BASEBAND SIGNALING AND PULSE SHAPING

Michael L. Honig, Northwestern University, and Melbourne Barton, Bellcore

Many physical communications channels, such as radio channels, accept a continuous-time waveform as input. Consequently, a sequence of source bits, representing data or a digitized analog signal, must be converted to a continuous-time waveform at the transmitter. In general, each successive group of bits taken from this sequence is mapped to a particular continuous-time pulse. In this chapter we discuss the basic principles involved in selecting such a pulse for channels that can be characterized as linear and time-invariant with finite bandwidth.

1. Communications System Model

Figure 1a shows a simple block diagram of a communications system. The sequence of source bits $\{b_i\}$ are grouped into sequential blocks (vectors) of m bits $\{\mathbf{b}_i\}$, and each binary vector $\mathbf{b}_i$ is mapped to one of 2^m pulses, $p(\mathbf{b}_i; t)$, which is transmitted over the channel. The transmitted signal as a function of time can be written as

$$s(t) = \sum_i p(\mathbf{b}_i; t - iT) \tag{1}$$

where $1/T$ is the rate at which each group of m bits, or pulses, are introduced to the channel. The information (bit) rate is therefore m/T.

The channel in Figure 1a can be a radio link, which may distort the input signal $s(t)$ in a variety of ways. For example, it may introduce pulse dispersion (due to finite bandwidth) and multipath, as well as additive background noise. The output of the channel is denoted as $x(t)$, which is processed by the receiver to determine estimates of the source bits. The receiver can be quite complicated; however, for the purpose of this discussion, it is sufficient to assume only that it contains a front-end filter and a sampler, as shown in Figure 1a. This assumption is valid for a wide variety of detection strategies. The purpose of the receiver filter is to remove noise outside of the transmitted frequency band, and to compensate for the channel frequency response.

A commonly used channel model is shown in Figure 1b, and consists of a linear, time-invariant filter, denoted as $G(f)$, followed by additive noise $n(t)$. The channel output is therefore

$$x(t) = [g(t) * s(t)] + n(t) \tag{2}$$

where $g(t)$ is the channel impulse response associated with $G(f)$, and "$*$" denotes convolution: $g(t) * s(t) = \int_{-\infty}^{\infty} g(t-\tau)s(\tau)d\tau$. This channel model accounts for all linear, time-invariant channel impairments, such as finite bandwidth, and time-invariant multipath. It does not account for time-varying impairments, such as rapid fading due to time-varying multipath. Nevertheless, this model can be considered valid over short time-periods during which the multipath parameters remain constant.

In Figure 1 it is assumed that all signals are **baseband** signals, which means that the frequency content is centered around $f = 0$ (DC). The channel passband therefore (partially) coincides with the transmitted spectrum. In general, this condition requires that the transmitted signal be modulated by an appropriate carrier frequency, and demodulated at the receiver. In that case, the model in Figure 1 still applies; however, *baseband-equivalent* signals must be derived from their modulated (passband) counterparts. *Baseband signaling and pulse shaping* refers to the way in which a group of source bits is mapped to a baseband transmitted pulse.

As a simple example of baseband signaling, we can take $m = 1$ (map each source bit to a pulse), assign a 0 bit to a pulse $p(t)$, and a 1 bit to the pulse $-p(t)$. Perhaps the simplest example of a baseband pulse is the *rectangular* pulse given by $p(t) = 1$, $0 < t \leq T$, and $p(t) = 0$ elsewhere. In this case, we can write the transmitted signal as

$$s(t) \;=\; \sum_i A_i p(t - iT) \tag{3}$$

where each symbol A_i takes on a value of +1 or −1, depending on the value of the ith bit, and $1/T$ is the *symbol rate*, namely, the rate at which the symbols A_i are introduced to the channel.

The preceding example is called *binary* **Pulse Amplitude Modulation (PAM)**, since the data symbols A_i are binary-valued, and they amplitude modulate the transmitted pulse $p(t)$. The information rate (bits per second) in this case is the same as the symbol rate $1/T$. As a simple extension of this signaling technique, we can increase m, and choose A_i from one of $M = 2^m$ values to transmit at bit rate m/T. This is known as *M-ary* PAM. For example, letting $m = 2$, each pair of bits can be mapped to a pulse in the set $\{p(t), -p(t), 3p(t), -3p(t)\}$.

In general, the transmitted symbols $\{A_i\}$, the baseband pulse $p(t)$, and channel impulse response $g(t)$ can be *complex-valued*. For example, each successive pair of bits might select a symbol from the set $\{1, -1, j, -j\}$, where $j = \sqrt{-1}$. This is a consequence of considering the baseband equivalent of passband modulation. (That is,

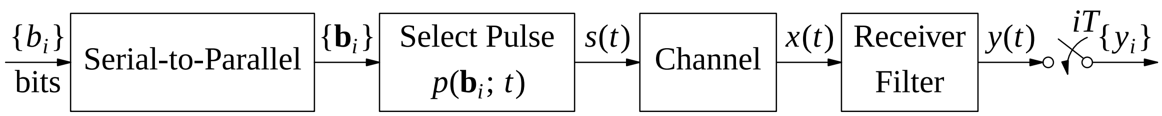

Figure 1a. Communication system model. The source bits are grouped into binary vectors, which are mapped to a sequence of pulse shapes.

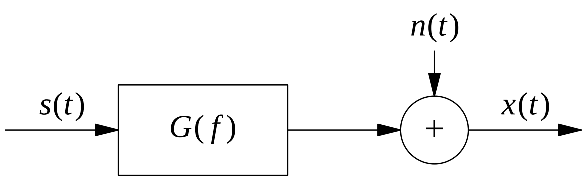

Figure 1b. Channel model consisting of a linear, time-invariant system (transfer function) followed by additive noise.

generating a transmitted spectrum which is centered around a carrier frequency f_c.) Here we are not concerned with the relation between the passband and baseband equivalent models, and simply point out that the discussion and results in this chapter apply to complex-valued symbols and pulse shapes.

As an example of a signaling technique which is not PAM, let $m = 1$ and

$$p(0;\, t) \;=\; \begin{cases} \sqrt{2}\sin(2\pi f_1 t) & 0 < t < T \\ 0 & \text{elsewhere} \end{cases} \qquad p(1;\, t) \;=\; \begin{cases} \sqrt{2}\sin(2\pi f_2 t) & 0 < t < T \\ 0 & \text{elsewhere} \end{cases} \tag{4}$$

where f_1 and $f_2 \neq f_1$ are fixed frequencies selected so that $f_1 T$ and $f_2 T$ (number of cycles for each bit) are multiples of 1/2. These pulses are *orthogonal*, namely, $\int_0^T p(1;\, t)p(0;\, t)dt = 0$. This choice of pulse shapes is called *binary* **Frequency-Shift Keying (FSK)**.

Another example of a set of orthogonal pulse shapes for $m = 2$ bits/T is shown in Figure 2. Because these pulses may have as many as three transitions within a symbol period, the transmitted spectrum occupies roughly four times the transmitted spectrum of binary PAM with a rectangular pulse shape. The spectrum is therefore ‘‘spread’’ across a much larger band than the smallest required for reliable transmission, assuming a data rate of $2/T$. This type of signaling is referred to as **spread-spectrum**. Spread-spectrum signals are more robust with respect to interference from other transmitted signals than are narrow-band signals.[1]

2. Intersymbol Interference and the Nyquist Criterion

Consider the transmission of a PAM signal illustrated in Figure 3. The source bits $\{b_i\}$ are mapped to a sequence of levels $\{A_i\}$, which modulate the transmitter pulse $p(t)$. The channel input is therefore given by (3) where $p(t)$ is the impulse response of the transmitter *pulse-shaping filter* $P(f)$ shown in Figure 3. The input to the transmitter filter $P(f)$ is the modulated sequence of delta functions $\sum_i A_i \delta(t - iT)$. The channel is represented by the transfer function $G(f)$ (plus noise), which has impulse response $g(t)$, and the receiver filter has transfer function $R(f)$ with associated impulse response

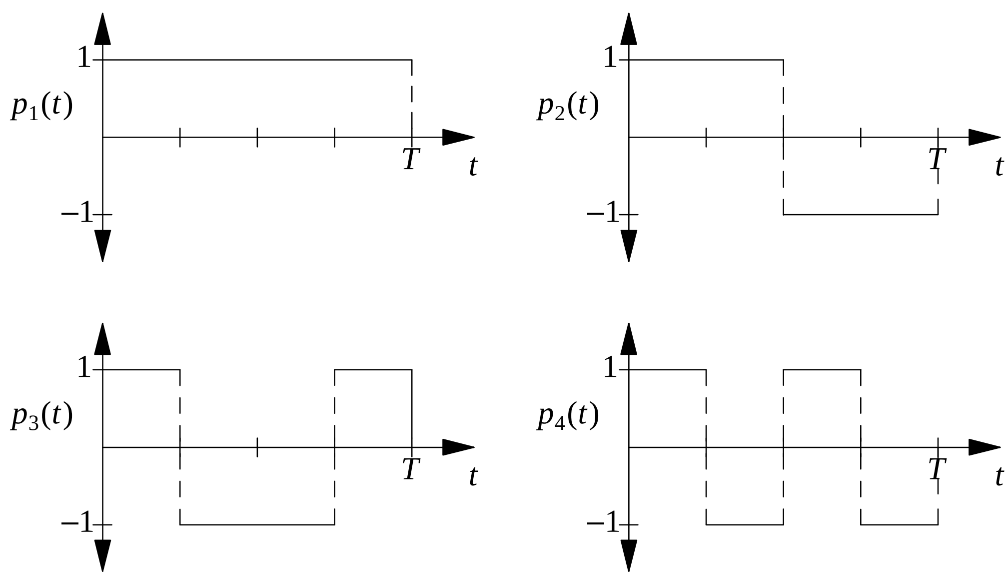

Figure 2. Four orthogonal spread-spectrum pulse shapes.

$r(t)$.

Let $h(t)$ be the overall impulse response of the combined transmitter, channel, and receiver, which has transfer function $H(f) = P(f)G(f)R(f)$. We can write

---

1. This example can also be viewed as coded binary PAM. Namely, each pair of two source bits are mapped to 4 coded bits, which are transmitted via binary PAM with a rectangular pulse. The current IS-95 air interface uses an extension of this signaling method in which groups of 6 bits are mapped to 64 orthogonal pulse shapes with as many as 63 transitions during a symbol.

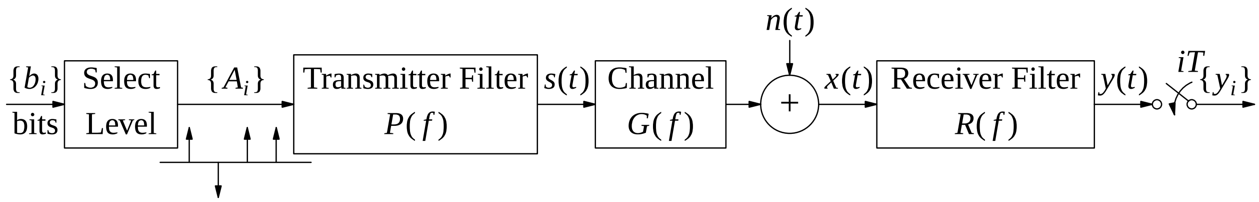

Figure 3. Baseband model of a Pulse Amplitude Modulation system.

$h(t) = p(t) * g(t) * r(t)$. The output of the receiver filter is then

$$y(t) = \sum_i A_i h(t - iT) + \tilde{n}(t) \tag{5}$$

where $\tilde{n}(t) = r(t) * n(t)$ is the output of the filter $R(f)$ with input $n(t)$. Assuming that samples are collected at the output of the filter $R(f)$ at the symbol rate $1/T$, we can write the kth sample of $y(t)$ as

$$\begin{aligned} y(kT) &= \sum_i A_i h(kT - iT) + \tilde{n}(kT) \\ &= A_k h(0) + \sum_{i \neq k} A_i h(kT - iT) + \tilde{n}(kT) \,. \end{aligned} \tag{6}$$

The first term on the right of (6) is the kth transmitted symbol scaled by the system impulse response at $t = 0$. If this were the only term on the right side of (6), we could obtain the source bits without error by scaling the received samples by $1/h(0)$. The second term on the right of (6) is called **intersymbol interference**, which reflects the view that neighboring symbols interfere with the detection of each desired symbol.

One possible criterion for choosing the transmitter and receiver filters is to minimize intersymbol interference. Specifically, if we choose $p(t)$ and $r(t)$ so that

$$h(kT) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}, \tag{7}$$

then the kth received sample is

$$y(kT) = A_k + \tilde{n}(kT). \tag{8}$$

In this case, the intersymbol interference has been eliminated. This choice of $p(t)$ and $r(t)$ is called a **zero-forcing** solution, since it ‘‘forces’’ the intersymbol interference to zero. Depending on the type of detection scheme used, a zero-forcing solution may not be desirable. This is because the probability of error also depends on the noise intensity, which generally increases when intersymbol interference is suppressed. However, it is instructive to examine the properties of the zero-forcing solution.

We now view (7) in the frequency domain. Since $h(t)$ has Fourier Transform

$$H(f) = P(f)G(f)R(f), \tag{9}$$

where $P(f)$ is the Fourier Transform of $p(t)$, the bandwidth of $H(f)$ is limited by the bandwidth of the channel $G(f)$. We will assume that $G(f) = 0$, $|f| > W$. The sampled impulse response $h(kT)$ can therefore be written as the inverse Fourier Transform

$$h(kT) = \int_{-W}^{W} H(f) e^{j2\pi fkT} df \ .$$

Through a series of manipulations, this integral can be rewritten as an inverse *Discrete* Fourier Transform:

$$h(kT) = T \int_{-1/(2T)}^{1/(2T)} H_{eq}(e^{j2\pi fT}) e^{j2\pi fkT} df \tag{10a}$$

where

$$H_{eq}(e^{j2\pi fT}) = \frac{1}{T} \sum_k H\left(f + \frac{k}{T} \right) = \frac{1}{T} \sum_k P\left(f + \frac{k}{T} \right) G\left(f + \frac{k}{T} \right) R\left(f + \frac{k}{T} \right). \tag{10b}$$

This relation states that $H_{eq}(z)$, $z = e^{j2\pi fT}$, is the Discrete Fourier Transform of the sequence $\{h_k\}$ where $h_k = h(kT)$. Sampling the impulse response $h(t)$ therefore changes the transfer function $H(f)$ to the *aliased* frequency response $H_{eq}(e^{j2\pi fT})$. From (10) and (6) we conclude that $H_{eq}(z)$ is the transfer function that relates the sequence of input data symbols $\{A_i\}$ to the sequence of received samples $\{y_i\}$, where $y_i = y(iT)$, in the absence of noise. This is illustrated in Figure 4. For this reason, $H_{eq}(z)$ is called the **equivalent discrete-time transfer function** for the overall system transfer function $H(f)$.

Since $H_{eq}(e^{j2\pi fT})$ is the Discrete Fourier Transform of the sequence $\{h_k\}$, the time-domain, or sequence condition (7) is equivalent to the frequency-domain condition

$$H_{eq}(e^{j2\pi fT}) = 1 \ . \tag{11}$$

This relation is called the **Nyquist criterion**. From (10b) and (11) we make the following observations:

1. To satisfy the Nyquist criterion, the channel bandwidth W must be at least $1/(2T)$. Otherwise, $G(f + n/T) = 0$ for f in some interval of positive length for all n, which implies that $H_{eq}(e^{j2\pi fT}) = 0$ for f in the same interval.

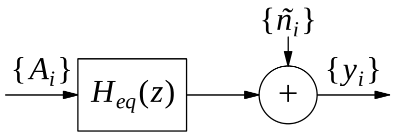

Figure 4. Equivalent discrete-time channel for the PAM system shown in Figure 3 ($y_i = y(iT)$, $\tilde{n}_i = \tilde{n}(iT)$).

2. For the minimum bandwidth $W = 1/(2T)$, (10b) and (11) imply that $H(f) = T$ for $|f| < 1/(2T)$ and $H(f) = 0$ elsewhere. This implies that the system impulse response is given by

$$h(t) = \frac{\sin(\pi t/T)}{\pi t/T} . \tag{12}$$

(Since $\int_{-\infty}^{\infty} h^2(t) = T$, the transmitted signal $s(t) = \sum_i A_i h(t - iT)$ has power equal to the symbol variance $E[|A_i|^2]$.) The impulse response in (12) is called a *minimum bandwidth* or *Nyquist* pulse. The frequency band $[-1/(2T), 1/(2T)]$ (i.e., the passband of $H(f)$) is called the **Nyquist band**.

3. Suppose that the channel is bandlimited to *twice* the Nyquist bandwidth. That is, $G(f) = 0$ for $|f| > 1/T$. The condition (11) then becomes

$$H(f) + H\left(f - \frac{1}{T}\right) + H\left(f + \frac{1}{T}\right) = T. \tag{13}$$

Assume for the moment that $H(f)$ and $h(t)$ are both real-valued, so that $H(f)$ is an even function of f ($H(f) = H(-f)$). This is the case when the receiver filter is the matched filter (see Section 3). We can then rewrite (13) as

$$H(f) + H\left(\frac{1}{T} - f\right) = T, \quad 0 < f < \frac{1}{2T}, \tag{14}$$

which states that $H(f)$ must have odd symmetry about $f = 1/(2T)$. This is illustrated in Figure 5, which shows two different transfer functions $H(f)$ that satisfy the Nyquist criterion.

4. The pulse shape $p(t)$ enters into (11) only through the product $P(f)R(f)$. Consequently, either $P(f)$ or $R(f)$ can be fixed, and the other filter can be adjusted, or adapted to the particular channel. Typically, the pulse shape $p(t)$ is fixed, and the receiver filter is adapted to the (possibly time-varying) channel.

Raised Cosine Pulse

Suppose that the channel is ideal with transfer function

$$G(f) = \begin{cases} 1, & |f| < W \\ 0, & |f| > W \end{cases} . \tag{15}$$

To maximize bandwidth efficiency, Nyquist pulses given by (12) should be used where $W = 1/(2T)$. However, this type of signaling has two major drawbacks. First, Nyquist pulses are noncausal and of infinite duration. They can be approximated in practice by introducing an appropriate delay, and truncating the pulse. However, the pulse decays very slowly, namely, as $1/t$, so that the truncation window must be wide. This is equivalent to observing that the ideal bandlimited frequency response given by (15) is difficult to approximate closely. The second drawback, which is more important, is the fact that this type of signaling is not robust with respect to sampling jitter. Namely, a small sampling offset ε produces the output sample

$$y(kT + \varepsilon) = \sum_i A_i \frac{\sin[\pi(k - i + \varepsilon/T)]}{\pi(k - i + \varepsilon/T)} . \tag{16}$$

Since the Nyquist pulse decays as $1/t$, this sum is not guaranteed to converge. A particular choice of symbols $\{A_i\}$ can therefore lead to very large intersymbol interference, no matter how small the offset. Minimum bandwidth signaling is therefore impractical.

The preceding problem is generally solved in one of two ways in practice:

1. The pulse bandwidth is increased to provide a faster pulse decay than $1/t$.
2. A *controlled* amount of intersymbol interference is introduced at the transmitter, which can be subtracted out at the receiver.

The former approach sacrifices bandwidth efficiency, whereas the latter approach sacrifices power efficiency. We will examine the latter approach in Section 5. The most common example of a pulse, which illustrates the first technique, is the **raised cosine pulse**, given by

$$h(t) = \left(\frac{\sin(\pi t/T)}{\pi t/T}\right)\left(\frac{\cos(\alpha\pi t/T)}{1 - (2\alpha t/T)^2}\right) \tag{17}$$

which has Fourier Transform

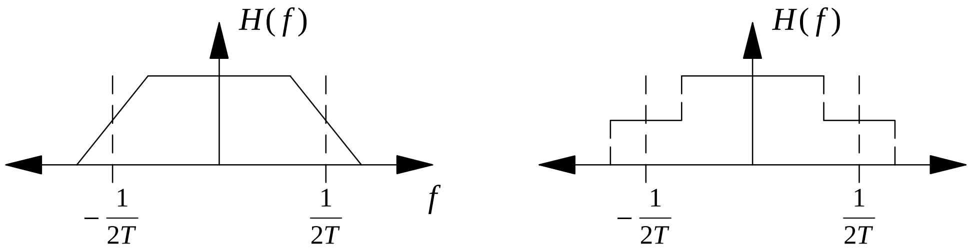

Figure 5. Two examples of frequency responses that satisfy the Nyquist criterion.

$$H(f) = \begin{cases} T & 0 \le |f| \le \dfrac{1-\alpha}{2T} \\ \dfrac{T}{2}\left\{1 + \cos\left[\dfrac{\pi T}{\alpha}\left(|f| - \dfrac{1-\alpha}{2T}\right)\right]\right\} & \dfrac{1-\alpha}{2T} \le |f| \le \dfrac{1+\alpha}{2T} \\ 0 & |f| > \dfrac{1+\alpha}{2T} \end{cases} \qquad (18)$$

where $0 \le \alpha \le 1$.

Plots of $p(t)$ and $P(f)$ are shown in Figure 6 for different values of α. It is easily verified that $h(t)$ satisfies the Nyquist criterion (7), and consequently $H(f)$ satisfies (11). When $\alpha = 0$, $H(f)$ is the Nyquist pulse with minimum bandwidth $1/(2T)$, and when $\alpha > 0$, $H(f)$ has bandwidth $(1+\alpha)/(2T)$ with a ‘‘raised cosine rolloff’’. The parameter α therefore represents the additional, or **excess bandwidth** as a fraction of the minimum bandwidth $1/(2T)$. For example, when $\alpha = 1$, we say that that the pulse is a raised cosine pulse with 100% excess bandwidth. This is because the pulse bandwidth, $1/T$, is twice the minimum bandwidth. Because the raised cosine pulse decays as $1/t^3$, performance is robust with respect to sampling offsets.

The raised cosine frequency response (18) applies to the combination of transmitter, channel, and receiver. If the transmitted pulse shape $p(t)$ is a raised cosine pulse, then $h(t)$ is a raised cosine pulse only if the combined receiver and channel frequency response is constant. Even with an ideal (transparent) channel, however, the optimum (matched) receiver filter response is generally not constant in the presence of additive Gaussian noise. An alternative is to transmit the *square-root raised cosine* pulse shape, which has frequency response $P(f)$ given by the square-root of the raised cosine frequency response in (18). Assuming an ideal channel, setting the receiver frequency response $R(f) = P(f)$ then results in an overall raised cosine system response $H(f)$.

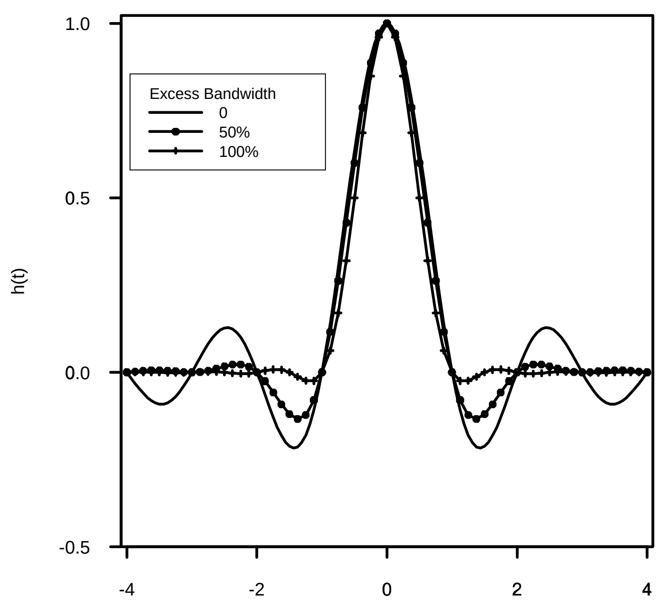

Fig. 6(a). Raised Cosine pulse.

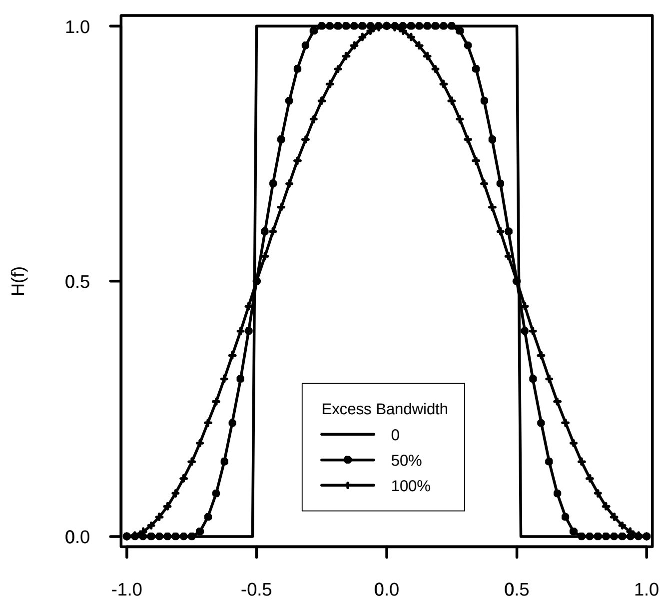

Fig. 6(b). Raised Cosine Spectrum.

3. Nyquist Criterion With Matched Filtering

Consider the transmission of an isolated pulse $A_0\delta(t)$. In this case the input to the receiver in Figure 3 is

$$x(t) \; = \; A_0\tilde{g}(t) + n(t) \tag{19}$$

where $\tilde{g}(t)$ is the inverse Fourier Transform of the combined transmitter-channel transfer function $\tilde{G}(f) = P(f)G(f)$. We will assume that the noise $n(t)$ is white with spectrum $N_0/2$. The output of the receiver filter is then

$$y(t) \; = \; r(t) * x(t) \; = \; A_0[r(t) * \tilde{g}(t)] + [r(t) * n(t)]. \tag{20}$$

The first term on the right-hand side is the desired signal, and the second term is noise. Assuming that $y(t)$ is sampled at $t = 0$, the ratio of signal energy to noise energy, or Signal-to-Noise Ratio (SNR) at the sampling instant, is

$$\text{SNR} \; = \; \frac{E[|A_0|^2]\left| \int_{-\infty}^{\infty} r(-t)\tilde{g}(t)dt \right|^2}{\frac{N_0}{2}\int_{-\infty}^{\infty} |r(t)|^2 dt} \; . \tag{21}$$

The receiver impulse response that maximizes this expression is $r(t) = \tilde{g}^*(-t)$ (complex conjugate of $\tilde{g}(-t)$), which is known as the **matched filter** impulse response. The associated transfer function is $R(f) = \tilde{G}^*(f)$.

Choosing the receiver filter to be the matched filter is optimal in more general situations, such as when detecting a sequence of channel symbols with intersymbol interference (assuming the additive noise is Gaussian). We therefore reconsider the Nyquist criterion when the receiver filter is the matched filter. In this case the baseband model is shown in Figure 7, and the output of the receiver filter is given by

$$y(t) \; = \; \sum_i A_i h(t - iT) + \tilde{n}(t) \tag{22}$$

where the baseband pulse $h(t)$ is now the impulse response of the filter with transfer function $|\tilde{G}(f)|^2 = |P(f)G(f)|^2$. This impulse response is the *autocorrelation* of the impulse response of the combined transmitter-channel filter $\tilde{G}(f)$:

$$h(t) \; = \; \int_{-\infty}^{\infty} \tilde{g}^*(s)\tilde{g}(s+t)ds \; . \tag{23}$$

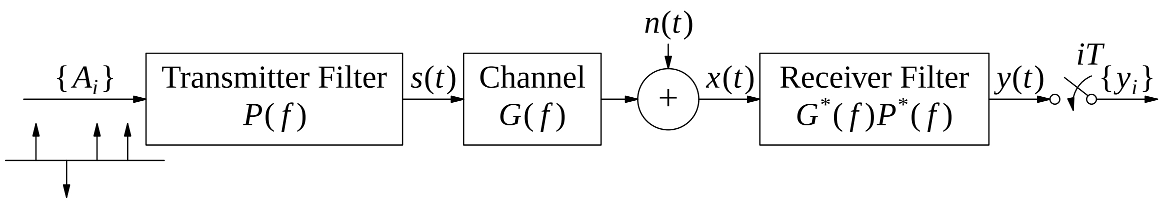

Figure 7. Baseband PAM model with a matched filter at the receiver.

With a matched-filter at the receiver, the equivalent discrete-time transfer function is

$$H_{eq}(e^{j2\pi fT}) = \frac{1}{T}\sum_k \left|\tilde{G}\left(f - \frac{k}{T}\right)\right|^2 = \frac{1}{T}\sum_k \left|P\left(f - \frac{k}{T}\right)G\left(f - \frac{k}{T}\right)\right|^2 , \tag{24}$$

which relates the sequence of transmitted symbols $\{A_k\}$ to the sequence of received samples $\{y_k\}$ in the absence of noise. Note that $H_{eq}(e^{j2\pi fT})$ is *positive, real-valued, and an even function of f*. If the channel is bandlimited to *twice* the Nyquist bandwidth, then $H(f) = 0$ for $|f| > 1/T$, and the Nyquist condition is given by (14) where $H(f) = |G(f)P(f)|^2$. The aliasing sum in (10b) can therefore be described as a ‘‘folding’’ operation in which the channel response $|H(f)|^2$ is ‘‘folded’’ around the Nyquist frequency $1/(2T)$. For this reason, $H_{eq}(e^{j2\pi fT})$ with a matched receiver filter is often referred to as the ‘‘folded channel spectrum’’.

4. Eye Diagrams

One way to assess the severity of distortion due to intersymbol interference in a digital communications system is to examine the **eye diagram**. The eye diagram is illustrated in Figure 8 for a raised cosine pulse shape with 25% excess bandwidth, and an ideal bandlimited channel. Figure 8a shows the data signal at the receiver

$$y(t) = \sum_i A_i\, h(t - iT) + n(t) \tag{25}$$

where $h(t)$ is given by (17), $\alpha = 1/4$, each symbol A_i is independently chosen from the set $\{\pm 1, \pm 3\}$, where each symbol is equally likely, and $n(t)$ is bandlimited white Gaussian noise. (The received SNR is 30 dB.) The eye diagram is constructed from the time-domain data signal $y(t)$ as follows (assuming nominal sampling times at kT, $k = 0, 1, 2, \cdots$):

1. Partition the waveform $y(t)$ into successive segments of length T starting from $t = T/2$.

2. Translate each of these waveform segments $(y(t),\ (k + 1/2)T \le t \le (k + 3/2)T,\ k = 0, 1, 2, \cdots)$ to the interval $[-T/2, T/2]$, and superimpose.

The resulting picture is shown in Figure 8b for the $y(t)$ shown in Figure 8a. (Partitioning $y(t)$ into successive segments of length iT, $i > 1$, is also possible. This would result in i successive eye diagrams.) The number of ‘‘eye openings’’ is one less than the number of transmitted signal levels. In practice, the eye diagram is easily viewed on an oscilloscope by by applying the received waveform $y(t)$ to the vertical deflection plates of the oscilloscope, and applying a saw-tooth waveform at the symbol rate $1/T$ to the horizontal deflection plates. This causes successive symbol intervals to be translated into one interval on the oscilloscope display.

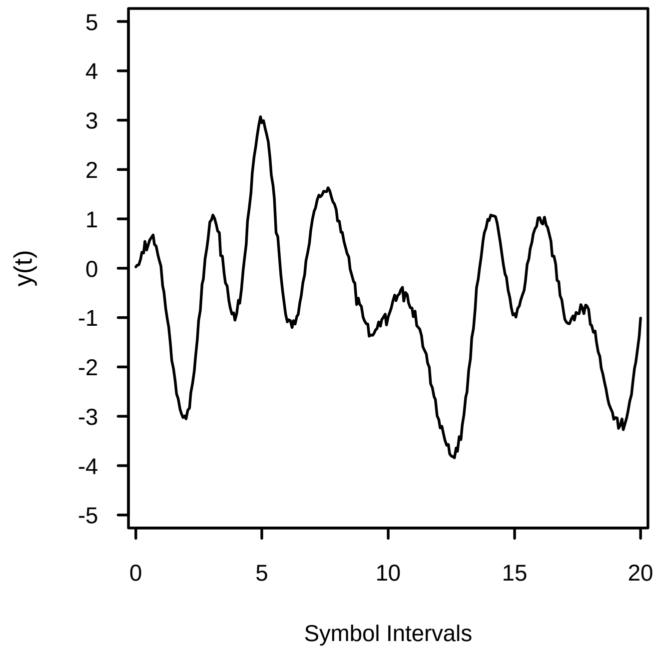

Figure 8a. Received signal $y(t)$.

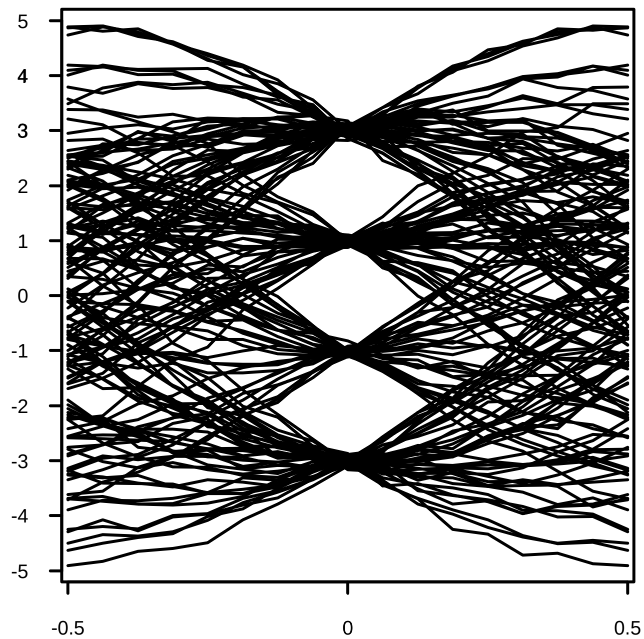

Fig. 8(b). Eye diagram for received signal shown in Fig. 8(a).

Each waveform segment $y(t)$, $(k + 1/2)T \le t \le (k + 3/2)T$, depends on the particular sequence of channel symbols surrounding A_k. The number of channel symbols which affects a particular waveform segment depends on the extent of the intersymbol interference, shown in (6). This in turn depends on the duration of the impulse response $h(t)$. For example, if $h(t)$ has most of its energy in the interval $0 < t < mT$, then each waveform segment depends on approximately m symbols. Assuming binary transmission, this implies that there are a total of 2^m waveform segments that can be superimposed in the eye diagram. (It is possible that only one sequence of channel symbols causes significant intersymbol interference, and this sequence occurs with very low probability.) In current digital wireless applications the impulse response typically spans only a few symbols.

The eye diagram has the following important features, which measure the performance of a digital communications system.

Vertical Eye Opening

The vertical ‘‘openings’’ at any time t_0, $-T/2 \le t_0 \le T/2$, represent the separation between signal levels with worst-case intersymbol interference, assuming that $y(t)$ is sampled at times $t = kT + t_0$, $k = 0, 1, 2 \cdots$. It is possible for the intersymbol interference to be large enough so that this vertical opening between some, or all, signal levels disappears altogether. In that case, the eye is said to be ‘‘closed’’. Otherwise, the eye is said to be ‘‘open’’. A closed eye implies that if the estimated bits are obtained by thresholding the samples $y(kT)$, then the decisions will depend primarily on the intersymbol interference, rather than on the desired symbol. The probability of error will therefore be close to 1/2. Conversely, wide vertical spacings between signal levels imply a large degree of immunity to additive noise. In general, $y(t)$ should be sampled at the times $kT + t_0$, $k = 0, 1, 2, \cdots$, where t_0 is chosen to maximize the vertical eye opening.

Horizontal Eye Opening

The width of each opening indicates the sensitivity to timing offset. Specifically, a very narrow eye opening indicates that a small timing offset will result in sampling where the eye is closed. Conversely, a wide horizontal opening indicates that a large timing offset can be tolerated, although the error probability will depend on the vertical opening.

Slope of the Inner Eye

The slope of the inner eye indicates sensitivity to timing jitter, or variance in the timing offset. Specifically, a very steep slope means that the eye closes rapidly as the timing offset increases. In this case, a significant amount of jitter in the sampling times significantly increases the probability of error.

The shape of the eye diagram is determined by the pulse shape. In general, the faster the baseband pulse decays, the wider the eye opening. For example, a rectangular pulse produces a box-shaped eye diagram (assuming binary signaling). The minimum bandwidth pulse shape (12) produces an eye diagram which is closed for all t except for $t = 0$. This is because, as shown earlier, an arbitrarily small timing offset can lead to an intersymbol interference term which is arbitrarily large, depending on the data sequence.

5. Partial-Response Signaling

To avoid the problems associated with Nyquist signaling over an ideal bandlimited channel, bandwidth and/or power efficiency must be compromised. Raised cosine pulses compromise bandwidth efficiency to gain robustness with respect to timing errors. Another possibility is to introduce a controlled amount of intersymbol at the transmitter, which can be removed at the receiver. This approach is called **partial-response (PR) signaling**. The terminology reflects the fact the sampled system impulse response does not have the ‘‘full-response’’ given by the Nyquist condition (7).

To illustrate PR signaling, suppose that the Nyquist condition (7) is replaced by the condition

$$h_k = \begin{cases} 1 & k = 0, 1 \\ 0 & \text{all other } k \end{cases} \tag{26}$$

The kth received sample is then

$$y_k = A_k + A_{k-1} + \tilde{n}_k , \tag{27}$$

so that there is intersymbol interference from one neighboring transmitted symbol. For now we focus on the spectral characteristics of PR signaling, and defer discussion of how to detect the transmitted sequence $\{A_k\}$ in the presence of intersymbol interference. The equivalent discrete-time transfer function in this case is the Discrete Fourier Transform of the sequence in (26):

$$H_{eq}(e^{j2\pi fT}) = \frac{1}{T}\sum_k H\left(f + \frac{k}{T}\right)$$

$$= 1 + e^{-j2\pi fT} = 2e^{-j\pi fT}\cos(\pi fT) . \tag{28}$$

As in the full-response case, for (28) to be satisfied, the *minimum* bandwidth of the channel $G(f)$ and transmitter filter $P(f)$ is $W = 1/(2T)$. Assuming $P(f)$ has this minimum bandwidth implies

$$H(f) = \begin{cases} 2Te^{-j\pi fT}\cos(\pi fT) & |f| < 1/(2T) \\ 0 & |f| > 1/(2T) \end{cases} \tag{29a}$$

and

$$h(t) = T(\operatorname{sinc}(t/T) + \operatorname{sinc}[(t-T)/T]) \tag{29b}$$

where $\operatorname{sinc} x = (\sin \pi x)/(\pi x)$. This pulse is called a *duobinary* pulse, and is shown along with the associated $H(f)$ in Figure 9. (Notice that $h(t)$ satisfies (26).) Unlike the ideal bandlimited frequency response, the transfer function $H(f)$ in (29a) is continuous, and is therefore easily approximated by a physically realizable filter. Duobinary PR was first proposed by Lender [1], and later generalized by Kretzmer [2].

The main advantage of the duobinary pulse (29b), relative to the minimum bandwidth pulse (12), is that signaling at the Nyquist symbol rate is feasible with *zero* excess bandwidth. Because the pulse decays much more rapidly than a Nyquist pulse, it is robust with respect to timing errors. Selecting the transmitter and receiver filters so that the overall system response is duobinary is appropriate in situations where the channel frequency response $G(f)$ is near zero, or has a rapid rolloff at the Nyquist band-edge $f = 1/(2T)$.

As another example of PR signaling, consider the *modified duobinary* partial-response

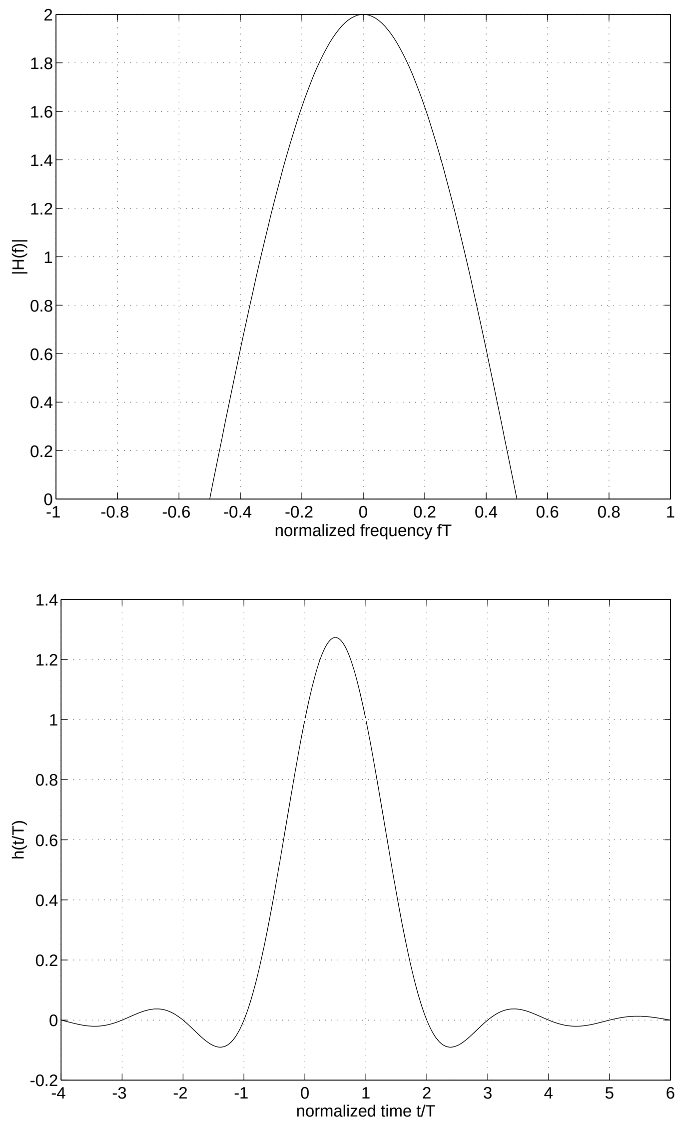

Figure 9. Duobinary frequency response and minimum bandwidth pulse.

$$h_k = \begin{cases} 1 & k = -1 \\ -1 & k = 1 \\ 0 & \text{all other } k \end{cases} \tag{30}$$

which has equivalent discrete-time transfer function

$$H_{eq}(e^{j2\pi fT}) = e^{j2\pi fT} - e^{-j2\pi fT}$$

$$= j2 \sin(2\pi fT). \tag{31}$$

With zero excess bandwidth the overall system response is

$$H(f) = \begin{cases} j2T \sin(2\pi fT) & |f| < 1/(2T) \\ 0 & |f| > 1/(2T) \end{cases} \tag{32a}$$

and

$$h(t) = T(\text{sinc}\,[(t+T)/T] - \text{sinc}\,[(t-T)/T]). \tag{32b}$$

These functions are plotted in Figure 10. This pulse shape is appropriate when the channel response $G(f)$ is near zero at both DC ($f = 0$) and at the Nyquist band-edge. This is often the case for wire (twisted-pair) channels where the transmitted signal is coupled to the channel through a transformer. Like duobinary PR, modified duobinary allows minimum bandwidth signaling at the Nyquist rate.

A particular partial-response is often identified by the polynomial

$$\sum_{k=0}^{K} h_k D^k,$$

where D (for "delay") takes the place of the usual z^{-1} in the z-Transform of the sequence $\{h_k\}$. For example, duobinary is also referred to as "$1 + D$" partial-response.

In general, more complicated system responses than those shown in Figures 9 and 10 can be generated by choosing more nonzero coefficients in the sequence $\{h_k\}$. However, this complicates detection because of the additional intersymbol interference which is generated.

Rather than modulating a PR pulse $h(t)$, a PR signal can also be generated by filtering the sequence of transmitted levels $\{A_i\}$. This is shown in Figure 11. Namely, the transmitted levels are first passed through a *discrete*-time (digital) filter with transfer function $P_d(e^{j2\pi fT})$ (where the subscript "d" stands for "discrete"). (Note that $P_d(e^{j2\pi fT})$ can be selected to be $H_{eq}(e^{j2\pi fT})$.) The outputs of this filter form the PAM signal, where the pulse shaping filter $P(f) = 1$, $|f| < 1/(2T)$, and is zero elsewhere. If the transmitted levels $\{A_k\}$ are selected independently and are identically distributed, then the transmitted spectrum is $\sigma_A^2 |P_d(e^{j2\pi fT})|^2$ for $|f| < 1/(2T)$ and is zero for $|f| > 1/(2T)$, where $\sigma_A^2 = E[|A_k|^2]$.

Shaping the transmitted spectrum to have nulls coincident with nulls in the channel response potentially offers significant performance advantages. However, by introducing intersymbol interference, PR signaling increases the number of received signal levels, which

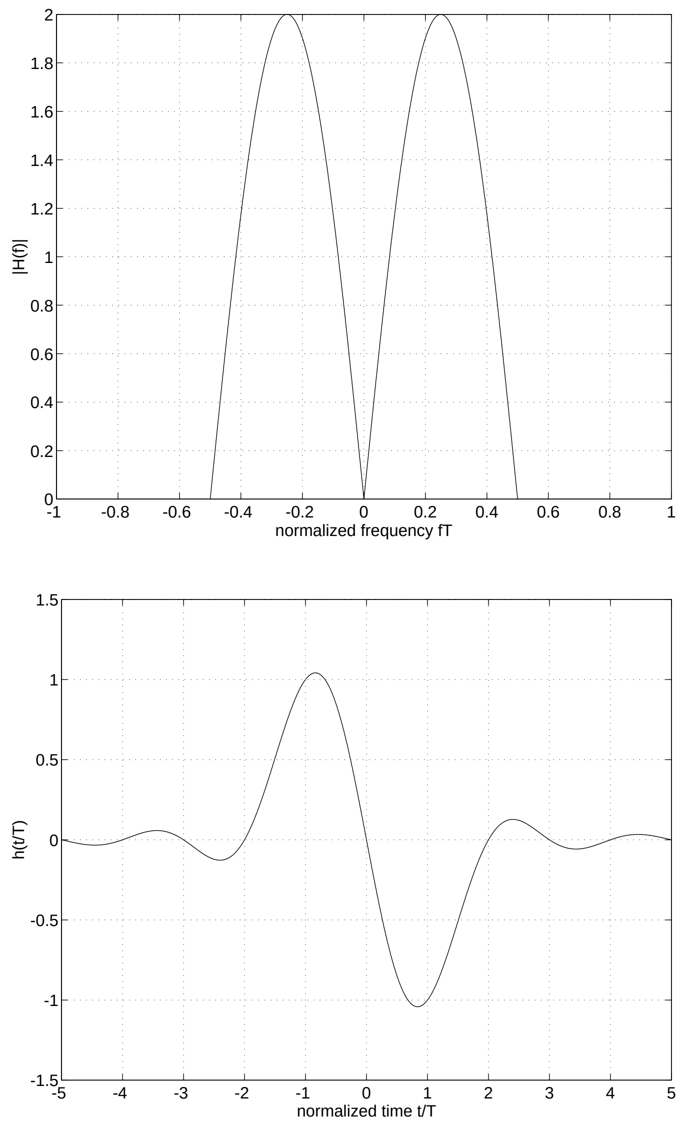

Figure 10. Modified duobinary frequency response and minimum bandwidth pulse.

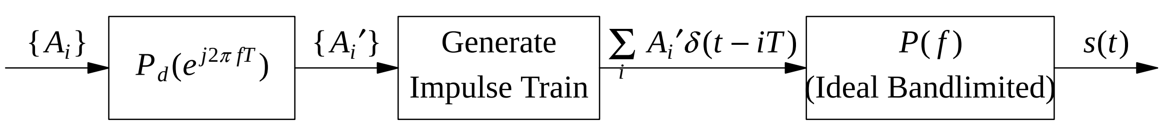

Figure 11. Generation of PR signal.

increases the complexity of the detector, and may reduce immunity to noise. For example, the set of received signal levels for duobinary signaling is $\{0, \pm 2\}$ from which the transmitted levels $\{\pm 1\}$ must be estimated. The performance of a particular PR scheme depends on the channel characteristics as well as the type of detector used at the receiver. We now describe a simple sub-optimal detection strategy.

Precoding

Consider the received signal sample (27) with duobinary signaling. If the receiver has correctly decoded the symbol A_{k-1}, then in the absence of noise A_k can be decoded by subtracting A_{k-1} from the received sample y_k. However, if an error occurs, then subtracting the preceding symbol estimate from the received sample will cause the error to propogate to successive detected symbols. To avoid this problem, the transmitted levels can be **precoded** in such a way as to compensate for the intersymbol interference introduced by the overall partial-response.

We first illustrate precoding for duobinary PR. The sequence of operations is illustrated in Figure 12. Let $\{b_k\}$ denote the sequence of source bits where $b_k \in \{0, 1\}$. This sequence is transformed to the sequence $\{b_k'\}$ by the operation

$$b_k' = b_k \oplus b_{k-1}', \tag{33}$$

where "$\oplus$" denotes modulo 2 addition (exclusive OR). The sequence $\{b_k'\}$ is mapped to the sequence of binary transmitted signal levels $\{A_k\}$ according to

$$A_k = 2b_k' - 1 . \tag{34}$$

That is, $b_k' = 0$ ($b_k = 1$) is mapped to the transmitted level $A_k = -1$ ($A_k = 1$). In the absence of noise the received symbol is then

$$y_k = A_k + A_{k-1} = 2(b_k' + b_{k-1}' - 1), \tag{35}$$

and combining (33) and (35) gives

$$b_k = (\tfrac{1}{2} y_k + 1) \bmod 2 . \tag{36}$$

That is, if $y_k = \pm 2$ then $b_k = 0$, and if $y_k = 0$ then $b_k = 1$. Precoding therefore enables the detector to make *symbol-by-symbol* decisions that do not depend on previous decisions. Table I shows a sequence of transmitted bits $\{b_i\}$, precoded bits $\{b_i'\}$, transmitted signal levels $\{A_i\}$, and received samples $\{y_i\}$.

The preceding precoding technique can be extended to multi-level PAM, and to other PR channels. Suppose that the PR is specified by $H_{eq}(D) = \sum_{k=0}^{K} h_k D^k$ where the coefficients are integers, and that the source symbols $\{b_k\}$ are selected from the set $\{0, 1, \ldots, M-1\}$.

Figure 12. Precoding for a PR channel.

$\{b_i\}$:		1	0	0	1	1	1	0	0	1	0
$\{b_i'\}$:	0	1	1	1	0	1	0	0	0	1	1
$\{A_i\}$:	-1	1	1	1	-1	1	-1	-1	-1	1	1
$\{y_i\}$:		0	2	2	0	0	0	-2	-2	0	2

Table I. Example of precoding for duobinary PR. What is shown are the source bits $\{b_i\}$, precoded bits $\{b_i'\}$, transmitted levels $\{A_i\}$, and received samples $\{y_i\}$.

These symbols are transformed to the sequence $\{b_k'\}$ via the precoding operation

$$b_k' = \left(b_k - \sum_{i=1}^{K} h_i b_{k-i}' \right) \mod M \tag{37}$$

Because of the modulo operation, each symbol b_k' is also in the set $\{0, 1, \ldots, M-1\}$. The kth transmitted signal level is given by

$$A_k = 2b_k' - (M-1) \tag{38}$$

so that the set of transmitted levels is $\{-(M-1), \ldots, (M-1)\}$ (i.e., a shifted version of the set of values assumed by b_k). In the absence of noise the received sample is

$$y_k = \sum_{i=0}^{K} h_i A_{k-i}, \tag{39}$$

and it can be shown that the kth source symbol is given by

$$b_k = \tfrac{1}{2}\left(y_k + (M-1) \cdot H_{eq}(1) \right) \mod M \tag{40}$$

Precoding the symbols $\{b_k\}$ in this manner therefore enables symbol-by-symbol decisions at the receiver. In the presence of noise, more sophisticated detection schemes (e.g., maximum-likelihood) can be used with PR signaling to obtain improvements in performance.

6. Additional Considerations

In many applications, bandwidth and intersymbol interference are not the only important considerations for selecting baseband pulses. Here we give a brief discussion of additional practical constraints that may influence this selection.

Average Transmitted Power and Spectral Constraints

The constraint on average transmitted power varies according to the application. For example, low average power is highly desirable for mobile wireless applications which use battery-powered transmitters. In many applications (e.g., digital subscriber loops as well as digital radio), constraints are imposed to limit the amount of interference, or crosstalk, radiated into neighboring receivers and communications systems. Because this type of interference is frequency dependent, the constraint may take the form of a ‘‘spectral mask’’, which specifies the maximum allowable transmitted power as a function of frequency. For example, crosstalk in wireline channels is generally caused by capacitive coupling, and *increases* as a function of frequency. Consequently, to reduce the amount of crosstalk generated at a particular transmitter, the pulse shaping filter generally attenuates high frequencies more than low frequencies.

In radio applications where signals are assigned different frequency bands, constraints on the transmitted spectrum are imposed to limit *adjacent-channel interference*. This interference is generated by transmitters assigned to adjacent frequency bands. A constraint is therefore needed to limit the amount of *out-of-band power* generated by each transmitter, in addition to an overall average power constraint. To meet this constraint, the transmitter filter in Figure 3 must have a sufficiently steep rolloff at the edges of the assigned frequency band. (Conversely, if the transmitted signals are *time-multiplexed*, then the duration of the system impulse response must be contained within the assigned time slot.)

Peak-to-Average Power

In addition to a constraint on average transmitted power, a *peak*-power constraint is often imposed as well. This constraint is important in practice for the following reasons:

1. The dynamic range of the transmitter is limited. In particular, saturation of the output amplifier will ‘‘clip’’ the transmitted waveform.

2. Rapid fades can severely distort signals with high peak-to-average power.

3. The transmitted signal may be subjected to nonlinearities. Saturation of the output amplifier is one example. Another example that pertains to wireline applications is the companding process in the voice telephone network [3]. Namely, the compander used to reduce quantization noise for pulse-code modulated voice signals introduces amplitude-dependent distortion in data signals.

The preceding impairments or constraints indicate that the transmitted waveform should have a low peak-to-average power ratio. The peak-to-average power ratio is minimized by using binary signaling with rectangular pulse shapes. However, this compromises bandwidth efficiency. In applications where peak-to-average ratio should be low, binary signaling with ‘‘rounded’’ pulses are often used.

Channel and Receiver Characteristics

The type of channel impairments encountered, and the type of detection scheme used at the receiver can also influence the choice of a transmitted pulse shape. For example, a constant amplitude pulse is appropriate for a fast fading environment with noncoherent detection. The ability to track channel characteristics, such as phase, may allow more bandwidth efficient pulse shapes in addition to multi-level signaling.

High-speed data communications over time-varying channels requires that the transmitter and/or receiver *adapt* to the changing channel characteristics. Adapting the transmitter to compensate for a time-varying channel requires a feedback channel through which the receiver can notify the transmitter of changes in channel characteristics. Because of this extra complication, adapting the receiver is often preferred to adapting the transmitter pulse shape. However, the following examples are notable exceptions.

1. The current IS-95 air interface for Direct-Sequence Code-Division Multiple-Access adapts the transmitter power to control the amount of interference generated, and to compensate for channel fades. This can be viewed as a simple form of adaptive transmitter pulse shaping in which a single parameter associated with the pulse shape is varied.

2. Multi-tone modulation divides the channel bandwidth into small subbands, and the transmitted power and source bits are distributed among these subbands to maximize the information rate. The received signal-to-noise ratio for each subband must be transmitted back to the transmitter to guide the allocation of transmitted bits and power [4].

In addition to multi-tone modulation, *adaptive precoding* (also known as *Tomlinson-Harashima* precoding [5],[6]) is another way in which the transmitter can adapt to the channel frequency response. Adaptive precoding is an extension of the technique described earlier for partial-response channels. Namely, the equivalent discrete-time channel impulse response is measured at the receiver, and sent back to the transmitter where it is used in a precoder. The precoder compensates for the intersymbol interference introduced by the channel, allowing the receiver to detect the data by a simple threshhold operation. Both multi-tone modulation and precoding have been used with wireline channels (voiceband modems and digital subscriber loops).

Complexity

Generation of a bandwidth-efficient signal requires a filter with a sharp cutoff. In addition, bandwidth-efficient pulse shapes can complicate other system functions such as timing and carrier recovery. If sufficient bandwidth is available, the cost can be reduced by using a rectangular pulse shape with a simple detection strategy (low-pass filter and threshold).

Tolerance to Interference

Interference is one of the primary channel impairments associated with digital radio. In addition to adjacent-channel interference described earlier, *co-channel interference* may be generated by other transmitters assigned to the *same* frequency band as the desired signal. Co-channel interference can be controlled through frequency (and perhaps time-slot) assignments and by pulse shaping. For example, assuming fixed average power, increasing the bandwidth occupied by the signal lowers the power spectral density, and decreases the amount of interference into a narrowband system that occupies part of the available bandwidth. Sufficient bandwidth spreading therefore enables wideband signals to be overlaid on top of narrowband signals without disrupting either service.

Probability of Intercept and Detection

The broadcast nature of wireless channels generally makes eavesdropping easier than for wired channels. A requirement for most commercial, as well as military applications is to guarantee the privacy of user conversations (low probability of intercept). An additional requirement, in some applications, is that determining whether or not communications is taking place must be difficult (low probability of detection). Spread spectrum waveforms are attractive in these applications since spreading the pulse energy over a wide frequency band decreases the power spectral density, and hence makes the signal less ‘‘visible’’. Power-efficient modulation combined with coding enables a further reduction in transmitted power for a target error rate.

7. Examples

We conclude this chapter with a brief description of baseband pulse shapes used in existing and emerging standards for digital mobile cellular and Personal Communications Services (PCS).

7.1 Global System for Mobile Communications (GSM)

The European GSM standard for digital mobile cellular communications operates in the 900 MHz frequency band, and is based on Time-Division Multiple-Access (TDMA) [7]. A special variant of binary FSK is used called *Gaussian Minimum-Shift Keying (GMSK)*. The GMSK modulator is illustrated in Figure 13. The input to the modulator is a binary PAM signal $s(t)$, given by (3), where the pulse $p(t)$ is a Gaussian function, and $|s(t)| < 1$. This waveform frequency modulates the carrier f_c, so that the (passband) transmitted signal is

$$w(t) \; = \; K \cos\left(2\pi f_c t + 2\pi f_d \int_{-\infty}^{t} s(\tau) d\tau \right).$$

The maximum frequency deviation from the carrier is $f_d = 1/(2T)$, which characterizes Minimum-Shift Keying. This technique can be used with a noncoherent receiver, which is easy to implement. Because the transmitted signal has a constant envelope, the data can be reliably detected in the presence of rapid fades, which are characteristic of mobile radio channels.

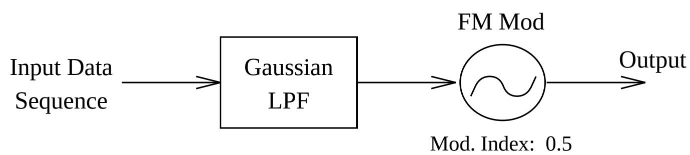

Figure 13. Generation of GMSK signal (LPF is Low-Pass Filter).

7.2 U.S. Digital Cellular (IS-54)

The IS-54 air interface operates in the 800 MHz band, and is based on TDMA [8]. The baseband signal is given by (3) where the symbols are complex-valued, corresponding to quadrature phase modulation. The pulse has a square-root raised cosine spectrum with 35% excess bandwidth.

7.3 Interim Standard-95

The IS-95 air interface for digital mobile cellular uses spread-spectrum signaling (CDMA) in the 800 MHz band [9]. The baseband transmitted pulse shapes are analogous to those shown in Figure 2, where the number of square pulses (“chips”) per bit is 128. To improve spectral efficiency the (wideband) transmitted signal is filtered by an approximation to an ideal low-pass response with a small amount of excess bandwidth. This shapes the chips so that they resemble minimum bandwidth pulses.

7.4 Personal Access Communications System (PACS)

Both PACS and the Japanese Personal Handi-Phone (PHP) system are TDMA systems which have been proposed for PCS, and operate near 2 Ghz [10]. The baseband signal is given by (3) with four complex symbols representing 4-phase quadrature modulation. The baseband pulse has a square-root raised cosine spectrum with 50% excess bandwidth.

References

1. A. Lender, ''The Duobinary Technique for High-Speed Data Transmission'', *AIEE Trans. on Commun. Electronics,* vol. 82, pp. 214-218, March 1963.

2. E. R. Kretzmer, ''Generalization of a Technique for Binary Data Communication'', *IEEE Tran. Commun. Tech.,* vol. COM-14, pp. 67-68, Feb. 1966.

3. I. Kalet and B. R. Saltzberg, ''QAM Transmission Through a Companding Channel − Signal Constellations and Detection'', *IEEE Trans. on Communications,* Vol. 42, No. 2-4, pp. 417-429, Feb.-April 1994.

4. J. A. C. Bingham, ''Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come'', *IEEE Commun. Magazine,* vol. 28, pp. 5-14, May 1990.

5. M. Tomlinson, ''New Automatic Equalizer Employing Modulo Arithmetic'', *Electron. Lett., vol. 7, pp. 138-139, March 1971.*

6. H. Harashima and H. Miyakawa, ''Matched-Transmission Technique for Channels With Intersymbol Interference,'', *IEEE Trans. Commun.,* vol. COM-20, pp. 774-780, Aug. 1972.

7. M. Rahnema, "Overview of the GSM System and Protocol Architecture," *IEEE Commun. Mag.,* pp. 92-100, Apr. 1993.

8. "Recommended Minimum Performance Standards for 800 MHz Dual-Mode Mobile Stations", (incorporating EIA/TIA 19B), EIA/TIA Project Number 2216, March 1991.

9. TIA/EIA/IS-95 "Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System," Telecommunication Industry Association, July 1993.

10. D. C. Cox, "Wireless Personal Communications: What Is It?", *IEEE Personal Communications,* Vol. 2, No. 2, pp. 20-35, April 1995.

Defining Terms

Baseband Signal: a signal with frequency content centered around DC.

Equivalent Discrete-Time Transfer Function: a discrete-time transfer function (z-Transform) which relates the transmitted amplitudes to received samples in the absence of noise.

Excess Bandwidth: that part of the baseband transmitted spectrum which is not contained within the Nyquist band.

Eye Diagram: superposition of segments of a received PAM signal, which indicates the amount of intersymbol interference present.

Frequency-Shift Keying: a digital modulation technique in which the transmitted pulse is sinusoidal, where the frequency is determined by the source bits.

Intersymbol Interference: the additive contribution (interference) to a received sample from transmitted symbols other than the symbol to be detected.

Matched-Filter: the receiver filter with impulse response equal to the time-reversed, complex conjugate impulse response of the combined transmitter filter-channel impulse response.

Nyquist Band: the narrowest frequency band that can support a PAM signal without intersymbol interference (the interval $[-1/(2T),\ 1/(2T)]$ where $1/T$ is the symbol rate).

Nyquist Criterion: a condition on the overall frequency response of a PAM system that ensures the absence of intersymbol interference.

Partial-Response Signaling: a signaling technique in which a controlled amount of intersymbol interference is introduced at the transmitter in order to shape the transmitted spectrum.

Precoding: a transformation of source symbols at the transmitter that compensates for intersymbol interference introduced by the channel.

Raised Cosine pulse: a pulse shape with Fourier Transform that decays to zero according to a raised cosine (see (18)). The amount of excess bandwidth is conveniently determined by a single parameter (α).

Pulse Amplitude Modulation (PAM): a digital modulation technique in which the source bits are mapped to a sequence of amplitudes that modulate a transmitted pulse.

Spread-spectrum: a signaling technique in which the pulse bandwidth is many times wider than the Nyquist bandwidth.

Zero-forcing criterion: a design constraint which specifies that intersymbol interference be eliminated.

For Further Information:

Baseband signaling and pulse shaping is fundamental to the design of any digital communications system, and is therefore covered in numerous texts on digital communications. For more advanced treatments see E. A. Lee and D. G. Messerschmitt, *Digital Communication*, Kluwer 1994, and J. G. Proakis, *Digital Communications*, McGraw-Hill 1995.