Applied Technology Institute (ATIcourses.com)
Stay Current In Your Field • Broaden Your Knowledge • Increase Productivity
349 Berkshire Drive • Riva, Maryland 21140
888-501-2100 • 410-956-8805
Website: www.ATIcourses.com • Email: ATI@ATIcourses.com

Boost Your Skills With ATIcourses.com!

ATI Provides Training In:

- Acoustic, Noise & Sonar Engineering
- Communications and Networking
- Engineering & Data Analysis
- Information Technology
- Radar, Missiles & Combat Systems
- Remote Sensing
- Signal Processing
- Space, Satellite & Aerospace Engineering
- Systems Engineering & Professional Development

Check Our Schedule & Register Today!

The Applied Technology Institute (ATIcourses.com) specializes in training programs for technical professionals. Our courses keep you current in state-of-the-art technology that is essential to keep your company on the cutting edge in today's highly competitive marketplace. Since 1984, ATI has earned the trust of training departments nationwide, and has presented On-site training at the major Navy, Air Force and NASA centers, and for a large number of contractors. Our training increases effectiveness and productivity.

Learn From The Proven Best!

Modulation, Power, and Bandwidth

Tradeoffs in Communication Systems Design

by Robert A. Nelson

Modulation is the process by which information is conveyed by means of an electromagnetic wave. The information is impressed on a sinusoidal carrier wave by varying its amplitude, frequency, or phase. Methods of modulation may be either analog or digital.

The power and bandwidth necessary for the transmission of a signal with a given level of quality depends on the method of modulation. There is a classic tradeoff between power and bandwidth that is fundamental to the efficient design of communication systems. This article will identify various methods of analog and digital modulation, describe their characteristics, and analyze their advantages and disadvantages. The scope of the discussion will be restricted to certain common types of modulation systems.

TYPES OF MODULATION

The carrier wave can be represented by the cosine function

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

A sinusoidal carrier wave thus has two fundamental properties: amplitude A and angle θ. Either of these parameters can be varied with time t to transmit information. Frequency and phase modulation are special cases of angle modulation.

In analog modulation the amplitude, frequency, or phase can take on a continuous range of values. The modulated parameter must faithfully follow all of the inflections of the signal to be transmitted. Any variation in this parameter due to propagation losses or interference will result in a distortion of the received demodulated signal.

The principal forms of analog modulation are amplitude modulation (AM) and frequency modulation (FM). These methods are familiar from their application to terrestrial broadcast radio and television.

In digital communication, the modulated parameter takes on only a discrete set of values, each of which represents a symbol. The symbol consists of one or more bits, or binary ones and zeroes. Since the demodulator must merely identify which amplitude, frequency, or phase state is most closely represented in the received signal during each symbol period, the signal can be regenerated without any distortion. Error correction coding is used to reduce bit transition errors caused by interference to meet a specified performance objective.

Two common forms of digital modulation used in satellite communication are phase shift keying (PSK), in which the carrier phase takes on one of a set of discrete values, and frequency shift keying (FSK), in which the frequency may have one of two or more discrete values.

FOURIER PRINCIPLE

A method of representing a time varying function in terms of an infinite trigonometric series was introduced by the eighteenth century French mathematician and physicist Jean Baptiste Fourier (1768 – 1830). According to the Fourier principle, an arbitrary periodic function defined over a specified interval can be represented as the sum of an infinite number of sine and cosine functions whose frequencies are integral multiples of the repetition rate, or fundamental frequency, and whose amplitudes depend on the given function. The frequencies above the fundamental frequency are called the harmonics. The frequency characteristics of a periodic function are determined by the amplitudes of the admixture of harmonics. To a communications engineer, the Fourier principle provides a method of understanding a complicated signal waveform in terms of the amplitudes of the individual harmonics.

For example, the musical sounds produced by a piano, trumpet, or clarinet all performing the tone of concert A (440 Hz) are distinguished by the harmonics that they produce. The fundamental frequency is 440 Hz, but the instruments sound different because they each produce a different set of harmonics. For high fidelity reproduction of these sounds, the range of frequencies should be as high as possible. In the case of the human ear, the frequency range is approximately between 50 Hz and 20,000 Hz. If this range is truncated by the limitations of the recording and reproduction equipment, then the original sound will appear to be distorted and will be easily detected as artificial.

In a typical toll-quality telephone channel, the bandwidth is about 4,000 Hz. This bandwidth is considered to be adequate for the transmission of clear speech. However, since all of the frequencies above 4,000 Hz are filtered out, certain subtle distinctions between similar sounds are lost. That is why, for example, the sounds for *m* and *n* or for *f* and *s* are easily confused over the telephone and we often find it necessary to use phonetics when spelling out a name, even though they are easily distinguished unconscientiously from their higher harmonics when spoken in person.

A mathematical generalization of a Fourier series is the Fourier transform. The Fourier transform permits the conversion of any continuous function in the time domain to a corresponding function in the frequency domain and vice versa. However, the Fourier transform and its inverse involve the use of complex variables. Thus to completely represent the spectrum of a time-dependent function, it is necessary to use the mathematical fiction of both positive and negative frequencies. Using the Fourier transform, one can analyze the frequency spectral content of any time-dependent signal. By a powerful mathematical theorem known as the Wiener-Khintchine Theorem, the power spectral density of a given function of time is the Fourier transform of its autocorrelation function.

FREQUENCY REGIMES

There are three frequency regimes that are involved in the transmission of a signal. These are the baseband frequencies, the intermediate frequency (IF) band, and the radio frequency (RF) band.

The baseband signals are the signals that carry the information, such as from a telephone, microphone, or video camera. The baseband is the range of frequencies generated by the original source of information. For sound, these frequencies are typically from 0 to a few kilohertz. For video, they may extend to a few megahertz.

The intermediate frequencies are the frequencies present in the signal that are produced after modulation and filtering.

The radio frequency band is the range of frequencies that are transmitted through space. The modulated signal is converted from the intermediate frequency regime to the radio frequency regime by frequency translation. The RF frequencies typically range from a few hundred to a few thousand kilohertz for terrestrial broadcasting and from 1 to 30 gigahertz for satellite communication. These satellite frequencies are in the microwave region, corresponding to wavelengths on the order of a few centimeters, and permit the use of antennas with reasonably sized physical dimensions.

AMPLITUDE MODULATION

With analog amplitude modulation (AM), the message signal $m(t)$ is used to modify the amplitude of the carrier wave. For 100 percent modulation, the amplitude becomes the time-dependent function

$$A(t) = A\,[\,1 + m(t)\,]$$

The angle is given by $\theta = \omega_c\, t + \phi$. The carrier angular frequency ω_c and phase ϕ (which can be taken to be zero) remain constant. Thus the transmitted signal assumes the mathematical form

$$\begin{aligned} s(t) &= A\,[\,1 + m(t)\,] \cos(\omega_c\, t) \\ &= A \cos(\omega_c\, t) + A\, m(t) \cos(\omega_c\, t) \end{aligned}$$

The carrier angular frequency ω_c is related to the frequency f_c by the relation $\omega_c = 2\pi\ f_c$, where ω_c is expressed in radians per second and f_c is expressed in hertz. Multiplication of the cosine function, which is generated in the local oscillator circuit of the modulator, by the message signal produces a spectrum that consists of two sidebands in addition to the frequency of the carrier.

By the Fourier principle, the message signal can be analyzed in terms of its individual sinusoidal components. Thus if the local oscillator generates a carrier $\cos(\omega_c\, t)$ at the intermediate frequency ω_c and it is modulated by one of the components of the message signal represented by $m(t) = \cos(\omega_m\ t)$ at frequency ω_m, then by a trigonometric identity the resulting waveform will be

$$\begin{aligned} \cos(\omega_c\, t) \cos(\omega\, t) &= \tfrac{1}{2} \cos(\omega_c + \omega_m)\, t \\ &\quad + \tfrac{1}{2} \cos(\omega_c - \omega_m)\, t \end{aligned}$$

The spectrum thus contains the two frequencies $\omega_c + \omega_m$ and $\omega_c - \omega_m$. For example, if the local oscillator generated cosine function at 64 kHz is multiplied by the original baseband signal comprising the set of the four frequencies 1, 2, 3, and 4 kHz, then the resulting spectrum would comprise the frequencies 65, 66, 67, and 68 kHz in the upper sideband and the frequencies 60, 61, 62, and 63 kHz in the lower sideband. Therefore, when the cosine function is multiplied by the message signal, two things happen: the frequencies are translated and the bandwidth is doubled.

In the type of amplitude modulation known as double sideband full carrier (DSB-FC) amplitude modulation, the modulated signal consists of the carrier wave with a time varying amplitude that forms an envelope. The spectrum consists of the carrier frequency, the upper sideband, and the lower sideband. The signal can be easily demodulated simply by passing the modulated signal through a filter to remove the high frequency components contributed by the carrier, leaving the low frequency components of the envelope representing the desired signal.

The transmitted power consists of the carrier power and the power in the sidebands. For 100 percent modulation by a sinusoidal message component, the power in the two sidebands together is one-half the power in the carrier. That is, the total power is three times the power in the sidebands. The sideband power is evenly divided between the two sidebands, giving them each one-fourth the carrier power. For example, full modulation of a 100 watt sinusoidal carrier will add 50 watts to the sidebands, with 25 watts in each sideband, resulting in a total transmitted power of 150 watts.

Since the carrier conveys no information while each sideband contains the same information, this form of modulation is wasteful in both power and bandwidth. The advantage is that only envelope detection is needed to demodulate the signal and the receiver can be built easily and inexpensively. The recovery circuit may be as simple as a diode followed by a low pass filter consisting of a resistor and capacitor in parallel. In US commercial AM radio, the baseband is filtered to 5 kHz and thus the bandwidth per channel is 10 kHz. The AM band extends from 535 kHz to 1705 kHz and the carriers are centered at 540 kHz to 1700 kHz in 10 kHz steps.

In double sideband supressed carrier (DSB-SC) amplitude modulation, both sidebands are transmitted but the carrier is removed. The bandwidth is twice the bandwidth of the baseband signal.

In single sideband (SSB) amplitude modulation the signal is generated by a balanced modulator and filter and the transmitted frequencies consist only of a single sideband. The bandwidth is therefore the same as that of the baseband signal. This method requires only one half the bandwidth as DSB-FC amplitude modulation while transmitting only a fraction of the power.

Envelope detection is not possible in either DSB-SC or SSB. Therefore, the receiver must recover the frequency and phase of the transmitter and is more complex and costly. In DSB-SC a small phase error causes a variation in amplitude, whereas in SSB it affects both amplitude and phase. SSB is thus well suited for voice communication, since the human ear is relatively insensitive to phase distortion, but it is not well adapted to other signals, such as video or digital. It is used in marine and citizens band radio. Before they were replaced by digital circuits, analog telephone channels were combined by frequency division multiplexing using SSB modulation.

FREQUENCY MODULATION

In analog frequency modulation (FM), the message signal is used to vary the frequency of the carrier. The deviation of the instantaneous frequency is directly proportional to the message signal. The amplitude of the carrier remains constant. The range of values of the frequency about the carrier center frequency is called the peak deviation Δf. The instantaneous angular frequency is

$$\omega(t) = d\theta/dt = \omega_c + \Delta\omega\, m(t)$$

where $\Delta\omega = 2\pi\, \Delta f$. For modulation by a sinusoid at the single frequency f_m, the message signal is $m(t) = \cos(\omega_m t)$, where $\omega_m = 2\pi\ f_m$. Then $\theta = \omega_c\, t + \beta \sin \omega_m\, t$ and the signal has the mathematical form

$$s(t) = A \cos(\omega_c\, t + \beta \sin \omega_m\, t)$$

where $\beta = \Delta f / f_m$. The parameter β, which is ratio of the peak deviation to the baseband modulation frequency, is a key property called the modulation index.

This expression for $s(t)$ can be expanded into an infinite series of discrete components involving the Bessel function of integral orders, which characteristically occur in mathematical physics when trigonometric functions of trigonometric functions are involved. The resulting spectrum is a distribution of "spikes." (The logo for Cisco Systems is based on this pattern and is also intended to resemble San Francisco's Golden Gate Bridge.) The amplitudes are determined only by the modulation index and become more uniform as the modulation index increases, e.g., for $\beta > 10$. For example, the telemetry, tracking, and control (TT&C) subsystem of a satellite generally uses FM with high modulation index to transmit three tones representing a binary one or zero and execute.

In general, the FM spectrum is a complex function of β consisting of multiple sidebands that occur at integral multiples of the modulating frequency on either side of the carrier rather than, as in AM, consisting of a single pair of sideband frequencies. The spectrum can be analyzed mathematically only in the simplest cases since FM is inherently nonlinear and superposition of individual source signals is not applicable

In principle, the required bandwidth is infinite, but in practice it is given approximately by Carson's rule,

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

where f_m is the highest baseband frequency. This well known empirical estimate for determining the practical bandwidth of FM was first suggested in an unpublished memorandum in 1939 by John Renshaw Carson, chief theoretical mathematician at Bell Laboratories. For example, in US commercial FM monaural radio, the highest baseband frequency is 15 kHz and the peak deviation is 75 kHz. Thus the modulation index β is 5 and the required bandwidth B is 180 kHz. Allowing for a 10 kHz guardband on each side, the channel bandwidth is 200 kHz. There are 100 channels, each 200 kHz wide, in the FM band from 88 MHz to 108 MHz.

Two characteristics of FM that are familiar to radio listeners are that the signal quality is much better than AM and that the signal drops out rapidly beyond the nominal range of the transmitter.

The better performance is due to the fact that the signal to noise ratio at the demodulator output is higher for wideband FM than for AM. It was Edwin Howard Armstrong who first recognized the noise-reducing potential of FM for radio broadcasting in the early 1930s. On theoretical grounds Carson had correctly rejected narrowband FM as inferior to AM for the reduction of noise, since he was principally interested in reducing the bandwidth of telephone circuits and hence increasing the system capacity. On the other hand, through experimental measurements Armstrong found that by *widening* the bandwidth the signal to noise ratio could be increased dramatically for radio. He designed and demonstrated the first FM radio circuits.

For a single-frequency sinusoidally modulated signal, the FM output signal to noise ratio at baseband S_b/N_b may be expressed

$$\begin{aligned} S_b/N_b &= 3\,\beta^2\,(B / 2\,f_m)\,(S/N) \\ &= 3\,\beta^2\,(\beta + 1)\,(S/N) \end{aligned}$$

where S/N is the input signal to noise ratio in the RF channel. Thus after demodulation the output signal to noise ratio is greater than the input signal to noise ratio by the factor $3\,\beta^2\,(\beta + 1)$. In contrast, when the same sideband power is transmitted, the output signal to noise ratio is *equal* to the input signal to noise ratio for all types of amplitude modulation. The FM noise density is $N_0 = N/B$. For a double-sideband AM system with the same noise density, the input noise power is $N' = 2\,f_m\,N_0$. If also the AM input signal power is $S' = S$, then

$$S_b/N_b = 3\,\beta^2\,S'/N'$$

Thus after demodulation the FM signal to noise ratio is greater than the corresponding AM signal to noise ratio by a factor of $3\,\beta^2$. This factor is called the "FM improvement."

From the theoretical relation above, it is seen that as long as $\beta > 0.6$, FM delivers better performance than AM for equal signal power and equal noise power density. However, the FM bandwidth is expanded to $2\,(\beta + 1)$ times the information bandwidth f_m , whereas the AM bandwidth is $2\,f_m$. This is a classic example of trading bandwidth for power. For example, when $\beta = 5$ the FM output signal to noise ratio is 75 times that of an equivalent AM system (19 dB higher), but the bandwidth is 6 times larger. Therefore, the modulation index must be sufficiently high that it provides the desired FM improvement, but it is limited by the need to preserve bandwidth through Carson's rule.

In addition, below a certain threshold input signal to noise ratio that increases somewhat with increasing β, the demodulated signal to noise ratio falls off precipitously. This property is why the range of an FM signal is limited. The existence of a threshold is characteristic of any system that reduces noise in exchange for extra bandwidth and becomes pronounced when the reduction is large. For wideband FM the threshold occurs at roughly 10 dB.

With analog frequency modulation the instantaneous frequency varies directly as the message signal and the phase varies as the integral of the message signal. Analog phase modulation is a closely related form of angle modulation where it is the phase that varies directly as the message signal and where the frequency varies directly as the derivative of the message signal.

TELEVISION

In the United States, the broadcast television standard is the NTSC (National Television System Committee) system. The video signal is modulated by a form of amplitude modulation called vestigial sideband (VSB) amplitude modulation, in which a portion of the lower sideband is transmitted with the upper sideband. VSB AM requires less bandwidth than DSB-SC, overcomes the problem of phase distortion present in SSB and the difficulty of filtering the low frequency content, and permits simple envelope detection. The highest luminance baseband frequency is 4.2 MHz. The upper sideband of the video signal has a bandwidth of 4.2 MHz, while the vestigial lower sideband has a bandwidth of 1.25 MHz. The color signal is transmitted on a separate subcarrier interlaced in the frequency domain with the luminance signal. The audio signal uses frequency modulation, with a highest baseband frequency of 10 kHz and a frequency deviation of 25 kHz. Thus the audio bandwidth is 70 kHz and is centered 4.5 MHz above the video carrier. The total bandwidth for both the video and audio signals is 6.0 MHz.

For the transmission of a television signal over a satellite, amplitude modulation would be severely affected by losses, various forms of interference, and nonlinearities in the transponder. Therefore, the video signal is frequency modulated along with the audio signal. The peak frequency deviation of video on the main carrier is 12 MHz and the modulation index is 2.86. By Carson's rule, the required bandwidth is 2 (12 MHz + 4.2 MHz) = 32.4 MHz. Thus a bandwidth of 36 MHz was originally chosen for a satellite transponder so that it could safely accommodate one analog FM television channel. Since the FM television channel occupies the entire transponder bandwidth, the transponder can be operated at full power without any intermodulation interference caused by the nonlinear transfer characteristic.

ANALOG TO DIGITAL CONVERSION

In digital communication, information is transmitted in the form of a continuous string of binary ones and zeroes. Thus it is necessary to convert the analog baseband signal, such as a sound or video recording, to a digital signal.

Pulse code modulation (PCM) is a conventional technique that converts an analog waveform into a sequence of binary numbers. The first step is to establish a set of discrete times at which the input signal is sampled. According to a classic theorem of sampling theory stated by Harry Nyquist of Bell Laboratories in 1933, the minimum sampling rate f_s is twice the highest baseband frequency f_m, or $f_s = 2\, f_m$. The next step is to represent each analog sample value by a binary number. If there are n bits per sample, then there can be $2^n - 1$ possible levels in each sample. The required bit rate is therefore $R_b = n\ f_s = 2\ n\ f_m$. The original signal waveform is recovered by using a low pass filter. The restriction of each sample to a discrete set of values results in a small amount of quantization noise. This encoding/decoding technique is essentially independent of the nature of the analog signal.

For example, in a conventional toll-quality telephone channel, the practical band extends from about 200 Hz to about 3400 Hz. Rounding up to 4,000 Hz, the Nyquist sampling rate is thus 8,000 samples per second. If 8 bits are allocated for each sample, resulting in 255 possible levels per sample, the required bit rate is 8 × 8,000 bits per second, or 64 kbps, which is the basis of the standard bit rate for a telephone channel. In a digital compact disc (CD) audio recording, the sampling rate is 44,100 samples per second to ensure a perceived bandwidth of more than 20 kHz. With 16 bits per sample for each of two separate stereo channels, the audio data rate is 1.411 Mbps.

In terrestrial cell phone and satellite mobile telephony systems, the bit rate can be as low as 2.4 kbps. This significantly lower bit rate is made possible because the voice coder (vocoder) is designed specifically for speech. The vocoder uses a model of the human vocal tract and synthesizes speech, much as a keyboard musical synthesizer can emulate the sounds of various musical instruments. Only a limited set of perceptually important parameters are transmitted, such as vowel sound, pitch, and level, resulting in fewer bits and smaller bandwidth. Although the speech is intelligible, the quality is below telephone standards. Other forms of sound, such as music, cannot be transmitted.

An NTSC digital television signal following the ITU-R Rec. 601 standard has 30 frames per second, 525 lines per frame, 858 samples for luminance and 429 samples for each of two color differences per line (so-called 4:2:2 component structure), and 8 bits per sample. Theoretically, the required bit rate is 216 Mbps. In practice, there are 480 active lines with 720 samples for luminance and 360 samples for each of two color differences per line, yielding 166 Mbps. The luminance sampling rates for these two formats are 13.5 MHz and 10.4 MHz, repectively, compared with the Nyquist sampling rate of 8.4 MHz for analog video. With compression the bit rate can be reduced to about 8 Mbps (MPEG-2 quality) or 1.5 Mbps (MPEG-1 quality).

PULSE SHAPING

The baseband digital symbols must be represented by a continuous string of pulses of some appropriate form. For example, a "1" may be represented by a positive rectangular pulse and a "0" may be represented by a negative rectangular pulse. This type of pulse train is called "Non-Return to Zero" (NRZ) pulse shaping, because the pulse remains at a constant amplitude over each full bit period. Numerous other pulse shapes are also used, in which "notches" are added to improve synchronization. However, since these pulses require greater bandwidth, the NRZ signal format is generally preferred in satellite communication systems.

Since the pulse train transmits information, each successive pulse is independent of those that came before it. Thus the probability of a given pulse representing a one or zero is random, and the sequence of NRZ pulses is a stochastic process. It can be shown that the autocorrelation function for this case is a triangle function. Thus by the Wiener-Khintchine Theorem, the power spectral density (or power per unit bandwidth at frequency f) is the Fourier transform of the triangle function and happens to be a function that has the form of $(\sin x/x)^2$ centered about 0, where $x = \pi f / R_b$.

In practice, the baseband pulse shapes are not nice, perfect rectangles with right angle corners. To produce such pulses, the bandwidth would have to be infinite. Instead, because of the finite bandwidth of the filter, the pulses are actually rounded "blips." The tails of these blips will tend to overlap, causing a phenomenon known as intersymbol interference (ISI).

Nyquist showed that the pulse shape that required the minimum bandwidth without intersymbol interference is one that in the time domain has the form of the function $\sin(\pi R_b t)/(\pi R_b t)$. For this function, the tails of the preceding and following pulses pass through zero at the peak of the present pulse. In the frequency domain, the Fourier transform looks like a rectangular brick wall. The minium required baseband bandwidth is one half the information bit rate, or $b = R_b / 2$.

But it is impossible to realize this pulse shape in an actual filter. Instead a form of pulse shaping called "raised cosine" filtering is used, characterized by a parameter called the rolloff ρ that is between 0 and 1. A typical value of ρ is 0.2. For zero rolloff, the raised cosine pulse shape reduces to the ideal $\sin x/x$ pulse shape. The actual baseband bandwidth is thus $b = k\ R_b\ /\ 2$, where $k = 1 + \rho$.

DIGITAL MODULATION

In digital communication, the carrier to noise density ratio is given by the relation

$$C/N_0 = R_b\ (\ E_b\ /\ N_0\ )$$

where R_b is the information bit rate. The quantity $E_b\ /\ N_0$ is the ratio of the energy per information bit E_b and the total noise density N_0 (noise power per unit bandwidth) and has fundamental importance. The value of $E_b\ /\ N_0$ is determined by three design factors: the bit error rate, the method of modulation, and the method of forward error correction coding.

If the phase is the parameter that is varied, the modulation is called phase shift keying (PSK). Two common forms of digital modulation used for satellite communication are binary phase shift keying (BPSK) and quaternary phase shift keying (QPSK). If the frequency is varied instead of the phase, the modulation is called frequency shift keying (FSK).

In BPSK modulation the carrier can have one of two phase states, 0° and 180°, which represent a binary one or zero. In a BPSK modulator, the baseband pulse train simply multiplies a cosine function generated by a local oscillator, usually at the intermediate frequency of 70 MHz. Multiplication of $\cos\ (\omega_c\ t)$ by a pulse of level + 1 representing a binary 1 leaves the phase of 0 unchanged. On the other hand, multiplication by a pulse of level – 1 representing 0 yields $-\cos\ (\omega_c\ t) = \cos\ (\omega_c\ t + \pi)$, which changes the phase by 180°. Coherent detection is needed for demodulation. In other words, the receiving circuit must recover the absolute phase of the transmitting circuit. This is usually done by either a Costas loop or a squaring loop.

In QPSK modulation the carrier can assume one of four phase states, consisting of 0°, 90°, 180°, and 270°, which represent the symbols 00, 01, 11, and 10. A QPSK modulator is usually thought of as two BPSK modulators that are out of phase by 90°.

As discussed for AM, forming a product with a cosine function results in a spectrum containing sums and differences of the oscillator frequency and each frequency in the baseband signal. Thus with NRZ pulse shaping, the BPSK spectral density consists of *two* $(\sin\ x/x)^2$ functions, one centered at 70 MHz and the other centered at – 70 MHz in the complex domain. The frequencies are thus translated and the bandwidth is doubled.

In general, the required occupied bandwidth for digital modulation, including forward error correction coding, is

$$B = \ k\ (\ R_b\ /\ m\ )(1\ /\ r\ )$$

where R_b is the bit rate, m is the number of bits per symbol, r is the code rate, and k is the bandwidth expansion factor used to minimize intersymbol interference. For example, if R_b = 64 kbps, m = 2 for QPSK modulation, r = 1/2, and k = 1.2, then B = 76.8 kHz.

For a given bit error rate, the value of $E_b\,/\ N_0$ required for transmission of both BPSK and QPSK signals is the same and is less than that required for other forms of digital modulation, such as FSK. Hence for a given information bit rate R_b the power is also the same. However, since each QPSK symbol consists of two bits while each BPSK symbol consists of only one bit, the bandwidth required for QPSK modulation is only half that for BPSK. This is the communications equivalent of a "free lunch." (Actually, the tradeoff is in the increased complexity of the QPSK modulator.) Therefore, until recently, QPSK has been the preferred form of digital modulation in satellite communications.

The trend in power and bandwidth does not continue to higher order forms of PSK modulation. For example, in 8-phase PSK (8PSK), there are three bits per symbol, comprising the set 000, 001, 011, 010, 110, 111, 101, and 100. Therefore, the bandwidth for 8PSK is one third that of BPSK and 2/3 that of QPSK. However, since the phase states are closer together and are harder to distinguish, the power required for 8PSK is higher.

The mapping sequences illustrated for QPSK and 8PSK are examples of Gray encoding, in which two symbols represented by neighboring phases differ by only one bit. This method is most often preferred because an error in the demodulator will likely be caused by choosing an adjacent phase state and thus will result in at most one errored bit.

It is possible to vary more than one parameter. In quadrature amplitude modulation (QAM), both the amplitude and phase are modulated. In 16QAM, for example, there are twelve possible phase states and three possible amplitudes. There are four bits per symbol, e.g., 0000, 0001, 0011, etc., and the required bandwidth is one fourth that of BPSK and one-half that of QPSK. However, the required power is much higher. This form of modulation has been used for computer modems and wireless cable television.

SUMMARY

Modulation may be described as the process by which information is impressed on an electromagnetic carrier wave for transmission from one point to another. This article has reviewed several forms of analog and digital modulation. In the design of a communication system, the choice of modulation is of fundamental importance and always involves a tradeoff between power and bandwidth.

In the past, frequency spectrum was relatively plentiful but the power available on a satellite was limited. A satellite typical of the 1980s had a power of less than 1 kW for a payload of 24 transponders. Today, the equation has been reversed. Spectrum is now scarce but a large spacecraft commonly provides 10 to 15 kW for up to 100 transponders. In addition, faster computer processors enable the use of more complex forward error correction coding techniques at high bit rates. Therefore, more spectrum efficient forms of digital modulation such as 8PSK and 16QAM are becoming more attractive, even though the power requirements are higher. Coupled with powerful coding methods such as concatenated Reed Solomon/Viterbi coding, these methods offer the prospect of enhanced spectral efficiency with virtually error-free digital signal transmission.

Dr. Robert A. Nelson, P.E. is president of Satellite Engineering Research Corporation, a satellite engineering consulting firm in Bethesda, Maryland. He is *Via Satellite*'s Technical Editor.