---->|Amplifier|----> Signal
```
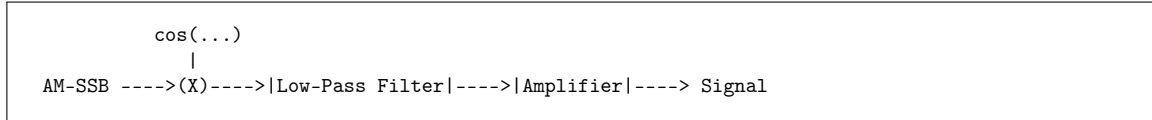
This filter doesnt need to be a very high order, like the transmitter has.

These receivers require extremely stable oscillators, good adjacent channel selectivity, and typically use a double conversion technique. Envelope detectors cannot be used since the envelope varies at twice the frequency of the AM envelope.

Stable oscillators are needed since the detected signal is proportional to the difference between the untransmitted carrier and the instantaneous side band. A small shift of 50 Hz makes the received signal unusable.

SSB receivers typically use fixed frequency tuning rather than continuous tuning as found on most radios. Crystal oscillators are often used to select the fixed frequency channels.

2.10 AM-VSB

As a compromise between AM-SSB and AM-DSB is AM-VSB. To make an AM-VSB signal, we pass an AM-DSB signal through a lowpass filter. Now, the trick is that we pass it through a low-order filter, so that some of the filtered sideband still exists. This filtered part of the sideband is called the "Vestige" of the sideband, hence the name "Vestigial Side Band".

AM-VSB signals then can get demodulated in a similar manner to AM-SSB. We can see when we remodulate the input signal, the two vestiges (the positive and negative mirrors of each other) over-lap each other, and add up to the original, unfiltered value!

AM-VSB is less expensive to implement then AM-SSB because we can use lower-order filters.

ⓘ **Information**

Broadcast television in North America uses AM-VSB

2.10.1 Transmitter

here we will talk about an AM-VSB transmitter circuit.

2.10.2 Receiver

here we will talk about an AM-VSB receiver circuit. (AM)

2.11 Frequency Modulation

w:Frequency modulation[4]

If we make the frequency of our carrier wave a function of time, we can get a generalized function that looks like this:

$$s_{FM} = A\cos(2\pi[f_c + ks(t)]t + \phi)$$

We still have a carrier wave, but now we have the value ks(t) that we add to that carrier wave, to send our data.

As an important result, ks(t) must be less than the carrier frequency always, to avoid ambiguity and distortion.
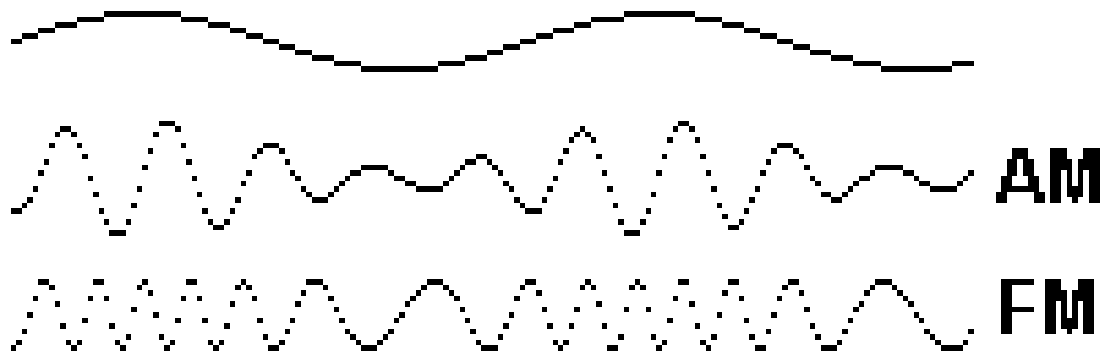
Figure 63

2.11.1 Deriving the FM Equation

Recall that a general sinusoid is of the form:

$$e_c = \sin\left(\omega_c t + \theta\right)$$

---

4   http://en.wikipedia.org/wiki/Frequency%20modulation

Frequency modulation involves deviating a carrier frequency by some amount. If a sine wave is used to deviate the carrier, the expression for the frequency at any instant would be:

$$\omega_i = \omega_c + \Delta\omega \sin(\omega_m t)$$

where:

$\omega_i =$ instantaneous frequency

$\omega_c =$ carrier frequency

$\Delta\omega =$ carrier deviation

$\omega_m =$ modulation frequency

This expression describes a signal varying sinusoidally about some average frequency. However, we cannot simply substitute this expression into the general equation for a sinusoid to get the FM equation. This is because the sine operator acts on angles, not frequency. Therefore, we must define the instantaneous frequency in terms of angles.

It should be noted that the modulation signal amplitude governs the *amount of carrier deviation* while the modulation signal frequency governs the *rate of carrier deviation.*

The term ω is an *angular velocity* (radians per second) and is related to frequency and angle by the following relationship:

$$\omega = 2\pi f = \frac{d\theta}{dt}$$

To find the angle, we must integrate ω with respect to time:

$$\int \omega dt = \theta$$

We can now find the instantaneous angle associated with the instantaneous frequency:

$$\theta = \int \omega_i dt = \int (\omega_c + \Delta\omega \sin(\omega_m t)) dt = \omega_c t - \frac{\Delta\omega}{\omega_m}\cos(\omega_m t) = \omega_c t - \frac{\Delta f}{f_m}\cos(\omega_m t)$$

This angle can now be substituted into the general carrier signal to define FM:

$$e_{fm} = \sin\left(\omega_c t - \frac{\Delta f}{f_m}\cos(\omega_m t)\right)$$

The *FM modulation index* is defined as the ratio of the carrier deviation to modulation frequency:

$$m_{fm} = \frac{\Delta f}{f_m}$$

Consequently, the FM equation is often written as:

$$e_{fm} = \sin\left(\omega_c t - m_{fm}\cos\left(\omega_m t\right)\right)$$

2.11.2 Bessel's Functions

w:Bessel function[5]

This is a very complex expression and it is not readily apparent what the sidebands of this signal are like. The solution to this problem requires a knowledge of *Bessel's function*s of the first kind and order p. In open form, it resembles:

$$J_p(x) = \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{x}{2}\right)^{2k+p}}{k!\,(k+p)!}$$

where:

$J_p(x) =$ Magnitude of the frequency component

$p =$ Side frequency number (not to be confused with sidebands)

$x =$ Modulation index

As a point of interest, Bessel's functions are a solution to the following equation:

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + \left(x^2 - p^2\right) = 0$$

Bessel's functions occur in the theory of cylindrical and spherical waves, much like sine waves occur in the theory of plane waves.

It turns out that FM generates an infinite number of side frequencies (in both the upper and lower sidebands). Each side frequency is an integer multiple of the modulation signal frequency. The amplitude of higher order side frequencies decreases rapidly and can generally be ignored.

The amplitude of the carrier signal is also a function of the modulation index and under some conditions, its amplitude can actually go to zero. This does not mean that the signal disappears, but rather that all of the broadcast energy is redistributed to the side frequencies.

A plot of the carrier and first five side frequency amplitudes as a function of modulation index resembles:

---

5   http://en.wikipedia.org/wiki/Bessel%20function

Bessels Functions of the First Kind
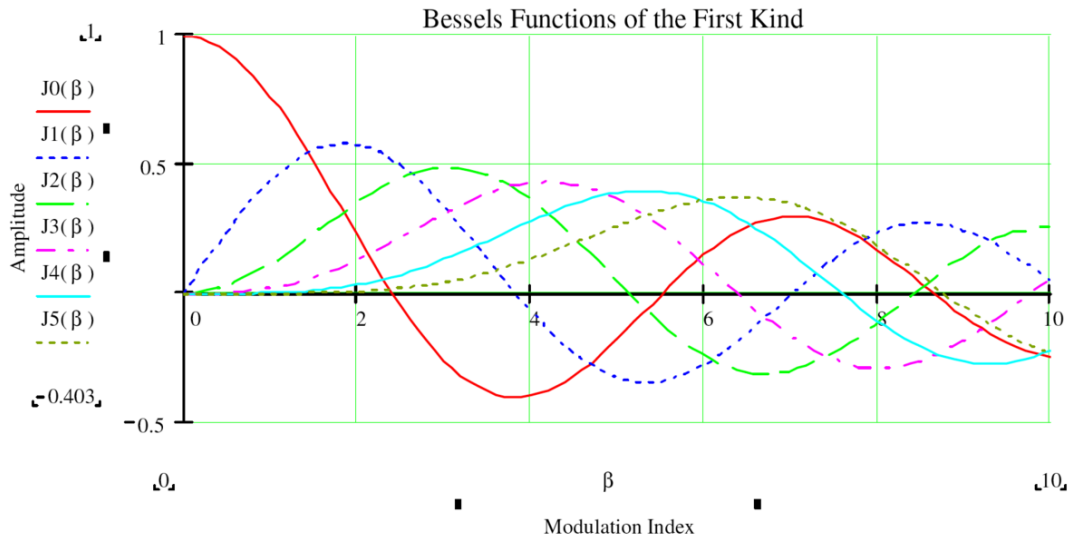
Figure 64

The Bessel coefficients have several interesting properties including:

$$J_0^2 + 2\left(J_1^2 + J_2^2 + J_3^2 + \cdots\right) = 1$$

One very useful interpretation of this is: J_0 represents the voltage amplitude of the carrier, J_1 represents the amplitude of the 1st side frequency, J_2 the 2nd side frequency etc. Note that the sum of the squares (power) remains constant.

2.11.3 FM Bandwidth

FM generates upper and lower sidebands, each of which contain an infinite number of side frequencies. However, the FM bandwidth is not infinite because the amplitude of the higher order side frequencies decreases rapidly. *Carson's Rule* is often used to calculate the bandwidth, since it contains more than 90% of the FM signal.

Carson's Rule

w: Carson's rule[6]

$$B_{fm} \approx 2\left(m_{fm} + 1\right)f_m = 2\left(\Delta f + f_m\right))$$

In commercial broadcast applications, the maximum modulation index $(m_{fm}) = 5$, the maximum, carrier deviation $(\Delta f) = 75$ kHz, and maximum modulation frequency $(f_m) = 15$ kHz. The total broadcast spectrum according to Carson's rule is 180 kHz, but an additional 20 kHz guard band is used to separate adjacent radio stations. Therefore, each FM radio station is allocated 200 kHz.

---

6   http://en.wikipedia.org/wiki/%20Carson%27s%20rule

2.11.4 Noise

In AM systems, noise easily distorts the transmitted signal however, in FM systems any added noise must create a frequency deviation in order to be perceptible.
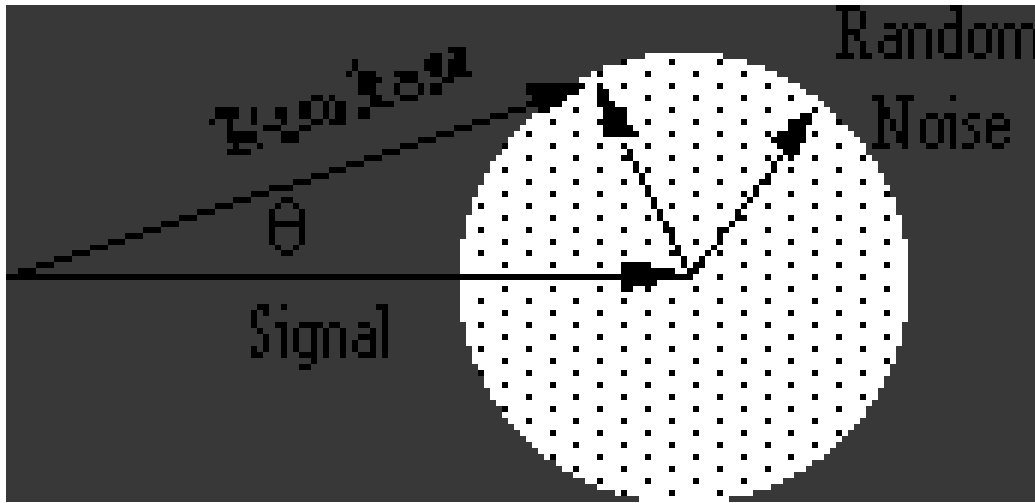
Figure 65

The maximum frequency deviation due to random noise occurs when the noise is at right angles to the resultant signal. In the worst case the signal frequency has been deviated by:

$$\delta = \theta f_m$$

This shows that the deviation due to noise increases as the modulation frequency increases. Since noise power is the square of the noise voltage, the signal to noise ratio can significantly degrade.
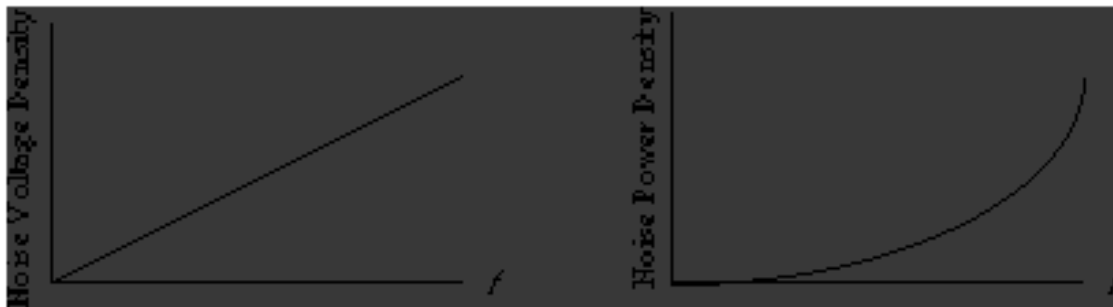
Figure 66

To prevent this, the amplitude of the modulation signal is increased to keep the S/N ratio constant over the entire broadcast band. This is called pre-emphasis.

2.11.5 Pre & De-emphasis

w:Preemphasis[7]

Increasing the amplitude of high frequency baseband signals in the FM modulator (transmitter) must be compensated for in the FM demodulator (receiver) otherwise the signal would sound quite tinny (too much treble).

The standard curves resemble:
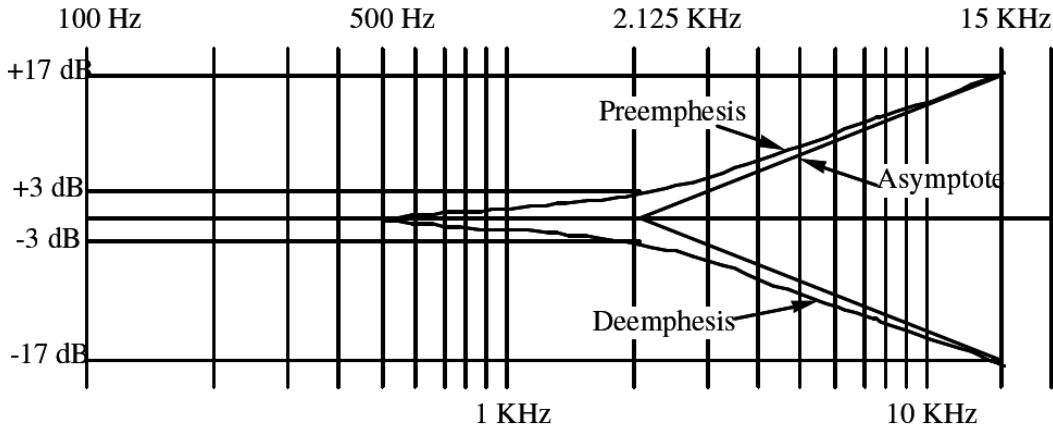
Figure 67

In commercial FM broadcast, the emphasis circuits consist of a simple RC network with a time constant of 75 μSec and a corner frequency of 2125 Hz.

Figure 68

---

7   http://en.wikipedia.org/wiki/Preemphasis
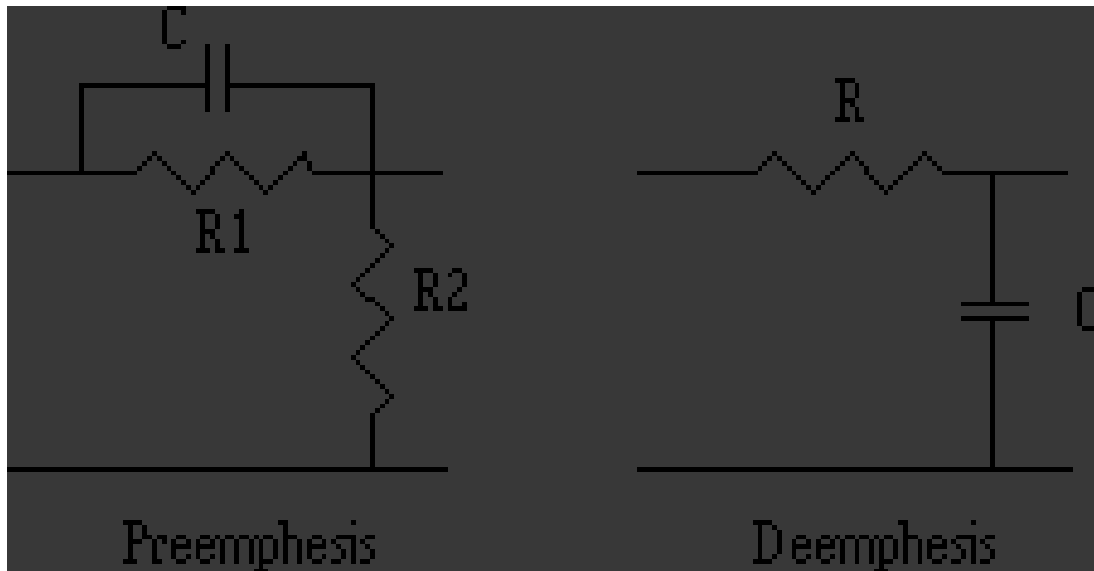
The magnitude of the pre-emphasis response is defined by:

$$\left| \frac{V_o}{V_{in}} \right| = 20 \log \left[\frac{R_2}{R_1 + R_2} \sqrt{1 + \left(\frac{f}{f_c} \right)^2} \right]$$

$$f_c = 2125 \, \text{Hz}$$

Figure 69

2.12 FM Transmission Power

The equation for the transmitted power in a sinusoid is a fundamental equation. Remember it.

Since the value of the amplitude of the sine wave in FM does not change, the transmitted power is a constant. As a general rule, for a sinusoid with a constant amplitude, the transmitted power can be found as follows:

$$P(t) = \frac{A^2}{2R_L}$$

Where A is the amplitude of the sine wave, and R_L is the resistance of the load. In a normalized system, we set R_L to 1.

The Bessel coefficients can be used to determine the power in the carrier and any side frequency:

$$P_C = P_T \left(J_0^2 + 2 \left(J_1^2 + J_2^2 + J_3^2 + \cdots \right) \right)$$

P_C is the power in the unmodulated carrier.

P_T is the total power and is by definition equal to the unmodulated carrier power.

As the modulation index varies, the individual Bessel coefficients change and power is redistributed from the carrier to the side frequencies.

2.13 FM Transmitters

FM Transmitters can be easily implemented using a VCO (see why we discussed Voltage Controlled Oscillators, in the first section?), because a VCO converts an input voltage (our input signal) to a frequency (our modulated output).

```
Signal ----->|VCO|-----> FM Signal
```

2.14 FM Receivers

Any angle modulation receiver needs to have several components:

1. A limiter, to remove abnormal amplitude values
2. bandpass filter, to separate the out-of-band noise.
3. A Discriminator, to change a frequency back to a voltage
4. A lowpass filter, to remove noise added by the discriminator.

A discriminator is essentially a differentiator in line with an envelope detector:

```
FM ---->|Differentiator|---->|Envelope Filter|----> Signal
```

Also, you can add in a blocking capacitor to remove any DC components of the signal, if needed. (FM)

2.15 Phase Modulation

Phase modulation is most commonly used to convey digital signals. All high performance modems today use phase modulation.

Similar to FM (frequency modulation), is Phase modulation. (We will show how they are the same in the next chapter.) If we alter the value of the phase according to a particular function, we will get the following generalized PM function:

$s_{PM} = A\cos(2\pi f_c t + \alpha s(t))$

It is important to note that the fact that $-\pi < \alpha s(t) \leq \pi$ for all values of t. If this relationship is not satisfied, then the phase angle is said to be *wrapped*.

2.15.1 BPSK Modulator

The binary phase shift keyed modulator is the simplest of app PSK modulators since it has only two output phase states. It is generally a multiplier which can either be an IC (integrated circuit) or ring modulator.
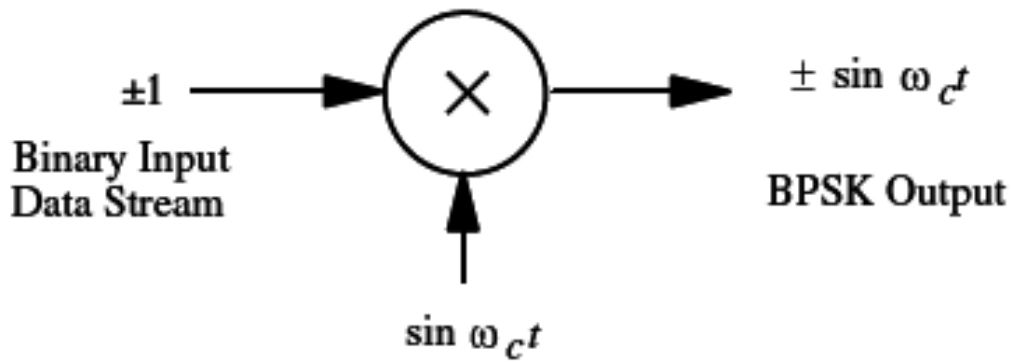
Figure 70

The output has two phase states:
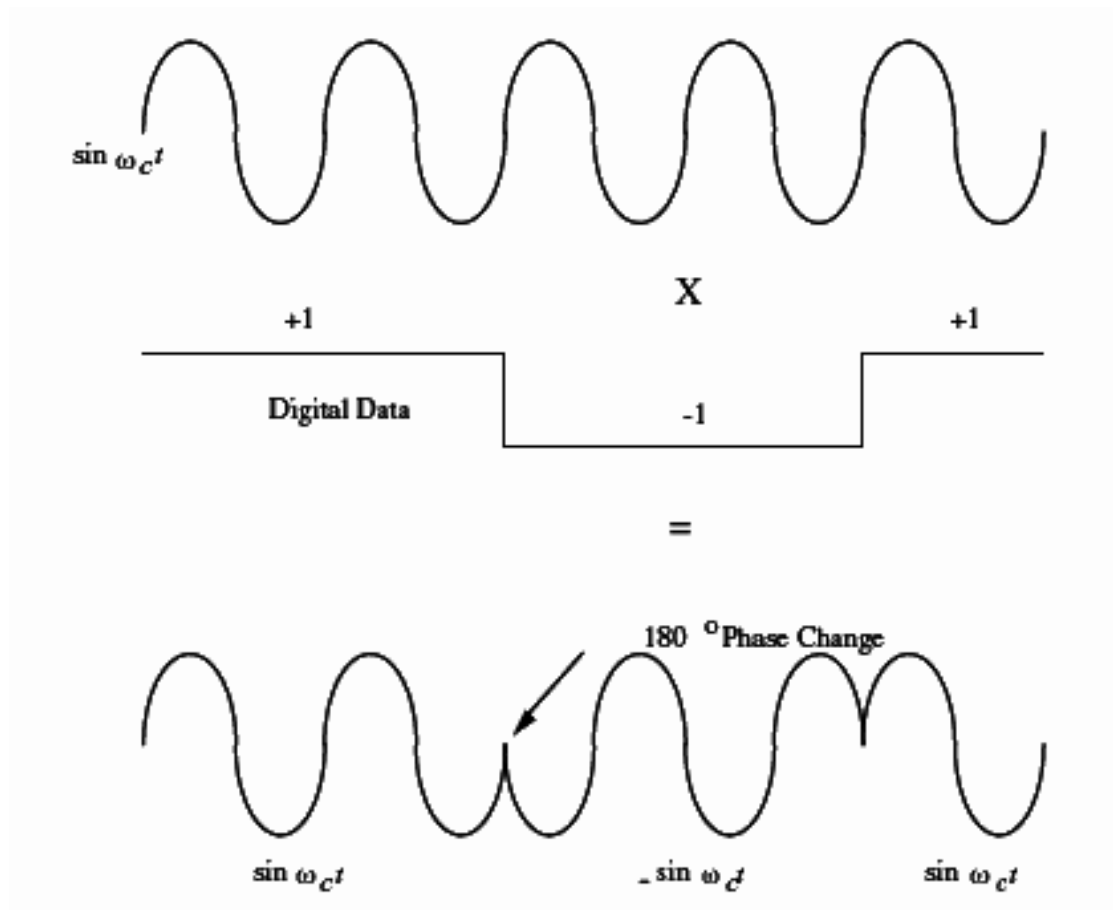
Figure 71

In the above illustration, the duration of each of the phase states corresponds to one signaling element or baud. The baud rate is therefor equal to the bit rate.

The spectrum of the BPSK signal will depend upon the data being transmitted, but it is very easy to sketch it for the highest data rate input.
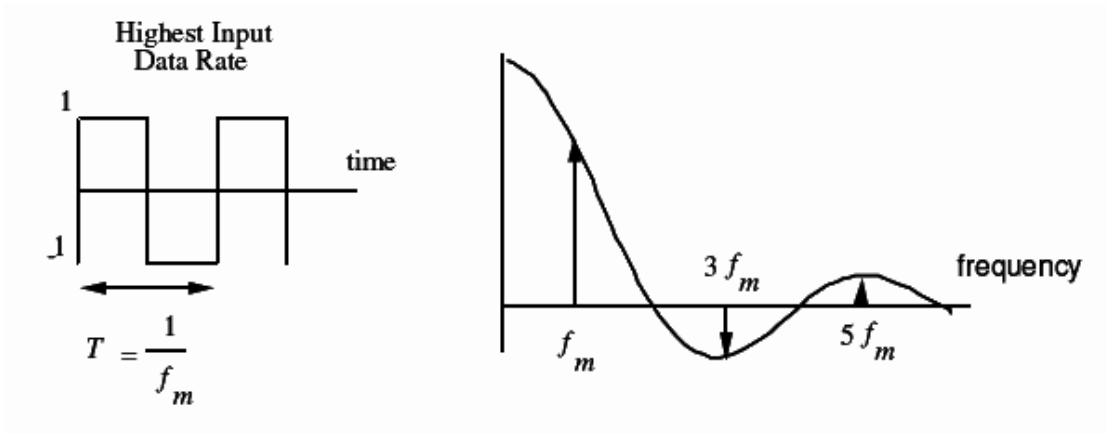
Figure 72

The resultant BPSK spectrum is:
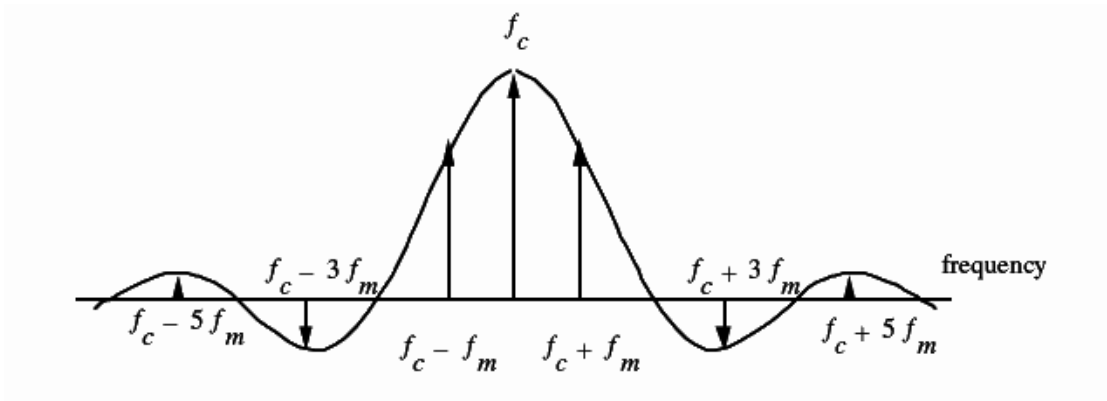
Figure 73

2.15.2 QPSK Modulators (4-PSK)

Quadrature modulation uses two data channels denoted I (in phase) and Q (quadrature phase) displaced by 90° with respect to each other. It may seem somewhat paradoxical, that although these two channels are combined prior to transmission, they do not interfere with each other.
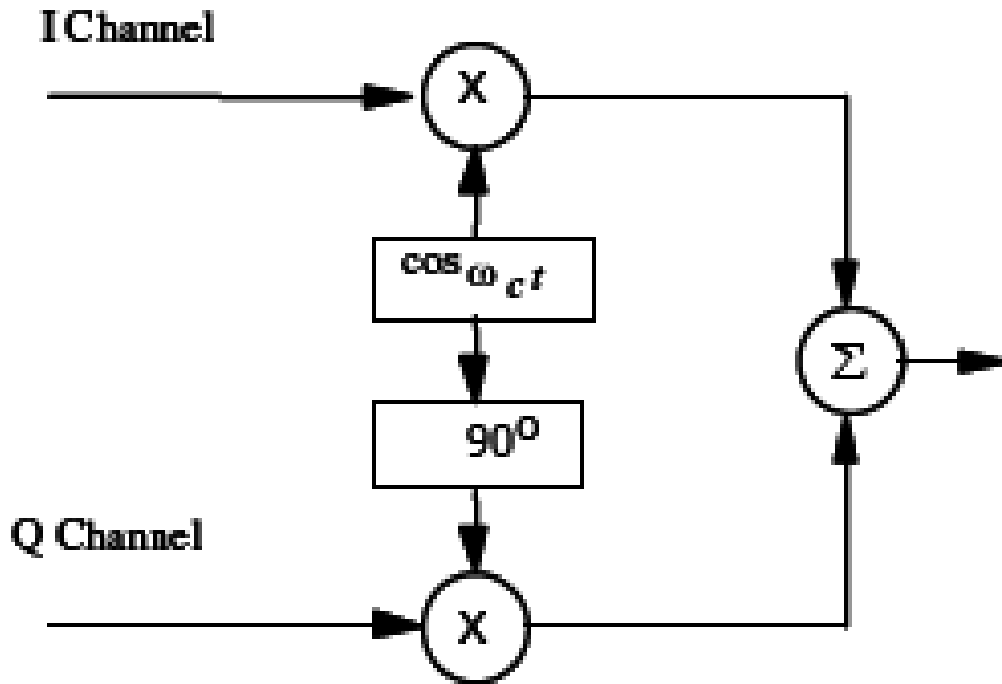
Figure 74

The receiver is quite capable of separating them because of their quadrature or orthogonal nature.
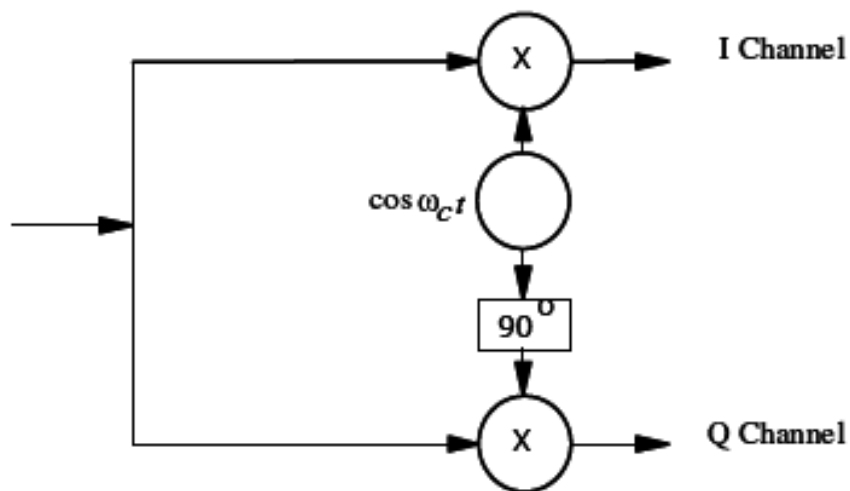
Figure 75

In the most basic configuration, there are 4 possible output phases. This suggests that each output symbol correspond to 2 bits of binary information. Since several bits can be encoded into a baud, the bit rate exceeds the baud rate.
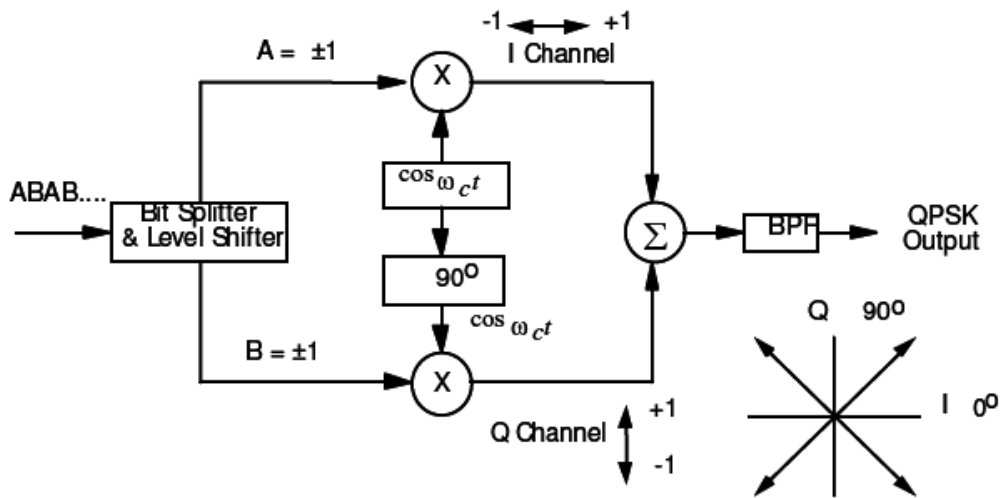
Figure 76

The first thing that happens in this circuit is that the incoming bits are organized into groups of 2 called dibits. They are separated into 2 data streams and kept constant over the dibit period.
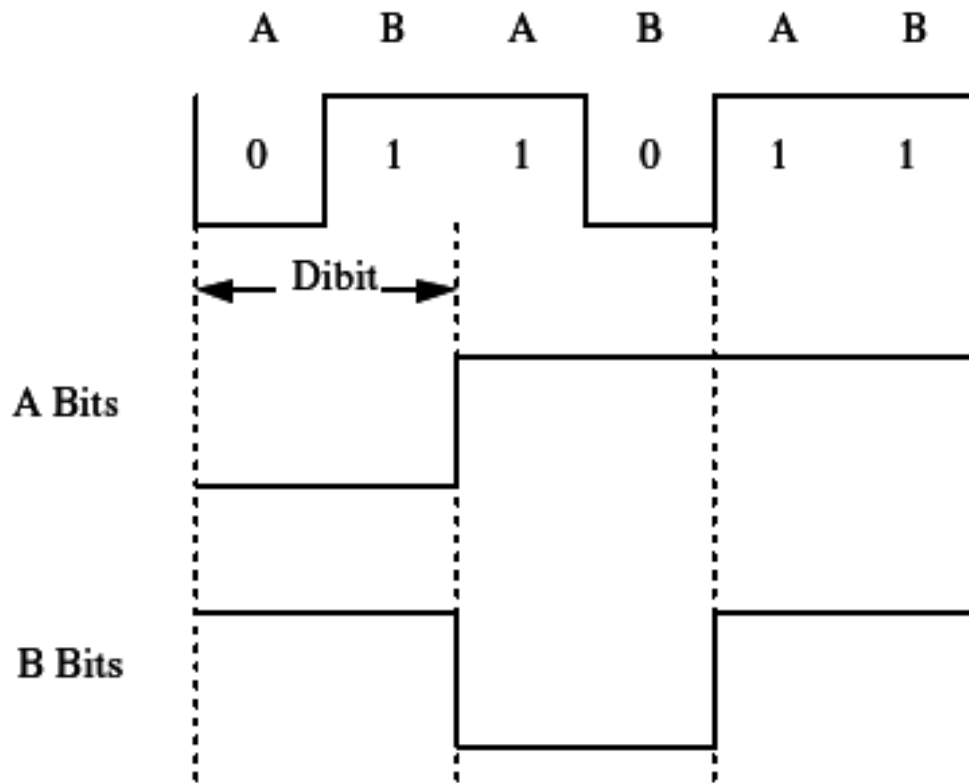
Figure 77

Each data stream is fed to a BPSK modulator. However, orthogonal carriers feed the two modulators. The output of the I channel modulator resembles:

A Bits

+1

-1

$-\cos \omega_c t$ $\cos \omega_c t$

Output of I Channel Modulator
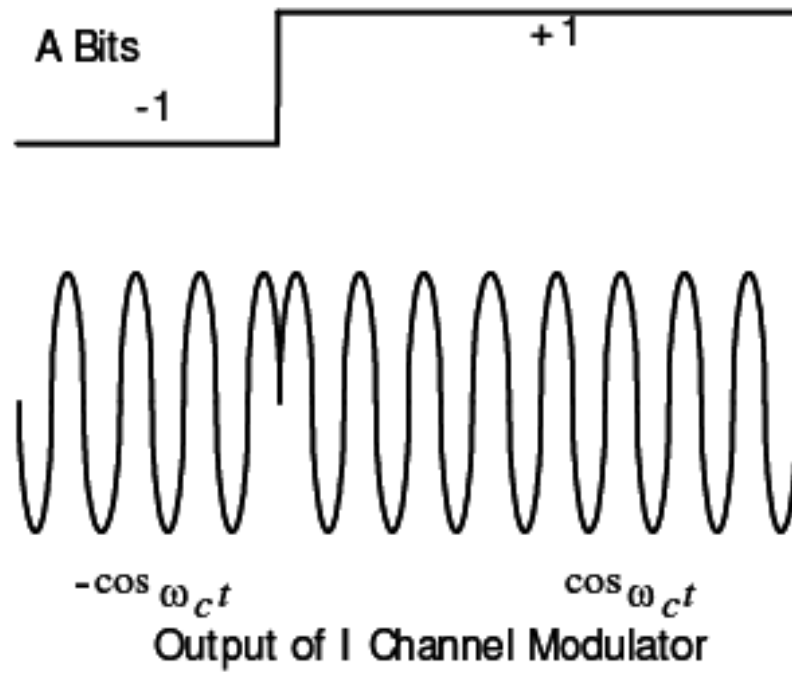
Figure 78

The output of the Q channel modulator resembles

B Bits

sin $\omega_c t$ $-$ sin $\omega_c t$ sin $\omega_c t$

Output of Q Channel Modulator
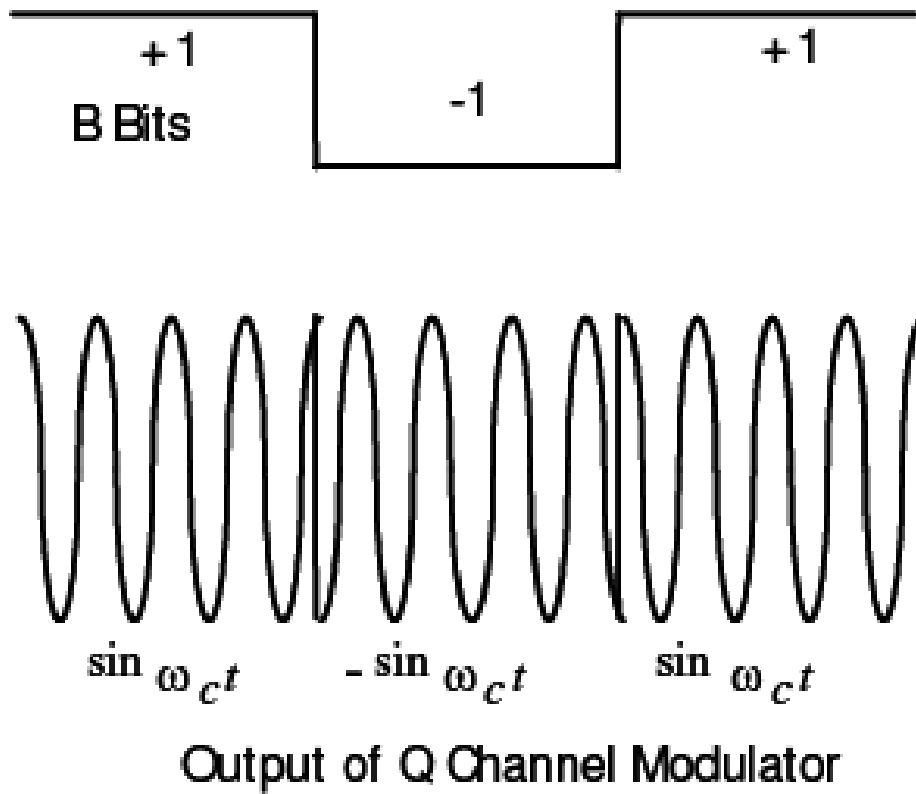
Figure 79

Combining the I and Q channels has the effect of rotating the output state by 45º.
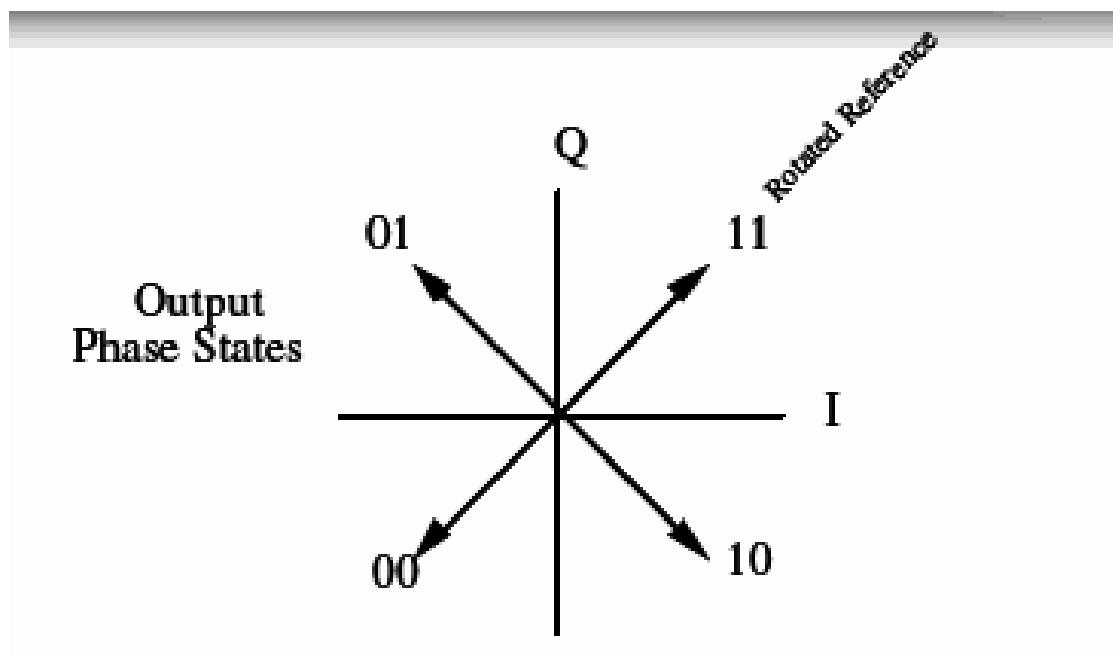
Figure 80

Rotating the output reference to 45º for the sake of clarity, the transmitted output for this particular data sequence is therefor:
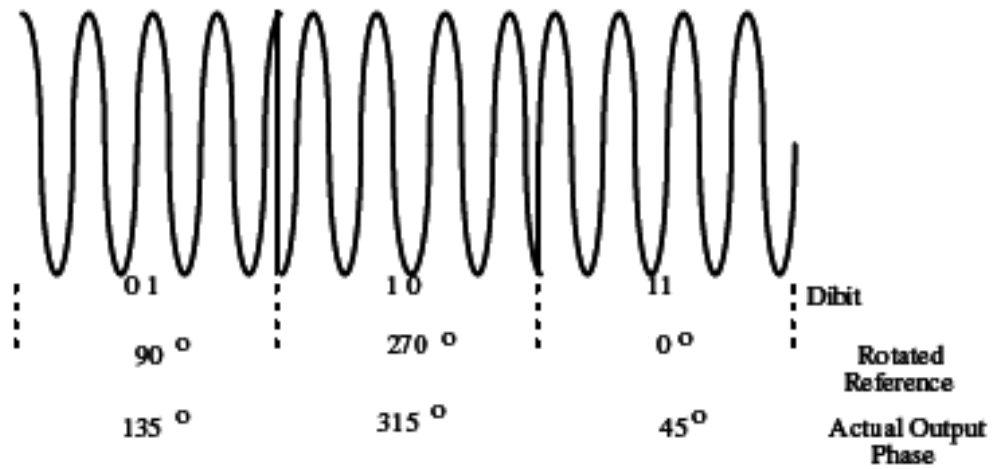
Figure 81

2.15.3 8-PSK

This process of encoding more bits into each output baud or phase state can be continued. Organizing binary bits into 3 bytes corresponds to 8 different conditions.

The output constellation diagram for the 8 different phase states is:
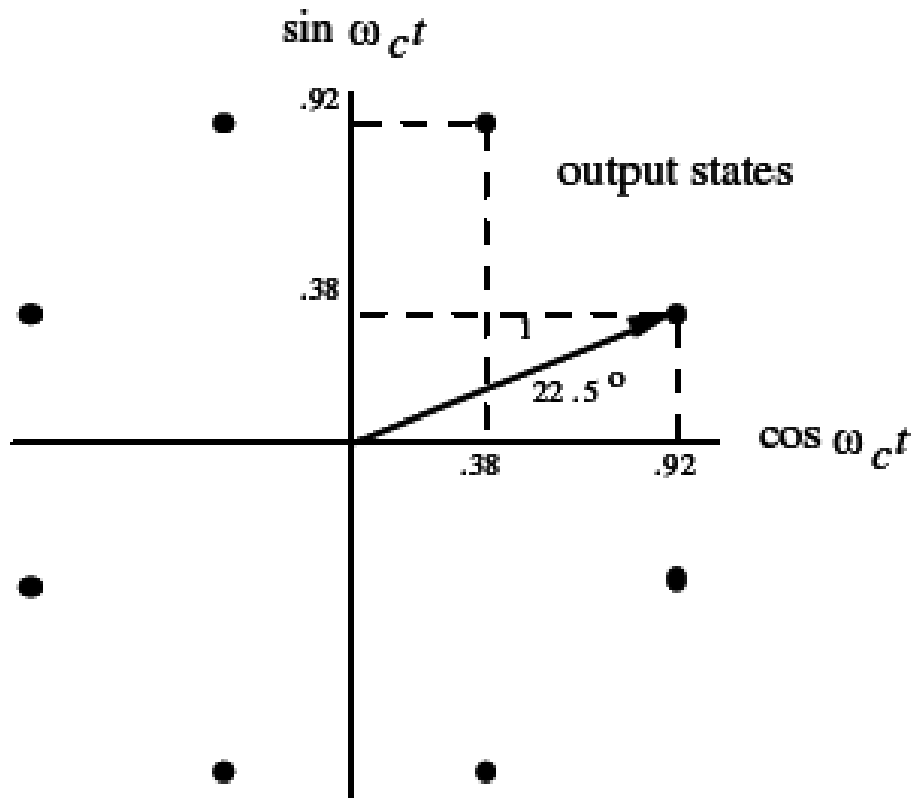
Figure 82

From this diagram it is readily apparent that two different amplitudes are needed on the I and Q channels. If the A bit is used to control the polarity of the I channel and the B bit the polarity of the Q channel, then the C bit can be used to define the two different amplitudes. In order to evenly space the phase states; the amplitudes must be $\pm$ 0.38 and $\pm$ 0.92. The magnitude of the I and Q channel signals must always be different. An inverter can be used to assure this condition.

The input bit stream is organized into 3 bit bytes. Each bit is sent to a different location to control a certain aspect of the modulator. The inputs to the 2 - 4 level converter are 0's or 1's but the output is $\pm$ 0.38 or $\pm$ 0.92, depending on the C bit.
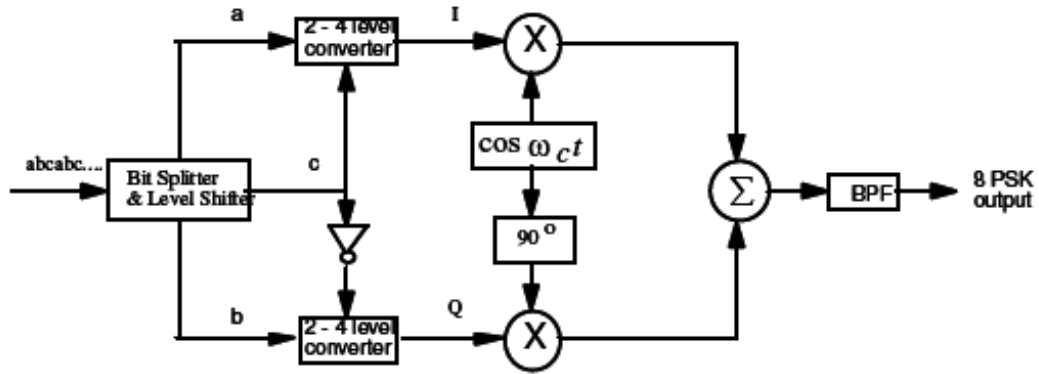
Figure 83

2.16 Wrapped/Unwrapped Phase

The phase angle is a circular quantity, with the restriction $0 = 2\pi$. Therefore, if we wrap the phase a complete 360 degrees around, the receiver will not know the difference, and the transmission will fail. When the phase exceeds 360 degrees, the phase value is said to be **wrapped**. It is highly difficult to construct a communication system that can detect and decode a wrapped phase value.

2.17 PM Transmitter

PM signals can be transmitted using a technique very similar to FM transmitters. The only difference is that we need to add a differentiator to it:

```
Signal ---->|Differentiator|---->|VCO|----> PM Signal
```
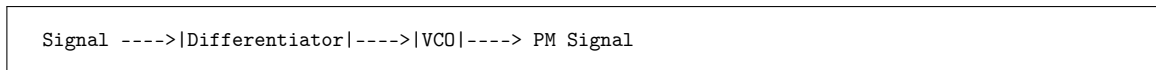
2.18 PM Receiver

PM receivers have all the same parts as an FM receiver, except for the 3rd step:

1. A limiter, to remove abnormal amplitude values
2. bandpass filter, to separate the out-of-band noise.
3. A Phase detector, to convert a phase back into a voltage
4. A lowpass filter, to remove noise added by the discriminator.

Phase detectors can be created using a Phase-Locked-Loop (again, see why we discussed them first?). (PM)

2.19 Concept

We can see from our initial overviews that FM and PM modulation schemes have a lot in common. Both of them are altering the angle of the carrier sinusoid according to some function. It turns out that we can go so far as to generalize the two together into a single modulation scheme known as **angle modulation**. Note that we will never abbreviate "angle modulation" with the letters "AM", because Amplitude modulation is completely different from angle modulation.

2.20 Instantaneous Phase

Let us now look at some things that FM and PM have of common:

$$s_{FM} = A\cos(2\pi[f_c + ks(t)]t + \phi)$$

$$s_{PM} = A\cos(2\pi f_c t + \alpha s(t))$$

What we want to analyze is the *argument of the sinusoid*, and we will call it Psi. Let us show the Psi for the bare carrier, the FM case, and the PM case:

$$\Psi_{carrier}(t) = 2\pi f_c t + \phi$$

$$\Psi_{FM}(t) = 2\pi[f_c + ks(t)]t + \phi$$

$$\Psi_{PM}(t) = 2\pi f_c t + \alpha s(t)$$

$$s(t) = A\cos(\Psi(t))$$

This Psi value is called the *Instantaneous phase* of the sinusoid.

2.21 Instantaneous Frequency

Using the Instantaneous phase value, we can find the *Instantaneous frequency* of the wave with the following formula:

$$f(t) = \frac{d\Psi(t)}{dt}$$

We can also express the instantaneous phase in terms of the instantaneous frequency:

$$\Psi(t) = \int_{-\infty}^{t} f(\lambda)d\lambda$$

Where the Greek letter "lambda" is simply a dummy variable used for integration. Using these relationships, we can begin to study FM and PM signals further.

2.22 Determining FM or PM

If we are given the equation for the instantaneous phase of a particular angle modulated transmission, is it possible to determine if the transmission is using FM or PM? it turns out that it is possible to determine which is which, by following 2 simple rules:

1. In PM, instantaneous phase is a linear function.
2. In FM, instantaneous frequency minus carrier frequency is a linear function.

For a refresher course on Linearity, there is a chapter on the subject in the Signals and Systems book[8] worth re-reading.

ⓘ Information

FM radio uses generalized "Angle Modulation"

2.23 Bandwidth

In a PM system, we know that the value $\alpha s(t)$ can never go outside the bounds of $(-\pi, \pi]$. Since sinusoidal functions oscillate between [-1, 1], we can use them as a general PM generating function. Now, we can combine FM and PM signals into a general equation, called *angle modulation*:

$$v(t) = A\sin(2\pi f_c t + \beta \sin(2\pi f_m t))$$

If we want to analyze the spectral components of this equation, we will need to take the Fourier transform of this. But, we can't integrate a sinusoid of a sinusoid, much less find the transform of it. So, what do we do?

It turns out (and the derivation will be omitted here, for now) that we can express this equation as an infinite sum, as such:

$$v(t) = A\sum_{n=-\infty}^{\infty} J_n(\beta)\sin[2\pi(nf_m + f_c)t]$$

But, what is the term $J_n(\beta)$? J is the *Bessel function*, which is a function that exists only as an open integral (it is impossible to write it in closed form). Fortunately for us, there are extensive tables tabulating Bessle function values.

---

8   http://en.wikibooks.org/wiki/Signals%20and%20Systems%2FTime%20Domain%20Analysis%2FLTI%20systems

2.24 The Bessel Function

The definition of the Bessel function is the following equation:

$J_n(\beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j[\beta sin\theta - n\theta]} d\theta$

The bessel function is a function of 2 variables, N and β.

Bessel Functions have the following properties:

- If n is even:

$$J_{-n}(\beta) = J_n(\beta)$$

- If n is odd:

$$J_{-n}(\beta) = -J_n(\beta)$$

- $J_{n-1} + J_{n+1} = \frac{2n}{\beta} J_n(\beta)$.

The bessel function is a relatively advanced mathematical tool, and we will not analyze it further in this book.

2.25 Carson's Rule

If we have our generalized function:

$$v(t) = A\sin(2\pi f_c t + \beta \sin(2\pi f_m t))$$

We can find the bandwidth BW of the signal using the following formula:

$$BW = 2(\beta + 1)f_m = 2(\Delta f + f_m)$$

where Δf is the maximum frequency deviation, of the transmitted signal, from the carrier frequency. It is important to note that Carson's rule is only an approximation (albeit one that is used in industry frequently).

2.26 Demodulation: First Step

Now, it is important to note that FM and PM signals both do the same first 2 steps during demodulation:

1. Pass the signal through a **limiter**, to remove amplitude peaks

2. Pass the signal through a bandpass filter to remove low and high frequency noise (as much as possible, without filtering out the signal).

Once we perform these two steps, we no longer have white noise, because we've passed the noise through a filter. Now, we say the noise is **colored**.

here is a basic diagram of our demodulator, so far:

```
        channel
  s(t) ---------> r(t) --->|Limiter|--->|Bandpass Filter|---->z(t)
```
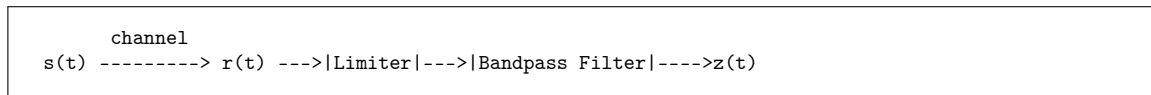
Where z(t) is the output of the bandpass filter.

2.27 Filtered Noise

To denote the new, filtered noise, and new filtered signal, we have the following equation:

$$z(t) = \gamma A \cos(\Psi(t)) + n_0(t)$$

Where we call the additive noise $n_0(t)$ because it has been filtered, and is not white noise anymore. $n_0(t)$ is known as **narrow band noise**, and can be denoted as such:

$$n_0(t) = \mathbf{W}(t)\cos(2\pi f_c t) + \mathbf{Z}(t)\sin(2\pi f_c t)$$

Now, once we have it in this form, we can use a trigonometric identity to make this equation more simple:

$$n_0(t) = \mathbf{R}(t)\cos(2\pi f_c t + \mathbf{\Theta}(t))$$

Where

$$\mathbf{R}(t) = \sqrt{\mathbf{W}(t)^2 + \mathbf{Z}(t)^2}$$

$$\mathbf{\Theta}(t) = tan^{-1}(\mathbf{Z}(t)/\mathbf{W(t)})$$

Here, the new noise parameter R(t) is a **rayleigh** random variable, and is discussed in the next chapter.

2.28 Noise Analysis

R(t) is a noise function that affects the amplitude of our received signal. However, our receiver passes the signal through a limiter, which will remove amplitude fluctuations from our signal. For this reason, R(t) doesnt affect our signal, and can be safely ignored for

now. This means that the only random variable that is affecting our signal is the variable $\Theta(t)$, "Theta". Theta is a uniform random variable, with values between pi and -pi. Values outside this range "Wrap around" because phase is circular.

3 Transmission

This page will discuss some of the fundamental basics of EM wave propagation.

3.1 Electromagnetic Spectrum

w:Electromagnetic_wave[1]
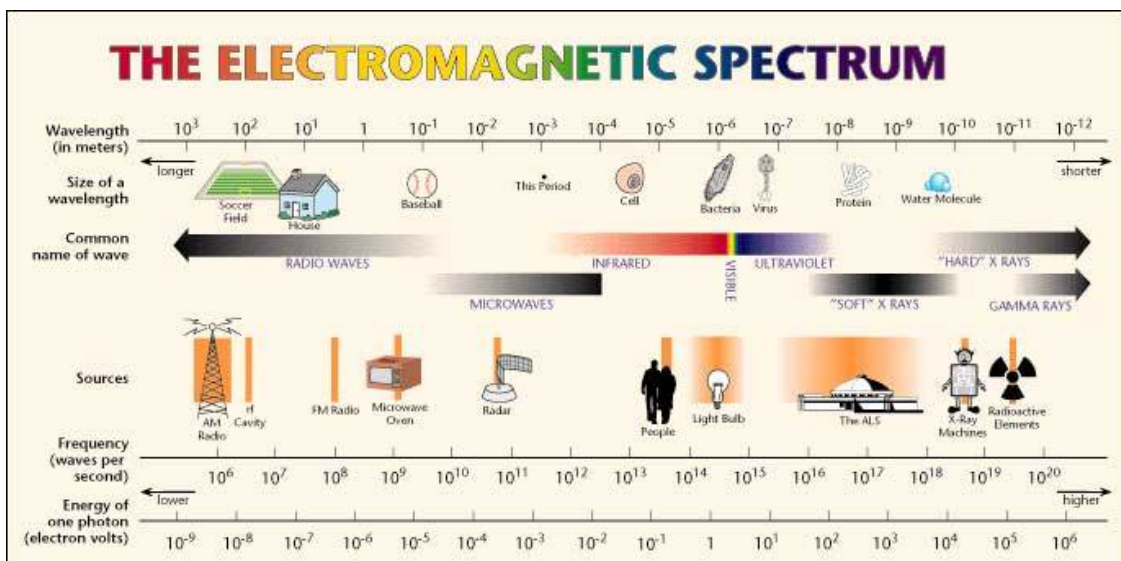
Figure 84

3.2 Radio Waves

w:Radio_waves[2]

Maxwell first predicted the existence of electromagnetic waves in the 19th century. He came to this conclusion by careful examination of the equations describing electric and magnetic phenomenon. It was left up to Hertz to create these waves, and Marconi to exploit them.

In spite of one hundred years of study, exactly what radio waves are and why they exist, remain somewhat of a mystery.

---

1   http://en.wikipedia.org/wiki/Electromagnetic_wave
2   http://en.wikipedia.org/wiki/Radio_waves

Electromagnetic waves in free space, or TEM waves, consist of electric and magnetic fields, each at right angles to each other and the direction of propagation.
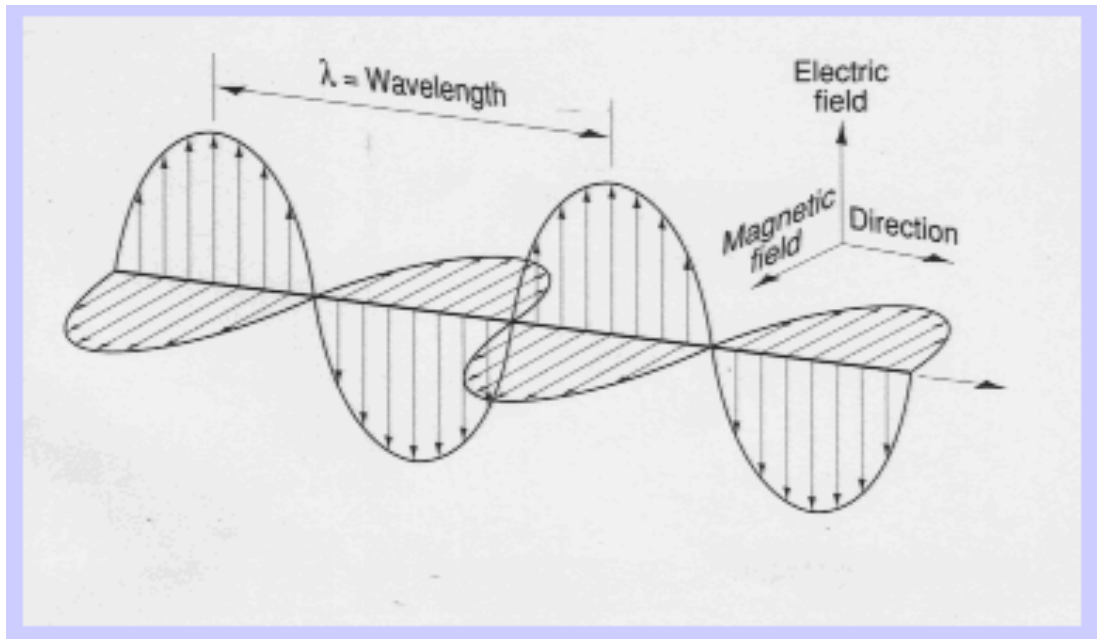
Figure 85

The relationship between wavelength and frequency is give by:

$$c = \lambda f$$

where c is the speed of light (approximately 300,000,000 m/s in vacuum), f is the frequency of the wave, and λ is the wavelength of the wave.

Radio waves can be reflected and refracted in a manner similar to light. They are affected by the ground terrain, atmosphere and other objects.

Maxwell's equations state that a time varying magnetic field produces an electric field and a time varying electric field produces a magnetic field. This is kind of a chicken and egg situation.

Radio waves propagate outward from an antenna, at the speed of light. The exact nature of these waves is determined by the transmission medium. In free space, they travel in straight lines, whereas in the atmosphere, they generally travel in a curved path. In a confined or guided medium, radio waves do not propagate in the TEM mode, but rather in a TE or TM mode.

Radio waves interact with objects in three principle ways:

Reflection – A radio wave bounces off an object larger than its wavelength.

Diffraction – Waves bend around objects.

Scattering – A radiowave bounces off an object smaller than its wavelength.

Because of these complex interactions, radio wave propagation is often examined in three distinct regions in order to simplify the analysis:

Surface (or ground) waves are located very near the earth's surface.

Space waves occur in the lower atmosphere (troposphere).

Sky waves occur in the upper atmosphere (ionosphere).

The boundaries between these regions are somewhat fuzzy. In many cases, it is not possible to examine surface waves without considering space waves.
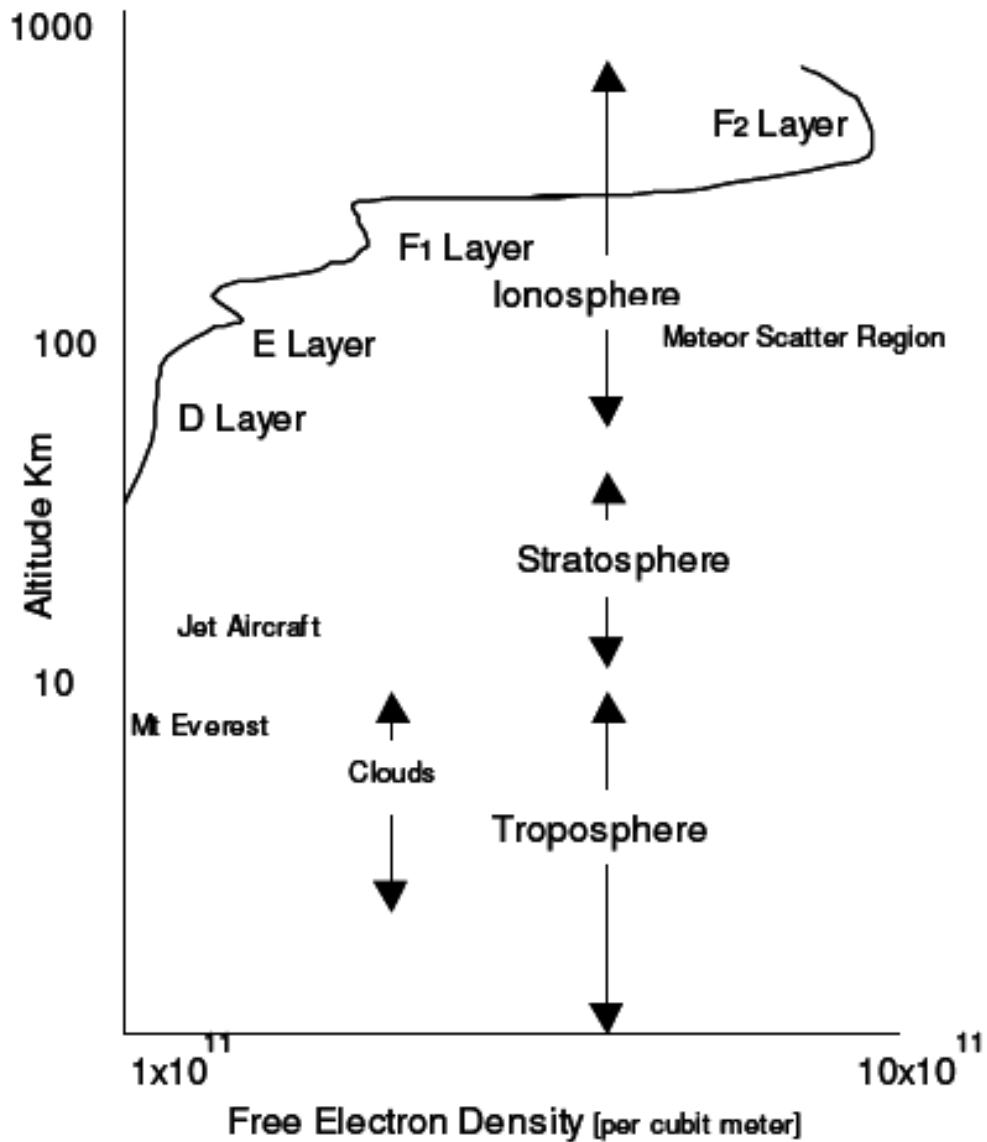
Figure 86

3.2.1 Common RF Band Designations

Frequency band name	Frequency	Wavelength
ELF - Extremely Low Frequency	$3-30$ Hz	$100000-10000$ km
SLF - Super Low Frequency	$30-300$ Hz	$10000-1000$ km
ULF - Ultra Low Frequency	$300-3000$ Hz	$1000-100$ km
VLF - Very Low Frequency	$3-30$ kHz	$100-10$ km
LF - Low Frequency	$30-300$ kHz	$10-1$ km
MF - Medium Frequency	$300-3000$ kHz	$1000-100$ m
HF - High Frequency	$3-30$ MHz	$100-10$ m
VHF - Very High Frequency	$30-300$ MHz	$10-1$ m
UHF - Ultra High Frequency	$300-3000$ MHz	$1000-100$ mm
SHF - Super High Frequency	$3-30$ GHz	$100-10$ mm
EHF - Extremely High Frequency	$30-300$ GHz	$10-1$ mm

3.2.2 Surface Waves

These are the principle waves used in AM, FM and TV broadcast. Objects such as buildings, hills, ground conductivity, etc. have a significant impact on their strength. Surface waves are usually vertically polarized with the electric field lines in contact with the earth.
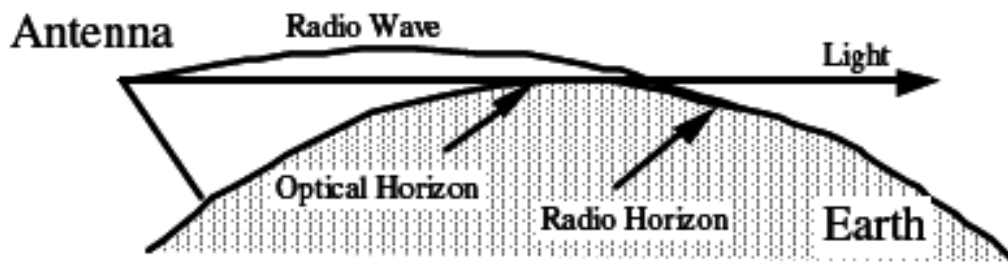
Figure 87

Refraction

Because of refraction, the radio horizon is larger than the optical horizon by about 4/3. The typical maximum direct wave transmission distance (in km) is dependent on the height of the transmitting and receiving antennas (in meters):

$$d_{\max} \approx \sqrt{17h_t} + \sqrt{17h_r} \quad \text{km}$$

However, the atmospheric conditions can have a dramatic effect on the amount of refraction.
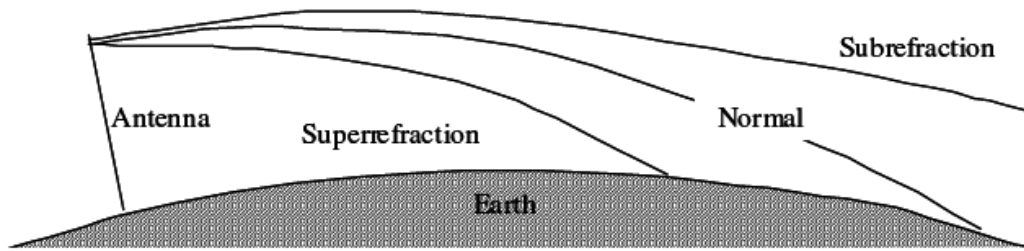
Figure 88

Super Refraction

In super refraction, the rays bend more than normal thus shortening the radio horizon. This phenomenon occurs when temperature increases but moisture decreases with height. Paradoxically, in some cases, the radio wave can travel over enormous distances. It can be reflected by the earth, rebroadcast and super refracted again.

Sub refraction

In sub refraction, the rays bend less than normal. This phenomenon occurs when temperature decreases but moisture increases with height. In extreme cases, the radio signal may be refracted out into space.

3.2.3 Space Waves

These waves occur within the lower 20 km of the atmosphere, and are comprised of a direct and reflected wave. The radio waves having high frequencies are basically called as space waves. These waves have the ability to propagate through atmosphere, from transmitter antenna to receiver antenna. These waves can travel directly or can travel after reflecting from earth's surface to the troposphere surface of earth. So, it is also called as Tropospherical Propagation. In the diagram of medium wave propagation, c shows the space wave propagation. Basically the technique of space wave propagation is used in bands having very high frequencies. E.g. V.H.F. band, U.H.F band etc. At such higher frequencies the other wave propagation techniques like sky wave propagation, ground wave propagation can't work. Only space wave propagation is left which can handle frequency waves of higher frequencies. The other name of space wave propagation is line of sight propagation. There are some limitations of space wave propagation.

1. These waves are limited to the curvature of the earth.
2. These waves have line of sight propagation, means their propagation is along the line of sight distance.

The line of sight distance is that exact distance at which both the sender and receiver antenna are in sight of each other. So, from the above line it is clear that if we want to

increase the transmission distance then this can be done by simply extending the heights of both the sender as well as the receiver antenna. This type of propagation is used basically in radar and television communication.

The frequency range for television signals is nearly 80 to 200MHz. These waves are not reflected by the ionosphere of the earth. The property of following the earth's curvature is also missing in these waves. So, for the propagation of television signal, geostationary satellites are used. The satellites complete the task of reflecting television signals towards earth. If we need greater transmission then we have to build extremely tall antennas.

Direct Wave

This is generally a line of sight transmission, however, because of atmospheric refraction the range extends slightly beyond the horizon.

Ground Reflected Wave

Radio waves may strike the earth, and bounce off. The strength of the reflection depends on local conditions. The received radio signal can cancel out if the direct and reflected waves arrive with the same relative strength and 180º out of phase with each other.

Horizontally polarized waves are reflected with almost the same intensity but with a 180º phase reversal.

Vertically polarized waves generally reflect less than half of the incident energy. If the angle of incidence is greater than 10º there is very little change in phase angle.

3.2.4 Sky Waves

These waves head out to space but are reflected or refracted back by the ionosphere. The height of the ionosphere ranges from 50 to 600 km.

Radio waves are refracted by the ionized gas created by solar radiation. The amount of ionization depends on the time of day, season and the position in the 11-year sun spot cycle. The specific radio frequency refracted is a function of electron density and launch angle.

A communication channel thousands of kilometers long can be established by successive reflections at the earth's surface and in the upper atmosphere. This ionospheric propagation takes place mainly in the HF band.

The ionosphere is composed of several layers, which vary according to the time of day. Each layer has different propagation characteristics:

D layer – This layer occurs only during the day at altitudes of 60 to 90 km. High absorption takes place at frequencies up to 7 MHz.

E layer – This layer occurs at altitudes of 100 to 125 km. In the summer, dense ionization clouds can form for short periods. These clouds called *sporadic E* can refract radio signals in the VHF spectrum. This phenomenon allows amateur radio operators to communicate over enormous distances.

F layer - This single nighttime layer splits into two layers (F1 and F2) during the day. The F1 layer forms at about 200 km and F2 at about 400 km. The F2 layer propagates most HF short-wave transmissions.

Because radio signals can take many paths to the receiver, multipath fading can occur. If the signals arrive in phase, the result is a stronger signal. If they arrive out of phase with each other, they tend to cancel.

Deep fading, lasting from minutes to hours over a wide frequency range, can occur when solar flares increase the ionization in the D layer.

The useful transmission band ranges between the LUF (lowest usable frequency) and MUF (maximum usable frequency). Frequencies above the MUF are refracted into space. Below the LUF, radio frequencies suffer severe absorption. If a signal is near either of these two extremes, it may be subject to fading.

Meteors create ionization trails that reflect radio waves. Although these trails exist for only a few seconds, they have been successfully used in communications systems spanning 1500 km.

The Aurora Borealis or Northern Lights cause random reflection in the 3 - 5 MHz region. Aurora causes signal flutter at 100 Hz to 2000 Hz thus making voice transmission impossible.

3.3 Fading and Interference

Radio signals may vary in intensity for many reasons.

3.3.1 Flat Earth Reflections (Horizontal Polarization)

There are at least two possible paths for radio waves to travel when the antennas are near the earth: direct path and reflected path. These two signals interact in a very complex manner. However, ignoring polarization and assuming a flat earth can produce some interesting mathematical descriptions.
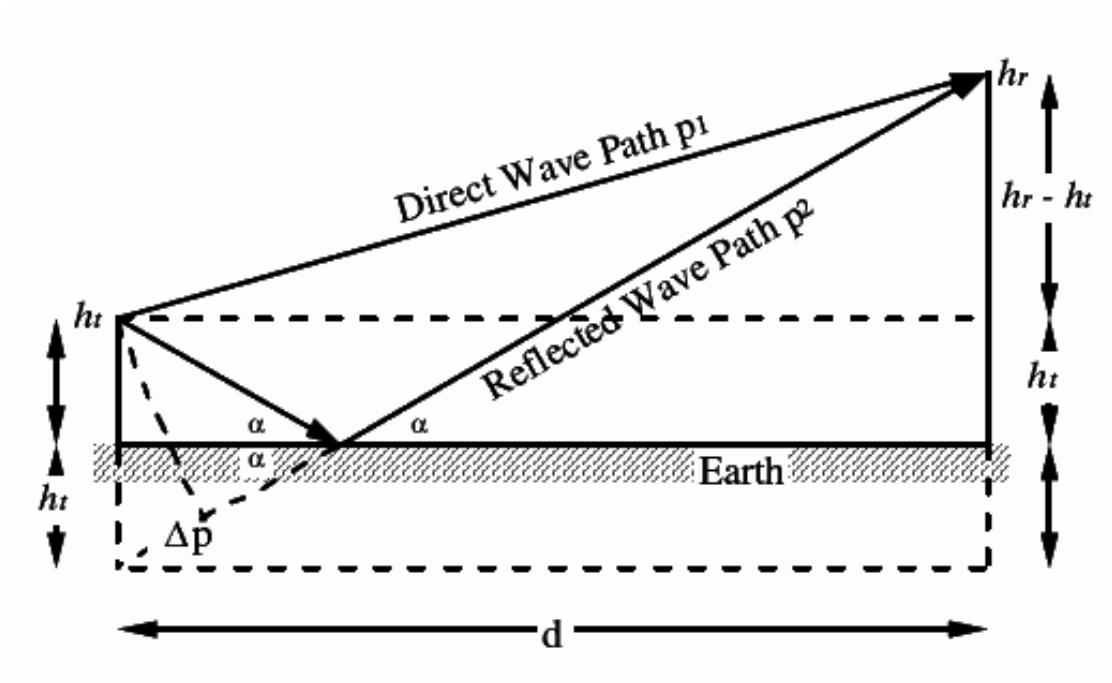
Figure 89

p_1 = direct wave path length

p_2 = reflected wave path length

$\Delta p = p_2$ - p_1 difference in path lengths

d = distance

From the geometry we can observe:

$$p_1^2 = (h_r - h_t)^2 + d^2$$

$$p_2^2 = (h_r + h_t)^2 + d^2$$

$$p_2^2 - p_1^2 = (h_r - h_t)^2 + d^2 - (h_r + h_t)^2 - d^2 = 4h_r h_t$$

$$(p_2 - p_1)(p_2 + p_1) = 4h_r h_t$$

But$\Delta p = (p_2 - p_1)$ and $d \approx p_1 \approx p_2$

$\Delta p 2d \approx 4h_r h_t$ therefore $\Delta p \approx \frac{2h_r h_t}{d}$

If the difference in the two paths Δp, is $1/2$ λ long, the two signals tend to cancel. If Δp is equal to λ, the two signals tend to reinforce. The path difference Δp therefore corresponds to a phase angle change of:

102

$$\varphi_p = \frac{2\pi}{\lambda}\Delta p = \frac{4\pi h_r h_t}{\lambda d}$$

The resultant received signal is the sum of the two components. The situation is unfortunately made more complex by the fact that the phase integrity of the reflected wave is not maintained at the point of reflection.

If we limit the examination of reflected waves to the horizontally polarized situation, we obtain the following geometry:
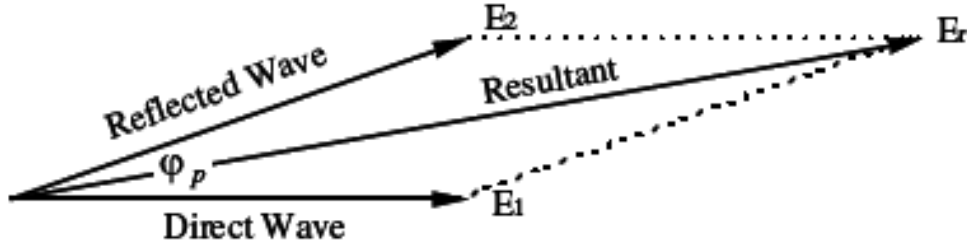
Figure 90

Applying the cosine rule to this diagram, we obtain a resultant signal of:

$$E_r = E_1\sqrt{2\left(1-\cos\varphi_p\right)} = 2E_1\sin\left(\frac{\varphi_p}{2}\right)$$

The signal strength of the direct wave is the unit distance value divided by the distance: $E_r = \frac{E_0}{d}$ Therefore, the received signal can be written as:

$$E_r = \frac{2E_0}{d}\sin\left(\frac{2\pi h_r h_t}{\lambda d}\right)$$

For small angles this can be approximated by:

$$E_r \approx \frac{2E_0}{d}\frac{2\pi h_r h_t}{\lambda d} = E_0\frac{4\pi h_r h_t}{\lambda d^2}$$

3.3.2 Multipath Fading

The received signal is generally a combination of many signals, each coming over a different path. The phase and amplitude of each component are related to the nature of the path. These signals combine in a very complex manner. Some multipath fading effects are characterized by delay spread, Rayleigh and Ricean fading, doppler shifting, etc. Fading is the most significant phenomenon causing signal degradation. There are several different categories of fading:

- Flat fading: the entire pass band of interest is affected equally (also known as narrow or amplitude varying channels).
- Frequency selective fading: certain frequency components are affected more than others (also known as wideband channels). This phenomenon tends to introduce inter-symbol interference.
- Slow fading: the channel characteristics vary at less than the baud rate.
- Fast fading: the channel characteristics vary faster than the baud rate.

Time Dispersion

Time dispersion occurs when signals arrive at different times. Signals traveling at the speed of light move about 1 foot in 1 nanosecond. This spreading tends to limit the bit rate over RF links.

Rayleigh Fading

The Rayleigh distribution can be used to describe the statistical variations of a flat fading channel. Generally, the strength of the received signal falls off as the inverse square of the distance between the transmitter and receiver. However, in cellular systems, the antennas are pointed slightly down and the signal falls of more quickly.
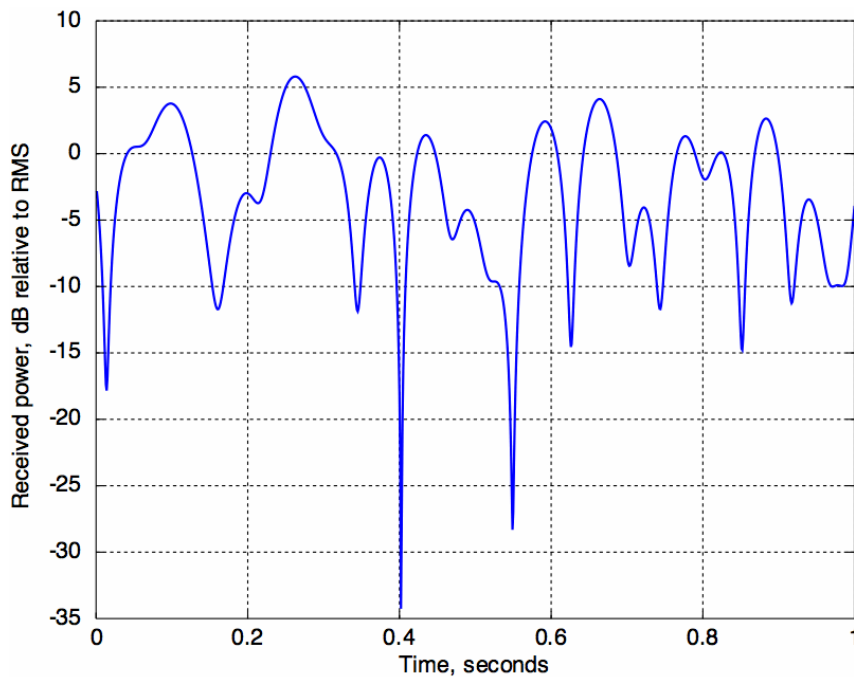
Figure 91

Ricean Fading

The Ricean distribution is used to describe the statistical variations of signals with a strong direct or line-of-sight component and numerous weaker reflected ones. This can happen in any multipath environment such as inside buildings or in an urban center.

A received signal is generally comprised of several signals, each taking a slightly different path. Since some may add constructively in-phase and others out of phase, the overall signal strength may vary by 40 dB or more if the receiver is moved even a very short distance.

Doppler Shift

A frequency shift is caused by the relative motion of the transmitter and receiver, or any object that reflects/refracts signal. This movement creates random frequency modulation. Doppler frequency shift is either positive or negative depending on whether the transmitter is moving towards or away from the receiver.

This Doppler frequency shift is given by:

$$f_d = \frac{v_m}{c} f_c$$

v_m is the relative motion of the transmitter with respect to the receiver, c is the speed of light and f_c is the transmitted frequency. In the multipath environment, the relative movement of each path is generally different. Thus, the signal is spread over a band of frequencies. This is known as the Doppler spread.

3.3.3 Atmospheric Diffraction

Radio waves cannot penetrate very far into most objects. Consequently, there is often a shadow zone behind objects such as buildings, hills, etc.

The radio shadow zone does not have a very sharp cutoff due to spherical spreading, also called Huygens' principle. Each point on a wavefront acts as it were a point source radiating along the propagation path. The overall wavefront is the vector sum of all the point sources or wavelets. The wavelet magnitude is proportional to $1 + \cos\theta$ where θ is measured from the direction of propagation. The amplitude is a maximum in the direction of propagation and zero in the reverse direction.

3.4 Reflection

Reflection normally occurs due to the surface of earth or building & hills which have large dimension relative to the wavelength of the propagation waves. The reflected wave changes the incident angle.

There is similarity b/w the reflection of light by a conducting medium. In both cases, angle of reflection is equal to angle of incidence. The equality of the angles of reflection & incidence follows the second law of reflection for light.

3.5 Diffraction

3.6 Path Loss

3.7 Rayleigh Fading

3.8 Rician Fading

3.9 Doppler Shift

This page is going to talk about the effect of noise on transmission systems.

3.10 Types of Noise

w:Electronic_noise[3]

Most man made electro-magnetic noise occurs at frequencies below 500 MHz. The most significant of these include:

- Hydro lines

- Ignition systems

- Fluorescent lights

- Electric motors

Therefore deep space networks are placed out in the desert, far from these sources of interference.

There are also a wide range of natural noise sources which cannot be so easily avoided, namely:

- *Atmospheric noise* - lighting < 20 MHz

- *Solar noise* - sun - 11 year sunspot cycle

- *Cosmic noise* - 8 MHz to 1.5 GHz

- *Thermal or Johnson noise.* Due to free electrons striking vibrating ions.

- *White noise* - white noise has a constant spectral density over a specified range of frequencies. Johnson noise is an example of white noise.

- *Gaussian noise* - Gaussian noise is completely random in nature however, the probability of any particular amplitude value follows the normal distribution curve. Johnson noise is Gaussian in nature.

---

3   http://en.wikipedia.org/wiki/Electronic_noise

- *Shot noise* - bipolar transistors

  $i_n = \sqrt{2qI_{dc}\Delta f}$

  where q = electron charge 1.6 x 10^{-19} coulombs

- *Excess noise, flicker, 1/f, and pink noise* < 1 KHz are Inversely proportional to frequency and directly proportional to temperature and dc current

- *Transit time noise* - occurs when the electron transit time across a junction is the same period as the signal.

Of these, only Johnson noise can be readily analysed and compensated for. The noise power is given by:

$$P_n = kTB$$

Where:

k = Boltzman's constant (1.38 x 10^{-23} J/K)

T = temperature in degrees Kelvin

B = bandwidth in Hz

This equation applies to copper wire wound resistors, but is close enough to be used for all resistors. Maximum power transfer occurs when the source and load impedance are equal.

3.10.1 Combining Noise Voltages

The instantaneous value of two noise voltages is simply the sum of their individual values at the same instant.

$$v_{totalinst} = v_{1inst} + v_{2inst}$$

This result is readily observable on an oscilloscope. However, it is not particularly helpful, since it does not result in a single stable numerical value such as one measured by a voltmeter.

If the two voltages are coherent [K = 1], then the total rms voltage value is the sum of the individual rms voltage values.

$$v_{totalrms} = v_{1rms} + v_{2rms}$$

If the two signals are completely random with respect to each other [K = 0], such as Johnson noise sources, the total power is the sum of all of the individual powers:

$$P_{totalrandomnoise} = P_{n1random} + P_{n2random}$$

A Johnson noise of power $P = kTB$, can be thought of as a noise voltage applied through a resistor, Thevenin equivalent.

Figure 92

An example of such a noise source may be a cable or transmission line. The amount of noise power transferred from the source to a load, such as an amplifier input, is a function of the source and load impedances.

Thermal Noise Source

Amplifier Input
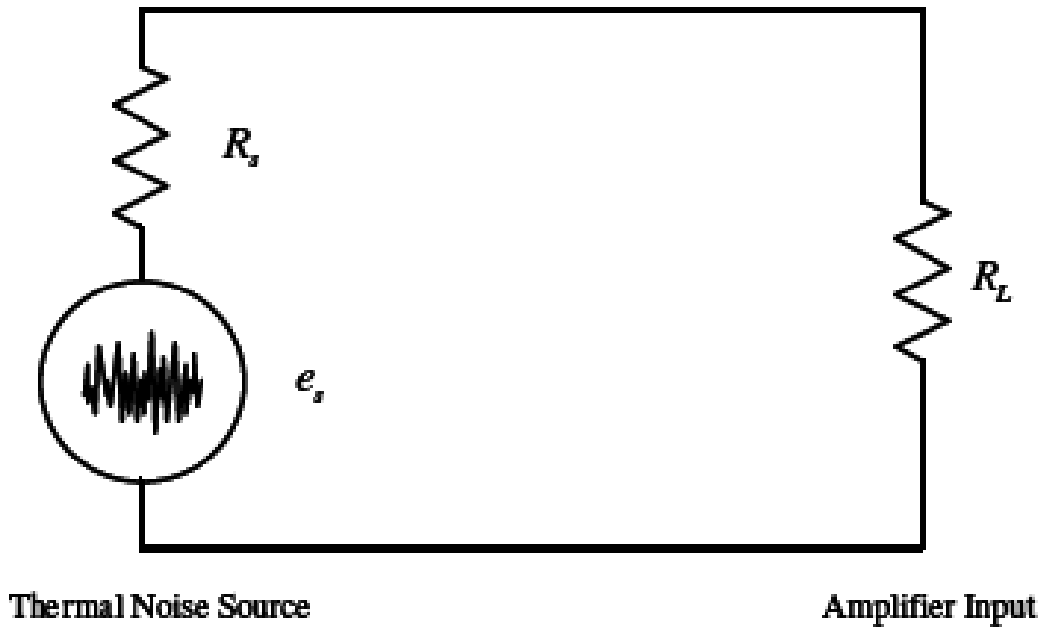
Figure 93

If the load impedance is 0 Ω, no power is transferred to it since the voltage is zero. If the load has infinite input impedance, again no power is transferred to it since there is no current. Maximum power transfer occurs when the source and load impedances are equal.

$$P_{L\max} = \frac{e_s^2}{4R_s}$$

The rms noise voltage at maximum power transfer is:

$$e_n = \sqrt{4RP} = \sqrt{4RkTB}$$

Thermal Noise Source
under matched load conditions
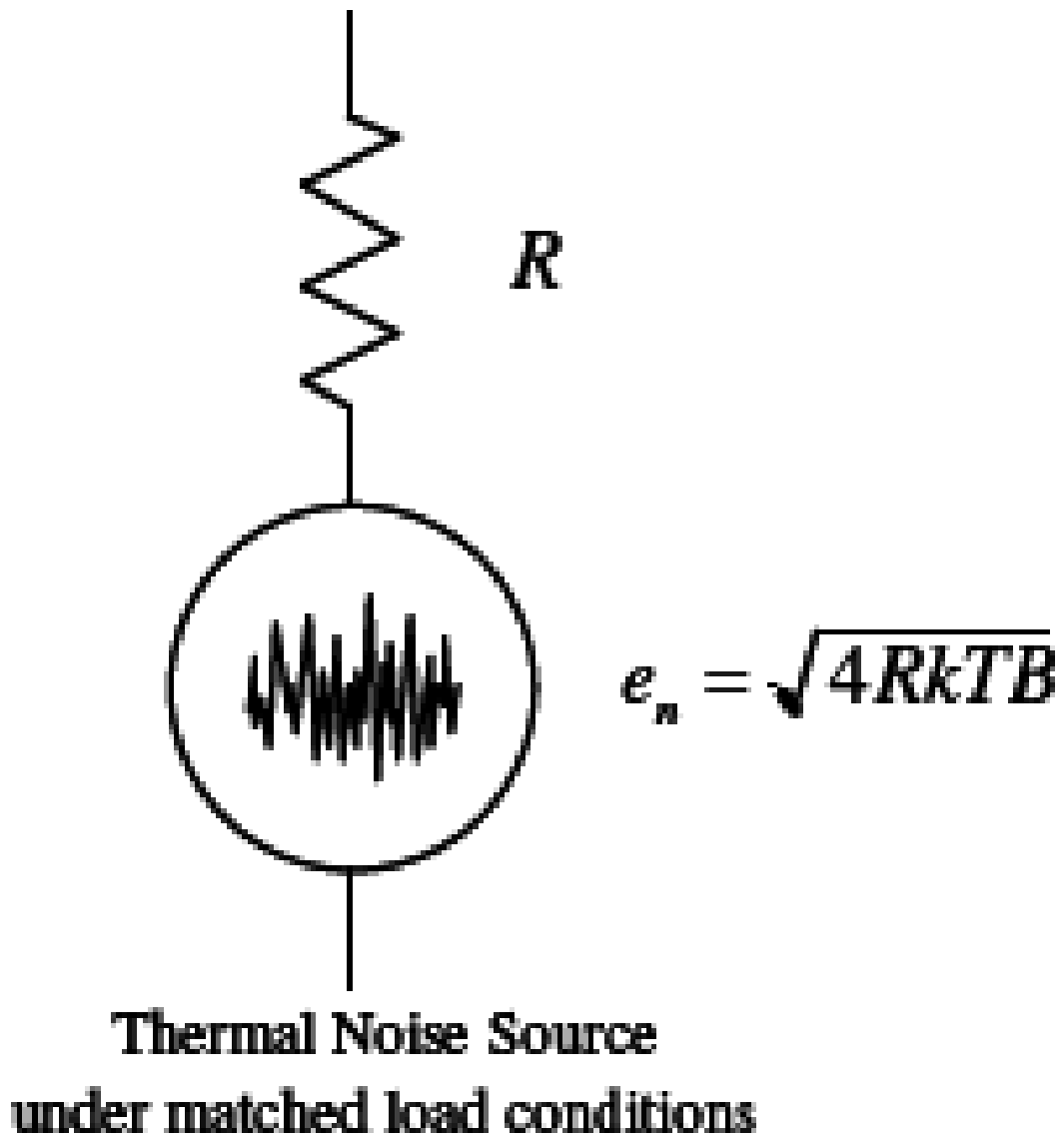
Figure 94

Observe what happens if the noise resistance is resolved into two components:

$$e_n^2 = 4RkTB = 4\left(R_1 + R_2\right)kTB = e_{n1}^2 + e_{n2}^2$$

From this we observe that random noise resistance can be added directly, but random noise voltages add vectorially:

$$E_{Total} = \sqrt{E_{n1}^2 + E_{n2}^2}$$

E_{n2}

E_{n1}
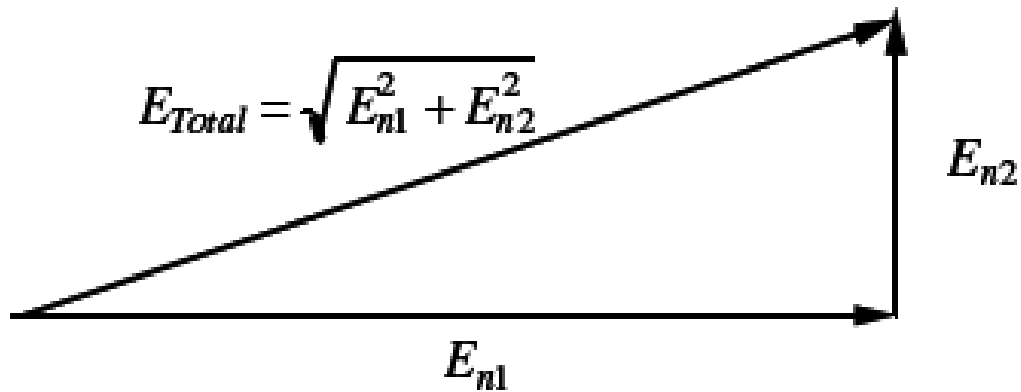
Figure 95

If the noise sources are not quite random, and there is some correlation between them [0 < K < 1], the combined result is not so easy to calculate:

$$P_{Total(\text{notquiterandom})} = \frac{E_1^2 + E_2^2 + 2KE_1E_2}{R_0} = P_1 + P_2 = 2K\sqrt{P_1 + P_2}$$

where

K = correlation $[0 \le K \le 1]$

R_0 = reference impedance

3.11 Noise Temperature

The amount of noise in a given transmission medium can be equated to thermal noise. Thermal noise is well-studied, so it makes good sense to reuse the same equations when possible. To this end, we can say that any amount of radiated noise can be approximated by thermal noise with a given **effective temperature**. Effective temperature is measured in Kelvin. Effective temperature is frequently compared to the **standard temperature**, T_o, which is 290 Kelvin.

In microwave applications, it is difficult to speak in terms of currents and voltages since the signals are more aptly described by field equations. Therefore, temperature is used to characterize noise. The total noise temperature is equal to the sum of all the individual noise temperatures.

3.12 Noise Figure

w:Noise_figure[4]

---

4   http://en.wikipedia.org/wiki/Noise_figure

The terms used to quantify noise can be somewhat confusing but the key definitions are:

Signal to noise ratio: It is either unitless or specified in dB. The S/N ratio may be specified anywhere within a system.

$$\frac{S}{N} = \frac{\text{signalpower}}{\text{noisepower}} = \frac{P_s}{P_n}$$

$$\left(\frac{S}{N}\right)_{dB} = 10\log\frac{P_s}{P_n}$$

Noise Factor (or Noise Ratio): $F = \frac{\left(\frac{S}{N}\right)_{in}}{\left(\frac{S}{N}\right)_{out}}$ (unit less)

Noise Figure: $NF = 10\log F = SNR_{in} - SNR_{out}$ dB

This parameter is specified in all high performance amplifiers and is measure of how much noise the amplifier itself contributes to the total noise. In a perfect amplifier or system, NF = 0 dB. This discussion does not take into account any noise reduction techniques such as filtering or dynamic emphasis.
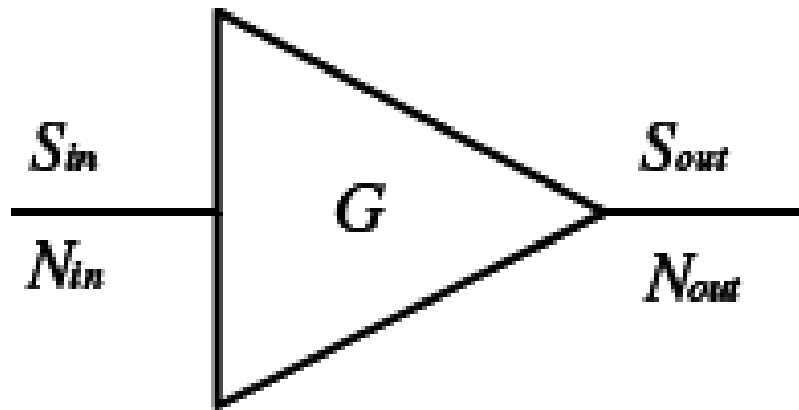
Figure 96

3.12.1 Friiss' Formula & Amplifier Cascades

It is interesting to examine an amplifier cascade to see how noise builds up in a large communication system.

$$F = \frac{\left(\frac{S}{N}\right)_{in}}{\left(\frac{S}{N}\right)_{out}} = \frac{S_{in}}{N_{in}} \times \frac{N_{out}}{S_{out}}$$

Amplifier gain can be defined as: $G = \frac{S_{out}}{S_{in}}$
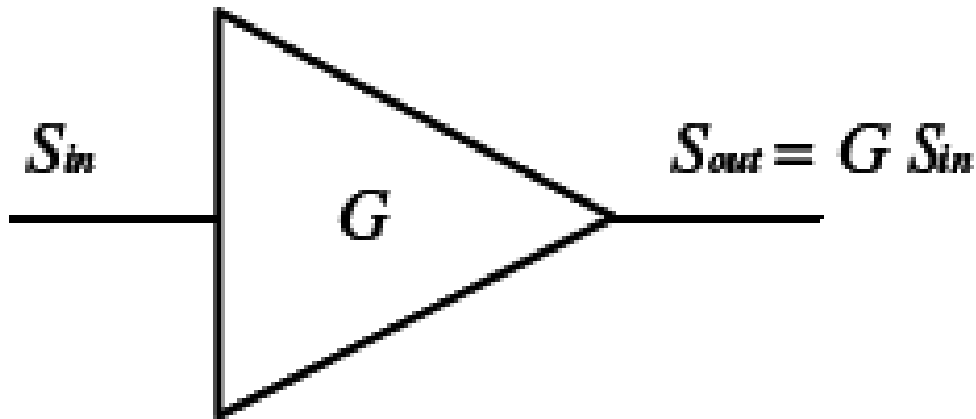
Figure 97

Therefore the output signal power is: $S_{out} = GS_{in}$

and the noise factor (ratio) can be rewritten as: $F = \frac{S_{in}}{N_{in}} \times \frac{N_{out}}{GS_{in}} = \frac{N_{out}}{GN_{in}}$

The output noise power can now be written: $N_{out} = FGN_{in}$

From this we observe that the input noise is increased by the noise ratio and amplifier gain as it passes through the amplifier. A noiseless amplifier would have a noise ratio (factor) of 1 or noise figure of 0 dB. In this case, the input noise would only be amplified by the gain since the amplifier would not contribute noise.

The minimum noise that can enter any system is the Johnson Noise:

$$N_{in(\text{minimum})} = kTB$$

Therefore the minimum noise that can appear at the output of any amplifier is:

$$N_{out(\text{minimum})} = FGkTB$$

The output noise of a perfect amplifier would be ($F = 1$):

$$N_{out(\text{perfect})} = GkTB$$

The difference between these two values is the noised created (added) by the amplifier itself:

$$N_{out(\text{added})} = N_{out(\text{minimum})} - N_{out(\text{perfect})} = FGkTB - GkTB = (F-1)GkTB$$

This is the additional (created) noise, appearing at the output.

The total noise out of the amplifier is then given by:

$$N_{total} = N_{out(\text{perfect})} + N_{out(\text{added})} = GkTB + (F-1)GkTB$$

If a second amplifier were added in series, the total output noise would consist the first stage noise amplified by the second stage gain, plus the additional noise of the second amplifier:

$$N_{total} = G_1 G_2 kTB + (F_1 - 1)G_1 G_2 kTB + (F_2 - 1)G_2 kTB$$

If we divide both sides of this expression by the common term: $G_1 G_2 kTB$

we obtain:

$$\frac{N_{total}}{G_1 G_2 kTB} = \frac{G_1 G_2 kTB + (F_1 - 1)G_1 G_2 kTB + (F_2 - 1)G_2 kTB}{G_1 G_2 kTB}$$

Recall: $F = \frac{N_{out}}{GN_{in}} = \frac{N_{total}}{G_1 G_2 kTB}$

Then: $F_{overall} = F_1 + \frac{F_2 - 1}{G_1}$

This process can be extended to include more amplifiers in cascade to arrive at:

Friiss' Formula

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} +$$

This equation shows that the overall system noise figure is largely determined by the noise figure of the first stage in a cascade since the noise contribution of any stage is divided by the gains of the preceding stages. This is why the 1st stage in any communication system should be an LNA (low noise amplifier).

3.13 Receiver Sensitivity

In a given bandwidth, W, we can show that the noise power N equals:

$$N = F(kT_0)W$$

From N, we can show that the sensitivity of the receiver is equal to

$$SNR \times N$$

3.14 Cascaded Systems

This page will discuss the topic of signal propagation through physical mediums, such as wires.

3.15 Transmission Line Equation

w:Transmission_line[5]

Many kinds of communication systems require signals at some point to be conveyed over copper wires.

The following analysis requires two assumptions:

- • A transmission line can be decomposed into small, distributed passive electrical elements

- • These elements are independent of frequency (i.e. although *reactance* is a function of frequency, resistance, capacitance and inductance are not)

These two assumptions limit the following analysis to frequencies up to the low MHz region. The second assumption is particularly difficult to defend since it is well known that the resistance of a wire increases with frequency because the conduction cross-section decreases. This phenomenon is known as the skin effect and is not easy to evaluate.
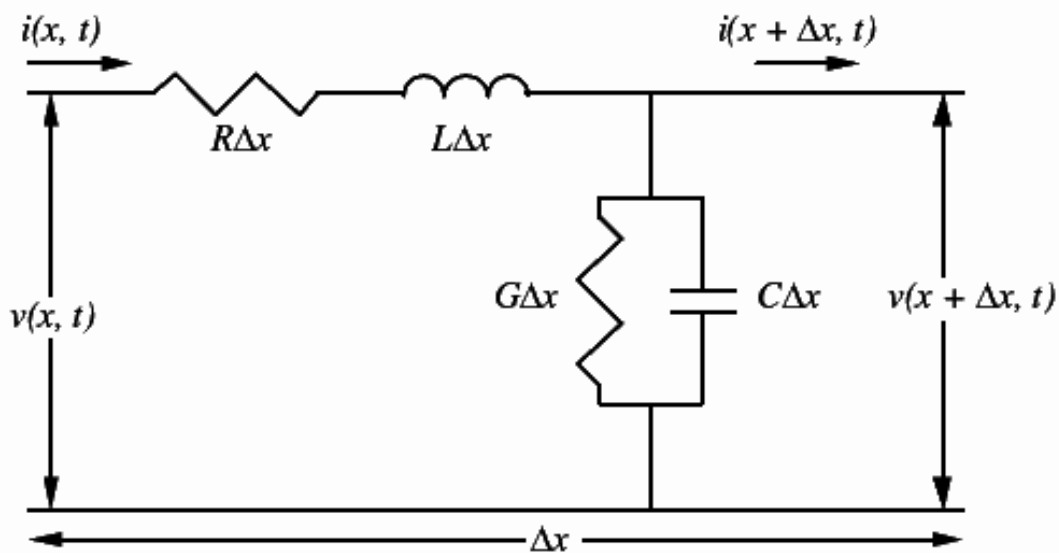
Distributed Element Model of a Transmission Line

Figure 98

The purpose behind the following mathematical manipulation is to obtain an expression that defines the voltage (or current) at any time (t) along any portion (x) of the transmission line. Later, this analysis will be extended to include the frequency domain.

Recall the characteristic equations for inductors and capacitors:

$v = L\frac{\partial i}{\partial t}$ and $i = C\frac{\partial v}{\partial t}$

---

5   http://en.wikipedia.org/wiki/Transmission_line

3.15.1 Kirchoff's Voltage Law

Kirchoff's voltage law (KVL) simply states that the sum of all voltage potentials around a closed loop equal zero. Or in other words, if you walked up a hill and back down, the net altitude change would be zero.

Applying KVL in the above circuit, we obtain:

$$v(x,t) = R\Delta x i(x,t) + L\Delta x \frac{\partial i}{\partial t}(x,t) + v(x+\Delta x,t)$$

Rearranging:

$$v(x,t) - v(x+\Delta x,t) = R\Delta x i(x,t) + L\Delta x \frac{\partial i}{\partial t}(x,t)$$

But the LHS (left hand side) of the above equation, represents the voltage drop across the cable element Δv, therefor:

$$\Delta v = R\Delta x i(x,t) + L\Delta x \frac{\partial i}{\partial t}(x,t)$$

Dividing through by Δx, we obtain:

$$\frac{\Delta v}{\Delta x} = Ri(x,t) + L\frac{\partial i}{\partial t}(x,t)$$

The LHS is easily recognized as a derivative. Simplifying the notation:

$$\frac{\partial v}{\partial x} = Ri + L\frac{\partial i}{\partial t}$$

This expression has both current and voltage in it. It would be convenient to write the equation in terms of current or voltage as a function of distance or time.

3.15.2 Simplifying the Equation (trust me)

The first step in separating voltage and current is to take the derivative with respect to the position x (**Equation 1**):

$$\frac{\partial^2 v}{\partial x^2} = R\frac{\partial i}{\partial x} + L\frac{\partial^2 i}{\partial x \partial t}$$

The next step is to eliminate the current terms, leaving an expression with voltage only. The change in current along the line is equal to the current being shunted across the line through the capacitance C and conductance G. By applying KCL in the circuit, we obtain the necessary information (**Equation 2**):

$$\frac{\partial i}{\partial x} = Gv + C\frac{\partial v}{\partial t}$$

Taking the derivative with respect to time, we obtain (**Equation 3**):

$$\frac{\partial^2 i}{\partial x \partial t} = G\frac{\partial v}{\partial t} + C\frac{\partial^2 v}{\partial t^2}$$

Substituting (**Equation 2**) and (**Equation 3**) into (**Equation 1**), we obtain the desired simplification:

$$\frac{\partial^2 v}{\partial x^2} = R\left[Gv + C\frac{\partial v}{\partial t}\right] + L\left[G\frac{\partial v}{\partial t} + C\frac{\partial^2 v}{\partial t^2}\right]$$

Collecting the terms, we obtain:

The Transmission Line Equation for Voltage

$$\frac{\partial^2 v}{\partial x^2} = RGv + (RC + LG)\frac{\partial v}{\partial t} + LC\frac{\partial^2 v}{\partial t^2}$$

This equation is known as the transmission line equation. Note that it has voltage at any particular location x as a function of time t.

Similarly for current, we obtain:

The Transmission Line Equation for Current

$$\frac{\partial^2 i}{\partial x^2} = RGi + (RC + LG)\frac{\partial i}{\partial t} + LC\frac{\partial^2 i}{\partial t^2}$$

But we're not quite done yet.

3.15.3 Solving the Transmission Line Equation

Historically, a mathematician would solve the transmission line equation for v by assuming a solution for v, substituting it into the equation, and observing whether the result made any sense. An engineer would follow a similar procedure by making an "educated guess" based on some laboratory experiments, as to what the solution might be. Today there are more sophisticated techniques used to find solutions. In this respect, the engineer may lag behind the mathematician by several centuries in finding applications for mathematical tools.

To solve the transmission line equation, we shall guess that the solution for the voltage function is of the form:

$$v(t) = e^{j\omega t}e^{-\gamma x}$$

The first term represents a unity vector rotating at an angular velocity of ω radians per second, in other words, a sine wave of some frequency. The second term denotes the sinusoid being modified by the transmission line, namely its amplitude decaying exponentially with distance. If we let γ be a complex quantity, we can also include any phase changes which occur as the signal travels down the line.

The sine wave is used as a signal source because it is easy to generate, and manipulate mathematically. Euler's Identity shows the relationship between exponential notation and trigonometric functions:

Euler's Identity

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t)$$

Going back to our educated guess, we will let $\gamma = \alpha + j\beta$, therefore:

$$v(t) = e^{j\omega t}e^{-(\alpha+j\beta)x} = e^{-\alpha x}e^{(\omega t-\beta x)j}$$

The term $e^{-\alpha x}$ represents the exponential amplitude decay as this signal travels down the line. α is known as the attenuation coefficient and is expressed in Nepers per meter.

The term $e^{(\omega t-\beta x)j}$ represents the frequency of the signal at any point along the line. β component is known as the phase shift coefficient, and is expressed in radians per meter.

Substituting our *educated guess*

$$v(t) = e^{j\omega t}e^{-(\alpha+j\beta)x}$$

into the transmission line equation for voltage, we obtain:

$$\frac{\partial^2}{\partial x^2}\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right] = RG\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right] + (RC+LG)\frac{\partial}{\partial t}\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right] + LC\frac{\partial^2}{\partial t^2}\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right]$$

This looks pretty intimidating, but if you can do basic differentials and algebra, you can do this!

Simplifying the Equation (trust me)

The idea now is to work through the math to see if we come up with a reasonable solution. If we arrive at a contradiction or an unreasonable result, it means that our educated guess was wrong and we have to do more experimenting and come up with a better guess as to how voltage and current travel down a transmission line.

Let's look at this equation one term at a time:

LHS = RHS Term 1 + RHS Term 2 + RHS Term 3

Starting with the left hand side (LHS) we get the following simplification:

$$\frac{\partial^2}{\partial x^2}\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right] = \frac{\partial}{\partial x}\left[-(\alpha+j\beta)e^{j\omega t}e^{-(\alpha+j\beta)x}\right] = (\alpha+j\beta)^2 e^{j\omega t}e^{-(\alpha+j\beta)x}$$

Believe it or not, the RHS Term 1 does not need simplifying.

Simplifying the RHS Term 2, we obtain:

$$(RC + LG)\frac{\partial}{\partial t}\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right] = (RC + LG)\,j\omega\left(e^{j\omega t}e^{-(\alpha+j\beta)x}\right)$$

Simplifying the RHS Term 3, we obtain:

$$LC\frac{\partial^2}{\partial t^2}\left[e^{j\omega t}e^{-(\alpha+j\beta)x}\right] = LC\frac{\partial}{\partial t}\left[j\omega e^{j\omega t}e^{-(\alpha+j\beta)x}\right] = -LC\omega^2 e^{j\omega t}e^{-(\alpha+j\beta)x}$$

Let's put it all back together:

$$(\alpha+j\beta)^2\,e^{j\omega t}e^{-(\alpha+j\beta)x} = RG\left(e^{j\omega t}e^{-(\alpha+j\beta)x}\right) + (RC+LG)\,j\omega\left(e^{j\omega t}e^{-(\alpha+j\beta)x}\right) - LC\omega^2\left(e^{j\omega t}e^{-(\alpha+j\beta)x}\right)$$

Note that each of the four terms contain the expression $e^{j\omega t}e^{-(\alpha+j\beta)x}$.

Therefore we end up with:

$$(\alpha+j\beta)^2 = RG + (RC+LG)\,j\omega - LC\omega^2$$

And this can be further simplified to:

Attenuation and Phase Shift Coefficients

$$\alpha + j\beta = \gamma = \sqrt{(R+j\omega L)(G+j\omega C)}$$

This result is not self contradictory or unreasonable. Therefore we conclude that our educated guess was right and we have successfully found an expression for attenuation and phase shift on a transmission line as a function of its distributed electrical components and frequency.

3.15.4 Lossless Transmission Line

Signal loss occurs by two basic mechanisms: signal power can be dissipated in a resistor [or conductance] or signal currents may be shunted to an AC ground via a reactance. In transmission line theory, a lossless transmission line does not dissipate power. Signals, will still gradually diminish however, as shunt reactances return the current to the source via the ground path. For the power loss to equal zero, R = G = 0. This condition occurs when the transmission line is very short. An oscilloscope probe is an example of a very short transmission line. The transmission line equation reduces to the voltage equation:

$$\frac{\partial^2 v}{\partial x^2} = LC\frac{\partial^2 v}{\partial t^2}$$

and the current equation:

$$\frac{\partial^2 i}{\partial x^2} = LC\frac{\partial^2 i}{\partial t^2}$$

To determine how sinusoidal signals are affected by this type of line, we simply substitute a sinusoidal voltage or current into the above expressions and solve as before, or we could take a much simpler approach. We could start with the solution for the general case:

$$\alpha + j\beta = \gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$$

Let $R = G = 0$, and simplify:

$$\alpha + j\beta = \sqrt{(j\omega L)(j\omega C)} = \omega (LC)^2 j$$

Equating the real and imaginary parts:

$$\alpha = 0$$

$$\beta = \omega\sqrt{LC}$$

This expression tells us that **a signal travelling down a lossless transmission line, experiences a phase shift directly proportional to its frequency.**

Phase Velocity

A new parameter, known as phase velocity, can be extracted from these variables:

$V_p = \frac{1}{\sqrt{LC}} = \frac{\omega}{\beta}$ meters per second

Phase velocity is the speed at which a fixed point on a wavefront, appears to move. In the case of wire transmission lines, it is also the velocity of propagation., typically: 0.24c < Vp < 0.9c .

The distance between two identical points on a wavefront is its wavelength (λ) and since one cycle is defined as 2π radians:

$\lambda = \frac{2\pi}{\beta}$ and $\omega = 2\pi f$

therefore:

$$V_p = \lambda f$$

In free space, the phase velocity is 3 x 10^8 meters/sec, the speed of light. In a cable, the phase velocity is somewhat lower because the signal is carried by electrons. In a waveguide transmission line, the phase velocity exceeds the speed of light.

3.15.5 Distortionless Transmission Line

A distortionless line does not distort the signal phase, but does introduce a signal loss. Since common transmission lines are not super conductors, the signal will decrease in amplitude but retain the same shape as the input. This characteristic is essential for long cable systems.

Phase distortion does not occur if the phase velocity V_{p} is constant at all frequencies.

By definition, a phase shift of 2π radians occurs over one wavelength λ.

Since

$$V_p = \lambda f \qquad \lambda = \frac{2\pi}{\beta} \qquad f = \frac{\omega}{2\pi}$$

Then:

$$V_p = \frac{2\pi}{\beta} \times \frac{\omega}{2\pi} = \frac{\omega}{\beta}$$

This tells us that in order for phase velocity V_{p} to be constant, the phase shift coefficient β, must vary directly with frequency ω.

Recall

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta$$

The problem now is to find β. This can be done as follows:

$$\gamma = \sqrt{\left(\frac{R + j\omega L}{j\omega L}\right)(j\omega L)\left(\frac{G + j\omega C}{j\omega C}\right)(j\omega C)} = j\omega\sqrt{LC}\sqrt{1 + \frac{R}{j\omega L}}\sqrt{1 + \frac{G}{j\omega C}}$$

It may seem that we have lost β, but do not give up. The 2nd and 3rd roots can be expanded by means of the Binomial Expansion.

Recall:

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

In this instance $n = 1/2$. Since the contribution of successive terms diminishes rapidly, γ is expanded to only 3 terms:

$$\gamma \approx j\omega\sqrt{LC}\left(1 + \frac{1}{2}\frac{R}{j\omega L} - \frac{1}{8}\left(\frac{R}{j\omega L}\right)^2\right)\left(1 + \frac{1}{2}\frac{G}{j\omega C} - \frac{1}{8}\left(\frac{G}{j\omega C}\right)^2\right)$$

This may seem complex, but remember it is only algebra and it will reduce down to simple elegance. Expanding the terms we obtain:

$$\gamma \approx j\omega\sqrt{LC}\left\{\begin{array}{l} 1 + \frac{1}{2}\frac{G}{j\omega C} - \frac{1}{8}\left(\frac{G}{j\omega C}\right)^2 + \frac{1}{2}\frac{R}{j\omega L} - \frac{1}{4}\frac{RG}{\omega^2 LC} \\ -\frac{1}{16}\frac{R}{j\omega L}\left(\frac{G}{j\omega C}\right)^2 - \frac{1}{8}\left(\frac{R}{j\omega L}\right)^2 \\ -\frac{1}{16}\left(\frac{R}{j\omega L}\right)^2\frac{G}{j\omega C} + \frac{1}{64}\left(\frac{R}{j\omega L}\right)^2\left(\frac{G}{j\omega C}\right)^2 \end{array}\right\}$$

Since $\gamma = \alpha + j\beta$, we merely have to equate the real and imaginary terms to find β.

$$\beta \approx \omega\sqrt{LC}\left\{1 + \underbrace{\frac{1}{8}\left(\frac{G}{\omega C}\right)^2 - \frac{1}{4}\frac{RG}{\omega^2 LC} + \frac{1}{8}\left(\frac{R}{\omega L}\right)^2}_{Difference\,of\,squares} + \underbrace{\frac{1}{64}\left(\frac{R}{\omega L}\right)^2\left(\frac{G}{\omega C}\right)^2}_{Very\,small}\right\}$$

Or

$$\beta \approx \omega\sqrt{LC}\left\{1 + \frac{1}{8}\left(\frac{R}{\omega L} - \frac{G}{\omega C}\right)^2\right\}$$

Note that if $\frac{R}{\omega L} = \frac{G}{\omega C}$ then $\beta \approx \omega\sqrt{LC}$

From this we observe that β is directly proportional to ω.

Therefore the **requirement for distortionless transmission is:**

$RC = LG$

w:Heaviside_condition[6]

This is one of the essential design characteristics for a broadband coax cable network.

If we equate the real terms, we obtain:

$$\alpha \approx \sqrt{RG}$$

So there is a reason to study algebra after all!

3.16 The Frequency Domain

Signal analysis is often performed in the frequency domain. This tells us how the transmission line affects the spectral content of the signals they are carrying.

To determine this, it is necessary to find the Fourier Transform of the transmission line equation. Recall:

$$\frac{\partial^2 v}{\partial x^2} = RGv + (RC + LG)\frac{\partial v}{\partial t} + LC\frac{\partial^2 v}{\partial t^2}$$

and recall (hopefully) the Fourier Transform (which converts the time domain to the frequency domain):

$$\mathbb{F}\{f(t)\} = F(\omega) = \int\limits_{-\infty}^{\infty} e^{-j\omega t} f(t)\,dt$$

---

6   http://en.wikipedia.org/wiki/Heaviside_condition

To prevent this analysis from 'blowing up', we must put a stipulation on the voltage function namely, that it vanishes to zero at an infinite distance down the line. This comprises a basic boundary condition.

$$\text{let} \quad v \to 0 \quad \text{as} \quad x \to \infty$$

This stipulation is in agreement with actual laboratory experiments. It is well known that the signal magnitude diminishes as the path lengthens.

Likewise, a time boundary condition, that the signal was zero at some time in the distant past and will be zero at some time in the distant future, must be imposed.

$$\text{let} \quad v \to 0 \quad \text{as} \quad t \to \infty$$

Although engineers have no difficulty imposing these restrictions, mathematical purists, are somewhat offended. For this and other reasons, other less restrictive transforms have been developed. The most notable in this context, is the Laplace transform, which does not have the same boundary conditions.

Having made the necessary concessions in order to continue our analysis, we must find the Fourier Transform corresponding to the following terms:

$$\mathbb{F}\{v\} \qquad \mathbb{F}\left\{\frac{\partial v}{\partial t}\right\} \qquad \mathbb{F}\left\{\frac{\partial^2 v}{\partial t^2}\right\}$$

$$\text{Let:} \qquad \mathbb{F}\{v\} = V$$

Then applying the transform on the derivative, we obtain:

$$\mathbb{F}\left\{\frac{\partial v}{\partial t}\right\} = \int_{-\infty}^{\infty} e^{-j\omega t}\frac{\partial v}{\partial t}dt$$

This equation can be solved by using integration by parts:

$$\int u\,dv = uv - \int v\,du$$

$$\text{let} \quad u = e^{-j\omega t} \qquad \therefore du = -j\omega e^{-j\omega t}$$

$$\text{and} \quad dv = \frac{\partial v}{\partial t} \qquad \therefore v = v$$

$$\therefore \mathbb{F}\left\{\frac{\partial v}{\partial t}\right\} = e^{-j\omega t}\,v|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v\left(-j\omega e^{-j\omega t}\right)dt$$

Applying the boundary conditions when t goes to infinity makes the 1st term disappear.

$$\therefore \mathbb{F}\left\{\frac{\partial v}{\partial t}\right\} = j\omega \int\limits_{-\infty}^{\infty} e^{-j\omega t} v \, dt$$

Note that the resulting integral is simply the Fourier Transform. In other words:

$$\mathbb{F}\left\{\frac{\partial v}{\partial t}\right\} = j\omega \mathbb{F}\left\{v\right\} = j\omega V$$

similarly:

$$\mathbb{F}\left\{\frac{\partial^2 v}{\partial t^2}\right\} = (j\omega)^2 \mathbb{F}\left\{v\right\} = (j\omega)^2 V$$

We can now write the transmission line equation in the frequency domain:

$$\frac{\partial^2 V}{\partial x^2} = RGV + (RC + LG) j\omega V + LC (j\omega)^2 V$$

where:

$$V = V(\omega) = \mathbb{F}\left\{v(t)\right\}$$

Rearranging the terms, we obtain:

$$\frac{\partial^2 V}{\partial x^2} = [RG + (RC + LG) j\omega + (j\omega L)(j\omega C)] V$$

or

$$\frac{\partial^2 V}{\partial x^2} = [(R + j\omega L)(G + j\omega C)] V$$

since:

$$\sqrt{(R + j\omega L)(G + j\omega C)} = \alpha + j\beta = \gamma$$

then

$$\frac{\partial^2 V}{\partial x^2} = \gamma^2 V$$

or

$$\frac{\partial^2 V}{\partial x^2} - \gamma^2 V = 0$$

This represents the most general form of the transmission line equation in the frequency domain. This equation must now be solved for V to observe how voltage (or current) varies with distance and frequency. This can be done by assuming a solution of the form:

$$V = \underbrace{Ae^{-\gamma x}}_{\text{forward wave}} + \underbrace{Be^{\gamma x}}_{\text{reverse wave}}$$

These terms represent an exponential decay as the signal travels down the transmission line. If we ignore any reflections, assuming that the cable is infinitely long or properly terminated, this simplifies to:

$$V = V_0 e^{-\gamma x}$$

To verify whether this assumption is correct, substitute it into the equation, and see if a contradiction occurs. If there is no contradiction, then our assumption constitutes a valid solution.

$$\frac{\partial^2}{\partial x^2} V_0 e^{-\gamma x} - \gamma^2 V_0 e^{-\gamma x} = 0$$

$$\frac{\partial}{\partial x} \left(-\gamma^2 V_0 e^{-\gamma x} \right) - \gamma^2 V_0 e^{-\gamma x} = 0$$

$$\gamma^2 V_0 e^{-\gamma x} - \gamma^2 V_0 e^{-\gamma x} = 0$$

$$0 = 0$$

Thus we validate the assumed solution. This tells us that in the frequency domain, the voltage or current on a transmission line decays exponentially:

$$V = V_0 e^{-\gamma x}$$

where:

$$\gamma = \sqrt{(R + j\omega)(G + j\omega)} = |\gamma| \angle \varphi = \alpha + j\beta$$

$$\gamma = \text{ propagation constant}$$

$$\alpha = \text{ attenuation coeficient}$$

$$\beta = \text{phase coefficient}$$

In exponential notation, a sinusoid may be represented by a rotating unity vector, of some frequency:

$$e^{j\omega t} = \cos \omega t + j \sin \omega t$$

Note that the magnitude of this function is 1, but the phase angle is changing as a function of t.

If we let: $V_0 = e^{j\omega t}$

Then: $V_0 = e^{j\omega t} e^{-\gamma x} = e^{j\omega t} e^{-(\alpha + j\beta)x} = \underbrace{e^{-\alpha x}}_{attenuation\, vs.\, x} \overbrace{e^{j(\omega t - \beta x)}}^{phase\, vs.\, t\, and\, x}$

This result is quite interesting because it is the same solution for the transmission line equation in the time domain. The term $e^{-\alpha x}$ represents an exponential decay. The signal is attenuated as length x increases. The amount of attenuation is defined as:

Attenuation in Nepers: $N = |\ln e^{-\alpha x}| = \alpha x$

Attenuation in dB: $= 20 \log e^{-\alpha x} \approx 8.68589 \alpha x$

This allows us to determine the attenuation at any frequency at any point in a transmission line, if we are given the basic line parameters of R, L, G, $\&$ C.

The term $e^{j(\omega t - \beta x)}$ represents a rotating unity vector since:

$$e^{j(\omega t - \beta x)} = \cos(\omega t - \beta x) + j \sin(\omega t - \beta x)$$

The phase angle of this vector is βx radians.

3.17 Characteristic Impedance

The characteristic impedance of a transmission line is also known as its surge impedance, and should not be confused with its resistance. If a line is infinitely long, electrical signals will still propagate down it, even though the resistance approaches infinity. The characteristic impedance is determined from its AC attributes, not its DC ones.

Recall from our earlier analysis:

This page will discuss Wireless EM wave propagation, and some basics about antennas.

3.18 Isotropic Antennas

In communication we talk about 'antennas'; insects have 'antennae'

An isotropic antenna radiates it's transmitted power equally in all directions. This is an ideal model; all real antennas have at least some directionality associated with them. However, it is mathematically convenient, and good enough for most purposes.

A radio antenna is an example of an isotropic antenna

3.18.1 Power Flux Density

If the transmitted power is spread evenly across a sphere of radius R from the antenna, we can find the power per unit area of that sphere, called the **Power Flux Density** using the Greek letter Φ (capital phi) and the following formula:

$$\Phi = \frac{P_T}{4\pi R^2}$$

Where P_T is the total transmitted power of the signal.

3.18.2 Effective Area

The **effective area** of an antenna is the equivalent amount of area of transmission power, from a non-ideal isotropic antenna that appears to be the area from an ideal antenna. For instance, if our antenna is non-ideal, and 1 meter squared of area can effectively be modeled as .5 meters squared from an ideal antenna, then we can use the ideal number in our antenna. We can relate the actual area and the effective area of our antenna using the *antenna efficiency* number, as follows:

$$\eta = \frac{A_e}{A}$$

The area of an ideal isotropic antenna can be calculated using the wavelength of the transmitted signal as follows:

$$A = \frac{\lambda^2}{4\pi}$$

3.18.3 Received Power

The amount of power that is actually received by a receiver placed at distance R from the isotropic antenna is denoted P_R, and can be found with the following equation:

$$P_R = \Phi_R A_e$$

Where Φ_R is the power flux density at the distance R. If we plug in the formula for the effective area of an ideal isotropic antenna into this equation, we get the following result:

$$P_R = \frac{P_T}{(4\pi R/\lambda)^2} = \frac{P_T}{L_P}$$

Where L_P is the path-loss, and is defined as:

$$L_P = \left(\frac{4\pi R}{\lambda}\right)^2$$

The amount of power lost across freespace between two isotropic antenna (a transmitter and a receiver) depends on the wavelength of the transmitted signal.

3.19 Directional Antennas

A directional antenna, such as a parabolic antenna, attempts to radiate most of its power in the direction of a known receiver.

A "satellite dish" is an example of a parabolic antenna

Here are some definitions that we need to know before we proceed:

Azimuth Angle

The Azimuth angle, often denoted with a θ (Greek lower-case Theta), is the angle that the direct transmission makes with respect to a given reference angle (often the angle of the target receiver) when looking down on the antenna from above.

Elevation Angle

The elevation angle is the angle that the transmission direction makes with the ground. Elevation angle is denoted with a φ (Greek lower-case phi)

3.19.1 Directivity

Given the above definitions, we can define the transmission gain of a directional antenna as a function of θ and φ, assuming the same transmission power:

$$G_T(\theta,\phi) = \frac{\Phi_{\theta,\,\phi}}{\Phi_{isotropic}}$$

3.19.2 Effective Area

The effective area of a parabolic antenna is given as such:

$$A_e = \eta\frac{\pi D^2}{4}$$

3.19.3 Transmit Gain

$$G_{max} = \frac{4\pi A_e}{\lambda^2}$$

If we are at the transmit antenna, and looking at the receiver, the angle that the transmission differs from the direction that we are looking is known as Ψ (Greek upper-case Psi), and we can find the transmission gain as a function of this angle as follows:

$$G(\Psi) = \left(\frac{2J_1((\pi D/\lambda)sin(\Psi))}{sin(\Psi)}\right)^2 \left(\frac{\lambda}{\pi D}\right)^2$$

Where $J_1(\ )$ denotes the first-order bessel function.

3.19.4 Friis Equation

The **Friis Equation** is used to relate several values together when using directional antennas:

$$P_R = \frac{P_T G_T G_R}{L_P}$$

The Friis Equation is the fundamental basis for **link-budget analysis**.

3.20 Link-Budget Analysis

If we express all quantities from the Friis Equation in decibels, and divide both sides by the noise-density of the transmission medium, N0, we get the following equation:

$$C/N_0 = EIRP - L_P + (G_R/T_e) - k$$

Where C/N0 is the received carrier-to-noise ratio, and we can decompose N0 as follows:

$$N_0 = kTe$$

k is Boltzmann's constant, (-228.6dBW) and Te is the effective temperature of the noise signal (in degrees Kelvin). EIRP is the "Equivalent Isotropic Radiated Power", and is defined as:

$$EIRP = G_T P_T$$

To perform a link-budget analysis, we add all the transmission gain terms from the transmitter, we add the receive gain divided by the effective temperature, and we subtract boltzman's constant and all the path losses of the transmission.

3.21 Further reading

- Jean-Claude Wippler. "What if you're out of wireless range?"[7]. 2013.

This page is all about **Space-Division Multiplexing** (SDM).

- What is SDM: When we want to transmit multiple messages, the goal is maximum reuse of the given resources: time and frequency. Time-Division Multiplexing (TDM), operates by dividing the time up into time slices, so that the available time can be reused. Frequency-Division Multiplexing (FDM), operates by dividing up the frequency into transmission bands, so that the frequency spectrum can be reused. However, if we remember our work with directional antennas, we can actually reuse both time and frequency, by transmitting our information along parallel channels. This is known as **Space-Division Multiplexing**.

3.22 Technical categorisations

3.22.1 Spatial Coding

3.23 Multipathing

3.24 Application systems

3.24.1 MIMO Systems

3.24.2 Smart antenna

---

7   `http://jeelabs.org/2013/06/05/what-if-youre-out-of-wireless-range/`

4 Digital Modulation

4.1 Definition

What is PAM? Pulse-Amplitude Modulation is "pulse shaping". Essentially, communications engineers realize that the shape of the pulse in the time domain can positively or negatively affect the characteristics of that pulse in the frequency domain. There is no one way to shape a pulse, there are all sorts of different pulse shapes that can be used, but in practice, there are only a few pulse shapes that are worth the effort. These chapters will discuss some of the common pulses, and will develop equations for working with any generic pulse.

4.2 Square Wave

The most logical way to transmit a digital signal is through a stream of pulses. One distinct pulse for a digital "1", and another distinct pulse for a digital "0". Intuitively, a square pulse will transmit this data, and there are a number of different ways to transmit the data using

The square wave is a basic choice for transmitting digital data because it is easy to transmit, and is generally easy to receive. If we take the fourier transform of a square wave, we get a sinc function. A sinc function is a never-ending function, which means that a square wave in the time domain has a very wide bandwidth. When using a square wave, there will always be a trade-off, because high-frequency components of the square wave will be attenuated by the channel, and the resultant waveform will be more prone to error on the other end.

4.2.1 Unipolar Square Wave

A unipolar square wave is a wave where a logical 1 is transmitted using a square pulse of height A. Then a logical 0 is transmitted with a 0 voltage.

4.2.2 Bipolar Square Wave

A bipolar square wave is a square wave where a 1 is transmitted with a pulse of height A/2, and a 0 is transmitted with a pulse of -A/2.

4.3 Other pulses

It turns out that simply by changing the shape of the pulse (changing it away from a square wave), we can create a signal with better properties: lower bandwidth, lower error rate, etc... This section will talk about a few different pulses.

4.4 Sinc

By the property of duality, however, we can see that if we have a sinc wave in the time domain, we will have a square-shape in the frequency domain. This is an interesting result, because it means that we can transmit a sinc shape with definite bandwidth bounds, and it can travel through a channel in its entirety without being attenuated, or losing any harmonics!

4.5 Comparison

Here we will show a basic comparison between square pulses and sinc pulses:

Metric	Square Pulse	Sinc Pulse
Bandwidth	Large bandwidth	Small bandwidth
Jitter	Not susceptible to Jitter	Very susceptible to Jitter
Noise	Very susceptible to Noise	Susceptible to Noise
ISS	Not affected by ISS	Affected heavily by ISS

4.6 slew-rate-limited pulses

Most systems that use RS232, RS485, or high speed clocks use slew-rate-limited drivers. Some use driver chips with integrated slew-rate limiting; other systems use digital driver chips that put sharp square pulses on their output pins, then an external slew-rate-limiting resistor[1] between that output pin and the rest of the system.

4.7 Raised-Cosine Rolloff

Sinc pulses use less bandwidth, but they are very susceptible to jitter. Conversely, Square pulses have a large bandwidth, but are very resistant to jitter. The decision seems hopeless, but the intrepid communications engineer will try to find a 3rd option. The 3rd option is called the "Raised-Cosine Rolloff" pulse, and has the best characteristics of each wave.

---

1   slew-rate-limiting resistor ^{http://massmind.org/techref/electips.htm#series_resistor}

Raised Cosine Rolloff (RCR) pulses are resistant to jitter, and at the same time have reasonably good bandwidth. Granted the bandwidth of an RCR pulse is wider than a sinc pulse, and the jitter resistance isn't as good as with a square wave, but this is a compromise, after all.

Pulse	Bandwidth	Jitter Resistance
Rectangular	rb/2	Excellent
Sinc		Bad
RCR		Good

4.8 Binary symmetric pulses

A system uses binary symmetric pulses if it transmits only 2 kinds of pulse, s_1 and s_0, and $s_1(t) = -s_0(t)$. In a symmetric case, our comparator circuits are very easy: If the correlator outputs a positive value, it's a binary 1. If it outputs a negative value, it's a binary 0.

4.9 Asymmetric Pulses

Asymmetric pulses are more difficult for many reasons:

1. The threshold where the comparator should test the value may not be zero
2. The correlation receiver needs to correlate *two different signals.*
3. The different pulses may have different power, and they may be different susceptible to noise.

4.10 Asymmetric Correlation Receiver

When asymmetric pulses are used, the receiver system must actually employ 2 coherent optimal receivers, each one tuned to receive one of the pulses. When two coherent optimal receivers are used, the receiver that is outputting the highest voltage value at the end of time T has received it's pulse.

… why not employ only 1 coherent optimal receiver, tuned to receiver neither s0 nor s1, but instead tuned to receive the difference (s1 - s0) ? …

Some PAM systems, such as Ethernet 100BASE-T2, use PAM-5 …

4.11 References

(PAM)

This page discusses the binary modulation schemes and "keying".

4.12 What is "Keying?"

Square waves, sinc waves, and raised-cosine rolloff waves are all well and good, but all of them have drawbacks. If we use an optimal, matched filter, we can eliminate the effect of jitter, so frankly, why would we consider square waves at all? Without jitter as a concern, it makes no sense to correct for jitter, or even take it into consideration. However, since the matched filter needs to look at individual symbols, the transmitted signal can't suffer from any intersymbol interference either. Therefore, we aren't using the sinc pulse.

Since the raised-cosine roll-off wave suffers from both these problems (in smaller amounts, however), we don't want to use that pulse either.

So the question is, what other types of pulses can we send?

It turns out that if we use some of the techniques we have developed using analog signal modulation, and implement a sinusoidal carrier wave, we can create a signal with no inter-symbol interference, very low bandwidth, and no worries about jitter. Just like analog modulation, there are 3 aspects of the carrier wave that we can change: the amplitude, the frequency, and the phase angle. Instead of "modulation", we call these techniques **keying** techniques, because they are operating on a binary-number basis.

There is one important point to note before continuing with this discussion: Binary signals **are not** periodic signals. Therefore, we cannot expect that a binary signal is going to have a discrete spectra like a periodic squarewave will have. For this reason, the spectral components of binary data are continuous spectra.

4.13 Amplitude Shift Keying

In an ASK system, we are changing the amplitude of the sine wave to transmit digitial data. We have the following cases:

- Binary 1: $A_1 \sin(f_c t)$
- Binary 0: $A_0 \sin(f_c t)$

The simplest modulation scheme sets A0 = 0V (turning the transmitter off), and setting A1 = +5V (any random non-zero number turns the transmitter on). This special case of ASK is called OOK (On-Off keying). Morse code uses OOK.

Another common special case of ASK sets A1 to some positive number, and A0 to the corresponding negative number A0 = -A1. We will mention this case again later.

In ASK, we have the following equation:

$$a(t)\sin(\omega t)$$

by the principal of duality, multiplication in the time domain becomes convolution in the frequency domain, and vice-versa. Therefore, our frequency spectrum will have the following equation:

$$A(j\omega) * \delta(t - \omega)$$

where the impulse function is the fourier-transform of the sinusoid, centered at the frequency of the wave. the value for A is going to be a sinc wave, with a width dependant on the bitrate. We remember from the Signals and Systems[2] book that convolution of a signal with an impulse is that signal centered where the impulse was centered. Therefore, we know now that the frequency domain shape of this curve is a sinc wave centered at the carrier frequency.

4.14 Frequency Shift Keying

In **Frequency Shift Keying** (FSK), we can logically assume that the parameter that we will be changing is the frequency of the sine wave. FSK is unique among the different keying methods in that data is never transmitted *at* the carrier frequency, but is instead transmitted at a certain offset from the carrier frequency. If we have a carrier frequency of f_c, and a frequency offset of Δf, we can transmit binary values as such:

- Binary 1: $A \sin((f_c + \Delta f)t)$
- Binary 0: $A \sin((f_c - \Delta f)t)$

Similar to ASK, we have FSK, which uses 2 different frequencies to transmit data. For now we will call them $\omega 1, \omega 2$. Using the same logic that we used above, the fourier representations of these waves will be (respectively):

$$A_1(j\omega) * \delta(t - \omega 1)$$

$$A_0(j\omega) * \delta(t - \omega 2)$$

With one sinc wave centered at the first frequency, and one sinc wave centered at the second frequency. Notice that A1 and A0 are the half-square waves associated with the 1s and the 0s, respectively. These will be described later.

4.14.1 Error Rate

The BER of coherent QPSK in the presence of gaussian and Rayleigh noise is as follows:

{| border=1

|- | Gaussian Noise || Rayleigh Fading |- | $\frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$ | $\frac{1}{2}\left(1 - \sqrt{\frac{\gamma_0}{2 + \gamma_0}}\right)$ |}

---

2   http://en.wikibooks.org/wiki/Signals%20and%20Systems

4.15 Phase Shift Keying

PSK systems are slightly different then ASK and FSK systems, and because of this difference, we can exploit an interesting little trick of trigonometry. PSK is when we vary the phase angle of the wave to transmit different bits. For instance:

- Binary 1: $A\sin(f_c t + \phi_1)$
- Binary 0: $A\sin(f_c t + \phi_0)$

If we evenly space them out around the unit-circle, we can give ourselves the following nice values:

- Binary 1: $A\sin(f_c t + 0)$
- Binary 0: $A\sin(f_c t + \pi)$

Now, according to trigonometry, we have the following identity:

$$\sin(f_c t + \pi) = -\sin(f_c t)$$

So in general, our equations for each signal (s) is given by:

- $s_1(t) = A\sin(f_c t)$
- $s_0(t) = -A\sin(f_c t)$

Which looks awfully like an ASK signal. Therefore, we can show that the spectrum of a PSK signal is the same as the spectrum of an ASK signal.

There are two commonally used forms of Phase Shift keying Modulation:

Binary Phase Shift Keying (BPSK)

Quadrature Phase Shift Keying (QPSK)

Binary Phase Shift keying is set out above.

4.15.1 QPSK

Quadrature Phase Shift Keying utilises the fact that a cosine wave is in quadrature to a sine wave, allowing 2 bits to be simultaneously represented.

- Binary 11: $A\sin(f_c t + 0) + \cos(f_c + \pi/2)$
- Binary 10: $A\sin(f_c t + 0) + \cos(f_c - \pi/2)$
- Binary 01: $A\sin(f_c t + \pi) + + \cos(f_c + \pi/2)$
- Binary 00: $A\sin(f_c t + \pi) + + \cos(f_c - \pi/2)$

QPSK has the advantage over BPSK of requiring half the transmission band width for the same data rate, and error probability.

4.15.2 Error Rate

The BER of coherent BPSK in the presence of gaussian and Rayleigh noise is as follows:

{| border=1

|- | Gaussian Noise || Rayleigh Fading |- | $\frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$ | $\frac{1}{2}\left(1 - \sqrt{\frac{\gamma_0}{1+\gamma_0}}\right)$ |}

4.16 Binary Transmitters

4.17 Binary Receivers

Now what if try to cram more information into a single bittime? If we take 2 bits at a time, and arrange them together, we can assign each set of 2 bits to a different symbol, and then we can transmit the different symbols.

4.18 Pronunciation

First off, "M-ary" is not pronounced like the female name "mary". If you say "mary" to a boardroom filled with engineers, they will laugh at you, and your boss will politely tell you later to start cleaning up your desk. "M-ary" is pronounced like "em airy".

4.19 Example: 4-ASK

Let us use the following scheme:

- "00" = +5V
- "01" = +1.66V
- "10" = -1.66V
- "11" = -5V

we can see now that we can transmit data *twice as fast* using this scheme, although we need to have a more complicated receiver, that can decide between 4 different pulses (instead of 2 different pulses, like we have been using).

4.20 Bits Per Symbol

All popular communication systems transmit an integer number of bits per symbol. We can relate the number of bits ("k") and the number of different symbols ("m") with the following equation:

$$m = 2^k$$

This causes the number of symbols to be a power of two.

With M-ary modulation techniques, the "symbols per second" rate can be much slower than the "bits per second" data rate.

4.21 QPSK

Quadrature phase shift keying (aka 4-PSK) is PSK modulation that has four points in the constellation.

4.22 CPFSK (MSK)

[MSK]minimum shift keying

4.23 DPSK

4.24 For further reading

- Wikipedia:Constellation_diagram[3]
- Wikipedia:Quadrature amplitude modulation[4]

The **quadrature amplitude modulation** (QAM) system of modulation is the most popular M-ary scheme[5].

4.25 Definition

Consider the case of a system with two carrier waves instead of a single carrier wave as we have considered with modulation schemes so far. One is a sine wave, and the other is a cosine wave of the same frequency. Since these two waves are orthogonal we can use them simultaneously in a single channel without losing the information of either. If both waves have the same frequency f we can write out the equation for a generic symbol, s:

$$s(t) = A_k \sin(ft) + B_k \cos(ft)$$

In this way, we can create multiple symbols by simply changing the values of A and B. This equation can be broken up into two parts:

- $A_k \sin(ft)$ Which is called the "in-phase" component of the equation.
- $B_k \cos(ft)$ Which is called the "quadrature" component of the equation.

An equation which is written as a sum of a sine plus a cosine is said to be in "quadrature form". If we combine the two components into a single waveform as such:

$$s(t) = \sqrt{A_k^2 + B_k^2} \cos(ft + \tan^{-1}(B_k/A_k))$$

---

3   http://en.wikipedia.org/wiki/Constellation_diagram
4   http://en.wikipedia.org/wiki/Quadrature%20amplitude%20modulation
5   Chapter 4.17 on page 137

This form is called the "Polar Form" of the equation.

> **ⓘ Information**
>
> 56K modems and Digital TV use QAM

4.26 Constellation Plots

If we make a graph with the X axis being the values for A, and the Y axis being the values for B, we get what is called a **Constellation Plot**. These plots are called constellation plots due to the similiarity in shape and layout with astronomical star charts. The A and B values for each symbol are plotted (the "stars") and various measurements between them are used to determine information from the system. On a constellation plot, we can see a number of rules:

1. The further apart the points are on the constellation plot, the less likely they are to be mistaken for each other in the presence of noise.
2. The closer the points are to the origin, the less power it takes to send.
3. The more points there are, the faster the data rate (bit rate) at a fixed symbol rate (more symbols)
4. The fewer points there are, the simpler and cheaper the hardware necessary to distinguish between them (fewer symbols, fewer thresholds in the receiver).

For these reasons there is no single "best" constellation plot, but it is up to the engineer to pick the points that are best for the system. In other words, trade offs need to be made between speed, performance, and cost of hardware. These tradeoffs can be made by placing the constellation points at different locations on the constellation plot.

4.27 Benefits of QAM

Increase the efficiency of transmission by utilising both amplitude and phase variations.

Reducing or eliminating intermodulation interference caused by a continuous carrier near the modulation sidebands.

4.28 For further reading

w:quadrature amplitude modulation[6] w:constellation diagram[7]

The **quadrature amplitude modulation** (QAM) system of modulation is the most popular M-ary scheme.

---

6   http://en.wikipedia.org/wiki/quadrature%20amplitude%20modulation
7   http://en.wikipedia.org/wiki/constellation%20diagram

4.29 Definition

Let us say that we have 2 carrier waves. One is a sine wave, and the other is a cosine wave. Since these two waves are orthogonal, we can use them simultaneously, without losing the information of either. If both waves have the same frequency, f, we can write out the equation for a generic symbol, s:

$$s(t) = A_k \sin(ft) + B_k \cos(ft)$$

In this way, we can create multiple symbols by simply changing the values of A and B. This equation can be broken up into two parts:

- $A_k \sin(ft)$ Which is called the "in-phase" component of the equation.
- $B_k \cos(ft)$ Which is called the "quadrature" component of the equation.

An equation which is written as a sum of a sine plus a cosine is said to be in "quadrature form". If we combine the two components into a single waveform as such:

$$s(t) = \sqrt{A_k^2 + B_k^2} \cos(ft + \tan^{-1}(B_k/A_k))$$

This form is called the "Polar Form" of the equation.

ⓘ Information

56K modems and Digital TV use QAM

4.30 Constellation Plots

If we make a graph with the X axis being the values for A, and the Y axis being the values for B, we get what is called a "Constellation Plot". If A and B have discrete values, then the constellation plot will show dots at points that correspond to values for A and B coordinates. It is called a constellation plot because the layout of the different points can look very similar to the layout of stars in the sky.

On a constellation plot, we can see a number of points:

1. The further apart the points are, the less likely they are to be mixed up
2. The closer the points are to the origin, the less power it takes to send.

1. The more points there are, the faster the data rate (bit rate) at a fixed symbol rate.
2. The fewer points there are, the simpler and cheaper the hardware necessary to distinguish between them.

For these two reasons, there is no single "best" constellation plot, but it is up to the engineer to pick the points that are best for the system. By placing the points manually, the engineer is able to make trade-offs between the power of the system, and the number of bits per symbol (and therefore the bitrate).

4.31 Benefits of QAM

4.32 For further reading

w:quadrature amplitude modulation[8] w:constellation diagram[9] (QAM)

1. REDIRECT Communication Systems/Noncoherent Receivers[10]

4.33 Line Codes

w:line code[11]

In addition to pulse shaping, there are a number of useful **line codes** that we can use to help reduce our errors or to have other positive effects on our signal.

Line coding consists of representing the digital signal to be transported, by an amplitude- and time-discrete signal, that is optimally tuned for the specific properties of the physical channel (and of the receiving equipment). The waveform pattern of voltage or current used to represent the 1s and 0s of a digital signal on a transmission link is called line encoding. The common types of line encoding are unipolar, polar, bipolar and Manchester encoding.

Line codes are used commonly in computer communication networks over short distances.

There are numerous ways digital information can be coded onto a transmission medium. Some of the more common include:

8   http://en.wikipedia.org/wiki/quadrature%20amplitude%20modulation
9   http://en.wikipedia.org/wiki/constellation%20diagram
10   http://en.wikibooks.org/wiki/Communication%20Systems%2FNoncoherent%20Receivers
11   http://en.wikipedia.org/wiki/line%20code

Signal	Comments
NRZ–L	Non-return to zero level. This is the standard positive logic signal format used in digital circuits. 1 forces a high level 0 forces a low level
NRZ–M	Non return to zero mark 1 forces a transition 0 does nothing
NRZ–S	Non return to zero space 1 does nothing 0 forces a transition
RZ	Return to zero 1 goes high for half the bit period 0 does nothing
Biphase–L	Manchester. Two consecutive bits of the same type force a transition at the beginning of a bit period. 1 forces a negative transition in the middle of the bit 0 forces a positive transition in the middle of the bit
Biphase–M	There is always a transition at the beginning of a bit period. 1 forces a transition in the middle of the bit 0 does nothing
Biphase–S	There is always a transition at the beginning of a bit period. 1 does nothing 0 forces a transition in the middle of the bit
Differential Manchester	There is always a transition in the middle of a bit period. 1 does nothing 0 forces a transition at the beginning of the bit
Bipolar	The positive and negative pulses alternate. 1 forces a positive or negative pulse for half the bit period 0 does nothing

Figure 99
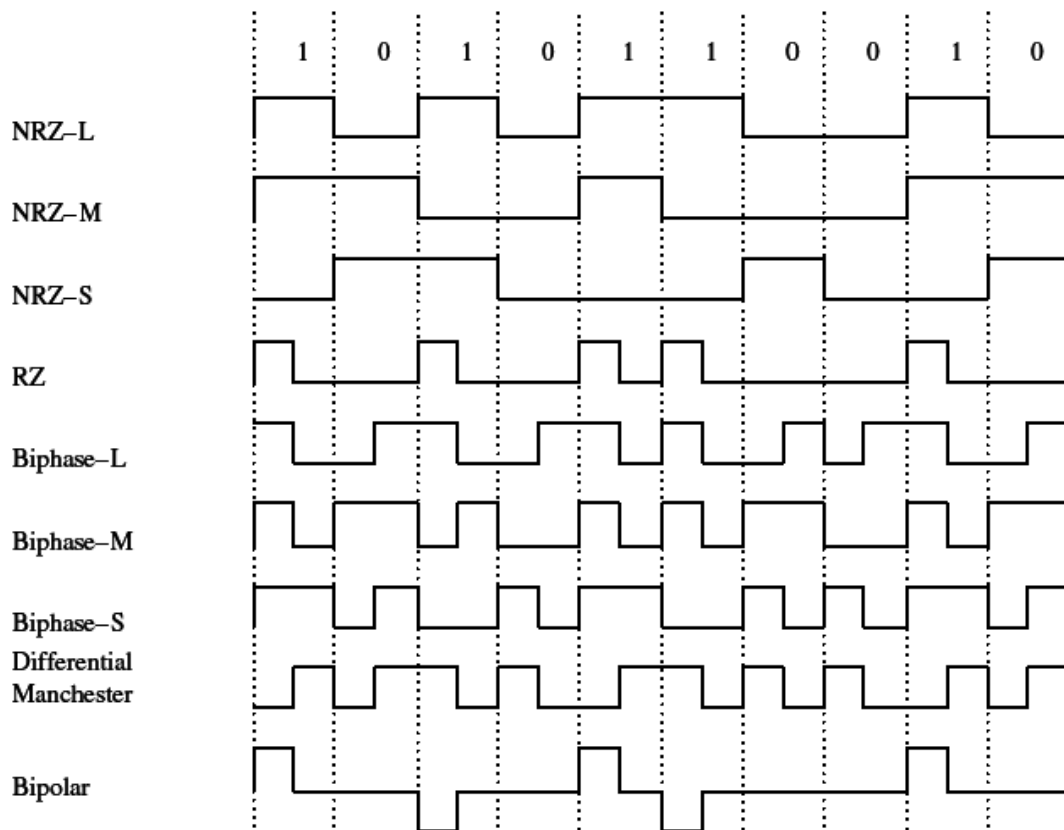
Figure 100

Each of the various line formats has a particular advantage and disadvantage. It is not possible to select one, which will meet all needs. The format may be selected to meet one or more of the following criteria:

- Minimize transmission hardware

- Facilitate synchronization

- Ease error detection and correction

- Minimize spectral content

- Eliminate a dc component

The Manchester code is quite popular. It is known as a self-clocking code because there is always a transition during the bit interval. Consequently, long strings of zeros or ones do not cause clocking problems.

4.34 Non-Return to Zero Codes (NRZ)

w:non-return-to-zero[12]

Non-Return to Zero (NRZ) codes are a bit awkwardly named, especially considering that the unipolar NRZ code does return to a zero value. In essence, an NRZ code is just a simple square wave, assigning one value to a binary 1, and another amplitude to a binary 0.

NRZ codes are more bandwidth efficient than bipolar ones. However, their spectral components go all the way down to 0 Hz. This prevents them from being used on transmission lines which are transformer coupled, or for some other reason cannot carry DC.
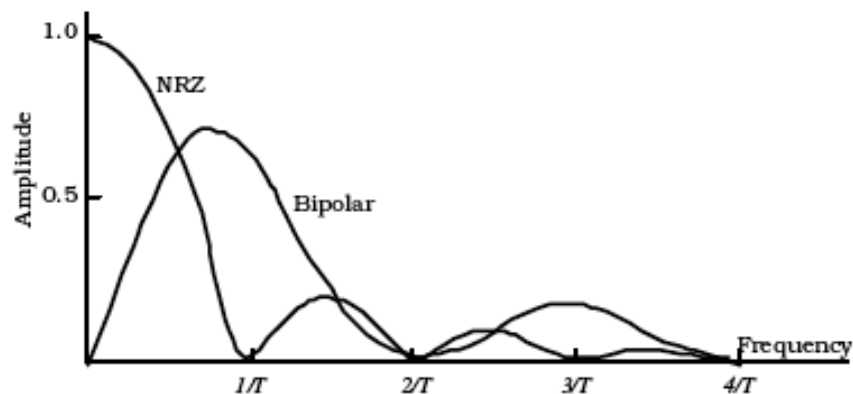
Figure 101

4.34.1 Unipolar NRZ

Unipolar NRZ is simply a square wave with +AV being a binary 1, and 0V being a binary 0. NRZ is convenient because computer circuits use unipolar NRZ internally, and it requires little effort to expand this system outside the computer. Unipolar NRZ has a DC term, but a relatively narrow bandwidth.

4.34.2 Bipolar NRZ

Bipolar NRZ operates using a bipolar voltage supply rail. Marks typically are represented using *negative* voltages (e.g., -9V), while spaces with *positive* voltages (e.g., +9V). For example, RS-232C/EIA-232 signaling relies on bipolar NRZ.

```
 -5V +5V -5V +5V -5V -5V +5V -5V -5V
  1   0   1   0   1   1   0   1   1
```

---

12   http://en.wikipedia.org/wiki/non-return-to-zero

Bipolar NRZ has similar bandwidth and DC balance issues as unipolar NRZ.

4.34.3 AMI

AMI (alternate mark inversion) is another example of a bipolar line code. Each successive mark is inverted and the average or DC level of the line is therefor zero.
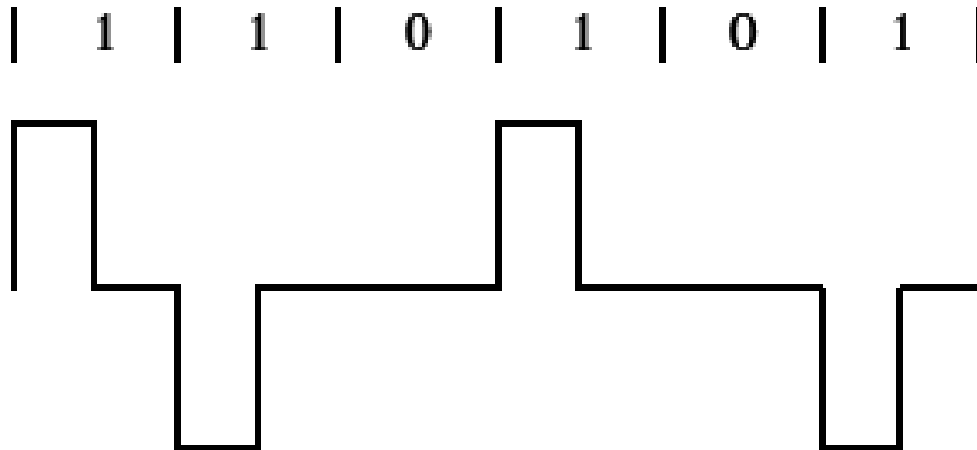
Figure 102

AMI is usually implemented as RZ pulses, but NRZ and NRZ-I variants exist as well.

One of the weaknesses of this approach is that long strings of zeros cause the receivers to lose lock. It is therefor necessary to impose other rules on the signal to prevent this. For example, combining NRZ-M with AMI yields MLT-3, the line coding system used with 100-base-T Ethernet.

4.34.4 CDI

The CDI(Conditioned Diphase Interface) bipolar line code is actually a slightly different form of the original FM line coding used in single-density disk drives and audio cassette tapes. Marks are encoded as alternate polarity full period pulses. Spaces are encoded by half a period at the negative voltage and half period at the positive voltage. This coding scheme has the advantage that it requires less logic to implement than HDB3.

| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |

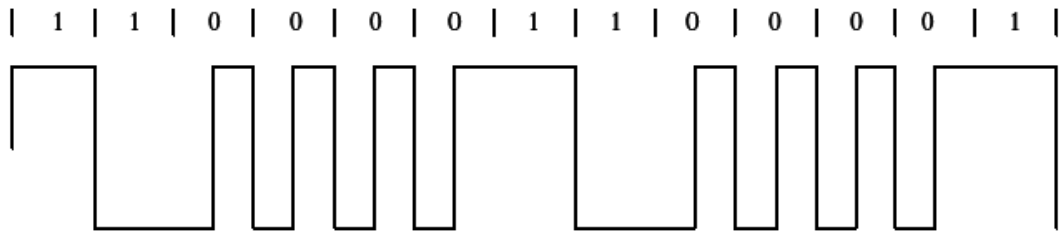**Figure 103**

4.35 Manchester

w:Manchester code[13]

Manchester codes were an invention that allows for timing information to be sent along with the data. In an NRZ code, if there is a long stream of ones or zeros, the receiver could conceivably suffer so much compound jitter that it would either lose or gain an entire bit time, and then be out of sync with the transmitter. This is because a long stream of 1s or 0s would not "change state" at all, but instead would simply stay at a single value. Manchester codes say that every single bittime will have a transition in the middle of the bit time, so that a receiver could find that transition, and "lock on" to the signal again, if it started to stray away from center. Because there are more transitions, however, manchester codes also require the highest bandwidth of all the line codes.

> **ⓘ Information**
>
> Ethernet LAN uses manchester codes

4.36 Differential Codes

Some communication channels (such as phase-modulated sine waves and differential twisted pairs) have the characteristic that transitions between the 2 symbols used can be easily distinguished, but when first starting to receive it is difficult to tell which of the 2 states it is in. For example, full-speed USB uses a twisted pair and transmits +3.3 V on one line and 0 V on the other line for "1", but 0 V on the one line and +3.3 V on the other line for "0". Because some cables have an extra half-twist in them, it is impossible for a device that was just plugged in to tell whether the symbol it is currently receiving is a "1" or a "0".

Differential codes still work, not even noticing when the 2 wires get switched.

---

13   http://en.wikipedia.org/wiki/Manchester%20code

Differential codes, in general, look exactly the same on a oscilloscope or spectrum analyzer as the non-differential code they are based on, and so use exactly the same bandwidth and have exactly the same bitrate.

Differential codes that work when the 2 wires get switched include:

- Differential Manchester encoding -- based on Manchester encoding
- Non-Return-to-Zero Inverted (NRZI) -- based on non-return-to-zero (NRZ)

(A few non-differential codes also work even when the 2 wires get switched -- such as bipolar encoding, and MLT-3 encoding).

4.36.1 Differential NRZ

4.36.2 Differential Manchester

Differential Manchester encoding, also called biphase mark code (BMC) or FM1, is a line code in which data and clock signals are combined to form a single 2-level self-synchronizing data stream. It is a differential encoding, using the presence or absence of transitions to indicate logical value. It has the following advantages over some other line codes: • A transition is guaranteed at least once every bit, allowing the receiving device to perform clock recovery. • Detecting transitions is often less error-prone than comparing against a threshold in a noisy environment. • Unlike with Manchester encoding, only the presence of a transition is important, not the polarity. Differential coding schemes will work exactly the same if the signal is inverted (wires swapped). (Other line codes with this property include NRZI, bipolar encoding, coded mark inversion, and MLT-3 encoding). • If the high and low signal levels have the same voltage with opposite polarity, coded signals have zero average DC voltage, thus reducing the necessary transmitting power and minimizing the amount of electromagnetic noise produced by the transmission line.

4.37 Comparison

Code	Bandwidth	Timing	DC value
Unipolar NRZ	Low bandwidth	No timing information	High DC component
Bipolar NRZ	Lower bandwidth	No timing information	No DC component
Differential NRZ	Lower bandwidth	No timing information	Little or no DC component
Manchester	High bandwidth	Good clock recovery	No DC component
Differential Manchester	Moderate bandwidth	Good clock recovery	No DC Component

4.38 further reading

- Wikipedia:differential coding[14]

---

14   http://en.wikipedia.org/wiki/differential%20coding

5 Analog Networks

1. REDIRECT Communication Networks/Circuit Switching Networks[1]

1. REDIRECT Communication Networks/Cable Television Network[2]

1. REDIRECT Communication Networks/Radio Communications[3]

1   http://en.wikibooks.org/wiki/Communication%20Networks%2FCircuit%20Switching%
    20Networks
2   http://en.wikibooks.org/wiki/Communication%20Networks%2FCable%20Television%20Network
3   http://en.wikibooks.org/wiki/Communication%20Networks%2FRadio%20Communications

6 Digital Networks

1. REDIRECT Communication Networks/Parallel vs Serial[1]

1. REDIRECT Communication Networks/OSI Reference Model[2]

1. REDIRECT Communication Networks/Channels[3]

1. REDIRECT Communication Networks/Internet[4]

1. REDIRECT Communication Networks/Ethernet[5] (IEEE 802.3)

1. REDIRECT Communication Networks/Wireless Internet[6]

---

1   http://en.wikibooks.org/wiki/Communication%20Networks%2FParallel%20vs%20Serial
2   http://en.wikibooks.org/wiki/Communication%20Networks%2FOSI%20Reference%20Model
3   http://en.wikibooks.org/wiki/Communication%20Networks%2FChannels
4   http://en.wikibooks.org/wiki/Communication%20Networks%2FInternet
5   http://en.wikibooks.org/wiki/Communication%20Networks%2FEthernet
6   http://en.wikibooks.org/wiki/Communication%20Networks%2FWireless%20Internet

7 Hybrid Networks

1. REDIRECT Communication Networks/Analog and Digital TV[1]

1. REDIRECT Communication Networks/Analog and Digital Telephony[2]

---

1   http://en.wikibooks.org/wiki/Communication%20Networks%2FAnalog%20and%20Digital%20TV
2   http://en.wikibooks.org/wiki/Communication%20Networks%2FAnalog%20and%20Digital%
    20Telephony

8 Advanced Internet

1. REDIRECT Communication Networks/Berkeley Socket API[1]

1. REDIRECT Communication Networks/IP Protocol and ICMP[2]

1. REDIRECT Communication Networks/TCP and UDP Protocols[3]

1. REDIRECT Communication Networks/HTTP Protocol[4]

w:File Transfer Protocol[5] **File Transfer Protocol** (**FTP**) is a standard network protocol used to exchange and manipulate files over a TCP/IP based network, such as the Internet[6]. FTP is built on a client-server architecture and utilizes separate control and data connections between the client and server applications. FTP is also often used as an application component to automatically transfer files for program internal functions. FTP can be used with user-based password authentication or with anonymous user access.

8.1 Purpose

Objectives of FTP, as outlined by its RFC, are:

- To promote sharing of files (computer programs and/or data).
- To encourage indirect or implicit use of remote computers.
- To shield a user from variations in file storage systems among different hosts.
- To transfer data reliably, and efficiently.
- To gain readability to the end user.

8.2 Connection methods

FTP runs over the Transmission Control Protocol (TCP). Usually FTP servers listen on the well-known port number 21 (IANA-reserved) for incoming connections from clients. A connection to this port from the FTP client forms the control stream on which commands are passed to the FTP server and responses are collected. FTP uses out-of-band control; it opens dedicated data connections on other port numbers. The parameters for the data

---

1   http://en.wikibooks.org/wiki/Communication%20Networks%2FBerkeley%20Socket%20API
2   http://en.wikibooks.org/wiki/Communication%20Networks%2FIP%20Protocol%20and%20ICMP
3   http://en.wikibooks.org/wiki/Communication%20Networks%2FTCP%20and%20UDP%20Protocols
4   http://en.wikibooks.org/wiki/Communication%20Networks%2FHTTP%20Protocol
5   http://en.wikipedia.org/wiki/File%20Transfer%20Protocol
6   http://en.wikibooks.org/wiki/Internet

streams depend on the specifically requested transport mode. Data connections usually use port number 20.

In *active mode*, the FTP client opens a dynamic port, sends the FTP server the dynamic port number on which it is listening over the control stream and waits for a connection from the FTP server. When the FTP server initiates the data connection to the FTP client it binds the source port to port 20 on the FTP server.

In order to use active mode, the client sends a PORT command, with the IP and port as argument. The format for the IP and port is "h1,h2,h3,h4,p1,p2". Each field is a decimal representation of 8 bits of the host IP, followed by the chosen data port. For example, a client with an IP of 192.168.0.1, listening on port 49154 for the data connection will send the command "PORT 192,168,0,1,192,2". The port fields should be interpreted as $p1{\times}256 + p2 =$ port, or, in this example, $192{\times}256 + 2 = 49154$.

In *passive mode*, the FTP server opens a dynamic port, sends the FTP client the server's IP address to connect to and the port on which it is listening (a 16-bit value broken into a high and low byte, as explained above) over the control stream and waits for a connection from the FTP client. In this case, the FTP client binds the source port of the connection to a dynamic port.

To use passive mode, the client sends the *PASV* command to which the server would reply with something similar to "227 Entering Passive Mode (127,0,0,1,192,52)". The syntax of the IP address and port are the same as for the argument to the PORT command.

In *extended passive mode*, the FTP server operates exactly the same as passive mode, however it only transmits the port number (not broken into high and low bytes) and the client is to assume that it connects to the same IP address that was originally connected to.

While data is being transferred via the data stream, the control stream sits idle. This can cause problems with large data transfers through firewalls which time out sessions after lengthy periods of idleness. While the file may well be successfully transferred, the control session can be disconnected by the firewall, causing an error to be generated.

The FTP protocol supports resuming of interrupted downloads using the REST command. The client passes the number of bytes it has already received as argument to the REST command and restarts the transfer. In some commandline clients for example, there is an often-ignored but valuable command, "reget" (meaning "get again"), that will cause an interrupted "get" command to be continued, hopefully to completion, after a communications interruption.

Resuming uploads is not as easy. Although the FTP protocol supports the APPE command to append data to a file on the server, the client does not know the exact position at which a transfer got interrupted. It has to obtain the size of the file some other way, for example over a directory listing or using the SIZE command.

In ASCII mode (see below), resuming transfers can be troublesome if client and server use different *end of line* characters.

8.3 Data format

While transferring data over the network, several data representations can be used. The two most common transfer modes are:

- ASCII mode
- Binary mode: In "Binary mode", the sending machine sends each file byte for byte and as such the recipient stores the bytestream as it receives it. (The FTP standard calls this "IMAGE" or "I" mode)

In ASCII mode, any form of data that is not plain text will be corrupted. When a file is sent using an ASCII-type transfer, the individual letters, numbers, and characters are sent using their ASCII character codes. The receiving machine saves these in a text file in the appropriate format (for example, a Unix machine saves it in a Unix format, a Windows machine saves it in a Windows format). Hence if an ASCII transfer is used it can be assumed plain text is sent, which is stored by the receiving computer in its own format. Translating between text formats might entail substituting the *end of line* and *end of file* characters used on the source platform with those on the destination platform, e.g. a Windows machine receiving a file from a Unix machine will replace the line feeds with carriage return-line feed pairs. It might also involve translating characters; for example, when transferring from an IBM mainframe to a system using ASCII, EBCDIC characters used on the mainframe will be translated to their ASCII equivalents, and when transferring from the system using ASCII to the mainframe, ASCII characters will be translated to their EBCDIC equivalents.

By default, most FTP clients use ASCII mode. Some clients try to determine the required transfer-mode by inspecting the file's name or contents, or by determining whether the server is running an operating system with the same text file format.

The FTP specifications also list the following transfer modes:

- EBCDIC mode - this transfers bytes, except they are encoded in EBCDIC rather than ASCII. Thus, for example, the ASCII mode server
- Local mode - this is designed for use with systems that are word-oriented rather than byte-oriented. For example mode "L 36" can be used to transfer binary data between two 36-bit machines. In L mode, the words are packed into bytes rather than being padded. Some FTP servers accept "L 8" as being equivalent to "I".

In practice, these additional transfer modes are rarely used. They are however still used by some legacy mainframe systems.

The text (ASCII/EBCDIC) modes can also be qualified with the type of carriage control used (e.g. TELNET NVT carriage control, ASA carriage control), although that is rarely used nowadays.

Note that the terminology "mode" is technically incorrect, although commonly used by FTP clients. "MODE" in RFC 959 refers to the format of the protocol data stream (STREAM, BLOCK or COMPRESSED), as opposed to the format of the underlying file. What is commonly called "mode" is actually the "TYPE", which specifies the format of the file rather than the data stream. FTP also supports specification of the file structure ("STRU"), which can be either FILE (stream-oriented files), RECORD (record-oriented files) or PAGE

(special type designed for use with TENEX). PAGE STRU is not really useful for non-TENEX systems, and RFC 1123 section 4.1.2.3 recommends that it not be implemented.

8.4 FTP return codes

FTP server return codes indicate their status by the digits within them. A brief explanation of various digits' meanings are given below:

- 1xx: Positive Preliminary reply. The action requested is being initiated but there will be another reply before it begins.
- 2xx: Positive Completion reply. The action requested has been completed. The client may now issue a new command.
- 3xx: Positive Intermediate reply. The command was successful, but a further command is required before the server can act upon the request.
- 4xx: Transient Negative Completion reply. The command was not successful, but the client is free to try the command again as the failure is only temporary.
- 5xx: Permanent Negative Completion reply. The command was not successful and the client should not attempt to repeat it again.
- x0x: The failure was due to a syntax error.
- x1x: This response is a reply to a request for information.
- x2x: This response is a reply relating to connection information.
- x3x: This response is a reply relating to accounting and authorization.
- x4x: Unspecified as yet
- x5x: These responses indicate the status of the Server file system vis-a-vis the requested transfer or other file system action.

8.5 Anonymous FTP

A host that provides an FTP service may additionally provide anonymous FTP access. Users typically login to the service with an 'anonymous' account when prompted for user name. Although users are commonly asked to send their email address in lieu of a password, little to no verification is actually performed on the supplied data.

As modern FTP clients typically hide the anonymous login process from the user, the ftp client will supply dummy data as the password (since the user's email address may not be known to the application). For example, the following ftp user agents specify the listed passwords for anonymous logins:

- Mozilla Firefox[7] (3.0.7) — `mozilla@example.com`
- KDE[8] Konqueror (3.5) — `anonymous@`
- wget (1.10.2) — `-wget@`
- lftp (3.4.4) — `lftp@`

---

7   `http://en.wikibooks.org/wiki/Mozilla%20Firefox`
8   `http://en.wikibooks.org/wiki/KDE`

8.6 Commands

Enter **ftp /?** in Windows, or **ftp --help** in Unix, to get the command parameters.

Once connected to a server, type **help** to display the different possible commands.

To manipulate the files with the mouth, download a good FTP client which will do the interface (for example this Filezilla doesn't need any installation[9]).

1. REDIRECT Communication Networks/Routing[10]

---

9   http://portableapps.com/apps/internet/filezilla_portable
10   http://en.wikibooks.org/wiki/Communication%20Networks%2FRouting

9 Coding and Security

1. REDIRECT Data Coding Theory/Transmission Codes[1]

1. REDIRECT Data Coding Theory/Spectrum Spreading[2]

1. REDIRECT Data Coding Theory/Data Compression[3]

1. REDIRECT Data Coding Theory/Hamming Codes[4]

---

1   http://en.wikibooks.org/wiki/Data%20Coding%20Theory%2FTransmission%20Codes
2   http://en.wikibooks.org/wiki/Data%20Coding%20Theory%2FSpectrum%20Spreading
3   http://en.wikibooks.org/wiki/Data%20Coding%20Theory%2FData%20Compression
4   http://en.wikibooks.org/wiki/Data%20Coding%20Theory%2FHamming%20Codes

10 Appendices

The 802 portion of the IEEE is responsible for publishing standards on a number of different protocols. Any chapter with a note such as (IEEE 802.3) is referring to the group below that has published the currently accepted standard for that topic.

Working Group	Task	Status
802.1	Higher Layer LAN Protocols Working Group	Active
802.2	Logical Link Control Working Group	Inactive
802.3	Ethernet Working Group	Active
802.4	Token Bus Working Group	Disbanded
802.5	Token Ring Working Group	Inactive
802.6	Metropolitan Area Network Working Group	Disbanded
802.7	Broadband TAG	Disbanded
802.8	Fiber Optic TAG	Disbanded
802.9	Isochronous LAN Working Group	Disbanded
802.10	Security Working Group	Disbanded
802.11	Wireless LAN Working Group	Active
802.12	Demand Priority Working Group	Inactive
802.14	Cable Modem Working Group	Disbanded
802.15	Wireless Personal Area Network (WPAN) Working Group	Active
802.16	Broadband Wireless Access Working Group	Active
802.17	Resilient Packet Ring Working Group	Active
802.18	Radio Regulatory TAG	Active
802.19	Coexistence TAG	Active
802.20	Mobile Broadband Wireless Access (MBWA) Working Group	Active
802.21	Media Independent Handoff Working Group	Active
802.22	Wireless Regional Area Networks	Active

11 Further reading

11.1 Wikibooks

- Communication Theory[1]
- Voice over IP[2]
- Internet Technologies[3]
- Networking:Ports and Protocols[4]
- Internet Engineering[5]
- Serial Communications Bookshelf[6]
- Analog and Digital Conversion[7]
- Wireless Mesh Networks[8]

11.2 Wikipedia Articles

- Wikipedia:History Of Internet[9]

11.3 Books

- Garcia and Widjaja, "Communication Networks: Fundamental Concepts and Key Architectures", Second Edition, McGraw Hill, 2004. ISBN 007246352X
- Stern and Mahmoud, "Communication Systems: Analysis and Design", Prentice Hall, 2004. ISBN 0130402680
- Silage, Dennis, "Digital Communication Systems Using SystemVue", DaVinci Engineering Press, 2006. ISBN 1584508507
- Haykin and Moher, "Modern Wireless Communications", Prentice Hall, 2005. ISBN 0130224723
- Gibson, Jerry D. "Principles of Digital and Analog Communications, Second Edition", Macmillan Publishing Company, 1989. ISBN 0023418605

---

1   http://en.wikibooks.org/wiki/Communication%20Theory
2   http://en.wikibooks.org/wiki/Voice%20over%20IP
3   http://en.wikibooks.org/wiki/Internet%20Technologies
4   http://en.wikibooks.org/wiki/Networking%3APorts%20and%20Protocols
5   http://en.wikibooks.org/wiki/Internet%20Engineering
6   http://en.wikibooks.org/wiki/Information%20technology%20bookshelf%23Serial%
     20Communications
7   http://en.wikibooks.org/wiki/Analog%20and%20Digital%20Conversion
8   http://en.wikibooks.org/wiki/Wireless%20Mesh%20Networks
9   http://en.wikipedia.org/wiki/History%20Of%20Internet

12 Contributors

Edits	User
15	Adrignola[1]
1	Aleksev[2]
78	Annieyc 2002[3]
28	Avicennasis[4]
3	Az1568[5]
1	Chazz[6]
1	CodeCat[7]
1	Cspurrier[8]
2	Darklama[9]
22	DavidCary[10]
1	Dirk Hünniger[11]
1	Everlong[12]
1	Frigotoni[13]
3	Ftiercel[14]
4	Gautamraj[15]
1	Glaisher[16]
2	Hagindaz[17]
4	JackPotte[18]
1	Jakec[19]
2	Jguk[20]
10	Jomegat[21]

1   http://en.wikibooks.org/wiki/User:Adrignola
2   http://en.wikibooks.org/wiki/User:Aleksev
3   http://en.wikibooks.org/wiki/User:Annieyc_2002
4   http://en.wikibooks.org/wiki/User:Avicennasis
5   http://en.wikibooks.org/wiki/User:Az1568
6   http://en.wikibooks.org/wiki/User:Chazz
7   http://en.wikibooks.org/wiki/User:CodeCat
8   http://en.wikibooks.org/wiki/User:Cspurrier
9   http://en.wikibooks.org/wiki/User:Darklama
10   http://en.wikibooks.org/wiki/User:DavidCary
11   http://en.wikibooks.org/wiki/User:Dirk_H%25C3%25BCnniger
12   http://en.wikibooks.org/wiki/User:Everlong
13   http://en.wikibooks.org/wiki/User:Frigotoni
14   http://en.wikibooks.org/wiki/User:Ftiercel
15   http://en.wikibooks.org/wiki/User:Gautamraj
16   http://en.wikibooks.org/wiki/User:Glaisher
17   http://en.wikibooks.org/wiki/User:Hagindaz
18   http://en.wikibooks.org/wiki/User:JackPotte
19   http://en.wikibooks.org/wiki/User:Jakec
20   http://en.wikibooks.org/wiki/User:Jguk
21   http://en.wikibooks.org/wiki/User:Jomegat

399	Jugandi[22]
1	Kar.ma[23]
2	L0g1c[24]
1	LlamaAl[25]
3	Mabdul[26]
2	Mattb112885[27]
1	Mike.lifeguard[28]
64	Panic2k4[29]
1	Prav001[30]
12	Priyankgandhi[31]
32	QuiteUnusual[32]
26	Recent Runes[33]
2	Reece[34]
1	Savh[35]
2	Scud43[36]
2	Sigma 7[37]
1	Sjlegg[38]
7	Tannersf[39]
1	Tegel[40]
1	Thenub314[41]
1	Upul[42]
18	Webaware[43]
264	Whiteknight[44]
26	Wknight8111[45]
4	Xania[46]

22 http://en.wikibooks.org/wiki/User:Jugandi
23 http://en.wikibooks.org/wiki/User:Kar.ma
24 http://en.wikibooks.org/wiki/User:L0g1c
25 http://en.wikibooks.org/wiki/User:LlamaAl
26 http://en.wikibooks.org/wiki/User:Mabdul
27 http://en.wikibooks.org/wiki/User:Mattb112885
28 http://en.wikibooks.org/wiki/User:Mike.lifeguard
29 http://en.wikibooks.org/wiki/User:Panic2k4
30 http://en.wikibooks.org/wiki/User:Prav001
31 http://en.wikibooks.org/wiki/User:Priyankgandhi
32 http://en.wikibooks.org/wiki/User:QuiteUnusual
33 http://en.wikibooks.org/wiki/User:Recent_Runes
34 http://en.wikibooks.org/wiki/User:Reece
35 http://en.wikibooks.org/wiki/User:Savh
36 http://en.wikibooks.org/wiki/User:Scud43
37 http://en.wikibooks.org/wiki/User:Sigma_7
38 http://en.wikibooks.org/wiki/User:Sjlegg
39 http://en.wikibooks.org/wiki/User:Tannersf
40 http://en.wikibooks.org/wiki/User:Tegel
41 http://en.wikibooks.org/wiki/User:Thenub314
42 http://en.wikibooks.org/wiki/User:Upul
43 http://en.wikibooks.org/wiki/User:Webaware
44 http://en.wikibooks.org/wiki/User:Whiteknight
45 http://en.wikibooks.org/wiki/User:Wknight8111
46 http://en.wikibooks.org/wiki/User:Xania

1 Xerol[47]
1 YMS[48]
1 Érico Júnior Wouters[49]

47 http://en.wikibooks.org/wiki/User:Xerol
48 http://en.wikibooks.org/wiki/User:YMS
49 http://en.wikibooks.org/wiki/User:%25C3%2589rico_J%25C3%25BAnior_Wouters

List of Figures

- GFDL: Gnu Free Documentation License. `http://www.gnu.org/licenses/fdl.html`

- cc-by-sa-3.0: Creative Commons Attribution ShareAlike 3.0 License. `http://creativecommons.org/licenses/by-sa/3.0/`

- cc-by-sa-2.5: Creative Commons Attribution ShareAlike 2.5 License. `http://creativecommons.org/licenses/by-sa/2.5/`

- cc-by-sa-2.0: Creative Commons Attribution ShareAlike 2.0 License. `http://creativecommons.org/licenses/by-sa/2.0/`

- cc-by-sa-1.0: Creative Commons Attribution ShareAlike 1.0 License. `http://creativecommons.org/licenses/by-sa/1.0/`

- cc-by-2.0: Creative Commons Attribution 2.0 License. `http://creativecommons.org/licenses/by/2.0/`

- cc-by-2.0: Creative Commons Attribution 2.0 License. `http://creativecommons.org/licenses/by/2.0/deed.en`

- cc-by-2.5: Creative Commons Attribution 2.5 License. `http://creativecommons.org/licenses/by/2.5/deed.en`

- cc-by-3.0: Creative Commons Attribution 3.0 License. `http://creativecommons.org/licenses/by/3.0/deed.en`

- GPL: GNU General Public License. `http://www.gnu.org/licenses/gpl-2.0.txt`

- LGPL: GNU Lesser General Public License. `http://www.gnu.org/licenses/lgpl.html`

- PD: This image is in the public domain.

- ATTR: The copyright holder of this file allows anyone to use it for any purpose, provided that the copyright holder is properly attributed. Redistribution, derivative work, commercial use, and all other use is permitted.

- EURO: This is the common (reverse) face of a euro coin. The copyright on the design of the common face of the euro coins belongs to the European Commission. Authorised is reproduction in a format without relief (drawings, paintings, films) provided they are not detrimental to the image of the euro.

- LFK: Lizenz Freie Kunst. `http://artlibre.org/licence/lal/de`

- CFR: Copyright free use.

- EPL: Eclipse Public License. `http://www.eclipse.org/org/documents/epl-v10.php`

Copies of the GPL, the LGPL as well as a GFDL are included in chapter Licenses[50]. Please note that images in the public domain do not require attribution. You may click on the image numbers in the following table to open the webpage of the images in your webbrower.

---

50   Chapter 13 on page 177

1	Adrignola, Jugandi	
2	Adrignola, Jugandi	
3	Adrignola, Jugandi	
4	Adrignola, Jugandi	
5	Adrignola, Jugandi	
6	Adrignola, Jugandi	
7	Adrignola, Jugandi	
8	Adrignola, Jugandi	
9	Adrignola, Jugandi	
10	Adrignola, Jugandi	
11	Adrignola, Jugandi	
12	Adrignola, Jugandi	
13	Adrignola, Jugandi	
14	Adrignola, Jugandi	
15	Adrignola, Jugandi	
16	Adrignola, Jugandi	
17	Adrignola, Jugandi	
18	Adrignola, Jugandi	
19	Adrignola, Jugandi	
20	Adrignola, Jugandi	
21	Adrignola, Jugandi	
22	Adrignola, Jugandi	
23	Adrignola, Jugandi	
24	Adrignola, Jugandi	
25	Adrignola, Jugandi	
26	Original uploader was Cuddlyable3[51] at en.wikipedia[52] Later versions were uploaded by Noclip[53] at en.wikipedia[54].	GFDL
27	Adrignola, Darklama, Jugandi, Red4tribe	
28	Original uploader was Krishnavedala[55] at en.wikipedia[56]	GFDL
29	Adrignola, Jugandi	
30	Adrignola, Jugandi	
31	Adrignola, Jugandi	
32	Adrignola, Jugandi, Mabdul	
33	Adrignola, Jugandi	
34	Adrignola, Jugandi	
35	Adrignola, Jugandi	
36	Adrignola, Jugandi	
37	Adrignola, Jugandi	
38	Adrignola, Jugandi	
39	Adrignola, Jugandi	
40	Adrignola, Jugandi	
41	Adrignola, Jugandi	
42	Adrignola, Jugandi	
43	Adrignola, Jugandi	

---

51 `http:///en.wikipedia.org/wiki/User:Cuddlyable3`

52 `http://en.wikipedia.org`

53 `http:///en.wikipedia.org/wiki/User:Noclip`

54 `http://en.wikipedia.org`

55 `http:///en.wikipedia.org/wiki/User:Krishnavedala`

56 `http://en.wikipedia.org`

44	Adrignola, Jugandi, Mike.lifeguard	
45	Adrignola, Jugandi	
46	Adrignola, Jugandi	
47	Adrignola, Jugandi	
48	Adrignola, Jugandi	
49	Adrignola, Jugandi	
50	Adrignola, Jugandi	
51	Adrignola, Jugandi	
52	Adrignola, Jugandi	
53	Adrignola, Jugandi	
54	Adrignola, Jugandi	
55	Adrignola, Jugandi	
56	Adrignola, Jugandi	
57	Adrignola, Jugandi	
58	Adrignola, Jugandi	
59	en:User:Jugandi[57]	
60	Adrignola, Jugandi	
61	Adrignola, Jugandi	
62	Adrignola, Jugandi, Mike.lifeguard	
63	Original uploader was Cuddlyable3[58] at en.wikipedia[59] Later versions were uploaded by Noclip[60] at en.wikipedia[61].	GFDL
64	Adrignola, Jugandi	
65	Adrignola, Jugandi, Red4tribe	
66	Adrignola, Jugandi, Red4tribe	
67	Adrignola, Jugandi	
68	Adrignola, Jugandi, Red4tribe	
69	Adrignola, Jugandi, Red4tribe	
70	Adrignola, Jugandi	
71	Adrignola, Jugandi	
72	Adrignola, Jugandi	
73	Adrignola, Jugandi	
74	Adrignola, Jugandi	
75	Adrignola, Jugandi	
76	Adrignola, Jugandi, Original uploader was Paweł Łucek[62] at pl.wikipedia[63]	GFDL
77	Adrignola, Jugandi	
78	Adrignola, Jugandi	
79	Adrignola, Jugandi	
80	Adrignola, Jugandi	
81	Adrignola, Jugandi	
82	Adrignola, Jugandi, Mike.lifeguard	
83	Adrignola, Darklama, Jugandi	

---

57 http:///en.wikibooks.org/wiki/en:User:Jugandi

58 http:///en.wikipedia.org/wiki/User:Cuddlyable3

59 http://en.wikipedia.org

60 http:///en.wikipedia.org/wiki/User:Noclip

61 http://en.wikipedia.org

62 http:///pl.wikipedia.org/wiki/User:Pawe%C5%82_%C5%81ucek

63 http://pl.wikipedia.org

84	unknown. Original uploader was Jugandi[64] at en.wikibooks[65]	PD
85	unknown. Original uploader was Jugandi[66] at en.wikibooks[67]	PD
86	Adrignola, Jugandi	
87	Adrignola, Jugandi	
88	Adrignola, Jugandi	
89	Adrignola, Jugandi	
90	Jugandi	
91	Jugandi	
92	Adrignola, Jugandi	
93	Adrignola, Jugandi	
94	Adrignola, Jugandi	
95	Adrignola, Jugandi	
96	Adrignola, Jugandi, Fayaz29[68]	
97	Adrignola, Jugandi	
98	Adrignola, Jugandi	
99	Adrignola, Jugandi	
100	Adrignola, Jugandi	
101	Adrignola, Jugandi	
102	Adrignola, Jugandi	
103	Adrignola, Jugandi	

---

64 `http:///en.wikibooks.org/wiki/en:User:Jugandi`
65 `http://en.wikibooks.org`
66 `http:///en.wikibooks.org/wiki/en:User:Jugandi`
67 `http://en.wikibooks.org`
68 `http:///w/index.php?title=User:Fayaz29&action=edit&redlink=1`

13 Licenses

13.1 GNU GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed. Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program--to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow. TERMS AND CONDITIONS 0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrightable work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion. 1. Source Code.

The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

A "Standard Interface" means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The "System Libraries" of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A "Major Component", in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The "Corresponding Source" for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work. 2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary. 3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures. 4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee. 5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

* a) The work must carry prominent notices stating that you modified it, and giving a relevant date. * b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices". * c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it. * d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate. 6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

* a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange. * b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge. * c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b. * d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a

different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements. * e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying. 7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

* a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or * b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or * c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or * d) Limiting the use for publicity purposes of names of licensors or authors of the material; or * e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or * f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way. 8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates

your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10. 9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so. 10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An "entity transaction" is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it. 11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. "Knowingly relying" means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is "discriminatory" if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law. 12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy

both those terms and this License would be to refrain entirely from conveying the Program. 13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such. 14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version. 15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION. 16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. 17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

END OF TERMS AND CONDITIONS How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively state the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.> Copyright (C) <year> <name of author>

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Also add information on how to contact you by electronic and paper mail.

If the program does terminal interaction, make it output a short notice like this when it starts in an interactive mode:

<program> Copyright (C) <year> <name of author> This program comes with ABSOLUTELY NO WARRANTY; for details type `show w'. This is free software, and you are welcome to redistribute it under certain conditions; type `show c' for details.

The hypothetical commands `show w' and `show c' should show the appropriate parts of the General Public License. Of course, your program's commands might be different; for a GUI interface, you would use an "about box".

You should also get your employer (if you work as a programmer) or school, if any, to sign a "copyright disclaimer" for the program, if necessary. For more information on this, and how to apply and follow the GNU GPL, see <http://www.gnu.org/licenses/>.

The GNU General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Lesser General Public License instead of this License. But first, please read <http://www.gnu.org/philosophy/why-not-lgpl.html>.

13.2 GNU Free Documentation License

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc. <http://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed. 0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference. 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "publisher" means any person or entity that distributes copies of the Document to the public.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses

following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License. 2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies. 3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document. 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

* A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission. * B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement. * C. State on the Title page the name of the publisher of the Modified Version, as the publisher. * D. Preserve all the copyright notices of the Document. * E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices. * F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below. * G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice. * H. Include an unaltered copy of this License. * I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence. * J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission. * K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein. * L. Preserve all

the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles. * M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version. * N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section. * O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version. 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements". 6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document. 7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate. 8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its

Title (section 1) will typically require changing the actual title. 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it. 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document. 11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing. ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright (C) YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

13.3 GNU Lesser General Public License