**Digital Communications: The ABCs Of Ones And Zeroes**

August 04, 2010 12:00 AM http://electronicdesign.com/article/communications/digital\_communications\_the\_abcs\_of\_ones\_and\_zeroes.aspx

Electronic Design

Louis E. Frenzel

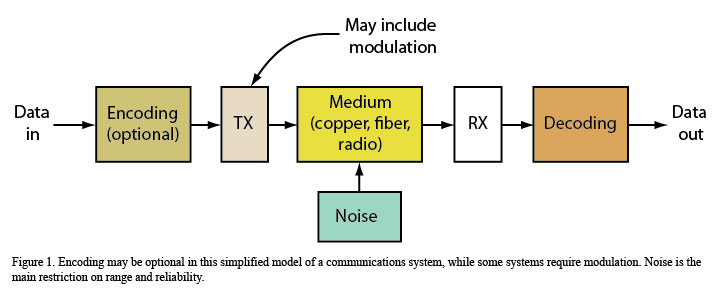
Electronic communications began as digital technology with Samuel Morse’s invention of the telegraph in 1845. The dots and dashes of his famous code were the binary ones and zeroes of the current through the long telegraph wires. Radio communications also started out as digital, with Morse code producing the off and on transmission of continuous-wave spark-gap pulses.

Then analog communications emerged with the telephone and amplitude-modulation (AM) radio, which dominated for decades. Today, analog is slowly fading away, found only in the legacy telephone system; AM and FM radio broadcasting; amateur, CB/family, and shortwave radios; and some lingering two-way mobile radios. Nearly everything else, including TV, has gone digital. Cell phones and Internet communications are digital. Wireless networks are digital.

Though the principles are generally well known, veteran members of the industry may have missed out on digital communications schooling. Becoming familiar with the basics broadens one’s perspective on the steady stream of new communications technologies, products, trends, and issues.

**The Fundamentals**

All communications systems consist of a transmitter (TX), a receiver (RX), and a [transmission medium (Fig. 1](http://electronicdesign.com/Content/15001/60956-fig-1.jpg)).



The TX and RX simply make the information signals to be transmitted compatible with the medium, which may involve modulation. Some systems use a form of coding to improve reliability. In this article, consider the information to be non-return-to-zero (NRZ) binary data. The medium could be copper cable like unshielded twisted pair (UTP) or coax, fiber-optic cable, or free space for wireless. In all cases, the signal is greatly attenuated by the medium and noise is superimposed. Noise rather than attenuation usually determines if the communications medium is reliable.

Communications falls into one of two categories—baseband or broadband. Baseband is the transmission of data directly over the medium itself, such as sending serial digital data over an RS-485 or I2C link. The original 10-Mbit/s Ethernet was baseband. Broadband implies the use of modulation (and in some cases, multiplexing) techniques. Cable TV and DSL are probably the best examples, but cellular data is also broadband.

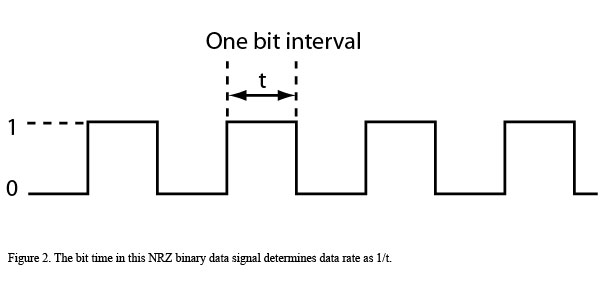
Communications may also be synchronous or asynchronous. Synchronous data is clocked as in SONET fiber-optical communications, while asynchronous methods use start and stop bits as in RS-232 and a few others.

Furthermore, communications links are simplex, half duplex, or full duplex. Simplex links involve one-way communications, or, simply, broadcasting. Duplex is two-way communications. Half duplex uses alternating TX and RX on the same channel. Full duplex means simultaneous (or at least concurrent) TX and RX, as in any telephone.

Topology is also fundamental. Point-to-point, point-to-multipoint, and multipoint-to-point are common. Networking features buses, rings, and mesh. They all don’t necessarily work for all media.

**Data Rate versus Bandwidth**

 Digital communications sends bits serially—one bit after another. However, you’ll often find multiple serial paths being used, such as four-pair UTP CAT 5e/6 or parallel fiber-optic cables. Multiple-input multiple-output (MIMO) wireless also implements two or more parallel bit streams. In any case, the [basic data speed (Fig. 2](http://electronicdesign.com/Content/15001/60956-fig-2.jpg))



or capacity C is the reciprocal of the bit time (t):

C = 1/t

C is the channel capacity or data rate in bits per second and t is the time for one bit interval. The symbol R for rate is also used to indicate data speed. A signal with a bit time of 100 ns has a data rate of:

C = 1/100 × 10–9 = 10 Mbits/s

The big question is how much bandwidth (B) is needed to pass a binary signal of data rate C. As it turns out, it’s the rise time (tR) of the bit pulse that determines the bandwidth:

B = 0.35/tR

B is the 3-dB bandwidth in megahertz and tR is in microseconds (µs). This formula factors in the effect of Fourier theory. For example, a rise time of 10 ns or 0.01 µs needs a bandwidth of:

B = 0.35/0.01 = 35 MHz

A more precise measure is to use the Shannon-Hartley theorem. Hartley said that the least bandwidth needed for a given data rate in a noise-free channel is just half the data rate or:

B = C/2

Or the maximum possible data rate for a given bandwidth is:

C = 2B

As an example, a 6-MHz bandwidth will allow a data rate up to 12 Mbits/s. Hartley also said that this figure holds for two-level or binary signals. If multiple levels are transmitted, then the data rate can be expressed as:

C = (2B)log2M

M indicates the number of multiple voltage levels or symbols transmitted. Calculating the base 2 logarithm is a real pain, so use the conversion where:

log2N = (3.32)log10N

Here, log10N is just the common log of a number N. Therefore:

C = 2B(3.32)log10N

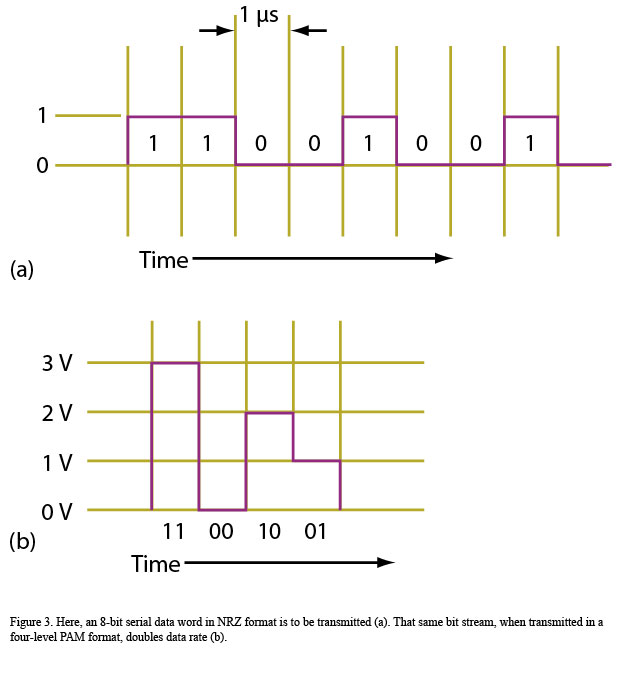
For binary or two-level transmission, the data rate for a bandwidth of 6 MHz is as given above:

C = 2(6)(3.32)log102 = 12 Mbits/s

With four voltage levels, the theoretical maximum data rate in a 6-MHz channel is:

C = 2(6)(3.32)log104 = 24 Mbits/s

To explain this, let’s consider multilevel transmission schemes. Multiple voltage levels can be transmitted over a baseband path in which each level represents two or more bits. Assume we want to transmit the [serial 8-bit byte (Fig. 3a](http://electronicdesign.com/Content/15001/60956-fig-3.jpg)).



Also assume a clock of 1 Mbit/s for a bit period of 1 µs. This will require a minimum bandwidth of:

B = C/2 = 1 Mbit/s/2 = 500 kHz

With four levels, two bits per level can be [transmitted (Fig. 3b](http://electronicdesign.com/Content/15001/60956-fig-3.jpg)). Each level is called a symbol. In this example, the four levels (0, 1, 2, and 3 V) transmit the same byte 11001001. This technique is called pulse amplitude modulation (PAM). The time for each level or symbol is 1 µs, giving a symbol rate—also called the baud rate—of 1 Msymbol/s. Therefore, the baud rate is 1 Mbaud, but the actual bit rate is twice that, or 2 Mbits/s. Note that it takes just half the time to transmit the same amount of data.

What this means is that for a given clock rate, eight bits of data can be transmitted in 8 µs using binary data. With four-level PAM, twice the data, or 16 bits, can be transmitted in the same 8 µs. For a given bandwidth, that translates to the higher data rate equivalent to 4 Mbits/s. Shannon later modified this basic relationship to factor in the signal-to-noise ratio (S/N or SNR):

C = (B)log2(1 + S/N)

or:

C = B(3.32)log10(1 + S/N)

The S/N is a power ratio and is not measured in dB. You will also hear S/N referred to as the carrier-to-noise ratio or C/N. C/N is usually defined as the S/N of a modulated or broadband signal. S/N is used at baseband or after demodulation.With a S/N of 20 dB or 100 to 1, the maximum data rate in a 6-MHz channel will be:

C = 6(3.32)log10(1 + 100) = 40 Mbits/s

With a S/N = 1 or 0 dB, the data rate drops to:

C = 6(3.32)log10(1 + 1) = 6 Mbits/s

This last example is why many engineers use the conservative rule of thumb that the data rate in a channel with noise is roughly equal to the bandwidth C = B.

If the speed through a channel with a good S/N seems to defy physics, that’s because the Shannon-Hartley formulas don’t specifically say that multiple levels or symbols can be used. Consider that:

C = B(3.32) log10(1 + S/N) = 2B(3.32) log10M

Here, M is the number of levels or symbols. Solving for M:

M = √(1 + S/N)

Take the case of a 40-Mbit/s data rate in a 6-MHz channel, if the S/N is 100. This will require multiple levels or symbols:

M = √(1 + 100) = 10

Theoretically, the 40-Mbit/s rate can be achieved with 10 levels.

Incidentally, the levels or symbols could be represented by something other than different voltage levels. They can be different phase shifts or frequencies or some combination of levels, phase shifts, and frequencies. Recall that quadrature amplitude modulation (QAM) is a combination of different voltage levels and phase shifts. QAM, the modulation of choice to achieve high data rates in narrow channels, is used in digital TV as well as wireless standards like HSPA, WiMAX, and Long-Term Evolution (LTE).

**Channel Impairments**

Data experiences many impairments during transmission, especially noise. The calculations of data rate versus bandwidth assume the presence of additive white Gaussian noise (AWGN).

Noise comes from many different sources. For instance, it emanates from thermal agitation, which is most harmful in the front end of a receiver. The sources are resistors and transistors, while other forms of noise come from semiconductors. Intermodulation distortion creates noise. Also, signals produced by mixing in nonlinear circuits create interfering signals that we treat as noise.

Other sources of noise include signals picked up on a cable by capacitive or inductive coupling. Impulse noise from auto ignitions, inductive kicks from motor or relay turn on or off, and power-line spikes are particularly harmful to digital signals. The 60-Hz “hum” induced by power lines is another example. Signals coupled from one pair of conductors to another within the same cable create “crosstalk” noise. In a wireless link, noise can come from the atmosphere (e.g., lightning) or even the stars themselves.

Because noise is usually random, its frequency spectrum is broad. Noise can be reduced by simply filtering to limit the bandwidth. Bandwidth narrowing obviously will affect data rate.

It’s also important to point out that noise in a digital system is treated differently from that in an analog system. The S/N or C/N is used for analog systems, but Eb/N0 usually evaluates digital systems. Eb/N0 is the ratio of the energy per bit to the spectral noise density. It’s typically pronounced as E sub b divided by N sub zero.

Energy Eb is signal power (P) multiplied by bit time t expressed in joules. Since data capacity or rate C (sometimes designated R) is the reciprocal of t, then Eb is P divided by R. N0 is noise power N divided by bandwidth B. Using these definitions, you can see how Eb/N0 is related to S/N:

Eb/N0= S/N (B/R)

Remember, you can also express Eb/N0 and S/N in dB.

The energy per bit is a more appropriate measure of noise in a digital system. That’s because the signal transmitted is usually during a short period, and the energy is averaged over that time. Typically, analog signals are continuous. In any case, Eb/N0 is often determined at the receiver input of a system using modulation. It’s a measure of the noise level and will affect the received bit error rate (BER). Different modulation methods have varying Eb/N0 values and related BERs.

Another common impairment is attenuation. Cable attenuation is a given thanks to resistive losses, filtering effects, and transmission-line mismatches. In wireless systems, signal strength typically follows an attenuation formula proportional to the square of the distance between transmitter and receiver.

Finally, delay distortion is another source of impairment. Signals of different frequencies are delayed by different amounts over the transmission channel, resulting in a distorted signal.

Channel impairments ultimately cause loss of signal and bit transmission errors. Noise is the most common culprit in bit errors. Dropped or changed bits introduce serious transmission errors that may make communications unreliable. As such, the BER is used to indicate the quality of a transmission channel.

BER, which is a direct function of S/N, is just the percentage of the number of error bits to the total transmitted bits over a given time period. It’s usually considered to be the probability of an error occurring in so many bits transmitted. One bit error per 100,000 transmitted is a BER of 10–5. The definition of a “good” BER depends on the application and technology, but the 10–5 to 10–12 range is a common target.

**Error Coding**

Error detection and correction techniques can help mitigate bit errors and improve BER. The simplest form of error detection is to use a parity bit, a check code sum, or cyclical redundancy check (CRC). These are added to the transmitted data. The receiver recreates these codes, compares them, and then identifies errors. If errors occur, an automatic repeat request (ARQ) message is sent back to the transmitter and the corrupted data is retransmitted. Not all systems use ARQ, but ARQ-less systems will typically employ some form of it.

Nonetheless, most modern communications systems go much further by using sophisticated forward error correction (FEC) techniques. Taking advantage of special mathematical encoding, the data to be transmitted is translated into a set of extra bits, which are then added to the transmission. If bit errors occur, the receiver can detect the failed bits and actually correct all or most of them. The result is a significantly improved BER.

Of course, the downsides are the added complexity of the encoding and the extra transmission time needed for the extra bits. This overhead is easily accommodated in more contemporary IC-based communications systems.

The many different types of FEC techniques available today fall into two groups: block codes and convolutional codes. Block codes operate on fixed groups of data bits to be transmitted, with extra coding bits added along the way. The original data may or may not be transmitted depending on the code type. Common block codes include the Hamming, BCH, and Reed-Solomon codes. Reed-Solomon is widely used, as is a newer form of block code called the low-density parity check (LDPC).

Convolutional codes use sophisticated algorithms. Examples include the Viterbi, Golay, and turbo codes. FEC is widely used in wireless and wired networking, cell phones, and storage media such as CDs and DVDs, hard-disk drives, and flash drives.

FEC will enhance the S/N. The BER improves with the use of FEC for a given value of S/N, an effect known as “coding gain.” Coding gain is defined as the difference between the S/N values for the coded and uncoded data streams of a given BER target. For instance, if a system needs 20 dB of S/N to achieve a BER of 10–6 without coding, but only 8 dB S/N when FEC is used, the coding gain is 20 – 8 = 12 dB.

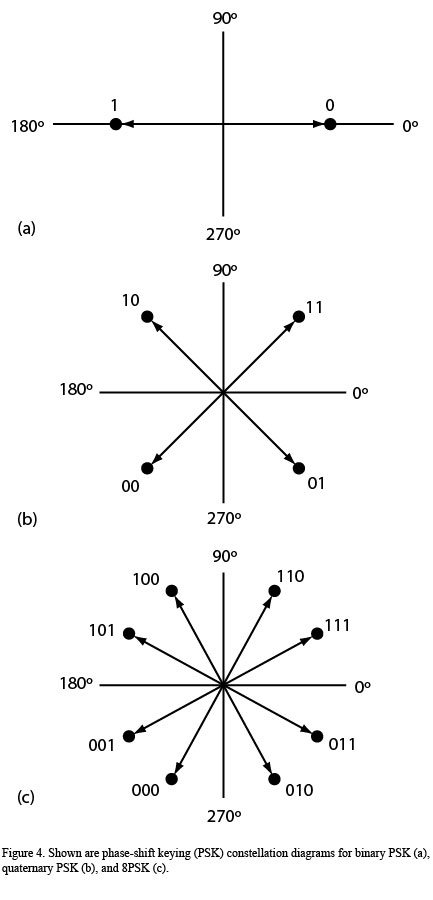
**Modulation**

Almost any modulation scheme may be used to transmit digital data. But in today’s more complex critical applications, the most widely used methods are some form of phase-shift keying (PSK) and QAM. Special modes like spread spectrum and orthogonal frequency division multiplexing (OFDM) are especially well adopted in the wireless space.

Amplitude-shift keying (ASK) and on-off keying (OOK) are generated by turning the carrier off and on or by shifting it between two carrier levels. Both are used for simple and less critical applications. Since they’re susceptible to noise, the range must be short and the signal strength high to obtain a decent BER.

Frequency-shift keying (FSK), which is very good in noisy applications, has several widely used variations. For instance, minimum-shift keying (MSK) and Gaussian-filtered FSK are the basis for the GSM cell-phone system. These methods filter the binary pulses to limit their bandwidth and thereby reduce the sideband range. They also use coherent carriers that have no zero-crossing glitches; the carrier is continuous. In addition, a multi-frequency FSK system provides multiple symbols to boost data rate in a given bandwidth. In most applications, PSK is the most widely used.

Binary phase-shift keying (BPSK) is another popular scheme. Plain-old BPSK is a favorite in which the 0 and 1 bits shift the carrier phase 180°. BPSK is best illustrated in a [constellation diagram (Fig. 4a](http://electronicdesign.com/Content/15001/60956-fig-4.jpg)).



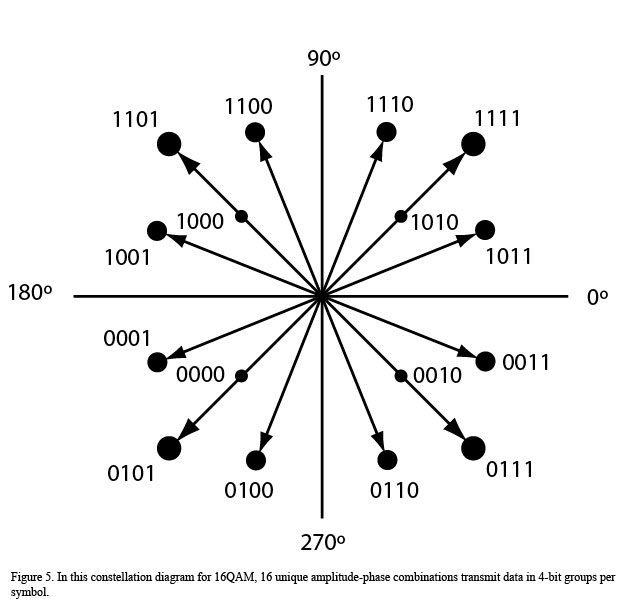
It shows an axis where each phasor represents the amplitude of the carrier and the direction represents the phase position of the carrier.

Quaternary, 4-ary, or quadrature PSK (QPSK) uses sine and cosine waves in four combinations to produce four different [symbols shifted 90° apart (Fig. 4b](http://electronicdesign.com/Content/15001/60956-fig-4.jpg)). It doubles the data rate in a given bandwidth but is very tolerant of noise.

Beyond QPSK is what’s called M-ary PSK or M-PSK. It uses many phases like 8PSK and 16PSK to produce eight or 16 unique phase shifts of the carrier, allowing for very high data rates in a [narrow bandwidth (Fig. 4c](http://electronicdesign.com/Content/15001/60956-fig-4.jpg)). For instance, 8PSK allows transmission of three bits per phase symbol, theoretically tripling the data rate in a given bandwidth.

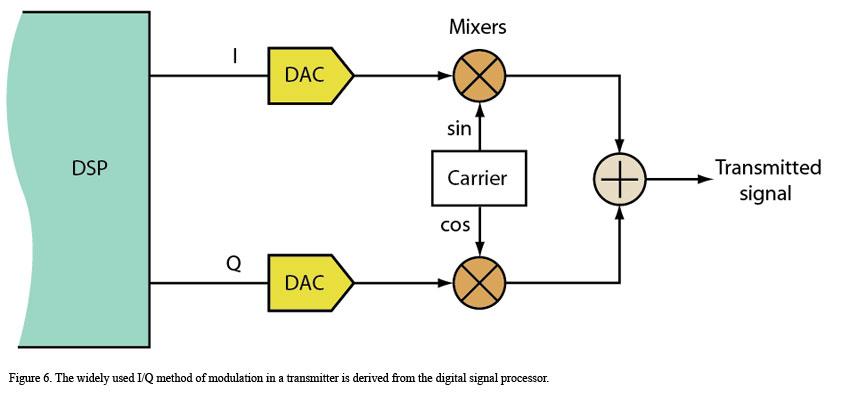
The ultimate multilevel scheme is QAM, which uses a mix of different amplitudes and phase shifts to define as many as 64 to 1024 or more different symbols. Thus, it reigns as the champion of getting high data rates in small bandwidths.

When using 16QAM, for example, each 4-bit group is represented by a phasor of a [specific amplitude and phase angle (Fig. 5](http://electronicdesign.com/Content/15001/60956-fig-5.jpg)).



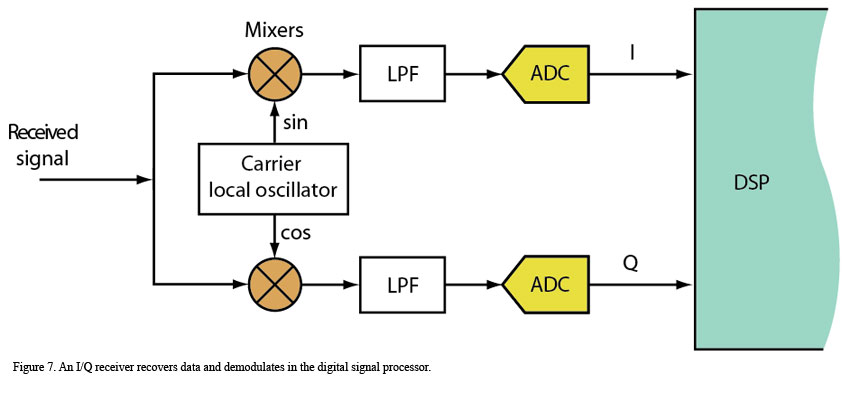
With 16 possible symbols, four bits can be transmitted per baud or symbol period. That effectively multiplies the data rate by four for a given bandwidth.

Today, most digital modulation and demodulation employs digital signal processing (DSP). The data is first encoded and then sent to the digital signal processor, whose software produces the correct bit streams. The bit streams are encoded in an I/Q or in-phase and quadrature format using a [mixer arrangement (Fig. 6](http://electronicdesign.com/Content/15001/60956-fig-6.jpg)).



Subsequently, the I/Q data is translated into analog signals by the digital-to-analog converters (DACs) and sent to the mixers, where it’s mixed with the carrier or some IF sine and cosine waves. The resulting signals are summed to create the analog RF output. Further frequency translation may be needed. The bottom line is that virtually any form of modulation may be produced this way, as long as you have the right DSP code. (Forms of PSK and QAM are the most common.)

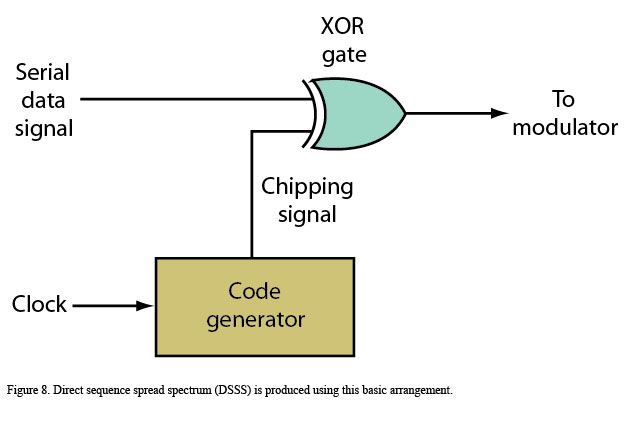
At the receiver, the antenna signal is amplified, downconverted, and sent to an [I/Q demodulator (Fig. 7](http://electronicdesign.com/Content/15001/60956-fig-7.jpg)).



The signal is mixed with the sine and cosine waves, then filtered to create the I and Q signals. These signals are digitized in analog-to-digital converters (ADCs) and sent to a digital signal processor for the final demodulation.

Most radio architectures use this I/Q scheme and DSP. It’s generally referred to as software-defined radio (SDR). The DSP software manages the modulation, demodulation, and other processing of the signal, including some filtering.

As mentioned earlier, two modulation schemes of special interest are spread spectrum and OFDM. These broadband wide-bandwidth schemes are also forms of multiplexing or multiple access. Spread spectrum, which is employed in many cell phones, allows multiple users to share a common bandwidth. It’s referred to as code division multiple access (CDMA). OFDM also uses a wide bandwidth to enable multiple users to access the same wide channel.



[Figure 8](http://electronicdesign.com/Content/15001/60956-fig-8.jpg) shows how the digitized serial voice, video, or other data is modified to produce spread spectrum. In this scheme, called direct sequence spread spectrum (DSSS), the serial data is sent to an exclusive OR gate along with a much higher chipping signal. The chipping signal is coded so it’s recognized at the receiver. Consequently, the narrow-band digital data (several kilohertz) is converted to a wider bandwidth signal that occupies a wide channel. In cell-phone cdma2000 systems, the channel bandwidth is 1.25 MHz and the chipping signal is 1.288 Mbits/s. Therefore, the data signal is spread over the entire band.

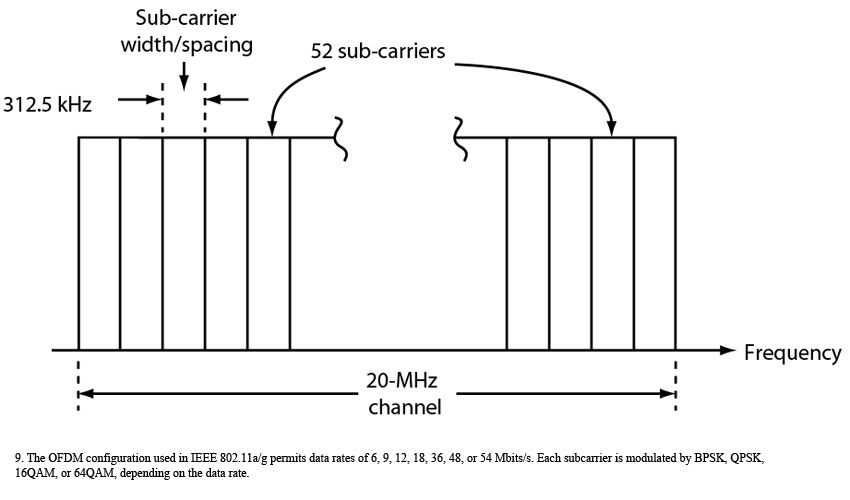
Spread spectrum can also be achieved with a frequency-hopping scheme called FHSS. In this configuration, the data is transmitted in hopping periods over different randomly selected frequencies, spreading the information over a wide spectrum. The receiver, knowing the hop pattern and rate, can reconstruct the data and demodulate it. The most commonly used example of FHSS is Bluetooth wireless.

Other data signals are processed the same way and transmitted in the same channel. Because each data signal is uniquely encoded by a special chipping-signal code, all of the signals are scrambled and pseudorandom in nature. They overlay one another in the channel. A receiver hears only a low noise level. Special correlators and decoders in the receiver can pick out the desired signal and demodulate it.

In OFDM, the high-speed serial data stream gets divided into multiple slower parallel data streams. Each stream modulates a very narrow sub-channel in the main channel. BPSK, QPSK, or different levels of QAM are used, depending on the desired data rate and the application’s reliability requirements.

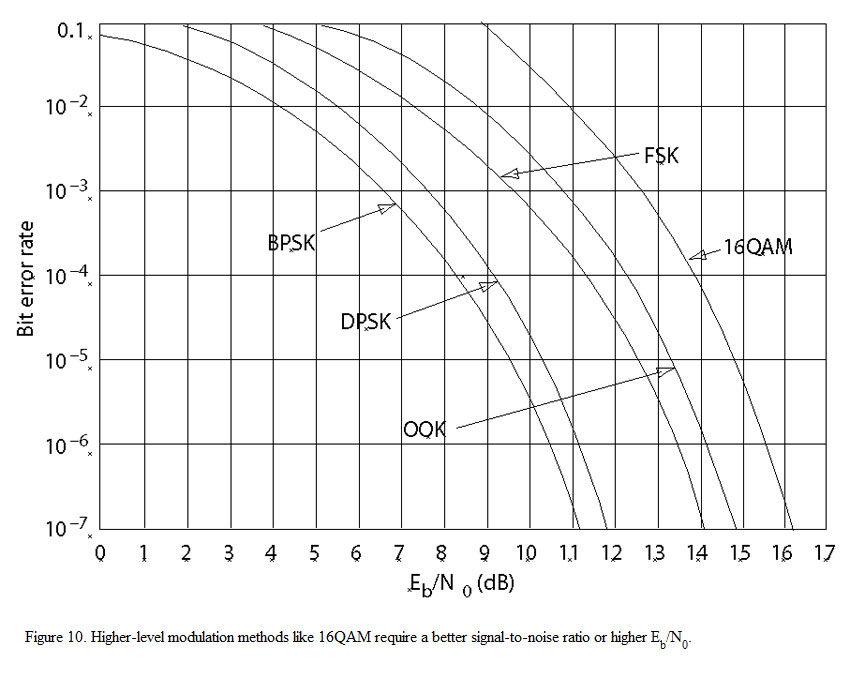
Multiple adjacent sub-channels are designed to be orthogonal to one another. Therefore, the data on one sub-channel doesn’t produce inter-symbol interference with an adjacent channel. The result is a high-speed data signal that’s spread over a wider bandwidth as multiple, parallel slower streams.

The number of sub-channels varies with each OFDM system, from 52 in Wi-Fi radios to as many as 1024 in cell-phone systems like LTE and wireless broadband systems such as WiMAX. With so many channels, it’s possible to divide the sub-channels into groups. Each group would transmit one voice or other data signal, allowing multiple uses to share the assigned bandwidth. Typical channel widths are 5, 10, and 20 MHz. To illustrate, the popular 802.11a/g Wi-Fi system uses an OFDM scheme to transmit data rates to 54 Mbits/s in a [20-MHz channel (Fig. 9](http://electronicdesign.com/Content/15001/60956-fig-9.jpg)).



All new cell-phone and wireless broadband systems use OFDM because of its high-speed capabilities and reliable communications qualities. Broadband DSL is OFDM, as are most power-line technologies. Implementing OFDM can be difficult to implement, which is where DSP steps in.

As indicated earlier, modulation methods vary in the amount of data they can transmit in a given bandwidth and how much noise they can withstand. One measure of this is the BER per given [Eb/N0 ratio (Fig. 10](http://electronicdesign.com/Content/15001/60956-fig-10.jpg)).



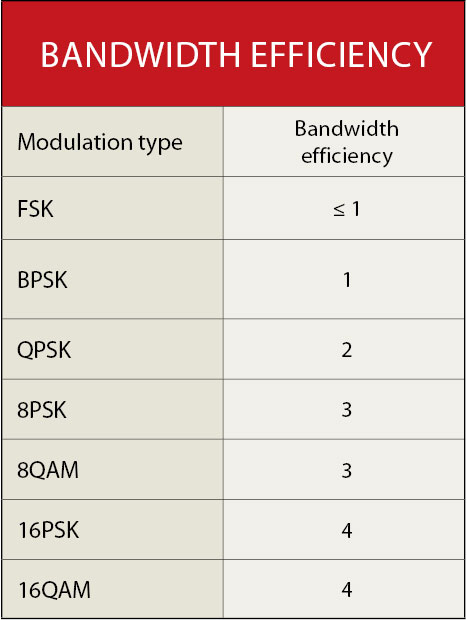
Simpler modulation schemes like BPSK and QPSK produce a lower BER for a low Eb/N0, making them more reliable in critical applications. However, different levels of QAM produce higher data rates in the same bandwidth, although a higher Eb/N0 is needed for a given BER. Again, the tradeoff is data rate against BER in a given bandwidth.

**Spectral Efficiency**

Spectral efficiency is a measure of how many bits can be transmitted at a given rate over a fixed bandwidth. It’s one way to compare the effectiveness of modulation methods. Spectral efficiency is stated in terms of bits per second per hertz of bandwidth, or (bits/s)/Hz. Though the measure usually excludes any FEC coding, it’s sometimes useful to include FEC in a comparison.

Remember 56k dial-up modems? They achieved an amazing 56 kbits/s in a 4-kHz telephone channel, and their spectral efficiency was 14 (bits/s)/Hz. Maximum throughput for an 802.11g Wi-Fi radio is 54 Mbits/s in a 20-MHz channel for a spectral efficiency of 2.7 (bits/s)/Hz. A standard, digital GSM cell phone does 104 kbits/s in a 200-kHz channel, making the spectral efficiency 0.53 (bits/s)/Hz. Add EDGE modulation and that jumps to 1.93 (bits/s)/Hz. And taking it to new levels, the forthcoming LTE cell phones will have a spectral efficiency of 16.32 (bits/s)/Hz in a 20-MHz channel.

Spectral efficiency shows just how much data can be crammed into a narrow bandwidth with different modulation methods.



The [table](http://electronicdesign.com/Content/15001/60956-table-1.jpg) compares the relative efficiencies of different modulation methods, where bandwidth efficiency is just data rate divided by bandwidth or C/B.

**Data Compression**

Data compression offers another way to transmit more data in a given bandwidth. Various mathematical algorithms are applied to reduce the original data to a fewer number of bits. This speeds the transmission and minimizes storage requirements. Reversing the algorithm at the receiver allows for data recovery.

Compression schemes can produce compression ratios of up to several hundred to one. These include the voice-compression schemes used in MP3 players, cell phones, Voice over Internet Protocol (VoIP) phones, and digital radios. Video also employs compression rather extensively. MPEG2 is used in digital TV, while MPEG4 and H.264 standards are found in mobile video and video-surveillance systems.

Copyright © 2010 Penton Media, Inc., All rights reserved.