UCL DEPARTMENT OF ELECTRONIC AND ELECRICAL ENGINEERING
INTEGRATED GRADUATE DEVELOPMENT PROGRAMME
MSC TELECOMMUNICATIONS

# *Sampling & Quantization*

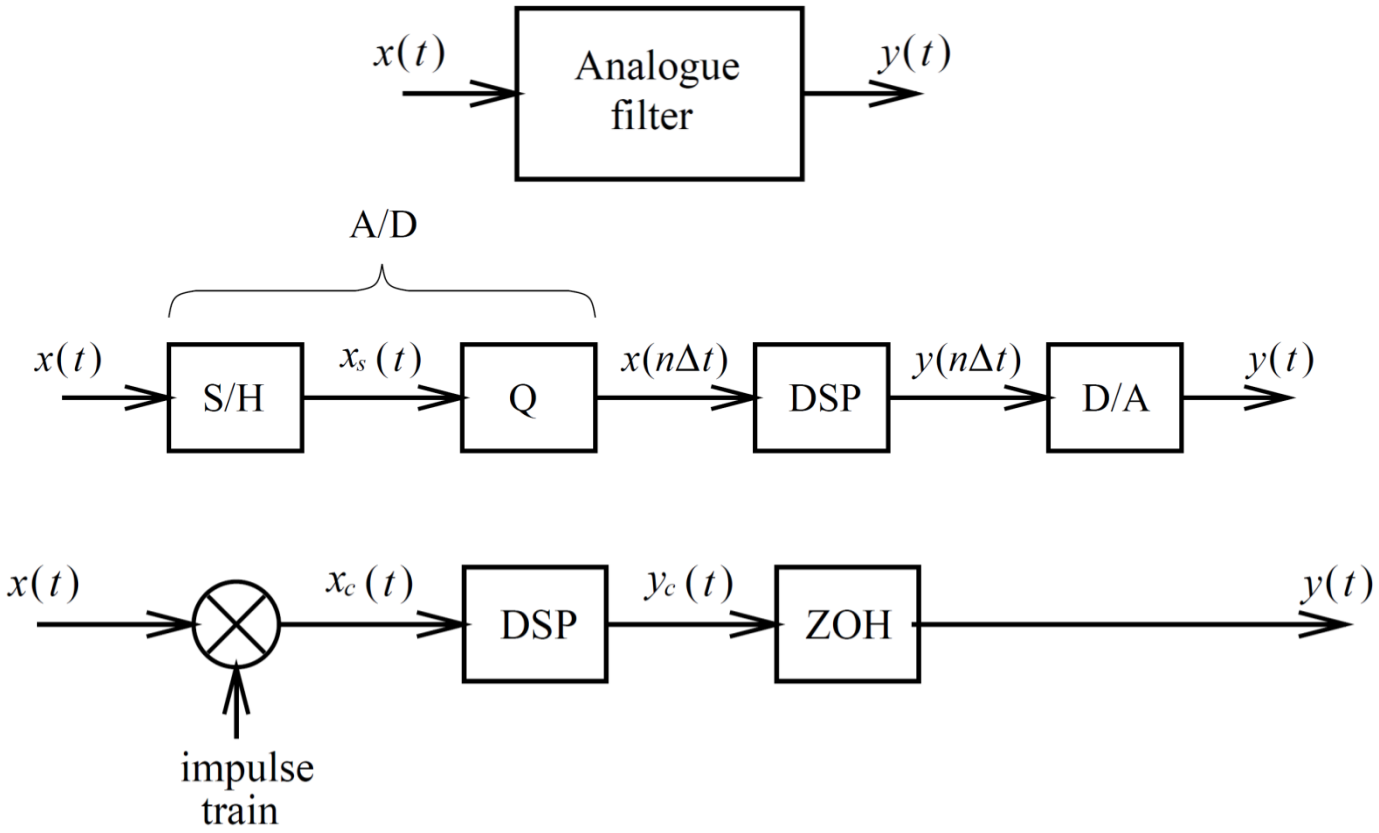
**Prof Yiannis Andreopoulos**
**University College London**

From analogue to digital systems
Sampling and the Sampling Theorem
Quantization: Unipolar/Bipolar/SAQ, error analysis

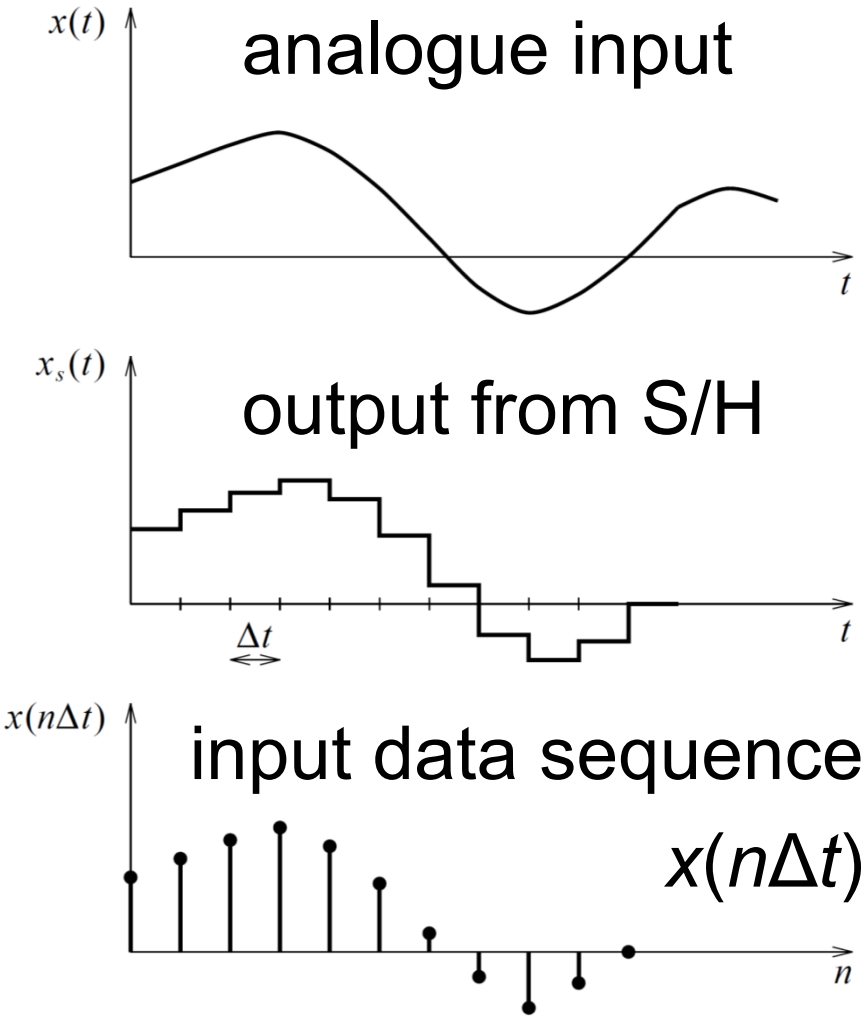
Acknowledgement: Some material adapted from: G. Halikias (City U.) and S. J. Orfanidis “Introduction to Signal Processing”, Prentice Hall

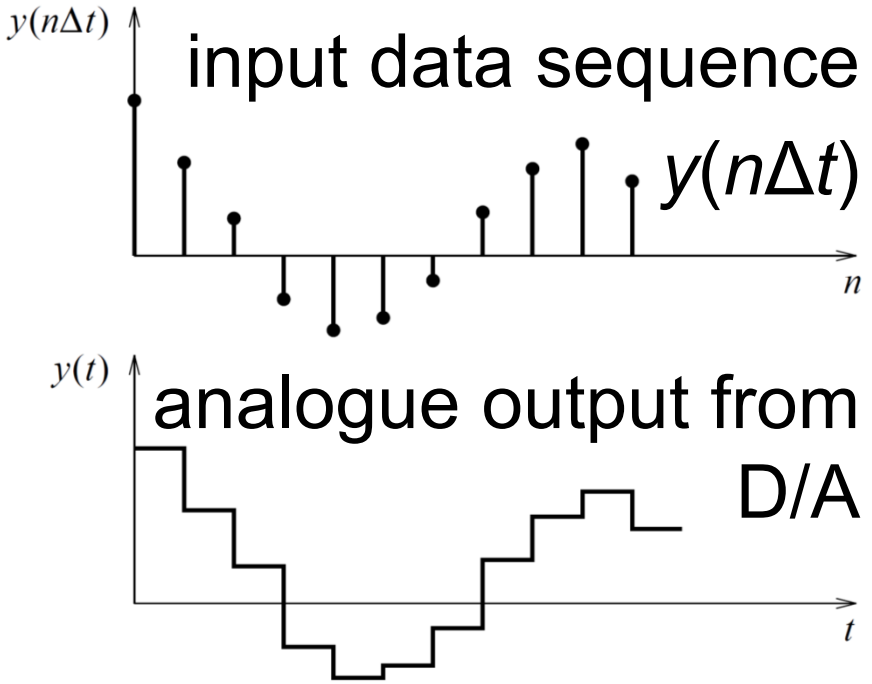
# Introduction

- Two major problems with analogue filters:
  *(i)* drift; *(ii)* limited functionality (and programmability)



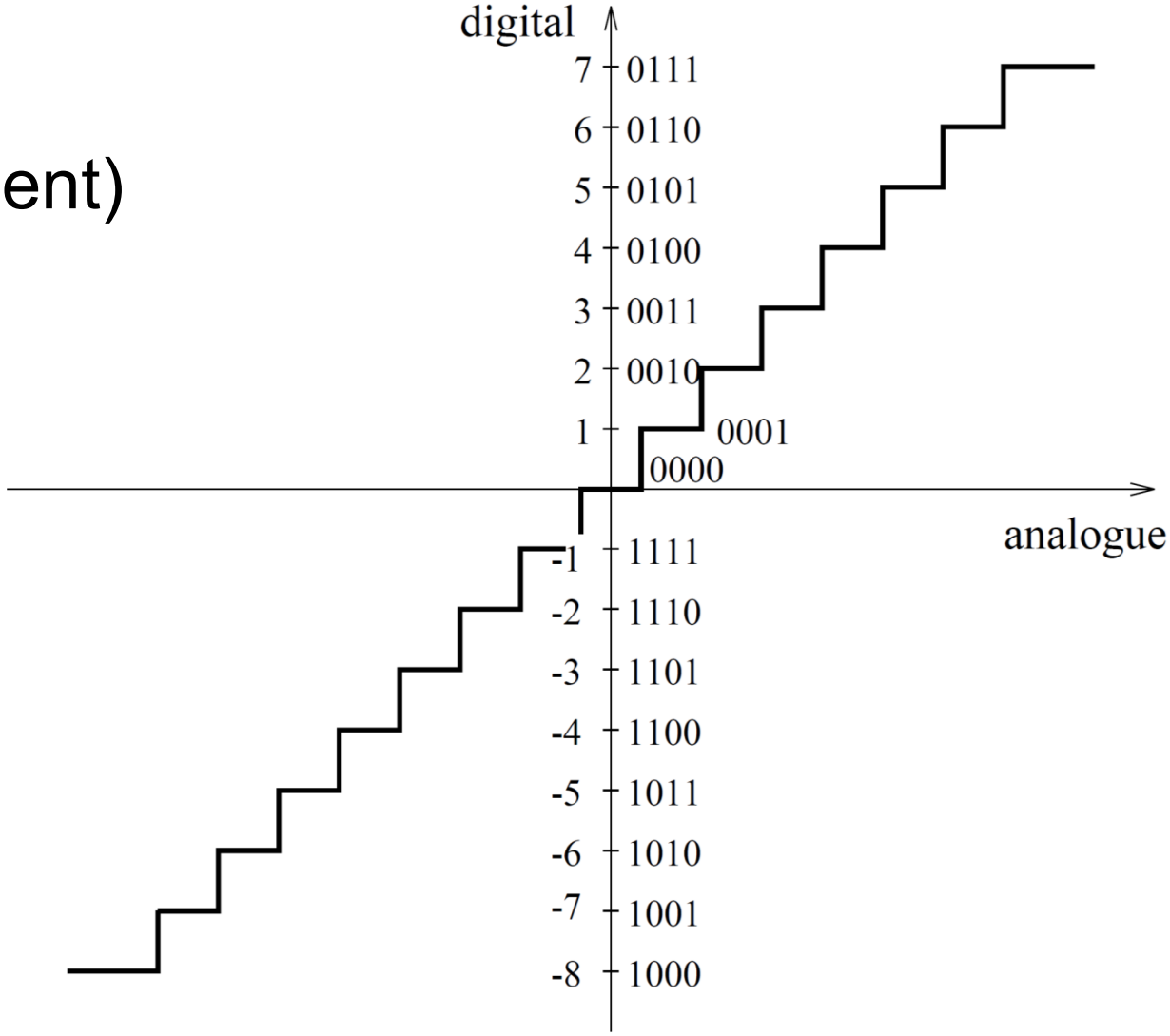
# Introduction (cont'd)



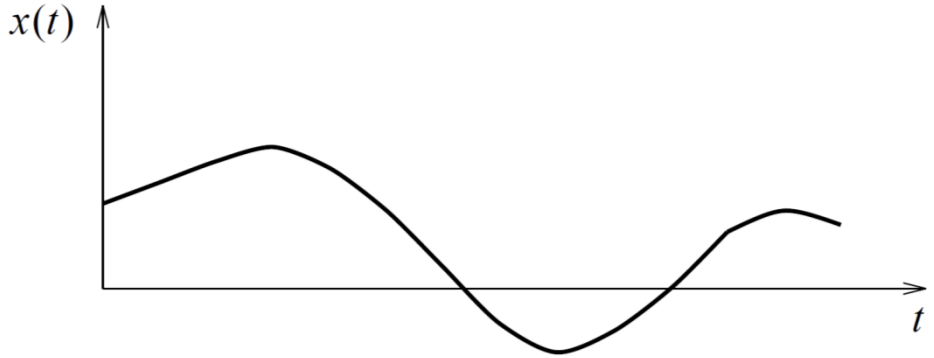



# Introduction (Quantisation)

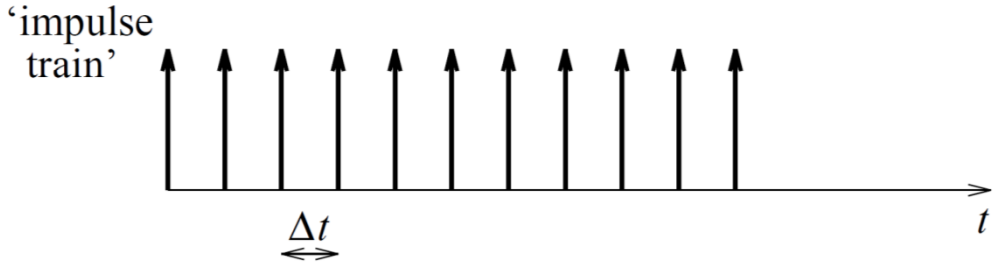
4-bit quantiser
(two's complement)
Details on this
come later…

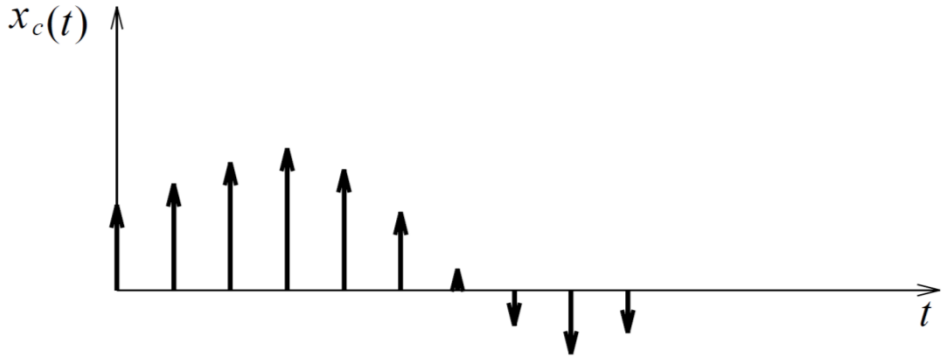


# Sampling




Periodic impulse train $\delta_T(t)$



# Sampling

Impulse train

$$\delta_T(t) = \cdots + \delta(t+2\Delta t) + \delta(t+\Delta t) + \delta(t) + \delta(t-\Delta t) + \delta(t-2\Delta t) + \cdots$$

$$= \sum_{n=-\infty}^{\infty} \delta(t-n\Delta t)$$

Time domain description

$$x_c(t) = x(t)\ \delta_T(t)$$

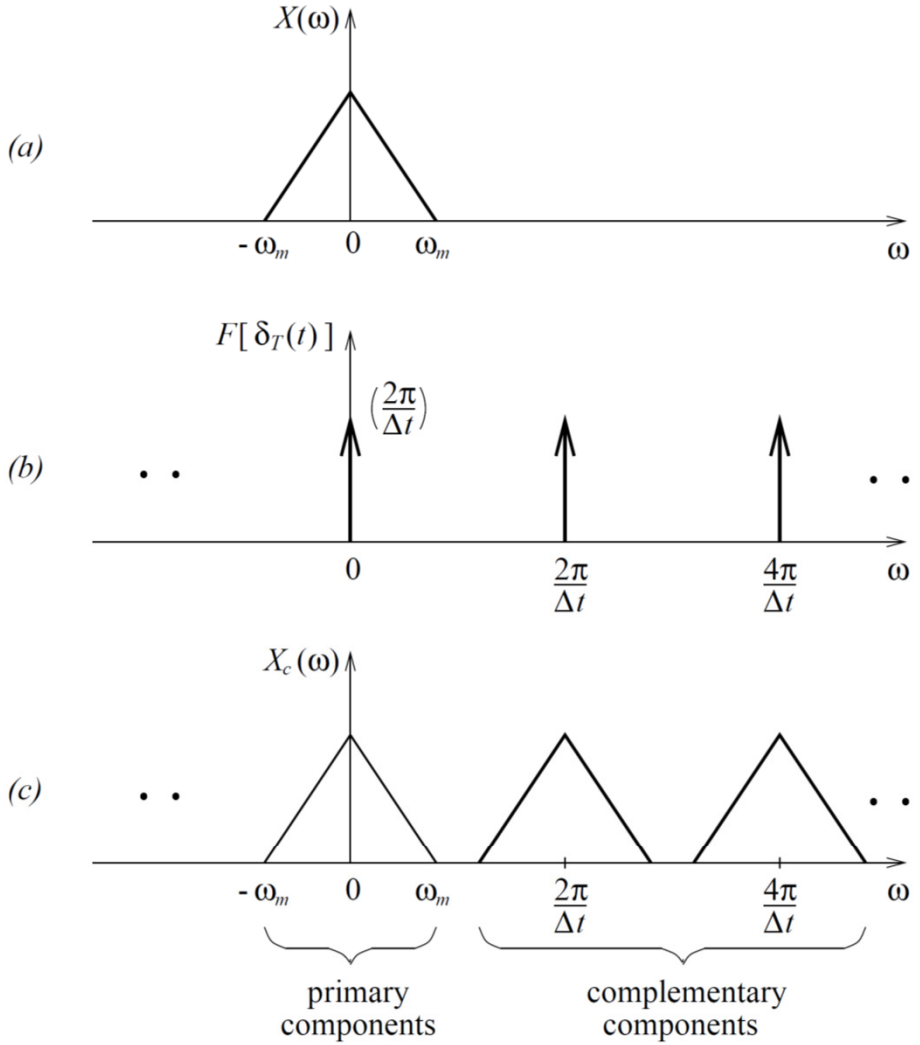
$$= x(t)\ [\ \delta(t) + \delta(t-\Delta t) + \delta(t-2\Delta t) + \cdots + \ ]$$

$$= x(t)\ \delta(t) \ + \ x(t)\ \delta(t-\Delta t) + x(t)\ \delta(t-2\Delta t) \cdots$$

$$= \ x(0)\ \delta(t) \ + \ x(\Delta t)\ \delta(t-\Delta t) + x(2\Delta t)\ \delta(t-2\Delta t) \cdots$$

$$= \sum_{n} \ x(n\Delta t)\ \delta(t-n\Delta t)$$

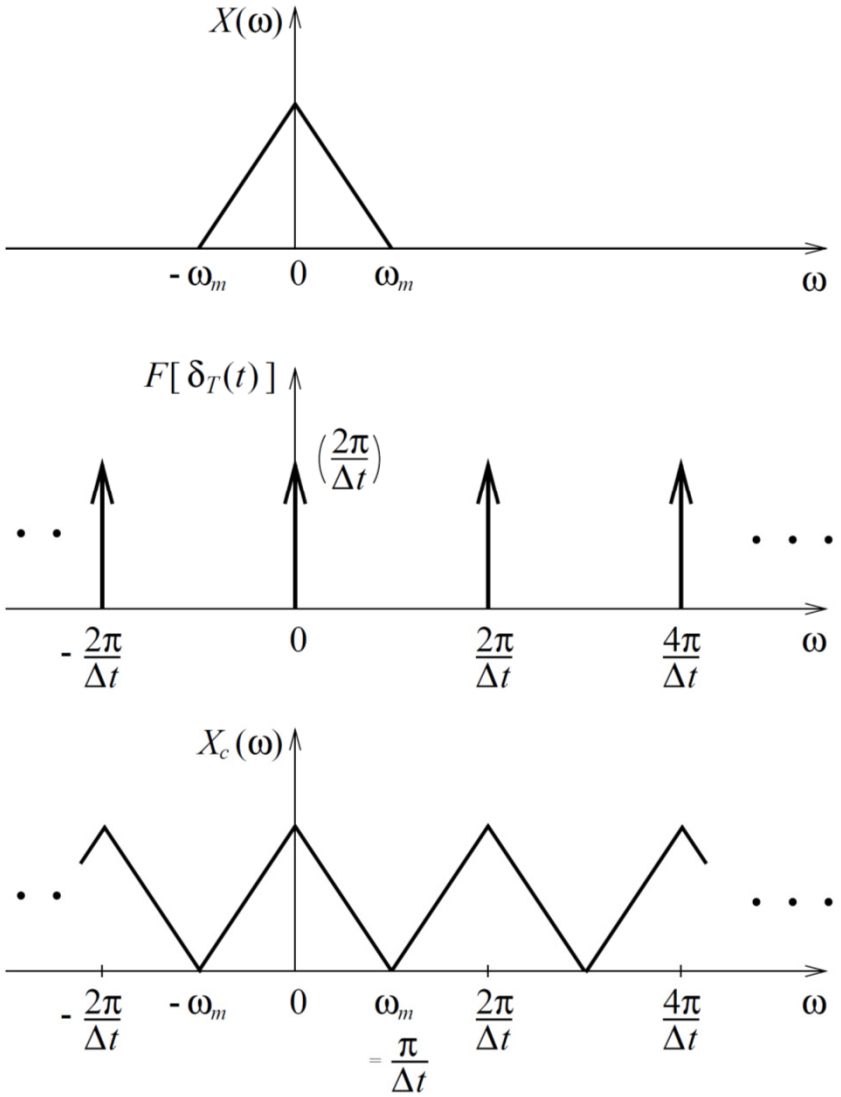
# Sampling




*Frequency content of a signal sampled at an adequate rate; (a) Fourier transform of analogue input; (b) Fourier transform of impulse train; (c) Fourier transform of sampled signal.*

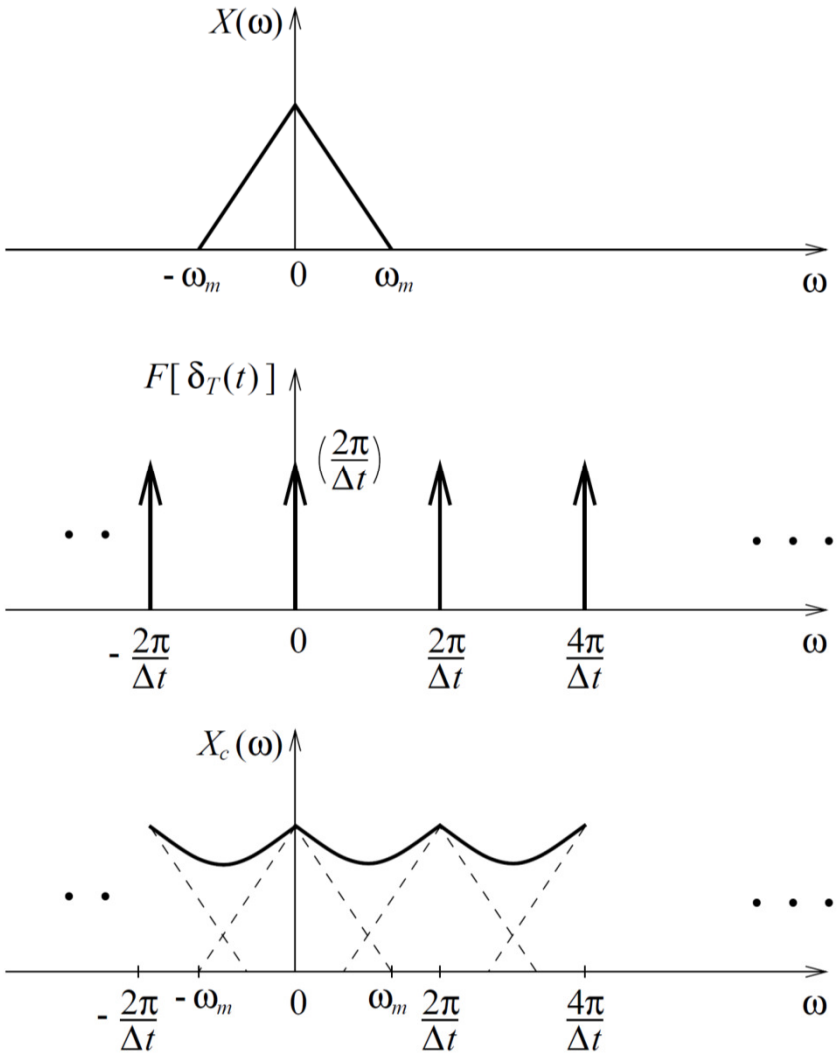
# Sampling

**Barely adequate sampling rate**



# Sampling

**Inadequate sampling rate**



# Sampling

**Sampling theorem – the Nyquist criterion.**

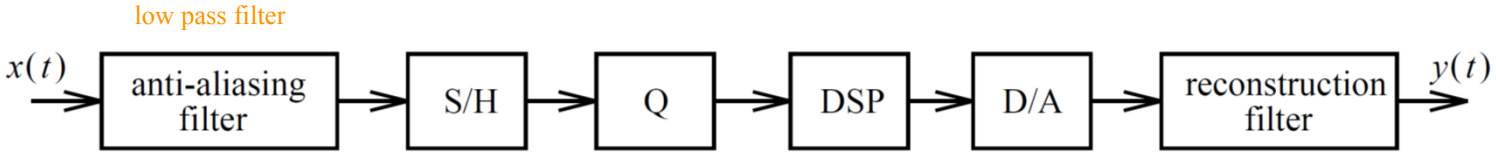
In order to prevent aliasing distortion, sampling must occur at a frequency greater than twice the highest frequency present. The sampling theorem can be stated more formally as:

- to prevent aliasing distortion, a bandlimited signal of bandwidth, $\omega_B$ rad/s, must be sampled at a rate of at least $2\omega_B$ rad/s.

where 'a bandlimited signal of bandwidth, $\omega_B$,' is a signal whose Fourier transform is zero at all values of $\omega$ except a region of width $\omega_B$ for positive values of $\omega$ (and a corresponding region for negative values of $\omega$).

# Sampling

**Practical sampled data systems**




*Block diagram of practical sampled data system.*

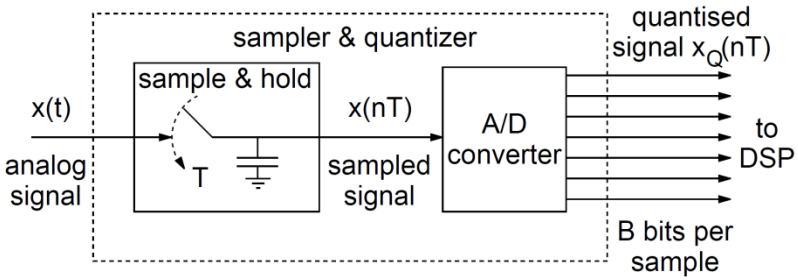
- Perfect reconstruction is not possible for two reasons:
  - it is impossible to construct an ideal LPF from a finite number of components;
  - if an analogue output is required a D/A converter/ZOH is always present.

- When designing a sampled data system it is not always possible to ensure that the Nyquist criterion is met.
  - an analogue LPF prior to the S/H whose function is to remove or attenuate all frequencies above half the sampling frequency.

# Quantization

**Signal quantisation:** Signal sampling and quantisation are necessary prerequisites for any digital processing operation on analog signals.

- After the signal is sampled, the capacitor in the sampler holds each sample $x(nT)$ for at most $T$ seconds. During this period the A/D converter converts $x(nT)$ to a quantised sample $x_Q(nT)$, represented by a finite number of bits, $B$ say.




- After digital processing, the resulting $B$-bit word is applied to the D/A converter, which converts it back to analog (typically staircase) form.

- The A/D converter is characterised by a full-scale *range* $R$. For a unipolar converter

$$0 \leq x_Q < R$$

while for a bipolar converter,

$$-\frac{R}{2} \leq x_Q < \frac{R}{2}$$

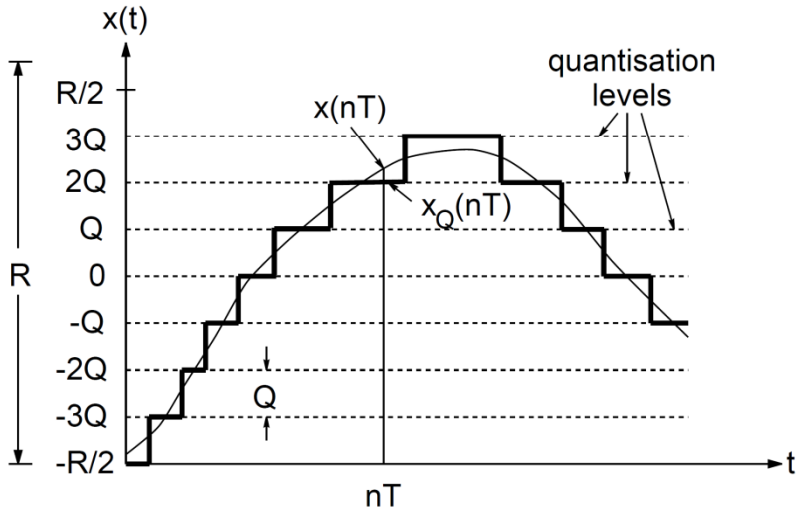
Typical value is $R = 10$ Volts.

- The quantised sample $x_Q(nT)$, represented by $B$ bits, can take only one of $2^B$ possible values.

- The range $R$ is divided equally between $2^B$ *quantisation levels*. The space between levels is the *quantisation width*:

$$Q = \frac{R}{2^B}$$

The full-scale range of the quantiser is 0 to $R - Q$ Volts (unipolar) or $-\frac{R}{2}$ to $\frac{R}{2} - Q$ Volts (bipolar).




- Typically quantisation is performed by *rounding*, i.e. replacing each value of $x(nT)$ by its nearest level. Alternatively, quantisation can be done by *truncation*, i.e. replacing each value by the level below it.

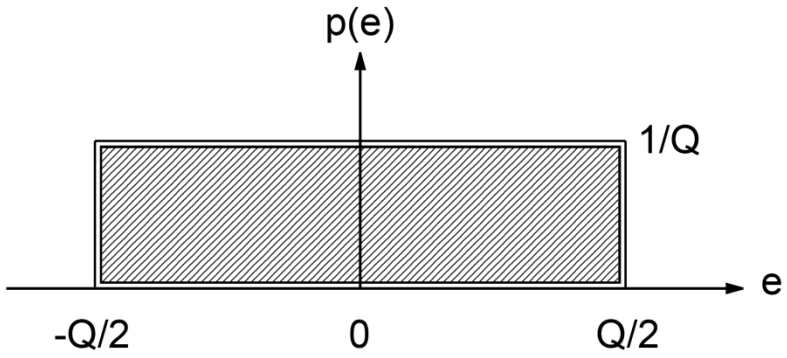
- The *quantisation error* is:

$$e(nT) = x_Q(nT) - x(nT)$$

i.e. difference between sampled and quantised value. For rounding a sample which lies in the range $[-R/2\ R/2)$ we have

$$-\frac{Q}{2} \leq e(nT) \leq \frac{Q}{2}$$

Thus maximum error is $e_{\max} = Q/2$. Assuming a uniform probability distribution of $e$ in the range $[-Q/2\ Q/2]$ we have




$$p(e) = \frac{1}{Q} \quad \text{if} \quad -\frac{Q}{2} \leq e \leq \frac{Q}{2}$$
$$= 0 \quad \text{otherwise}$$

Thus the expected-value (statistical average) of $e$ and its variance are

$$E[e] = \int_{-Q/2}^{Q/2} ep(e)\, de = 0$$
$$E[e^2] = \int_{-Q/2}^{Q/2} e^2 p(e) de = \frac{Q^2}{12}$$

respectively. Thus the standard deviation ("r.m.s. value") of $e$ is $e_{\text{rms}} = Q/\sqrt{12}$.

- Thinking of $R$ and $Q$ as the ranges of the signal and quantisation noise, the SNR (signal to noise ratio) of the qauntisation process can be written as:

$$\text{SNR} = 20\log_{10}\left(\frac{R}{Q}\right) = 20\log_{10}(2^B)$$
$$= B \cdot 20\ \log_{10} 2 = 6B\ \text{dB}.$$

which defines the *dynamic range* of the quantiser.

- **Example:** In a digital audio application the signal is sampled at 44 kHz and each sample quantised using an A/D converter having a full-range scale of 10 Volts. Determine the

number of bits $B$ if the rms quantisation error must be kept below $50$ microvolts. Then determine the actual rms error and the bit rate in bits/sec. What is the dynamic range of the quantiser?

Here

$$e_{\rm rms} = \frac{Q}{\sqrt{12}} = \frac{R}{2^B\sqrt{12}}$$

Hence

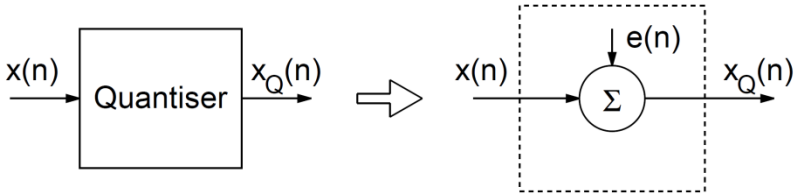
$$B = \log_2\left[\frac{R}{e_{\rm rms}\sqrt{12}}\right] = \log_2\left[\frac{10}{50\cdot 10^{-6}\sqrt{12}}\right]$$
$$= 15.82$$

Hence we need $B = 16$ bits, corresponding to $2^{16} = 65536$ quantisation levels, and thus $e_{\rm rms} = R/2^B\sqrt{12} = 44$ microvolts. The bit rate is $B \cdot f_s = 16 \cdot 44 = 704$ kbits/sec. The dynamic range of the quantiser is $6B = 6 \cdot 16 = 96$ dB. (Compare with the dynamic range of the human ear which is $\approx$ 100 dB).

- A simple additive-noise model of the error quantisation process is given by

$$x_Q(n) = x(n) + e(n)$$

where $e(n)$ is a zero-mean white-noise process with variance $\sigma_e^2 = Q^2/12$.




The assumption that $e(n)$ is white, means

that its auto-correlation function is

$$R_{ee}(k) = E[e(n+k) \cdot e(n)] = \sigma_e^2 \text{ for } k = 0$$
$$= 0 \text{ for } k \neq 0$$

It is further assumed that the signal and noise sequences are uncorrelated, i.e. the cross-correlation function

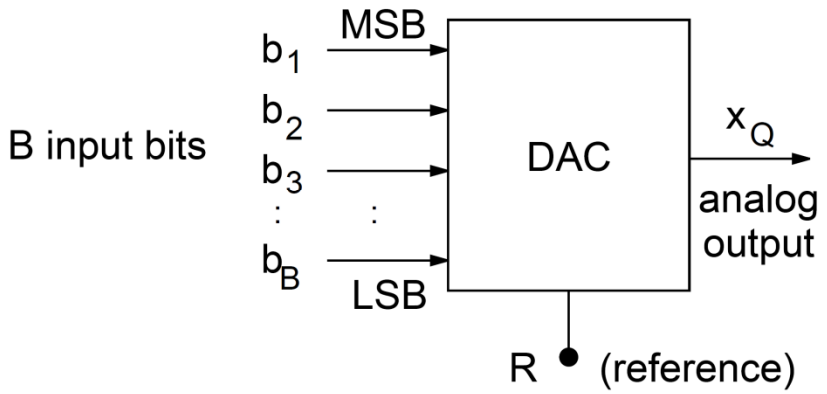
$$R_{ex}(k) = E[e(n+k)x(n)] = 0 \text{ for all } k$$

This model may be used to analyse the effects of quantisation on various filter structures.

**D/A converters:** We consider $B$-bit DAC with full-scale range $R$. Given $B$ input bits of 0's and 1's, $\mathbf{b} = [b_1, b_2, \ldots, b_B]$, the converter outputs an analog value $x_Q$ at one of the $2^B$ quantisation levels within range $R$:

LSB

- If converter is unipolar, $x_Q \in [0, R)$.
- If converter is bipolar, $x_Q \in [-R/2,\ R/2)$.




Analog signal $x_Q$ depends on the type of converter and coding convention used. The three widely available converter types are:

- *unipolar natural binary*,
- *bipolar offset binary*, and
- *bipolar two's complement*.

For a *unipolar natural binary converter*,

$$x_Q = R(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_B 2^{-B})$$

Note that for this type of converter:

- Minimum level is $0$, reached when all bits are $0$.
- Smallest non-zero level is $2^{-B}R = Q$, when $\mathbf{B} = [0, 0, \ldots, 1]$.
- Most significant bit pattern $[1, 0, 0, \ldots, 0]$ corresponds to $x_Q = R/2$.
- Maximum level is reached when all bits are $1$, i.e. $b = [1, 1, \ldots, 1]$. This is given by

$$\begin{aligned} x_Q &= R(2^{-1} + 2^{-2} + \ldots + 2^{-B}) \\ &= R2^{-1}(1 + 2^{-1} + 2^{-2} + \ldots + 2^{-(B-1)}) \\ &= R2^{-1}\left(\frac{1 - 2^{-B}}{1 - 2^{-1}}\right) \\ &= R(1 - 2^{-B}) \end{aligned}$$

In terms of the quantisation width,

$$x_Q = R2^{-B}(\underbrace{b_1 2^{B-1} + b_2 2^{B-2} + \ldots + b_B}_{\text{m}}) = Qm$$

where $m$ is integer with binary representation $(b_1, b_2, \ldots, b_B)$.

The *bipolar offset binary* converter is obtained by shifting $x_Q$ down by half-scale:

$$x_Q = R(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_B 2^{-B} - 0.5)$$

- Minimum and maximum values are obtained by shifting corresponding natural binary values down by $R/2$, i.e.

$$x_{\mathrm{Qmin}} = 0 - \frac{R}{2} = -\frac{R}{2}$$
$$x_{\mathrm{Qmax}} = (R - Q) - \frac{R}{2} = \frac{R}{2} - Q$$

- Analog value can also be expressed in terms of $Q$ as

$$x_Q = Qm'$$

where

$$m' = m - \frac{1}{2}2^B = m - 2^{B-1}$$

- An unnatural property of the offset binary code is that the level $x_Q = 0$ is represented by the non-zero bit pattern $\mathbf{b} = [1, 0, 0, \ldots, 0]$.

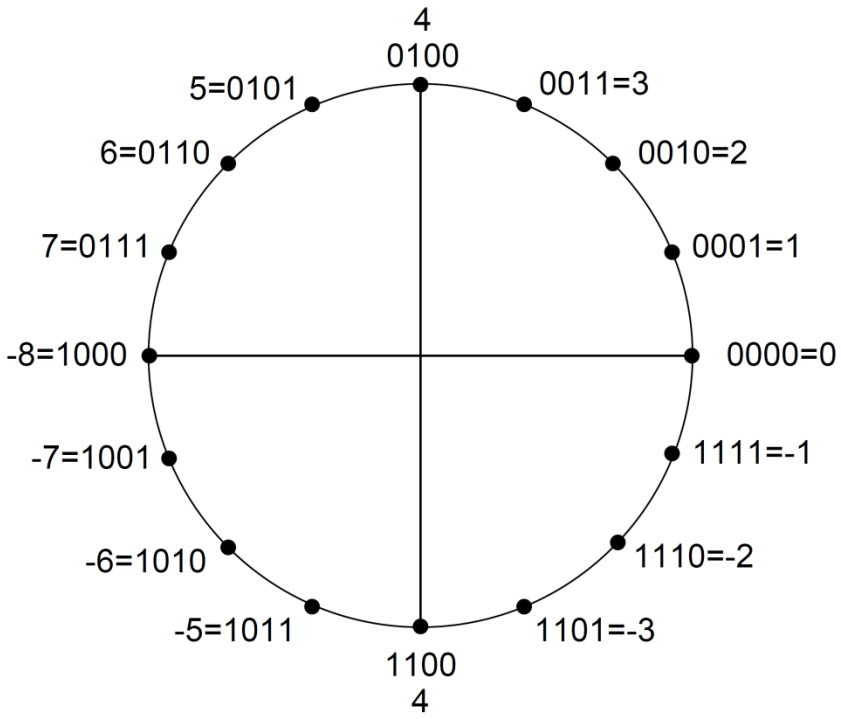
This is corrected by the *two's complement code* (next).

The *bipolar two's complement converter* code is obtained from the offset binary converter by complementing the most significant bit, i.e.

$$x_Q = R(\bar{b}_1 2^{-1} + b_2 2^{-2} + \ldots + b_B 2^{-B} - 0.5)$$

Thus the zero pattern $\mathbf{b} = [0, 0, \ldots, 0]$ now corresponds to $0$.

The two's complement code is best understood by wrapping the natural binary code around a circle. Positive integers appear in the upper part, negative integers in the lower part of the circle.

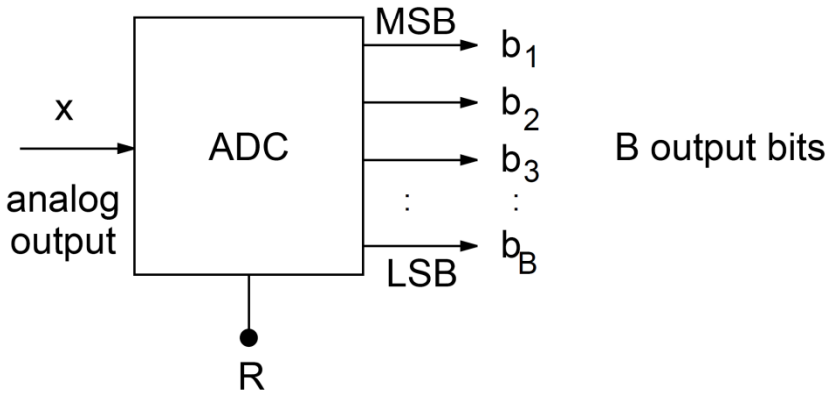



The negative of any integer $m$ in the upper semicircle is obtained by complementing all bits and adding 1. This corresponds to reflection with respect th the horizontal axis.

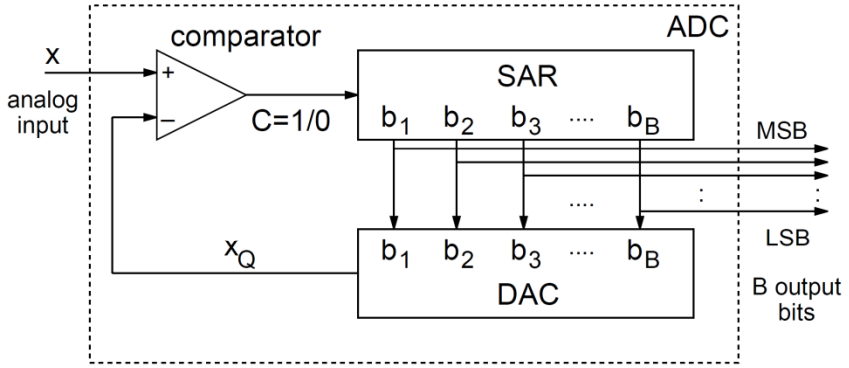
**Example:** The natural binary code of $m = 3$ is $0011$. Then

$$m_{2c} = \bar{m} + 1 = (\overline{0011}) + (0001)$$
$$= (1100) + (0001) = (1101) = -3$$

**A/D converters:** A/D converters quantize an analog value $X$ into $B$ bits $[b_1, b_2, \ldots, b_B]$. A popular type is the "*successive approximation*" converter, which essentially consists of D/A converter inside a feedback loop.



For natural binary or offset binary coding, the successive-approximation ADC works as follows:




- Initially all bits in the successive approximation register (SAR) are cleared to $0$.

- Starting with the MSB $b_1$, each bit is turned $on$ in sequence, and a test is performed to determine whether the bit should be kept $on$, or turned $off$.

- The control logic puts the correct value of each bit in the right slot in the SAR register. Then, leaving all tested bits set at their correct values, the next bit is turned $on$ in the SAR and the process continues.

- At the end of each test, the SAR bit vector $\mathbf{B}$ is applied to the input of the DAC which produces an analog quantised value $x_Q$. This is then compared to the analog input $x$ to the ADC. If $x \geq x_Q$ the bit under test is kept on, otherwise it is turned off.

- After $B$ tests the SAR contains the quantised bits of $x$. Here quantisation is done by $truncation$. To implement $rounding$, $x_Q$ should be compared with $x + \frac{Q}{2}$ for each test.

Note that the successive approximation

essentially implements a *binary search* algorithm within range $R$.

truncation

**Example:** Convert the analog values $x = 3.5$ Volts to its offset binary representation using truncation, assuming $B = 4$ bits and $R = 10$ volts.

4*0.625=2.5

b1 = 1 represents 0

b1 is not used

| test | $b_1b_2b_3b_4$ | $x_Q$ | Pass/Fail |
|---|---|---|---|
| $b_1$ | 1000 | 0.000 | Pass |
| $b_2$ | 1100 | 2.500 | Pass |
| $b_3$ | 1110 | 3.750 | Fail |
| $b_4$ | 1101 | 3.125 | Pass |
| | 1101 | 3.125 | |

Note that although $3.5$ is closer to $3.750$ than $3.125$, the result is still truncated down to $3.125$ Volts.

rounding

**Example:** Convert the analog values $x = 3.5$ Volts to its offset binary representation using rounding, assuming $B = 4$ bits and $R = 10$ volts.

The quantisation width is $Q = 10/2^4 = 0.625$ Volts. We first shift $x = 3.5$ to $x + Q/2 = 3.8125$.

compare this value

| test | $b_1b_2b_3b_4$ | $x_Q$ | Pass/Fail |
|---|---|---|---|
| $b_1$ | 1000 | 0.000 | Pass |
| $b_2$ | 1100 | 2.500 | Pass |
| $b_3$ | 1110 | 3.750 | Pass |
| $b_4$ | 1101 | 4.375 | Fail |
| | 1110 | 3.750 | |

b1 is not used

The algorithm for two's complement coding is slightly different. Because the MSB is complemented it must treated separately from

other bits. Note that the MSB bit determines the sign ($1$ means negative, $0$ positive). Apart from the MSB, two's complement and offset binary codes are identical. Thus, we need to complement the MSB and proceed as before with the remaining digits.

# *Summary*

- Sampling and AD and DA conversion
- Quantization: Unipolar/Bipolar/SAQ, error analysis, examples