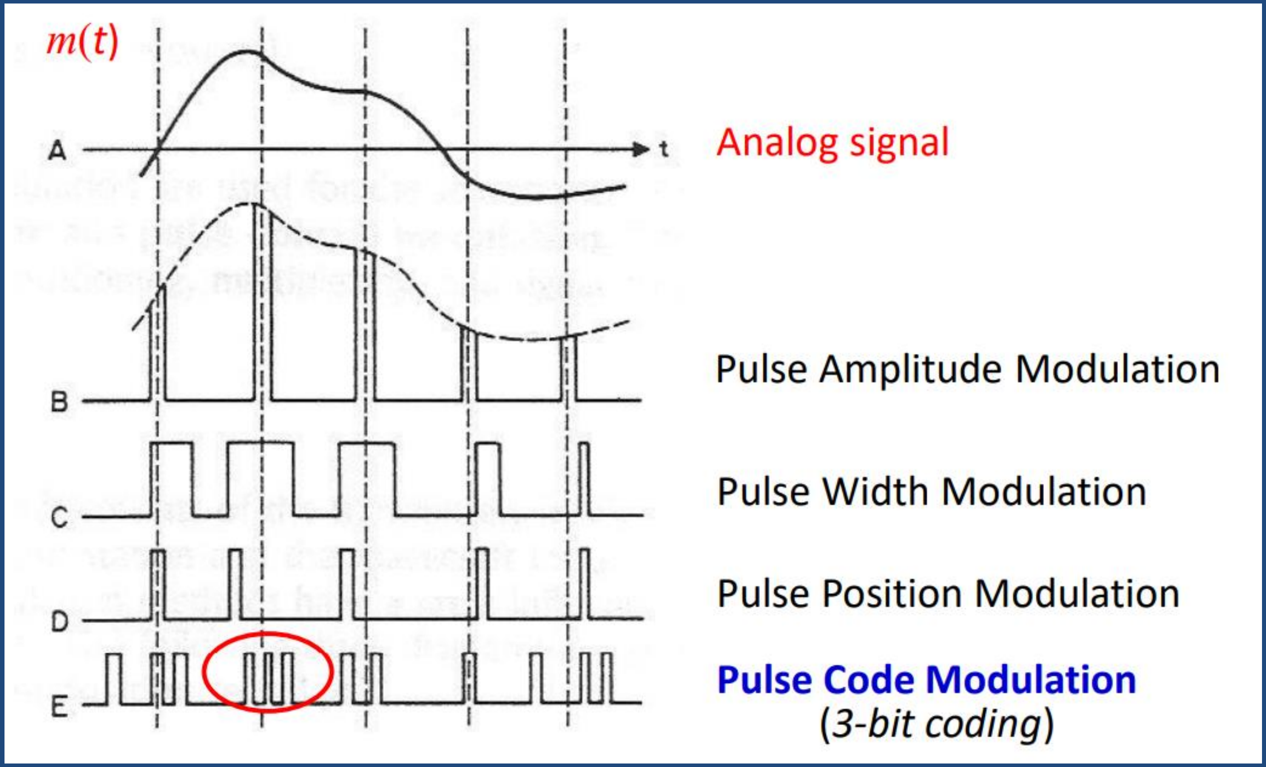
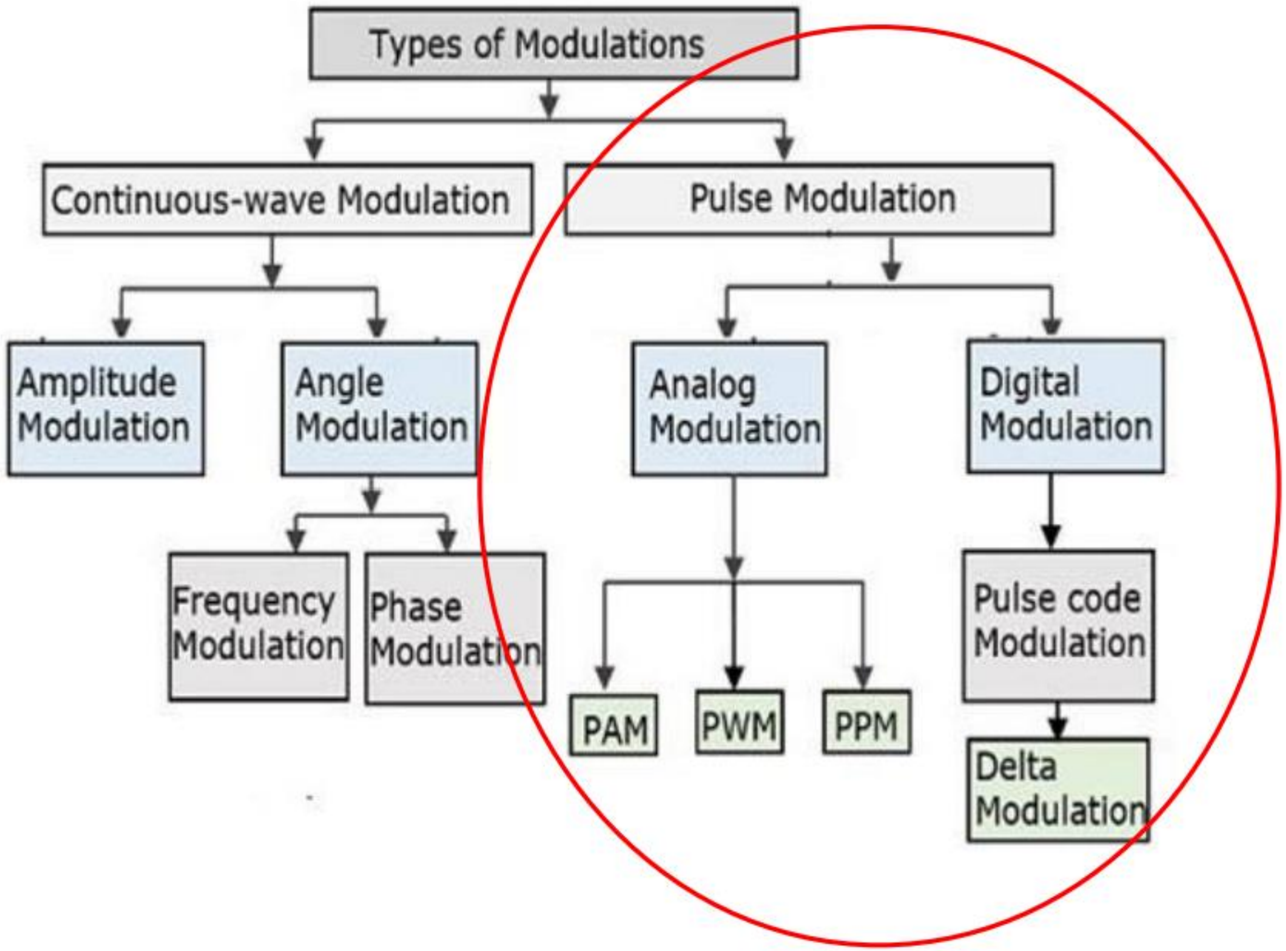
LECTURE 5:
ANALOG TO DIGITAL CONVERSION



This lecture

Reading Proakis and Salehi, Chapter 7



Roadmap

- Thus far we have covered analogue modulation techniques: namely 4 types of amplitude modulation and 2 types of angle modulation (focusing mainly on FM)
- It is time to go digital. The conversion from analogue to digital is accomplished in two steps:
    1. sampling - the conversion of a signal from continuous to discrete *in time*
    2. quantization - the conversion of the signal samples from continuous to discrete *in amplitude*
- following step 1, we will consider transmission schemes of sampled (but not quantized) signals
    - Pulse Amplitude Modulation (PAM)
    - Pulse Position Modulation (PPM)
    - Pulse Width Modulation (PWM)
- PAM, PPM and PWM are *analogue* modulation schemes
- as we will see, the sampling process is lossless (if done correctly)

Roadmap

- the transition to digital is complete following the discretization of the sample levels
- this process is *lossy*, and as a result there are a different methods aiming to reduce the loss of fidelity
- following the discussion of quantization, we will consider techniques of transmitting digital data over a communication link, which is intrinsically analogue
  - Pulse-Code Modulation (pulses denoting bit values)
  - Delta-Modulation / Delta-Sigma Modulation
- In PCM, there are a number of ways to represent the bit sequence using pulses - these are called *line coding*

Roadmap

- Having discussed the digital signal representation and transmission, we will then look the receiver methods
- in order to recover the digital content, the incoming waveform (pulses) must be sampled, and the sampler output interpreted as binary data
- we will consider the detection of pulses corrupted by additive noise
- this will lead to a key result in communication theory: *Matched Filter receiver*
- we will consider how to design the pulses to mitigate dispersive effects of the channel - *Intersymbol Interference* (ISI)
- this will lead another key result: *Nyquist's Criterion for Distortionless Transmission*

Outlines

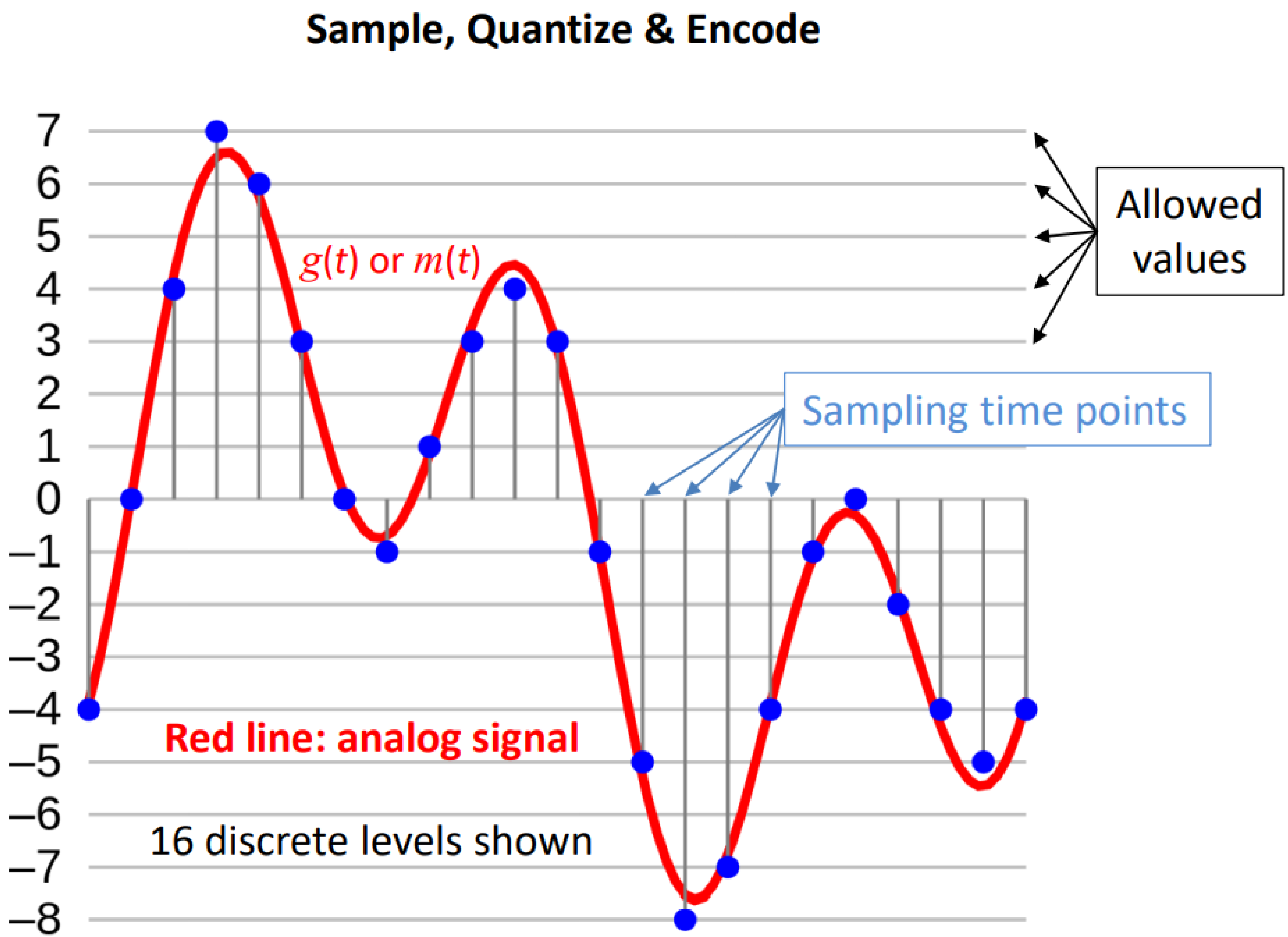
- Background: Sampling
- Pulse Amplitude Modulation
- Pulse Position/Width Modulation
- Background: Quantization
- Pulse Code Modulation
- Line Codes

Why digital? The need for sampling

1. Digital is more robust than analog to noise and interference†
2. Digital is more viable to using regenerative repeaters
3. Digital hardware more flexible by using microprocessors and VLSI
4. Can be coded to yield extremely low error rates with error correction
5. Easier to multiplex several digital signals than analog signals
6. Digital is more efficient in trading off SNR for bandwidth
7. Digital signals are easily encrypted for security purposes
8. Digital signal storage is easier, cheaper and more efficient
9. Reproduction of digital data is more reliable without deterioration
10. Cost is coming down in digital systems faster than in analog systems and DSP algorithms are growing in power and flexibility

† Analog signals vary continuously and their value is affected by all levels of noise.

Picture: Practical sampling process (uniform sample interval)



Background: The Sampling Process

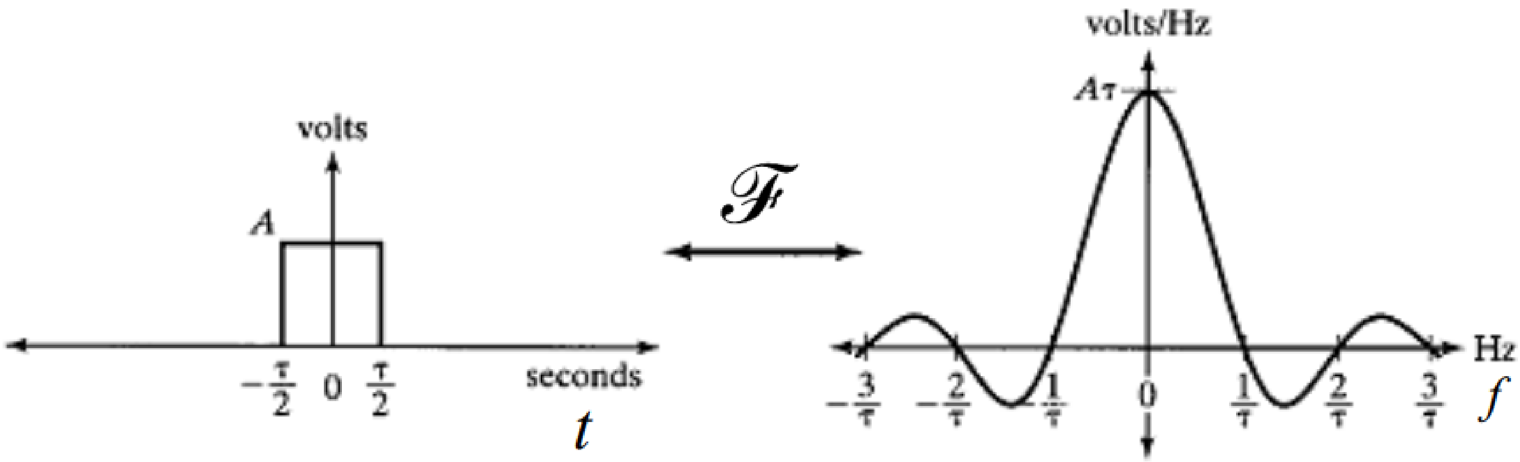
- Recall that a signal that is bandlimited to B Hz, that is $G(f) = 0$ for $|f| > B$ can be reconstructed exactly from its samples taken uniformly at a rate of $f_s = 2B$ Hz.
- Note that the above applies to *low-pass signals*; a signal with spectrum in $f_c - B/2 < |f| < f_c + B/2$ band also has a bandwidth B, yet requires a different sampling technique.
- Let's look a the details...

Remember this…

An absolutely band-limited waveform can’t be absolutely time-limited,

and

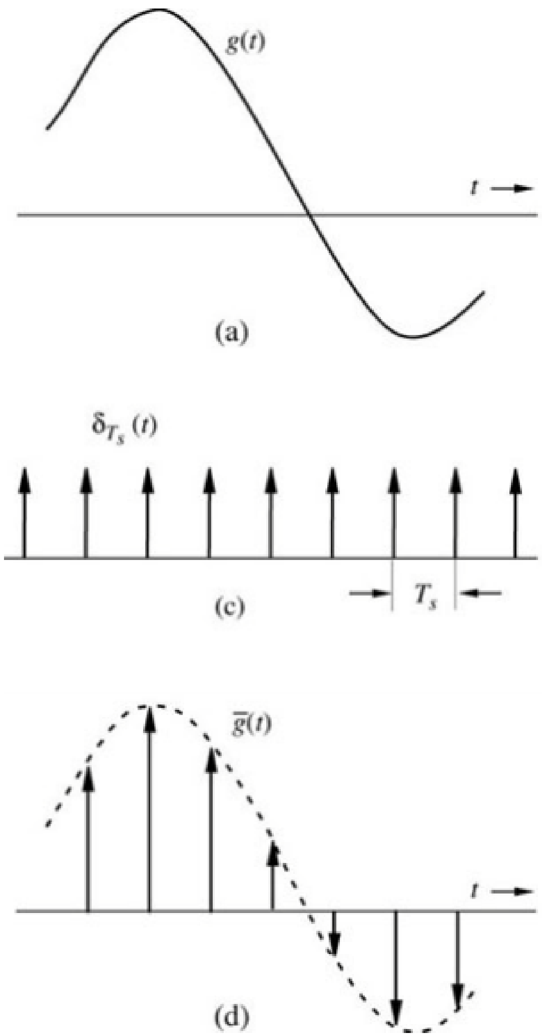
an absolutely time-limited waveform can’t be absolutely band-limited.



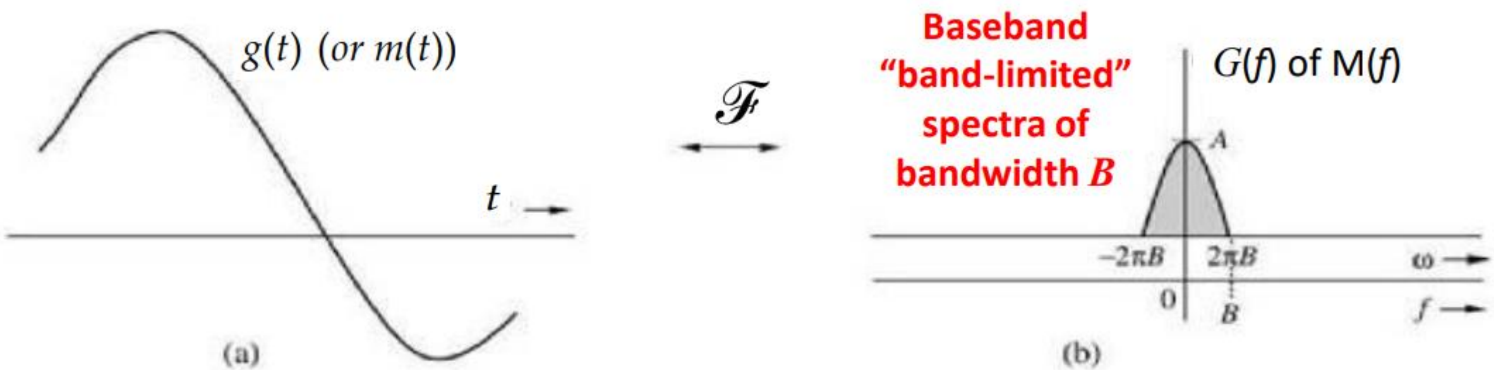
Background: The Sampling Process

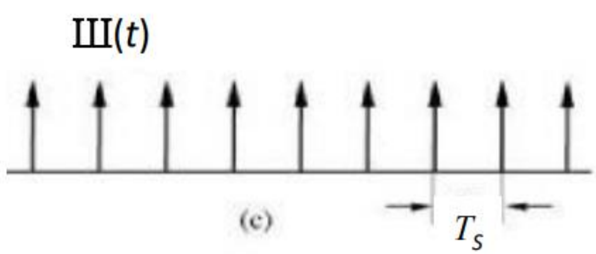
- Consider a continuous signal $g(t)$, which is sampled at a uniform rate ($f_s = 1/T_s$), once every T_s seconds
- The sampling process can be accomplished by multiplying $g(t)$ by a impulse train $\delta_{T_s}(t)$
- The resulting sampled signal $\bar{g}(t)$ can be mathematically described by a periodic sequence of Dirac delta functions, weighted by the values $\{g(nT_s)\}$

$$\bar{g}(t) = g(t)\delta_{T_s}(t) = \sum_{n=-\infty}^{\infty} g(nT_s)\delta(t - nT_s)$$

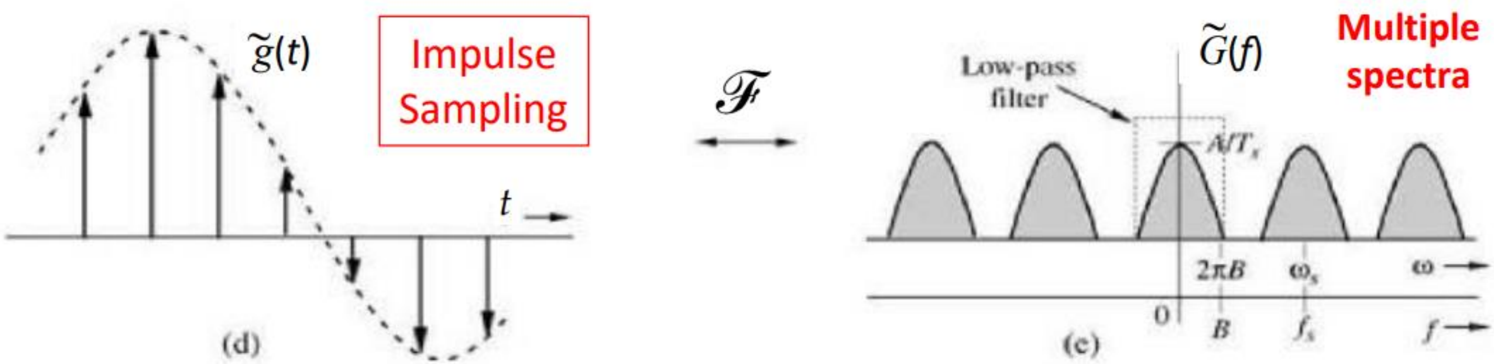


The sampling process





$G(f)$ consists of $G(f)$, scaled by the constant $1/T_S$, repeated periodically with period $f_S = (1/T_S)$ as shown here.



The Sampling Process

- one can show that $\bar{g}(t)$ has a Fourier transform of

$$\bar{g}(t) \leftrightharpoons \frac{1}{T_s} \sum_{n=-\infty}^{\infty} G(f - nf_s)$$

- That is, *the process of uniformly sampling a continuous-time signal results in a periodic spectrum with the period equal to the sampling rate*
- Thus, provided that $G(\omega) = 0$ is bandlimited to B Hz, the original signal $g(t)$ can be recovered by passing $\bar{g}(t)$ through an ideal lowpass filter of bandwidth B Hz.

Side note on Fourier Transform:

The Fourier transform (*i.e.*, spectrum) of $f(t)$ is $F(\omega)$:

$$F(\omega) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt$$

$$f(t) = \mathcal{F}^{-1}\{F(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega$$

Therefore, $f(t) \Leftrightarrow F(\omega)$ is a Fourier Transform pair

The Sampling Process

- The sampling theorem states that *A band-limited signal of finite energy, which has frequency content less than* B *Hz, is completely described by (and can be completely recovered from) the values of the signal at times separated by* $1/2B$ *seconds*
- The sampling rate of $2B$ samples per second, for a signal bandwidth B, is referred to as the *Nyquist rate*
- The reconstruction of $g(t)$ from $g_\delta(t)$ is accomplished by passing the later through an *ideal* lowpass filter
- This is equivalent to a convolution of $g_\delta(t)$ with a sinc function

Maximum Information Rate

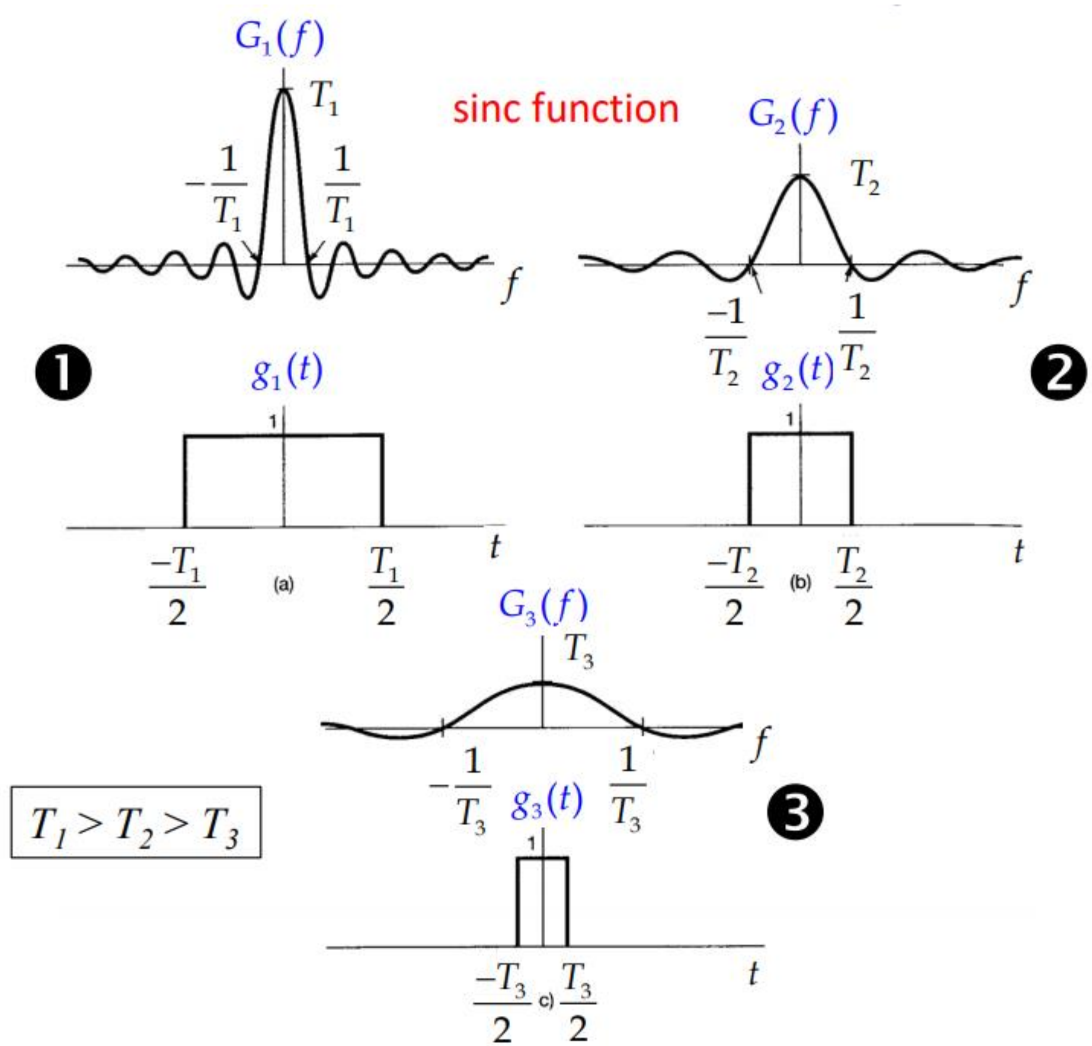
- The sampling theorem allows us to derive a key relationship of communications: the maximum rate (for binary transmission) of information that can be transmitted over a channel with bandwidth B Hz.

Theorem (Maximum Information Rate)

Given a noiseless channel with bandwidth B Hz, we can transmit a maximum of $2B$ independent, error-free pieces of information per second.

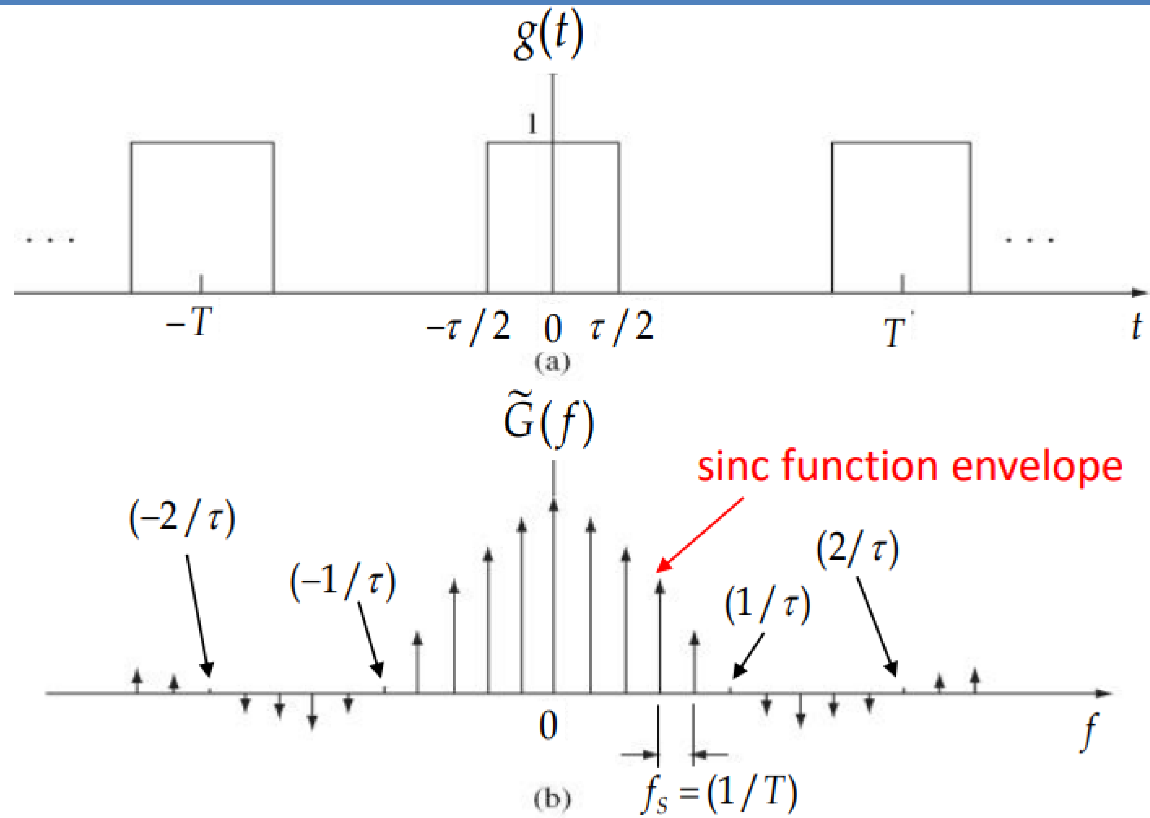
- If a rate of signal transmission is $2B$ then signal with bandwidth no greater than B is *sufficient* to carry the signal rate.
- Assuming no noise, a channel of bandwidth B Hz can transmit a signal of bandwidth B Hz error-free.
- Note however, that this can be extended by using M signal levels to represent multiple bits (ex Quaternary, i.e. $M = 4$), making the limit $C = 2B \log_2 M$ (more on this later)
- Recall that for a *noisy* channel, Shannon tells us $C = B \log_2(1 + \mathrm{SNR})$

More explanation: Sinc



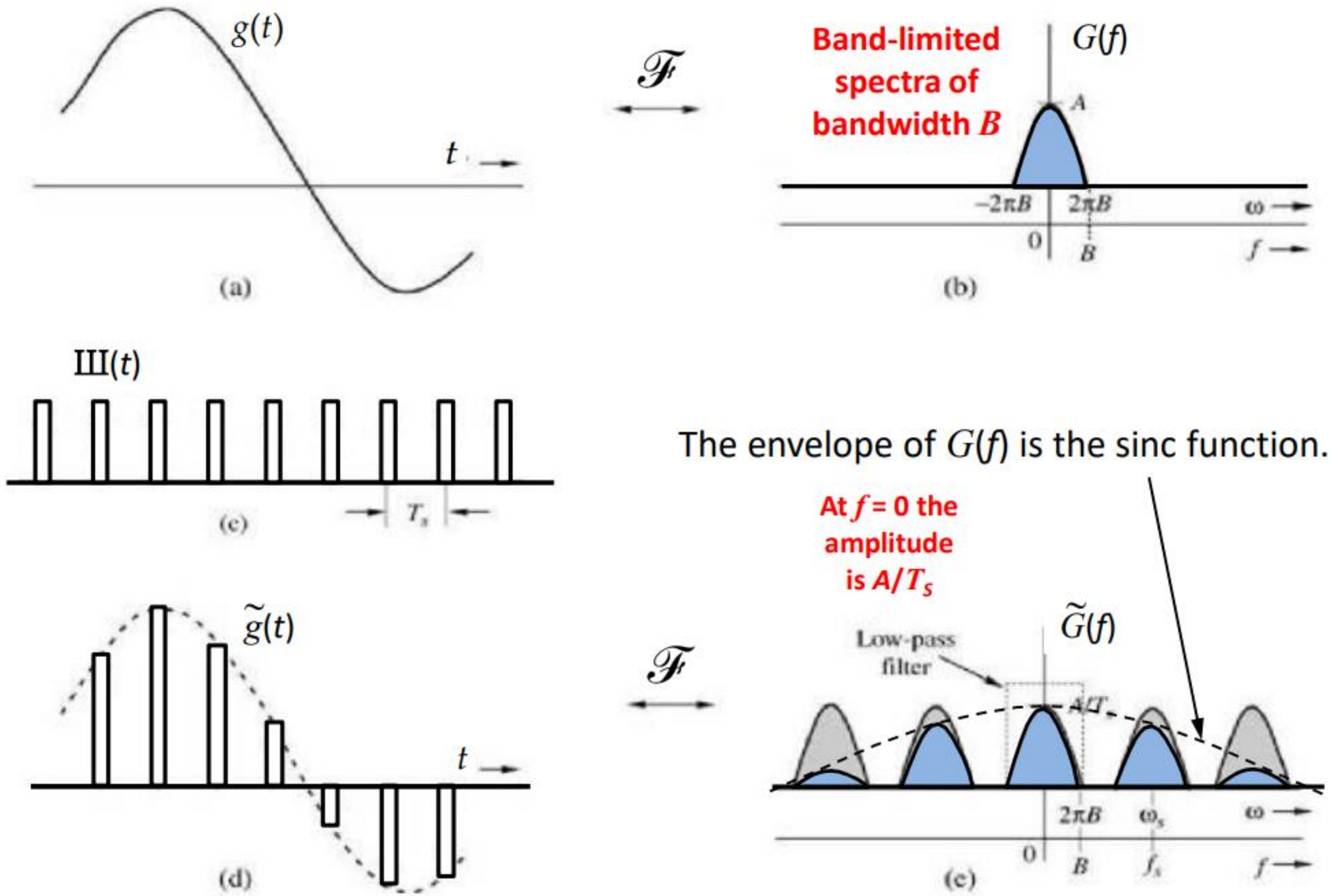
Narrower spectrum corresponds to wider pulse.

FT of periodic pulse train



For a pulse train we have discrete frequencies (rather than a continuum of frequencies) spaced as $(1/T)$ from each other. Thus, we have a comb of sinusoids at . . . $-2f_S$, $-f_S$, 0, f_S, $2f_S$, . . . , with amplitudes given by the sinc function associated with pulses of width τ.

FT of sampled signal



Can g(t) be reconstructed back from samples?

Answer: **<u>Yes</u>**, assuming no overlap, meaning that $f_s > 2B$, so the the sampling interval T_s must be < $(1/2B)$.

Minimum Sampling Rate:

The Minimum Sampling Rate $f_s = 2B$ is the ***Nyquist rate***. Oversampling occurs when the rate exceeds the Nyquist rate.

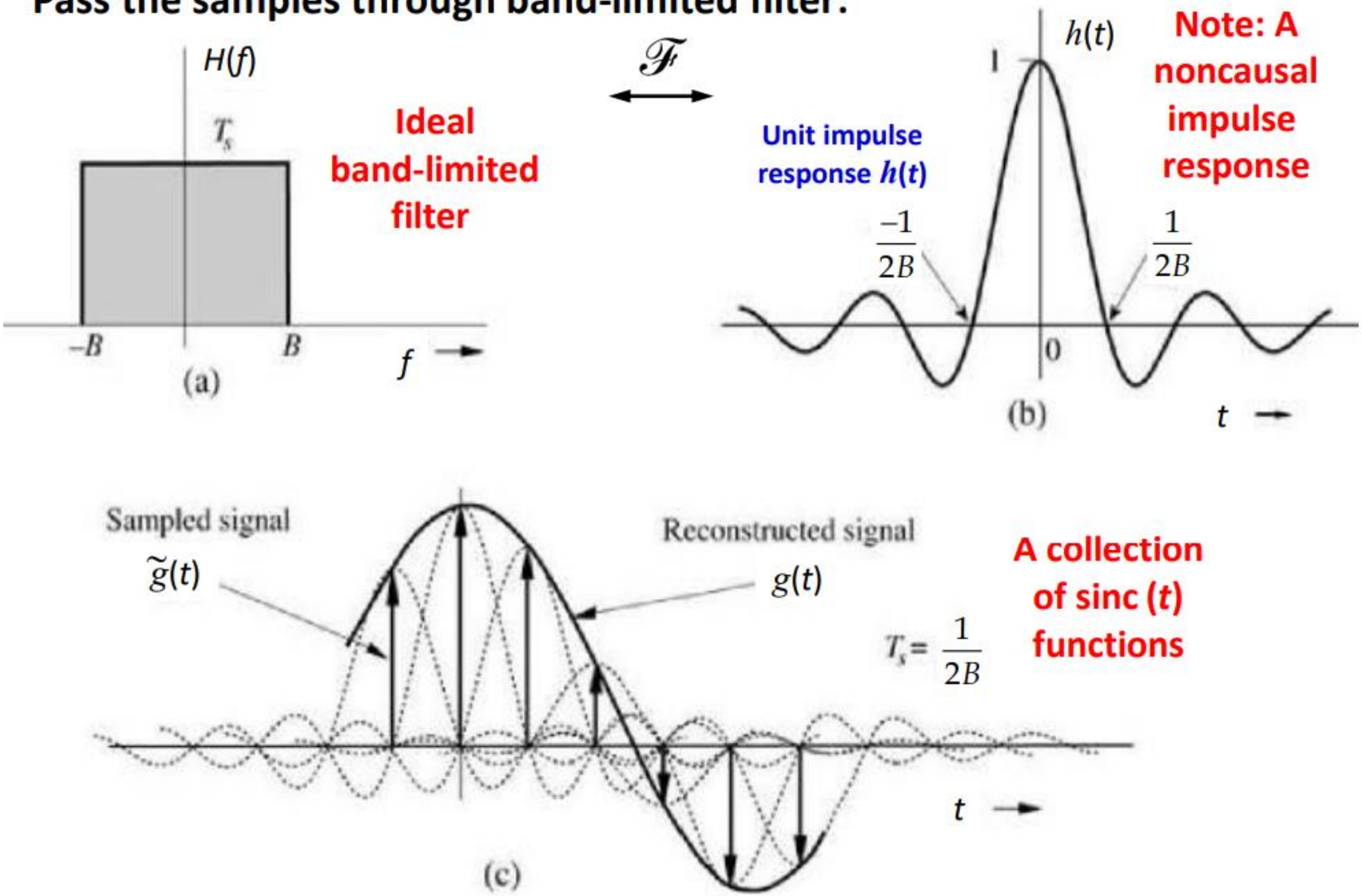
Interpolation:

Signal reconstruction is called ***Interpolation***. We can recover $g(t)$ by sending the samples through a band-limited filter of bandwidth B (Hz).

How we do this is illustrated in the next slide . . .

Ideal interpolation (graphically)

Pass the samples through band-limited filter:



Ideal interpolation (mathematically)

Mathematics for the previous slide in reconstruction of $g(t)$:

$$h(t) = 2BT_S \cdot \mathrm{sinc}(2\pi Bt)$$

At the Nyquist rate, with $2BT_S = 1$, then $h(t)$ becomes

$$h(t) = \mathrm{sinc}(2\pi Bt)$$

$$g(t) = \sum_k g(kT_S) \cdot h(t - kT_S)$$

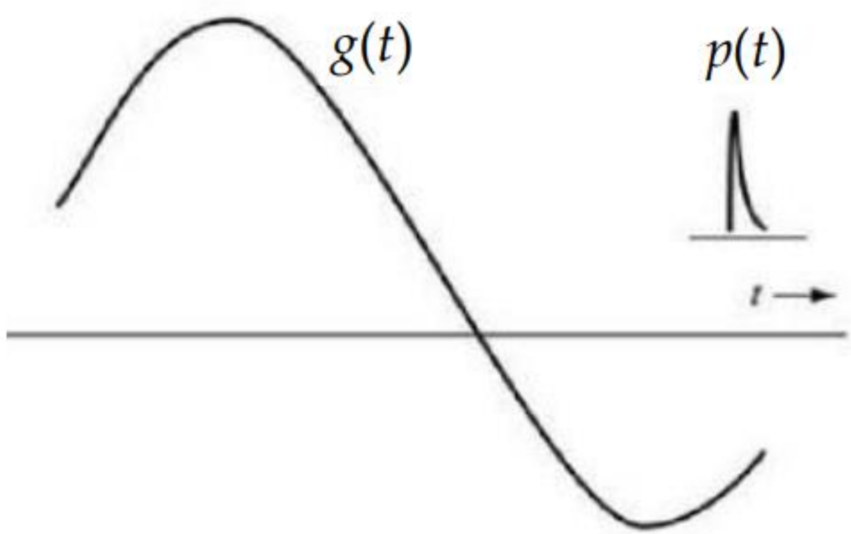
$$g(t) = \sum_k g(kT_S) \cdot \mathrm{sinc}\left[2\pi B(t - kT_S)\right]$$

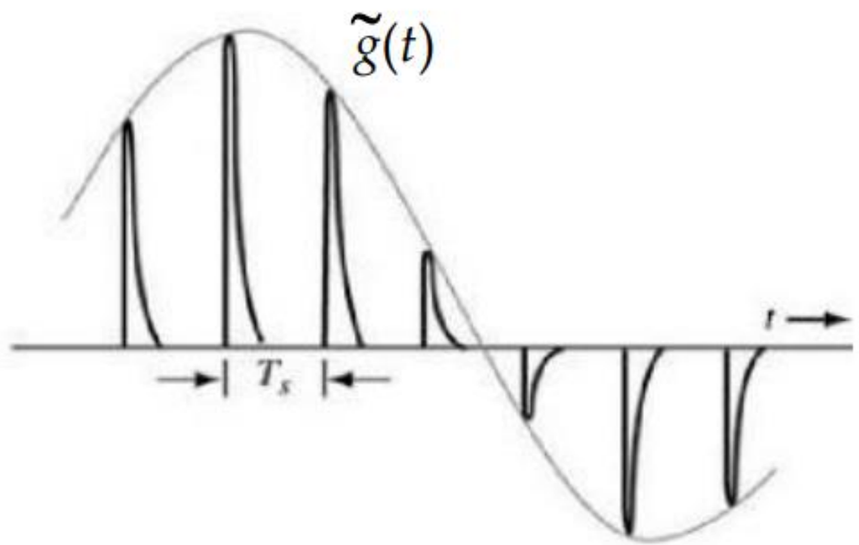
Example: Reconstruction of non-ideal samples

Impulse functions are not physically realizable, so a sampling pulse can not be truly instantaneous. Here $p(t)$ is the reconstruction pulse.

$$\tilde{g}(t) = \sum_n g(nT_S)p(t - nT_S) = p(t) * \overbrace{\left[\sum_n g(nT_S)\delta(t - nT_S) \right]}^{\text{Convolution operation}}$$

$$= p(t) * g(t)$$





Example: Reconstruction of non-ideal samples

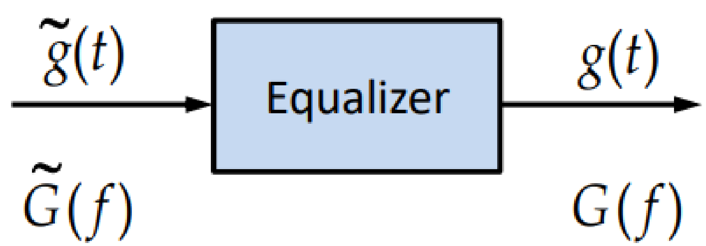
To reconstruct $g(t)$ from non-ideal sampling, we must filter $\tilde{g}(t)$ by using a special filter known as an ***equalizer.*** In the frequency domain,

$$\tilde{G}(f) = P(f)\frac{1}{T_S}\sum_n G(f - nf_S)$$

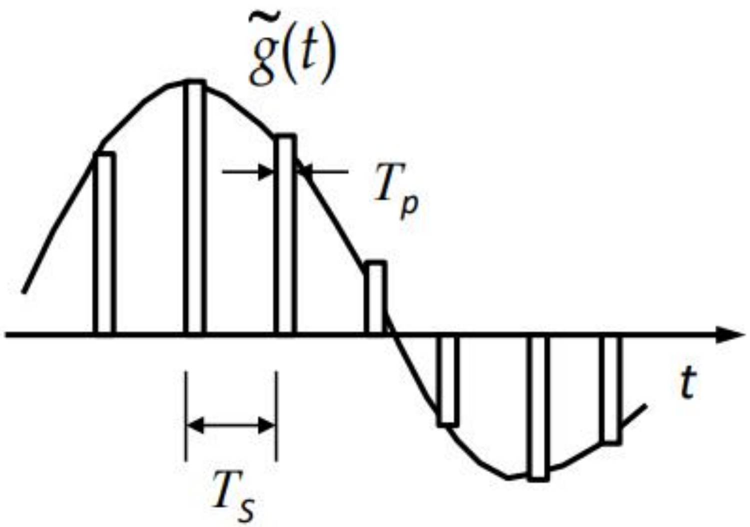
We denote the equalizer's transfer function as $E(f)$. The operation we use is to apply the filter to $\tilde{G}(f)$, hence

$$G(f) = E(f)\tilde{G}(f) = E(f)P(f)\frac{1}{T_S}\sum_n G(f = nf_S)$$

This relationship shows that the equalizer must remove all the shifted replicas $G(f - nf_S)$ in the summation except for the low-pass term at $n = 0$.



Signal recovery observation



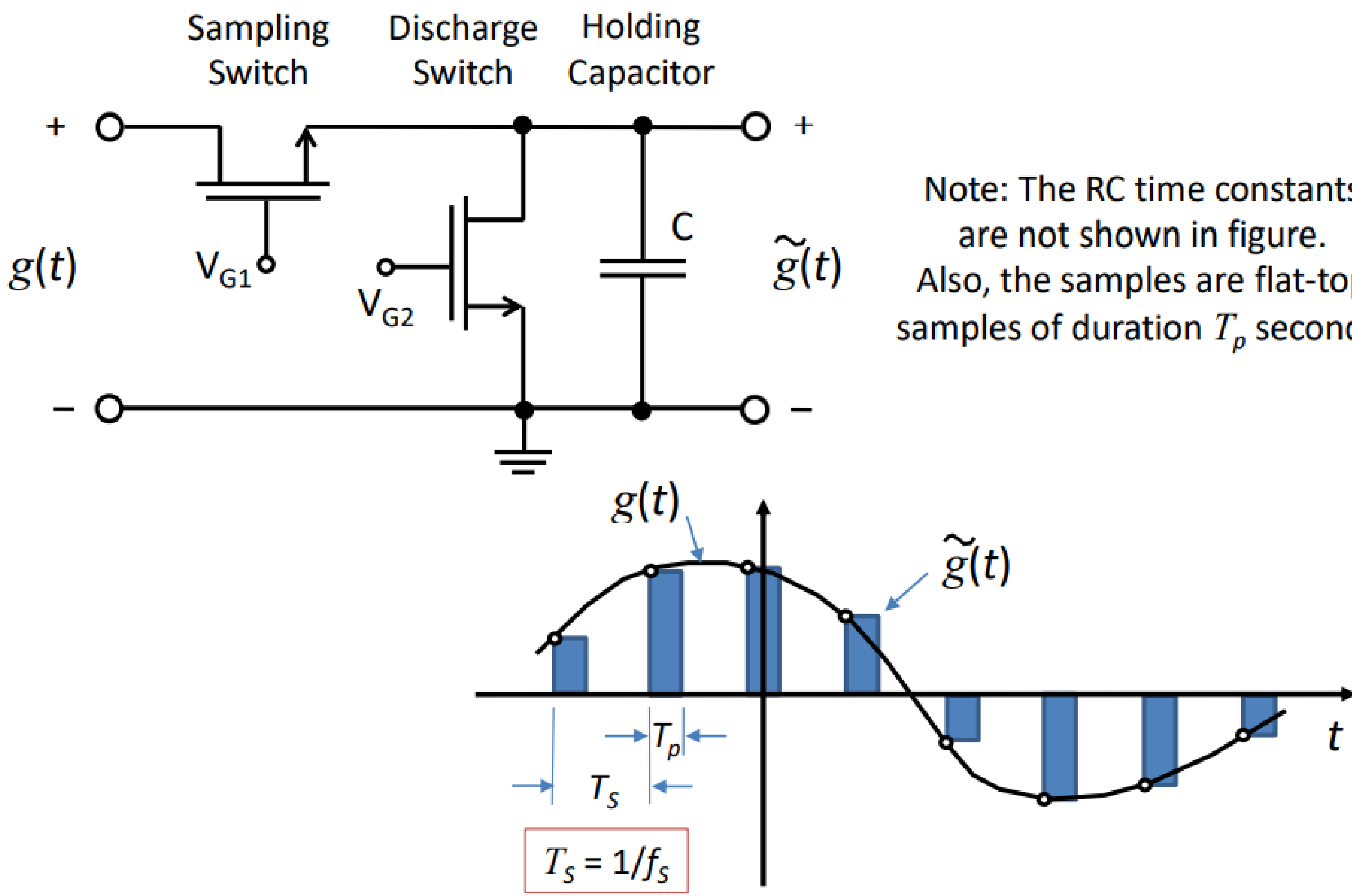
Generally we use <u>short</u> <u>pulses</u> of duration T_p to sample a function $g(t)$. This does a better job of signal interpolation without excessive demands upon an equalizer filter. In fact, often the equalizer can be omitted.

In practice it is impossible to precisely recover a band-limited signal g(t) from its samples, even if the sampling rate is higher than the Nyquist rate.

The sampling theorem rests upon the signal being strictly band-limited. All practical signals are time-limited, therefore can't be precisely band-limited.

Sample and hold circuit

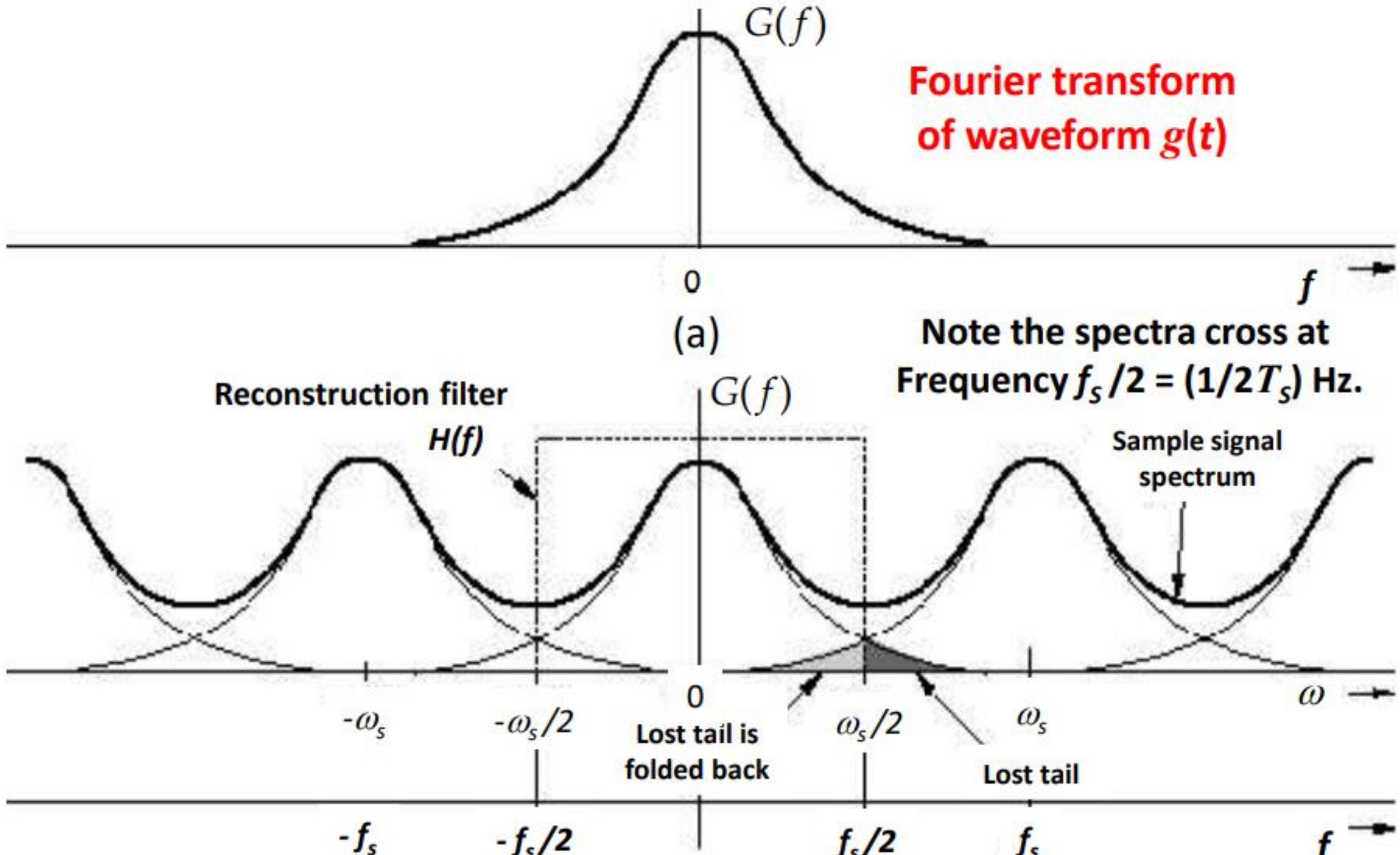
Basic sample and hold circuit and general waveform



Note: The RC time constants are not shown in figure. Also, the samples are flat-top samples of duration T_p seconds.

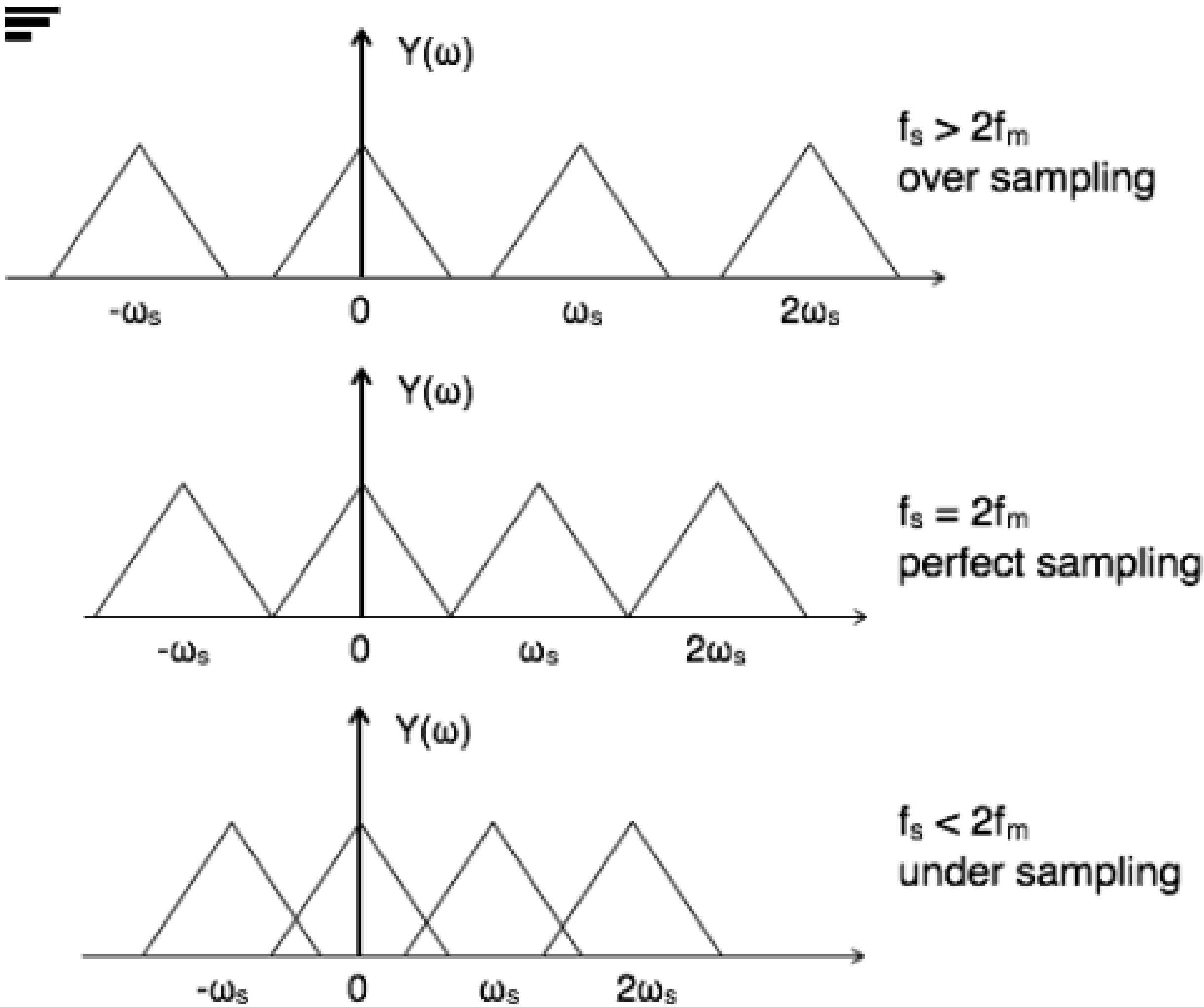
Aliasing (or spectral folding)

Because signals are not band-limited, they have long tails in the frequency domain as shown in $G(f)$. Sampling at higher rates does not eliminate spectral overlapping of repeated spectral cycles as shown in (b).



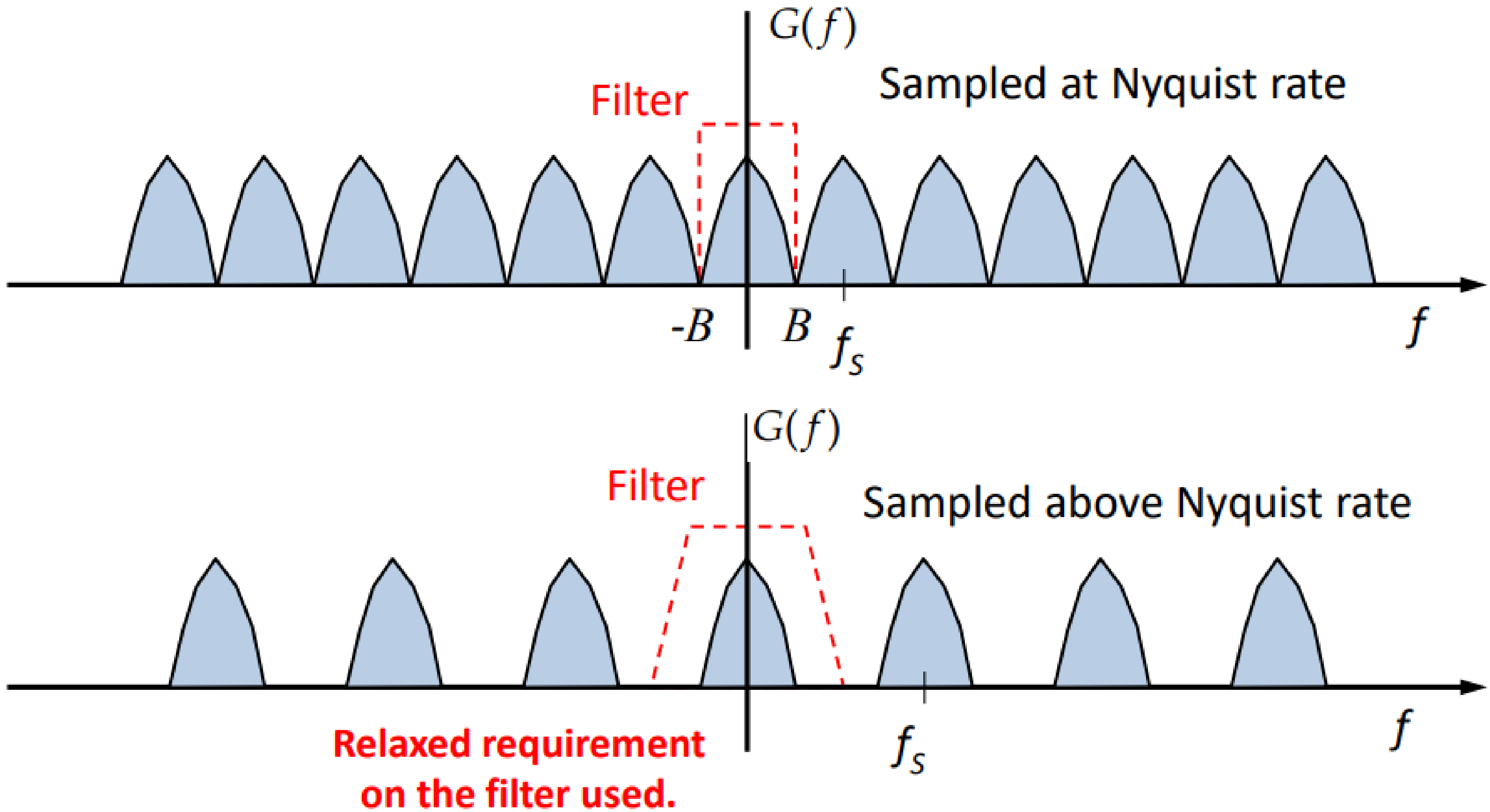
After Lathi & Ding, 4th ed., 2009; p. 311   (b)

Another view of Aliasing



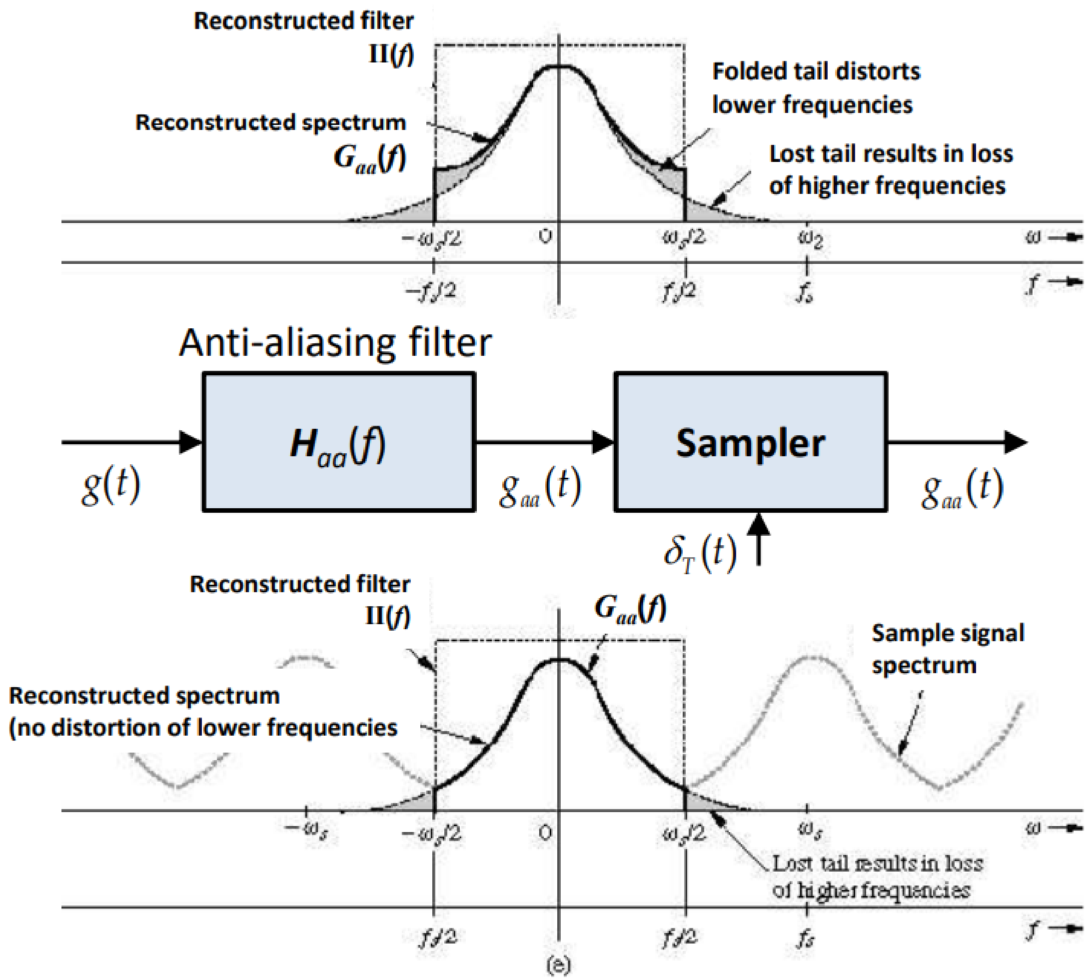
What can we do to avoid aliasing?

We can **oversample**, that is, we can sample at a rate exceeding the Nyquist rate. This is illustrated below:



What can we do to Reduce aliasing?

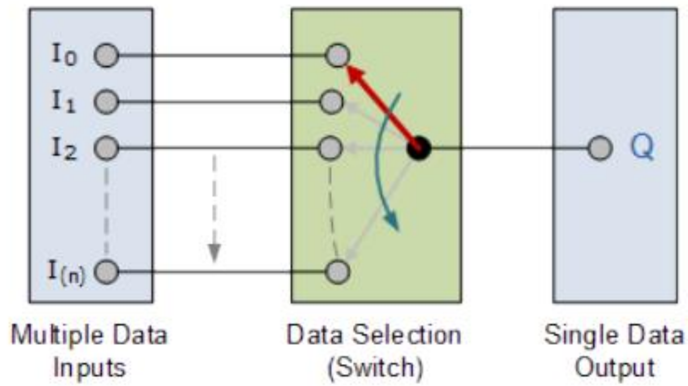
We could place an **anti-aliasing filter** in front of the sampler.



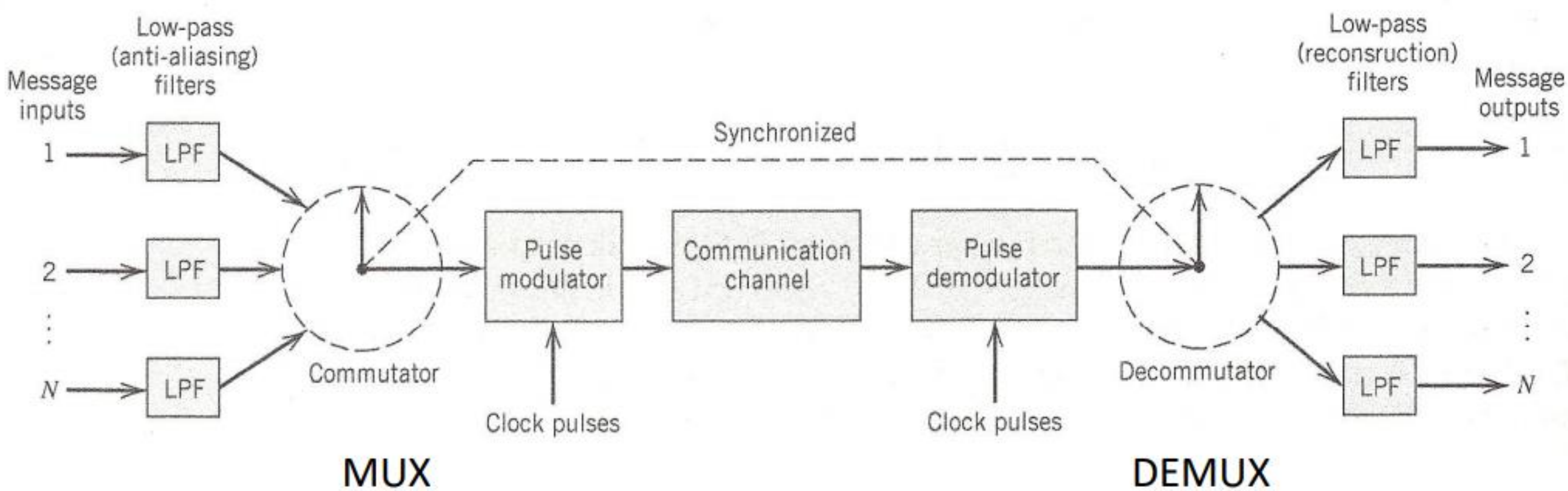
After Lathi & Ding, 4th ed., 2009; p. 311.

Multiplexing and demultiplexing

Time sharing a transmission medium.

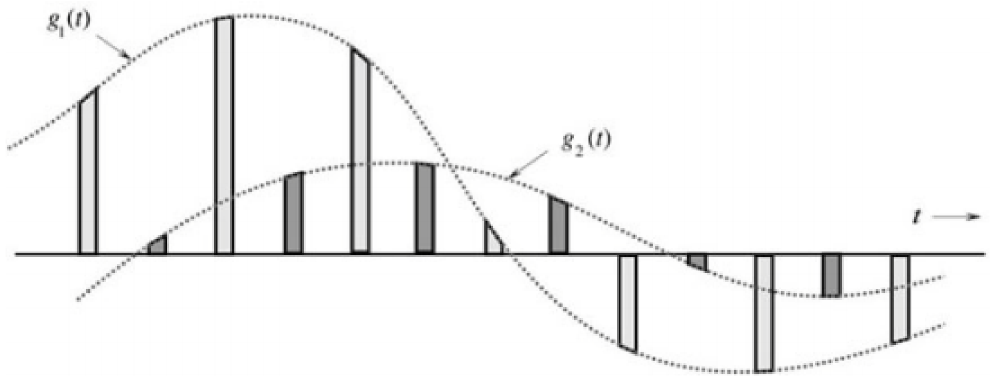


Time Division Multiplexed (TDM) Output



Time Division Multiplexing

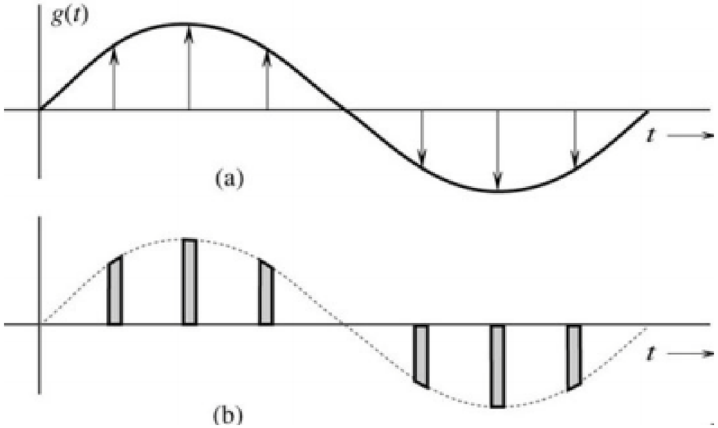
- An advantage of pulse modulation is that it allows for simultaneous transmission of several signal on time-sharing basis



- Simultaneous transmission of multiple signals = *multiplexing*
- Simultaneous transmission by time division = *time division multiplexing*

Pulse Amplitude Modulation (PAM)

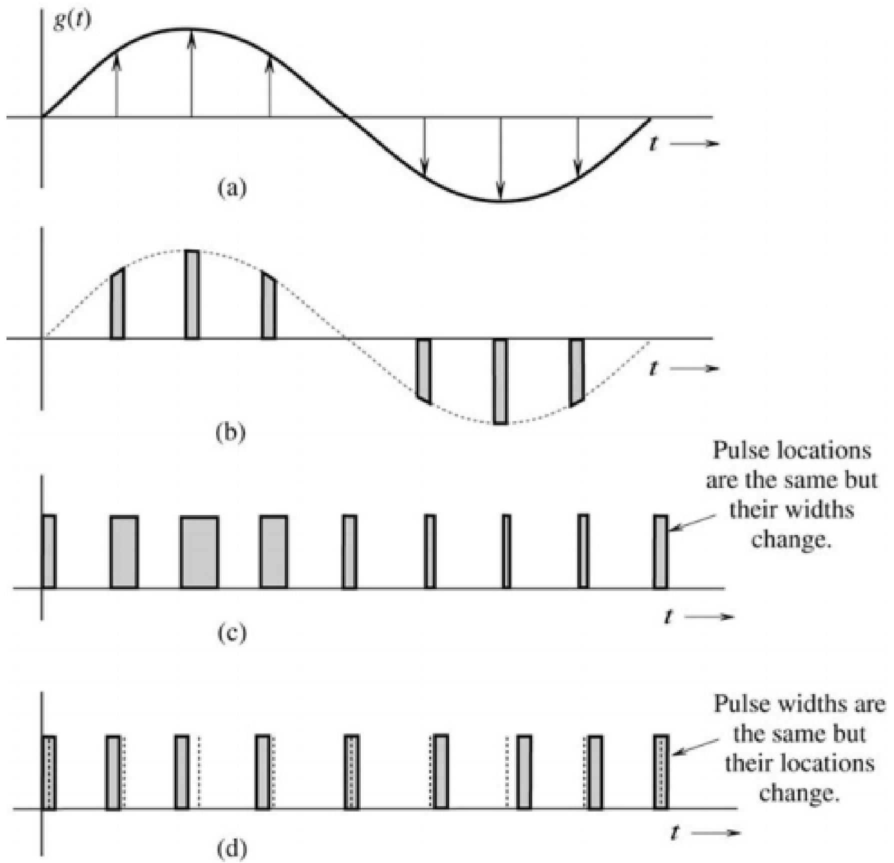
- Pulse-Amplitude Modulation (PAM) is a transmission method where the **amplitude** of regularly spaced pulses is proportional to the corresponding sample of the message signal



- There are two operations in the generation of PAM signal
  1. instantaneous sampling of $m(t)$ at the sampling rate $f_s = 1/T_s$, chosen in accordance to the sampling theorem
  2. lengthening of the duration of each sample to some constant value $T < T_s$
- these operations are jointly referred to as *sample and hold*
- why bother lengthening the samples? lengthening of the samples in necessary to avoid excessive channel bandwidth, since bandwidth is inversely proportional to pulse duration

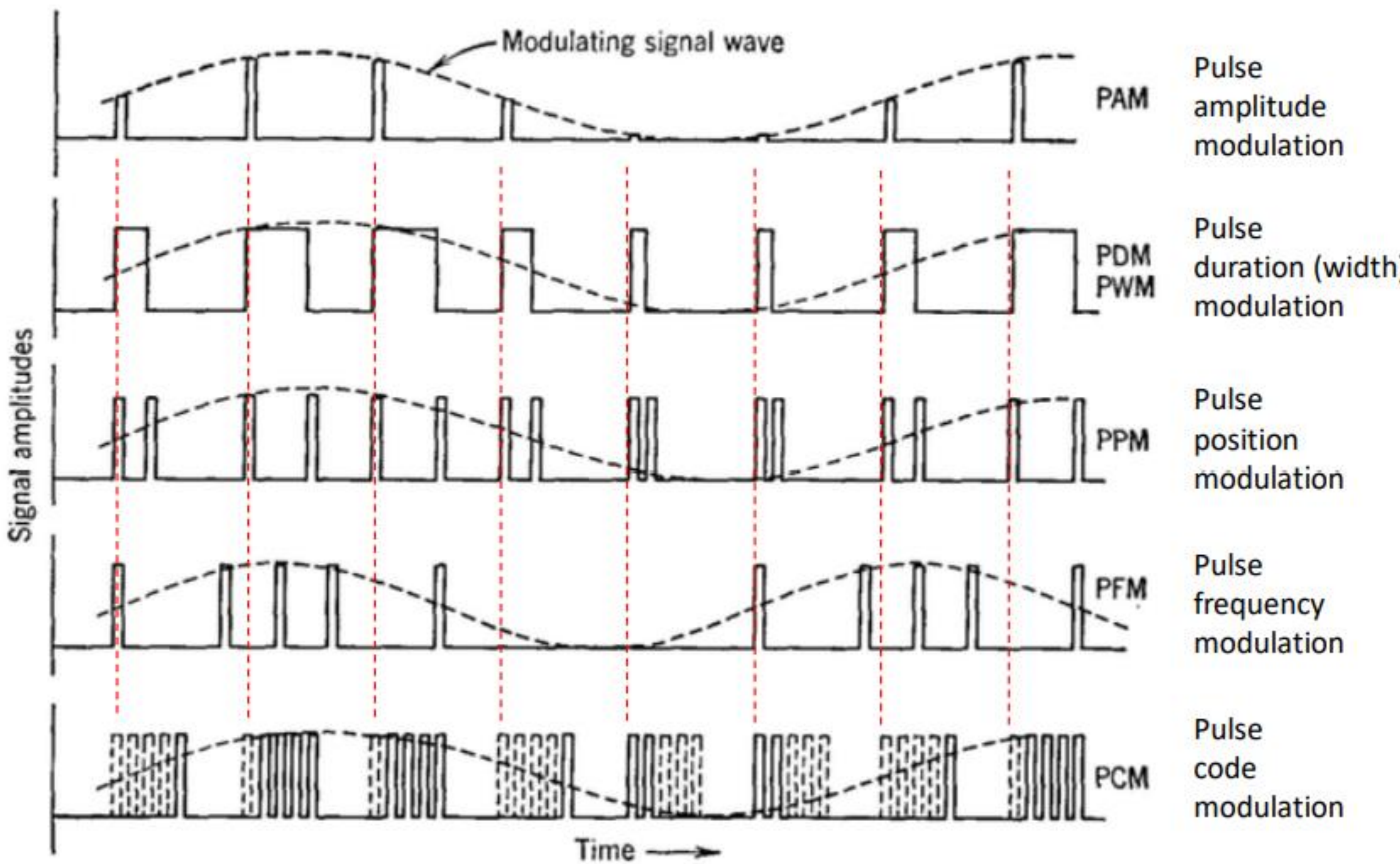
Pulse Duration Modulation / Pulse Position Modulation

- In *pulse-duration modulation* (PDM) the samples of the message signal are used to vary the *duration* of the individual pulses
- this is often referred to as *pulse-width* or *pulse-length* modulation



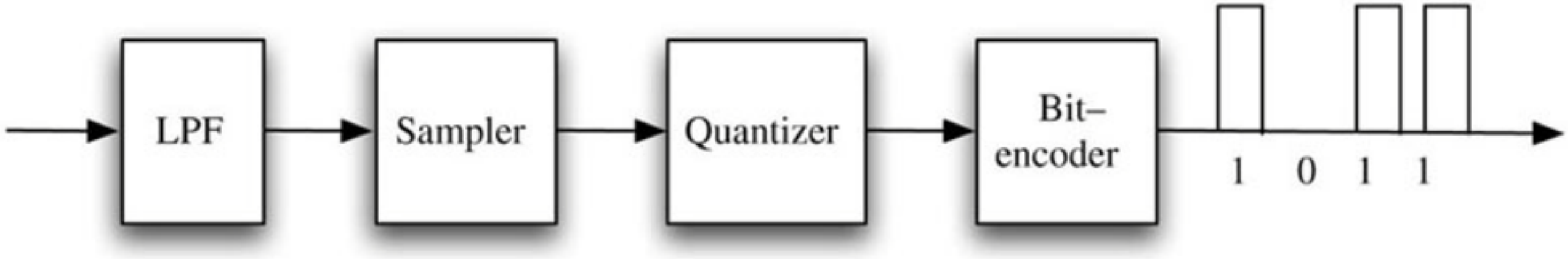
- In *pulse-position modulation* (PPM) the samples of the message signal are used to vary the *position* of the individual pulses
- (a) message; (b) PAM; (c) PWM; (d) PPM

Pulse modulation in picture

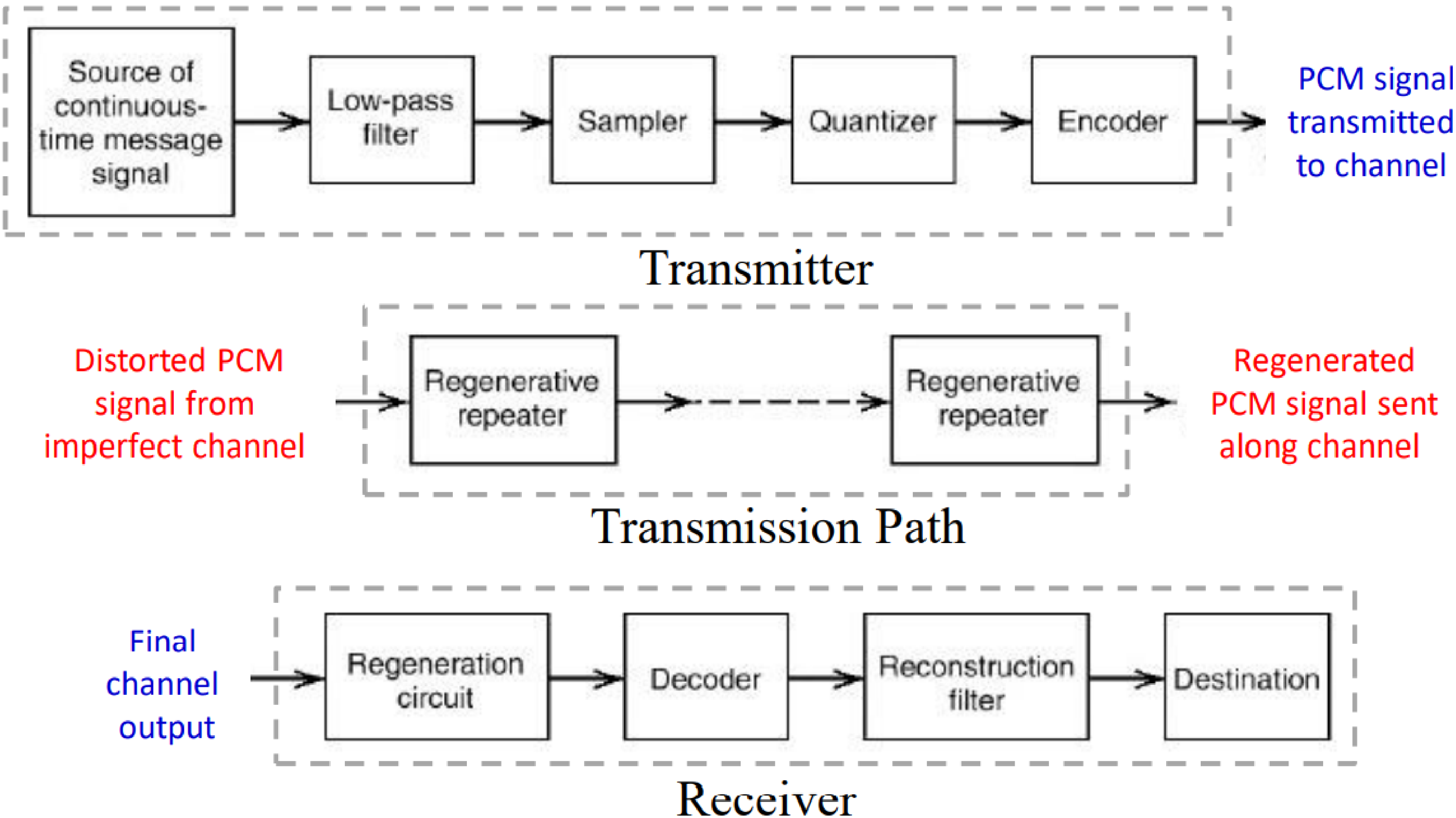


Pulse Code Modulation - Preview

- In digital systems, the pulses constituting PAM need to be converted to a binary data stream
- PAM amplitudes can take on an infinite number of values, which must be *quantized* into a finite number of levels
- These are then converted to binary and used to *encode* a pulse stream (in any number of ways)
- The resulting process is referred to Pulse Code Modulation (PCM)

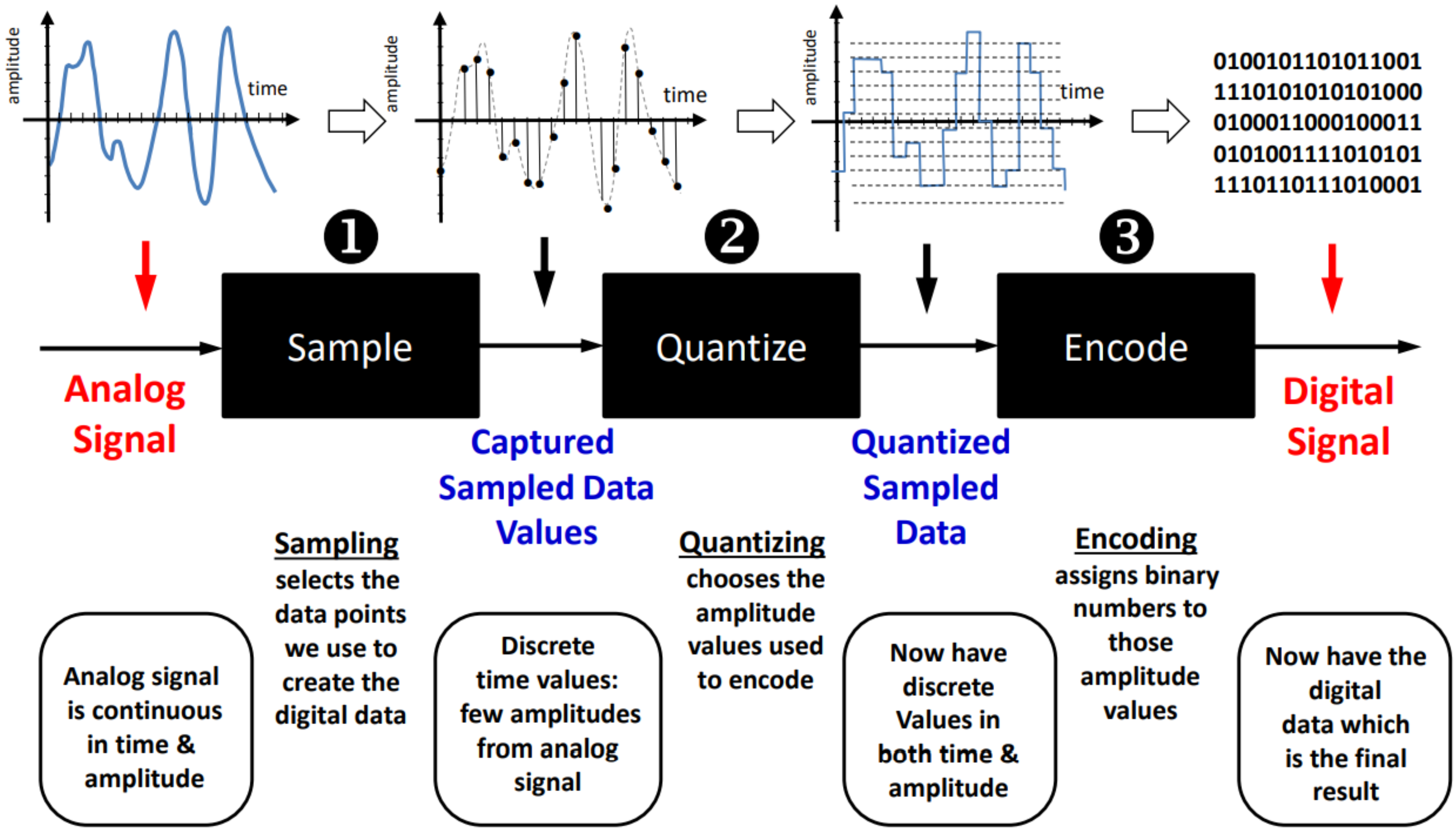


Typical PCM communication system



Analog to digital conversion process

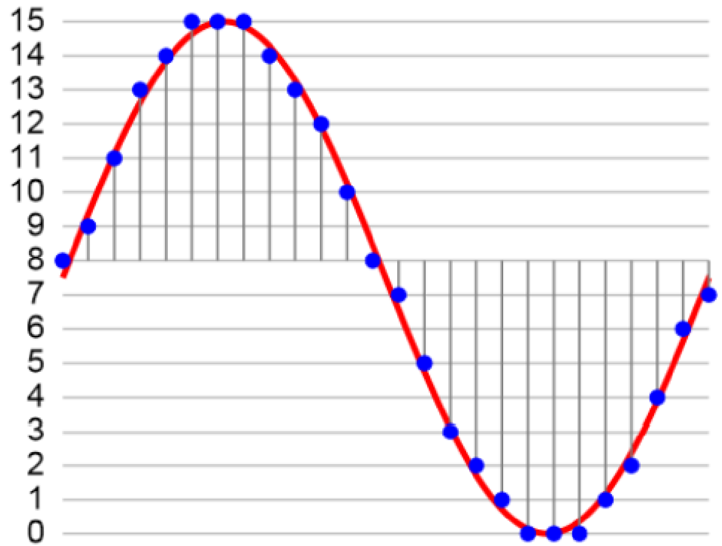
Three Step Process



Note: “Discrete time” corresponds to the timing of the sampling.

PC in picture

Pulse-code modulation (PCM) is used to digitally represent sampled analog signals. It is the standard form of digital audio in computers, CDs, digital telephony and other digital audio applications. The amplitude of the analog signal is sampled at uniform intervals and each sample is quantized to its nearest value within a predetermined range of digital levels.



Four-bit coding (16 discrete levels)

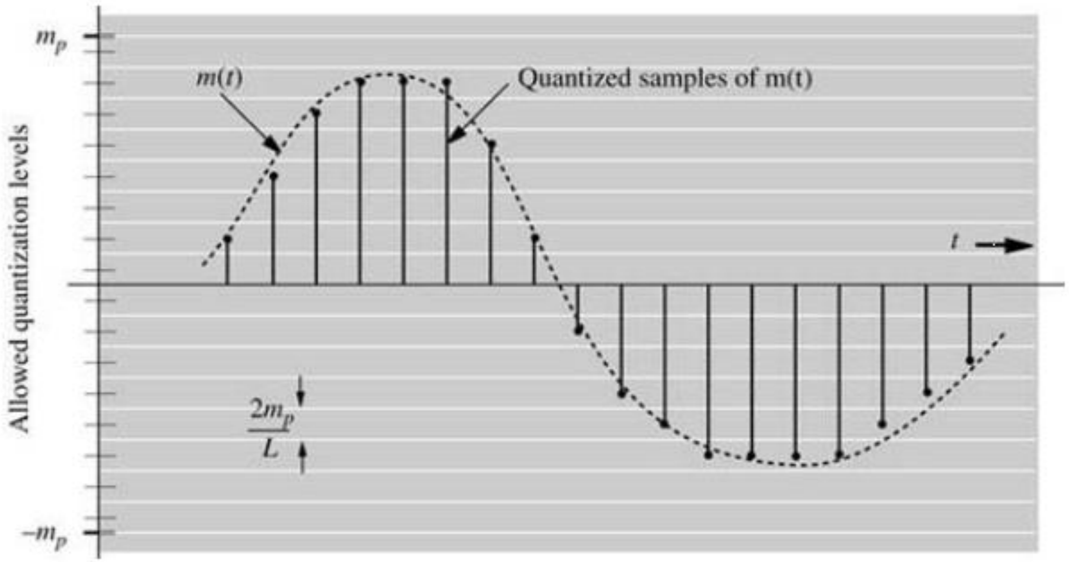
Example:

Digit	Binary equivalent	Pulse code waveform
0	0000	
1	0001	
2	0010	
3	0011	
4	0100	
5	0101	
6	0110	
7	0111	
8	1000	
9	1001	
10	1010	
11	1011	
12	1100	
13	1101	
14	1110	
15	1111	

Example

Quantization

- The process of quantization is a *lossy one* - associated with it is *quantization noise*
- The analog signal samples are rounded off to the closest permissible numbers - *quantization levels*



- The analog signal $m(t)$ is assumed to lie in a range of $(-m_p, m_p)$, which is partitioned into L subintervals of width $\Delta\nu = 2m_p/L$
- Each signal sample is rounded off to the midpoint of an interval, resulting in one of L values

Quantization

- One can show (see Lathi) that the quantisation noise power by N_q, that is

$$N_q = \frac{m_p^2}{3L^2}$$

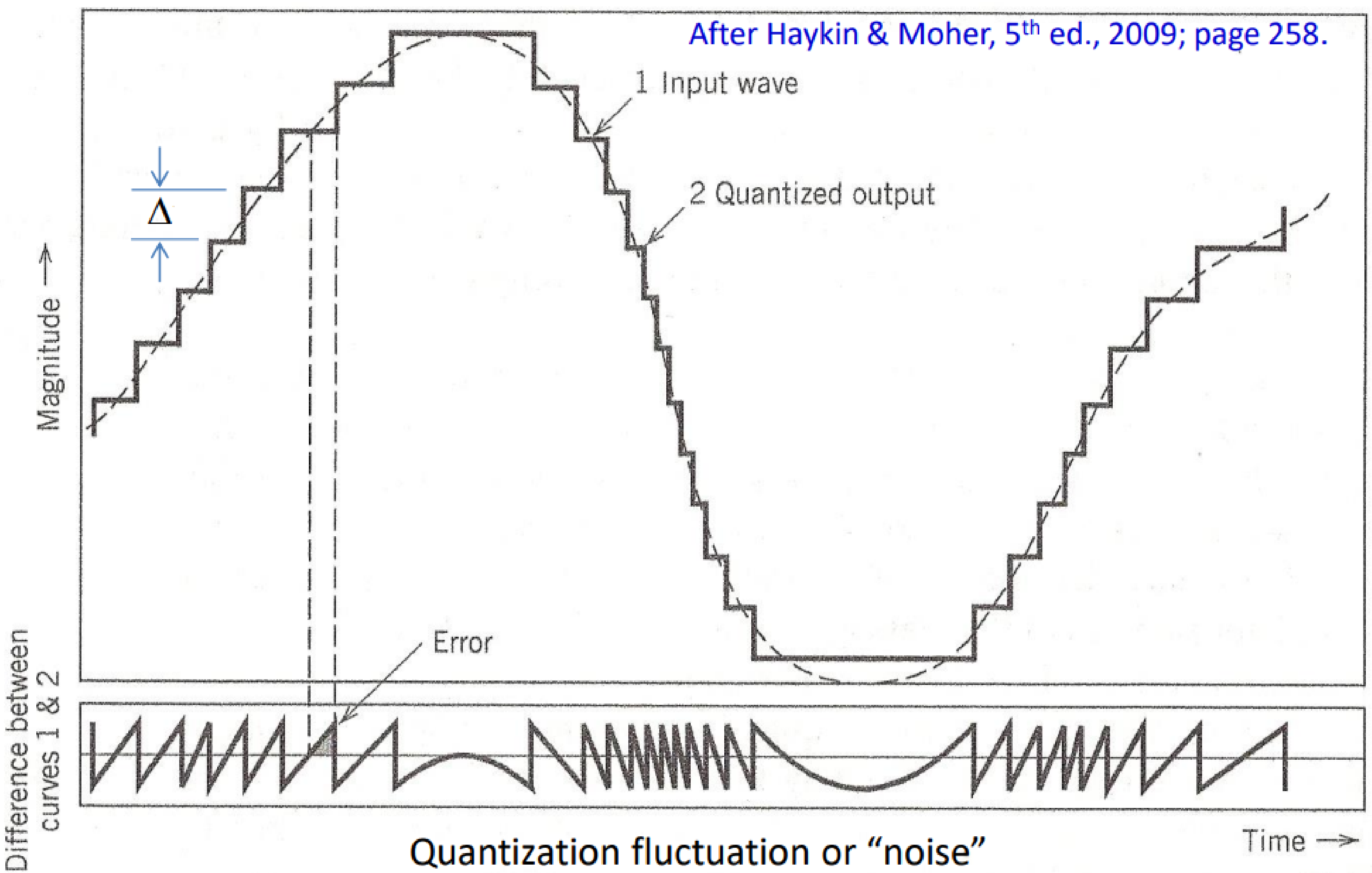
- The *desired* signal power is $S_0 = \widetilde{m^2(t)}$, and thus the *signal to noise ratio* (SNR) is

$$\frac{S_0}{N_0} = \frac{S_0}{N_q} = 3L^2 \frac{\widetilde{m^2(t)}}{m_p^2}$$

- The SNR is an measure of the quality of signal.
- The above includes only contribution of quantization noise, but later we will consider other sources, such as noise introduced by the channel

Quantization noise

Quantization noise $q(t) = m(t) - m_q(t)$



Quantization fluctuation or “noise”

More details: Quantization noise

The quantized levels are separated by $\boxed{\Delta = \frac{2m_p}{L}}$

The maximum error for any sample point's quantized value is at most ½Δ.

The "time average" **mean-square quantization error** is

$$\overline{q^2} = \langle q^2 \rangle = \frac{m_p^2}{3L^2} = \frac{(\Delta)^2}{12}.$$

Let $N_q = \overline{q^2}$. Thus, N_q is proportional to the fluctuation of the error signal. This is usually called the **quantization noise power**.

Next we define $m(t) = m_q(t) + q(t)$.

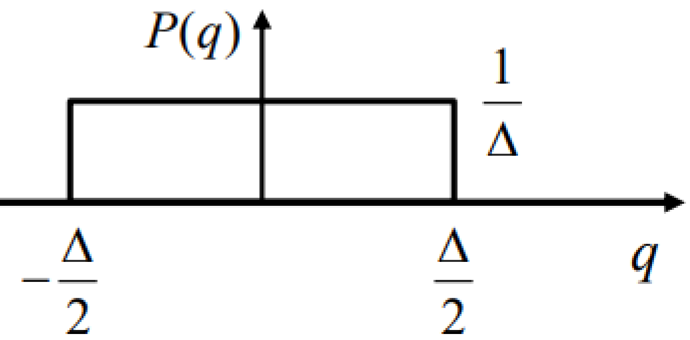
The signal (or message) power S_0 is proportional to the mean square of $m(t)$,

$$S_0 = \sigma^2 = \langle m^2(t) \rangle; \text{ and if } m(t) \text{ is sinusoidal, } S_0 = \frac{m_p^2}{2} = \frac{A_m^2}{2}$$

Note: $\langle \cdots \rangle$ denotes a time average.

Where Does (Δ^2/12) Come From?

Probability Density Function



This assumes the quantization error is uniformly distributed.

$$E(q^2) = \int_{\Delta/2}^{\Delta/2} \frac{1}{\Delta} q^2 dq = \frac{1}{\Delta}\left(\frac{q^3}{3}\right)\Bigg|_{-\Delta/2}^{\Delta/2} = \frac{1}{3\Delta}\left[\left(\frac{\Delta}{2}\right)^3 - \left(\frac{-\Delta}{2}\right)^3\right] = \frac{\Delta^2}{12}$$

$$\text{But,} \quad N_q = q^2 = \langle q^2 \rangle = E(q^2)$$

SQNR

We want a measure of the quality of received signal (that is, the ratio of the strength of the received signal power S_0 relative to the strength of the noise power N_q due to quantization).

This is the **Signal-to-Quantization Noise Ratio** (SQNR) and is given by

$$SQNR = \frac{S_0}{N_q} = \frac{\langle m^2(t) \rangle}{\left(m_p^2 \big/ 3L^2 \right)} = 3L^2 \left(\frac{\langle m^2(t) \rangle}{m_p^2} \right)$$

It is usually expressed in decibels,

Note: $\left(\frac{\langle m^2(t) \rangle}{m_p^2} \right) \cong \frac{1}{2}$

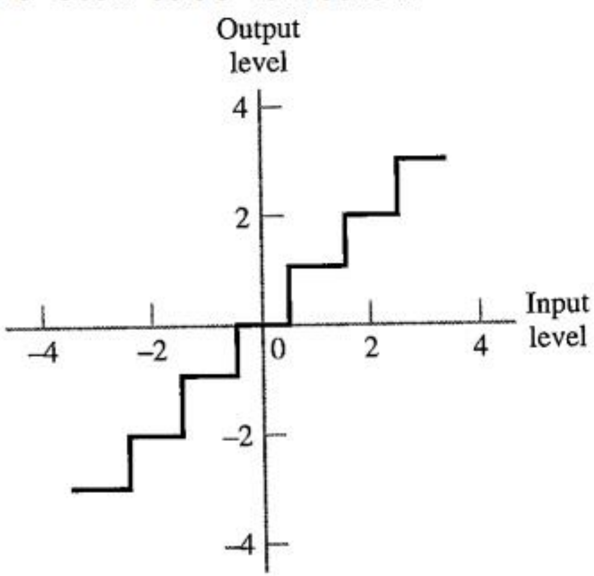
$$SQNR_{dB} = 10 \cdot \log_{10} \left(\frac{S_0}{N_q} \right) \cong 10 \cdot \log_{10} \left(\frac{3L^2}{2} \right)$$

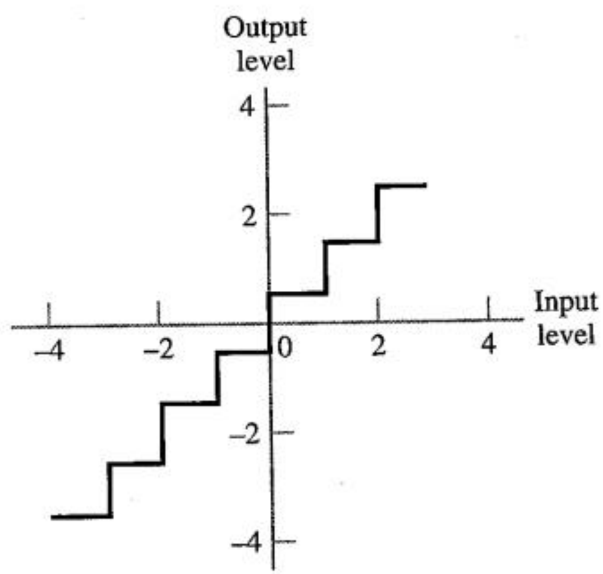
Conclusion:

To reduce the quantization error relative to the message signal level, use smaller quantization steps Δ.

Quantization

- The process of quantization can be described by a mapping called *quantizer characteristic*
- it is a staircase function

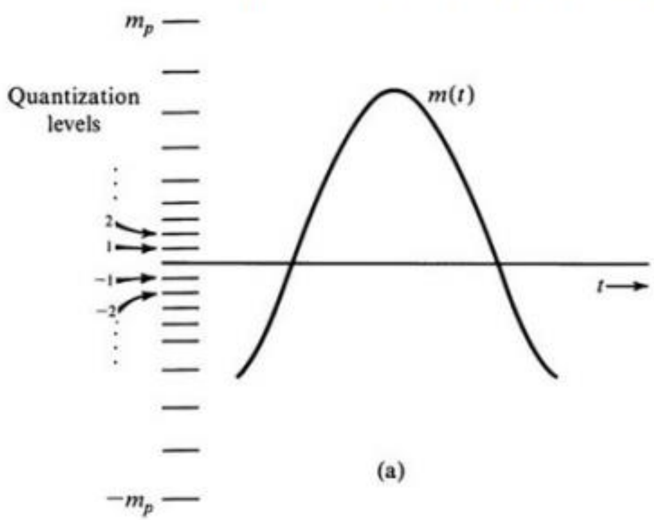




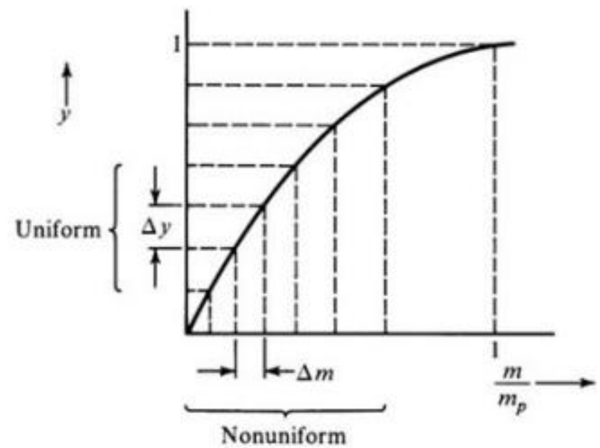
- Uniform quantizers have equally spaced representation levels

Nonuniform Quantization

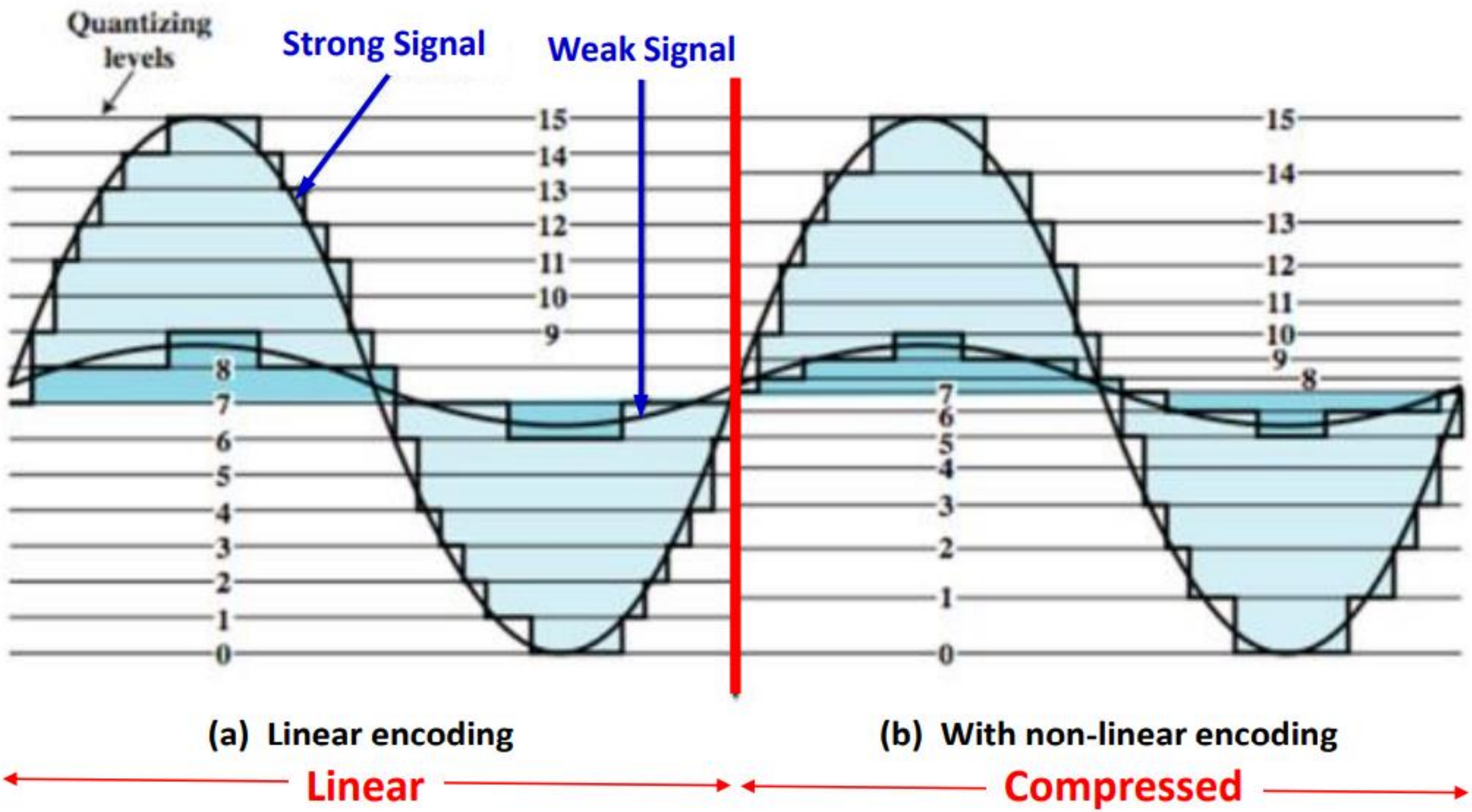
- The SNR depends on the number of quantization levels, and also on the signal power.
- In practice the signal power varies greatly
- For speech signal most of the time the amplitude is low, resulting in low (poor) SNR
- A common solution to this is to use *non-uniform* quantization steps
- this is equivalent to passing the message signal through a *compressor*



(a)



Dilemma of strong and weak signal



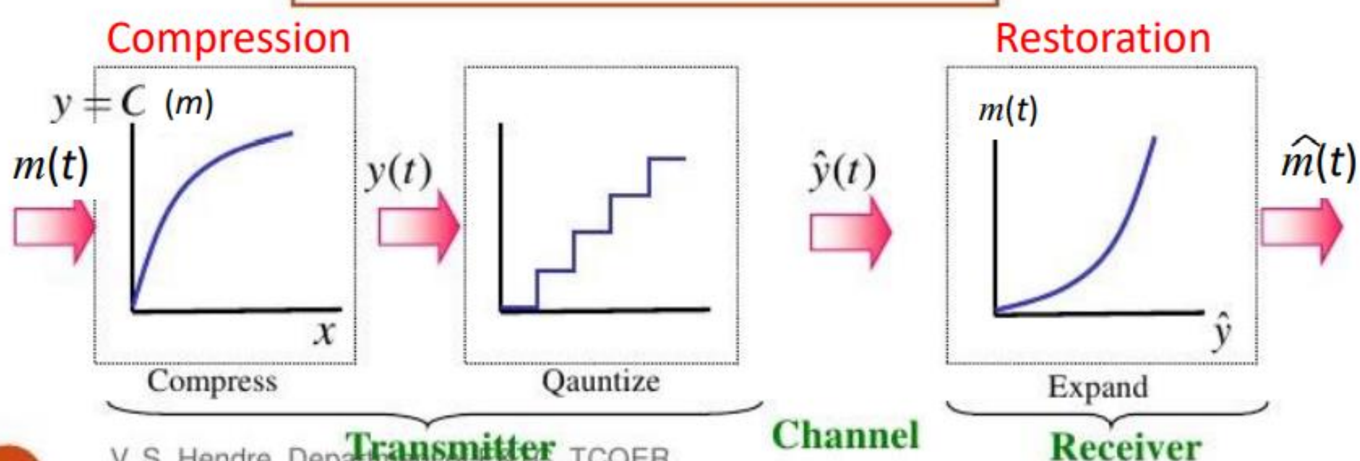
Note different encoding levels on each side.

How to solve? Use a "compandor"

Non-uniform quantization….process

- At the transmitter Uniformly quantizing the "compressed" signal.
- At the receiver, an inverse compression/expansion characteristic, called "expansion" is employed to avoid signal distortion.

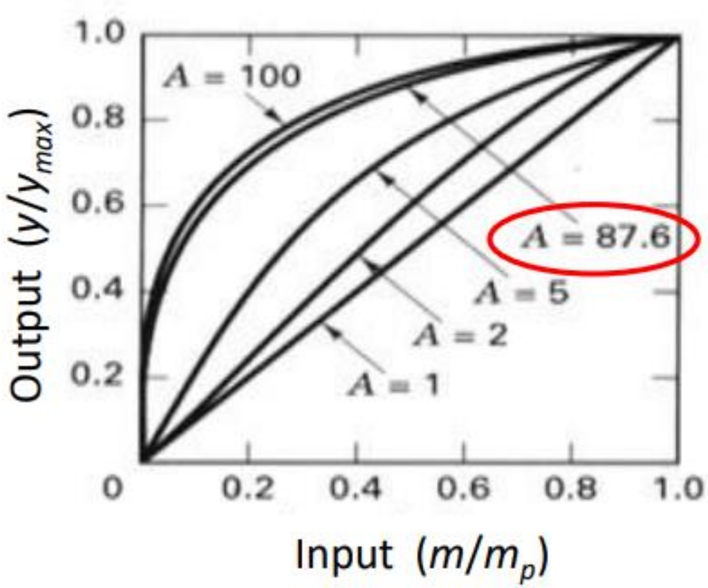
compression+expansion ⇨ companding



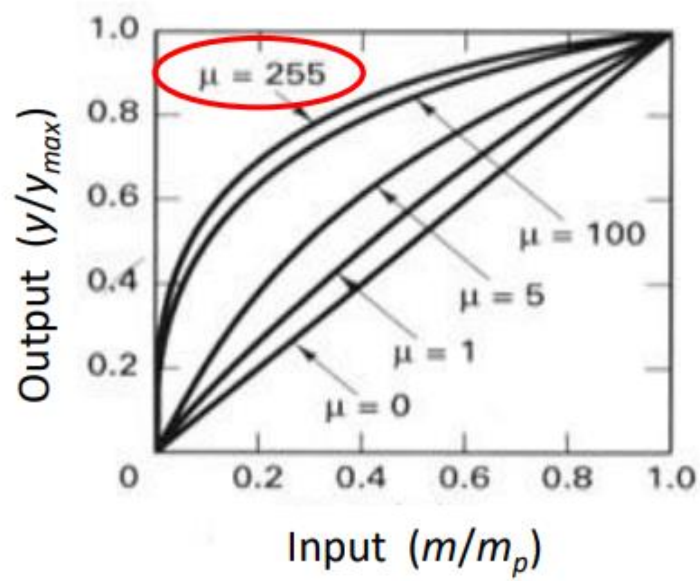
V. S. Hendre Department of E&TC, TCOER, Pune

Companding laws

A-Law Companding (Europe)



μ-Law Companding (North America)

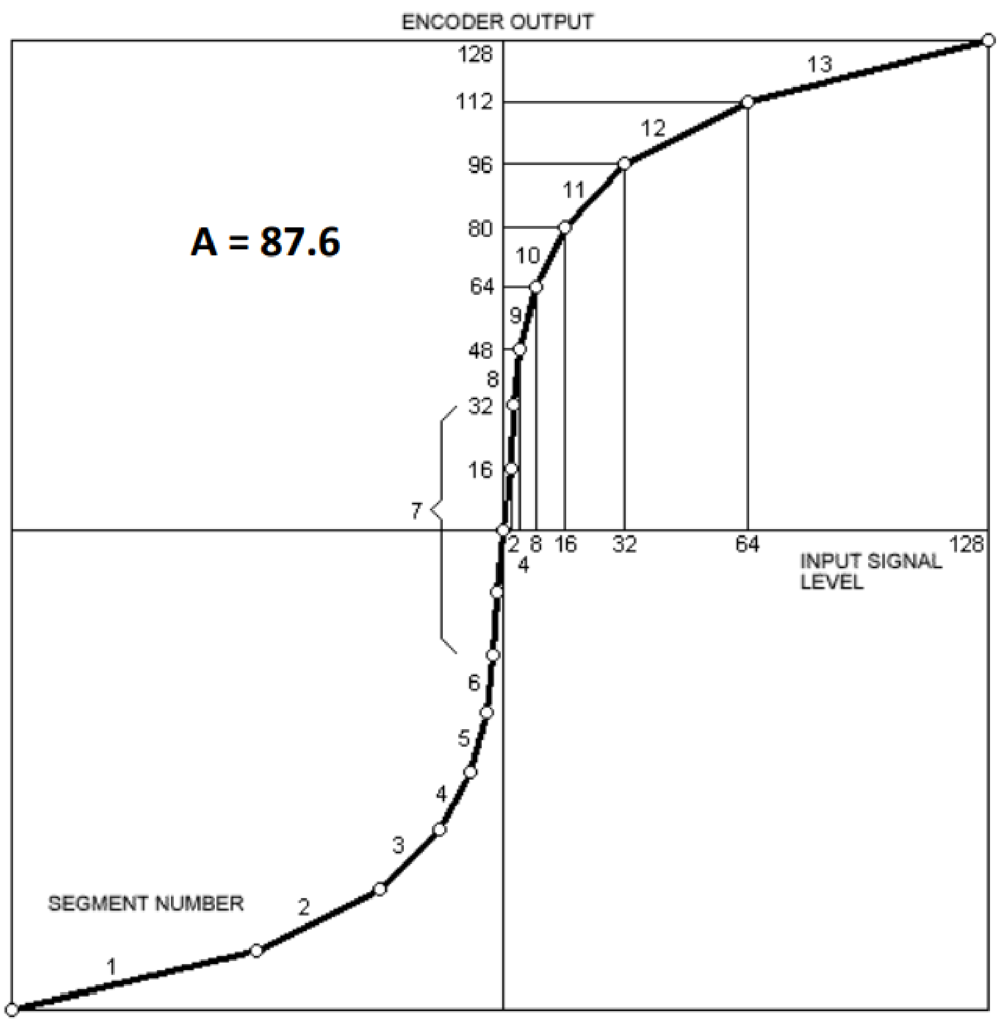


$$y = \frac{A}{1+\log_e A}\left(\frac{m}{m_p}\right) \quad for \quad 0 \le \frac{m}{m_p} \le \frac{1}{A}$$

$$y = \frac{A}{1+\log_e A}\left(1+\log_e\left(\frac{Am}{m_p}\right)\right) \quad for \quad \frac{1}{A} \le \frac{m}{m_p} \le 1$$

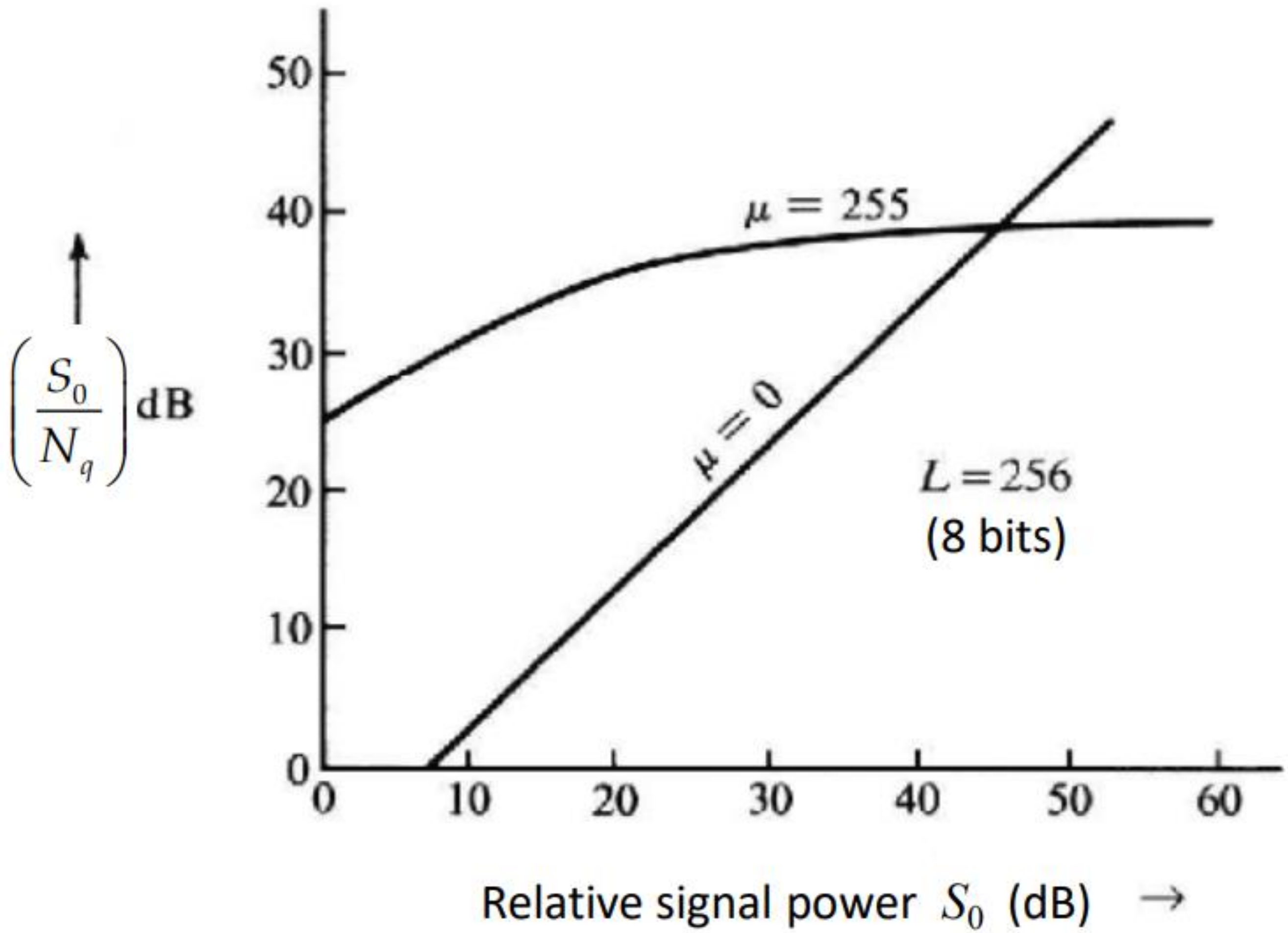
$$\boxed{y = \frac{1}{\log_e(1+\mu)}\log_e\left(1+\frac{\mu m}{m_p}\right) \quad for \quad 0 \le \frac{m}{m_p} \le 1}$$

A-law shown graphically



Flattening the SNR ratio

For optimal S_0/N_q ratio in North America μ = 255 is used.
An approximately constant S_0/N_q ratio is the most desirable.



Nonuniform Quantization

- For a μ-law compandor, the compression curve is is given by

$$y = \frac{\log(1 + \mu m / m_p)}{\log(1 + \mu)}$$

- the parameter μ dictates the amount of compression and is typically grater than 100.
- The resulting SNR is practically *independent* of the input power

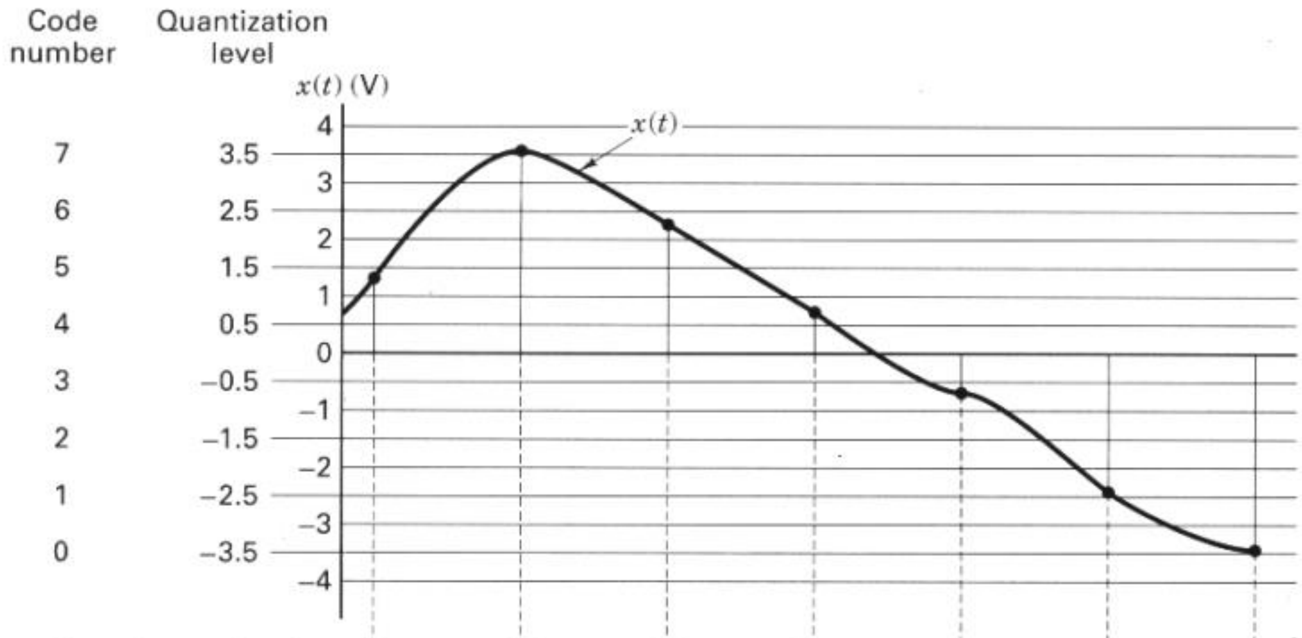
$$\frac{S_0}{N_0} \approx \frac{3L^2}{[\ln(1 + \mu)]^2}$$

Pulse-Code Modulation

- Pulse-Code Modulation (PCM) - transmission of a message using sequence of coded pulses representing the sampled and quantized signal
- at the transmitter
  1. **low-pass filtering**: used prior to sampling to prevent aliasing of the message
  2. **sampling**: sample the message at a rate of twice the message bandwidth
  3. **quantization**: use a quantizer (usually non-uniform) assign the sample values to discrete levels
  4. **encoding**: represent each discrete level in binary form
- PCM is obtained from quantized PAM signals, by *encoding* each quantized sample onto a digital *word*

Pulse-Code Modulation: example

- Consider an example of PCM of an analog voltage signal



Natural sample value	1.3	3.6	2.3	0.7	−0.7	−2.4	−3.4
Quantized sample value	1.5	3.5	2.5	0.5	−0.5	−2.5	−3.5
Code number	5	7	6	4	3	1	0
PCM sequence	101	111	110	100	011	001	000

PCM bandwidth

In binary PCM, we have a group of n bits corresponding to L levels with n bits. Thus,

$$L = 2^n \quad \text{or} \quad n = \log_2(L)$$

Signal $m(t)$ is band-limited to B Hz, which requires $2B$ samples per second.

For $2nB$ elements of information, we must transfer $2nB$ bits/second. Thus, the minimum bandwidth B_T needed to transmit $2nB$ bits/second is

$$B_T = nB \text{ Hz}$$

Practically speaking, we usually choose the transmission bandwidth to be a little higher than the minimum bandwidth required.

Exponential increase of the output SNR

$$\frac{S_0}{N_q} = 3\left(\frac{\langle m^2(t)\rangle}{m_p^2}\right)L^2$$

$\langle \cdots \rangle$ denotes time average

The number of levels L can be expressed as $L^2 = 2^{2n}$ where $n = \log_2(L)$ and is the number of bits to generate L levels. The SNR can now be expressed as

$$\frac{S_0}{N_q} = 3\left(\frac{\langle m^2(t)\rangle}{m_p^2}\right)(2)^{2n}$$

Using the expression for bandwidth, $B_T = nB$, then we arrive at

$$\frac{S_0}{N_q} = 3\left(\frac{\langle m^2(t)\rangle}{m_p^2}\right)(2)^{2B_T/B}$$

Taking the logarithm gives

$$\left(\frac{S_0}{N_q}\right)_{dB} = 10\log_{10}\left(\frac{S_0}{N_q}\right) = 10\log_{10}\left[3\left(\frac{\langle m^2(t)\rangle}{m_p^2}\right)\right] + 10\cdot 2n\log_{10}(2) \cong (\alpha + 6n)\,\mathrm{dB}$$

Example: SNR

Given a sinusoidal modulating signal $m(t)$ of amplitude A_m into a load resistance R = 1 ohm, find the signal-to-quantization noise ratio (sometimes called SNR):

$$P_{ave} = \frac{A_m^2}{2} \quad \text{and set} \quad m_{\max} = A_m$$

Setting $m_{max} = A_m$

$$\frac{S_0}{N_q} = 3\left(\frac{\langle m^2(t)\rangle}{m_p^2}\right)(2)^{2n} = \left(\frac{3P_{ave}}{m_{\max}^2}\right)(2)^{2n} = \left(\frac{3}{2}\right)(2)^{2n}$$

$$\boxed{10 \cdot \log_{10}\left(\frac{S_0}{N_q}\right) \cong (1.76 + 6n)\,\text{dB}}$$

L	n	SNR
32	5	31.8 dB
64	6	37.8 dB
128	7	43.8 dB
256	8	49.8 dB

Example: Bandwidth

Problem: A band-limited signal $m(t)$ of 3 kHz bandwidth is sampled at rate of 33⅓ % higher than the Nyquist rate. The maximum allowable error in the sample amplitude (*i.e.*, the maximum quantization error) is 0.5% of the peak amplitude m_p. Assume binary encoding. Find the minimum bandwidth of the channel to transmit the encoded binary signal.

Solution:

Solution:

The Nyquist rate is R_N = 2 x 3000 Hz = 6000 Hz (samples/second), but the actual rate is 33⅓ % higher, so the sample rate is 6000 Hz + (⅓ x 6000) = 8000 Hz.

The quantization step is Δ and the maximum quantization error is plus/minus Δ/2. Hence, we can write

$$\frac{\Delta}{2} = \frac{m_p}{L} = \frac{0.5}{100} m_p \quad \Rightarrow \quad L \geq 200$$

For binary coding, L, must be a power of two; therefore, knowing that L = 2^7 = 128 and 2^8 = 256, we must choose n = 8 to guarantee better than a 0.5% error.

Solution continue:

Having chosen n = 8 to guarantee < 0.5% error, to find the bandwidth required we start with

$$\text{Total number of bits per second } C = 8 \text{ bits} \times 8000 \text{ Hz}$$
$$= 64{,}000 \text{ bits/second}$$

However, we know we can transmit 2 bits/Hz of bandwidth[¶], so it requires a bandwidth B_T of

$$\boxed{B_T = C/2 = 32{,}000 \text{ Hz} = 32 \text{ kHz}}$$

If 24 such signals are multiplexed onto a single line (known as a **T1** Line in the Bell telephone system), then

$$C_{T1} = 24 \times 64 \text{ kb/s} = 1.536 \text{ Mb/s; so the bandwidth} = 768 \text{ kHz} \quad \Leftarrow$$

[¶]A maximum of $2B$ independent elements of information per second can be transmitted, error-free, over a noiseless channel of bandwidth B Hz.

Example: SQNR

We are given a signal $m(t) = 2 \cdot \cos(2\pi \cdot 250\, t)$ as a single-tone signal input.

(a) Find the SQNR with 8-bit PCM.

Solution:

For 8-bit encoding, L = 2^n where n = 8, therefore, the number of levels = 256. The amplitude A_m of the sinusoidal waveform means that m_p = 2 volts. The total signal swing possible (- m_p to + m_p) will be $2m_p$ = 4 volts, therefore, the average signal power is P_{ave} = $[(A_m)^2/2]$ = $[2^2/2]$ = 2 watts.

The interval Δ = $[2m_p/L]$ = 4 volts/256 levels = 1.5625×10^{-2} volt.

Now we can find the SQNR (signal-to-quantized noise ratio)

Using for the quantization noise N_q = $[\Delta^2/12]$, and taking P_{ave} = 2 W, the SNR is given by

$$\left(\frac{S_O}{N_q}\right) = \left(\frac{P_{ave}}{N_q}\right) = \left(12\frac{2}{\Delta^2}\right) = \left(\frac{24}{(1.5625\times 10^{-2})^2}\right) = 98,304$$

$$SNR_{dB} = 10\cdot\log_{10}(98,304) = 49.93 \text{ dB} \quad \Leftarrow$$

Example: Number of bits needed

We are given a signal $m(t)$ = 2·cos(2π·250 t) as the signal input.

(b) If the minimum SNR is to be at least 36 dB, how many bits n are needed to encode the signal (*i.e.*, find n)? Other parameters such as signal power remain the same as in part (a) on previous slide.

Solution:

Note that 36 dB is numerically equivalent to 3,981.
Remembering that the interval is $\Delta = [2m_p/L]$ and $2m_p = 4$ volts.

$$3{,}981 = \frac{6\,m_p^2}{\Delta^2} = \frac{24}{\Delta^2}, \quad \Delta^2 = \frac{24}{3981} = 0.006029 \quad \text{and} \quad \Delta = 0.07764 \text{ volt}$$

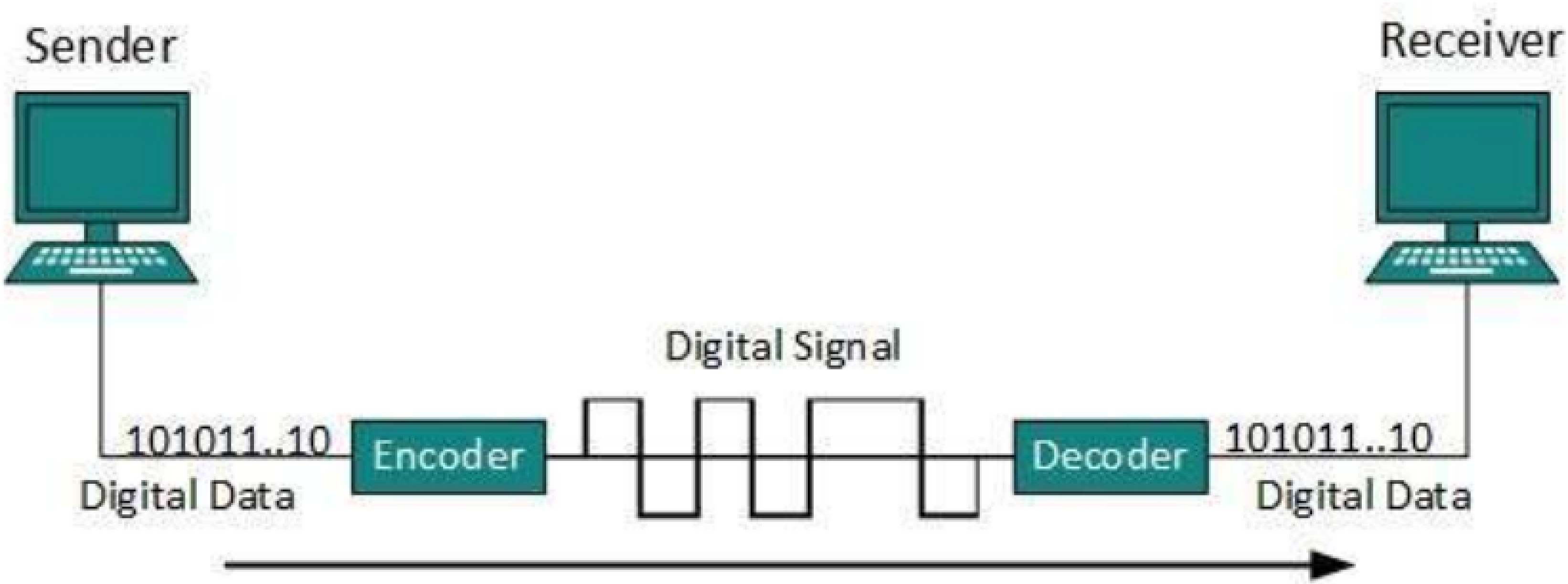
Therefore, we can determine the number of levels L and then find n.

$$L = \frac{2m_p}{\Delta} = \frac{4}{0.07764} = 51.5$$

The lowest integer number of bits n that will give at least 51.5 levels is $n = 6$ because $2^6 = 64$ levels. **So the answer is 6 bits.**

A look at Line Code

A **line code** is an assignment of a symbol or pulse to each zero or one to be transmitted. **Digital data is represented by baseband data formats known as line codes.**



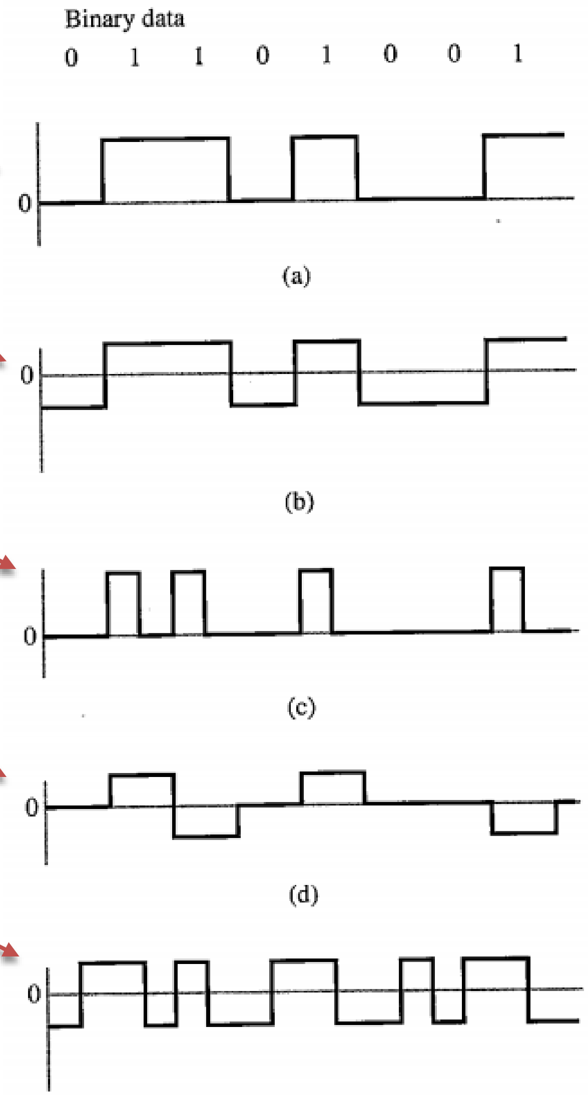
In other words, what shape will the waveform take?

Line Codes

- Strictly speaking, PCM as just described is a technique of digital representation of analog information
- PCM produces a sequence of bits - but bits themselves are not suitable for transmission over a physical medium; we must represent the binary stream as electrical pulses: this is the purpose of *line coding*
- There is a variety of linecodes, and the choice will depend on the design criteria. Some parameters/characteristics that differentiate the various PCM codes:
    - absence or presence of DC component
    - noise immunity
    - bandwidth characteristics
    - error detection capabilities
    - synchronisation requirements

Line Codes

1. **Unipolar (on-off) Nonreturn-to-Zero (NRZ)**: 1 represented by a pulse of amplitude A, 0 by absence of a pulse
2. **Polar NRZ**: 1 and 0 represented by pulses of amplitudes A and $-A$ respectively
3. **Unipolar (on-off) Return-to-Zero (RZ)**: $1 =$ rectangular pulse of amplitude A for half the symbol duration; $0 =$ no pulse
4. **Bipolar** NRZ positive and negative pulse of magnitude A use alternately to represent 1's; $0 =$ no pulse
5. **Split Phase / Manchester Code**: 1 represented by a pulse of amplitude A followed by a pulse of amplitude $-A$, both of half symbol duration; 0 represented the same way but with the pulses reversed



(we will come back to line coding again in Lec 6)

PCM Improvements

- PCM inefficient- high data rate required
- Improvements exploit the characteristics of the source signal
- Analog messages contain a lot of redundancy in their sample values - in other words, each sample is not independent.
- Exploiting this redundancy significantly reduces the bit rate required to transmit the message.
  - Differential PCM (DPCM)
  - Adaptive DPCM (ADPCM)
  - (Sigma-)Delta Modulation

Differential Pulse Code Modulation

- Rather then transmitting sample value k, transmit the difference between successive samples

$$d[k] = m[k] - m[k-1]$$

- The receiver, having previously reconstructed previous samples, simply adds $d[k]$ to obtain the current value
- Key: the difference between successive samples is generally much smaller than the samples themselves
- Thus, the peak amplitude of the transmitted signal m_p is reduced significantly, consequently reducing the quantization noise (or required bit rate for a fixed SNR)

Differential Pulse Code Modulation

- The scheme can be further improved by the system estimating/predicting the kth sample value $\hat{m}[k]$ from several previous samples
- The transmitter sends the *difference between the estimate and the actual value*

$$d[k] = m[k] - \hat{m}[k]$$

- The receiver, used the same method for the predicted sample $\hat{m}[k]$, and uses $d[k]$ to obtain the true value $m[k]$
- The difference between $m[k]$ and $\hat{m}[k]$ is even smaller than between $m[k]$ and $m[k-1]$, thus further improving the noise performance

Differential Pulse Code Modulation

- At the heart of the prediction techniques is the fact that a function value can be expressed as an infinite sum of derivatives of that function at a (different) point.
- Taylor series:

$$\begin{aligned} m(t+T_s) &= m(t) + T_s\dot{m}(t) + \frac{T_s^2}{2!}\ddot{m}(t) + \frac{T_s^3}{3!}\dddot{m}(t) + \dots \\ &\approx m(t) + T_s\dot{m}t \quad \textbf{for small } T_s \end{aligned}$$

- Letting $m[k] = m(kT_s)$, and noting that $\dot{m}(t) \approx \frac{m[k]-m[k-1]}{T_s}$

$$\begin{aligned} m[k+1] &\approx m[k] + T_s\left[\frac{m[k]-m[k-1]}{T_s}\right] \\ &= 2m[k] - m[k-1] \end{aligned}$$

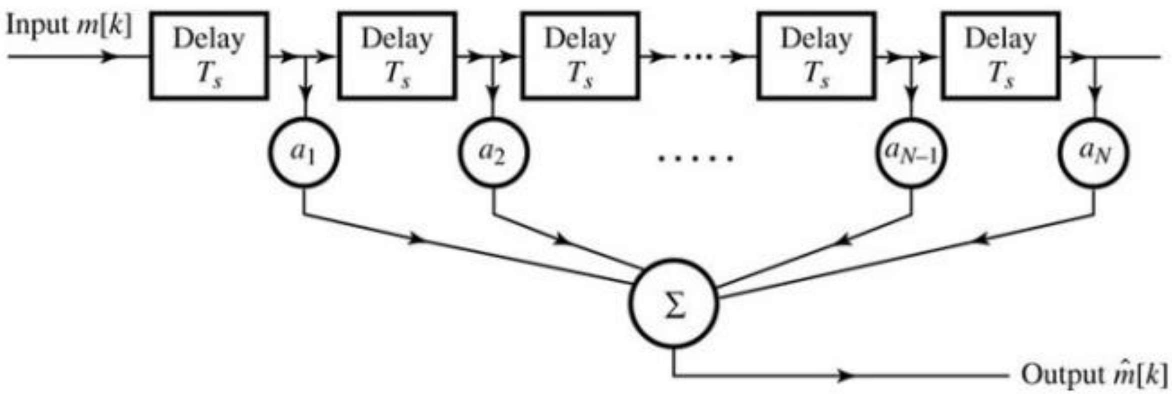
- This is a crude prediction of sample $k+1$ from the two previous samples.
- The approximation can be improved using higher order derivatives and thus more past samples. General prediction formula:

$$\begin{aligned} m[k] &\approx a_1m[k-1] + a_2m[k-2] + \ldots + a_Nm[k-N] \\ \hat{m}[k] &= a_1m[k-1] + a_2m[k-2] + \ldots + a_Nm[k-N] \end{aligned}$$

Differential Pulse Code Modulation

$$\hat{m}[k] = a_1 m[k-1] + a_2 m[k-2] + \ldots + a_N m[k-N]$$

- The above is an expression for the Nth order predictor
- a_j are referred to as *prediction coefficients*
- This *linear predictor* is a *transversal filter* or a *tapped delay line*, with tap gains equal to the prediction coefficients:



- The above is a simple prediction technique. A more sophisticated one uses prediction coefficients based on minimum mean squared error criterion (more on this at the end of the course)

Differential Pulse Code Modulation

- Minor complication: the receiver operates on the *quantized* values of past samples $m[k-1], m[k-2], \ldots$, ie $m_q[k-1], m_q[k-2], \ldots$.
- Thus, it can only calculate $\hat{m}_q[k]$, that is the estimate of the quantized sample $m_q[k]$
- To reduce the error in reconstruction, the transmitter instead should use the difference between the true sample and the *quantized* value of the previous one, ie

$$d[k] = m[k] - \hat{m}_q[k]$$

- This value must be quantized (and transmitted) giving

$$d_q[k] = d[k] + q[k]$$

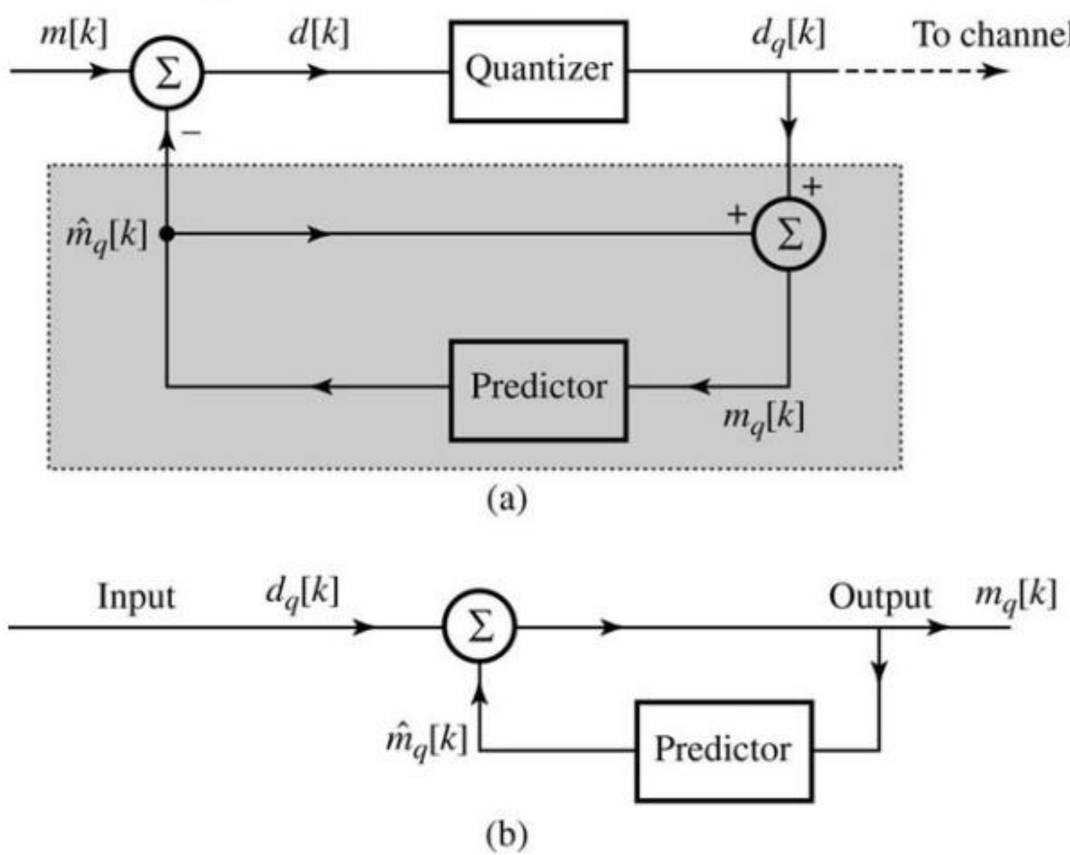
where $q[k]$ is the quantization error.

Differential Pulse Code Modulation

- The predictor output is $\hat{m}_q[k]$, is fed back to the predictor, giving an input of

$$m_q[k] = \hat{m}_q[k] + d_q[k] = m[k] - d[k] + d_q[k] = m[k] + q[k]$$

- This shows that $m_q[k]$ is indeed the quantized version of $m[k]$



DPCM SNR improvement

How much better is DPCM with regard to SNR?

To determine this, define m_p and d_p as the peak amplitudes of $m(t)$ and $d(t)$, respectively. Assuming the same number of steps L for both, then the quantization step Δ in DPCM is reduced in magnitude by d_p/m_p.

The quantization noise is proportional to $(\Delta)^2$ – the quantization noise power is reduced by a factor $(d_p/m_p)^2$ and the SNR is therefore increased by $(m_p/d_p)^2$.

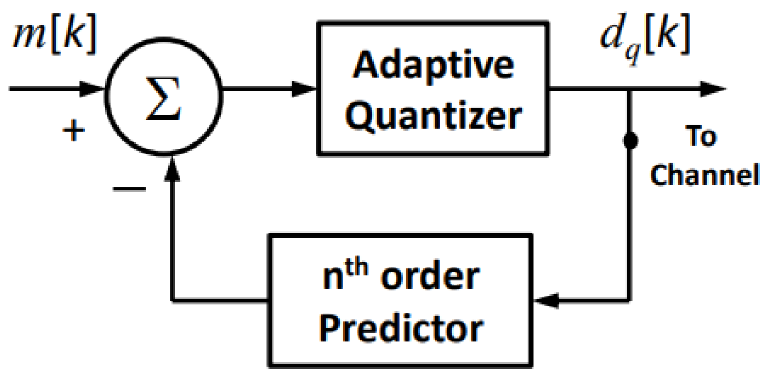
Maintaining the same SNR, the number of bits can be reduced.

<u>Example</u>:
The AT&T telephone system sometimes operates at 32 kbits/s (or even 24 kbits/s) when using DPCM. [The telephone system was initially designed to use a 64 kbits/second data rate.]

Adaptive DPCM

Adaptive differential PCM (**ADPCM**) can further improve upon DPCM by Incorporating an adaptive quantizer (variable Δ) at encoding.

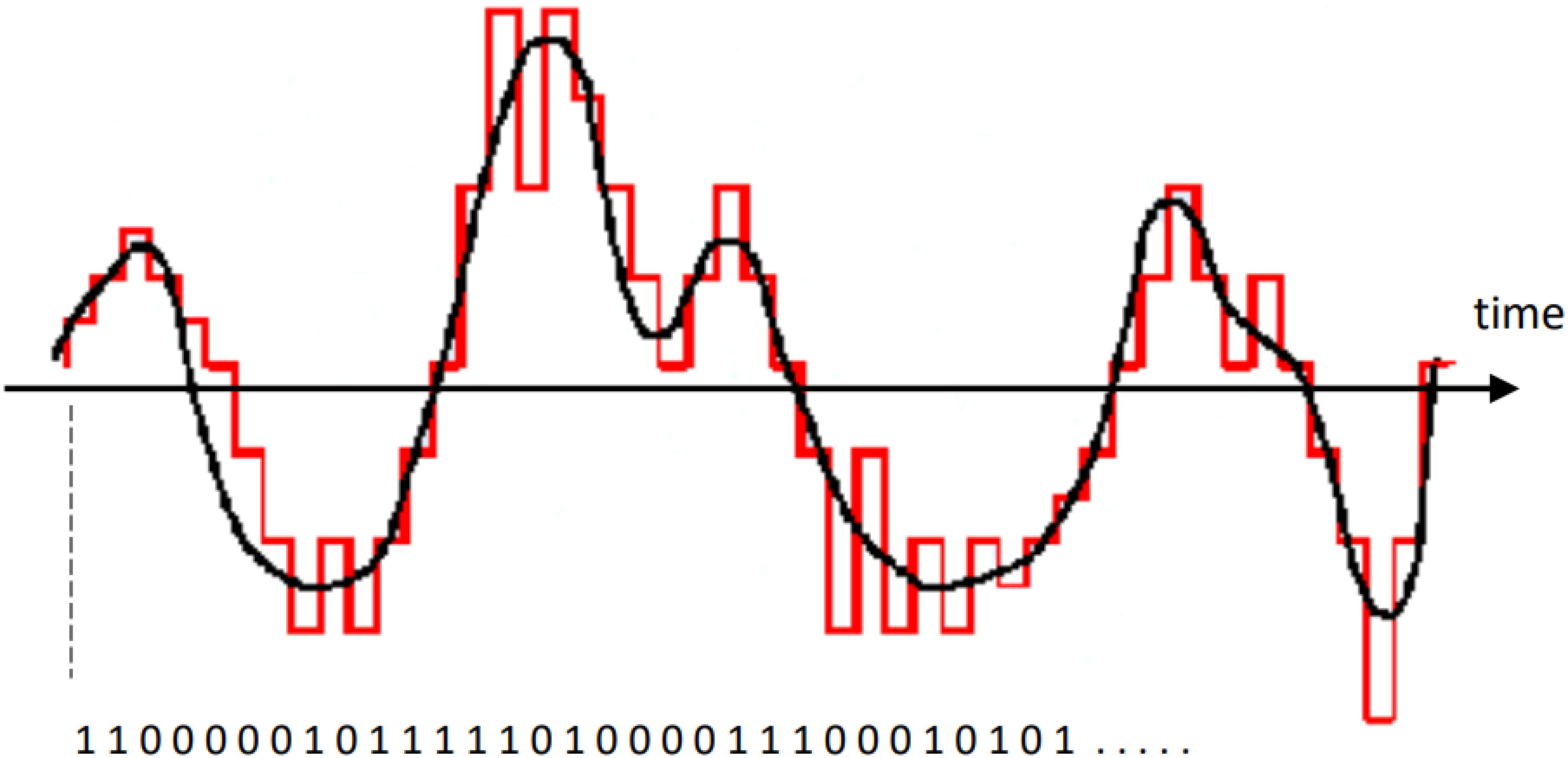
The quantized prediction error $d_q[k]$ is a good measure of the predicted error size – it can be used to change Δ which minimizes $d_q[k]$. When $d_q[k]$ fluctuates around large positive or negative values, the prediction error is large and Δ needs to increase, but when $d_q[k]$ fluctuates around zero (small values), then Δ needs to decrease.



After Lathi and Ding, 4th edition, 2009; page 345.

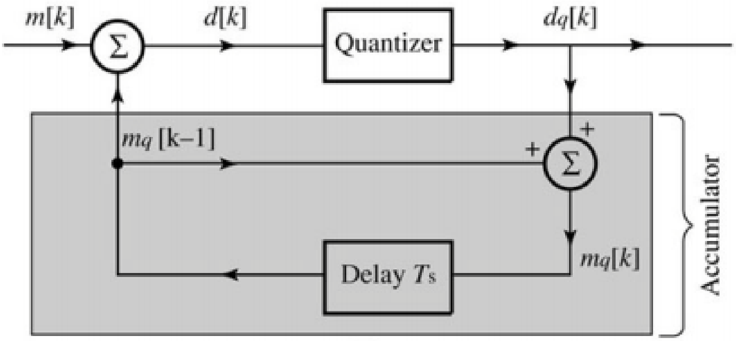
<u>Example</u>: An 8-bit PCM sequence can be encoded into a 4-bit ADPCM sequence at the same sampling rate. This reduces the channel bandwidth by one-half with no loss in quality.

Example of adaptive DPCM waveform

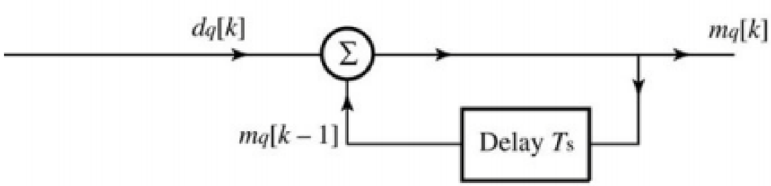


Delta Modulation

- Delta Modulation is a special case of DPCM
- DM exploits the correlation between samples even further by oversampling the signal which results in very small prediction errors
- Such values of $d_q[k]$ can be transmitted using one bit only, ie $L = 2$ quantization levels. Thus DM is a 1-bit DPCM.
- DM uses a first order predictor, such that $m_q[k] = m_q[k-1] + d_q[k]$



(a)



(b)

Delta Modulation

- Note that we have

$$m_q[k] = m_q[k-1] + d_q[k]$$

and similarly

$$m_q[k-1] = m_q[k-2] + d_q[k-1]$$

which upon substitution gives

$$m_q[k] = m_q[k-2] + d_q[k] + d_q[k-1]$$

which, upon repeating iteratively, can be compactly written as

$$m_q[k] = \sum_{m=0}^{k} d_q[m]$$

- Thus, the DM receiver is simply an *accumulator* (or *adder*)

Key attributes of DM

1. Delta modulation (DM) is the **simplest method** for analog-to-digital conversion (ADC).

2. Delta modulation uses **1-bit** per sampling period (T_S) – it is a 1-bit ADC

3. Delta modulation requires a **sampling rate much greater than the Nyquist rate** (commonly four or more times the Nyquist rate).

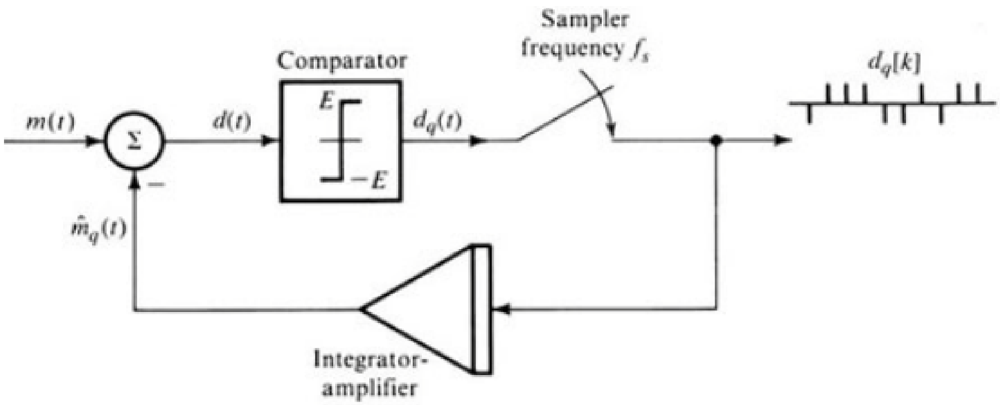
4. DM is closely related to differential PCM.

5. In DM we use a first-order predictor (one time delay T_S is the predictor).

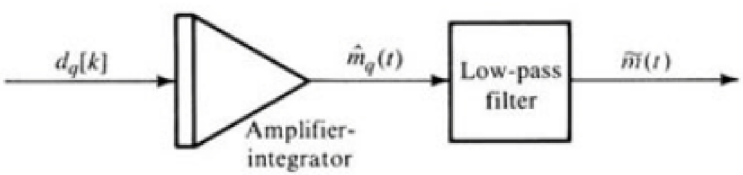
6. DM uses very simple hardware and is very low cost for that reason.

7. The transmitted output is a binary stream of $\pm\Delta$ pulses at f_S. It gives a stepwise approximation $m_q(t)$ to the message signal $m(t)$.

Practical implementation of DM

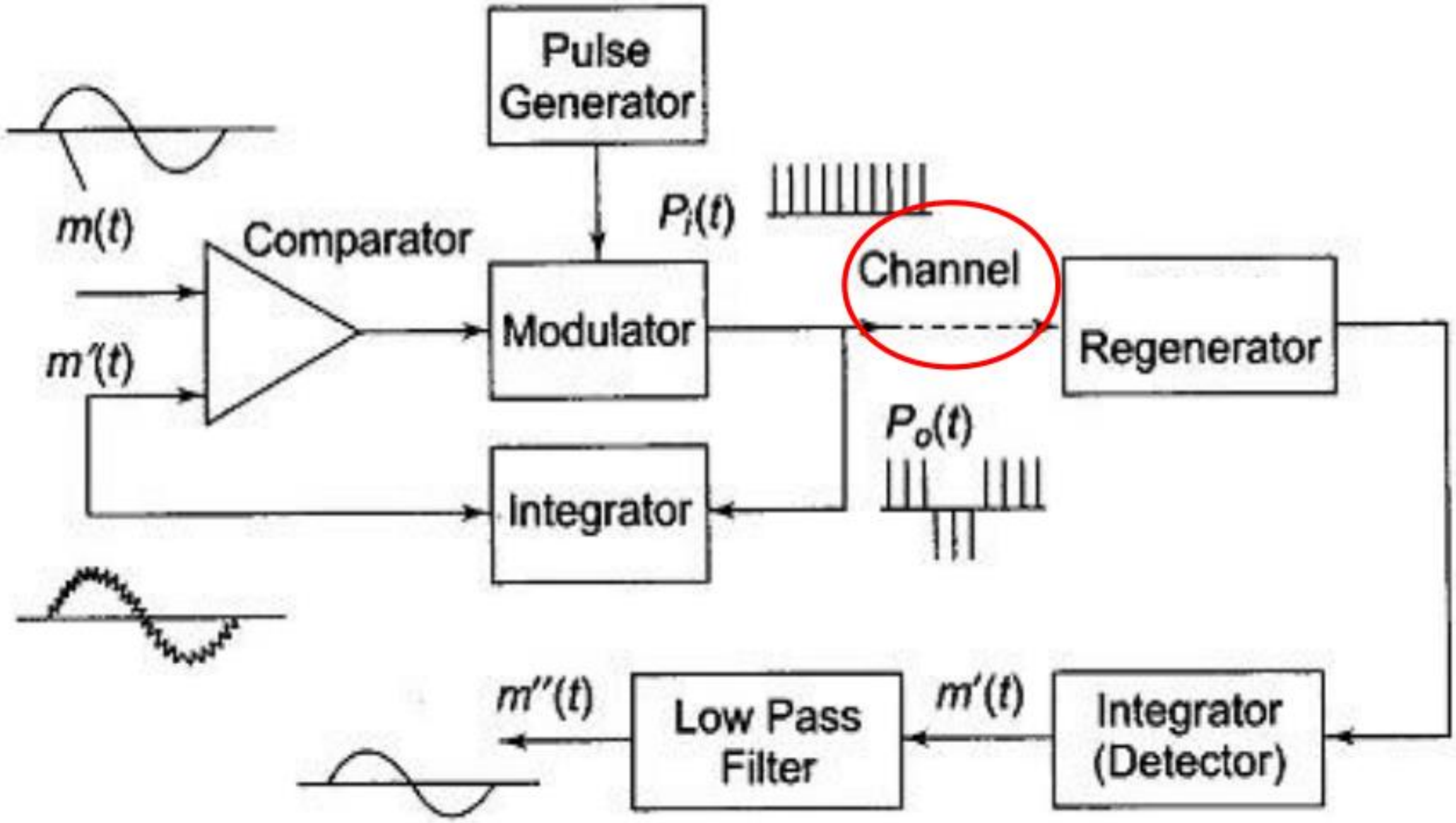


(a)



(b)

A DM transmission system



Example: PCM vs DM

Problem:

A one kilohertz (1 kHz) signal $m(t)$ is sampled at 8 kHz with 12-bit encoding for PCM transmission.

(a) How many bits are transmitted per second in in PCM? What is the bandwidth required in this case?

(b) Now switch to using DM with 8 kHz sampling. How many bits are transmitted per second using DM? What is the bandwidth required in using DM?

Solution:

Solution:

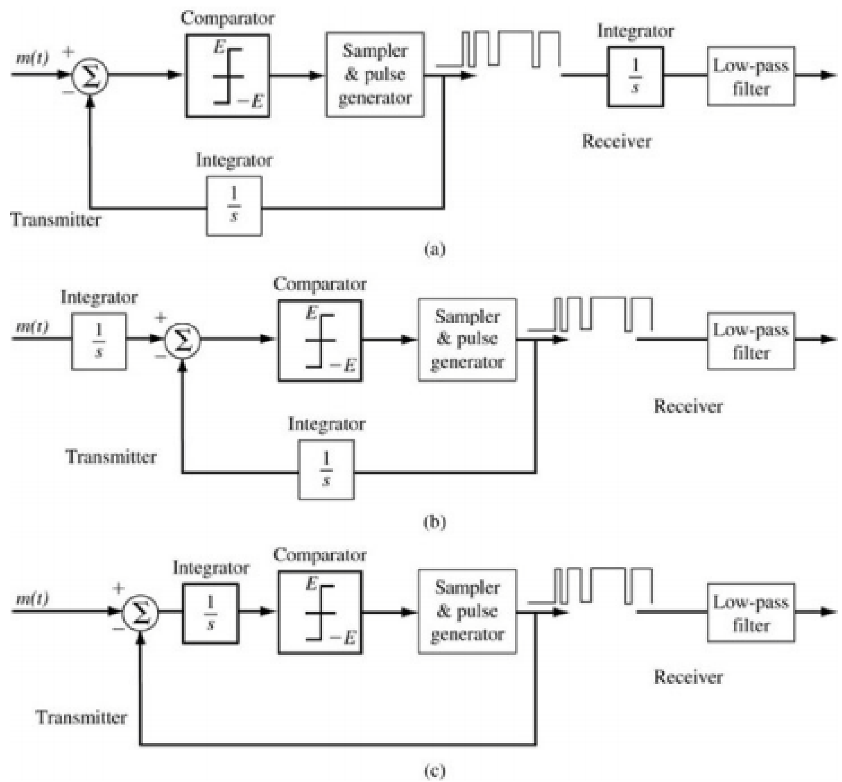
The signal frequency is f_m = 1 kHz and the sampling rate is 8 kHz.

(a) For PCM we have 8,000 samples per second and 12 bits per sample; which equals 96,000 bits/second. The bandwidth is one-half of this giving 48,000 Hz.

(b) Now for DM we have 1 bit per sample at 8,000 samples per second. Thus, we have 8,000 bits per second and a bandwidth of 4,000 Hz.

Sigma-Delta Modulation

- DM essentially transmits the derivative of the signal; at the receiver, this requires an integrator (adder).
- This will result in the noise being accumulated - major drawback of DM
- Solution: move the integrator from the receiver to the transmitter



Summary table

S.NO	Parameter of Comparison	Pulse Code Modulation (PCM)	Delta Modulation (DM)	Adaptive Delta Modulation (ADM)	Differential Pulse Code Modulation (DPCM)
1.	Number of bits	It can use 4,8, or 16 bits per sample.	It uses only one bit for one sample	It uses only one bit for one sample	Bits can be more than one but are less than PCM.
2.	Levels and step size	The number of levels depends on number of bits. Level size is fixed.	Step size is kept fixed and cannot be varied.	According to the signal variation, step size varies.	Number of levels is fixed.
3.	Quantization error and distortion	Quantization error depends on number of levels used.	Slope overload distortion and granular noise are present.	Quantization noise is present but other errors are absent.	Slope overload distortion and quantization noise is present.
4.	Transmission bandwidth	Highest bandwidth is required since numbers of bits are high.	Lowest bandwidth is required.	Lowest bandwidth is required.	Bandwidth required is less than PCM.
5.	Feedback	There is no feedback in transmitter or receiver.	Feedback exists in transmitter.	Feedback exists.	Feedback exists.
6.	Complexity of Implementation	System is complex.	Simple	Simple	Simple

… another table

Parameters	PCM	DPCM	DM
Number of bits/sample	4, 8, 16, and so on.	Typically 4 to 6 bits < PCM	1 bit
Bandwidth	Highest required	Lower than PCM	Lowest
Step Size	Fixed	Fixed	Fixed
Sampling Rate	8 kHz	8 kHz	64 to 128 kHz
Bit Rate	56 and 64 kbps	32 to 48 kbps	64 to 128 kbps
Quantization Error	Depends upon q	Slope overload and granular noise	Slope overload and granular noise