Signal Design for Band-Limited Channels

Wireless Information Transmission System Lab.
Institute of Communications Engineering
National Sun Yat-sen University
We consider the problem of signal design when the channel is band-limited to some specified bandwidth of $W$ Hz.

The channel may be modeled as a linear filter having an equivalent low-pass frequency response $C(f)$ that is zero for $|f| > W$.

Our purpose is to design a signal pulse $g(t)$ in a linearly modulated signal, represented as

$$v(t) = \sum_n I_n g(t - nT)$$

that efficiently utilizes the total available channel bandwidth $W$.

When the channel is ideal for $|f| \leq W$, a signal pulse can be designed that allows us to transmit at symbol rates comparable to or exceeding the channel bandwidth $W$.

When the channel is not ideal, signal transmission at a symbol rate equal to or exceeding $W$ results in inter-symbol interference (ISI) among a number of adjacent symbols.
Fig. (a) is a band-limited pulse having zeros periodically spaced in time at $\pm T$, $\pm 2T$, etc.

If information is conveyed by the pulse amplitude, as in PAM, for example, then one can transmit a sequence of pulses, each of which has a peak at the periodic zeros of the other pulses.
Characterization of Band-Limited Channels

- However, transmission of the pulse through a channel modeled as having a linear envelope delay $\tau(f)$ [quadratic phase $\theta(f)$] results in the received pulse shown in Fig. (b), where the zero-crossings that are no longer periodically spaced.
A sequence of successive pulses would no longer be distinguishable. Thus, the channel delay distortion results in ISI.

It is possible to compensate for the nonideal frequency-response of the channel by use of a filter or equalizer at the demodulator.

Fig. (c) illustrates the output of a linear equalizer that compensates for the linear distortion in the channel.
The equivalent low-pass transmitted signal for several different types of digital modulation techniques has the common form

\[ v(t) = \sum_{n=0}^{\infty} I_n g(t - nT) \]

where \( \{I_n\} \): discrete information-bearing sequence of symbols. 

\( g(t) \): a pulse with band-limited frequency-response \( G(f) \), i.e., \( G(f) = 0 \) for \( |f| > W \).

This signal is transmitted over a channel having a frequency response \( C(f) \), also limited to \( |f| \leq W \).

The received signal can be represented as

\[ r(t) = \sum_{n=0}^{\infty} I_n h(t - nT) + z(t) \]

where \( h(t) = \int_{-\infty}^{\infty} g(\tau)c(t - \tau)d\tau \) and \( z(t) \) is the AWGN.
Suppose that the received signal is passed first through a filter and then sampled at a rate $1/T$ samples/s, the optimum filter from the point of view of signal detection is one matched to the received pulse. That is, the frequency response of the receiving filter is $H^*(f)$.

We denote the output of the receiving filter as

$$y(t) = \sum_{n=0}^{\infty} I_n x(t - nT) + v(t)$$

where

$x(t)$: the pulse representing the response of the receiving filter to the input pulse $h(t)$.

$v(t)$: response of the receiving filter to the noise $z(t)$.  

If \( y(t) \) is sampled at times \( t = kT + \tau_0, k = 0, 1, \cdots \), we have
\[
y(kT + \tau_0) = y_k = \sum_{n=0}^{\infty} I_n x(kT - nT + \tau_0) + v(kT + \tau_0)
\]
or, equivalently,
\[
y_k = \sum_{n=0}^{\infty} I_n x_{k-n} + v_k, \quad k = 0, 1, \ldots
\]
where \( \tau_0 \): transmission delay through the channel.

The sample values can be expressed as
\[
y_k = x_0 \left( \frac{1}{x_0} \sum_{n=0}^{\infty} I_n x_{k-n} \right) + v_k, \quad k = 0, 1, \ldots
\]
We regard $x_0$ as an arbitrary scale factor, which we arbitrarily set equal to unity for convenience, then

$$y_k = I_k + \sum_{n=0}^{\infty} I_n x_{k-n} + \nu_k$$

where

$I_k$: the desired information symbol at the $k$-th sampling instant.

$$\sum_{n=0}^{\infty} I_n x_{k-n} : \text{ISI}$$

$\nu_k$: additive Gaussian noise variable at the $k$-th sampling instant.
The amount of ISI and noise in a digital communication system can be viewed on an oscilloscope.

For PAM signals, we can display the received signal $y(t)$ on the vertical input with the horizontal sweep rate set at $1/T$.

The resulting oscilloscope display is called an **eye pattern**.

Eye patterns for binary and quaternary amplitude-shift keying:

- The effect of ISI is to cause the eye to close.
- Thereby, reducing the margin for additive noise to cause errors.
Effect of ISI on eye opening:

- ISI distorts the position of the zero-crossings and causes a reduction in the eye opening.
- Thus, it causes the system to be more sensitive to a synchronization error.
For PSK and QAM, it is customary to display the “eye pattern” as a two-dimensional scatter diagram illustrating the sampled values \( \{y_k\} \) that represent the decision variables at the sampling instants.

Two-dimensional digital “eye patterns.”
Assuming that the band-limited channel has ideal frequency-response, i.e., \( C(f) = 1 \) for \( |f| \leq W \), then the pulse \( x(t) \) has a spectral characteristic \( X(f) = |G(f)|^2 \), where

\[
x(t) = \int_{-W}^{W} X(f) e^{j2\pi ft} df
\]

We are interested in determining the spectral properties of the pulse \( x(t) \), that results in no inter-symbol interference.

Since

\[
y_k = I_k + \sum_{n=0}^{\infty} I_n x_{k-n} + v_k
\]

the condition for no ISI is

\[
x(t = kT) \equiv x_k = \begin{cases} 
1 & (k = 0) \\
0 & (k \neq 0) 
\end{cases} \quad (\ast)
\]
Nyquist pulse-shaping criterion (Nyquist condition for zero ISI)

The necessary and sufficient condition for $x(t)$ to satisfy

$$x(nT) = \begin{cases} 1 & (n = 0) \\ 0 & (n \neq 0) \end{cases}$$

is that its Fourier transform $X(f)$ satisfy

$$\sum_{m=-\infty}^{\infty} X(f + m/T) = T$$
Proof:

- In general, $x(t)$ is the inverse Fourier transform of $X(f)$. Hence,

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df$$

- At the sampling instant $t = nT$,

$$x(nT) = \int_{-\infty}^{\infty} X(f)e^{j2\pi fnT} df$$
Breaking up the integral into integrals covering the finite range of $1/T$, thus, we obtain

$$x(nT) = \sum_{m=-\infty}^{\infty} \int_{(2m-1)/2T}^{(2m+1)/2T} X(f) e^{j2\pi fnT} df$$

$$= \sum_{m=-\infty}^{\infty} \int_{-1/2T}^{1/2T} X(f' + m/T) e^{j2\pi f'nT} df'$$

$$= \int_{-1/2T}^{1/2T} \left[ \sum_{m=-\infty}^{\infty} X(f + m/T) \right] e^{j2\pi fnT} df$$

$$= \int_{-1/2T}^{1/2T} B(f) e^{j2\pi fnT} df$$

(1)

where we define $B(f)$ as $B(f) = \sum_{m=-\infty}^{\infty} X(f + m/T)$
Design of Band-Limited Signals for No ISI
The Nyquist Criterion

- Obviously $B(f)$ is a periodic function with period $1/T$, and, therefore, it can be expanded in terms of its Fourier series coefficients $\{b_n\}$ as

$$B(f) = \sum_{n=-\infty}^{\infty} b_n e^{j2\pi nfT} \quad (2)$$

where

$$b_n = T \int_{-1/2T}^{1/2T} B(f) e^{-2\pi nfT} df \quad (3)$$

- Comparing (1) and (3), we obtain

$$x(-nT) = \int_{-1/2T}^{1/2T} B(f) e^{-j2\pi fnT} df = \frac{1}{T} b_n$$

$$b_n = Tx(-nT) \quad (4)$$

Recall that the conditions for no ISI are (from *):

$$x(t = kT) \equiv x_k = \begin{cases} 1 & (k = 0) \\ 0 & (k \neq 0) \end{cases}$$
Therefore, the necessary and sufficient condition for

\[ y_k = I_k + \sum_{n=0}^{\infty} I_n x_{k-n} + v_k \]

to be satisfied is that

\[ b_n = \begin{cases} 
T & (n = 0) \\
0 & (n \neq 0) 
\end{cases} \]

which, when substituted into \( B(f) = \sum_{n=-\infty}^{\infty} b_n e^{j2\pi nfT} \), yields

\[ B(f) = T \]

or, equivalently \( \sum_{m=-\infty}^{\infty} X(f + m/T) = T \)

This concludes the proof of the theorem.
Suppose that the channel has a bandwidth of $W$. Then $C(f) \equiv 0$ for $|f| > W$ and $X(f) = 0$ for $|f| > W$.

When $T < 1/2W$ (or $1/T > 2W$),

Since $B(f) = \sum_{-\infty}^{+\infty} X(f + n/T)$ consists of nonoverlapping replicas of $X(f)$, separated by $1/T$, there is no choice for $X(f)$ to ensure $B(f) \equiv T$ in this case and there is no way that we can design a system with no ISI.
Design of Band-Limited Signals for No ISI
The Nyquist Criterion

◊ When \( T = 1/2W \), or \( 1/T = 2W \) (the Nyquist rate), the replications of \( X(f) \), separated by \( 1/T \), are shown below:

\[
\sum_{n=-\infty}^{\infty} X(f+n/T)
\]

◊ In this case, there exists only one \( X(f) \) that results in \( B(f) = T \), namely,

\[
X(f) = \begin{cases} 
T & (|f| < W) \\
0 & (\text{otherwise})
\end{cases}
\]

which corresponds to the pulse

\[
x(t) = \sin\left(\frac{\pi t}{T}\right) \equiv \text{sinc}\left(\frac{\pi t}{T}\right)
\]
The smallest value of $T$ for which transmission with zero ISI is possible is $T = 1/2W$, and for this value, $x(t)$ has to be a sinc function.

The difficulty with this choice of $x(t)$ is that it is noncausal and nonrealizable.

A second difficulty with this pulse shape is that its rate of convergence to zero is slow.

The tails of $x(t)$ decay as $1/t$; consequently, a small mistiming error in sampling the output of the matched filter at the demodulator results in an infinite series of ISI components.

Such a series is not absolutely summable because of the $1/t$ rate of decay of the pulse, and, hence, the sum of the resulting ISI does not converge.

\[
1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = ?
\]

\[
1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots = ?
\]
When $T > 1/2W$ (or $1/T < 2W$), $B(f)$ consists of overlapping replications of $X(f)$ separated by $1/T$:

\[
\sum_{n=-\infty}^{\infty} X(f + n/T)
\]

In this case, there exist numerous choices for $X(f)$ such that $B(f) \equiv T$. 
A particular pulse spectrum, for the $T > 1/2W$ case, that has desirable spectral properties and has been widely used in practice is the \textit{raised cosine spectrum}.

\begin{align}
X_{rc}(f) &= \begin{cases}
T & \text{for } 0 \leq |f| \leq \frac{1-\beta}{2T} \\
\frac{T}{2} \left[ 1 + \cos \left( \frac{\pi T}{\beta} \left( |f| - \frac{1-\beta}{2T} \right) \right) \right] & \text{for } \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \\
0 & \text{for } |f| > \frac{1+\beta}{2T}
\end{cases}
\end{align}

$\beta$: roll-off factor. ($0 \leq \beta \leq 1$)
The bandwidth occupied by the signal beyond the Nyquist frequency $1/2T$ is called the *excess bandwidth* and is usually expressed as a percentage of the Nyquist frequency.

- $\beta = 1/2 \Rightarrow$ excess bandwidth = 50%.
- $\beta = 1 \Rightarrow$ excess bandwidth = 100%.

The pulse $x(t)$, having the raised cosine spectrum, is

$$x(t) = \frac{\sin(\pi t / T)}{\pi t / T} \frac{\cos(\pi \beta t / T)}{1 - 4 \beta^2 t^2 / T^2}$$

$$= \sin c(\pi t / T) \frac{\cos(\pi \beta t / T)}{1 - 4 \beta^2 t^2 / T^2}$$

- $x(t)$ is normalized so that $x(0) = 1$. 
Pulses having a raised cosine spectrum:

- For $\beta = 0$, the pulse reduces to $x(t) = \text{sinc}(\pi t/T)$, and the symbol rate $1/T = 2W$.
- When $\beta = 1$, the symbol rate is $1/T = W$. 
Design of Band-Limited Signals for No ISI
The Nyquist Criterion

- In general, the tails of $x(t)$ decay as $1/t^3$ for $\beta > 0$.
- Consequently, a mistiming error in sampling leads to a series of ISI components that converges to a finite value.
- Because of the smooth characteristics of the raised cosine spectrum, it is possible to design practical filters for the transmitter and the receiver that approximate the overall desired frequency response.
- In the special case where the channel is ideal, i.e., $C(f) = 1$, $|f| \leq W$, we have
  \[ X_{rc}(f) = G_T(f)G_R(f) \]
  where $G_T(f)$ and $G_R(f)$ are the frequency responses of the two filters.
If the receiver filter is matched to the transmitter filter, we have
\[ X_{rc}(f) = G_T(f) \quad G_R(f) = |G_T(f)|^2. \]
Ideally,
\[ G_T(f) = \sqrt{|X_{rc}(f)|} e^{-j2\pi f t_0} \]
and \( G_R(f) = G_T^*(f) \), where \( t_0 \) is some nominal delay that is required to ensure physical realizability of the filter.

Thus, the overall raised cosine spectral characteristic is split evenly between the transmitting filter and the receiving filter.

An additional delay is necessary to ensure the physical realizability of the receiving filter.
Square Root Raised Cosine Filter

- The cosine roll-off transfer function can be achieved by using identical square root raised cosine filter $\sqrt{X_{rc}(f)}$ at the transmitter and receiver.
- The pulse $\text{SRRC}(t)$, having the square root raised cosine spectrum, is

$$
\text{SRRC}(t) = \frac{\sin \left( \pi \frac{t}{T_C} (1 - \beta) \right) + 4\beta \frac{t}{T_C} \cos \left( \pi \frac{t}{T_C} (1 + \beta) \right)}{\pi \frac{t}{T_C} \left(1 - \left(4\beta \frac{t}{T_C}\right)^2\right)}
$$

where $T_C$ is the inverse of chip rate ($\approx 0.2604167\mu s$) and $\beta = 0.22$ for WCDMA.
Design of Band-limited Signals with Controlled ISI
-- Partial-Response Signals

◊ It is necessary to reduce the symbol rate $1/T$ below the Nyquist rate of $2W$ symbols/s to realize practical transmitting and receiving filters.

◊ Suppose we choose to relax the condition of zero ISI and, thus, achieve a symbol transmission rate of $2W$ symbols/s.

◊ By allowing for a controlled amount of ISI, we can achieve this symbol rate.

◊ The condition for zero ISI is $x(nT)=0$ for $n\neq 0$.

◊ Suppose that we design the band-limited signal to have controlled ISI at one time instant. This means that we allow one additional nonzero value in the samples $\{x(nT)\}$. 
Design of Band-limited Signals with Controlled ISI -- Partial-Response Signals

◊ One special case that leads to (approximately) physically realizable transmitting and receiving filters is the duobinary signal pulse:

\[
x(nT) = \begin{cases} 
1 & (n = 0, 1) \\
0 & \text{(otherwise)} 
\end{cases}
\]

◊ Using Equation (4) in Page 17

\[
b_n = T x(-nT)
\]

\[
b_n = \begin{cases} 
T & (n = 0, -1) \\
0 & \text{(otherwise)} 
\end{cases}
\]

◊ When substituted into Equation (2) in Page 17, we obtain:

\[
B(f) = T + T e^{-j2\pi f T}
\]
It is impossible to satisfy the above equation for $1/T > 2W$.

For $T = 1/2W$, we obtain

$$X(f) = \begin{cases} \frac{1}{2W} \left(1 + e^{-j\pi f/W}\right) & (|f| < W) \\ 0 & \text{(otherwise)} \end{cases}$$

This pulse is called a *duobinary signal pulse*.
Time-domain and frequency-domain characteristics of a duobinary signal.

- **Modified duobinary signal pulse:**

\[
x\left(\frac{n}{2W}\right) = x(nT) = \begin{cases} 
1 & (n = -1) \\
-1 & (n = 1) \\
0 & \text{(otherwise)}
\end{cases}
\]
Design of Band-limited Signals with Controlled ISI
-- Partial-Response Signals

- The corresponding pulse $x(t)$ is given as

$$x(t) = \text{sinc} \frac{\pi (t + T)}{T} - \text{sinc} \frac{\pi (t - T)}{T}$$

- The spectrum is given by

$$X(f) = \begin{cases} 
\frac{1}{2W} (e^{j\pi f/W} - e^{-j\pi f/W}) = \frac{j}{W} \sin \frac{\pi f}{W} & |f| \leq W \\
0 & |f| > W 
\end{cases}$$

Zero at $f=0$. 

$$|X(f)| = \frac{1}{W} \left\{ \begin{array}{ll} 
1 & f = 0 \\
1 \sin \frac{\pi f}{W} & f \neq 0 
\end{array} \right.$$
Other physically realizable filter characteristics are obtained by selecting different values for the samples \( \{x(n/2W)\} \) and more than two nonzero samples.

As we select more nonzero samples, the problem of unraveling the controlled ISI becomes more cumbersome and impractical.

When controlled ISI is purposely introduced by selecting two or more nonzero samples from the set \( \{x(n/2W)\} \), the class of band-limited signal pulses are called \textit{partial-response signals}:

\[
x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2W}\right) \text{sinc}\left[2\pi W \left(t - \frac{n}{2W}\right)\right]
\]

\[
X(f) = \begin{cases} 
\frac{1}{2W} \sum_{n=\infty}^{2W} x\left(\frac{n}{2W}\right) e^{-jnf/W} & (|f| \leq W) \\
0 & (|f| > W)
\end{cases}
\]
Data Detection for Controlled ISI

- Two methods for detecting the information symbols at the receiver when the received signal contains controlled ISI:
  - **Symbol-by-symbol** detection method.
    - Relatively easy to implement.
  - **Maximum-likelihood criterion** for detecting a sequence of symbols.
    - Minimizes the probability of error but is a little more complex to implement.

- The following treatment is based on PAM signals, but it is easily generalized to QAM and PSK.

- We assume that the desired spectral characteristic $X(f)$ for the partial-response signal is split evenly between the transmitting and receiving filters, i.e., $|G_T(f)| = |G_R(f)| = |X(f)|^{1/2}$.
Symbol-by-symbol suboptimum detection

For duobinary signal pulse, \( x(nT) = 1 \), for \( n = 0, 1 \), and is zero otherwise.

The samples at the output of the receiving filter (demodulator) have the form

\[
y_m = B_m + v_m = I_m + I_{m-1} + v_m
\]

where \( \{I_m\} \) is the transmitted sequence of amplitudes and \( \{v_m\} \) is a sequence of additive Gaussian noise samples.

Consider the binary case where \( I_m = \pm 1 \), \( B_m \) takes on one of three possible values, namely, \( B_m = -2, 0, 2 \) with corresponding probabilities \( \frac{1}{4}, \frac{1}{2}, \frac{1}{4} \).

If \( I_{m-1} \) is the detected symbol from the \((m-1)\)th signaling
Symbol-by-symbol suboptimum detection

interval, its effect on $B_m$, the received signal in the $m$th signaling interval, can be eliminated by subtraction, thus allowing $I_m$ to be detected.

Major problem with this procedure is error propagation: if $I_{m-1}$ is in error, its effect on $B_m$ is not eliminated but, in fact, is reinforced by the incorrect subtraction.

Error propagation can be avoided by precoding the data.

The precoding is performed on the binary data sequence prior to modulation.

From the data sequence $\{D_n\}$, the precoded sequence $\{P_n\}$ is given by:

$$P_m = D_m \oplus P_{m-1}, \quad m = 1, 2,...$$
Symbol-by-symbol suboptimum detection

Set $I_m = -1$ if $P_m = 0$ and $I_m = 1$ if $P_m = 1$, i.e., $I_m = 2P_m - 1$.

The noise-free samples at the output of the receiving filter are given by

$$B_m = I_m + I_{m-1} = (2P_m - 1) + (2P_{m-1} - 1) = 2(P_m + P_{m-1} - 1)$$

$$P_m + P_{m-1} = \frac{1}{2} B_m + 1$$

Since $D_m = P_m \oplus P_{m-1}$, it follows that the data sequence $D_m$ is obtained from $B_m$ using the relation:

$$D_m = \frac{1}{2} B_m + 1 \pmod{2}$$

Consequently, if $B_m = \pm 2$, then $D_m = 0$, and if $B_m = 0$, $D_m = 1$. 
Data Detection for Controlled ISI

◊ Symbol-by-symbol suboptimum detection

◊ Binary signaling with duobinary pulses

<table>
<thead>
<tr>
<th>Data sequence ( D_n )</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precoded sequence ( P_n )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Transmitted sequence ( I_n )</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Received sequence ( B_n )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>-2</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Decoded sequence ( D_n )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

◊ The extension from binary PAM to multilevel PAM signaling

◊ The \( M \)-level amplitude sequence \( \{ I_m \} \) results in a noise-free sequence

\[
B_m = I_m + I_{m-1}, \quad m = 1, 2, \ldots
\]
Symbol-by-symbol suboptimum detection which has $2M-1$ possible equally spaced levels.

The amplitude levels are determined from the relation:

$$I_m = 2P_m - (M - 1)$$

where $\{P_m\}$ is the precoded sequence that is obtained from an $M$-level data sequence $\{D_m\}$ according to the relation

$$P_m = D_m \bigodot P_{m-1} \pmod{M}$$

where the possible values of the sequence $\{D_m\}$ are 0, 1, 2, ⋯, $M-1$.

In the absence of noise, the samples at the output is given by:

$$B_m = I_m + I_{m-1} = 2[P_m + P_{m-1} - (M - 1)]$$
Data Detection for Controlled ISI

◊ Symbol-by-symbol suboptimum detection

◊ Hence \( P_m + P_{m-1} = \frac{1}{2} B_m + (M - 1) \)

◊ Since \( D_m = P_m + P_{m-1} \) (mod \( M \)), it follows that

\[
D_m = \frac{1}{2} B_m + (M - 1) \pmod{M}
\]

◊ Four-level signal transmission with duobinary pulses

<table>
<thead>
<tr>
<th>Data sequence ( D_m )</th>
<th>0 0 1 3 1 2 0 3 3 2 0 1 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precoded sequence ( P_m )</td>
<td>0 0 0 1 2 3 3 1 2 1 1 3 2 2</td>
</tr>
<tr>
<td>Transmitted sequence ( I_m )</td>
<td>-3 -3 -3 -1 1 3 3 -1 1 -1 -1 3 1 1</td>
</tr>
<tr>
<td>Received sequence ( B_n )</td>
<td>-6 -6 -4 0 4 6 2 0 0 -2 2 4 2</td>
</tr>
<tr>
<td>Decoded sequence ( D_m )</td>
<td>0 0 1 3 1 2 0 3 3 2 0 1 0</td>
</tr>
</tbody>
</table>

(M=4)
Symbol-by-symbol suboptimum detection

In the case of the modified duobinary pulse, the controlled ISI is specified by the values $x(n/2W)=-1$, for $n=1$, $x(n/2W)=1$, for $n=-1$, and zero otherwise.

The noise-free sampled output from the receiving filter is given as:

$$B_m = I_m - I_{m-2}$$

Where the $M$-level sequence $\{I_m\}$ is obtained by mapping a precoded sequence according to

$$I_m = 2P_m - (M - 1)$$

and

$$P_m = D_m \oplus P_{m-2} \pmod{M}$$
Data Detection for Controlled ISI

◊ Symbol-by-symbol suboptimum detection
  ◊ From these relations, it is easy to show that the detection rule for recovering the data sequence \( \{D_m\} \) from \( \{B_m\} \) in the absence of noise is
  \[
  D_m = \frac{1}{2} B_m \pmod{M}
  \]
  ◊ The precoding of the data at the transmitter makes it possible to detect the received data on a symbol-by-symbol basis without having to look back at previously detected symbols. Thus, error propagation is avoided.
  ◊ The symbol-by-symbol detection rule is **not** the optimum detection scheme for partial-response signals. Nevertheless, it is relatively simple to implement.
Data Detection for Controlled ISI

- Maximum-likelihood Sequence Detection
  - Partial-response waveforms are signal waveforms with memory. This memory is conveniently represented by a trellis.
  - The trellis for the duobinary partial-response signal for binary data transmission is illustrated in the following figure.

- The first number on the left is the new data bit and the number on the right is the received signal level.
Maximum-likelihood Sequence Detection

The duobinary signal has a memory of length $L=1$. In general, for $M$-ary modulation, the number of trellis states is $M^L$.

The optimum maximum-likelihood sequence detector selects the most probable path through the trellis upon observing the received data sequence $\{y_m\}$ at the sampling instants $t=mT$, $m=1,2,\cdots$.

The trellis search is performed by the Viterbi algorithm.

For the class of partial-response signals, the received sequence $\{y_m, 1\leq m \leq N\}$ is generally described statistically by the joint PDF $p(y_N|I_N)$, where $y_N=[y_1 \ y_2 \ \cdots \ y_N]'$ and $I_N=[I_1 \ I_2 \ \cdots \ I_N]'$ and $N>L$. 
Data Detection for Controlled ISI

- **Maximum-likelihood Sequence Detection**
  - When the additive noise is zero-mean Gaussian, \( p(y_N | I_N) \) is a multivariate Gaussian PDF, i.e.,

  \[
  p(y_N | I_N) = \frac{1}{(2\pi \det C)^{N/2}} \exp \left[ -\frac{1}{2} \left( y_N - B_N \right)' C^{-1} \left( y_N - B_N \right) \right]
  \]

  where \( B_N = [B_1 \ B_2 \ \cdots \ B_N]' \) is the mean of the vector \( y_N \) and \( C \) is the \( N \times N \) covariance matrix of \( y_N \).

- The ML sequence detector selects the sequence through the trellis that maximizes the PDF \( p(y_N | I_N) \).

- Taking the natural logarithms of \( p(y_N | I_N) \):

  \[
  \ln p(y_N | I_N) = -\frac{1}{2} N \ln(2\pi \det C) - \frac{1}{2} \left( y_N - B_N \right)' C^{-1} \left( y_N - B_N \right)
  \]
Maximum-likelihood Sequence Detection

Given the received sequence \( \{y_m\} \), the data sequence \( \{I_m\} \) that maximizes \( \ln p(y_N|I_N) \) is identical to the sequence \( \{I_N\} \) that minimizes \( (y_N-B_N)' C^{-1}(y_N-B_N) \), i.e.,

\[
\hat{I}_N = \arg \min_{I_N} \left[ (y_N-B_N)' C^{-1}(y_N-B_N) \right]
\]

The metric computations in the trellis search are complicated by the correlation of the noise samples at the output of the matched filter for the partial-response signal.

In the case of the duobinary signal waveform, the correlation of the noise sequence \( \{v_m\} \) is over two successive signal samples.
Data Detection for Controlled ISI

- Maximum-likelihood Sequence Detection
  - Hence, \( v_m \) and \( v_{m+k} \) are correlated for \( k=1 \) and uncorrelated for \( k>1 \).
  - If we ignore the noise correlation by assuming that \( E(v_m v_{m+k}) = 0 \) for \( k>0 \), the computation can be simplified to

\[
\hat{I}_N = \arg \min_{I_N} \left[ (y_N - B_N)' (y_N - B_N) \right] = \arg \min_{I_N} \left[ \sum_{m=1}^{N} \left( y_m - \sum_{k=0}^{L} x_k I_{m-k} \right)^2 \right]
\]

where

\[
B_m = \sum_{k=0}^{L} x_k I_{m-k}
\]

and \( x_k = x(kT) \) are the sampled values of the partial-response signal waveform.
In this section, we perform the signal design under the condition that the channel distorts the transmitted signal.

We assume that the channel frequency-response \( C(f) \) is known for \(|f| \leq W\) and that \( C(f) = 0 \) for \(|f| > W\).

The filter responses \( G_T(f) \) and \( G_R(f) \) may be selected to minimize the error probability at the detector.

The additive channel noise is assumed to be Gaussian with power spectral density \( \Phi_{nn}(f) \).
Signal Design for Channels with Distortion

- For the signal component at the output of the demodulator, we must satisfy the condition

\[
G_T(f)C(f)G_R(f) = X_d(f)e^{-j2\pi ft_0}, \quad |f| \leq W
\]

where \(X_d(f)\) is the desired frequency response of the casecade of the modulator, channel, and demodulator, and \(t_0\) is a time delay that is necessary to ensure the physical realizability of the modulation and demodulation filter.

- The desired frequency response \(X_d(f)\) may be selected to yield either zero ISI or controlled ISI at the sampling instants.

- We shall consider the case of zero ISI by selecting \(X_d(f) = X_{rc}(f)\), where \(X_{rc}(f)\) is the raised cosine spectrum with an arbitrary roll-off factor.
The noise at the output of the demodulation filter may be expressed as
\[ v(t) = \int_{-\infty}^{\infty} n(t-\tau) g_R(\tau) d\tau \]
where \( n(t) \) is the input to the filter.

Since \( n(t) \) is zero-mean Gaussian, \( v(t) \) is zero-mean Gaussian, with a power spectral density
\[ \Phi_{vv}(f) = \Phi_{nn}(f) |G_R(f)|^2 \]

For simplicity, we consider binary PAM transmission. Then, the sampled output of the matched filter is
\[ y_m = x_0 I_m + v_m = I_m + v_m \]
where \( x_0 \) is normalized to unity, \( I_m = \pm d \), and \( v_m \) represents the noise term.
Signal Design for Channels with Distortion

- $v_m$ is a zero-mean Gaussian with variance

$$\sigma_v^2 = \int_{-\infty}^{\infty} \Phi_{mm}(f) |G_R(f)|^2 df$$

- The error probability is given by

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{d/\sigma_v}^{\infty} e^{-y^2/2} dy = Q\left(\sqrt{\frac{d^2}{\sigma_v^2}}\right)$$

- The probability of error is minimized by maximizing the ratio $d^2/\sigma_v^2$.

- There are two possible solutions for the case in which the additive Gaussian noise is white so that $\Phi_{nn}(f) = N_0/2$.

- 1st solution: pre-compensate for the total channel distortion at the transmitter, so that the filter at the receiver is matched to the received signal.
Signal Design for Channels with Distortion

- The transmitter and receiver filters have the magnitude characteristics

\[ |G_T(f)| = \frac{\sqrt{X_{rc}(f)}}{|C(f)|}, \quad |f| \leq W \]  \hspace{1cm} (5)

\[ |G_R(f)| = \sqrt{X_{rc}(f)}, \quad |f| \leq W \]

- The phase characteristic of the channel frequency response \( C(f) \) may also be compensated at the transmitter filter.

- For these filter characteristics, the average transmitted power is

\[ P_{av} = \frac{E(I_m^2)}{T} \int_{-W}^{W} g_T^2(t) \, dt = \frac{d^2}{T} \int_{-W}^{W} |G_T(f)|^2 \, df = \frac{d^2}{T} \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|^2} \, df \]
Hence,

\[ d^2 = P_{av} T \left[ \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|^2} df \right]^{-1} \]  

(6)

The noise at the output of the receiver filter is \( \sigma_v^2 = N_0/2 \) and, hence, the SNR at the detector is

\[ \frac{d^2}{\sigma_v^2} = \frac{2P_{av} T}{N_0} \left[ \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|^2} df \right]^{-1} \]
2nd solution: As an alternative, suppose we split the channel compensation equally between the transmitter and receiver filters, i.e.,

\[
|G_T (f)| = \frac{\sqrt{X_{re} (f)}}{|C(f)|^{1/2}}, \quad |f| \leq W
\]

\[
|G_R (f)| = \frac{\sqrt{X_{re} (f)}}{|C(f)|^{1/2}}, \quad |f| \leq W
\]  

(7)

The phase characteristic of \(C(f)\) may also be split equally between the transmitter and receiver filter.

The average transmitter power is

\[
P_{av} = \frac{d^2}{T} \int_{-W}^{W} \frac{X_{re} (f)}{|C(f)|} df
\]
The noise variance at the output of the receiver filter is

\[ \sigma_v^2 = \frac{N_0}{2} \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|} df \]

The SNR at the detector is

\[ \frac{d^2}{\sigma_v^2} = \frac{2P_{av}T}{N_0} \left[ \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|} df \right]^{-2} \] (8)

From Equations (6) (P.54) and Equation (8), we observe that when we express the SNR \( d^2/\sigma_v^2 \) in terms of the average transmitter power \( P_{av} \), there is a loss incurred due to channel distortion.
In the case of the filters given by Equation (5) (P.53), the loss is
\[ 10 \log \left( \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|^2} df \right) \]

In the case of the filters given by Equation (7) (P.55), the loss is
\[ 10 \log \left( \int_{-W}^{W} \frac{X_{rc}(f)}{|C(f)|^2} df \right)^2 \]

When \( C(f) = 1 \) for \( |f| \leq W \), the channel is ideal and \( \int_{-W}^{W} X_{rc}(f) df = 1 \) so that no loss is incurred.

When there is amplitude distortion, \( |C(f)| < 1 \) for some range of frequencies in the band \( |f| \leq W \) and there is a loss in SNR.

It can be shown that the filters given by Equation (7) (P.55) result in the smaller SNR loss.