What's New in Communications

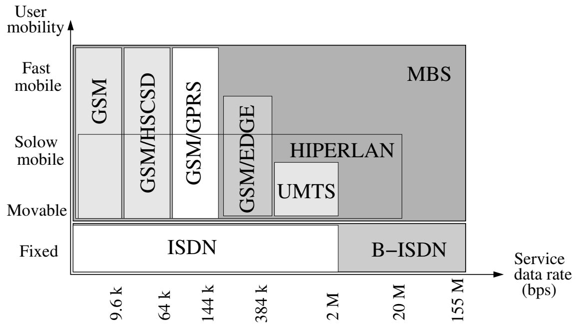
- Imagine a few scenarios:
  - In holiday, use your fancy mobile phone to take picture and send it to a friend
  - In airport waiting for boarding, switch on your laptop and go to your favourite web side
  - Or play game via Internet with your mobile phone
- These are not *Research*, they are not even *Development*, they are *Developed*
- Do you know these words:

  CDMA, multicarrier, OFDM, space-time processing, MIMO, iterative or turbo coding, intelligent network, smart antenna
- There are plenty of opportunities for R & D
- In this introductory course, we only go through some A B C

Research Methods, ELEC6021 (EZ619) S Chen

Wireless and Mobile Networks

- Current/future: 2G GSM, 3G UMTS (universal mobile telecommunication system), and MBS (mobile broadband system) being developed for 4G mobile system



Some improved 2G, HSCSD: high-speed circuit switched data, GPRS: general packet radio service, EDGE: enhanced data rates for GSM evolution. Also, HIPERLAN: high performance radio local area network

Research Methods, ELEC6021 (EZ619) S Chen

For Those Keen to Read

- Any good text books on digital communications, e.g.

  I.A. Glover and P.M. Grant, *Digital Communications*. 2nd edition, Pearson, 2004.

  J.D. Gibson, *Principles of Digital and Analog Communications*. 2nd edition, McMillan, 1993
- More advanced level, e.g.

  L. Hanzo, W. Webb and T. Keller, *Single- and Multi-Carrier Quadrature Amplitude Modulation*. Wiley, 2000.

  L. Hanzo, M. Münster, B.J. Choi and T. Keller, *OFDM and MC-CDMA*. Wiley, 2003

  A. Paulraj, R. Nabar and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003

Research Methods, ELEC6021 (EZ619) S Chen

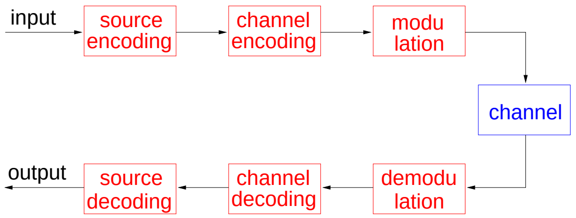
General Comments on Communications

- Aim of telecommunications: to communicate information between geographically separated locations via a communications channel of adequate quality (at certain rate reliably)

- The transmission will be based on digital data, which is obtained from (generally) analogue quantities by
  1. sampling (Nyquist: sampling with at least twice the maximum frequency), and
  2. quantisation (introduction of quantisation noise through rounding off)
- Transmitting at certain rate requires certain spectral **bandwidth**
- Here channel means whole system, which has certain **capacity**, the maximum rate that can be used to transmit information through the system reliably

General Transmission Scheme

- A digital transmission scheme generally involves:



- Input/output are considered digital (analogue sampled/quantised)
- CODEC, MODEM, channel (transmission medium)
- Your 2G mobile phone, for example, contains a pair of transmitter and receiver (transceiver), consisting of a CODEC and MODEM

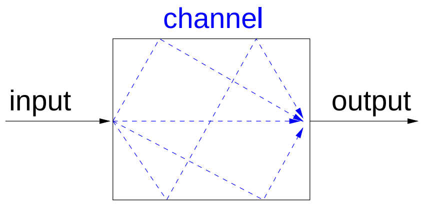
Source Encoding / Decoding

- A digital source is characterised by:
  - source alphabet; symbol rate; symbol probabilities; and probabilistic interdependence of symbols
- Source coding is about how to code symbols (samples) into bits; source decoding is about how to code bits back into symbols (samples)
- Ideally, source encoding should remove any redundancy from the signal to be transmitted, and the source decoder has to restore the original source signal distortionless (lossless coding)
- In practice, a small (potentially imperceptible) error may be allowed in the coding/encoding (lossy coding)

  e.g. speech coding can be lossy, but medical diagnosis images cannot tolerate lossy

Channel Encoding / Decoding

- The channel exhibits impairments:
  - channel distortion (e.g. multipath and fading, insufficient bandwidth), and
  - noise (e.g. thermal noise in receiver amplifier circuit)

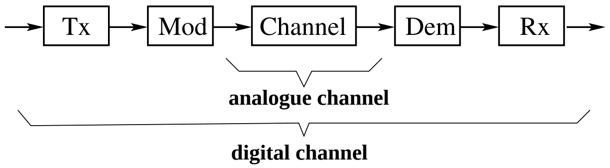


- Depending on the severity of impairment, transmission errors may occur
- Channel coding adds redundancy to the transmitted signal to allow error detection and/or correction on the receiving side

Channel Characteristics

- Channel may introduce amplitude and phase distortion
  - Bandwidth B, signal-to-noise ratio $\frac{S}{N}$, and
  - maximum rate for possible error-free transmission (channel capacity C)

$$C = B \cdot \log_2 \left(1 + \frac{S}{N}\right) \qquad \text{[bits/sec]}$$

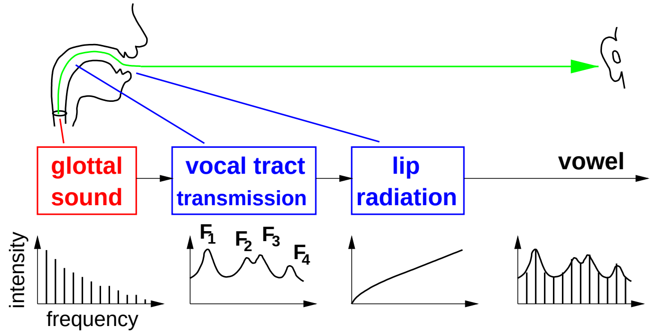


- Channel characteristics depend on design of transmission system:
  - transmission power, bandwidth, speed (data rate), reliability (error rate)

Modulation / Demodulation

- Modulation requires:
  - encoding of a bit stream into a symbol stream;
  - filtering (pulse shaping) to limit the bandwidth;
  - modulation with a carrier frequency
- Demodulation requires:
  - carrier recovery: correct carrier phase have to be found;
  - synchronisation (timing recovery): correct sampling instances have to be found;
  - equalisation: to cope with channel distortions

Human Speech



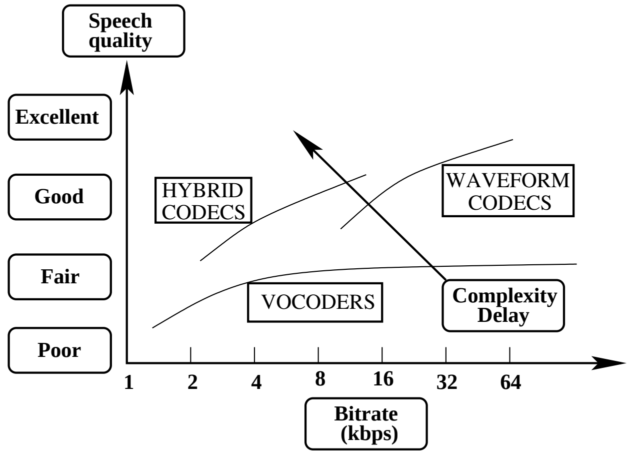
- Speech is either voiced or unvoiced, which results in quasi-periodic or noise-like signals, respectively; both types have a degree of redundancy
  - Voiced speech: vocal folds vibrate at a fundamental frequency (100 . . . 150 Hz for male, 200 . . . 300 Hz for female)
  - Articulation: spectral shaping through the specific resonances of the vocal tract (peaks: formant frequencies, which determine the vowel /a/,/i/,etc.)

Source Coding for Speech

- Methods of source coding for speech:
  - Waveform coding tries to adapt the quantiser characteristic
    $\longrightarrow$ good quality, moderate compression possible
  - Predictive coding tries to quantise only non-redundant information (i.e. the prediction error and predictive model coefficients)
    $\longrightarrow$ good quality, reasonable compression
  - Analysis-by-synthesis is based on a model of the vocal tract
    $\longrightarrow$ up to a very high compression ratio at an arbitrarily poor quality
- Speech codecs:
  - Waveform codecs (waveform coding method)
  - Vocoders (analysis-by-synthesis and predictive coding methods)
  - Hybrid codecs (trade-off between waveform coding and vocoding)

Quality versus Bitrate

- Classification of speech codecs



Waveform Codecs

◯ Time domain waveform coding:

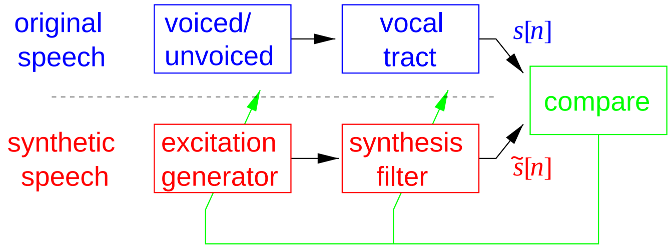
- ITU standard 64 kbps PCM: sampling at 8 kHz with 8 bits quantiser
- ITU standard G.721, 32 kbps adaptive differential PCM (ADPCM)
- ITU standards G.726 and G.727
  - G.726 codecs: variable-rate for rates 16 – 40 kbps, allowing the network to adjust quality/bitrate on instantaneous requirement
  - G.727 codecs: core-bits and enhancement bits, allowing the network to drop enhancement bits under heavy loads and to keep them when in light loads

◯ Frequency domain waveform coding:

- Sub-band coding and adaptive transform coding

Analysis-by-Synthesis Coding

- The speech is divided into small segments of 20ms duration;
- for each segment, a *parametric* model (excitation and synthesis filter coeff.) is sought for the speech generation:



- excitation generation: either noise (unvoiced) or a pulse train (voiced);
- the synthesis filter matches the spectral shape of the harmonics for voiced excitation, and the vocal tract

Linear Predictive Coding

- General aim of predictive coding is to make prediction error or residual as $\mathit{unpredictable}$ as possible
- Linear predictive coding (LPC) employs more complex excitation models than used in analysis-by-synthesis
  - LPC coefficients and prediction residuals are quantised to give designed bitrate
  - e.g. standard 13 kbps GSM speech codec: 8 LPC coefficients are encoded with 36 bits/20 ms update interval $\rightarrow$ 1.8 kbps
- Short-term prediction: speech analysis segments ≤ 20 ms duration
- Long-term prediction (LTP): LTP synthesis filter models the fine structure of the speech spectrum, after short-term prediction
  - When employing a LTP, the residual error becomes truly unpredictable $\rightarrow$ code excited linear predictive coding (CELP) codecs

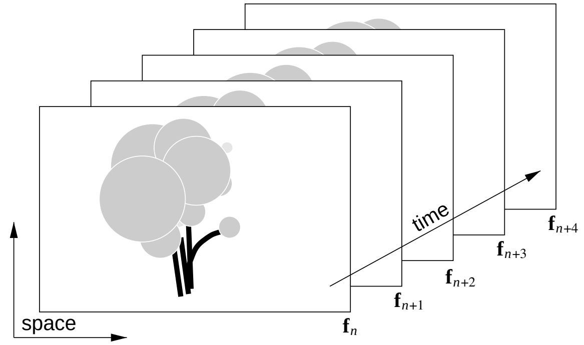
Half-Rate GSM Speech Codec

- 5.6 kbps vector sum excited linear predictive (VSELP) codec, similar to CELP
- Four synthesis modes, depending on grade of voicing detected in the speech
- Decisions on which excitation mode to use is based on LTP gain: high – highly correlated voiced; low – noise-like uncorrelated unvoiced
- 5 bits for overall frame energy spans a dynamic range of 64 dB, when using a stepsize of 2 dB and 32 steps
- Residual error is coded as a codeword in a code book

parameter	bits/frame
LPC coefficients	28
LPC interpolation flag	1
excitation mode	2
mode 0:	
codebook 1 index	$4 \times 7 = 28$
codebook 2 index	$4 \times 7 = 28$
modes 1, 2, 3:	
LTPD (subframe 1)	8
Δ LTPD (subframes 2, 3, 4)	$3 \times 4 = 12$
codebook 3 index	$4 \times 9 = 36$
frame energy E_F	5
excitation gain-related quantity $[E_{S_1} E_1]$	$4 \times 5 = 20$
total no of bits	112 bits/20 ms
bitrate	5.6 kbps

Video Source Coding

- Video signal: luminance signal of two spatial and a temporal dimension:



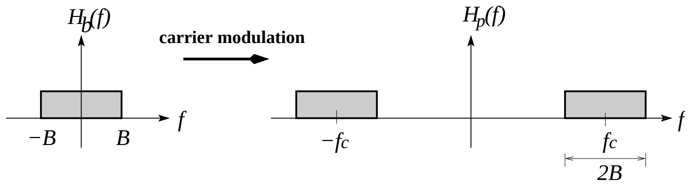
- $\mathbf{f}_n$ is frame at time n;
- there are intra- (i.e. spatial) and inter-frame correlations, i.e. correlations within $\mathbf{f}_n$ and between $\mathbf{f}_n$ and past frames $\mathbf{f}_{n-1}$, $\mathbf{f}_{n-2}$, etc.

Video Compression

- A pixel in a colour image can be presented
  - An independent intensity (luminance), and two colour (chrominance) signals known as hue and saturation; or
  - Three colour signals, intensity values of red, green and blue
- MPEG standards, basic blocks are
  - Motion compensation
  - DCT
  - Variable length coding
- MPEG-1: 1.5 Mbps; MPEG-2: 2 to 8 Mbps, now upto 30 Mbps for digital TV
- MPEG-4: initiated in 1993, for interactive multimedia applications
- MPEG-7: started in 1997, for large stored database

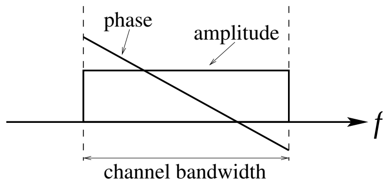
Channel Characteristics (I)

- Passband channel and baseband (remove modulator/demodulator) equivalence:



baseband channel bandwidth $B \leftrightarrow$ passband channel bandwidth $2B$

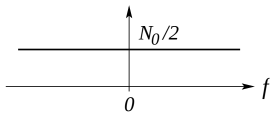
- Channel has *finite* bandwidth, ideally phase is linear and amplitude is flat:



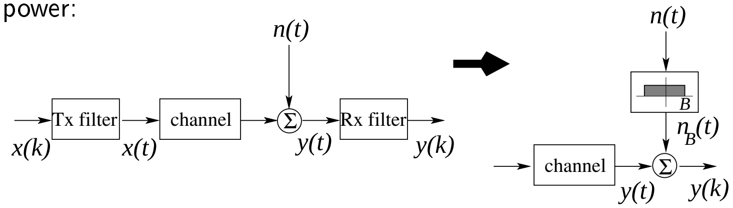
- Bandwidth is the most precise resource
- Two killer factors: multipath and fading

Channel Characteristics (II)

- Channel noise: AWGN with a constant power spectrum density (PSD);



- Power is the area under PSD, so WN has infinitely large power;
- but communication channels are bandlimited, so noise is also bandlimited and has a finite power:



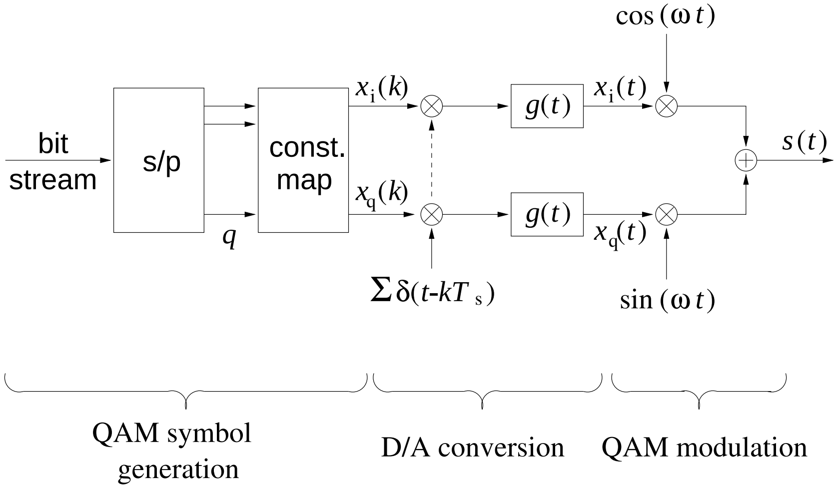
- Channel has certain *capacity*

Digital Modulation

- Aim of MODEM: transmit bit stream at certain rate reliably

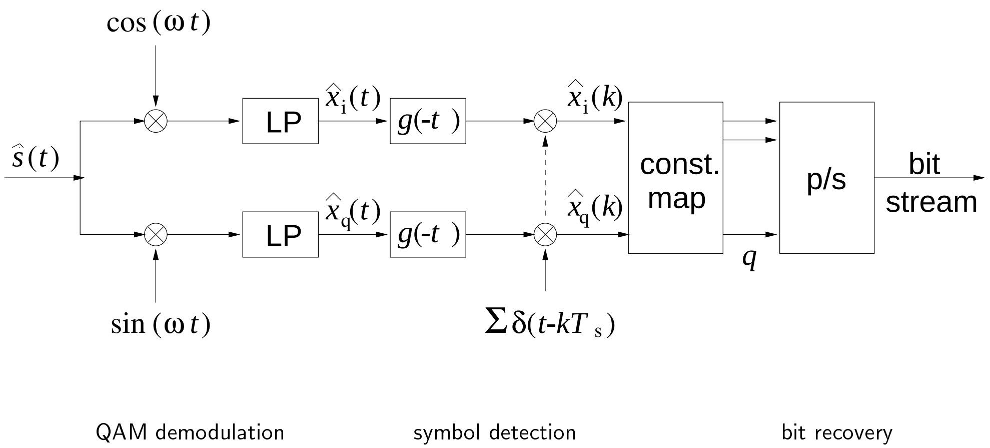
- Carrier communications
  - Low frequency signals (baseband) cannot travel far
  - Most spectrum resources (channels) are in RF bands (high frequency passband)
- Carrier: $A\cos(2\pi f_c t + \phi)$, three quantities, amplitude, frequency and phase
- Basic digital modulation forms: amplitude shift keying (ASK), frequency shift keying (FSK), and phase shift keying (PSK)
- We will consider quadrature amplitude modulation (QAM), which can be viewed as a combination of ASK and PSK

Quadrature Amplitude Modulation



Note: e.g., odd bits go to form $x_{\rm i}(k)$ and even bits to form $x_{\rm q}(k)$; $x_{\rm i}(k)$ and $x_{\rm q}(k)$ are inphase and quadrature components of the $x_{\rm i}(k) + jx_{\rm q}(k)$ QAM symbol; $x_{\rm i}(k)$ and $x_{\rm q}(k)$ are M-ary symbols.

Quadrature Amplitude Demodulation

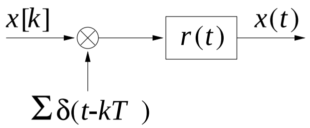


Basic circuits: carrier recovery, timing-recovery, detection

Pulse Shaping

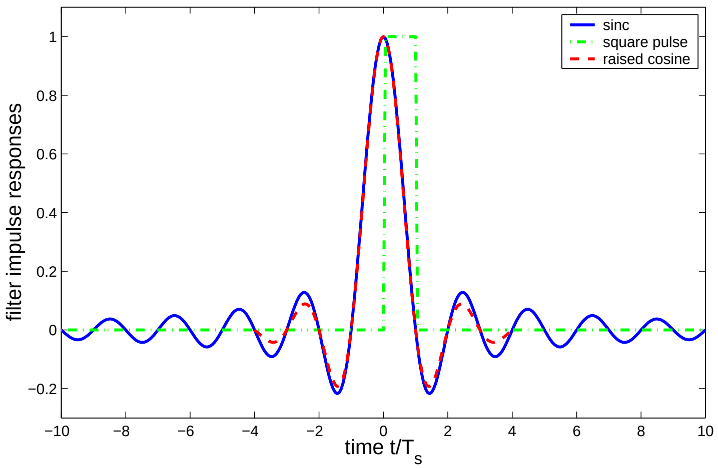
- Channel has finite bandwidth, and transmitted digital signal has to be pulse shaped
- To transmit at symbol rate f_s needs certain bandwidth B_T and B_T depends on which pulse shaping used

$$x(t) = r(t) \star \sum_{k=-\infty}^{+\infty} x[k]\delta(t - kT_s) = \sum_{k=-\infty}^{+\infty} x[k] \cdot r(t - kT_s)$$



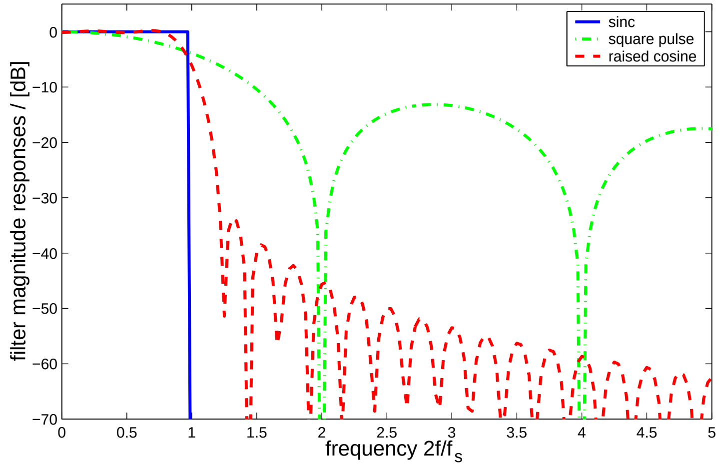
- Pulse shaping filter $r(t)$ would allow to retrieve the original digital data $x[k]$ from $x(t)$: by limiting the bandwidth of $x(t)$ to $B_T \leq$ channel bandwidth

Pulse Shaping — Time Domain



- Impulse response of all these filters have regular symbols-spaced zero-crossing (Nyquist system), but different supports; raised cosine shown is a truncated one

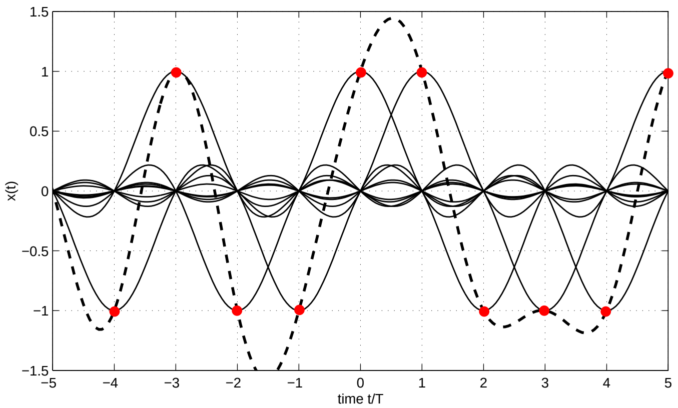
Pulse Shaping — Frequency Domain



- Spectrum of the square pulse produces considerable excess bandwidth beyond the symbol rate f_s; sinc impractical to realize; truncated raised cosine easy to realize

Pulse Shaping — Example

- Binary (± 1) $x[k]$, each is transmitted as a sinc pulse; the peak of different shifted sinc functions coincide with zero crossings of all other sincs:



- At receiver, sampling at correct symbol rate enables recovery of transmitted $x[k]$

Transmit and Receive Filters

- Pulse shaping fulfills two purposes: limit the transmitted bandwidth, and enable to recover the correct sample values of transmitted symbols; such a pulse shaping $r(t)$ is called a Nyquist system

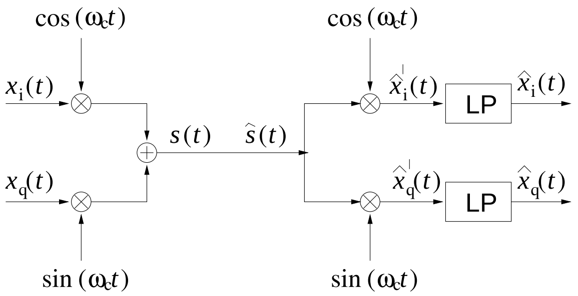
  1. sinc has a (passband) bandwidth $B_T = f_s$, (infinite) raised cosine has $f_s \leq B_T \leq 2f_s$ depending on roll-off factor
  2. a Nyquist time pulse have regular zero-crossing at symbol-rate spacings to avoid interference with neighboring pulses at correct sampling instances

- The Nyquist system $r(t)$ is separated into transmit filter $g(t)$ and receive filter $g(-t)$ (square-root Nyquist systems)

  1. the filter $g(-t)$ in the receiver is also called a matched filter (to $g(t)$); $g(t)$ and $g(-t)$ are basically identical (square-root of $r(t)$)
  2. this division of $r(t)$ enables suppression of out-of-band noise and results in the maximum received SNR

QAM Modulator / Demodulator

- Recall modulator and demodulator of the QAM scheme:



- Modulation of in-phase and quadrature components to carrier frequency ω_c:

$$x_\mathrm{i}(t) \cdot \cos(\omega_c t) \quad \text{and} \quad x_\mathrm{q}(t) \cdot \sin(\omega_c t)$$

- The transmitted signal is: $s(t) = x_\mathrm{i}(t) \cdot \cos(\omega_c t) + x_\mathrm{q}(t) \cdot \sin(\omega_c t)$

QAM — Demodulation

- To explain the demodulation, assume perfect transmission $\hat{s}(t) = s(t)$
- Demodulation for the "in-phase" component:

$$\begin{aligned} \hat{x}'_\mathrm{i}(t) &= s(t) \cdot \cos(\omega_c t) = (x_\mathrm{i}(t) \cdot \cos(\omega_c t) + x_\mathrm{q}(t) \cdot \sin(\omega_c t)) \cdot \cos(\omega_c t) \\ &= x_\mathrm{i}(t) \cdot \frac{1}{2} \cdot \big(1 + \cos(2\omega_c t)\big) + x_\mathrm{q}(t) \cdot \frac{1}{2} \cdot \sin(2\omega_c t) \end{aligned}$$

- If the lowpass filter LP is selected appropriately (cut-off frequency $\leq \omega_c$), the components modulated at frequency $2\omega_c$ can be filtered out, and hence:

$$\hat{x}_\mathrm{i}(t) = \mathsf{LP}\big(\hat{x}'_\mathrm{i}(t)\big) = \frac{1}{2} x_\mathrm{i}(t)$$

- A similar calculation can be performed for the demodulation of $\hat{x}_\mathrm{q}(t)$:

$$\hat{x}'_\mathrm{q}(t) = x_\mathrm{i}(t) \cdot \frac{1}{2} \cdot \sin(2\omega_c t) + x_\mathrm{q}(t) \cdot \frac{1}{2} \cdot \big(1 - \cos(2\omega_c t)\big)$$

Modulation — Complex Notation

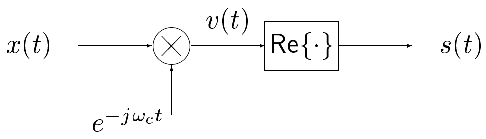
- The modulation/demodulation can be expressed in complex notation — in-phase and quadrature components are "real" and "imaginary" part of the signal:

$$x(t) = x_\mathrm{i}(t) + j \cdot x_\mathrm{q}(t)$$

- The transmitted signal is obtained by taking the real part only of a complex carrier ($e^{-j\omega_c t}$) modulated signal:

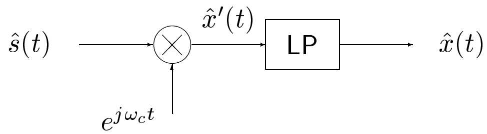
$$s(t) = \mathsf{Re}\{v(t)\} = \mathsf{Re}\{x(t) \cdot e^{-j\omega_c t}\} = x_\mathrm{i}(t) \cdot \cos(\omega_c t) + x_\mathrm{q}(t) \cdot \sin(\omega_c t)$$

- Flow graph:



Demodulation — Complex Notation

- Flow graph for the complex demodulation scheme:



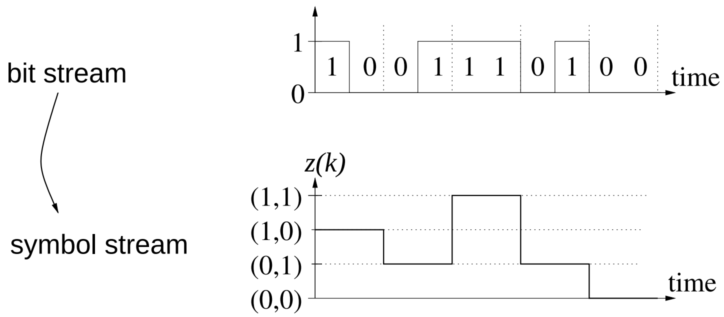
- The demodulated signal:

$$\begin{aligned} \hat{x}'(t) &= e^{j\omega_c t} \cdot s(t) \\ &= (\cos(\omega_c t) + j \sin(\omega_c t)) \cdot (x_\mathrm{i}(t) \cdot \cos(\omega_c t) + x_\mathrm{q}(t) \cdot \sin(\omega_c t)) \\ &= x_\mathrm{i}(t) \cdot \frac{1}{2}\big(1 + \cos(2\omega_c t) + j \sin(2\omega_c t)\big) \;\; + \\ &\quad\; j x_\mathrm{q}(t) \cdot \frac{1}{2}\big(1 - \cos(2\omega_c t) - j \sin(2\omega_c t)\big) \end{aligned}$$

- The lowpass filter (LP) will again remove components modulated at $2\omega_c$

Bits to Symbols

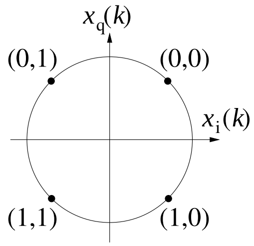
- The bit stream to be transmitted is serial / parallel multiplexed onto a stream $z[k]$ of symbols of q bits/symbol (discrete 2^q levels)
- Example for $q = 2$ bits/symbol: 4-ary modulation



- Symbol rate is half of bit rate, and required bandwidth is half (but transmit power has to be increased); symbol stream is then pulse shaped ...

Mapping to Constellation Pattern

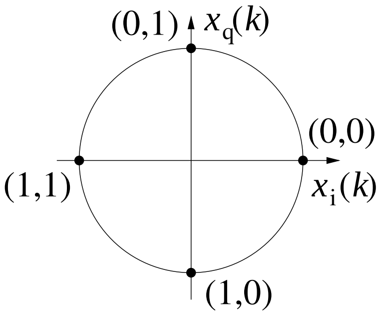
- The symbols $z[k]$ are translated into values for the in-phase and quadrature components, $x_{\mathrm{i}}[k]$ and $x_{\mathrm{q}}[k]$, by assigning them to points in a *constellation pattern*
- Example for a case of $q = 2$ bits/symbol:

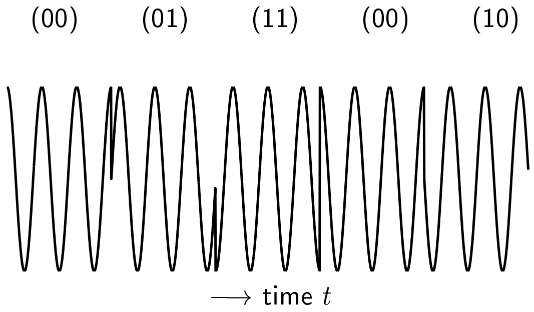


- From the constellation pattern, the values $x_{\mathrm{i}}[k]$ and $x_{\mathrm{q}}[k]$ are determined
- In the receiver, the constellation point (and therefore the transmitted symbol) is determined from $\hat{x}_{\mathrm{i}}[k]$ and $\hat{x}_{\mathrm{q}}[k]$

Phase Shift Keying (PSK)

- Phase shift keying (PSK): carrier phase is used to carry symbol information
- Example of QPSK (minimum phase separation $\frac{\pi}{2}$) constellation pattern and transmitted signal $s(t)$:

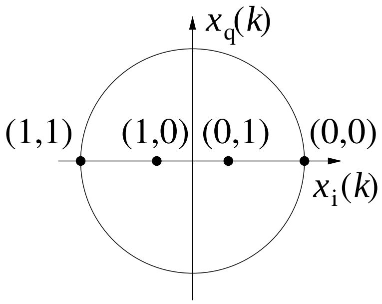


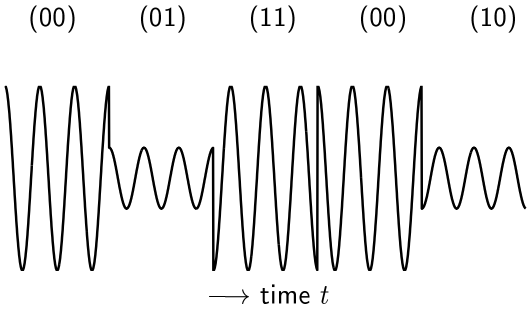


- There are means to smooth discontinuities in the signal $s(t)$

Amplitude Shift Keying (ASK)

- Amplitude shift keying (ASK): carrier amplitude is used to carry symbol information
- Example of 4-ary constellation pattern and transmitted signal $s(t)$:

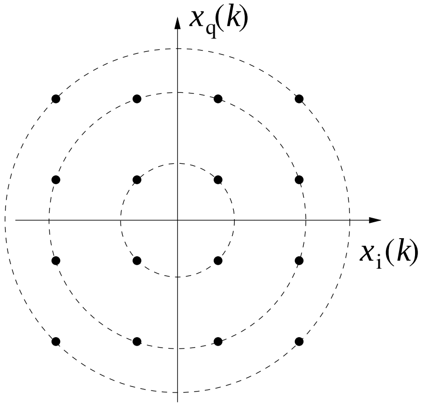




- Note: (i) the quadrature component is not used; (ii) this is not "purely" ASK, as a phase shift of π is exploited in the modulation scheme.

Combined ASK / PSK

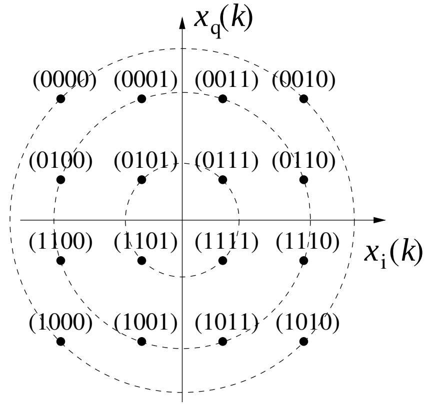
- QAM: combines features of PSK and ASK, and is bandwidth very efficient
- Example of 16-QAM (4 bits per symbol):



- Depending on the channel quality, 64-QAM (6 bits/symbol), or 256-QAM (8 bits/symbol) are possible

Gray Mapping

- If noise or distortions are likely to cause misclassification in the receiver, Gray code mapping can minimize the bit error rate:



- Adjacent constellation points only vary in a single bit (minimum Hamming distance)

Carrier Recovery — Phase Offset

- Previously, we assume

$$\hat{s}(t) = s(t) = x_{\mathrm{i}}(t) \cdot \cos(\omega_c t) + x_{\mathrm{q}}(t) \cdot \sin(\omega_c t)$$

  so that we can use $e^{j\omega_c t}$ to remove carrier in demodulator

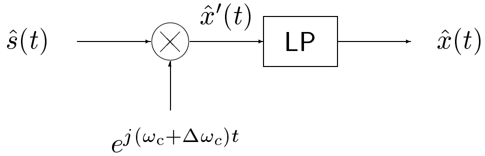
- Most likely, the transmitted signal having traveled to the receiver will accumulate a phase offset φ:

$$\hat{s}(t) = x_{\mathrm{i}}(t) \cdot \cos(\omega_c t + \varphi) + x_{\mathrm{q}}(t) \cdot \sin(\omega_c t + \varphi)$$

- Thus, the receiver has to "recover" the carrier $e^{j(\omega_c t + \varphi)}$ (in fact the phase φ) in oder to demodulate the signal correctly
- Usually, this is done by means of some phase lock loop based carrier recovery

Carrier Recovery — Frequency Offset

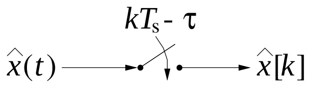
- Tx and Rx frequency generators are unlikely to match exactly. Consider demodulation with a Rx local "carrier" having a frequency offset $\Delta\omega_c$:



- Even assuming $\hat{s}(t) = s(t)$, the demodulated signal prior to sampling is $\hat{x}(t) = x(t) \cdot e^{j\Delta\omega_c t}$, not $\hat{x}(t) = x(t)$
- The effect of carrier frequency mismatch is $\Delta\omega_c t$ and, like the phase difference φ, it has to be compensated at the receiver
- $\Delta\omega_c t + \varphi$ is called carrier offset, and has to be "recovered" in order to demodulate the signal correctly

Synchronisation

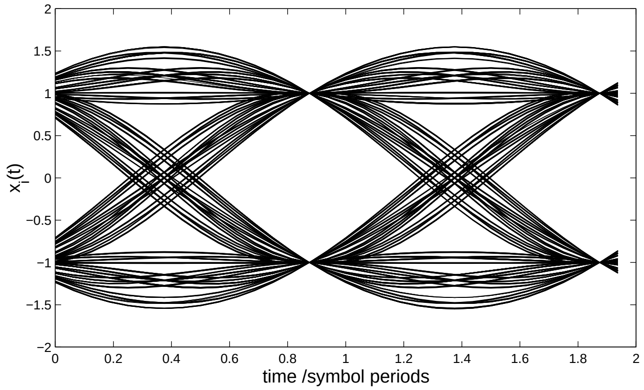
- The process of selecting the correct sampling instances is called synchronisation, also known as timing or clock recovery
- Tx and Rx clocks are likely to have mismatch, clock recovery tries to synchronise the receiver clock with the symbol-rate transmitter clock to obtain samples at appropriate instances
- This is equivalent to replacing the impulse train $\sum \delta(t - kT_\text{s})$ by $\sum \delta(t - kT_\text{s} - \tau)$ with $0 \leq \tau \leq T_\text{s}$:



- The demodulated signal can be *oversampled*, and from the distribution (histogram) of the sample sets for different τ, the one with the smallest deviation from discrete levels (depending on the QAM mode, 16-QAM, 64-QAM, etc.) is chosen

Eye Diagram — Perfect Channel

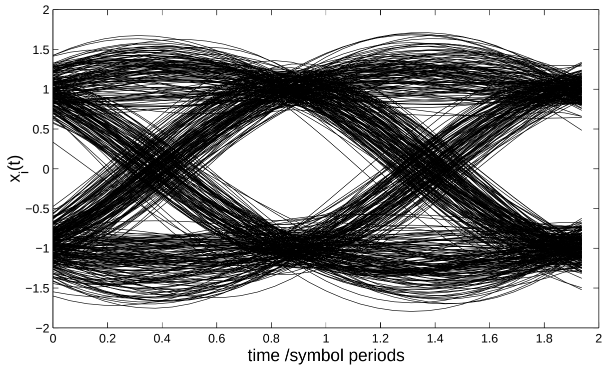
- We are looking at stacked 2 symbol period intervals of the demodulated signal $\hat{x}_\text{i}(t)$ in a QPSK scheme ($\hat{x}_\text{i}(t)$ is BPSK):



- This is called an eye diagram; ideal sampling of $\hat{x}_\text{i}(k)$ will sample the crossing points $\hat{x}_\text{i}(t) = \pm 1 \quad \longrightarrow$ clock/timing recovery ($\tau \approx 0.85 T_s$)

Eye Diagram — Noisy Channel

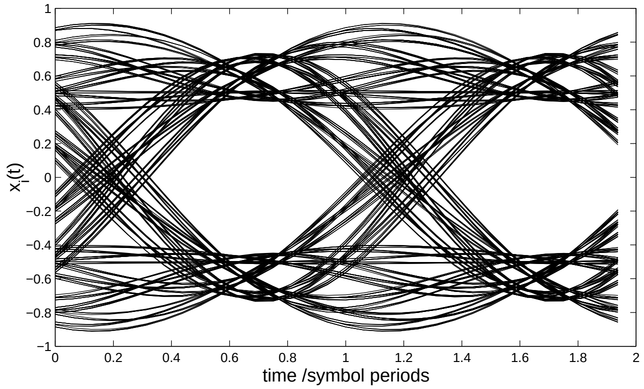
- With channel noise at 3dB SNR, the eye diagram looks different:



- But as long as the sampling points can be clearly determined and the eye is "open", $\hat{x}_\text{i}[k]$ will correctly resemble $x_\text{i}[k]$. At higher noise levels, misclassifications can occur if the eye is "closed"

Eye Diagram — Distorting Channel

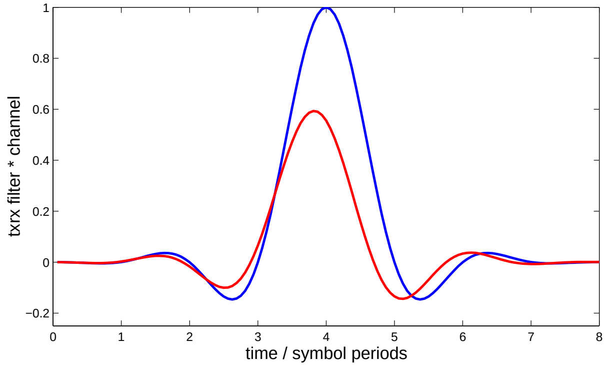
- The channel is non-ideal with an impulse response $c(t) = \delta(t) - \frac{1}{2} \cdot \delta(t - T_s/4)$, where T_s is the symbol period:



- The eye diagram is distorted; this together with noise effect will make the eye completely closed, leading to misclassification due to intersymbol interference

Intersymbol Interference (ISI)

- The response of an ideal pulse shaping filter with regular symbol-spaced zero crossings, and the same system in combination with the channel impulse response $c(t) = \delta(t) - \frac{1}{2} \cdot \delta(t - T_s/4)$:



- The system, Tx-filter – channel – Rx filter, has lost the property of a Nyquist system; peaks of the function no longer coincide with zero crossings of neighbouring pulses

Equalisation

- For an ideal channel without ISI, sampled receiver output is

$$\hat{x}[k] = c \cdot x[k - \Delta] + n[k]$$

- If the channel has sever amplitude and phase distortion, sampling alone is unable to recover the correct symbols; This is because now (note the ISI)

$$\hat{x}[k] = \sum_{i=0}^{n_c} c_i \cdot x[k - i] + n[k]$$

- An equaliser is required to combat the channel distortion, and a typical linear equaliser is defined by:

$$y[k] = \sum_{l=0}^{n_w} w_l \cdot \hat{x}[k - l]$$

- The equaliser soft output $y[k]$ is used to determine the transmitted symbol $x[k - \Delta]$

Equalisation Design

- Let the z-transforms of the channel and equaliser be $C(z)$ and $W(z)$; Similarly define $X(z)$, $\hat{X}(z)$ and $Y(z)$

$$\xrightarrow{X(z)} \boxed{C(z)} \xrightarrow{\hat{X}(z)} \boxed{W(z)} \xrightarrow{Y(z)}$$

- Zero-forcing: we want to find an equaliser $W(z)$ such that $Y(z)$ is a delayed version of the transmitted signal, $Y(z) = z^{-\Delta} \cdot X(z)$
  - The solution is

$$W(z) \cdot C(z) \ = \ z^{-\Delta} \qquad \text{or} \quad W(z) \ = \ z^{-\Delta} \cdot C^{-1}(z)$$

  - Completely eliminate ISI, but amplify the noise too much

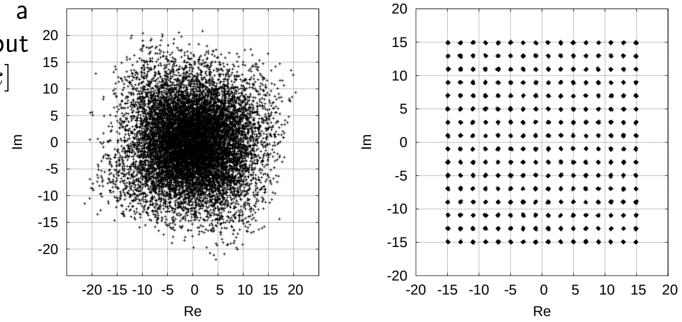
- Minimum mean square error: choose the equaliser $W(z)$ to minimise

$$\text{MSE} = \mathcal{E}[|x[k - \Delta] - y[k]|^2]$$

Adaptive Equalisation

- Training: periodically provide the receiver with $x[k]$ so that equaliser can adjust its weights w_i using $\{x[k], \hat{x}[k]\}$
  - This happens for example in your GSM mobile phone
  - Training requires extra bandwidth and may not always be possible

- Blind: equaliser has to adjust its weights having only the received signal $\{\hat{x}[k]\}$
  - Constellations of a blind equaliser's input $\hat{x}[k]$ and output $y[k]$

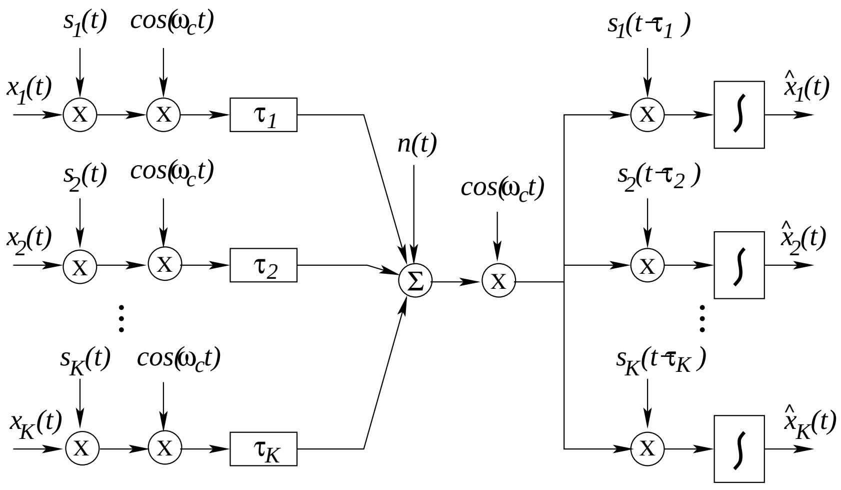


User / Channel Separation

- System has large capacity and supports many users; different users or "channels" much separate in some way:
  - FDMA: separation is in frequency domain, different users have different carriers and channels occupy different frequency bands
  - TDMA: same carrier or frequency band, but separation is in time domain, different users occupy different time slots
  - CDMA: no physical separation, i.e. same carrier and time slot, but virtual separation is in "code" domain, different users have different spreading codes
  - SDMA: if no separation in frequency or time or code domain, separation can still be achieved in spatial domain – this further enhances spectrum efficiency
- CDMA is a key technology in 3G systems, and SDMA has an important role in 4G systems

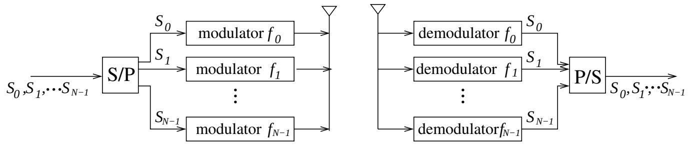
DS-CDMA System

- DS-CDMA uplink, supporting K users

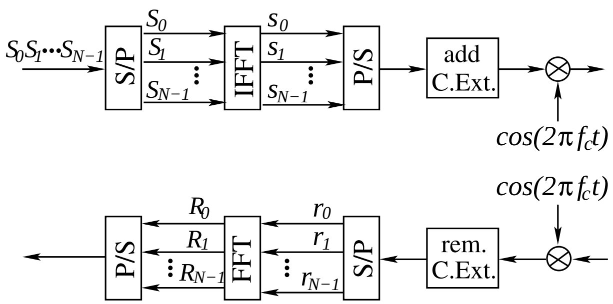


OFDM

- Channel is divided into many subchannels, each supports a subcarrier; Data stream is divided into several bit streams, which are used to modulate several carriers

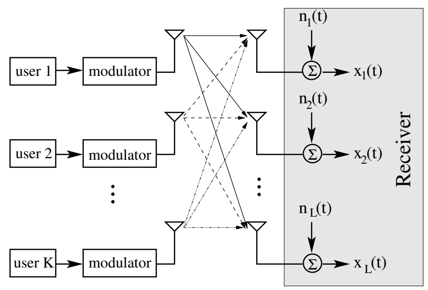


- OFDM transceiver

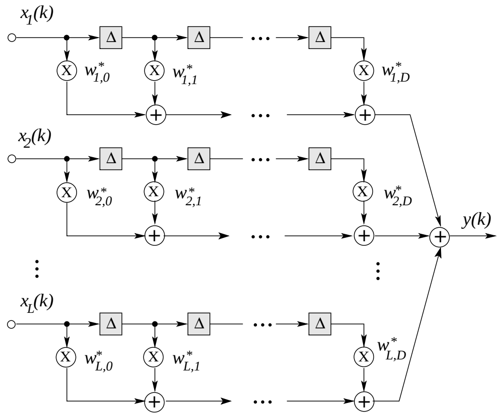


MIMO

- SDMA induced MIMO system



- Space-time equalisation



Channel Capacity and Channel Coding

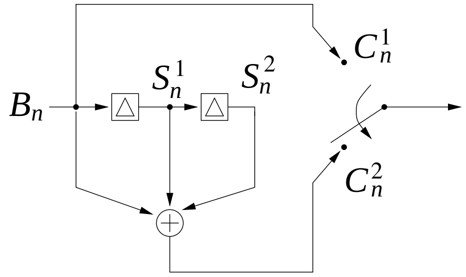
- Even in case of a noisy channel, there is no obstruction of reliable transmission, but only a limitation of the rate at which transmission can take place
- Reliable transmission can be established by introducing redundancy into the transmitted message to achieve arbitrarily low bit errors
- The channel capacity C forms a limit for the coding rate R:

$$R = \frac{K}{N} \leq C$$

- $R = K/N$ is the ratio between the number of true (information) bits (K) and the number of bits after redundancy has been introduced by channel coding (N)
- Example: binary symmetric channel, $P(1) = P(0)$ with $p_e = 0.1 \longrightarrow C \approx 0.5$ $\longrightarrow N \geq 2K$, i.e. at least 100% redundancy needs to be incorporated into the transmitted bit stream in order to obtain an error-free reception

Convolutional Coding (CC)

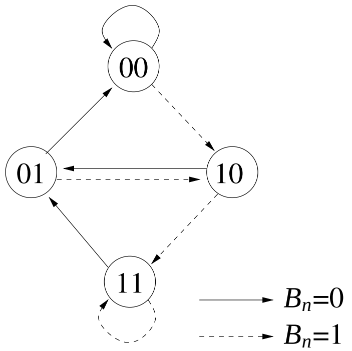
- A convolutional coder CC(N, K, L) takes a K-bit information symbol, which is shifted into an L-bit shift register (holding a memory of $L+1$ bits!) and forms by convolution over this memory N output bits. Hence coding rate $R = K/N$
- Example: block diagram of a CC(2,1,2) half-rate ($R = \frac{1}{2}$) systematic coder:

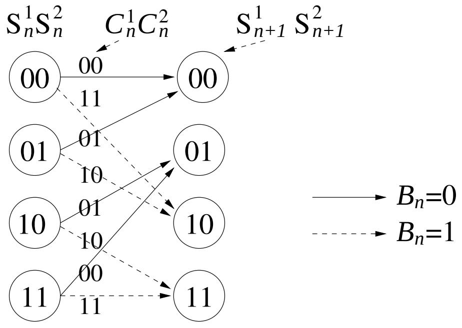


- Output bits are defined by N generator polynomials; here: $g_1 = [1\ 0\ 0]$, $g_2 = [1\ 1\ 1]$

CC — State and State Transition Diagrams

- With the previous CC(2,1,2) example, a state diagram (left) can be associated

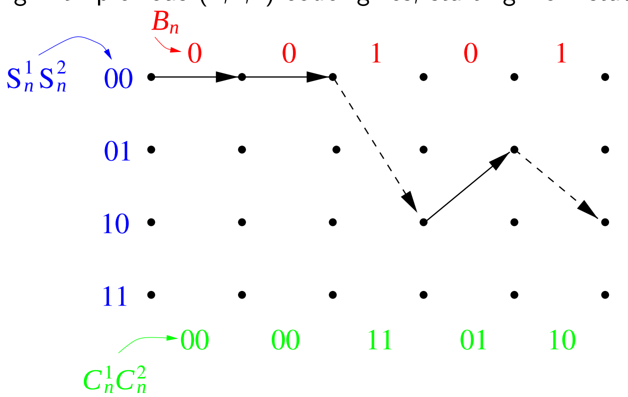




- The state transition diagram (right) indicates the time dimension, and the code output with each iteration of the CC

Trellis Path

- Consider the data sequence $B_n = [0\ \ 0\ \ 1\ \ 0\ \ 1]$
- Channel coding with previous (2,1,2) coder gives, starting from state $S_n^1 S_n^2 = 00$:



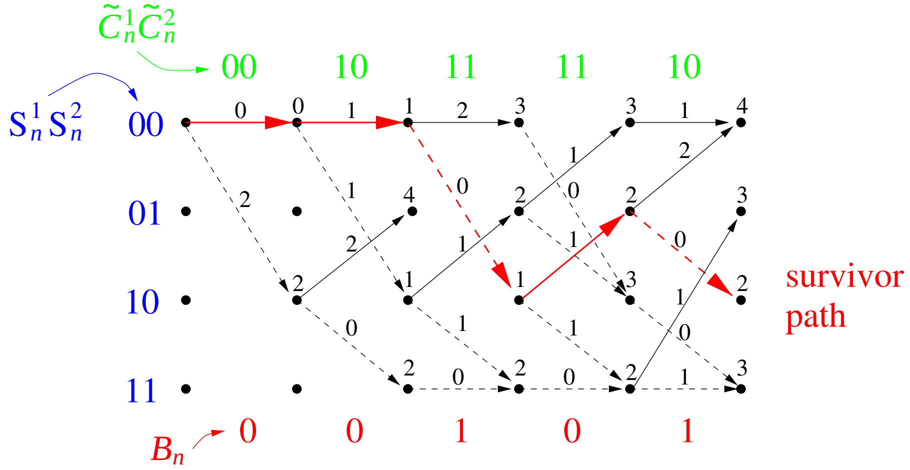
- This can be exploited for the channel decoding

Viterbi Decoding (I)

- In our example, the transmitted bit sequence is $C_n^1 C_n^2 = [00\ 00\ 11\ 01\ 10]$
- Assume, the received sequence is $\tilde{C}_n^1 \tilde{C}_n^2 = [00\ 10\ 11\ 11\ 10]$, i.e. two bit errors have occurred
- Viterbi decoding inspects all possible paths in the trellis diagram
- A metric to select the best (or "survivor") path is the minimum Hamming distance (number of different bits)
- Therefore, starting from state 00, for each possible path the Hamming distance is evaluated and accumulated
- If paths are joined, the ones with the larger Hamming distance are discarded; for equal distance, a random decision is made

Viterbi Decoding (II)

- Trellis graphs for received sequence $\tilde{C}_n^1 \tilde{C}_n^2 = [00\ 10\ 11\ 11\ 10]$:

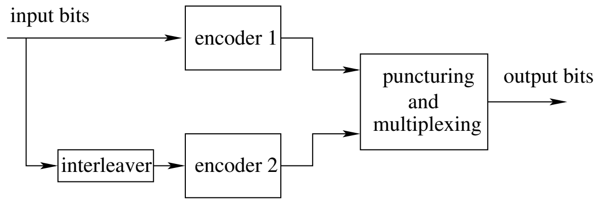


Viterbi Decoding (III)

- The final survivor path determines the "most likely" transmitted sequence; hence the Viterbi decoder is referred to as a *maximum likelihood* (ML) detector
- If the correct sequence is yielded, the cumulative Hamming distance of the survivor path gives the number of occurred bit errors
- If too many bit errors have occurred or bit errors are awkwardly distributed, Viterbi decoding will fail
- The above hard-decision decoding can be improved by soft-decision decoding, whereby "finer" levels of probability are assigned to the branches in the trellis diagram
- More generally, demodulator can output soft-decision, thus decoder has soft-input and can output soft-decision (SISO)

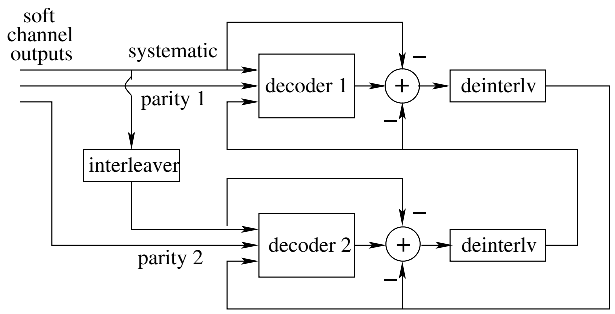
Turbo Coding

- Turbo encoder, assuming systematic $C(2, 1, 2)$



- Turbo decoder

  iterative decoding



Adaptive Equalisation

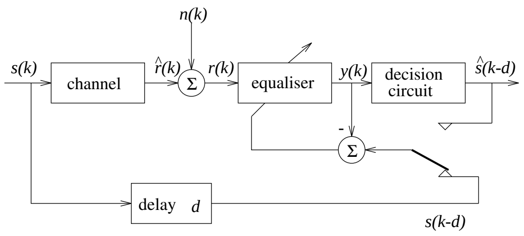
- Recall the framework of adaptive equalisation with two operation modes, training and decision-directed, where the channel model is:

$$r(k) = \sum_{i=0}^{n_c} c_i s(k-i) + n(k)$$

symbols are N-ary

$s(k) \in \{s_i = 2i-N-1,\ 1 \le i \le N\}$

and AWGN $n(k)$ has variance σ_n^2



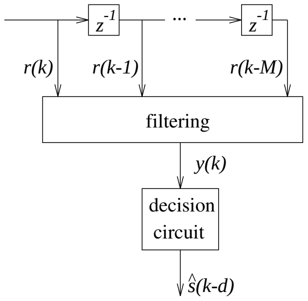
- We will first discuss symbol-decision equalisers and follow it by an introduction to the MLSE
- An equaliser decision delay d is necessary for coping with non-minimum phase channels
- The zero-forcing equaliser $H_E(z)$ inverses the channel $H_C(z)$: $H_C(z)H_E(z) \approx z^{-d}$
- Solving this gives the linear equaliser's weights. Although this zero-forcing equaliser completely eliminates ISI, it suffers from a serious noise enhancement problem
- The most popular designs are the linear equaliser and decision feedback equaliser based on the mean square error criterion

Linear Transversal Equaliser

- The linear equaliser is given by:

$$y(k) = \sum_{i=0}^{M} w_i r(k-i) = \mathbf{w}^T \mathbf{r}(k)$$

where $\mathbf{r}(k) = [r(k) \cdots r(k-M)]^T$ and M is the equaliser order



- Typical design is based on mean square error with the MMSE solution: $\hat{\mathbf{w}} = \mathbf{R}^{-1}\mathbf{p}$, where $\mathbf{R} = \mathrm{E}[\mathbf{r}(k)\mathbf{r}^T(k)]$ and $\mathbf{p} = \mathrm{E}[\mathbf{r}(k)s(k-d)]$

- Adaptive implementation typically adopts the LMS:

$$\tilde{\mathbf{w}}(k+1) = \tilde{\mathbf{w}}(k) + \mu \mathbf{r}(k) e(k) \quad \text{with} \quad e(k) = \begin{cases} y(k) - s(k-d), & \text{training} \\ y(k) - \hat{s}(k-d), & \text{decision-directed} \end{cases}$$

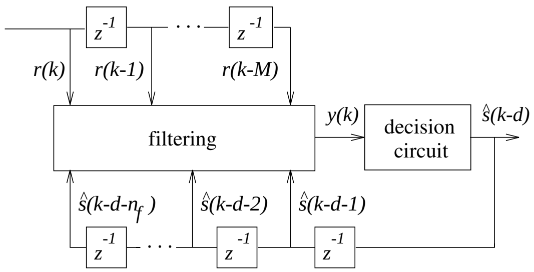
Decision Feedback Equaliser

- The DFE consists of a feedforward filter and a feedback filter:

$$y(k) = \mathbf{w}^T \mathbf{r}(k) + \mathbf{b}^T \hat{\mathbf{s}}(k-d)$$

$$= \sum_{i=0}^{M} w_i r(k-i) + \sum_{i=1}^{n_f} b_i \hat{s}(k-d-i)$$

The DFE generally outperforms the LTE in terms MSE and BER



- Assuming equaliser decisions $\hat{\mathbf{s}}(k-d)$ are correct, the feedback filter $\mathbf{b}^T\hat{\mathbf{s}}(k-d)$ eliminates a large proportion of ISI without enhancing noise and the feedforward filter $\mathbf{w}^T\mathbf{r}(k)$ takes care the remaining ISI
- **Error propagation**. Occasionally error occurs in symbol detection, i.e. $\hat{s}(k-d) \neq s(k-d)$, it is fed back and will affect subsequent symbol detections $\rightarrow$ further burst errors
- **Choice of structure parameters**. A simple practical rule: feedforward filter covers entire channel dispersion, i.e. $M = n_c$; decision delay is set to $d = n_c$; and feedback filter order $n_f = n_c$

Blind Equalisation

- In blind equalisation, there is no training, an equaliser has to estimate the transmitted symbols and/or channel based only on the received samples $r(k)$

- There are three classes of blind equalisation algorithms
  - **Joint data and channel estimation**: e.g. using blind or super trellis search techniques. This produces the best results but can be computationally prohibitive
  - **Higher-order statistics based methods**: to identify the channel using $r(k)$ only, 2nd order statistic is insufficient as it is phase blind. Higher-order statistics based methods can overcome this problem. This approach produces very good results but computational cost can be very expensive
  - **Bussgang-type adaptive FIR filters**: optimize some non-MSE type cost functions using stochastic gradient, computationally very simple

- We will discuss the 3rd class. Since there is no desired response $s(k-d)$ for the adaptive filter, one has to "invent" some substitute $\rightarrow$ the resulting non-MSE cost functions generally have local minima, and this often causes problems

Constant Modulus Algorithm

- We will consider the general QAM case and use complex notations, e.g. the channel taps $c_i = c_{R,i} + jc_{I,i}$, the received signals $r(k) = r_R(k) + jr_I(k)$, the symbols $s(k) = s_R(k) + js_I(k)$, and the equaliser weights $w_i = w_{R,i} + jw_{I,i}$
- Define the constant $\Delta_2 = E[|s(k)|^4]/E[|s(k)|^2]$, and consider adaptive filter or blind equaliser:
$$y(k) = \mathbf{w}^T\mathbf{r}(k)$$
with $\mathbf{w} = [w_0 \cdots w_M]^T$ and $\mathbf{r}(k) = [r(k) \cdots r(k-M)]^T$
- Although QAM symbols do not fall on the constant modulus circle of radius $\sqrt{\Delta_2}$, by penalizing equaliser output $y(k)$ which deviates from this circle, the correct symbol constellation can be restored
- This leads to the CMA, which is the most popular blind equaliser for high-order QAM signalling, as it has simple computational requirements similar to those of the LMS

CMA (continue)

- The CMA can be viewed to adjust $\mathbf{w}$ by minimizing the non-convex cost function
$$\bar{J}_{\text{CMA}}(\mathbf{w}) = E[(|y(k)|^2 - \Delta_2)^2]$$
using a stochastic gradient method, i.e. actually through minimizing $(|y(k)|^2 - \Delta_2)^2$
- At sample k, given $y(k) = \mathbf{w}^T(k)\mathbf{r}(k)$, the equaliser weights are updated using:
$$\left.\begin{aligned} \epsilon(k) &= y(k)(\Delta_2 - |y(k)|^2) \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu\epsilon(k)\mathbf{r}^*(k) \end{aligned}\right\}$$
where μ is a very small positive adaptive gain and $\mathbf{r}^*(k)$ is the conjugate of $\mathbf{r}(k)$
- Compare this with the LMS, where $\epsilon(k) = s(k-d) - y(k)$
- There are many solutions $\mathbf{w}_s$ that minimize the cost function $\bar{J}_{\text{CMA}}(\mathbf{w})$. One of them, $\mathbf{w}_{\text{opt}}$, restores the correct signal constellation and is corresponding to the MMSE solution
- The weight vectors that minimize $\bar{J}_{\text{CMA}}(\mathbf{w})$ are thus
$$\mathbf{w}_s = \exp(j\phi)\mathbf{w}_{\text{opt}}, \; 0 \le \phi < 2\pi$$
- This undesired phase shift cannot be resolved by the CMA (all blind equalisers suffer more or less a similar problem), and must be eliminated by other means, e.g. using differential encoding

Complex Variable Derivative

- Complex-valued variable derivative is defined as
$$\frac{\partial J(\mathbf{w})}{\partial w_i} = \frac{1}{2}\left(\frac{\partial J}{\partial w_{R,i}} + j\frac{\partial J}{\partial w_{I,i}}\right)$$
- Note that $y(k) = w_0 r(k) + \cdots + w_M r(k-M)$ and
$$J(\mathbf{w}) = \frac{1}{2}(|y(k)|^2 - \Delta_2)^2 = \frac{1}{2}(y(k)y^*(k) - \Delta_2)^2$$
- Hence we have
$$\frac{\partial J}{\partial w_i} = \frac{1}{2}\cdot 2(y(k)y^*(k) - \Delta_2)\frac{\partial y(k)y^*(k)}{\partial w_i} = (|y(k)|^2 - \Delta_2)\left(\frac{\partial y(k)}{\partial w_i}y^*(k) + y(k)\frac{\partial y^*(k)}{\partial w_i}\right)$$
- Note
$$\frac{\partial y(k)}{\partial w_i} = \frac{1}{2}\left(\frac{\partial y(k)}{\partial w_{R,i}} + j\frac{\partial y(k)}{\partial w_{I,i}}\right)$$
$$\frac{\partial y(k)}{\partial w_{R,i}} = r_R(k-i) + jr_I(k-i) \quad \text{and} \quad \frac{\partial y(k)}{\partial w_{I,i}} = -r_I(k-i) + jr_R(k-i)$$
- This leads to
$$\frac{\partial y(k)}{\partial w_i} = 0$$

Complex Variable Derivative (continue)

- Note
$$\frac{\partial y^*(k)}{\partial w_i} = \frac{1}{2}\left(\frac{\partial y^*(k)}{\partial w_{R,i}} + j\frac{\partial y^*(k)}{\partial w_{I,i}}\right)$$
$$\frac{\partial y^*(k)}{\partial w_{R,i}} = r_R(k-i) - jr_I(k-i) \quad \text{and} \quad \frac{\partial y^*(k)}{\partial w_{I,i}} = -r_I(k-i) - jr_R(k-i)$$
- This leads to
$$\frac{\partial y^*(k)}{\partial w_i} = r^*(k-i)$$
- Therefore, the gradient
$$\nabla J(\mathbf{w}) = \left[\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}\right]^T = y(k)(|y(k)|^2 - \Delta_2)\mathbf{r}^*(k)$$
- Using
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\left(-\nabla J(\mathbf{w}(k))\right)$$
- leads to
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu y(k)(\Delta_2 - |y(k)|^2)\mathbf{r}^*(k)$$