MODULATION ANALYSIS FORMULA

When conducting a modulation analysis, the engineer is concerned primarily with two things, the input signal-to-noise ratio and the output signal-to-noise ratio of the signal processing path.

These are defined as

CNR = Predetection (input) signal-to-noise power (watts)

SNR = Postdetection (output) signal-to-noise ratio (power) (watts)

The universal equation is

$$\mathrm{SNR} = NOPQR(E_b/\eta) \tag{1}$$

This equation relates all of the factors. The five sequential letters *NOPQR* are a mnemonic device.

N stands for number of bits per symbol. Some modulation methods combine two or more bits into a symbol [quadrature phase shift keying (QPSK), multiple phase shift keying (MPSK), quadrature amplitude modulation (QAM), minimum frequency shift keying (MFSK)]. This combining results in the terms “dibits,” “tribits,” and so on. *N* is expressed in bits/s/Hz. For the relationship between bits/symbol *N* and modulation states *M*, see Eq. (2):

$$2^N = M \tag{2}$$

O stands for other. This term takes in other special factors. For FM it is 3/2. When convolutional or Viterbi coding is used, it can assume several values. It is a numeric value.

P stands for power. All modulation methods start at full power then lose power as the number of *modulation states M,* or the error angle β changes. It is a voltage-squared relationship; that is use

$$\sin^2, \quad \left[\frac{1}{(\sqrt{M}-1)}\right]^2, \quad \beta^2$$

Q stands for bandwidth (BW) efficiency. This term is also called “processing gain,” or the ratio of input noise power to output noise power. It is expressed as

$$\text{Sampling rate/Filter bandwidth} = Q = Gp$$

or

$$\text{Noise power in/Noise power out} = Q$$

Both *N* and *Q* are expressed in bits/s/Hz. In most methods, $Q = 1$. It can acquire a numerical value other than one when the noise power relationship is altered by filtering or spreading as in minimum shift keying (MSK), MFSK, variable phase shift keying (VPSK), or quadrature partial response shift keying (QPRSK).

R stands for reduction factor. The effect of phase noise can be reduced by additional baseband filtering—for example, by using a phase-locked loop (PLL) as a tracking filter. It is equal to

$$\text{Nyquist BW/Filter BW} = R$$

It may appear to be related to *Q* and sometimes has a numerical value $= \frac{1}{2}Q$, but it is a different and separate effect (phase noise reduction ratio). It is a dimensionless numerical value.

E_b = Bit energy, in joules. Its value is

E_b = Signal power/(bits/s)

η = Noise power in watts, in a 1 Hz bandwidth. Its value is

$$\eta = \text{Total noise power/Filter noise bandwidth}$$

CNR, the input signal/noise ratio, usually noted *C*/*N*, is NE_b/η, where N = bits/symbol.

When *N* is varied, E_b/η = CNR/*N*. Bit energy is lost as *N* increases.

The group *NOPQR* can be used to calculate the equivalent of Shannon’s limit, as is noted later.

PREENCODING THE DATA: NRZ LINE CODE VERSUS BIPHASE MODULATION

Data can be encoded ahead of the modulator. This can drastically alter the spectrum and the analysis of the signal. If the preencoding does not change the time periods for the one and zero bits, the method is referred to as “non-return-to-zero (NRZ) line-coded.” If the bit time is altered, it generally becomes “biphase”-coded data.

Most modulation methods are NRZ line code. That is, the data bits are unaltered prior to modulation. When the bits are encoded to some form other than simple ones and zeros, usually to prevent 0 Hz from appearing, a biphase code results. Typical of the codes which alter the ones and zeros are Manchester and modified frequency modulation (MFM). The encoding used for Ethernet is a biphase code. The encoding used for double-density disk recording (MFM) is a biphase code. The slip codes used in VPSK are variations of MFM.

NRZ codes result in a spectrum that extends from 0 Hz upward. When using dibits, and so on, the bandwidth shrinks toward 0 Hz. They always occupy the entire Nyquist BW. A biphase code, on the other hand, causes the spectrum to center on a frequency near 0.5 bit rate. The biphase codes MFM and VPSK have spectrums that extend from 0.5 bit rate down. 0 Hz is avoided. MFM extends from 0.25 BR to 0.5 BR, occupying the upper half of the Nyquist BW, while VPSK can occupy the upper 1/5 or less. Very minimum shift keying (VMSK) modulation uses an aperture code to reduce the bandwidth used to less than 1/10 the Nyquist bandwidth.

Figure 1 (a) shows the typical spectral pattern when $Q = R = 1$. This is the pattern for all methods other than single–sideband or biphase modulation. The area marked *W* varies with *N*. This is for NRZ line-coded data. Biphase is not the same as bipolar.

In Fig. 1(a), the total RF bandwidth = data rate, double sideband = f_b, when $N = 1$. The sampling rate is *W*. The ratio of $f_b/W = N$ as shown. If *N* is greater than 1, the RF bandwidth = *W* for NRZ line code methods.

The Nyquist bandwidth is usually defined as the minimum bandwidth at baseband that can be used to pass the information (assuming it is all used). It is also one-half the double-sideband sampling rate, or *W*/2. (The data rate clock is twice the frequency of the data.) It is not the bandwidth needed to pass the data at RF where it can be modified by *Q*, as in MSK,

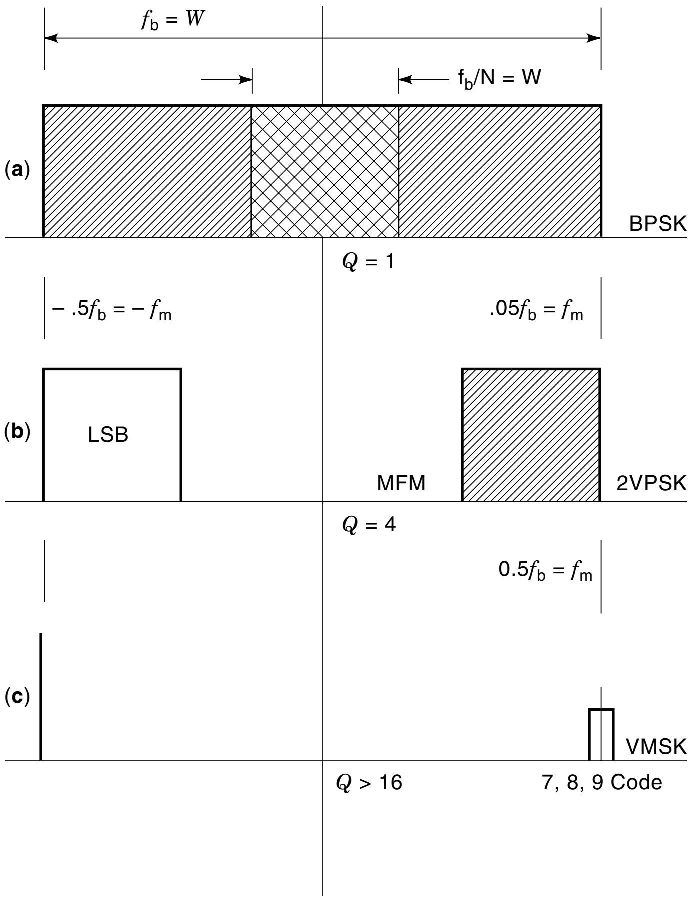

Figure 1. Typical spectral pattern when $Q = R = 1$.

Gaussian minimum shift keying (GMSK), or VPSK. Nor does it express the baseband bandwidth needed when biphase modulation is used. It is interpreted here to mean all of the band from 0 Hz to a frequency $f_m = W/2$, not the actual bandwidth needed.

There is a relationship here that cannot be altered.

(Sampling rate) × (Bits/Symbol) = Bit rate or:

$$WN = \text{Bit rate } (f_b)$$

Some confusion is possible here since these terms can be expressed as double-sideband RF, or baseband, or single sideband. *Be careful when selecting a Nyquist BW.*

When double-sideband RF terms are used: f_b = bit rate and W = sample rate = 2 × Nyquist BW, but in baseband terms for data transmission, the transmission BW (f_m) is the modulation frequency and is equal to 1/2 the bit rate if $N = 1$. The *unmodified* Nyquist BW is therefore f_m.

If the data rate is X bits per second, then it must be sampled X times per second to extract any useful data. By combining bits into dibits or tribits, the values of $2f_m$ and W shrink together in a 1:1 relationship while f_b is fixed. W shrinks according to

$$f_b/N = W$$

This is inviolate. In all transmission methods, the bandpass filters must pass the frequency W (or f_m), but not necessarily a "bandwidth" = W. Expressed differently, to pass a tone of 1 kHz, the filter must pass 1 kHz, but there is no need to pass 500 Hz or 287 Hz. This is related to the terms Q and R.

W is the sampling rate. In some cases it can be the bandwidth used, but very rarely is in practice. Never use bandwidth as a term in calculating Shannon's limit.

Figures 1(b) and 1(c) show the spectrum where $N = 1$, but Q and R have values other than 1. This is the spectrum for biphase modulation as opposed to NRZ line code modulation. W very definitely does not equal the bandwidth used.

The spectrum in Figs. 1(b) and 1(c) differ from the spectrum in Fig. 1(a) in that the data information is located at or slightly below f_m and all of the spectrum not in use below f_m can be filtered off. This yields some spectacular results as will be shown later. In this case, $f_b = W = N = 1$.

MODULATION METHODS COMPARED

Constellations and Eye Patterns

All modulation methods other than bipolar phase shift keying (BPSK), multilevel amplitude shift keying (MASK), and FM involve a phase rotation utilizing two axes in quadrature (I and Q). This results in a "constellation" pattern used in the illustrative figures. This constellation is relative to a reference frequency. Using the constellations makes it easier to measure and interpret the modulation methods. In some cases it is desirable to use an "eye" pattern to analyze the results being obtained in practice, but eye patterns have little use in theoretical analysis except for VPSK. Figure 2 shows the performance of various modulation types.

AM with Suppressed Carrier (AM-SC) and Bipolar Phase Shift Keying (BPSK)

There are no AM methods in general use for data transmission that use AM in the manner used for audio. The data methods are transmitted with a suppressed carrier so that the two contra-rotating sideband vectors add vectorially to

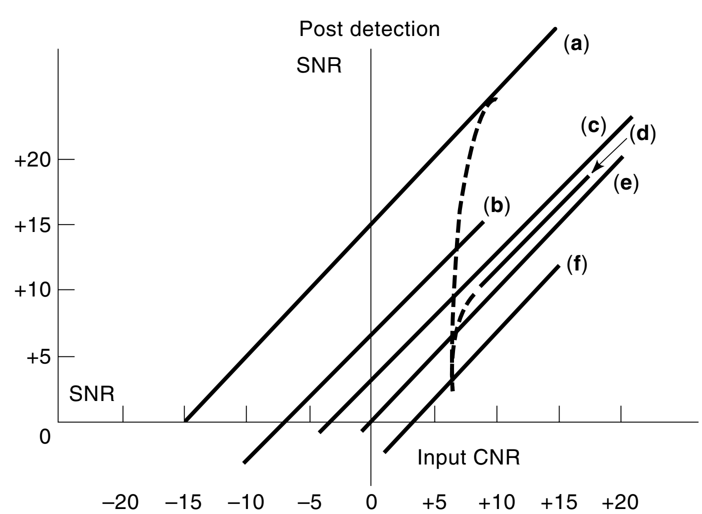

Figure 2. Performance of various modulation types. (a) FM, $\beta = 5$; (b) VMSK; (c) SSB-AM/16FSK; (d) FM, $\beta = 1$; (e) AM; (f) 16QAM.

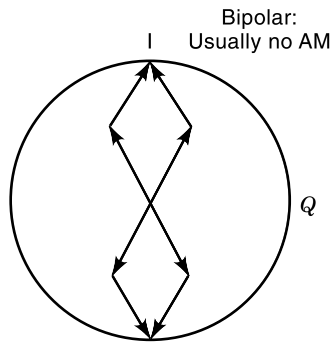

Figure 3. Bipolar modulation vectors.

produce a positive and a negative voltage output. They thus become bipolar in nature (Fig. 3).

Bipolar AM with suppressed carrier is indistinguishable from bipolar phase shift keying (BPSK). In both cases the resultant detected signal from the RF signal is a result of the added sideband vectors—a bipolar signal at baseband.

The spectrum for AM and BPSK is shown in Fig. 1(a). The information rate and the sampling rate are the same, so $N = 1$. The spectrum shown in Figs. 1(b) and 1(c) does not apply, so $Q = 1$ and $R = 1$. Without convolutional coding, $O = 1$. The only factor remaining is P, which is 1 for 100% AM modulation.

Equation (1) for AM becomes

$$\text{SNR} = (1)(1)(1)(1)(1)E_b/\eta = E_b/\eta$$
$$\text{SNR} = E_b/\eta$$

and since NE_b/η = CNR, the input CNR = the output SNR.

In the real world, this would be true if an ideal filter were being used and the signal transmitted with suppressed carrier. The SNR is reduced as the filter bandwidth increases from ideal.

BPSK is the simplest of all data modulation methods. It is also known as phase reversal keying (PRK) or 2PSK. To implement BPSK, one can use an XOR gate or a double balanced mixer. The signal consists of phase 1 at 0° for a digital one and phase 2 at 180° for a digital 0.

The modulation angle is 180°, or ±90°. In analyzing the signal, the error angle ($\beta = 90°$) is used. If the signal path resulted in a phase error of 90°, one could not tell a 1 from a 0.

The power P is determined from the sine of the error angle (sine 90° = 1) (same value as for 100% AM modulation). This is an NRZ line-coded method [Fig. 1(a)]. $Q = R = N = O = 1$.

There are no dibits, and so on, so $N = 1$.

$$\text{SNR} = (\text{sine } 90°)^2(1)(1)(1)(1)E_b/\eta$$
$$= E_b/\eta = \text{CNR}$$

BPSK is the method against which all other digital modulation methods are compared. Again, real-world conditions come into play. The bandwidth should be that shown in Fig. 1(a), but the harmonics of a square-wave signal used to drive the biphase modulator cause considerably high-level out-of-band products that must be filtered off.

Differential Phase Shift Keying

Differential phase shift keying utilizes a delay line and XOR gate to compare the last bit in the stream with the present bit. The difference between the two is determined, and the result is used by the modulator. The signal remains NRZ line-coded.

This method has the advantage that a coherent carrier is not required to recover the data. A reference carrier that can be as much as 6 ppm off from the actual suppressed carrier can be used. The SNR relationship is unchanged, but there is a slight increase in error rate since the errors occur in pairs. The differential preencoding method is also applicable to QPSK.

Multilevel Amplitude Methods

By adding a digitally controlled attenuator to a BPSK modulator, the amplitude can be varied in multiple levels as well as the phase. There are two multilevel amplitude methods in use: MASK and QAM. QAM (quadrature amplitude modulation) is ASK (amplitude shift keying) applied to two axes (I and Q). The analysis procedure is the same except that for amplitude methods the value for P is $(1/(\sqrt{M} - 1))^2$. To present an example, assume we wish to analyze 4ASK and 16QAM. Figure 1 applies, and it is still an NRZ line-coded method: $Q = R = O = 1$.

$$\text{SNR} = \left[\frac{1}{(\sqrt{M} - 1)}\right]^2 NE_b/\eta = PNE_b/\eta$$

For 4ASK, $N = 2$ and $M = 4$ [see Eq. (2)].

$$\text{SNR} = (1/3)^2 2E_b/\eta = (2/9)E_b/\eta$$

ASK modulation is little used, since the bandwidth efficiency can be doubled by using QAM, modulating to four different levels on each axis (I and Q). The number of modulation distances is squared while there are twice as many bits per symbol N. ($N = 4$)($M = 16$). The constellation for MASK is shown in Fig. 4.

For 16 QAM (a 4 × 4 dot pattern) we have

$$\text{SNR} = (4/9)E_b/\eta = 2 \times \text{Eq. (1d)}$$

E_b/η must be raised by 3.5 dB to maintain the SNR level, but the CNR level must be raised by 6 dB above the E_b/η level.

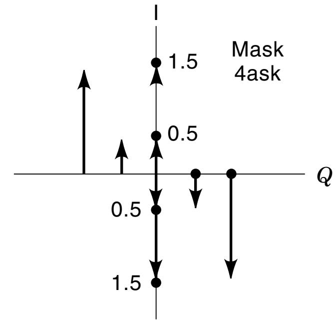

Figure 4. MASK. Each position represents 2 bits.

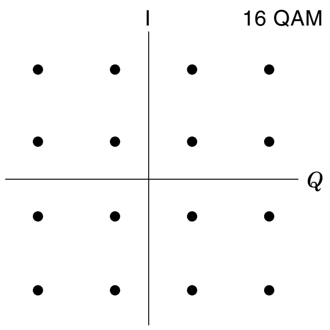

Figure 5. 16 QAM. Each position represents 4 bits. (0010,0110,1101, etc).

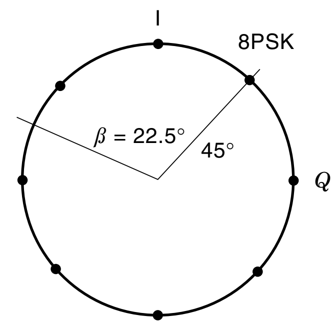

Figure 7. Each position represents 3 bits: 001, 101, 110, and so on (3 bits/symbol).

The constellation for 16QAM is shown in Fig. 5. The constellation for 4ASK is only one of these columns (Fig. 4).

QPSK

Quadrature phase shift keying is used to reduce bandwidth, theoretically by 2/1. In the real world it is much less. Dibits are used to compress the transmitted bandwidth. The spectrum is that of Fig. 1(a), with f_m and W both cut in half, so the ratio is the same: $N = 2$, $Q = R = O = 1$. There are four possible phase points, each 90° apart, as shown in Fig. 6. The error angle is 45° (sine 45° = 0.7). Equation (1) becomes

$$\begin{aligned} \text{SNR} &= (0.7)^2(1)(1)(1)(2)E_b/\eta = (0.5)\,2E_b/\eta \\ \text{SNR} &= E_b/\eta \qquad \text{(the same as for BPSK)} \end{aligned} \tag{1a}$$

CNR = $2E_b/\eta$. We need twice as much input signal-to-noise power. Since $N = 2$, the theoretical bandwidth efficiency is 2 bits/s/Hz.

Multiple Phase Shift Keying (MPSK) and Vestigial Sideband PSK (VSB)

Multiple phase shift keying (MPSK) is an extension of QPSK; or the reverse is true, QPSK is 4PSK. Assume that instead of dibits, we use tribits—that is, three bits per symbol. Figure 1 still applies, but the bandwidth and the sample rate are reduced to 1/3 f_b ($N = 3$). The bandwidth efficiency is nominally 3 bits/s/Hz.

When using multiple bits/symbol, the relationship

$$2^N = M \tag{2}$$

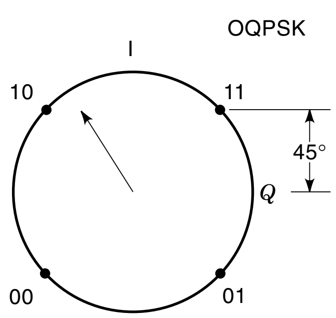

Figure 6. OQPSK (offset QPSK). Each position represents 2 bits.

applies. If there are three bits per symbol (N), then there are eight modulation states (M), or eight points on the modulation constellation. The error angle is 180/8 = 22.5° (sin 22.5° = 0.38). Therefore P becomes 0.38 squared. $N = 3$ and $Q = R = O = 1$.

$$\begin{aligned} \text{SNR} &= \beta^2 N E_b/\eta \\ \text{SNR} &= (0.38)^2 N E_b/\eta = (0.146)3E_b/\eta \\ &= 0.44E_b/\eta \qquad \text{for } N = 3 \end{aligned} \tag{1b}$$

The level of E_b/η must be raised 3.6 dB to maintain the SNR, and CNR must be raised 3 times or 4.7 dB to maintain the same SNR level. The 3.6 dB is also approximately the value of Shannon/s limit for 8PSK.

Figure 7 shows the constellation for MPSK.

Other MPSK levels can be analyzed using the same formula. The error angle β is π/M. The value of N is in bits per symbol, and also the nominal bandwidth efficiency is in bits/s/Hz.

The coming high-definition TV standard calls for vestigial sideband PSK (VSB). Two standards are proposed, vestigial 8PSK and vestigial 16 PSK. The analysis is the same as above except that $Q = 2$. 8VSB would have SNR = $0.88E_b/\eta$.

FM

For narrow band FM ($\beta < 0.8$), the spectrum is the same as that seen in Fig. 1 for AM. If it is broadband FM ($\beta > 0.8$), the spectrum spreads according to Carson's rule:

$$BW = 2(f_m + \text{deviation})$$

A 15 kHz signal deviated 75 kHz occupies a bandwidth of 180 kHz.

P = the square of the modulation index β. The sample rate $W = f_b$ for narrow band FM (NBFM) and the bits/symbol $N = 1$. Since the spectrum of Fig. 1(b,c) does not appear, $Q = R = N = 1$ and $O = 3/2$. Let the modulation index β equal 0.5, then Eq. (1) becomes

$$\begin{aligned} \text{SNR} &= 3/2\beta^2 E_b/\eta \\ \text{SNR} &= (0.5)(0.5)(3/2)E_b/\eta \\ &= (0.25)(3/2)E_b/\eta = 0.375E_b/\eta \end{aligned}$$

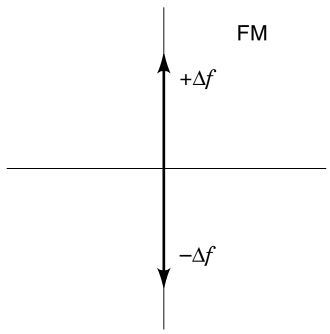

Figure 8. FM vectors.

and since $N = 1$:

$$\text{SNR} = 0.375\,\text{CNR}$$

There is a loss compared with AM (no loss if $\beta = 0.8$ or more).

When the equivalent of FM is produced by the two-tone method (MSK), the bandwidth is supposedly 1.5 times that for BPSK. When narrow band FM is used, it is the same as for BPSK ($BW = 2f_m$).

If the modulation index, β, is 5, as in wideband FM, then

$$\begin{aligned}\text{SNR} &= (5)(5)(3/2)\,\text{CNR}\\ &= 75/2\,\text{CNR} = 37.5\,\text{CNR}\end{aligned}$$

which is an improvement of 16 dB in SNR over AM at the expense of six times as much bandwidth.

This can be seen in Fig. 2. Figure 2 is of importance in several other modulation methods or combination methods as will be shown later. The constellation is shown in Fig. 8.

The sloping lines in Fig. 2 show the possible extension of SNR down to the 0 dB level, which is far below the CNR = 0 level. At this level it is assumed the signal cannot be separated from the noise and there is a 50/50 chance of error. In the real world, there is an FM knee below which the system cannot operate. This is shown in Fig. 2. The FM knee does not apply to AM, and there is good reason to believe that it does not apply to all phase shift keying methods.

Shannon's limit for a method in which $N = 1$ is 0 dB, but the curves in Fig. 2 show a possible much lower limit, which can actually be obtained by certain signal processing methods, particularly those using phase-locked loops (to be discussed later).

In the real world, the use of a square-wave data input to modulate FM causes extensive harmonics of the modulating frequency plus numerous Bessel products to appear that cause the bandwidth used to spread far beyond that indicated by Carson. Extensive filtering must be used to get rid of this out-of-band radiation. For this reason, straight FM is little used for data. MSK or GMSK are preferred methods.

Preliminary Filtering: Gaussian MSK

To reduce the sideband energy, the incoming data can be filtered to reduce the amplitude level of the higher frequencies in the data. These filters usually have a characteristic such as BT = 0.5 or BT = 0.3. T in this case is $1/f_b$ and B equals the 3 dB bandwidth of the filter so the filter actually has a 3 dB roll-off at $\text{BW}/f_b = 0.5$ or more at $\text{BW}_b/f = 0.3$. When $\text{BW} = f_b$, the 3 dB points are at f_m.

Using a filter alone with an FM voltage-controlled oscillator (VCO) and a carefully controlled modulation index will create a four-phase effect also accomplished by code generated by MSK. For $N = 1$, the constellation resembles that for QPSK (Fig. 9).

$$\text{SNR} = (0.7)^2 E_b/\eta = 0.5 E_b/\eta$$

This is first level binary Gaussian minimum shift keying. The Bessel products are removed or reduced, so that the BW is only slightly larger than that for theoretical BPSK.

MSK (Code-Generated)

Minimum frequency shift keying is related to filtered binary FM in that two frequencies are involved to obtain a Gaussian response, but the frequencies are chosen very carefully to change smoothly and have a 180° phase relationship at the sampling time. While this can be done with a VCO and filtering, it is usually done with a more sophisticated direct digital synthesis (DDS) chip and a computer program. It can be shown that if two frequencies $f_c + f_b/4$ and $f_c - f_b/4$ are used, the spectrum spread is minimized. The modulation index is 0.5, and it must be maintained as close to 0.5 as possible to enable a reference signal recovery and prevent phase overshoot. This is known as minimum shift keying (MSK). The bandwidth occupied supposedly complies with Carson's rule and is 1.5 times the bandwidth occupied by the Nyquist bandwidth of a similar AM signal ($Q = 1/1.5$), but it can also be shown that most of the energy lies within 1.2 times the Nyquist bandwidth. $N = R = 1$, $O = 3/2$. This method can use two separate detecting filters that are made very narrow to pass only the upper or lower frequency. As far as the receiver is concerned, the noise bandwidth is halved so Q becomes $2/1.5\, f_m/Bf = R = 1$.

$$\text{SNR} = 3/2\beta^2(2/1.5)E_b/\eta = 2\beta^2 E_b/\eta = 0.5 E_b/\eta$$

The power value β must be maintained at 0.5 to reduce spread and make recovery of a coherent carrier possible, but a curious thing happens along the way. The signal acquires the attributes of a QPSK signal with both I and Q components and a similar constellation. The effective error angle β increases to sine 45° = 0.7. This can be used instead of the FM analysis with a modulation index = 0.5. The bandwith is considered to be 2 f_b. While this is consistent with the MFSK

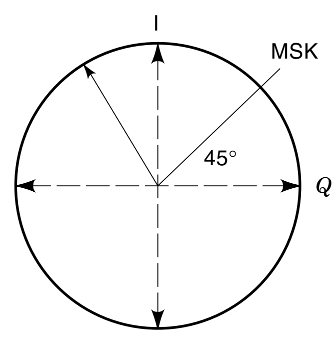

Figure 9. Minimum shift keying.

equations that follow, it is not correct for the frequency spread for MSK where $N = 1$ and one accepts Carson's rule. The spread is 1.5/1.

$$\mathrm{SNR} = (0.7)^2(1/2)(2)E_\mathrm{b}/\eta$$
$$\mathrm{SNR} = 0.5E_\mathrm{b}/\eta$$

The bandwidth efficiency in bits/s/Hz is no longer the value for N, but it is $0.66N$ or $0.5N$ according to the method used in the analysis. This points out a trap for the unwary. Do not assume that N is always the bandwidth efficiency in bits/s/Hz.

The constellation is shown in Fig. 9.

Quadrature Partial Response Keying

Using a cosine filter with QPSK creates a nine-point constellation, like that shown in Fig. 10. A quadrature partial response system (QPRS) is shown above, with a pattern that could be called 9QAM (3 × 3 dot pattern).

This constellation can be analyzed by the method used for QAM, where

$$\left[\frac{1}{(\sqrt{M}-1)}\right]^2 = 1/2$$

If it were QPSK, the value of the error angle would be 0.7 instead of 0.5, so there is a loss in power.

$$\mathrm{SNR} = (0.5)^2 NE_\mathrm{b}/\eta = 0.25NE_\mathrm{b}/\eta$$
$$= 0.5E_\mathrm{b}/\eta \quad \text{(as a first try to be corrected)}$$

However, the bandwidth is reduced to $2/\pi$ Nyquist. The bandwidth needed is not always the Nyquist bandwidth defined as $W/2$. ($Q = \pi/2$), so

$$\mathrm{SNR} = 0.79E_\mathrm{b}/\eta = \beta^2 QNE_\mathrm{b}/\eta$$

There is a theoretical loss of 2 dB [compared with Eq. (1a)] in exchange for the lower sideband levels, but this partial response QPSK is better than binary GMSK in both data rate and SNR.

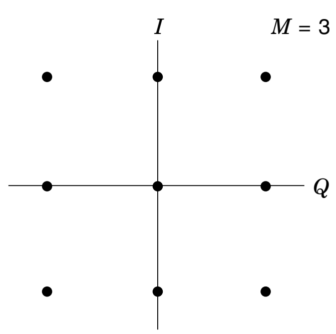

Figure 10. Quadrature partial response.

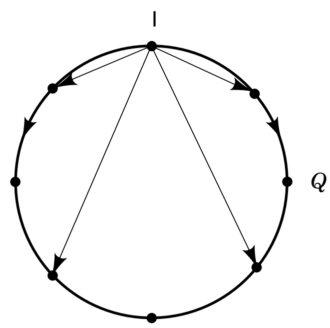

Figure 11. $\pi/4$ DQPSK. The pattern rotates CW and CCW.

(π/4) Differential Quadrature Phase Shift Keying (DQPSK)

(π/4)DQPSK is a variation of QPSK. The constellation is shown in Fig. 11. In this method, there are no repeat dibits in the same phase position as there would be with ordinary QPSK. This is done by adding to or subtracting 45° from the normal phase shift, or when a dibit is repeated, resulting in phase shifts of 45° and 135° only. $N = 2$, the error angle appears to be 45° as in QPSK, and the bandwidth is the same.

The essential difference between (π/4)DQPSK and partial response QPSK is that the latter is created by filtering and the former by a computer program that offsets the phase. (π/4)DQPSK utilizes the DPSK preencoding to eliminate the need for a coherent reference carrier, which simplifies the circuitry.

$$\mathrm{SNR} = 1.0E_\mathrm{b}/\eta = \beta^2 QNE_\mathrm{b}/\eta$$
$$\mathrm{SNR} = 1.0E_\mathrm{b}/\eta$$

The value of CNR is still 3 dB above the E_b/θ value ($N = 2$). See Appendix Note.

Tamed FM

Tamed FM is an orphan method whereby several different levels of FM deviation are used according to the incoming data pattern. This changes a one axis FM method to a form of PSK, instead of to a form of QPSK, as in MSK. The constellation (Fig. 7) resembles that of 8PSK, where the error angle is $\pi/8$.

The encoding sums three bit inputs and uses the sum to set the frequency/phase deviation. For three bits of the same polarity in the incoming stream, the phase is shifted $\pi/2$. For alternating polarity, the change is 0, and for 110, 100, 011, and 001 it is $\pi/4$. Time enters into the detected output bit; just wait and the bit will change with respect to a reference frequency. The wait cycle is 3 bits long, so there is the equivalent of $N = 3$ in the result. The equation for 8PSK applies, with $N = 3$.

$$\mathrm{SNR} = \beta^2 NE_\mathrm{b}/\eta$$
$$\mathrm{SNR} = 0.44E_\mathrm{b}/\eta$$

which is about 1 dB worse than ordinary MSK. The advantage is that the spectrum is almost totally free of any unwanted sideband overshoot and conforms to the ideal NBFM spectrum. The disadvantage is that a reference frequency must be established and maintained. Tamed FM is little used. GMSK and QPSK seem to be preferred.

A Variation of Eq. (1).

$$\mathrm{SNR} = NOPQR(E_b/\eta)$$

can be rewritten

$$\mathrm{SNR} = \beta^2 QRNOE_b/\eta$$

Substituting values, we obtain

$$\mathrm{SNR} = \beta^2 (W/\mathrm{Bi}) \cdot (f_m/\mathrm{Bf}) \cdot (f_b/W) \cdot O \cdot E_b/\eta$$
$$\mathrm{SNR} = \beta^2 (f_b/\mathrm{Bi}) \cdot (f_m/\mathrm{Bf}) \cdot O \cdot E_b/\eta = \beta^2 QRE_b/\eta$$

where

$$f_b/\mathrm{Bi} = Q \quad \text{and} \quad f_m/\mathrm{Bf} = R \qquad (3)$$

where Bi is the RF noise BW and Bf is the filter BW. This variation is important in analyzing the following methods. Do not be concerned that the term N is missing; the units (bits/s/Hz) are retained in the ratios, as in the following examples.

MFSK

MSK utilizes the $f_b/4$ frequency shift (or a multiple of it) corresponding to a 90° phase shift at sampling time, to convert what would otherwise be $a \pm 180°$ modulation method to the equivalent of QPSK, with an error angle of 45°. If additional shifts separated $\pm f_b/2$ are added to the initial shift, they create a constellation pattern similar to MASK along two axes. MSK can be adapted to multiple bits per symbol. Unlike MPSK, the error angle β remains fixed at 45° degrees and the bandwidth spreads instead of decreasing. This improves the error rate.

With the extra steps, the frequency spread is BW = (M/N)f_b, where M is given by Eq. (2) (for $N = 3$, $M = 8$). The constellation shown in Fig. 12 is like that for MSK in phase, with M points along the axes. The difference from M-ASK is that it is now a multilevel frequency change instead of an amplitude change and the phasor is on two different axes (I and Q).

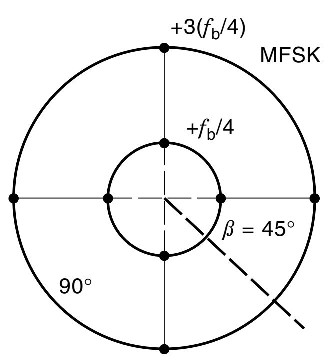

Figure 12. Minimum frequency shift keying.

The formula for MPSK [Eq. (1b)] applies, except that β remains fixed at 0.7, regardless of N. However, the formula can be derived differently.

$$\mathrm{SNR} = (0.7)^2 NE_b/\eta$$
$$\mathrm{SNR} = 0.5NE_b/\eta$$

In Eq. (3), the bandwidth is $(M/N)f_b$. There are M very narrow band filters required, one for each M position, with each having a relative bandwidth of 1/M. The noise bandwidth degrades by 1/(M/N) and is improved by M due to the narrow band filters for each frequency position. O is not used. The missing N is restored.

$$\mathrm{Bi} = M/Nf_b \times 1/M \; f_b/\mathrm{Bi} = N$$
$$\mathrm{SNR} = (0.7)^2(1/(M/N))(M)E_b/\eta$$
$$\mathrm{SNR} = 0.5NE_b/\eta$$

For 16FSK, the spread is $(16/4)f_b = 4f_b$, or 4 times that needed for BPSK.

The SNR is

$$\mathrm{SNR} = (4/2)E_b/\eta = 2E_b/\eta$$

There is an improvement in SNR of 3 dB compared with BPSK or QPSK, gained by sacrificing bandwidth.

Biphase Modulation Methods

Biphase encoding is commonly used at baseband. Ethernet, for example, uses Manchester biphase encoding. Double-density disk recording uses Miller or MFM encoding.

At the present time there is only one truly biphase baseband modulation method in use with an RF modulator. The "biphase modulation" spectrum differs from that of the NRZ line code spectrum in that it clusters the information-containing part of the spectrum at or below 0.5 bit rate and extends toward 0 Hz, while the NRZ method extends from 0 Hz up toward 0.5 bit rate. This is shown in Fig. 1(b,c). The values Q and R both become involved in biphase modulation.

VMSK Modulation

Very minimum shift keying is a new method that encodes the incoming bits to an end-to-end pulse width modulation pattern using a fractional bit stretch for each bit. The phase of the pulses changes after each stretch, so the method is an AM or a bipolar (BPSK) method at the input, but the bits coming in are stretched by a small amount instead of the usual repeated bit pattern, and the phase changes 180° plus the stretch angle after each stretched bit, or one-bit width plus a small time fraction. The alternating phase and narrow frequency band result in a biphase spectrum instead of the NRZ line-coded spectrum occupied by other methods [Fig. 1(b,c)].

The data bits are stretched according to an algorithm that says: (1) If there is no change from the last bit polarity, switch the phase 180° after each normal bit width. (2) If there is a 0–1 change, stretch the bit by a predetermined fraction of the bit width and then reverse the phase. (3) If there is a 1–0

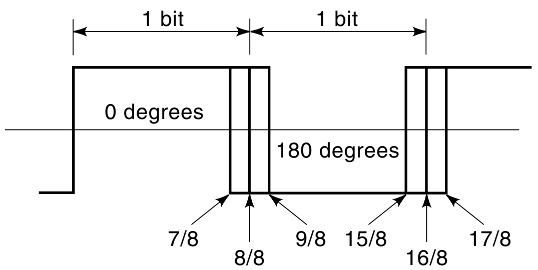

Figure 13. 7,8,9 Aperture code.

change, decrease the bit width by the fraction, then reverse phase.

VMSK modulation is a two-step process. First, the data are encoded using a biphase code known as aperture coding, which uses only a part of the baseband spectrum located at the upper limit of the Nyquist bandwidth. In the baseband form it can be used as a power line modem, or for FM-subcarrier adapter (SCA). A phase noise improvement R takes place at baseband.

Figure 13 shows the aperture coding for 7,8,9 VMSK. The bits are varied 1/8 of a bit width to indicate a change.

Figure 14 shows the spectrums involved. Figure 14(a) shows the baseband spectrum. Figure 14(b) shows the modulated double-sideband spectrum. Figure 14(c) shows the single sideband (SSB) spectrum actually transmitted [equivalent to Fig. 1(b,c)]. Figure 14(d) shows the double sideband (DSB) spectrum restored in the product detector. Figure 14(e) shows the restored baseband spectrum after detection. Note that it is a perfect restoration of the original. The Nyquist bandwidth is restored, although only partly used. Carson's rule is observed in that f_m and the sample rate W are within the bandwidth passed.

The signal is then transmitted SSB-AM-SC. (A double-balanced SSB-AM modulator is used.) The RF bandwidth used is the upper sideband shown in Fig. 1(b,c). The area ratio between the full spectrum and the part transmitted is the factor Q.

Assume 7,8,9 VMSK. The variation is 1/8 of a bit width. Thus the possible pulse widths are 7/8, 8/8, and 9/8 bit widths. The possible Fourier frequencies are 8/15, 8/16, and 8/17 times the bit rate. Calculating these values, the spectrum extends from 0.470 to 0.5333 bit rate. A spread of 0.063 bit rate. This is the only spectrum an IF filter needs to pass. The RF transmission improvement is:

Q =Sample rate/filter BW, or 1/0.063 = 16 = Q

R = (Nyquist BW/Filter BW) is a factor that has considerable noise reduction importance. In the case of VMSK it has a numerical value of $\frac{1}{2}Q$.

N = 1. The value for P is equal to $\frac{1}{2}$ the stretch converted into radians. In the 7,8,9 VMSK case, the stretch is 1/8 of the bit width. Each bit represents 180 degrees of the signal rate, hence the modulation angle is 22.5 degrees and the error angle is 11.25 degrees.

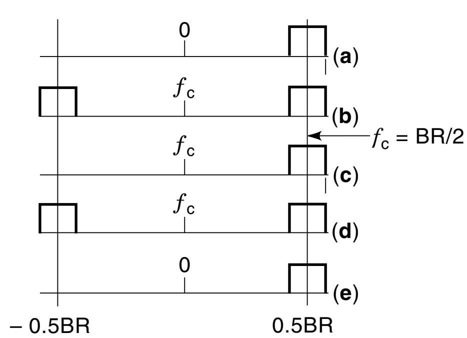

Figure 14. Spectrum for VMSK.

Since neither Q nor R are commonly associated with these RF equations, but are very important in MFSK and VMSK analysis, it should be pointed out that they have been around for 60 years or more and are generally applied to radio noise reduction and phase-locked loops, particularly in connection with tracking filters. They also apply to tone filtering in a noisy background. See Ref. 1, (pp. 57–59) and Ref. 2 (p. 324). They are definitely separate items which combine in MFSK and VMSK to result in the dramatic SNR improvement.

$$\text{SNR} = \beta^2 QRE_b/\eta$$

For 7,8,9 VMSK:

$$\beta^2 = \pi/Q,\ N = 1,\ Q = 16,\ R = 8$$

$$\begin{aligned}\text{SNR} &= (\pi/16)^2(16)(8)(E_b/\eta)\\ &= 0.5\pi^2(E_b/\eta)\\ &= 4.9(E_b/\eta)\end{aligned}$$

This is quite a change from most other methods where E_b/η must be increased to maintain SNR. Only the FM, MFSK, and VMSK methods have this SNR improvement. VMSK is transmitted SSB-AM with suppressed carrier, but the amplitude does not change and the signal can be limited, so it can also be considered SSB-FM-SC. It has an SNR improvement approximately equal to that of 16FSK. However, the bandwidth required is much less. N = 1 for VMSK at all levels. Shannon's limit is 0 dB.

This again points out the fallacy of using N bits/s/Hz as the bandwidth efficiency. In this case, N = 1 and Q is the bandwidth efficiency.

The equation is independent of the bandwidth efficiency. It doesn't matter if it is 10 bits/s/Hz or 20 bits/s/Hz, the equation and the SNR value is the same. The equivalent 0 dB SNR is 7.0 dB below 0 dB CNR. N = 1, hence Shannon's limit calculated the usual way is 0 dB, but the actual limit appears to be −7.0 dB as shown in Fig. 2.

The power lost with increased compression due to the decreasing error angle beta is regained in exactly the same ratio by the noise bandwidth reduction, or improvement in Q.

VMSK and its predecessor are the only modulation methods that do not lose power with increasing bandwidth compression (bandwidth efficiency in bits/s/Hz). They are also the only methods that have ever achieved such high compression ratios as 10/1 or 16/1 bits/s/Hz. VMSK is ideally suited for very small aperture terminal (VSAT) use where power is limited, or for radio links where the bandwidth is limited by law.

Measured values for VMSK are very close to the theoretical value for BPSK, even at 16 bits/s/Hz or higher.

Figure 14 shows the relative E_b/η values for the various methods at a bit error rate of 10^{-6}. VMSK would normally show a CNR value of 3.5 dB, but the detector uncertainty costs 2 dB to 3 dB. Hence it is shown as the equivalent of BPSK at all levels of compression.

APPENDIX

Bandwidth Efficiency

Most authors in the past were inclined to equate the bandwidth efficiency in bits/s/Hz with the value N. This is a mistake, as has been pointed out several times. The actual bandwidth efficiency can be considerably less than indicated by N. In practice, QPSK can have a bandwidth efficiency varying from 0.7 to 1.4 bits/s/Hz, depending on whether or not it is hard-limited. N, on the other hand, is 2 bits/s/Hz. In GMSK the efficiency is 0.66N. In MFSK it is much less than N.

The efficiency varies according to whether or not limiting is used and several other factors. The values for SNR must be corrected accordingly, preferably by changing Q. The values given for VMSK and QPSK above are measured values. See Ref. 3 for a comprehensive listing.

Shannon's Limit

Shannon's channel capacity equation is the inequality

$$\frac{f_b}{W} \leqq \log_2 \left[1 + \frac{f_b}{W}\left(\frac{E_b}{\eta}\right)\right]$$

If the right-hand side is greater than the left-hand side, there is adequate channel capacity. This has since been applied to mean that if SNR = 1, the lower signal limit has been reached. CNR = NE_b/n, so there is a rise in the limit as dibits, tribits, and so on, are used. Figure 15 shows Shannon's limit.

Shannon's equation is held to be absolutely, incontestably correct for the parameters used.

Note that

$$(\text{Sampling rate W})(\text{Bits/Symbol } N) = \text{Data rate } f_b$$

or

$$WN = f_b$$

or

$$f_b/W = N$$

This relationship is inviolate. W and f_b can be considered as bandwidths, but they are actually rates, with limits as bandwidths. Double-sideband W equals 2× Nyquist BW when considered a bandwidth, and f_b is the transmission bandwidth for a double-sided signal when N = 1.

There is an interpretation difficulty that arises in methods such as MFSK where the bandwidth does not equal the sampling rate W. If an attempt is made to use the actual bandwidth as W, the result is very difficult to visualize and generally incorrect. One ends up with fractional bits per symbol in MFSK. In the case of 7,8,9 VMSK, using bandwidth for W implies that 16 bits per symbol are being used, when that is clearly not the case. Only one bit per symbol is being used in VMSK.

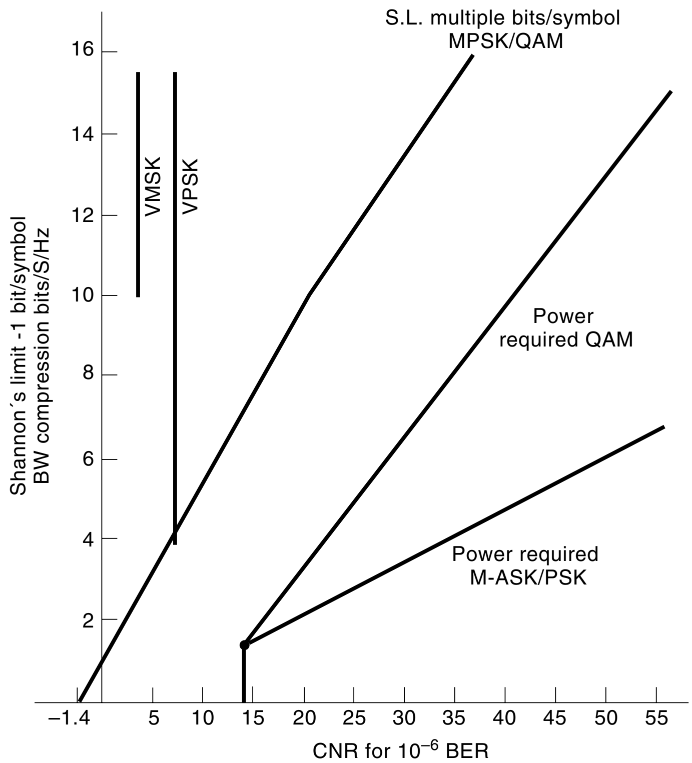

Figure 15. CNR required for 10^{-6} bit error rate using various modulation methods.

The failure in Shannon's equation to set an accurate lower limit in all cases rests on the fact that one need not use all of the Nyquist bandwidth. A narrow filter can be used to pass only the portion of the bandwidth being used.

In Fig. 2, there is a point where the SNR is 0 dB, which is far below the 0 dB CNR that Shannon's limit calculates. Is this level achievable? Yes, by reducing filter bandwidth. In space probes and other uses of FM/FM telemetry, a narrow bandwidth filter is used to restore a carrier down to very near that low limit, which would be far below Shannon's limit for FM with a modulation index of 5. Best (1, Chap. 7 and Taub and Schilling (2, p. 324) both calculate the improvements in SNR, which are referred to here as Q and the R factor or R effect. They are integral characteristics of the tracking filter, a device in common use. Shannon's limit should not be applied to FM, or the SNR limit either for that matter. The conditions are too indefinite.

Q and R are based on bandwidth, while N and W in the Shannon's limit equation are based on rates.

Carson's rule would also appear to be based on the Nyquist bandwidth, but a single frequency can be transmitted SSB-SC in a much narrower bandwidth than the Nyquist bandwidth if a coherent carrier is reconstructed and inserted in the product detector. Only 1 Hz of bandwidth is needed to pass a 1 kHz tone SSB-SC. It is not necessary to use all of Carson's BW or the Nyquist BW. Such a tone with a narrow band filter, such as those used with MFSK, would be detectable far below Shannon's limit.

A substitute for Shannon's limit (the SNR limit) can be obtained from the universal equation. Simply equate E_b/η to 1 and the resulting value for SNR is the SNR limit, which is approximately equal to Shannon's limit for the higher values of N. It does not apply to FM where there is an FM knee. (See Fig. 2.)

For example:

$$1024\,QAM \text{ has } N = 10 \quad \text{and} \quad \beta = 1/(32 - 1)$$
$$\text{SNR} = (1/1000)\,NE_b/\eta$$

With $E_b/\eta = 1$, SNR becomes 1/100, or −20 dB, which is Shannon's limit for 1024 QAM. From this equation it can be determined that CNR must be 10 times greater than E_b/η (10 dB). To obtain CNR, multiply the E_b/η value by N. The E_b/η value for 10^{-6} bit error rate (BER) is 10.5 dB above Shannon's limit (30 dB), and the CNR level required is 10 dB above the E_b/η value (40 dB).

NOTES ON REFERENCES

The basic equation is from Bellamy [4, Eq. (4.6), p. 193). The equation has been expanded to include all applicable factors. The values for the error angles and error distances P are from Refs. 2 and 4. 3/2 for the value of O is from Ref. 4 and the literature in general. Q and R are obvious as bandwidth ratios, but are calculated by Best (1, p. 59) in a single equation. Shannon's limit is in the form used in Ref. 4. The constellations are from Refs. 2–8 plus other sources. Confirmation of the SNR/CNR ratios calculated here is found in all of the references. A more extensive explanation of VMSK is given in Ref. 10. Feher in Ref. 5 uses the equation SNR = β^2 (Bit Rate/Bandwidth)E_b/η. It is to be noted that the (bit rate/bandwidth) term is the same as QR in the universal equation SNR = β^2QR(E_b/η).

BIBLIOGRAPHY

1. R. E. Best, *Phase Locked Loops,* New York: McGraw-Hill, 1984.
2. H. Taub and D. Schilling, *Principles of Communications Systems,* New York: McGraw-Hill, 1986.
3. C. S. Palmer and K. Feher, Performance of pi/4 SQAM in a hard limited channel, *IEEE Trans. Broadcast.,* **39**: 301–306, 1993.
4. J. C. Bellamy, *Digital Telephony,* New York: Wiley, 1982.
5. K. Feher, *Wireless Digital Communications: Modulation and Spread Spectrum Applications,* Upper Saddle River, NJ: Prentice-Hall, 1995.
6. Y. Guo and K. Feher, A spectrally efficient constant envelope modulation technique for PCS and mobile communications, *IEEE Trans. Broadcast.,* **40**: 177–183, 1994.
7. J. Borowski and K. Feher, Nonobvious correlation properties of quadrature GMSK, *IEEE Trans. Broadcast.,* **41**: 69–75, 1995.
8. D. Douros and K. Venema, *Philips Semiconductors Wireless Communications SourceBook,* 1996.
9. B. Stryzack and H. R. Walker, Improve data transmission using single sideband FM with suppressed carrier, *RF Microw. Mag.,* Nov. 1994 (Wireless Systems Designs Supplement).
10. C. S. Koukourlis et al., Spectrally efficient biphase modulation, *Appl. Microw. Wireless,* May 1998.

H. R. WALKER
Pegasus Data Systems

MODULATION BY ELECTROABSORPTION. See ELECTROABSORPTION.