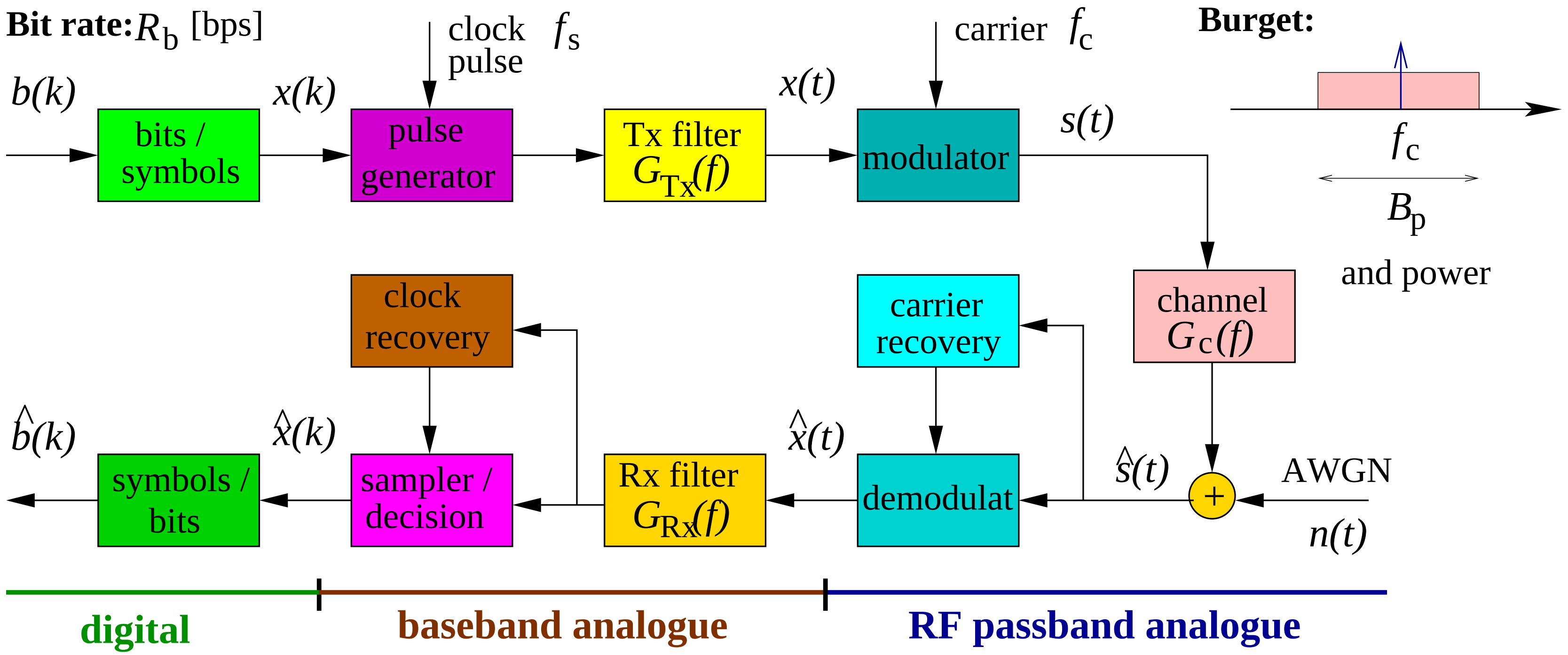
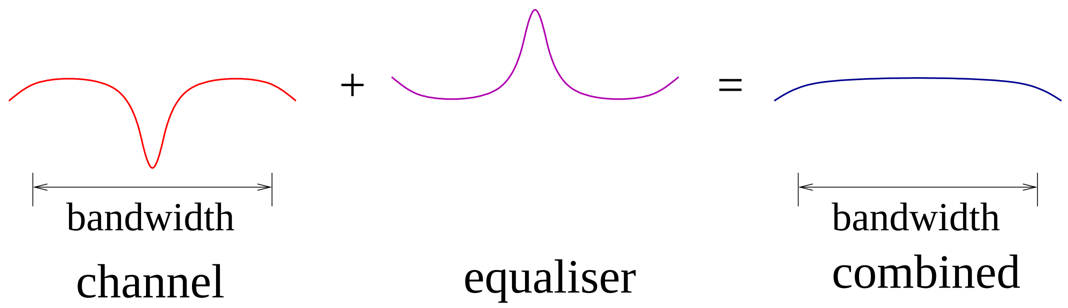
MODEM Revision - Non-dispersive Channel



- Information theory underpins every components of MODEM
- Given bit rate R_b [bps] and resource of channel bandwidth B_p and power budget
  - Select a modulation scheme (bits to symbols map) so that symbol rate can fit into required baseband bandwidth of $B = B_p/2$ and signal power can met power budget
  - Pulse shaping ensures bandwidth constraint is met and maximizes receive SNR
  - At transmitter, baseband signal modulates carrier so transmitted signal is in required channel
  - At receiver, incoming carrier phase must be recovered to demodulate it, and timing must be recovered to correctly sampling demodulated signal

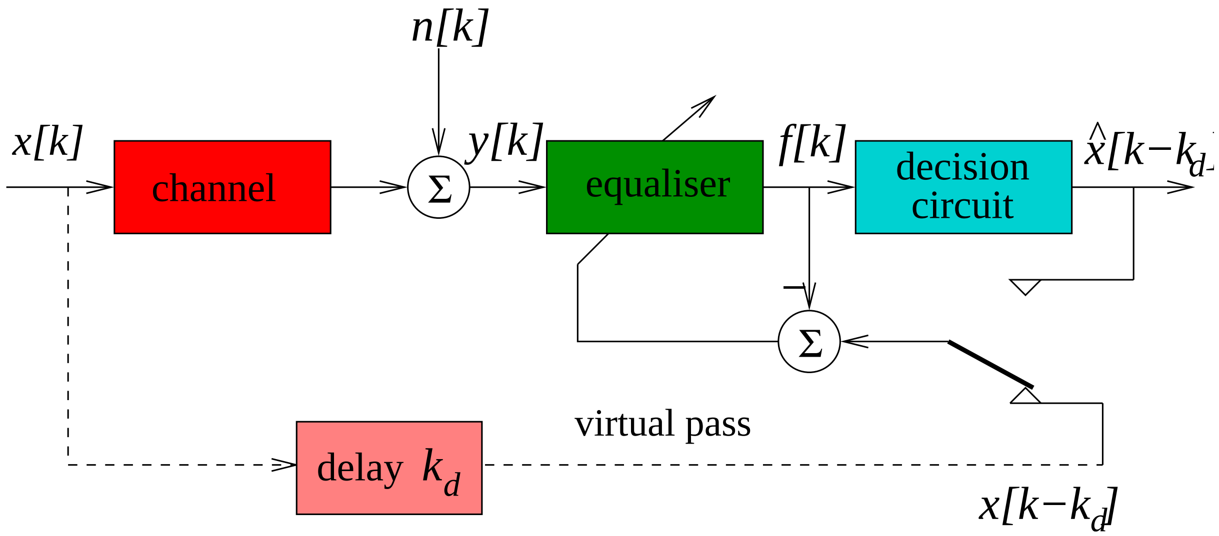
MODEM Revision - Dispersive Channel

- For dispersive channel, equaliser at receiver aims to make combined channel/equaliser a Nyquist system again



  – Design of equaliser is a **trade off** between eliminating ISI and not enhancing noise too much

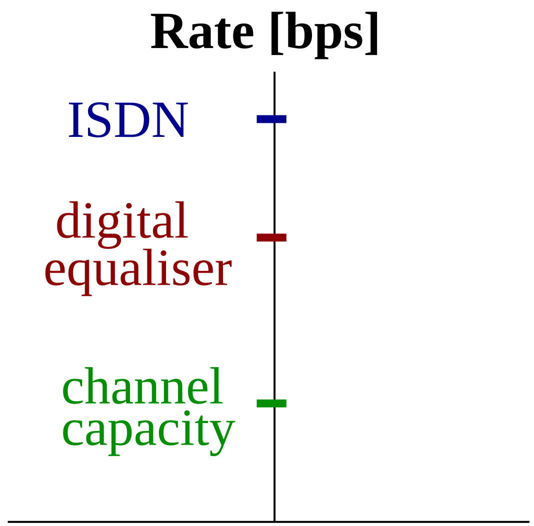
- Generic framework of adaptive equalisation with training and decision-directed modes



- Linear equaliser design issues: minimum mean square error equaliser, and adaptive least mean square algorithm

Test How Good You Really Are

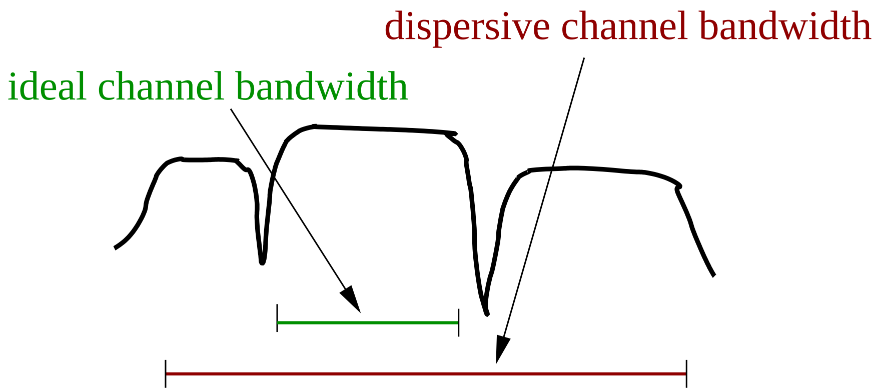
- In communication, we have 'Bible' – **channel capacity**, maximum possible rate might be achieved
  - Scientists and engineers endeavour to get close to this capacity
- In the late 1980s, the author of the book ISDN Explained (first edition) claimed in this figure:
  - BP telephone with digital equaliser achieves a higher rate than the channel capacity allowed
  - and with then suppose-to-come ISDN, the rate will be even higher
- Now channel capacity is the highest possible achieved rate, and how can BP telephone line with a digital equaliser overtakes it!?
- Was the author wrong? or was there any explanation?



- Channel capacity is indeed the maximum possible rate that may be achieved
  - The trouble was at that time BP telephone line with a digital equaliser did achieve the rate (red bar at figure) appeared to be higher than the supposed channel capacity (green bar at figure)
- There was also a personal story behind this, in 1990, I was interviewed for a lecturer post at Imperial College
  - A big professor asked me to explain how can the rate (red bar at figure) higher than the capacity
  - I could not ask, and you might guess, I did not get the job
- Incidentally, the author in the second edition removed this figure, and ISDN never really took off

Explain The Explained

- **The fact**: the channel capacity is the upper limit, highest possible rate, and practical systems' rates are far lower than the capacity
  - Today (2014) best near-capacity systems operates a few dBs away from channel capacity
- Channel capcity used in the book is for Gaussian signal, is it because practical signal is non-Gaussian?
  - Practical digital signal's capacity is different, but this is unlikely the reason to explain the figure



- Let us write down the channel capacity for AWGN channel with Gaussian signal

$$C = B_p \log_2 \left(1 + \frac{S_P}{N_P} \right) \quad \text{[bps]}$$

  - B_p: channel bandwidth, S_P: signal power, N_P: noise power
- **Channel capacity** in the book is for the ideal channel with the bandwidth (green band)
- **Rate** of BP telephone line with a digital equaliser in the book is for a much wider channel bandwidth (red band)
  - This channel is dispersive (non-ideal), hence equaliser is required
  - In fact, the channel capacity for this dispersive channel with 'red-band' channel bandwidth is much larger than the rate of BP telephone line with a digital equaliser given in the book

Example 1

The transmit and receive filters of a digital communication system have been designed to form a Nyquist system, but the engineer did it only wrote down the transfer function of the receive filter:

$$G_R(f) = \begin{cases} 1, & |f| \le 1000 \text{ Hz} \\ \sqrt{\frac{-|f|+3000}{2000}}, & 1000 \text{ Hz} < |f| \le 3000 \text{ Hz} \\ 0, & |f| > 3000 \text{ Hz} \end{cases}$$

(a) Determine the transfer function of the transmit filter $G_T(f)$.

(b) Determine the required baseband transmission bandwidth. What is the transmission (symbol) rate of this system? What is the roll-off factor of the system?

(c) The channel noise is an additive white Gaussian noise (AWGN) with a flat power spectra density (PSD) $\Phi_n(f) = N_o/2$ watts/Hz for all f. Determine the noise power at the receiver output.

(d) The carrier of this system is 2 MHz. Determine the required passband channel bandwidth and the frequency components of the transmitted radio-frequency (RF) signal.

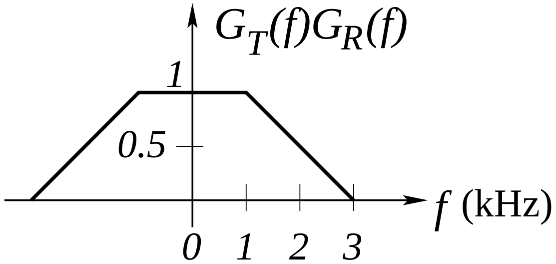
Keys to Remember

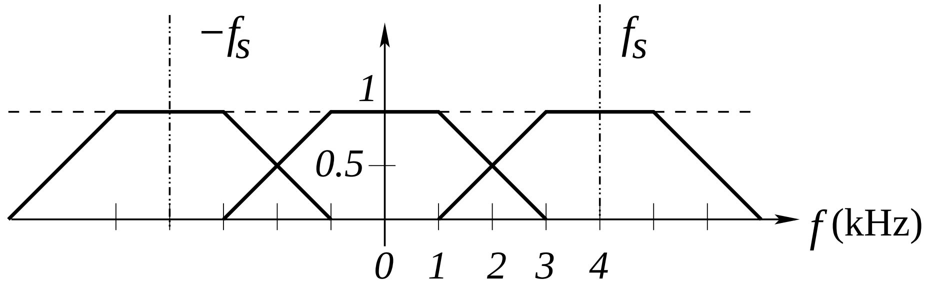
- Transmit and receive filter pair (pulse shaping): 1. achieve zero ISI, and 2. maximise receive SNR
  1. Combined impulse response of Tx and Rx filters has regular zero crossing at symbol-rate spacing $\longrightarrow$ by checking the combined transfer function using Nyquist criterion
  2. Rx filter matches to Tx filter, hence both Rx filter and Tx filter are the squared root of the desired Nyquist filter

(1.a) Tx/Rx (pulse shaping) filters:

$$\Phi_R(f) = G_R(f) \cdot G_T(f) = \text{Nyquist system} \quad \text{and} \quad G_T(f) = G_R(f)$$

(1.b) The baseband bandwidth $B = 3$ kHz.



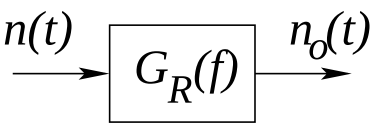


$$\sum_{k=-\infty}^{\infty} \Phi_R(f - kf_s) = \text{constant} \Rightarrow f_s = 4 \ \text{ kHz}$$

The minimum baseband bandwidth for zero ISI is $B_{\min} = \frac{f_s}{2} = 2$ kHz. The roll-off factor

$$\gamma = \frac{B - B_{\min}}{B_{\min}} = \frac{3-2}{2} = 0.5$$

(1.c) Noise passes through $G_R(f)$:

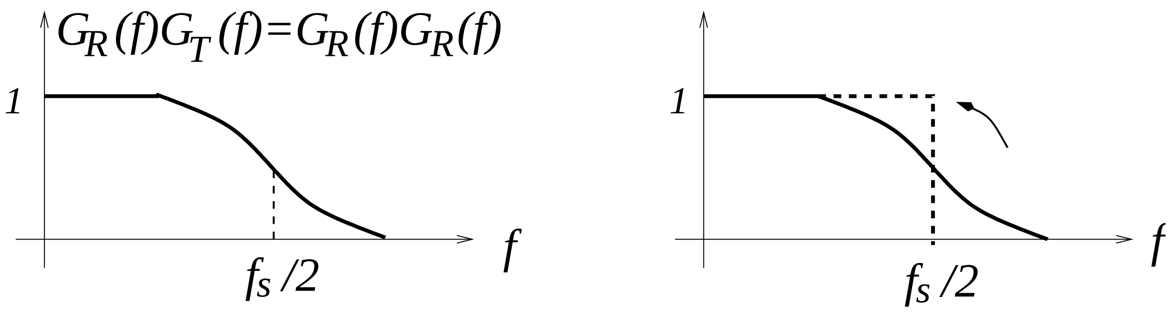


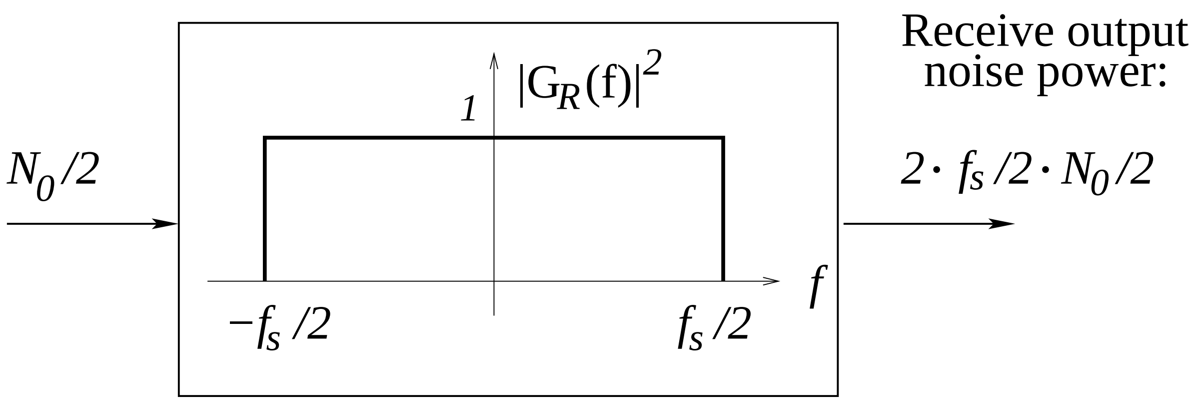
Channel noise PSD $\Phi_n(f) = N_o/2$

Receiver output noise PSD $\Phi_{n_o}(f) = |G_R(f)|^2 \Phi_n(f)$

$$P_{no} = \int_{-\infty}^{\infty} \Phi_{no}(f)df \Rightarrow$$ correct but bad, not fully understand subtle aspects of pulse shaping

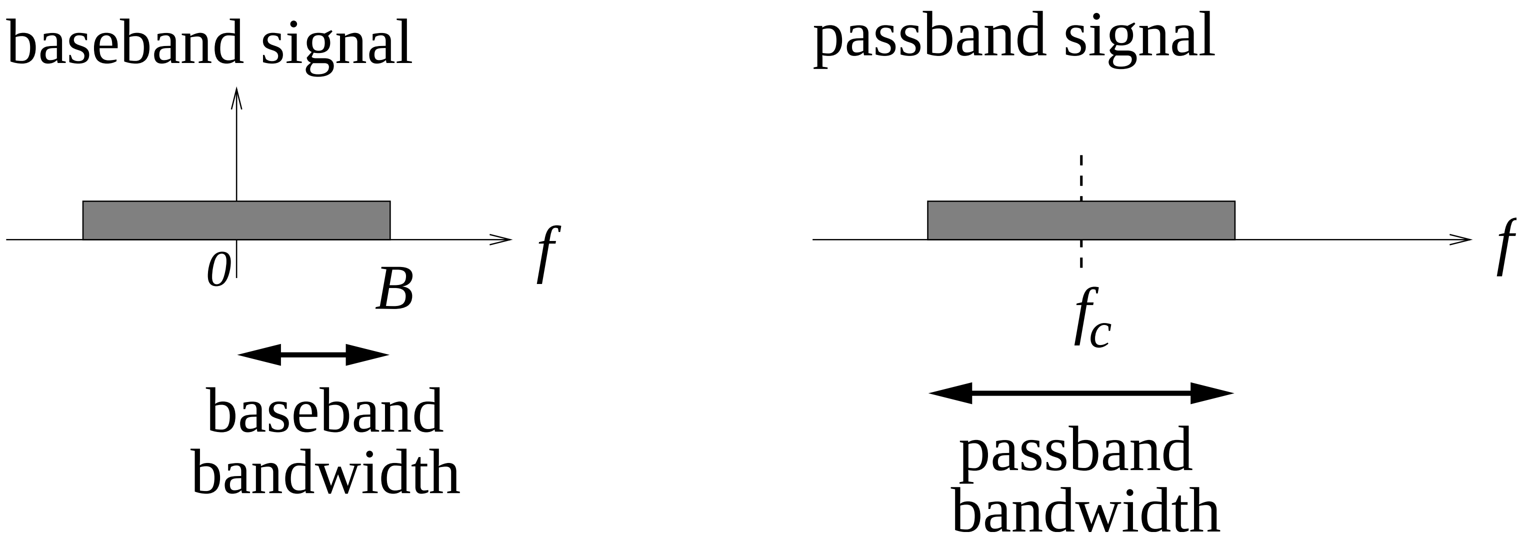
According to Nyquist criterion, combined $G_R(f)G_T(f)$ should be "symmetric" around $\frac{f_s}{2}$:



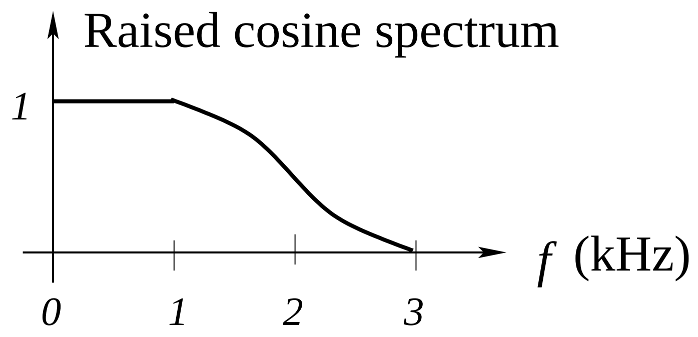


$$P_{no} = 2 \times \frac{f_s}{2} \times \frac{N_0}{2} = f_s \times \frac{N_0}{2} = \frac{N_o}{2} \times 4000 \text{ watts}$$

(1.d) The required passband channel bandwidth $B_p = 2B = 6$ kHz, centered at 2 MHz, and the transmitted RF signal contains frequency components in 1.997 MHz to 2.003 MHz.



- In Example 1, if $G_R(f)$ is changed to a square-root of raised cosine spectrum. Could you do it?



Example 2

In a digital communication system, the bit rate is $R_b = 20$ Mbps, and the raised cosine pulse shaping is employed with the roll off factor of 0.2.

1. Compute the required passband channel bandwidth if the binary phase shift keying (BPSK) modulation scheme is employ.
2. If the passband channel bandwidth is 12 MHz, can you use the above BPSK modulation scheme? Design an appropriate modulation scheme that can be used for this channel.
3. Discuss what is the penalty or cost for your design, in comparison to the BPSK scheme.

Keys to Remember

- Channel capacity tells us to transmit at required rate needs corresponding **bandwidth** and **power** resource
  - Bits-to-symbols mapping or modulation scheme is chosen so that the transmission can fit into the available channel bandwidth and power constraints
  - Higher the modulation scheme order, smaller the required bandwidth but higher the required signal power, and vice versus
- Given transmission rate f_s, the raised cosine pulse shaping with the roll off factor γ requires the baseband bandwidth

$$B = \frac{f_s}{2}(1 + \gamma)$$

  and the required passband channel bandwidth is $B_p = 2B$

1. BPSK modulation is one bit per symbol, hence its symbol rate or transmission rate is $f_s = R_b = 20$ MSymbols/s. Thus the required baseband bandwidth is

$$B = \frac{f_s}{2}(1+\gamma) = \frac{20}{2}(1.2) = 12 \text{ MHz}$$

The required passband channel bandwidth is therefore $B_p = 2B = 24$ MHz.

2. BPSK modulation cannot be used, as it requires 24 MHz passband channel bandwidth, but the channel only offers $B_p = 12$ MHz.

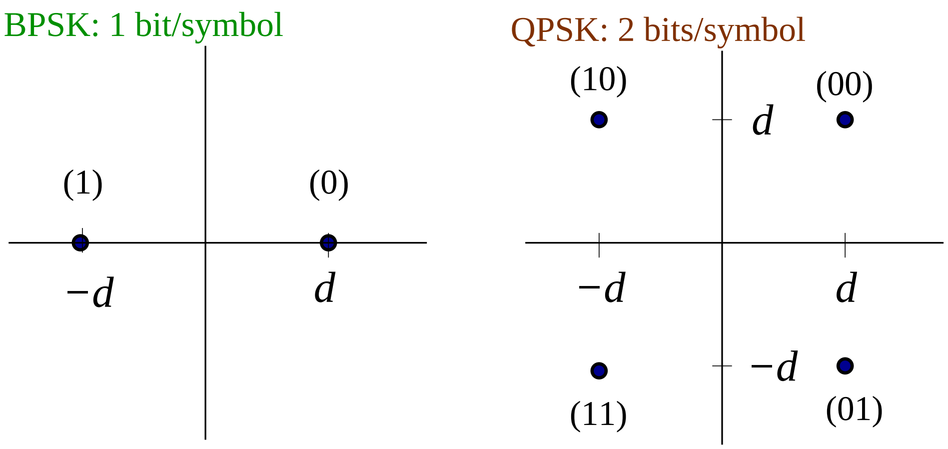
Using a two bits per symbol modulation scheme, such as QPSK, then the symbol rate is $f_s = 10$ MSymbols/s, and the required baseband bandwidth is

$$B = \frac{f_s}{2}(1+\gamma) = \frac{10}{2}(1.2) = 6 \text{ MHz}$$

and the required passband channel bandwidth is therefore $B_p = 2B = 12$ MHz.

3. QPSK requires only half of the channel bandwidth, compared with BPSK
   - but requires more power to achieve the same level of bit error rate performance
   - BPSK: average symbol energy $E_s = d^2$
   - QPSK: average symbol energy $E_s = 2d^2$

This will become more clear in the next example



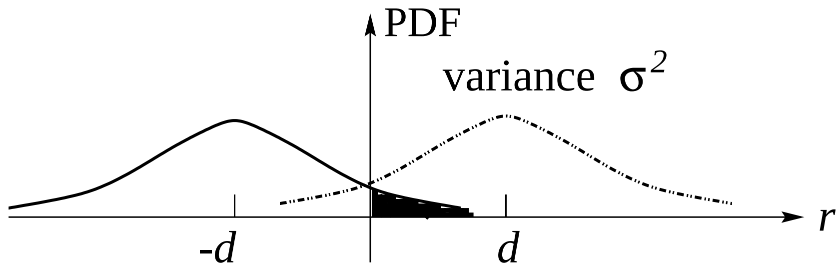
Example 3

1. Write down the bit error rate (BER) of the BPSK communication system. You should express the BER as a function of the channel signal to noise ratio (SNR), where SNR $= \frac{E_s}{N_0}$, E_s denotes the average symbol energy, and $N_0/2$ is the two-side power spectral density of the channel AWGN.
   Based on this BER expression, comment why in the pulse shaping design, the receive filter is matched to the transmit filter.
2. Based on the BPSK result, write down the BER of the QPSK communication system as a function of the channel SNR $= \frac{E_s}{N_0}$.
   By comparing the QPSK result with the BPSK result, what observation can you make?

1. The bit error rate of the BPSK system is given by

$$P_e = Q\left(d/\sigma\right)$$

where $Q(\ )$ is the Gaussian error function.



- Average symbol energy of BPSK is $E_s = d^2$
- Noise power is $\sigma^2 = N_0/2$
- With the definition of SNR $= \frac{E_s}{N_0}$, the BER of the BPSK system

$$P_e = Q\left(d/\sigma\right) = Q\left(\sqrt{\frac{2E_s}{N_0}}\right)$$

$Q(\ )$ function is monotonically decreasing function, as the SNR increases.

- Pulse shaping aims: i) achieving zero ISI, and ii) maximising receive SNR.

ii) is achieved by matching the receive filter to the transmit filter, and hence it is equivalently to minimise the BER

2. QPSK: I and Q components are both BPSK
   - Applying BPSK result to both I and Q yields

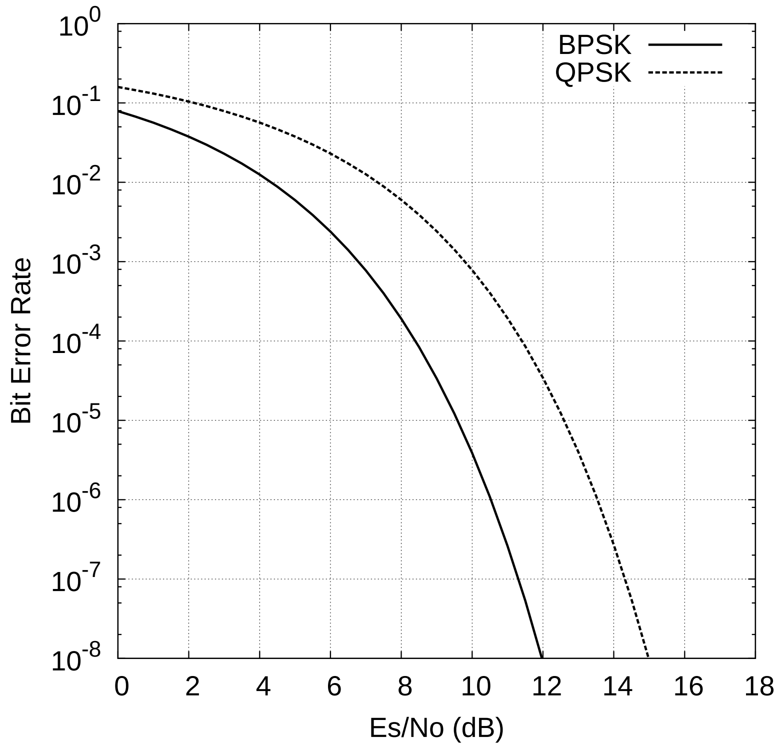
$$P_{e,I} = Q\left(d/\sigma\right) \text{ and } P_{e,Q} = Q\left(d/\sigma\right)$$

   - **Average error rate** for QPSK is then

$$P_e = \frac{1}{2}\left(P_{e,I} + P_{e,Q}\right) = Q\left(d/\sigma\right)$$

   - QPSK average symbol energy is $E_s = 2d^2$, and Noise power is $2\sigma^2 = N_0$, leading to

$$P_e = Q\left(d/\sigma\right) = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

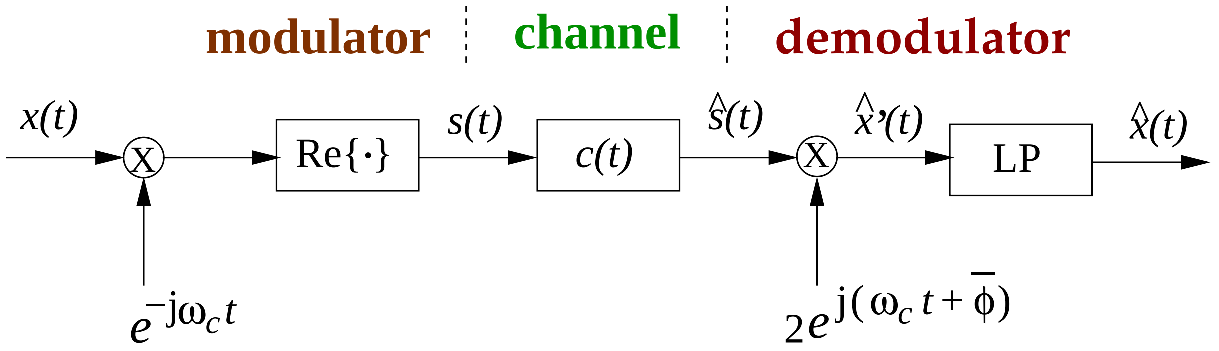


Remember 'no free lunch': For same bit rate R_b, QPSK bandwidth is half of BPSK, but requires higher signal power (factor of 2 or 3 dB) to achieve same level of BER

- This is essentially what channel capacity tells us: trade off between bandwidth and signal power

Example 4

Based on a QAM signal $x(t) = x_\mathrm{i}(t) + \mathrm{j}x_\mathrm{q}(t)$, the transmitted signal in the following figure is given by $s(t) = x_\mathrm{i}(t)\cos\omega_c t + x_\mathrm{q}(t)\sin\omega_c t$.



(a) With a channel impulse response $c(t) = \delta(t - 0.5T_s)$ where T_s is the symbol period, and local carrier of the phase $\bar{\phi}$ as well as a suitably selected lowpass filter LP, show that the receiver output is given by

$$\hat{x}(t) = x(t - 0.5T_s) \cdot e^{\mathrm{j}(\bar{\phi} + \omega_c 0.5T_s)}$$

Sketch the magnitude response of the lowpass filter LP.

(b) What is the best value of $\bar{\phi}$ for the demodulator? Name the component in the receiver that is used to lock into this optimal phase offset.

(c) In the receiver, $\hat{x}(t)$ is sampled at $t = kT_s + \tau$ to produce $\hat{x}[k]$. Determine the best value of τ for the sampler. Name the component in the receiver that is used to find this optimal sampling offset.

(4.a) The receiver input is

$$\begin{aligned}\hat{s}(t) &= x_{\mathrm{i}}(t-0.5T_s)\cos\omega_c(t-0.5T_s) + x_{\mathrm{q}}(t-0.5T_s)\sin\omega_c(t-0.5T_s) \\ &= x_{\mathrm{i}}(\tilde{t})\cos\omega_c\tilde{t} + x_{\mathrm{q}}(\tilde{t})\sin\omega_c\tilde{t}\end{aligned}$$

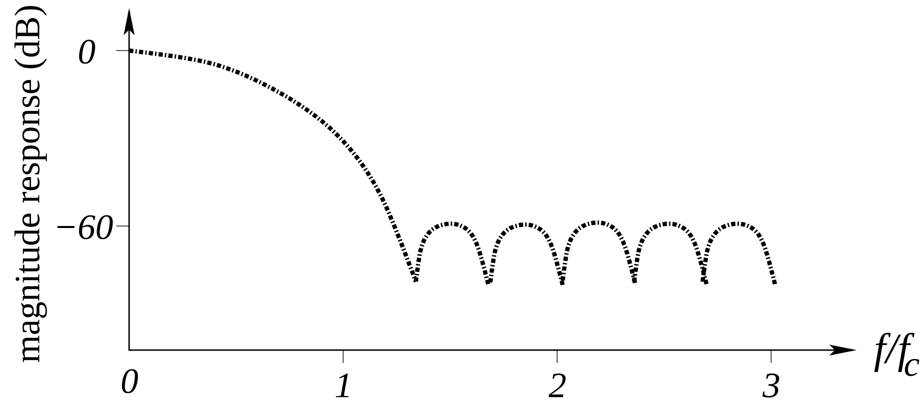
where $\tilde{t} = t - 0.5T_s$.
The demodulator output before LP is

$$\begin{aligned}\hat{x}^{'}(t) &= \hat{s}(t)\cdot 2\cdot e^{\mathrm{j}(\omega_c t+\bar{\phi})} = 2\cdot\hat{s}(t)\cdot e^{\mathrm{j}\omega_c(t-0.5T_s)}\cdot e^{\mathrm{j}(\bar{\phi}+\omega_c 0.5T_s)} \\ &= 2\cdot\hat{s}(t)\cdot e^{\mathrm{j}\omega_c\tilde{t}}\cdot e^{\mathrm{j}(\bar{\phi}+\omega_c 0.5T_s)} \\ &= 2\cdot\left(x_{\mathrm{i}}(\tilde{t})\cos\omega_c\tilde{t} + x_{\mathrm{q}}(\tilde{t})\sin\omega_c\tilde{t}\right)\cdot\left(\cos\omega_c\tilde{t} + \mathrm{j}\sin\omega_c\tilde{t}\right)\cdot e^{\mathrm{j}(\bar{\phi}+\omega_c 0.5T_s)} \\ &= \left\{x_{\mathrm{i}}(\tilde{t})\cdot(1+\cos 2\omega_c\tilde{t} + \mathrm{j}\sin 2\omega_c\tilde{t})\right. \\ &\quad \left. +\mathrm{j}x_{\mathrm{q}}(\tilde{t})\cdot(1-\cos 2\omega_c\tilde{t} - \mathrm{j}\sin 2\omega_c\tilde{t})\right\}\cdot e^{\mathrm{j}(\bar{\phi}+\omega_c 0.5T_s)}\end{aligned}$$

A suitably chosen LP will remove the components modulated at $2\omega_c$

$$\hat{x}(t) = \mathrm{LP}\{\hat{x}^{'}(t)\} = (x_{\mathrm{i}}(\tilde{t}) + \mathrm{j}x_{\mathrm{q}}(\tilde{t}))\cdot e^{\mathrm{j}(\bar{\phi}+\omega_c 0.5T_s)} = x(t-0.5T_s)\cdot e^{\mathrm{j}(\bar{\phi}+\omega_c 0.5T_s)}$$

Sketch of a suitable LP magnitude response



(4.b) The best value for $\bar{\phi}$ is $\bar{\phi} + \omega_c 0.5T_s = 0$ or $\bar{\phi} = -\omega_c 0.5T_s$. The carrier recovery circuit in the receiver is used to lock the local oscillator's phase $\bar{\phi}$ into this optimal phase offset.

(4.c) The best value for τ is $kT_s + \tau - 0.5T_s = kT_s$ or $\tau = 0.5T_s$. In this noise-free case, $\hat{x}[k] = x[k]$, the transmitted symbol sequence. The timing recovery (or clock recovery or synchronisation) circuit in the receiver is used to find this optimal sampling offset.

Coherent System

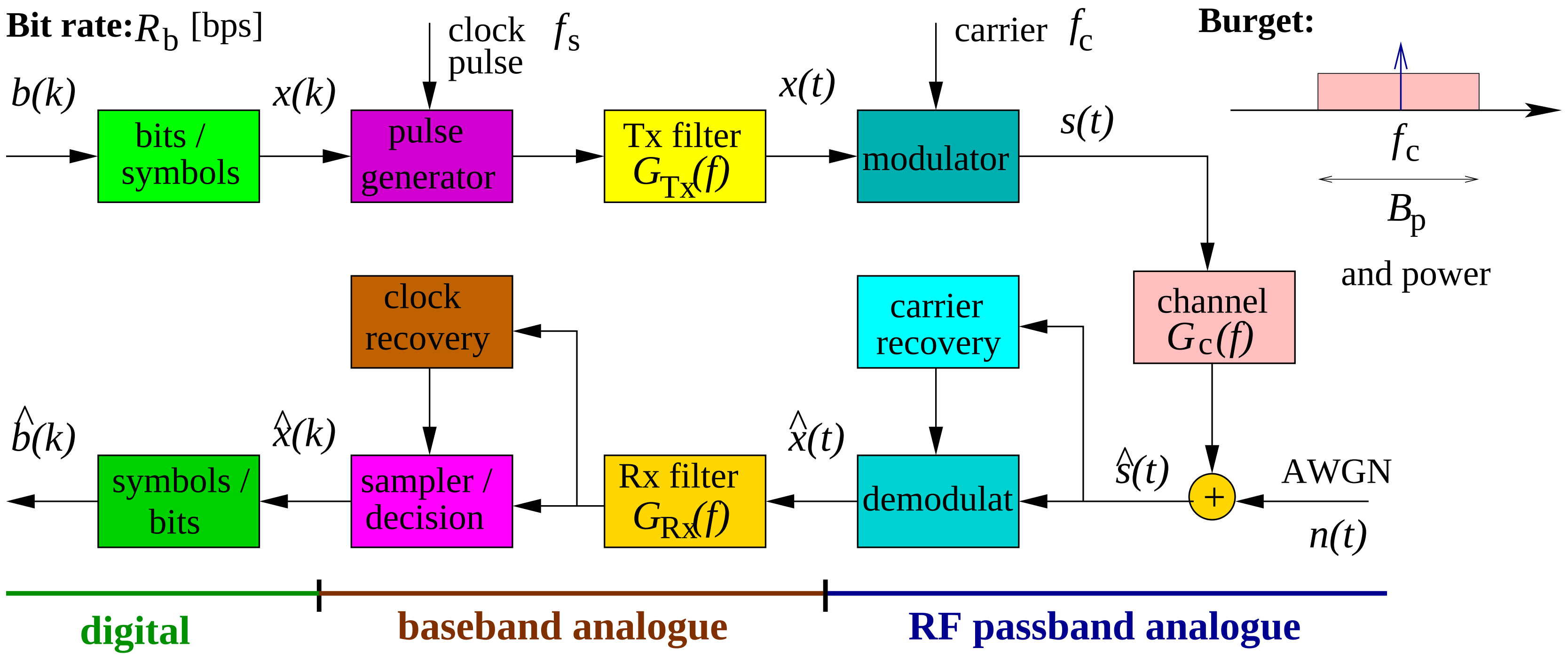
- In the above example, incoming carrier has a phase $\phi = -\omega_c 0.5T_s$, carrier recovery locks the local carrier's phase $\bar{\phi} = \phi$, in order to demodulate correctly
- System without carrier recovery is called non-coherent system, where demodulated signal is

$$\hat{x}(t) = x(t) \cdot e^{\mathrm{j}\Delta\phi} \text{ with } \Delta\phi = \bar{\phi} - \phi \neq 0$$

  – Other means must be employed to remove the influence of the unknown carrier phase ϕ

Example 5

A typical modem is illustrated below, where carrier recovery shown is required for coherent receivers



(a) Explain with the aid of key mathematical equations how the differential phase shift keying (DPSK) based non-coherent receiver works and therefore demonstrates no carrier recovery is necessary for non-coherent receivers.

(b) Compare the non-coherent and coherent receivers in terms of performance and implementation complexity

(b) If the channel is dispersive, an equaliser is required at the receiver. Draw the schematic of adaptive baseband channel equalisation. Explain the two operational modes of the adaptive equaliser.

(5.a) **Differential encoding** at transmitter for transmission

– PSK symbols $\{x_k\}$ are differentially encoded

$$c_k = \begin{cases} 1, & k = 0 \\ x_k \cdot c_{k-1}, & k \geq 1 \end{cases}$$

– Transmitted $\{c_k\}$: $|c_{k-1}|^2 = \text{con}$, and phase of $c_k \cdot c_{k-1}^*$ uniquely determines phase of x_k

Non-coherent detection at receiver

– Receiver samples

$$y_k = c_k \cdot |h| \cdot e^{j\phi} + n_k$$

$|h|$: magnitude of combined channel tap, $\phi \neq 0$: unknown phase

– Differential decoding leads to recovered symbols

$$\hat{x}_k = \frac{y_k \cdot y_{k-1}^*}{|c_{k-1}|^2} = \frac{y_k \cdot y_{k-1}^*}{\text{con}}$$

– For convenience, assuming $|h|^2 = 1$ (or $|h|^2$ is known), then

$$\frac{y_k \cdot y_{k-1}^*}{\text{con}} = \frac{c_k \cdot c_{k-1}^* \cdot e^{j(\phi - \phi)}}{\text{con}} + \frac{n_k \cdot n_{k-1}^*}{\text{con}} + \frac{c_k}{\text{con}} \cdot e^{j\phi} \cdot n_{k-1}^* + n_k \cdot e^{-j\phi} \cdot \frac{c_{k-1}^*}{\text{con}}$$

– Noting magnitudes of $\frac{c_k}{\text{con}}$ and $\frac{c_{k-1}^*}{\text{con}}$ are 1, $\frac{n_k \cdot n_{k-1}^*}{\text{con}}$ is much smaller than the last two terms, while $e^{j\phi} \cdot n_{k-1}^*$ and $n_k \cdot e^{-j\phi}$ have the same variance as n_k,

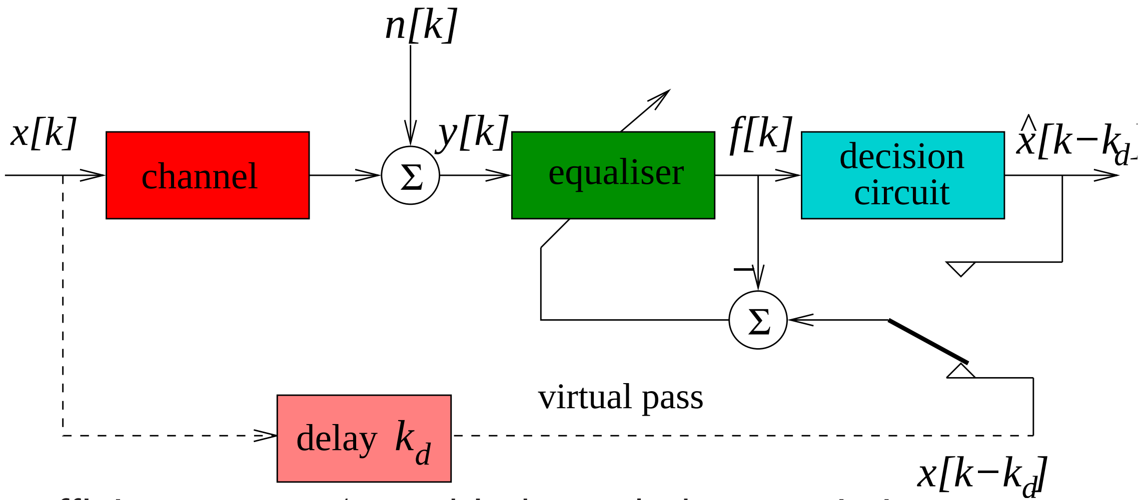
$$\hat{x}_k \approx x_k + 2n_k$$

– Unknown channel phase ϕ is cancelled out, and no carrier recovery is needed at the receiver

(5.b) Compared with coherent detection of $\hat{x}_k \approx x_k + n_k$

- Non-coherent detection **doubles** noise or its SNR is 3 dB worse off
- Does not need expensive and complicated carrier recovery circuit

(5.c) The generic framework of **adaptive** equalisation:



Equaliser sets its coefficients w_i to 'match' channel characteristics

- *Training mode*: Transmitter transmits a prefixed sequence known to receiver, the equaliser uses locally generated symbols $x[k]$ as the desired response to adapt w_i
- *Decision-directed mode*: the equaliser assumes the decisions $\hat{x}[k - k_d]$ are correct and uses them to substitute for $x[k - k_d]$ as the desired response

MMSE Solution and Adaptive LMS Algorithm

Consider dispersive channel

$$y[k] = \sum_{i=0}^{N_C} c_i \cdot x[k - i]$$

and linear equaliser

$$f[k] = \sum_{i=0}^{M} w_i^* \cdot y[k-i]$$

with weight vector $\boldsymbol{w} = \left[w_0 \; w_1 \cdots w_M\right]^{\mathrm{T}}$ and decision delay k_d, i.e. at sampling instance k it estimates transmitted symbol at $k - k_d$

- Equaliser input vector $\boldsymbol{y}_k$ can be expressed as:

$$\boldsymbol{y}_k = \boldsymbol{C} \cdot \boldsymbol{x}_k + \boldsymbol{n}_k$$

  - $\boldsymbol{x}_k = \left[x[k] \;\; x[k-1] \cdots x[k-L]\right]^{\mathrm{T}}$, with length $L = N_c + M$ and symbol power $\mathcal{E}\left\{|x[k]|^2\right\} = \sigma_d^2$
  - Channel AWGN vector $\boldsymbol{n}_k = \left[n[k] \; n[k-1] \cdots n[k-M]\right]^{\mathrm{T}}$ with noise power $\mathcal{E}\left\{|n[k]|^2\right\} = 2\sigma_n^2$
  - $(L+1) \times (L+1)$ CIR convolution matrix has Toeplitz form

$$\boldsymbol{C} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{N_c} & 0 & \cdots & 0 \\ 0 & c_0 & c_1 & \cdots & c_{N_c} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & c_0 & c_1 & \cdots & c_{N_c} \end{bmatrix} = \left[\boldsymbol{c}_0 \; \boldsymbol{c}_1 \cdots \boldsymbol{c}_{k_d} \cdots \boldsymbol{c}_L\right]$$

- The MMSE equalisation solution is given by

$$\boldsymbol{w}_{\mathrm{opt}} = \left(\boldsymbol{C} \cdot \boldsymbol{C}^{\mathrm{H}} + \frac{2\sigma_n^2}{\sigma_d^2} \cdot \boldsymbol{I}_{L+1}\right)^{-1} \cdot \boldsymbol{c}_{k_d}$$

- LMS algorithm: given an initial weight vector $\boldsymbol{w}_0 = \left[w_0[0]\ w_1[0] \cdots w_M[0] \right]^{\mathrm{T}}$, at sample k:

  1. filter output:

$$f[k] = \boldsymbol{w}_k^{\mathrm{H}} \cdot \boldsymbol{y}_k = \sum_{i=0}^{M} w_i^*[k] \cdot y[k-i]$$

  2. estimation error:

$$e[k] = d[k] - f[k]$$

  3. weight adaptation:

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu \cdot e^*[k] \cdot \boldsymbol{y}_k$$

  μ is called adaptive step size, and $d[k]$ is desired response

$$d[k] = \begin{cases} x[k-k_d], & \text{Training mode} \\ \hat{x}[k-k_d], & \text{Decision-directed mode} \end{cases}$$