Modulation Methods – Frequency Modulation

William Sheets K2MQJ   **Rudolf F. Graf KA2CWL**

The use of frequency modulation (called FM) is another method of adding intelligence to a carrier signal. While simple on-off (as used in Morse Code) keying and amplitude modulation were the first methods used to modulate a carrier, the use of FM had to await the availability of stable RF sources with a narrow frequency spectrum. The vacuum tube oscillator was the first RF source that could be frequency modulated. FM broadcasting started out before World War 2 as a means of obtaining "staticless" radio transmissions and better audio fidelity. Early FM broadcast stations used the 42-50 MHz band for their FM transmissions. At that time, 42-50 MHz was considered ultra high frequency. The elimination of amplitude modulation enabled the receiver to utilize high gain, so as to deliberately overload the amplifier stages in the IF circuitry, which had the effect of limiting (clipping) the received signal amplitude. This effectively suppressed AM noise spikes, and FM detector circuits could be also designed to reject AM signals to some degree. The fact that much man made noise is AM in nature, and that natural atmospheric static levels are lower at 42 MHz than at AM broadcast frequencies allowed very quiet reception with little if any static. Also, it was feasible to transmit a wider band of audio frequencies as larger amounts of spectrum space allowed wider signal bandwidths. Additionally, the bandwidth limitations of modulation transformers used in in AM transmitters at the time could be circumvented. A number of FM receivers were made for 42-50 MHz, and some surviving sets can still be found in forgotten attics and in antique radio collections. However, after WW2, the FM band was relocated to its present location at 88-108 MHz. However, there is still some FM broadcasting around 68 MHz in the former Soviet Union, and additionally, 76-88 MHz is also used in the Far East. And narrow band (±5 kHz deviation) FM is also widely used for TV audio and for 2 way VHF and UHF communications.

An FM (abbreviation for Frequency Modulation) signal is simply any RF signal that is modulated by changing the frequency of the signal in accordance with the modulation. Since frequency is measured in cycles per second (Hz). and cycles per second can be thought of as the number of times a sine wave goes through a complete cycle. (0 to 360 degrees) in one second, one can consider frequency as the rate of change of phase (degrees) of a sinusoidal waveform per unit time (second), or mathematically speaking, $d\theta/dt$, or the rate of change of phase angle with respect to time. Similarly, the reverse relationship says that total phase change in a time period equals the frequency multiplied by time. Frequency modulation is therefore related to phase modulation. Phase modulation is that modulation in which the relative phase of a carrier wave, with respect to a reference, is varied in accordance with a modulating signal. But in order to change the phase of a signal, there has to be an instantaneous change in frequency. As an example, if we want to change the phase of a carrier from 0 degrees to 180 degrees, we must "reverse" one cycle (See Figure 1). This would involve raising or lowering the frequency for a very short instant so the relationship of the phase with respect to the original phase could change.

If we want to vary the frequency of a signal, the instantaneous phase of the signal

must continuously change. For example, suppose we want to sweep from 100 MHz to 101 MHz in one second. This implies a frequency change of 1 MHz per second, (or 1 Hz per microsecond). In one microsecond, the 100 MHz carrier will go through 100 cycles. In this brief period, 1 cycle must be added so that after one second, one million cycles per second has been added. This can be done by advancing the phase of the carrier 1/100 cycle, or 3.6 degrees, every cycle, so that after 1 microsecond, 101 cycles will have been gone through, and the new <u>instantaneous</u> frequency will be 100,000,001 cycles (100 MHz + 1 Hz) per second. Notice that a <u>constant rate of change of phase has been introduced to produce a "sliding" frequency</u>. If this rate of change of phase is kept up indefinitely, the frequency would continue to sweep upward by 1 MHz per second, so that at t=l second the carrier frequency would be 101 MHz, at t=2 seconds, 102 MHz, etc. If we retarded the phase by the same 3.6 degrees, we would have produced a <u>decrease</u> in frequency of 1 MHz per second (in this example). If you did not follow this, or cannot understand the concept, simply remember that a change in frequency implies a phase change, and vice—versa.

The deviation of an FM signal is the component of change in carrier frequency that the modulating signal produces. Note that, in general, the <u>frequency</u> of the modulating signal is <u>NOT</u> the same as nor necessarily related to the <u>deviation</u>. In the USA, FM broadcast stations by law are permitted +75 kHz deviation, which is defined by the standards for FM broadcasting as 100% modulation. Both a 20 Hz audio signal and a 20 kHz audio signal can produce 75 kHz deviation. The amplitude of the modulating signal (program audio) produces the deviation. If one volt of audio produces +75 kHz deviation, then one tenth of a volt would produce +7.5 kHz deviation. Deviation and modulation frequency are <u>independent</u> variables.

Note that much wider deviations are possible. An AM transmitter has definite physical limits on modulation (Overmodulation will cut off the carrier and clip on positive peaks as the modulator is power limited). An FM transmitter has no such kind of physical limits. The limit of modulation is set as that which results in excessive modulation nonlinearity (Instantaneous frequency vs modulating voltage) in the modulation, or in how much frequency deviation the FM receiver will accept before it distorts the signal. This is related to receiver bandwidth. There is no theoreticall reason why our FM stations could not have used 100 kHz, or even 1 MHz deviation, other than needing more spectrum space. However, greater receiver sensitivity is possible with narrower bandwidths. Wider deviations are commonly used at microwave frequencies, for data and video transmission. Satellite television transmission uses 10 to 15 MHz deviation for the video signal, with a receiver bandwidth of 30 to 36 MHz.

The ratio of deviation to modulation frequency is called the <u>modulation index</u>. The Greek letter β (beta) is usually used to represent this quantity. It represents the maximum phase shift of the carrier wave.

$$\beta = \text{Deviation} / \text{Modulation frequency} = \text{Modulation Index}$$

Note that deviation and modulation frequency are in the same units of frequency (KHz, MHz, etc).

In general, the higher the value of β the better the signal to noise ratio of an FM

system compared to an AM system. In a typical FM broadcast transmitter situation, with a 1 kHz audio signal at 50% modulation (37.5 kHz deviation):

$$\beta = \frac{37.5 \text{ KHz}}{1 \text{ kHz}} = 37.5$$

FM is often done by applying audio to a frequency determining network component whose value is voltage dependent, such as a varactor diode. This is called Direct FM, since the frequency of the oscillator is being directly controlled by the modulating signal. Related mathematically to FM is PM, or phase modulation. In this case the phase of the RF carrier is modulated by the applied modulating signal. Phase modulation is generated by passing an RF signal through a phaseshift network whose characteristics are modulated by the modulating signal. It can also be generated indirectly using a balanced modulator and RF carrier phase shift networks. See Fig 2. From this system FM can be generated by simply integrating the modulating signal. In practice an R-C circuit can be used to integrate the modulating signal. It must be stated that this system is practically suitable for only small amounts of phase deviation, about 10-15 degrees of phaseshift. (0.2 radians or so). Using this method to generate FM is called Indirect FM. To get practical amounts of deviation when indirect FM is to be used, a frequency multiplier usually is needed. Often the degree of modulation requires a large multiplication factor. What is done here is that the signal is multiplied, mixed down to a much lower frequency, and multiplied up again. The deviation produced by the modulator is also multiplied by the multiplication factor, but unaffected by addition and subtraction. Therefore, if one starts out with a 1 MHz modulated signal, multiplies it by a factor of 64, then mixes the resultant 64 MHz signal down to 1 MHz again using a 63 MHz oscillator and mixer, the new 1 MHz signal can be multiplied by say 96 to get 96 MHz, in the middle of the FM broadcast band, with an effective multiplication factor of 64 X 96 = 6144. This would allow a modulation index of 1228. This would allow full 75 kHz deviation with an audio modulation frequency as low as 61 Hz, a situation that practically would seldom arise. One must start out with a high quality and low phase noise RF signal, as any FM noise on the original carrier would get multiplied by this 6144 factor as well.

Without going through the derivation, the signal to noise ratio improvement gained by using an FM system as compared to an AM system of the same bandwidth is related by the FM improvement factor

$$S / N = 3\beta^2$$

This is valid as long as the FM system is operating above a threshold level which is typically 8 to 10 dB carrier to noise ratio. The receiver must have a certain minimum signal for this to happen. Otherwise the advantage of FM over AM is rapidly lost for weaker signals. For a typical 2 way FM communications system with a 13 kHz bandwidth, the noise level (Assuming a 3 dB receiver noise figure) would be around –128 dBm, slightly less than 0.08 microvolts into 50 ohms. The carrier level would have to be about 8 to 10 dB stronger, or about 0.2 microvolts to achieve threshold. This is a typical rated sensitivity for many 2 way FM communications receivers, such as used for public service, and amateur FM. An stereo receiver with 250 kHz bandwidth would need a signal of about 1 microvolt at threshold, and you would probably not want to listen until the S/N ratio exceeded about 30 dB minimum. This increases rapidly above threshold, and about 2 microvolts or so would do. Receivers are also specified in dB quieting, which is a measure of how much the receiver

noise decreases on receipt of a carrier. This can be misleading as it does not take into account anything to do with the modulation. A measurement called SINAD (ratio of SIGNAL / (NOISE + DISTORTION) is often used, and 12 dB SINAD, the often used standard ratio corresponds approximately to 20 dB quieting for many FM receivers in practice. This is done by adjusting the carrier level from an FM signal generator feeding the receiver until the audio signal out of the receiver is 12 dB above the receiver output when the modulation at that carrier level is reduced to zero. This is about minimum signal level for reasonably intelligible speech output.

Note that, for constant modulation signal level, β increases at low audio frequencies, and decreases at high audio frequencies. Thus the signal-to-noise ratio of an FM system is poorer for high audio frequencies than for low frequencies. This is the reason <u>pre-emphasis</u> is used in FM broadcasting, so as to boost the modulation index at the high end of the audio spectrum. It turns out that, additionally, the noise spectrum of a detected output, in an FM receiver, has a rising noise vs frequency characteristic. If we employ <u>de-emphasis</u> at the receiver, we not only correct for the pre-emphasis at the FM broadcast transmitter (which would make the audio sound "tinny"), but also pick up <u>additional</u> signal-to-noise improvement at the receiver as well. Pre-emphasis and de-emphasis are used in both broadcasting and commercial FM two-way radio (which usually uses +5 kHz deviation) to materially increase signal to noise ratio. This gives FM its excellent quality and "clean" sound. Since most static is AM in a nature, very little static gets into the system. An FM receiver deliberately limits the incoming signal to a constant level before detection further reducing AM static rejection.

The occupied bandwidth of an FM signal, at first glance, appears to be simply the peak to peak deviation. However, this is not generally true. A 75 kHz deviation FM broadcast signal requires somewhat more bandwidth. It is necessary to know this required bandwidth for various reasons, among them channel spacing, necessary receiver bandwidth, and signal-to-noise ratio considerations. For very high modulation indices, the necessary bandwidth is very close to the peak-to-peak deviation. As an example, this would be true for, say, a 100% modulated FM signal (75 kHz in commercial broadcasting) with a low audio frequency modulation (i.e., 20 Hz). However, the situation changes for low modulation indices. At a modulation index of 10 the bandwidth required would be about 2.8 times 75 kHz, or 210 kHz. At a modulation index of 5 (for example 15 kHz audio, with 75 kHz deviation) about 3.3 times 75 kHz, or 247 kHz bandwidth, would be required. (See Figure 3). This bandwidth change is due to the sidebands generated in FM. The sidebands, as in the AM case, are separated by the modulation frequency from the carrier. However, depending on the modulation index, the sidebands vary in amplitude. They appear, reach a maximum, then, at higher modulation indices , some sidebands disappear. In fact, the carrier disappears at a modulation Index of 2.4. This means, if we apply a tone of about 31 kHz to an FM transmitter and adjust the level of this tone to produce a deviation of 75 kHz, the carrier will actually null out. Of course, the FM signal has not disappeared. All of its energy is now contained in sidebands spaced 31 kHz from the carrier , at +31 kHz, +62 kHz, +93 kHz, etc. (This is somewhat like what happens in an SSB signal, where all the energy is in one sideband, with the carrier and other redundant sideband being suppressed.) As a matter of fact, the null of the carrier at a modulation index of 2.4 can be used to measure the deviation of an FM transmitter, using a spectrum analyzer. See Fig 3 for a typical FM signal spectrum.

The sidebands vary in amplitude in a pattern that can be expressed mathematically by Bessel functions of the first kind. A given sideband N, where N is the order (carrier N=0,

first sidebands N=l, second N=2 etc.) of the Bessel function. We will not discuss this further, except to state that Figure 5-4 is an approximate illustration of how these functions appear. Further discussion would require the reader to have a background in analytic geometry and integral calculus. We cannot assume this and instead refer the interested reader to a more advanced text on modulation theory.

As a rule of thumb that works out in practice, the required bandwidth is <u>approximately</u> twice the deviation plus the highest modulating frequency. This figures out to about 240 kHz for an FM Stereo receiver, and bandwidths of 230 to 280 kHz are used in practice. Note that the FM detector must be very linear over this bandwidth to reduce distortion on the recovered audio and to maintain good stereo separation.. As another example, commercial 2-way FM radio used for police, fire, taxicab, etc and also 2 meter amateur radio FM use ±5 kHz deviation with audio restricted to 3 kHz (3000 Hz). These receivers use 12 to 13 kHz bandwidth IF filters. This is twice the deviation plus the highest modulation frequency, as is seen from the figures given. The various components of an FM broadcast (stereo) signal are as follows:

A.   Audio baseband 0-15 kHz. This is comprised of the <u>sum</u> of left and right audio channels, and is the basic program audio a monophonic FM radio would receive.

B.   Stereo baseband (19 kHz and 23-53 kHz). This consists of a pilot carrier at 19 kHz, and a DSB (double sideband) suppressed carrier AM signal centered at 38 kHz. The 38 kHz carrier is suppressed, and a low level pilot carrier at 19 kHz is used to regenerate the 38 kHz suppressed carrier. In this way the 38 kHz DSB AM signal is recovered and detected. It has the <u>difference</u> of left and right channel audio. This difference signal is zero for mono transmission.

B.   SCA and RDS subcarriers at 57, 67, and 92 kHz with up to 7.5 kHz deviation.

Adding this up, twice the deviation plus the highest modulating freqiency would be twice 75 kHz plus 92 ± 7.5 kHz, or about 250 kHz.
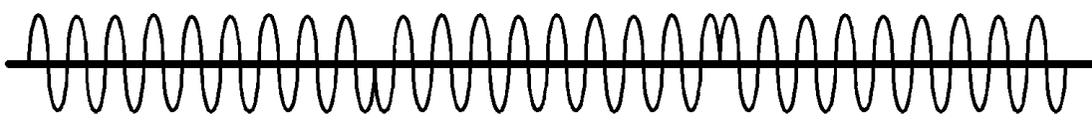
We have tried to present some of the basics of FM without going into the mathematics behind it. For more information on FM theory and practice, the reader should consult a good textbook on communications, and be prepared to go into the mathematics behind it all.
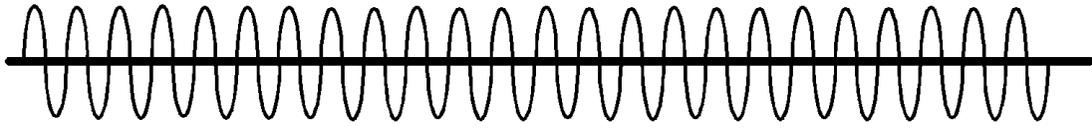
A) UNMODULATED CARRIER WAVE
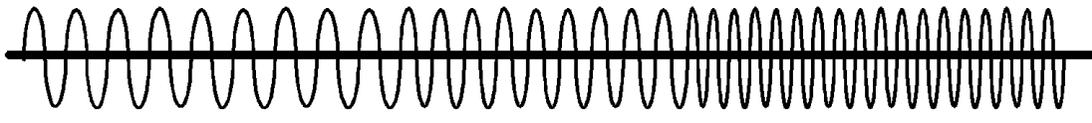
B) CARRIER WAVE WITH TWO STEPS IN FREQUENCY, MODULATED BY STAIRCASE WAVEFORM

C) PHASE MODULATED CARRIER WAVE, SQUARE WAVE MODULATING SIGNAL
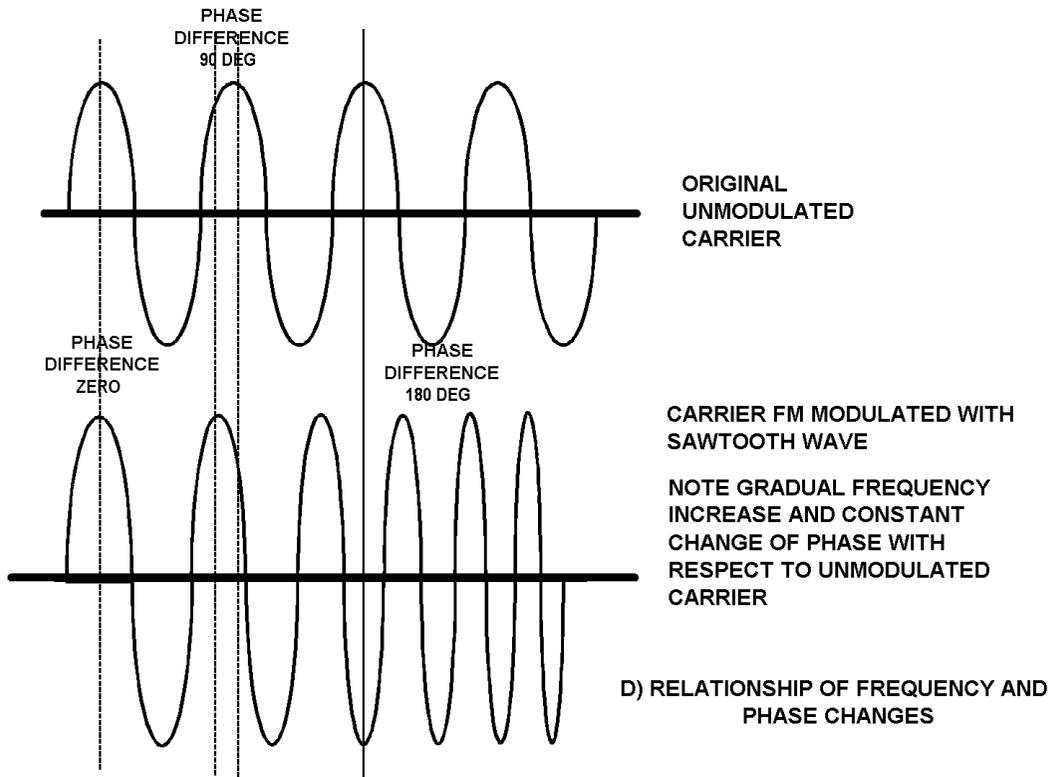
PHASE DIFFERENCE 90 DEG

ORIGINAL
UNMODULATED
CARRIER

PHASE DIFFERENCE ZERO

PHASE DIFFERENCE 180 DEG

CARRIER FM MODULATED WITH
SAWTOOTH WAVE

NOTE GRADUAL FREQUENCY
INCREASE AND CONSTANT
CHANGE OF PHASE WITH
RESPECT TO UNMODULATED
CARRIER

D) RELATIONSHIP OF FREQUENCY AND
     PHASE CHANGES

FIG 1 FREQUENCY AND PHASE MODULATION
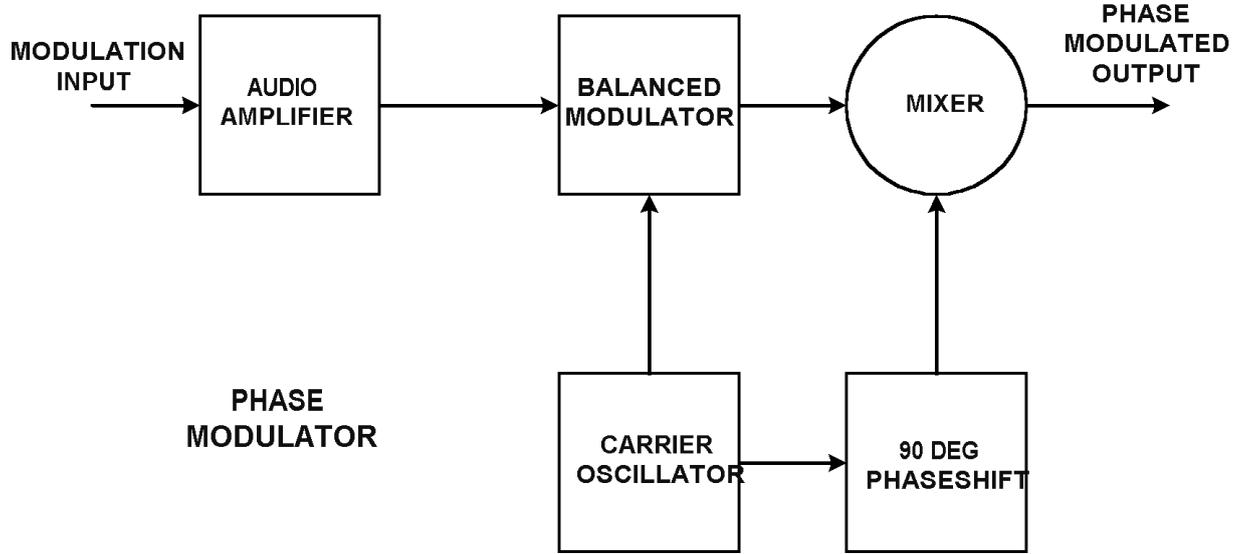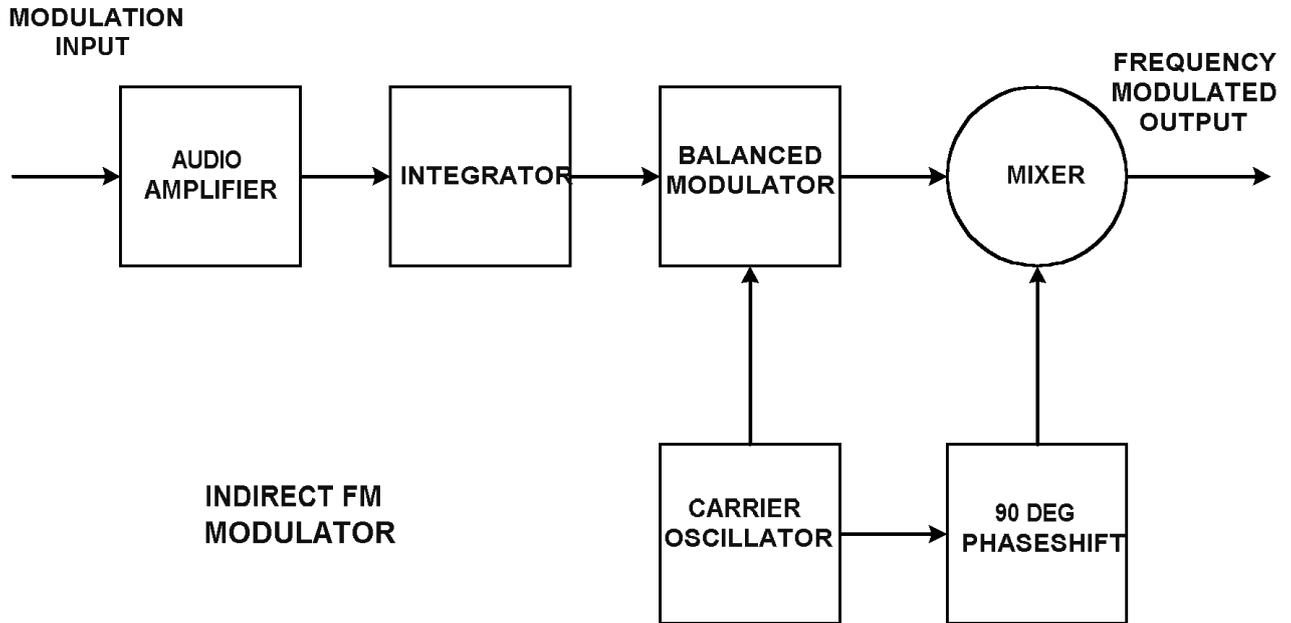
FIG 2 PHASE MODULATOR AND INDIRECT FM GENERATION

CARRIER = fc

SMALL DEVIATION
B = 0.2

APPEARS LIKE
AN AM SIGNAL

fc - fm   fc + fm

BANDWIDTH →| |←

CARRIER = fc
DECREASED AMPLITUDE

MORE DEVIATION
B = 1
2nd SET OF
SIDEBANDS HAVE
APPEARED

fc - fm   fc + fm

fc - 2fm   fc + 2fm

BANDWIDTH →| |←

MORE DEVIATION
B = 5
MORE SIDEBANDS APPEAR

CARRIER = fc
DECREASED AMPLITUDE

fc -8fm   fc + 8fm

|←   DEVIATION   →|

|←   BANDWIDTH   →|

MORE DEVIATION
B = 10

fc -14fm   fc +14fm

|←   DEVIATION   →|

|←   BANDWIDTH   →|

|← DEVIATION →|

POSSIBLE APPEARANCE
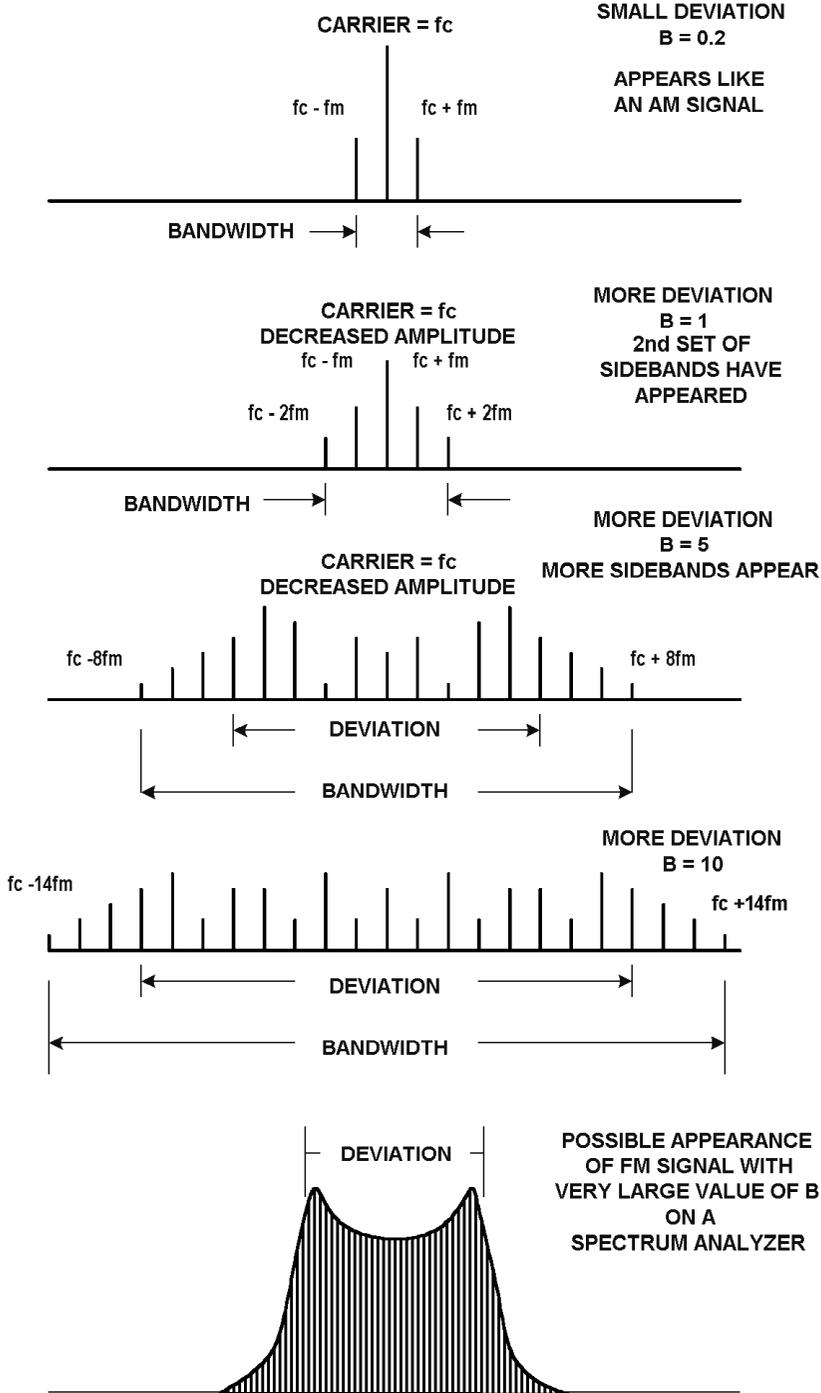OF FM SIGNAL WITH
VERY LARGE VALUE OF B
ON A
SPECTRUM ANALYZER

FIG 3
FM SIGNAL SPECTRUM BEHAVIOR AS DEVIATION
INCREASES