ADVANCED COMMUNICATION SYSTEMS

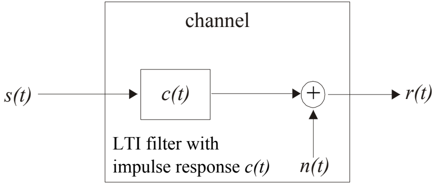
ELEN90051 (LECTURER: MARGRETA KUIJPER)

Digital Communication through Bandlimited Channels

1st Semester 2018

Material based on Chapter 9 of "Digital Communications" by Proakis & Salehi, 2008, and Chapter 8 of "Communication Systems Engineering" by Proakis & Salehi, 2002, and Chapter 5 of "Digital Communications—a Discrete-Time Approach" by Michael Rice, 2009

DIGITAL COMMUNICATION THROUGH BANDLIMITED CHANNELS



- The concept of Inter Symbol Interference (ISI)
- Transmitter design—Pulse shaping for zero ISI
  - Bandlimited channels
  - Inter symbol interference
  - Nyquist criterion
- Receiver design—Channel equalization
  - Maximum likelihood sequence detector – Viterbi algorithm
  - Linear equalization – Zero-forcing and MMSE equalizer

Bandlimited Channels

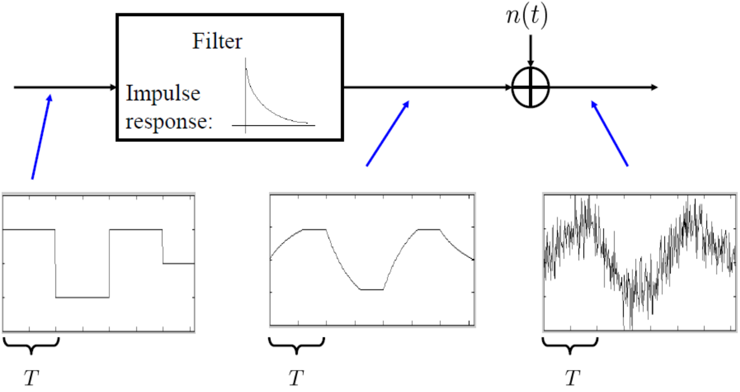
- Consider a rectangular pulse

$$g_T(t) = \begin{cases} 1 & 0 \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$$

(here the subscript T indicates "Transmitter")

- Question: What is the signal's bandwidth?
- Answer: The Fourier transform of $g_T(t)$ is a sinc function. Therefore its frequency response $G(f)$ extends over all frequencies $f \in \mathbb{R}$.
Therefore we need a channel with infinite bandwidth to transmit $g_T(t)$ without distortion. In fact, up until now we implicitly assumed that we were dealing with an infinite bandwidth ideal channel.

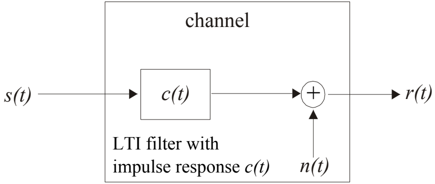
Question: What happens if we transmit $g_T(t)$ over a bandlimited channel?

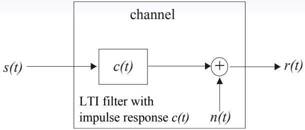


Answer: Then the signal is distorted, so may end up outside of the signal space, even in the noiseless case.

The concept of ISI

- Ideally, channels should have *infinite* bandwidth with *flat* amplitude response and *linear* phase response (i.e, constant delay for all frequencies).
- However, in practice, channels have *finite* bandwidth, often *non-flat* amplitude response and *nonlinear* phase response.
- Signal distortions in amplitude and phase due to bandwidth limitations of the channel results in inter symbol interference (ISI), which will be explained next.

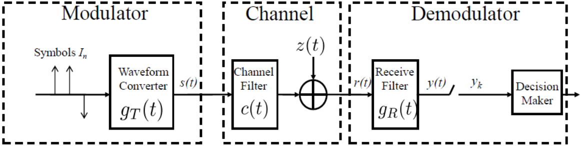


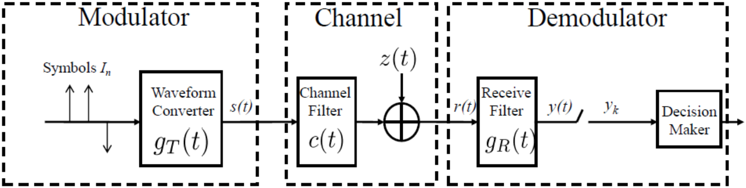


- Let's assume that the channel frequency response $C(f)$ is zero for $|f| > W$ and not necessarily flat for $|f| \leq W$.
- The transmitted signal $s(t)$ is of the form

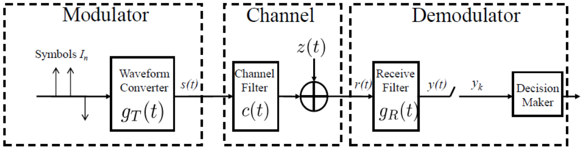
$$s(t) = \sum_{n=0}^{\infty} I_n g_T(t - nT)$$

where $R = 1/T$ is the symbol rate and $\{I_n\}$ is the transmitted sequence of symbols.





What is ISI? Intuitively, neighbouring symbols smear into each other when they pass through a non-ideal bandlimited channel. We will now explain this mathematically.



- If $s(t) = \sum_{n=0}^{\infty} I_n g_T(t - nT)$ is transmitted through a bandlimited channel $c(t)$, then the received signal is

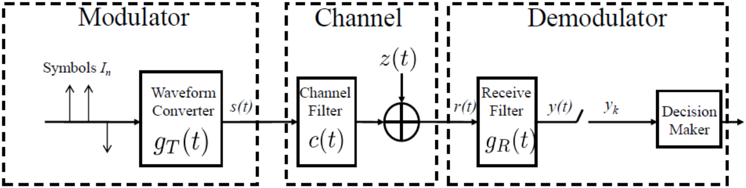
$$r(t) = \sum_{n=0}^{\infty} I_n h(t - nT) + z(t)$$

where $h(t) = (g_T \star c)(t) = \int_{-\infty}^{\infty} g_T(\tau)c(t-\tau)d\tau$ and $z(t)$ is the AWGN.

- The signal demodulator output is

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

where $x(t) = (g_T \star c \star g_R)(t)$ and $\nu(t) = (z \star g_R)(t)$ is filtered noise.



The receive filter output $y(t)$ is sampled at rate $1/T$, which results in

$$\begin{aligned} y(kT) &= \sum_{n=0}^{\infty} I_n x((k-n)T) + \nu(kT) \quad \text{for } k = 0, 1, 2, \ldots \\ &= \underbrace{I_k x(0)}_{\text{Desired symbol}} + \underbrace{\sum_{n=0, n\neq k}^{\infty} I_n x((k-n)T)}_{\text{ISI}} + \nu(kT) \end{aligned}$$

Pulse shaping for zero ISI

- Assume the channel frequency response $C(f)$ is known, with channel bandwidth W
- Can we design $g_T(t)$ and $g_R(t)$ to eliminate ISI, so that we have zero ISI?
- For zero ISI, we need

$$x(kT) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

  where $x(t) = (g_T \star c \star g_R)(t)$ with corresponding frequency response $X(f) = G_T(f)C(f)G_R(f)$.
- We next present the Nyquist pulse shaping criterion which gives a necessary and sufficient condition for zero ISI.

Nyquist Criterion

Nyquist Pulse-Shaping Criterion: A necessary and sufficient condition for $x(t)$ to satisfy

$$x(kT) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

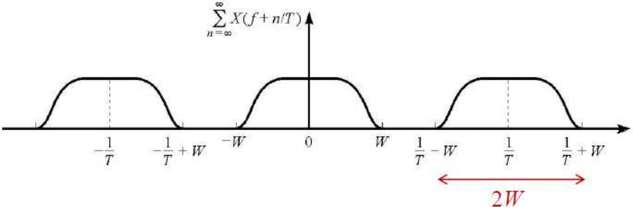
is that its frequency response $X(f)$ satisfies

$$\sum_{m=-\infty}^{\infty} X\left(f + \frac{m}{T}\right) = T$$

See textbook pp 605–606 for a proof.

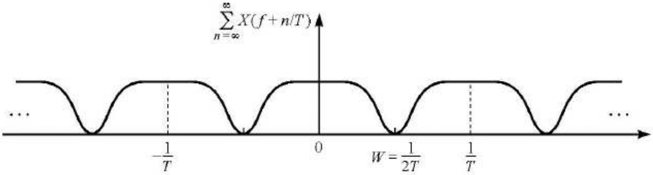
The above condition says that when you add up shifted versions of the frequency response $X(f)$ (by $\frac{1}{T}, \frac{2}{T}, \ldots$), the result should be flat (with value T).

- When is the Nyquist criterion satisfied?
- *Case 1:* $R > 2W$ where $R = \frac{1}{T}$ is the symbol transmission rate, and W is the baseband channel bandwidth



- From the above figure: there is no overlap between the shifted versions of $X(f)$ so that the Nyquist condition $\sum_{m=-\infty}^{\infty} X\left(f+\frac{m}{T}\right) = T$ cannot be satisfied.
- Therefore we conclude that zero ISI cannot be achieved when $R > 2W$.

- *Case 2:* $R = 2W$

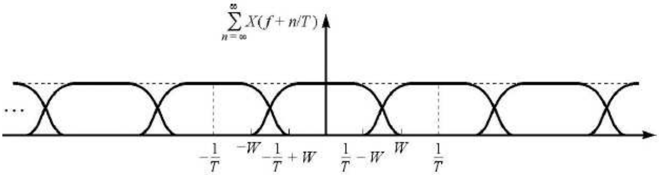


- From the above figure: the shifted versions of $X(f)$ touch at a point, therefore it is possible to satisfy the Nyquist criterion $\sum_{m=-\infty}^{\infty} X\left(f+\frac{m}{T}\right) = T$.
- In fact, the only choice for $X(f)$ when $R = 2W$ is the rectangular shape:

$$X(f) = \begin{cases} T & |f| \leq W \\ 0 & |f| > W \end{cases}$$

which corresponds to the sinc pulse $x(t) = \frac{\sin(\pi t/T)}{\pi t/T} = \operatorname{sinc}\left(\frac{t}{T}\right)$.

- *Case 3:* $R < 2W$



- From the above figure: the shifted versions of $X(f)$ overlap, therefore there are many choices for $X(f)$ to satisfy the Nyquist condition $\sum_{m=-\infty}^{\infty} X\left(f+\frac{m}{T}\right) = T$.
- To avoid ISI, the maximum symbol rate is $R = 2W$, i.e., the shortest signaling interval is

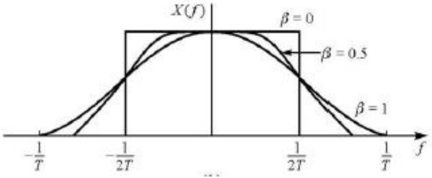
$$T = \frac{1}{2W}.$$

- *Case 3:* $R < 2W$–continued:
- A popular choice (but not the only one) for $X(f)$ when $R < 2W$ is the pulse with so-called raised cosine spectrum:

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

where $\beta \in [0,1]$ is the so-called roll-off factor. The bandwidth of this pulse is given by

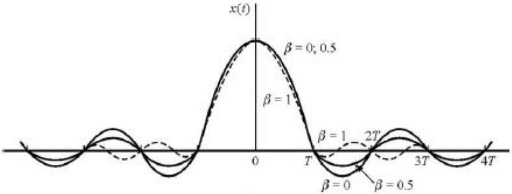
$$(1+\beta) \cdot \frac{1}{2T}.$$



In the time domain, the signal that corresponds to the pulse with raised cosine spectrum is

$$x_{rc}(t) = \operatorname{sinc}\left(\frac{t}{T}\right) \frac{\cos(\pi\beta t/T)}{1-(2\beta t/T)^2},$$

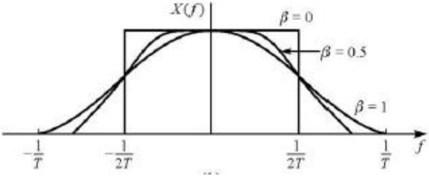
which reduces to the sinc pulse when $\beta = 0$.



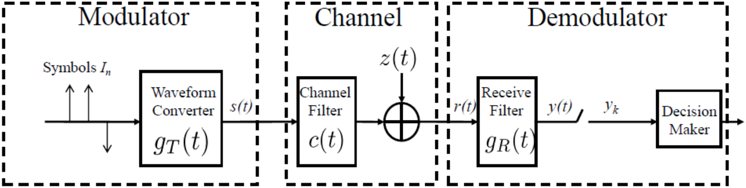
The side lobes of the pulse that corresponds to $\beta = 1$ are smaller than the side lobes of a sync pulse ($\beta = 0$). Small side lobes are better when there are timing errors because they then lead to smaller ISI components.

Choosing β for baseband channels

- We call $2W$ the Nyquist frequency; we see that $\beta = 0$ corresponds to a symbol rate at Nyquist frequency, with the rectangular spectrum of the sync pulse.
- For a larger value of β we need bandwidth beyond the Nyquist frequency: recall that
$$W = \frac{1+\beta}{2T};$$
we say that $\beta = 0.5$ involves an excess bandwidth of 50%; similarly $\beta = 1$ involves an excess bandwidth of 100%.
- Thus choosing the value of β is a trade-off between robustness against timing errors and transmission speed.



PULSE SHAPING FOR ZERO ISI



- We have $X(f) = G_T(f)C(f)G_R(f)$ which needs to satisfy the Nyquist criterion for zero ISI
- Raised-cosine spectrum is one possible choice for $X(f)$ when $R \leq 2W \Rightarrow$ choose $X(f) = X_{rc}(f)$ (we only need knowledge of the channel bandwidth W for this).
- $G_T(f)$ determines the transmitted pulse shape, $C(f)$ is the channel which we know but cannot control, $G_R(f)$ is the receive filter
- How to choose $G_T(f)$ and $G_R(f)$?

Two common choices for $G_T(f)$ and $G_R(f)$:

- *Design 1:*

$$|G_T(f)| = \frac{|X(f)|^{\frac{1}{2}}}{|C(f)|} \quad \text{and} \quad |G_R(f)| = |X(f)|^{\frac{1}{2}}$$

  This design compensates for the channel at the transmitter

- *Design 2:*

$$|G_T(f)| = \frac{|X(f)|^{\frac{1}{2}}}{|C(f)|^{\frac{1}{2}}} \quad \text{and} \quad |G_R(f)| = \frac{|X(f)|^{\frac{1}{2}}}{|C(f)|^{\frac{1}{2}}}$$

  This design compensates for the channel at both the transmitter and the receiver; it can be shown to be optimal in terms of error probability for white Gaussian noise (see textbook 9.2-4)

Both designs use $P(f) := X(f)^{\frac{1}{2}}$ which is called **square-root raised-cosine (SRRC)**, sometimes also abbreviated as RRC.

Its time domain plot is given in the next figure (from the reference book by Michael Rice, where the roll-off factor is denoted by α instead of β):

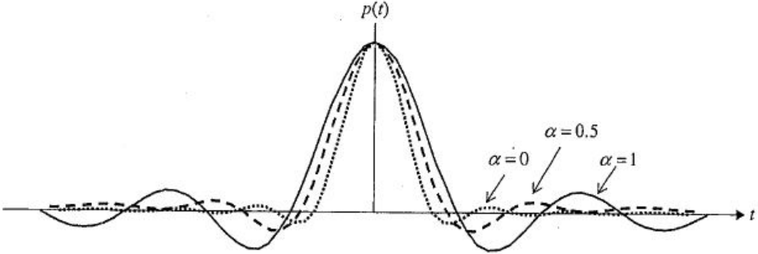
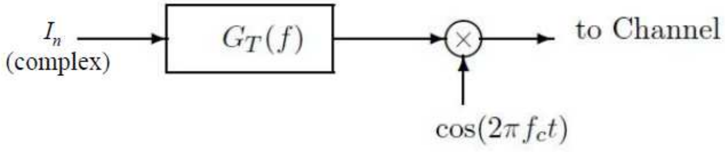


Figure A.2.5 Square-root raised-cosine (SRRC) pulse shape for excess bandwidths of 0%, 50%, and 100%.

CHOOSING β FOR BANDPASS CHANNELS

- For baseband channels, Nyquist criterion says that zero ISI can be achieved if symbol transmission rate $R \leq 2W$, where W is the channel bandwidth.
- For bandpass channels with bandwidth W, the previous analysis still holds if $s(t)$, $c(t)$, $z(t)$, $r(t)$, $y(t)$ are the (complex) baseband equivalents. In this case the Nyquist criterion says that zero ISI can be achieved if $R \leq W$ (due to different ways of defining bandwidth).



CHOOSING β FOR BANDPASS CHANNELS

- In this case the Nyquist frequency equals the channel bandwidth W; now $\beta = 0$ corresponds to a symbol rate at Nyquist frequency, with the rectangular spectrum of the sync pulse.
- For a larger value of β we need bandwidth beyond the Nyquist frequency: for bandpass channels we have

$$W = \frac{1+\beta}{T};$$

  again we say that $\beta = 0.5$ involves an excess bandwidth of 50%; similarly $\beta = 1$ involves an excess bandwidth of 100%.
- Again, choosing the value of β is a trade-off between robustness against timing errors and transmission speed.

Bandpass Example—Questions

A voice-band telephone channel has a passband characteristic in the frequency range $300 < f < 3000$ Hz. Assume that the channel has an ideal frequency response characteristic within the range $300 < f < 3000$ Hz.

- Using PAM, select a symbol rate and a constellation size to achieve 9600 bits/sec signal transmission with zero ISI.
- If a raised cosine pulse is used for the transmitter pulse $g_T(t)$, select the roll-off factor β.

Bandpass Example—Answers

A voice-band telephone channel has a passband characteristic in the frequency range $300 < f < 3000$ Hz. Assume that the channel has an ideal frequency response characteristic within the range $300 < f < 3000$ Hz.

- Using PAM, select a symbol rate and a constellation size to achieve 9600 bits/sec signal transmission with zero ISI. (Answer: the channel bandwidth equals $W = 3000 - 300 = 2700$ Hz, which is an upperbound for the symbol rate $9600/\log_2 M$, therefore $M = 16$ is the smallest possible M)
- If a raised cosine pulse is used for the transmitter pulse $g_T(t)$, select the roll-off factor β. (Answer: the Nyquist frequency equals $9600/\log_2 16 = 2400$ symbols/sec. Thus, we have excess bandwidth of $2700 - 2400 = 300$ Hz which is $\frac{1}{8}$th of 2400, thus $\beta = \frac{1}{8}$)

Channel Equalization

- In pulse shaping for zero ISI, we assumed that $C(f)$ is known, thus can be compensated for at the transmit and receive filters.
- However, in practice, $C(f)$ is often not known exactly (particularly at the transmitter), as a result some ISI still remains.
- A useful diagnostic tool is a so-called eye diagram. The amount of remaining ISI can be observed from the opening of the eye, see next few pages.
- A channel equalizer compensates for the effects of an unknown channel, in particular ISI, see later pages.

EYE DIAGRAM

- An eye diagram is a time-domain modulo-T plot of the receiver's response $y(t)$ when a modulated random symbol sequence is sent through the channel (without noise).

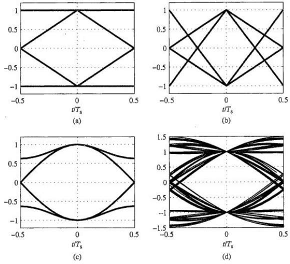


Figure 5.2.4 Eye diagrams for (a) The NRZ pulse shape, (b) the MAN pulse shape, (c) the HS pulse shape, and (d) the SRRC pulse shape with 50% excess bandwidth.

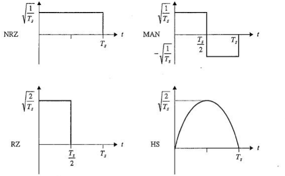


Figure A.1.1 The non-return-to-zero (NRZ), the return-to-zero (RZ), Manchester (MAN), and half-sine (HS) pulse shapes as shown.

- As you can see from the above figure, the timing is shifted in an eye diagram, so that the optimum sampling time instant is at $t = 0$.
- The resulting overall plot is shaped as **an eye** because the symbol sequence is random, here the modulo-T plot is sometimes positive, sometimes negative.

- As the eye closes, ISI increases; as the eye opens, ISI is decreasing.
- Other diagnostics are: noise margin (=eye opening), sensitivity-to-timing error (=eye width), see figure below. The figure spans two signaling intervals, clearly channel distortions lead to a smaller eye opening width and also a smaller eye opening at $t = 0$.

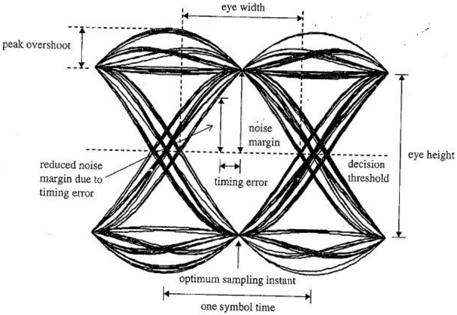
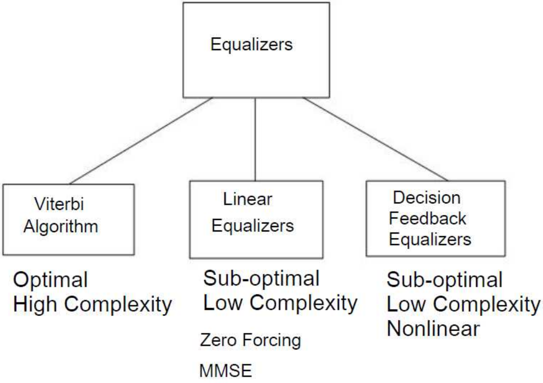


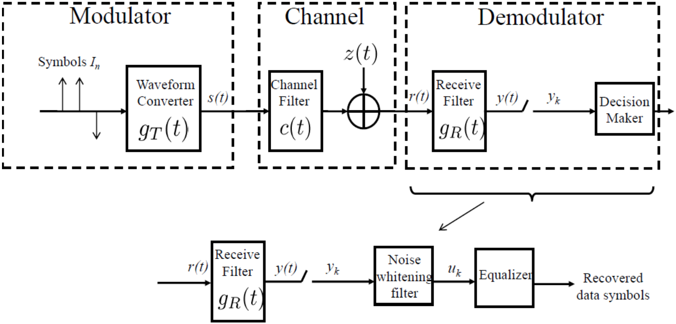
Figure 5.2.5 An eye diagram illustrating the important diagnostic characteristics. The diagram was generated using 200 random binary PAM symbols and the SRRC pulse shape with 50% excess bandwidth.

Channel Equalization



- Equalizers are useful in detecting data in the presence of ISI in many types of channels, e.g., dial-up telephone channels, wireless radio channels, underwater acoustic channels, etc.

Equivalent DT model for ISI



- Recall that the sampled output of the receive filter is given by

$$y_k = I_k x_0 + \sum_{n=0, n\neq k}^{\infty} I_n x_{k-n} + \nu_k.$$

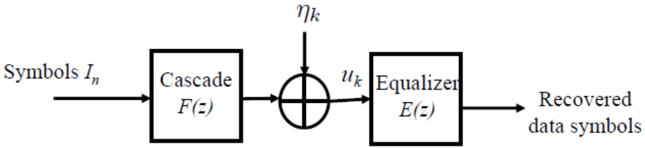
This is called the **equivalent Discrete Time model for ISI**. Here ν_k is filtered non-white noise; to whiten it, we pass y_k through a noise whitening filter.

- The output of the noise whitening filter can be written as

$$u_k = f_0 I_k + \sum_{n=1}^{\infty} f_n I_{k-n} + \eta_k$$

where η_k is white noise and f_n are the filter coefficients of the cascade of filters that we used (transmitter, channel, receiver, noise whitening).

- Thus, an equivalent model of our transmitter-equalizer system is:

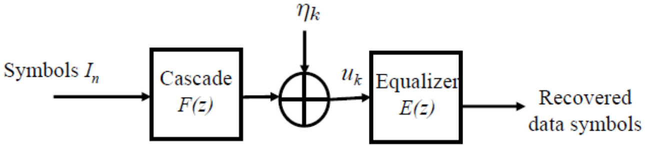


- In practice, we often assume that ISI only affects a finite number of neighbouring symbols. This results in a simpler FIR (=finite impulse response) model:

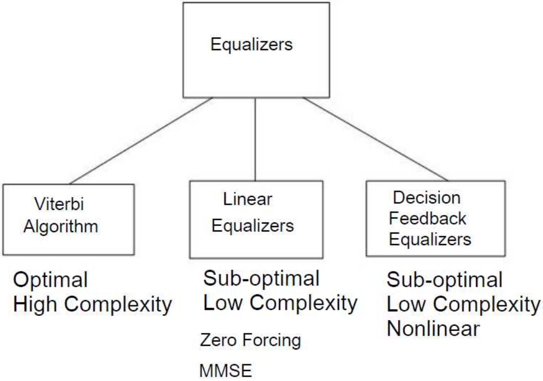
$$u_k = f_0 I_k + \sum_{n=1}^{L} f_n I_{k-n} + \eta_k$$

where L is the channel memory $=$ the number of ISI components.

- The aim of the equalizer is to recover the transmitted symbol sequence $\{I_k\}$ from the received sequence of values $\{u_k\}$.



RECALL: OVERVIEW OF CHANNEL EQUALIZATION



- Next we will present Viterbi Equalization.

Maximum Likelihood Sequence Detector

- When the received sequence is $\{u_0, u_1, \ldots, u_K\}$, the optimal detector is the **MAP sequence detector** given by

$$[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] = \arg \max_{i_0, \ldots, i_K} P\left[I_0 = i_0, I_1 = i_1, \ldots I_K = i_K | u_0, u_1, \ldots, u_K\right]$$

- Assuming equiprobable symbols, this equals the **ML sequence detector** given by

$$[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] = \arg \max_{i_0, \ldots, i_K} p\left[u_0, u_1, \ldots u_K | I_0 = i_0, I_1 = i_1, \ldots, I_K = i_K\right]$$

- Due to causality, we can write this as

$$\begin{aligned}
[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] &= \arg \max_{i_0, \ldots, i_K} p\left[u_0 | I_0 = i_0\right] p\left[u_1, \ldots u_K | I_0 = i_0, I_1 = i_1, \ldots, I_K = i_K\right] \\
&= \arg \max_{i_0, \ldots, i_K} p\left[u_0 | I_0 = i_0\right] p\left[u_1 | I_0 = i_0, I_1 = i_1\right] p\left[u_2, \ldots u_K | I_0 = i_0, I_1 = i_1, \ldots, I_K = i_K\right] \\
&= \arg \max_{i_0, \ldots, i_K} p\left[u_0 | I_0 = i_0\right] p\left[u_1 | I_0 = i_0, I_1 = i_1\right] p\left[u_2 | I_0 = i_0, I_1 = i_1, I_2 = i_2\right] \\
&\quad \times p\left[u_3 | I_0 = i_0, I_1 = i_1, I_2 = i_2, I_3 = i_3\right] \times \ldots \times p\left[u_K | I_0 = i_0, I_1 = i_1, \ldots, I_K = i_K\right]
\end{aligned}$$

- If the channel memory is L, then this ML sequence detector is simplified as

$$[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] = \arg\max_{i_0,\ldots,i_K} \prod_{k=0}^{K} p\left[u_k | I_{k-L} = i_{k-L}, \ldots, I_k = i_k\right]$$

which can be written in terms of log-likelihood functions as

$$\begin{aligned}[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] &= \arg\max_{i_0,\ldots,i_K} \ln\left(\prod_{k=0}^{K} p\left[u_k | I_{k-L} = i_{k-L}, \ldots, I_k = i_k\right]\right) \\ &= \arg\max_{i_0,\ldots,i_K} \sum_{k=0}^{K} \ln\left(p\left[u_k | I_{k-L} = i_{k-L}, \ldots, I_k = i_k\right]\right)\end{aligned}$$

- If $\eta_k \sim \mathcal{N}(0, \sigma^2)$, then $u_k = \sum_{n=0}^{L} f_n I_{k-n} + \eta_k$ has distribution $u_k \sim \mathcal{N}(\sum_{n=0}^{L} f_n I_{k-n}, \sigma^2)$ and the ML sequence detector is given by

$$\begin{aligned}[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] &= \arg\max_{i_0,\ldots,i_K} \sum_{k=0}^{K} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(u_k - \sum_{n=0}^{L} f_n i_{k-n})^2}{2\sigma^2}\right] \\ &= \arg\min_{i_0,\ldots,i_K} \sum_{k=0}^{K} \left(u_k - \sum_{n=0}^{L} f_n i_{k-n}\right)^2\end{aligned}$$

Example ML Sequence Detector

Suppose that binary PAM modulation is used and that the channel memory $L = 1$. This means that the current signaling interval experiences leakage from only the previous signaling interval. Assume that the received sequence is (u_0, u_1, u_2, u_3). Therefore, our model $u_k = \sum_{n=0}^{L} f_n I_{k-n} + \eta_k$ results in

$$
\begin{aligned}
u_0 &= f_0 I_0 + \eta_0, \\
u_1 &= f_0 I_1 + f_1 I_0 + \eta_1, \\
u_2 &= f_0 I_2 + f_1 I_1 + \eta_2, \\
u_3 &= f_0 I_3 + f_1 I_2 + \eta_3.
\end{aligned}
$$

How to compute the most likely sequence (i_0, i_1, i_2, i_3)?

- The ML sequence detector produces:

$$[\widehat{I}_0, \widehat{I}_1, \widehat{I}_2, \widehat{I}_3] = \arg \max_{i_0, i_1, i_2, i_3} [\ln p\,(u_0 | I_0 = i_0) + \ln p\,(u_1 | I_0 = i_0, I_1 = i_1) + \ln p\,(u_2 | I_1 = i_1, I_2 = i_2) + \ln p\,(u_3 | I_2 = i_2, I_3 = i_3)]$$

- For binary PAM, the detector needs to go through all 16 possible combinations of $\{i_0 = \pm 1, i_1 = \pm 1, i_2 = \pm 1, i_3 = \pm 1\}$ to find the one that maximizes $\sum_{k=0}^{3} \ln\left(p\left[u_k | I_{k-1} = i_{k-1}, I_k = i_k\right]\right)$

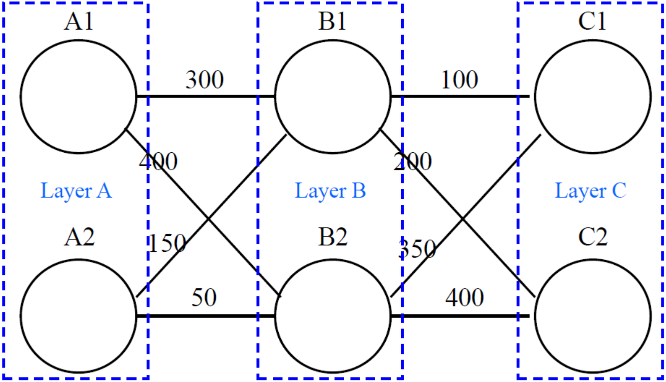
  This is a valid approach, however a faster approach is to use the so-called **Viterbi algorithm**, which is explained on the next few pages.

Viterbi Algorithm

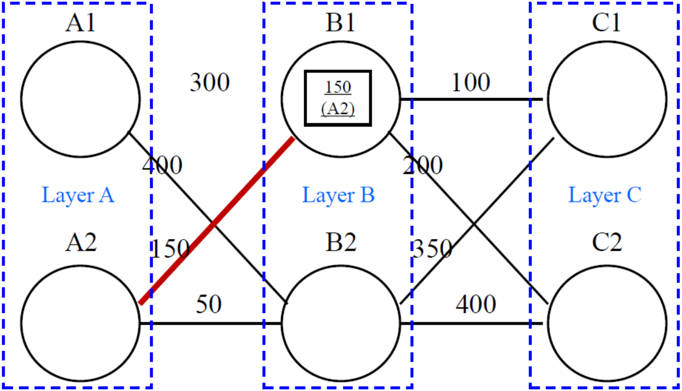
- The Viterbi algorithm finds the shortest (or longest) path in a trellis
- it was proposed in 1967 by Andrew Viterbi for the decoding of convolutional codes
- Its applications now also include channel equalization, maximum likelihood decoding of hidden Markov models, speech & text recognition, etc.

Viterbi Algorithm

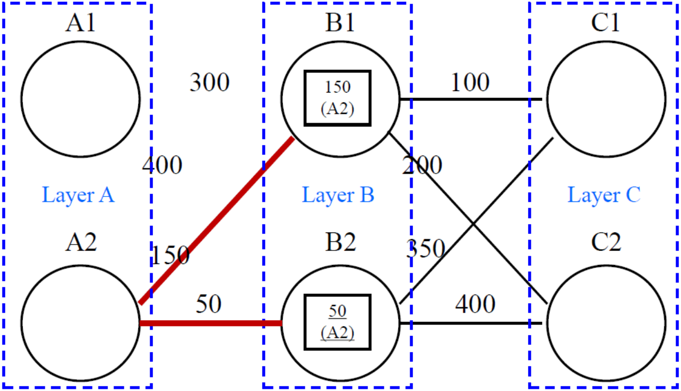
- Main idea: For a trellis with different individual path lengths, how to find the shortest 2-hop path length from A to C?
- For each node in B and C, calculate the shortest path length from A and discard all other paths.



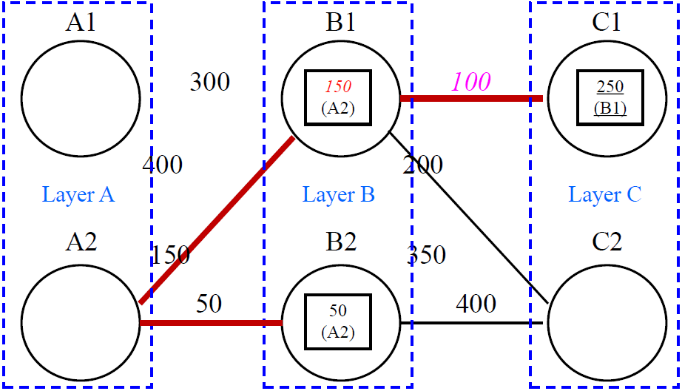
- For node B1, shortest path length from A is 150 via A2 ⇒ Discard path from A1 to B1.



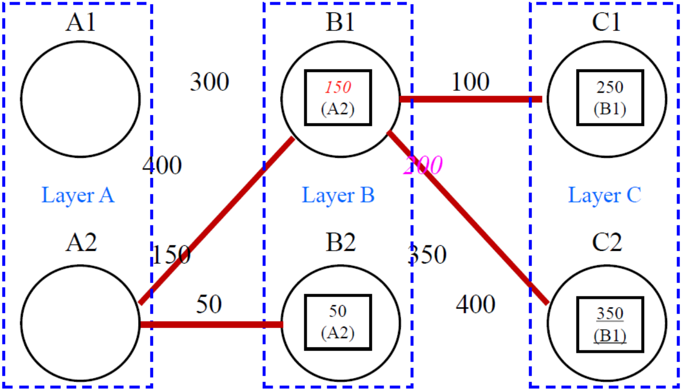
- For node B2, shortest path length from A is 50 via A2 ⇒ Discard path from A1 to B2.



- For node C1, shortest path length from A is 250 (=150 + 100) via B1 $\Rightarrow$ Discard path from B2 to C1.

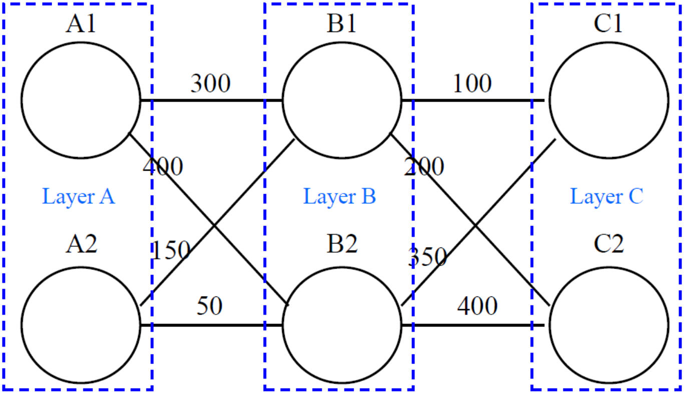


- For node C2, shortest path length from A is 350 (=150 + 200) via B1 ⇒ Discard path from B2 to C2.



- Finally, comparing the shortest distances from A to C1 and C2, the shortest path is A2→B1→C1 with total length 250.

- Exercise: Find the longest 2-hop path length from A to C? (Answer: A1→B2→C2 with total length 800)



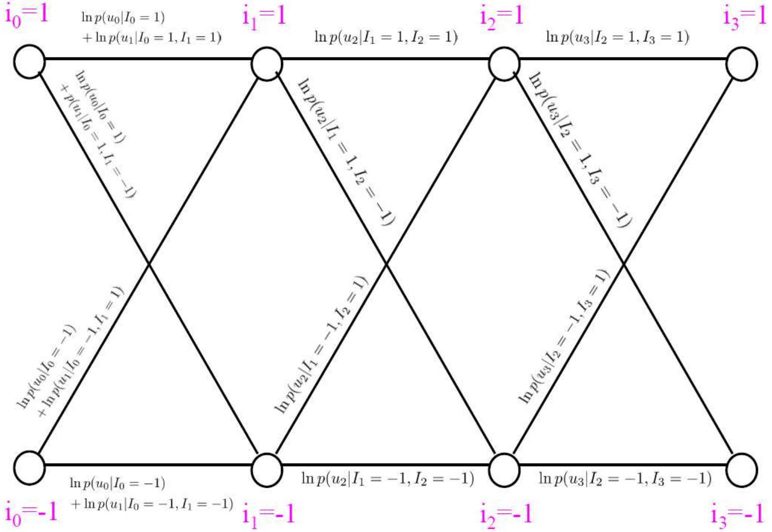
- Consider again example: $L = 1$ and received sequence $\{u_0, u_1, u_2, u_3\}$ with optimal decision rule

$$[\widehat{I}_0, \widehat{I}_1, \widehat{I}_2, \widehat{I}_3] = \arg \max_{i_0, i_1, i_2, i_3} [\ln p(u_0 | I_0 = i_0) + \ln p(u_1 | I_0 = i_0, I_1 = i_1) + \ln p(u_2 | I_1 = i_1, I_2 = i_2) + \ln p(u_3 | I_2 = i_2, I_3 = i_3)]$$

  and 16 possible combinations of $\{i_0, i_1, i_2, i_3\}$ for binary PAM.

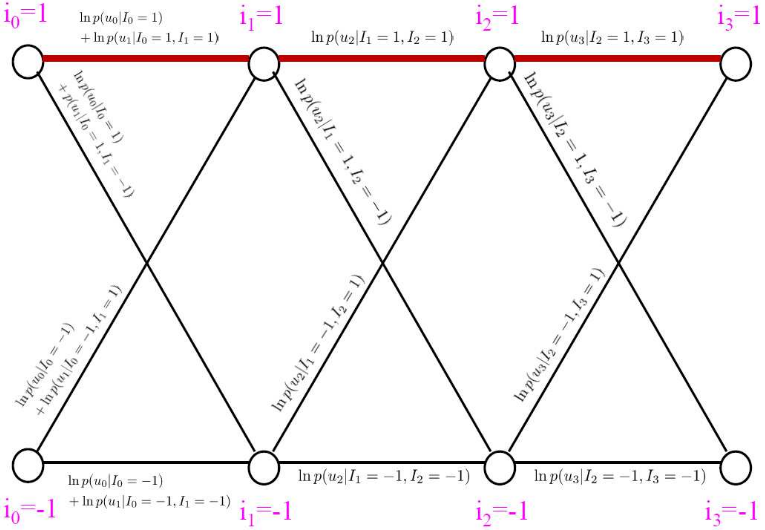
- To apply the Viterbi algorithm, we can construct a trellis with nodes $\{i_0 = \pm 1, i_1 = \pm 1, i_2 = \pm 1, i_3 = \pm 1\}$ and path lengths are the log-likelihood functions $\ln(p[u_k | I_{k-1} = i_{k-1}, I_k = i_k])$ where the most likely transmitted sequence $\{\widehat{I}_0 = i_0, \widehat{I}_1 = i_1, \widehat{I}_2 = i_2, \widehat{I}_3 = i_3\}$ corresponds to the longest path length through the trellis.

Trellis diagram for binary PAM with $L = 1$ and $\{u_0, u_1, u_2, u_3\}$.



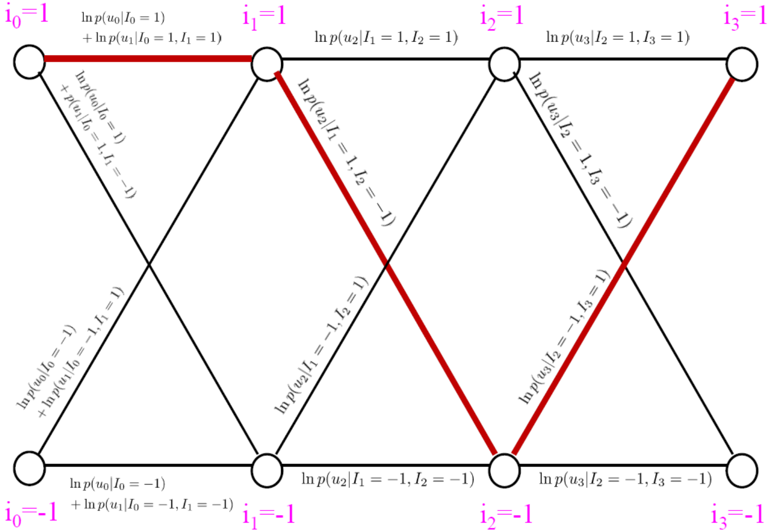
Total path length for $\{I_0 = 1, I_1 = 1, I_2 = 1, I_3 = 1\}$ is

$$\ln p\left(u_0|I_0 = 1\right) + \ln p\left(u_1|I_0 = 1, I_1 = 1\right) + \ln p\left(u_2|I_1 = 1, I_2 = 1\right) + \ln p\left(u_3|I_2 = 1, I_3 = 1\right)$$



Total path length for $\{I_0 = 1, I_1 = 1, I_2 = -1, I_3 = 1\}$ is

$$\ln p\left(u_0|I_0 = 1\right) + \ln p\left(u_1|I_0 = 1, I_1 = 1\right) + \ln p\left(u_2|I_1 = 1, I_2 = -1\right) + \ln p\left(u_3|I_2 = -1, I_3 = 1\right)$$



- Example: For $L = 1$ and $\{u_0, u_1, u_2, u_3\}$, with receive filter output $u_k = I_k + 0.5I_{k-1} + \eta_k$, $I_k \in \{-1, 1\}$, $\eta_k \sim \mathcal{N}(0, \sigma^2)$, and $I_k = 0$ for $k \leq 0$, we have

$$
\begin{aligned}
u_0 &= I_0 + \eta_0, \\
u_1 &= I_1 + 0.5I_0 + \eta_1, \\
u_2 &= I_2 + 0.5I_1 + \eta_2, \\
u_3 &= I_3 + 0.5I_2 + \eta_3.
\end{aligned}
$$

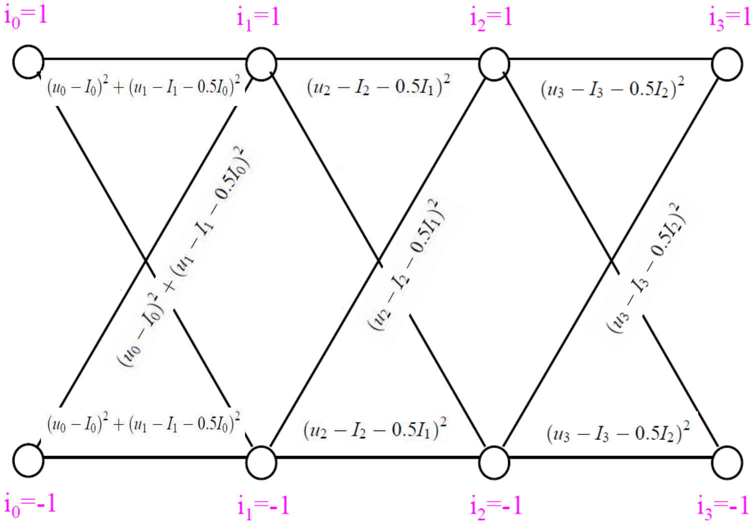
- Suppose the receive sequence from the sampled filter output is $\{u_0 = 0.25, u_1 = -0.8, u_2 = -1.6, u_3 = 0.45\}$, what is the most likely transmitted sequence $\{\widehat{I}_0, \widehat{I}_1, \widehat{I}_2, \widehat{I}_3\}$?

- Given that $\eta_k \sim \mathcal{N}(0, \sigma^2)$, then $u_k \sim \mathcal{N}(I_k + 0.5I_{k-1}, \sigma^2)$ and ML sequence detector is

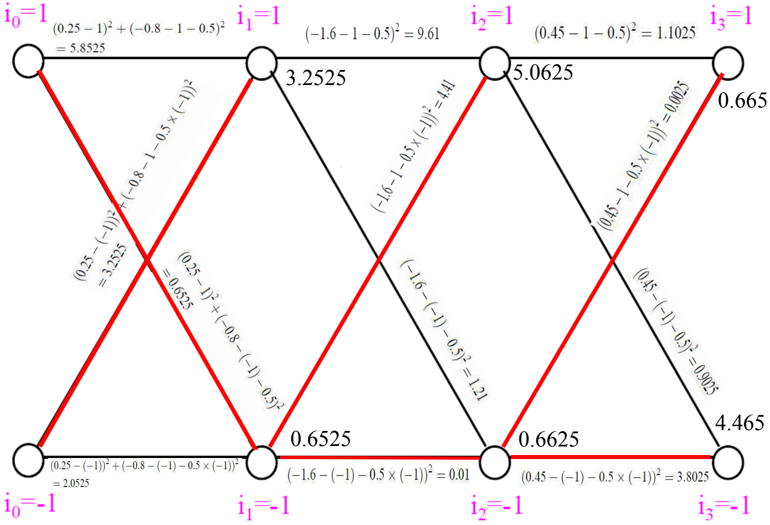
$$[\widehat{I}_0, \widehat{I}_1, \ldots, \widehat{I}_K] = \color{red}{\arg \min_{i_0, \ldots, i_K}} \sum_{k=0}^{K} (u_k - I_k - 0.5I_{k-1})^2$$

- To apply the Viterbi algorithm, we employ a trellis with nodes $\{i_0 = \pm 1, i_1 = \pm 1, i_2 = \pm 1, i_3 = \pm 1\}$ and path lengths $(u_k - I_k - 0.5I_{k-1})^2$ where the most likely transmitted sequence $\{\widehat{I}_0 = i_0, \widehat{I}_1 = i_1, \widehat{I}_2 = i_2, \widehat{I}_3 = i_3\}$ corresponds to the shortest path length through the trellis.
- The Viterbi algorithm finds this shortest path in an efficient way, per node the algorithm keeps track of the total survivor path length.

Trellis diagram for binary PAM with $L = 1$ and $\{u_0, u_1, u_2, u_3\}$



Most likely transmitted sequence is $\{\widehat{I}_0, \widehat{I}_1, \widehat{I}_2, \widehat{I}_3\} = \{1, -1, -1, 1\}$.



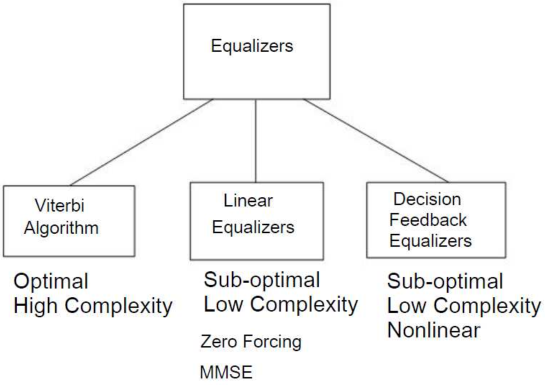
- In the previous example, the Viterbi algorithm compares 4 numbers at each stage over 3 stages with a total of 12 operations. Compare this with exhaustive search where all possible 16 sequences need to be compared.
- The Viterbi algorithm is more efficient when detecting longer sequences. For $L = 1$ and a received sequence of length 10, i.e., $\{u_0, u_1, \ldots, u_9\}$, the Viterbi algorithm compares 4 numbers at each stage over 9 stages with a total of 36 operations. Compare this with exhaustive search where all possible $2^{10} = 1024$ sequences need to be compared.

- In general, for M-ary signalling, the trellis diagram has M^L states $S_k = (I_{k-1}, I_{k-2}, \ldots, I_{k-L})$ at a given time (previous example had 2 states i.e., $M = 2$, $L = 1$).
- For each new received signal u_k, Viterbi algorithm computes the M^{L+1} metrics

$$\sum_{k=0}^{K} \ln\left(p\left[u_k | I_{k-L} = i_{k-L}, \ldots, I_k = i_k\right]\right)$$

- For large L, implementation of the Viterbi algorithm can (still) be computationally intensive $\Rightarrow$ consider some suboptimal approaches e.g., linear equalization and decision feedback equalization

Recall: Overview of Channel Equalization



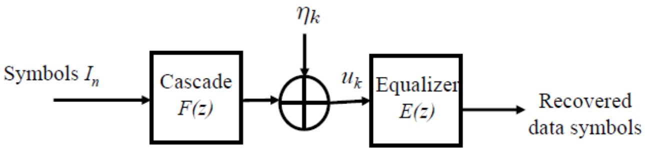
- Next we will present Linear Equalizers.

- Recall our FIR equivalent Discrete Time model for ISI:

$$u_k = f_0 I_k + \sum_{n=1}^{L} f_n I_{k-n} + \eta_k$$

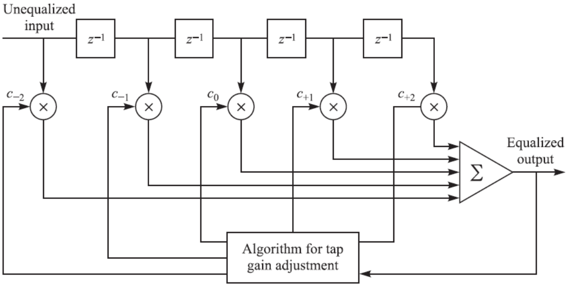
where L is the channel memory $=$ the number of ISI components.

- Recall that the aim of the equalizer is to recover the transmitted symbol sequence $\{I_k\}$ from the received sequence of values $\{u_k\}$.



- An obvious strategy would be to try and invert the FIR model. Ideally $E(z) \cdot F(z) = 1$, can we achieve this? Note: in the next pages the notation $C(z)$ is used instead of $E(z)$.

ZERO-FORCING EQUALIZATION



- Suppose that our linear equalizer's output is given as $\widehat{I}_k = \sum_{j=-K}^{K} c_j \, u_{k-j}$ where the c_j's are $2K + 1 \geq L$ filter tap coefficients of the equalizer to be designed—ideally $C(z) \cdot F(z) = 1$.
- Linear equalizers have computational complexity that **grows linearly** with channel memory L (vs exponential growth for the Viterbi algorithm).

- Given input $u_k = f_0 I_k + \sum_{n=1}^{L} f_n I_{k-n} + \eta_k$, the output of the linear filter equalizer is given by

$$\begin{aligned}\widehat{I}_k &= \sum_{j=-K}^{K} c_j \, u_{k-j} \\ &= q_0 I_k + \sum_{j=-K, j\neq k}^{K} q_{k-j} I_j + \sum_{j=-K}^{K} c_j \eta_{k-j}\end{aligned}$$

where $q_n = (f \star c)_n$.

- To satisfy zero-forcing criteria, choose filter tap coefficients c_j such that

$$q_n = \sum_{j=-K}^{K} c_j \, f_{n-j} = \begin{cases} 1 & n = 0 \\ 0 & n = \pm 1, \pm 2, \ldots, \pm K \end{cases}$$

- In general, some q_n's in the range $n = K+1, \ldots, K+L-1$ will be still non-zero due to finite filter length.

EXAMPLE–QUESTIONS

- Binary PAM is used to transmit information over an unequalized linear filter channel with impulse response

$$f_m = \begin{cases} 0.3 & m = \pm 1 \\ 0.9 & m = 0 \\ 0 & \text{otherwise} \end{cases}$$

- Design a three-tap zero-forcing linear equalizer so that the output is $q_1 = 0$, $q_0 = 1$, and $q_{-1} = 0$
- Determine the values of q_2 and q_{-2} with this design.

Example–answers

- The zero-forcing criteria with $2K+1=3$ taps are given by

$$\begin{aligned} q_1 &= c_{-1}f_2 + c_0 f_1 + c_1 f_0 = 0, \\ q_0 &= c_{-1}f_1 + c_0 f_0 + c_1 f_{-1} = 1, \\ q_{-1} &= c_{-1}f_0 + c_0 f_{-1} + c_1 f_{-2} = 0, \end{aligned}$$

where $q_n = \sum_{j=-1}^{1} c_j f_{n-j}$. The set of linear equations to be solved:

$$\begin{bmatrix} f_0 & f_{-1} & 0 \\ f_1 & f_0 & f_{-1} \\ 0 & f_1 & f_0 \end{bmatrix} \begin{bmatrix} c_{-1} \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} c_{-1} \\ c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} -0.4762 \\ 1.4286 \\ -0.4762 \end{bmatrix}$$

- Using these equalizer filter taps $c_1 = -0.4762$, $c_0 = 1.4286$, and $c_{-1} = -0.4762$ and the channel impulse response

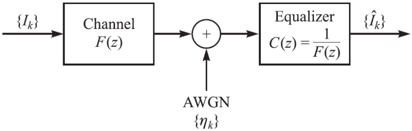
$$f_m = \begin{cases} 0.3 & m = \pm 1 \\ 0.9 & m = 0 \\ 0 & \text{otherwise,} \end{cases}$$

the equalizer outputs for $n = \pm 2$ are

$$\begin{aligned} q_2 &= c_{-1} f_3 + c_0 f_2 + c_1 f_1 \\ &= c_1 f_1 = -0.14286 \\ q_{-2} &= c_{-1} f_{-1} + c_0 f_{-2} + c_1 f_{-3} \\ &= c_{-1} f_{-1} = -0.14286 \end{aligned}$$

which means that still some residual ISI components remain—these are not eliminated by our ZF filter with finite number of filter taps.

MORE TAPS—PRO'S AND CON'S



- If an infinite number of filter taps c_j would be possible, then we would simply invert the channel response $F(z)$:

$$q_n = \sum_{j=-\infty}^{\infty} c_j f_{n-j} = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

- This implies $Q(z) = C(z)F(z) = 1$ or $C(z) = \frac{1}{F(z)}$. In this case no residual ISI components remain.
- Thus, in practice, more taps means less residual ISI in the noiseless case. However, in the presence of noise, amplification of the noise component η_k by the equalizer is possible if $F(z)$ is very small in certain frequencies.

SUMMARY

Pulse shaping for zero ISI

- Channels with finite bandwidth W cause intersymbol interference (ISI) at the receiver.
- Zero ISI can be achieved in bandlimited channels when frequency response $X(f) = G_T(f)C(f)G_R(f)$ satisfies Nyquist criterion $\sum_{m=-\infty}^{\infty} X\left(f + \frac{m}{T}\right) = T$.
- zero ISI can only be achieved if the symbol transmission rate R satisfies $R \leq 2W$, where W is the baseband channel bandwidth.
- Raised cosine spectrum is a popular choice for $X(f)$.

Summary

Channel Equalization

- ISI still occurs when the channel response $C(f)$ is not known exactly. To counter this remaining ISI, a channel equalizer can be used to detect the transmitted sequence $\{I_1, \ldots, I_K\}$ from the received sequence $\{u_1, \ldots, u_K\}$ that is corrupted by ISI.
- the optimal equalizer is the ML sequence detector which utilizes the Viterbi algorithm to make decisions $\{\widehat{I}_1, \ldots, \widehat{I}_K\}$ (assuming equiprobable symbols).
- Suboptimal equalizers offer lower computational complexity, e.g., linear growth for linear equalizers, such as ZF equalizer, compared with exponential growth for Viterbi equalizers.
- A ZF equalizer inverts the channel response but may amplify the noise.

Another Type of Linear Equalizer

- A Minimum Mean Squared Error (MMSE) Equalizer minimizes the MSE of the equalizer output defined as

$$J(\mathbf{c}) = \mathbb{E}[|I_k - \widehat{I}_k|^2] = \mathbb{E}\left[\left|I_k - \sum_{j=-K}^{K} c_j \, u_{k-j}\right|^2\right]$$

- When the noise is small for all frequencies, then the MMSE equalizer is identical to a zero-forcing equalizer.
- When the noise is large at frequencies where $F(z)$ is small, then the MMSE equalizer does not produce large noise amplification like the zero-forcing equalizer.

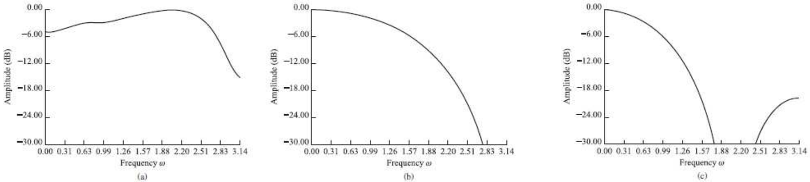


FIGURE 9.4–6
Amplitude spectra for the channels shown in Figure 9.4–5a, b, and c, resp

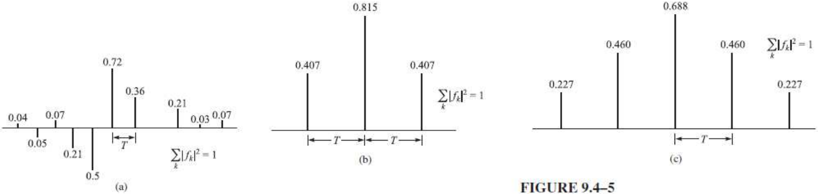


FIGURE 9.4–5
Three discrete-time channel characteristics.

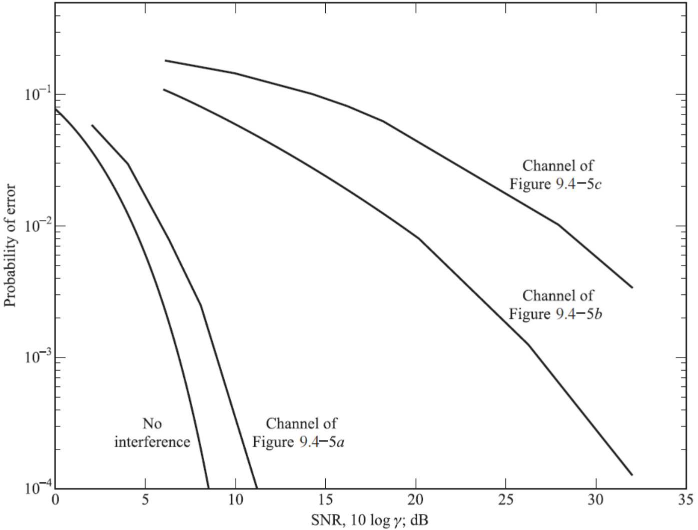
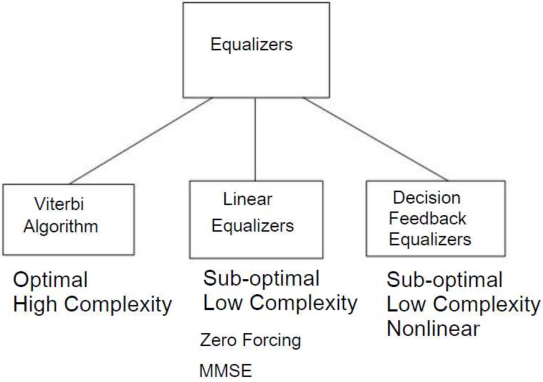


FIGURE 9.4–4
Error rate performance of linear MSE equalizer. Thirty-one taps in transversal equalizer.

Recall: Overview of Channel Equalization



Other types of equalizers are non-linear equalizers, such as so-called Decision Feedback Equalizers (*not covered here*).