Signal Encoding Techniques

The notes in this document are based almost entirely on Chapter 5 of the textbook [Sta05]. Rappaport's text is also a good reference for wireless signal propagation [Rap95].

1 Digital modulation: Digital data on analog signals

There are two important factors that govern the effectiveness of a digital modulation technique: *power efficiency* and *bandwidth efficiency* (also called spectral efficiency). The power efficiency of a digital modulation technique is measured by the E_b/N_0 ratio required at the receiver for a certain bit error rate. The bandwidth efficiency is measured by the ratio of the information rate to the bandwidth. There is a natural tradeoff between the two efficiency parameters. A modulation technique with superior error control has higher power efficiency, but lower bandwidth efficiency since the different data symbols are encoded by signals separated in the signal space.

In addition to the above two factors, digital modulation techniques need to be simple and cheap to implement, especially in the context of wireless communications since the vast majority of the wireless devices (e.g., cellphones, PDAs, wireless sensor nodes) are heavily resource-constrained.

1.1 Amplitude-shift keying (ASK)

Amplitude-Shift Keying modulation captures a bit by the amplitude of the carrier signal. Thus, if the carrier signal is $A\cos(2\pi f_c t)$ and the digital data is given by $d(t)$, then in ASK

$$s(t) = \begin{cases} A\cos(2\pi f_c t) & d(t) = 1 \\ 0 & d(t) = 0 \end{cases}$$

1.2 Frequency-shift keying (FSK)

Frequency-shift keying encodes the digital data using multiple frequencies around the center frequency f_c of the carrier signal. In Binary Frequency-Shift Keying (BFSK), we have

$$s(t) = \begin{cases} A\cos(2\pi f_1 t) & d(t) = 1 \\ A\cos(2\pi f_2 t) & d(t) = 0, \end{cases}$$

where $f_1, f_2 = f_c \pm f_d$.

In Multi-level Frequency Shift Keying (MFSK), we have

$$s_i(t) = A\cos(2\pi f_i t) \; 1 \le i \le M,$$

where $s_i(t)$ is the signal for the ith signal element, $f_i = f_c + (2i - 1 - M)f_d$, f_c is the carrier frequency, f_d is the difference frequency, and $M = 2^L$ is the number of different signal elements.

If T seconds is the bit period of the transmission signal, then each output signal element is held for $T_s = LT$ seconds. The minimum frequency separation required is $2f_d = 1/T_s$. Consequently the bandwidth required is M/T_s.

1.3 Phase-shift keying (PSK)

In Phase Shift Keying, the digital data is encoded by the phase of the carrier signal. In BPSK, we have

$$s(t) = \begin{cases} A\cos(2\pi f_c t) & d(t) = 1 \\ -A\cos(2\pi f_c t) & d(t) = 0, \end{cases}$$

Alternatively, if we have binary representation as $\{-1, 1\}$ instead of $\{0, 1\}$, we obtain $s(t) = Ad(t)\cos(2\pi f_c t)$.

In Differential Phase Shift Keying (DPSK), we encode a 1 by a change of phase, and 0 by maintaining phase. Just as in MFSK, we can use multi-level signaling in PSK. The QPSK modulation technique uses the following encoding.

$$s(t) = \begin{cases} A\cos(2\pi f_c t + \pi/4) & 11 \\ A\cos(2\pi f_c t + 3\pi/4) & 01 \\ A\cos(2\pi f_c t - 3\pi/4) & 00 \\ A\cos(2\pi f_c t - \pi/4) & 10 \end{cases}$$

The Quadrature Amplitude Modulation technique (QAM) is a combination of ASK and PSK techniques. It can also be seen as a logical extension of QPSK. The core idea is that it is possible to send two different signals simultaneously on the same carrier frequency, one phase-shifted by $\pi/2$ with respect to the other.

The digital data signal $d(t)$ is split into two signals, a signal $d_o(t)$ consisting of the odd bits and a signal $d_e(t)$ consisting of the even bits. In QAM, the encoding is given by

$$s(t) = d_o(t)\cos(2\pi f_c t) + d_e(t)\sin(2\pi f_c t).$$

1.4 Geometric representation of digital modulation techniques

Suppose the digital data is expressed in terms of M symbols. So the digital modulation uses M signal elements given by the set $S = \{s_i(t) : 1 \leq i \leq M\}$. Each signal can be viewed as a point in a vector space. Given a basis of this vector space, any point can be expressed as a linear combination of the basis elements. Thus, if the basis elements are $\phi_1(t), \ldots \phi_N(t)$, then we can write

$$s_i(t) = \sum_{j=1} N s_{ij}\phi_i(t).$$

The basis elements satisfy the standard orthogonality condition

$$\int_{-\infty}^{\infty} \phi_j(t)\phi_k(t)dt = 0, \ j \neq k.$$

Furthermore, the energy (per unit resistance) of each basis element can be normalized to 1.

$$\int_{-\infty}^{\infty} \phi_j^2(t)dt = 1.$$

Let us now analyze the phase shift keying techniques using this geometric representation. In BPSK modulation, the basis consists of only one element. If T_b is the time interval for a bit, then we have

$$\phi_1(t) = \sqrt{\frac{2}{T_b}} \cos(2\pi f_c t), 0 \leq t \leq T_b,$$

where T_b is the time period per bit. And we can write the two signal elements, corresponding to the two bits, as

$$s_1(t) = \sqrt{E_b}\phi_1(t) \;\; s_2(t) = -\sqrt{E_b}\phi_1(t),$$

where E_b is energy per bit. Thus, BPSK can be represented geometrically by two points on a line, $(\sqrt{E_b})$ and $(-\sqrt{E_b})$.

In QPSK, the four different signal elements can be expressed in terms of two basis elements

$$\begin{aligned} \phi_1(t) &= \sqrt{\frac{1}{T_b}} \cos(2\pi f_c t), 0 \leq t \leq 2T_b \\ \phi_2(t) &= \sqrt{\frac{1}{T_b}} \sin(2\pi f_c t), 0 \leq t \leq 2T_b. \end{aligned}$$

We define each basis element for $2T_b$ seconds since each signal element consists of 2 bits. Now each of the four signals can be expressed as

$$s_i(t) = \pm\sqrt{E_b}\phi_1(t) \pm \sqrt{E_b}\phi_2(t).$$

Thus QPSK can be expressed geometrically by four points on a 2-D plane (in which each dimension represents a basis element) $(\sqrt{E_b}, \sqrt{E_b})$, $(-\sqrt{E_b}, \sqrt{E_b})$, $(\sqrt{E_b}, -\sqrt{E_b})$, and $(-\sqrt{E_b}, -\sqrt{E_b})$.

QAM can also be easily expressed geometrically by 16 points on a 2-D plane. Frequency-shift keying schemes can be represented using higher-dimensional spaces by having one dimension for each of the frequencies.

The geometric representation provides good intuition for the power efficiency of a modulation technique. The more greater the maximum distance between two signal points, the higher the power efficiency since the more immune is the encoding technique to noise. Thus, QAM is more bandwidth-efficient than QPSK but less power-efficient. QPSK has the same power-efficiency as BPSK but better bandwidth-efficiency.

In frequency-shift keying schemes, increasing the number of signal elements increases the power efficiency at the expense of bandwidth efficiency since the different signal elements remain well-separated in the signal space (occupying orthogonal dimensions). In phase-shift keying schemes, increasing the number of signal elements decreases the power efficiency at the expense of bandwidth efficiency.

1.5 Bandwidth efficiency of digital modulation techniques

The following are approximate relationships between the transmission bandwidth B_T of a modulation technique and the information rate R. For ASK, BPSK, and BFSK, we have

$$B_T = (1 + r)R,$$

where $0 < r < 1$ is a parameter related to the filtering technique used in the modulation.

For MPSK, we have

$$B_T = \left(\frac{1 + r}{\log_2 M}\right) R,$$

and for MFSK, we have

$$B_T = \left(\frac{(1 + r)M}{\log_2 M}\right) R.$$

2 Analog modulation: Analog data on analog signals

There are two principles reasons to modulate analog data over analog signals. First, the frequency spectrum of the underlying data may be different than that needed for transmission. In the case of wireless transmissions, especially, it is improbable to transmit baseband signals without modulation. The other reason is that analog modulation permits frequency division multiplexing since we can transmit data of different users over different bands by appropriately modulating them.

2.1 Amplitude modulation

In Amplitude Modulation (AM), the analog data is encoded in the amplitude of the signal. Specifically, if the analog data is given by $m(t)$, then the encoded signal is given by

$$s(t) = (1 + n_a m(t)) \cos(2\pi f_c t),$$

where n_a is the ratio of the amplitude of the signal to that of the data and is called the amplitude modulation index. The frequency spectrum of this signal is twice as wide as the bandwidth of $m(t)$. To see this, suppose $m(t) = 2\cos(2\pi f t)$. Then $s(t)$ can be written as

$$s(t) = cos(2\pi f_c t) + \frac{n_a}{2} \left(\cos(2\pi(f_c + f)t) + \cos(2\pi(f_c - f)t\right).$$

As the above equation shows, any frequency component at frequency f in the data $m(t)$ shows up at frequencies $f_c + f$ and $f_c - f$ in the encoded signal. This is also easily seen in the figures given in the textbook. Such an AM encoding is referred to as Double Side-Banded and requires a transmission bandwidth B_T that is twice the bandwidth B of the original signal. Single-side banded AM encoding filters out one of the bands, and has $B_T = B$.

2.2 Angle modulation

Modulating the analog signal by varying its phase or frequency can be viewed as two variants of *angle modulation*. If the analog data is given by $m(t)$, then an angle modulation encoded signal is given by

$$s(t) = A\cos(2\pi f_c t + \phi(t),$$

where $\phi(t) = n_p m(t)$ for phase modulation and $\phi'(t) = n_f m(t)$ for frequency modulation. Here n_p refers to a parameter called the phase modulation index, n_f is the frequency modulation index and $\phi'(t)$ is the derivative of $\phi(t)$ wrt t.

One property of angle modulation is that it has constant power regardless of how the analog data varies. This makes it less susceptible to noise. The transmission bandwidth required is given by $B_T = 2(\beta + 1)B$, where $\beta = n_p A$ for phase modulation and $\beta = n_f A/(2\pi B)$ for frequency modulation.

3 Digitization: Analog data on digital signals

There are two popular ways of digitizing analog data. *Pulse code modulation* (PCM) is a process in which the analog data is sampled at a rate given by Nyquist's sampling theorem (that is, twice the bandwidth of the analog data). If we store each sample to full precision, then we can recover the entire analog data (assuming it has a finite bandwidth). This is, however, unrealistic. In PCM, we divide the range of the analog data to a number of intervals (a process called quantization), and replace a sample by the nearest (or closest higher/lower) quantized level.

The number of quantized levels determines the resolution of digitization. More the number of levels, the more precise (or higher resolution) the digitization process. That is, the error between the analog data value at any time, and the quantized value that we replace it by is small. The quantization step in PCM may be either uniform (all intervals are identical size) or non-uniform in which the intervals are smaller at lower analog data values (resulting in more uniform relative precision).

An alternative digitization technique is *Delta Modulation*, in which rather than encoding the actual analog data values, we encode the variation in data. We select a step size s and modulation rate r, and every $1/r$ seconds increase (resp., decrease) the digital signal value by s if the analog data value increases (resp., decreases) in $1/r$ seconds. Clearly, the resolution of delta modulation depends on the choice of r and s; higher the r, better the resolution. The effect of s is more subtle and depends on how the analog data varies. For instance, if s is too small, the digital signal may be underestimating (resp., overestimating) the analog data when it is increasing (resp., decreasing).

References

[Rap95] Theodore S. Rappaport. *Wireless Communications: Principles and Practice*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995.

[Sta05] William Stallings. *Wireless Communications and Networks*. Prentice Hall Professional Technical Reference, 2005. 2nd edition.