UNIT II – PULSE MODULATION

Introduction

In the simplest model of a telephone speech communication there is a direct, dedicated, physical connection between the two participants in the conversation, and this link is held for the duration of the conversation. The analogue electrical signal produced by the telephone at either end is sent on to connection without modification.

In Pulse Amplitude Modulation (PAM), the unmodified electrical signal is not sent on to the connection. Instead, short samples of the signal are taken at regular intervals, and these samples are sent on to the connection. The amplitude of each sample is identical to the signal voltage at the time when the sample was taken. Typically, 8,000 samples are taken per second, so that the interval between samples is 125μs, and the duration of each sample is approximately 4μs.

Because each sample is very short (~4μs) there is a lot of time between samples (~121μs). Samples from other conversations are put into this “spare time”. Usually, the samples from 32 separate conversations are put on to a single line. This process is called Time Division Multiplexing (TDM).

Each sample is very short and will be distorted as it travels across a communications network. In order to reconstruct the original analogue, signal the only information the receiver needs to have about a sample is its amplitude, but if this is distorted then all information about the sample has been lost. To overcome this problem, the pulse is not transmitted directly, instead its amplitude is measured and converted into an 8 binary number - a sequence of 1s and 0s. At the receiver end, the receiver merely needs to detect if a 1 or a 0 has been received so that it can still recover the amplitude of a PAM pulse even if the 1s and 0s used to describe it have been distorted.

The process of converting the amplitude of each pulse into a stream of 1s and 0s is called Pulse Code Modulation (PCM)

Note that the process of PAM and PCM (but without the use of TDM) is essentially used to store music and speech on CDs, but with a higher sample rate, more bits per sample and complex error correction mechanisms.

Some terms are:

Sampling : The process of measuring the amplitude of a continuous-time signal at discrete instants. It converts a continuous-time signal to a discrete-time signal.

Quantizing : Representing the sampled values of the amplitude by a finite set of levels. It converts a continuous-amplitude sample to a discrete-amplitude sample.

Encoding : Designating each quantized level by a (binary) code.

Sampling and quantizing operations transform an analogue signal to a digital signal. Use of quantizing and encoding distinguishes PCM from analogue pulse modulation methods.

The quantizing and encoding operations are usually performed in the same circuit at the transmitter, which is called an Analogue to Digital Converter (ADC). At the receiver end the decoding operation converts the (8 bit) binary representation of the pulse back into an analogue voltage in a Digital to Analogue Converter (DAC)

Low Pass Sampling:

Consider a band-limited signal with no frequency components above a certain frequency f_m. The sampling theorem states that this signal can be recovered completely from a set of samples of its amplitude, if the samples are taken at the rate of $\mathbf{f_s > 2f_m}$ samples per second.

This is often called the uniform sampling theorem for baseband or low-pass signals.

The minimum sampling rate, $2f_m$ samples per second, is called the Nyquist sampling rate (or Nyquist frequency); its reciprocal $1/(2f_m)$ (measured in seconds) is called the Nyquist interval.

$f_s = 2 * f_m$ is called the Nyquist sampling rate.

For telephone speech the standard sampling rate is 8 kHz (or one sample every 125 μs).

Sampling Methods

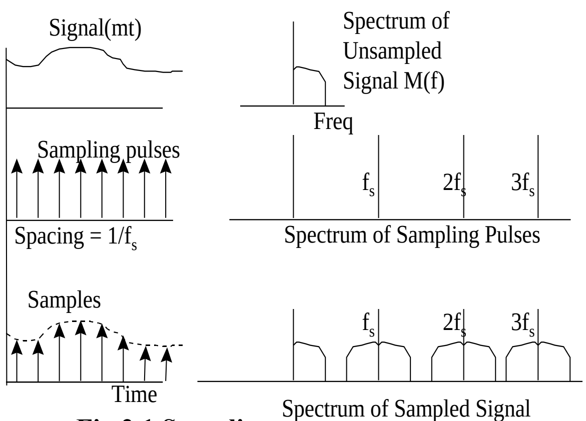


Fig 2.1 Sampling

Suppose we have an arbitrary signal (the **baseband signal m(t)) which has a spectrum M(f).** Take infinitesimally short samples of the signal m(t) at a uniform rate once every t_s seconds i.e. at a frequency f_s. This is the ideal form of sampling, it is called instantaneous (or impulse) sampling.

In effect the signal m(t) is multiplied by a train of impulses giving rise to a train of pulses as in the lower line of the diagram. The train of sampling impulses has a frequency spectrum consisting of all harmonics or multiples of f_s and all are at the same amplitude.

This **sampled signal has a spectrum as shown where M(f) is repeated unattenuated periodically and appears around all multiples of the sampling frequency ($f_s = 1/t_s$)**.

To recover m(t) from the sampled signal we need only pass the sampled signal through a low pass filter with a stop frequency of $f_s/2$. All of the higher frequency components will be dropped. In the diagram, if f_s is greater than twice the highest frequency in m(t) the repetitions of the sampled spectra around the harmonics of the f_s do not overlap.

Flat - top Sampling

An Analogue to Digital Converter requires that the sample value be held constant for a fixed time until the conversion is completed. This requires a flat-top sampled signal. This has approximately the same repeated frequency spectrum as with the instantaneous sampling above, but with each repetition slightly spread out.

The simplest and most common sampling method is performed by a functional block termed a Sample and Hold (S/H) circuit.

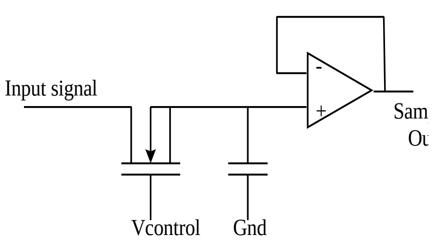


Fig 2.2 Sample and Hold

The output from the circuit must be held at a constant level for the sampling duration. V control switches the MOSFET ON until the charge on C is equal to the amplitude of the sampled voltage. V control then goes LOW, the MOSFET is OFF and the charge is held by the capacitor. The charge held on the capacitor puts a voltage across the capacitor, and it is held at that value until the next time that V control switches the MOSFET ON. This is called a sample and hold circuit and is usually used as the input to an ADC.

Aliasing Error

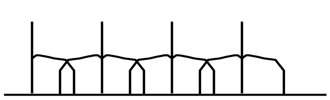


Fig 2.3 Spectrum of sampled signal

If a signal is under sampled (sampled at a rate below the Nyquist rate), the spectrum consists of overlapping repetitions of the sampled spectrum. Because of the overlapping tails a single repetition of the spectrum no longer has the complete information about the unsampled signal, and it is no longer possible to recover it from the sampled signal. To recover the original signal at the receiving end the sampled signal is passed through a lowpass filter with a cut off of $f_s/2$, we get a spectrum that is not the sampled signal but is a different version due to:

- Loss of the tail of the sampled signal spectrum beyond $f_s/2$
- This same tail appears inverted, or folded, onto the spectrum at the cut-off frequency.

This tail inversion is known as aliasing, (or spectral folding or foldover distortion).

The aliasing distortion can be eliminated by cutting the tail (i.e. filtering) of the sampled signal beyond $f > f_s/2$ before the signal is sampled. By so doing, the overlap of successive cycles in the sampled signal is avoided. The only error in the recovery of the unsampled signal is that caused by the missing tail above $f_s/2$.

It is simpler to consider aliasing by considering a single frequency component of m(t). We will look at the frequency f_m and it is sampled at a rate f_s. The diagrams show the frequencies which will be present in the sampled signal. There will be frequency components at f_m, $f_s - f_m$, $f_s + f_m$, $2 f_s - f_m$, $2 f_s + f_m$, $3 f_s - f_m$, $3 f_s + f_m$, etc. etc.

In the first case f_m is very much less than f_s, so that $f_s - f_m$ is much higher than the cut off of the filter ($f_s/2$).

In the second case f_m is below, but close to $f_s/2$, so that a sharp cut off filter is required to ensure that f_m is passed but $f_s - f_m$ is stopped.

In the third case f_m is higher than $f_s/2$, so that $f_s - f_m$ is less than $f_s/2$. The low pass filter with a cutoff of $f_s/2$ will therefore block f_m (the actual signal frequency) but will pass a signal with frequency $f_s - f_m$.

This is **aliasing**

Strictly speaking, a band limited signal does not exist in reality. It can be shown that if a signal is time limited it cannot be band limited. All physical signals are necessarily time limited because they begin at some finite instant and must terminate at some other finite instant. Hence, all practical signals are theoretically non band limited.

A real signal contains a finite amount of energy, therefore its frequency spectrum must decay at higher frequencies. Most of the signal energy resides in a finite band, and the spectrum at higher frequencies contributes little. The error introduced by cutting off the tail beyond a certain frequency B can be made negligible by making B sufficiently large.

Thus, for all practical purposes a signal can be considered to be essentially band limited at some value B, the choice of which depends upon the accuracy desired. A practical example of this is a speech signal. Theoretically, a speech signal, being a finite time signal, has an infinite bandwidth. But frequency components beyond 3400 Hz contribute a small fraction of the total energy. When speech signals are transmitted by PCM they are first passed through a lowpass filter of bandwidth of 3500 Hz. (This filter is called an **anti aliasing filter**). Higher sampling rates (i.e. 8000 samples/sec) permits recovery of the signal from its samples using relatively simple filters i.e. it allows for guard bands between the repetitions of the spectrum (otherwise recovering signals sampled at the Nyquist rate would require very sharp cut-off (ideal) filters).

In summary, aliasing distortion produces frequency components in the desired frequency band that did not exist in the original waveform. Aliasing problems are not confined to speech digitisation processes. The potential for aliasing is present in any sample data system.

Motion picture taking, for example, is another sampling system that can produce aliasing. A common example occurs when filming a rotating wheel. Often the sampling process (the picture refresh rate) is too slow to keep up with the wheel movements and spurious rotational rates are produced. If the wheel rotates 355^0 between frames, it looks to the eye as if it has moved backwards 5^0.

Sampling:

A message signal may originate from a digital or analog source. If the message signal is analog in nature, then it has to be converted into digital form before it can transmit by digital means. The process by which the continuous-time signal is converted into a discrete – time signal is called Sampling. Sampling operation is performed in accordance with the sampling theorem.

Sampling Theorem For Low-Pass Signals:-

Statement:- "If a band –limited signal g(t) contains no frequency components for $|f| > W$, then it is completely described by instantaneous values g(kT) uniformly spaced in time with period $T \leq 1/2W$. If the sampling rate, fs is equal to the Nyquist rate or greater ($fs \geq 2W$), the signal g(t) can be exactly reconstructed.

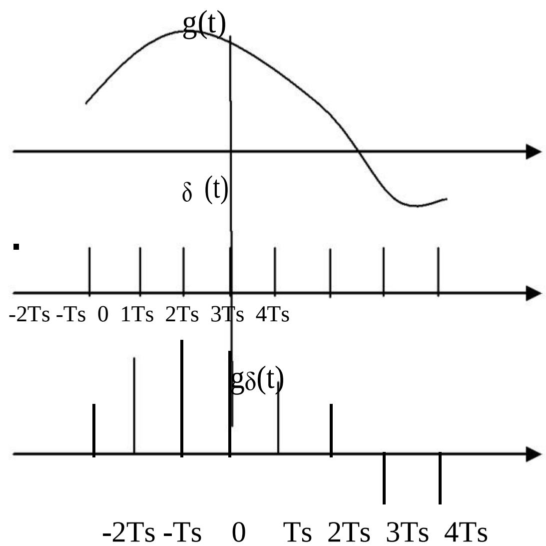


Fig 2.4 Sampling Process

Proof:-

Part - I If a signal x(t) does not contain any frequency component beyond W Hz, then the signal is completely described by its instantaneous uniform samples with sampling interval (or period) of $Ts < 1/(2W)$ sec.

Part – II The signal x(t) can be accurately reconstructed (recovered) from the set of uniform instantaneous samples by passing the samples sequentially through an ideal (brick-wall) lowpass filter with bandwidth B, where $W \leq B < fs - W$ and $fs = 1/(Ts)$. $\{x(nTs)\} \equiv xs(t) = \Sigma\, x(t).\delta(t- nTs)$
where $x(nTs) = x(t)|_{t=nTs}$, δ(t) is a unit pulse singularity function and „n" is an integer. The continuous-time signal x(t) is multiplied by an (ideal) impulse train to obtain {x(nTs)} and can be rewritten as,

$$xs(t) = x(t).\Sigma\, \delta(t- nTs) \qquad \text{-------------1.2}$$

Now, let X(f) denote the Fourier Transform F(T) of x(t), i.e.

$$X(f) = \int_{-\infty}^{+\infty} x(t).\exp(-j2\pi ft)dt$$

-------------1.3

Now, from the theory of Fourier Transform, we know that the F.T of $\Sigma \quad \delta$ (t- nTs), the impulse train in time domain, is an impulse train in frequency domain:

.

$$F\{\Sigma \quad \delta(t-nTs)\} = (1/Ts).\Sigma \quad \delta(f-n/Ts) = fs.\Sigma \quad \delta(f-nfs) \qquad \text{-----1.4}$$

If Xs(f) denotes the Fourier transform of the energy signal xs(t), we can use convolution property:

$$Xs(f) = X(f)*F\{\Sigma \quad \delta(t-nTs)\}$$
$$= X(f)*[fs.\Sigma \quad \delta(f-nfs)]$$
$$= fs.X(f)*\Sigma \quad \delta(f-nfs)$$

$$= f_s.\int_{-\infty}^{+\infty} X(\lambda).\Sigma\,\delta(f-nf_s-\lambda)d\lambda = f_s.\,\Sigma\int X(\lambda).\delta(f-nf_s-\lambda)d\lambda = f_s.\Sigma\,X(f-nf_s)$$

--------------1.5

This equation, when interpreted appropriately, gives an intuitive proof to Nyquist‟s theorems as stated above and also helps to appreciate their practical implications. Let us note that while writing Eq.(1.5), we assumed that x(t) is an energy signal so that its Fourier transform exists.

With this setting, if we assume that x(t) has no appreciable frequency component greater than W Hz and if fs > 2W, then Eq.(1.5) implies that Xs(f), the Fourier Transform of the sampled signal xs(t) consists of infinite number of replicas of X(f), centered at discrete frequencies n.fs, -∞ < n < ∞ and scaled by a constant fs= 1/Ts

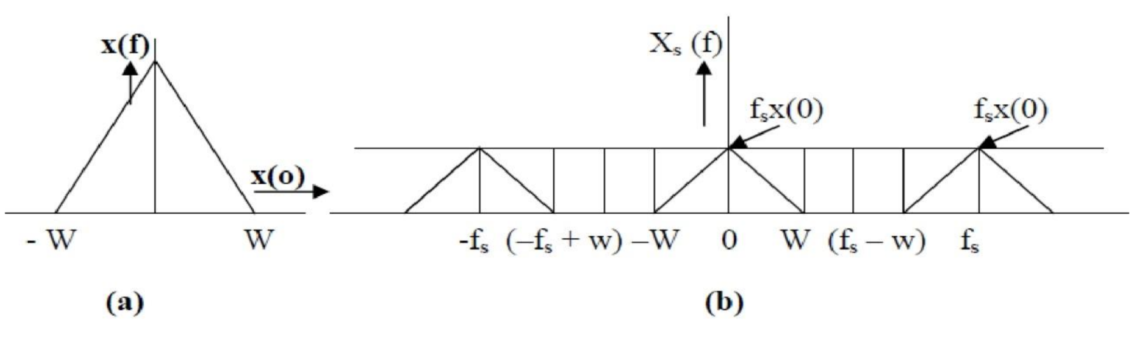


Fig 2.5 Indicates that the bandwidth of this instantaneously sampled wave xs(t) is infinite while the spectrum of x(t) appears in a periodic manner, centered at discrete frequency values n.fs.

Part – I of the sampling theorem is about the condition fs > 2.W i.e. (fs − W) > W and (− fs + W) < − W. As seen from Fig. 1.2.1, when this condition is satisfied, the spectra . of xs(t), centered at f = 0 and f = ± fs do not overlap and hence, the spectrum of x(t) is present in xs(t) without any distortion. This implies that xs(t), the appropriately sampled version of x(t), contains all information about x(t) and thus represents x(t).

The second part suggests a method of recovering x(t) from its sampled version xs(t) by using an ideal lowpass filter. As indicated by dotted lines in Fig. 1.2.1, an ideal lowpass filter (with brick-wall type response) with a bandwidth $W \leq B < (fs - W)$, when fed with xs(t), will allow the portion of Xs(f), centered at f = 0 and will reject all its replicas at f = n fs, for $n \neq 0$.

This implies that the shape of the continuous time signal xs(t), will be retained at the output of the ideal filter. If the sampling rate, $fs \geq 2fu$, exact reconstruction is possible in which case the signal g(t) may be considered as a low pass signal itself.

Sampling of Band Pass Signals:

Consider a band-pass signal g(t) with the spectrum shown in figure:

Band width = B

Upper Limit = f_u

Lower Limit = f_l

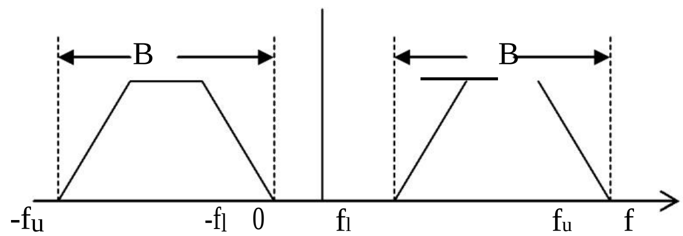


Fig 2.6 Spectrum of sampled signal

Example-1 :

Consider a signal g(t) having the Upper Cutoff frequency, fu = 100KHz and the Lower Cutoff frequency fl = 80KHz.

The ratio of upper cutoff frequency to bandwidth of the signal g(t) is

f_u / B = 100K / 20K = 5.

Therefore we can choose m = 5.

Then the sampling rate is $f_s = 2f_u / m$ = 200K / 5 = 40KHz

Example-2 :

Consider a signal g(t) having the Upper Cutoff frequency, fu = 120KHz and the Lower Cutoff frequency fl = 70KHz.

The ratio of upper cutoff frequency to bandwidth of the signal g(t) is

fu / B = 120K / 50K = 2.4

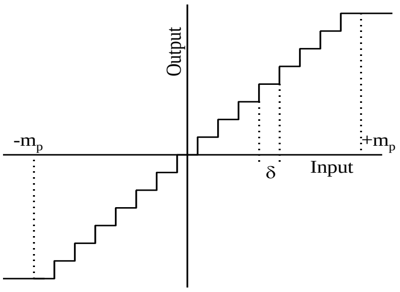
Therefore we can choose m = 2. ie.. m is an integer less than (fu /B).

Then the sampling rate is fs = 2fu / m = 240K / 2 = 120KHz.

Quantization

This is the process of setting the sample amplitude, which can be continuously variable to a discrete value. Look at Uniform Quantization first, where the discrete values are evenly spaced.

Uniform Quantization



We assume that the amplitude of the signal m(t) is confined to the range ($-m_p$, $+m_p$). This range ($2m_p$) is divided into L levels, each of step size δ, given by

$$\delta = 2\, m_p / L$$

A sample amplitude value is approximated by the midpoint of the interval in which it lies. The input/output characteristics of a uniform quantizer is shown.

Fig 2.7 Qualtization

Types of Quantizers:

1. Uniform Quantizer
2. Non- Uniform Quantizer

<u>Uniform Quantizer:</u>

In Uniform type, the quantization levels are uniformly spaced, whereas in non-uniform type the spacing between the levels will be unequal and mostly the relation is logarithmic.

Types of Uniform Quantizers: (based on I/P - O/P Characteristics)

1. Mid-Rise type Quantizer
2. Mid-Tread type Quantizer

In the stair case like graph, the origin lies the middle of the tread portion in Mid – Tread
type where as the origin lies in the middle of the rise portion in the Mid-Rise type.

Mid – tread type: Quantization levels – odd number.
Mid – Rise type: Quantization levels – even number.

<u>Quantization Noise and Signal-to-Noise:</u>

“The Quantization process introduces an error defined as the difference between the input signal, x(t) and the output signal, yt). This error is called the Quantization Noise.”

$$q(t) = x(t) – y(t)$$

Quantization noise is produced in the transmitter end of a PCM system by rounding off sample values of an analog base-band signal to the nearest permissible representation levels of the quantizer. As such quantization noise differs from channel noise in that it is signal dependent.

Let „Δ" be the step size of a quantizer and L be the total number of quantization levels.

Quantization levels are 0, ± ., ± 2 ., ±3 . . . .

The Quantization error, Q is a random variable and will have its co-sample values bounded by [-(Δ/2) < q < (Δ/2)]. is small, the quantization error can be assumed to a uniformly distributed random variable.

Consider a memory less quantizer that is both uniform and symmetric.

L = Number of quantization levels X = Quantizer input

Y = Quantizer output

The output y is given by

$$Y=Q(x)$$

which is a staircase function that befits the type of mid tread or mid riser quantizer of interest.

<u>Non – Uniform Quantizer:</u>

In Non – Uniform Quantizer the step size varies. The use of a non – uniform quantizer is equivalent to passing the baseband signal through a compressor and then applying the www compressed signal to a uniform quantizer. The resultant signal is then transmitted.

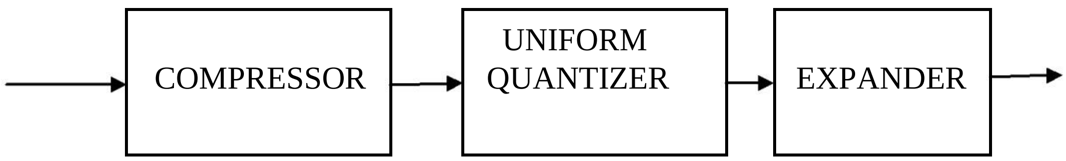


Fig: 2.8 Model of Non-Uniform Quantizer

At the receiver, a device with a characteristic complementary to the compressor called <u>Expander</u> is used to restore the signal samples to their correct relative level.

The Compressor and expander take together constitute a <u>Compander.</u>

Compander = Compressor + Expander

Advantages of Non – Uniform Quantization :

1. Higher average signal to quantization noise power ratio than the uniform quantizer when the signal pdf is non uniform which is the case in many practical situation.
2. RMS value of the quantizer noise power of a non – uniform quantizer is substantially proportional to the sampled value and hence the effect of the quantizer noise is reduced.

Companding

In a uniform or linear PCM system the size of every quantization interval is determined by the SQR requirement of the lowest signal to be encoded. This interval is also for the largest signal - which therefore has a much better SQR.

Example: A 26 dB SQR for small signals and a 30 dB dynamic range produces a 56 dB SQR for the maximum amplitude signal.

In this way a uniform PCM system provides unneeded quality for large signals. In speech the max amplitude signals are the least likely to occur. The code space in a uniform PCM system is very inefficiently utilised.

A more efficient coding is achieved if the quantization intervals increase with the sample value. When the quantization interval is directly proportional to the sample value (assign small quantization intervals to small signals and large intervals to large signals) the SQR is constant for all signal levels.

With this technique fewer bits per sample are required to provide a specified SQR for small signals and an adequate dynamic range for large signals (but still with the SQR as for the small signals). The quantization intervals are not constant and there will be a non linear relationship between the code words and the values they represent.

Originally to produce the non linear quantization the baseband signal was passed through a non-linear amplifier with input/output characteristics as shown before the samples were taken. Low level signals were amplified and high level signals were attenuated. The larger the sample value the more it is **comp**ressed before encoding. The PCM decoder ex**pands** the compressed value using an inverse compression characteristic to recover the original sample value. The two processes are called **companding**.

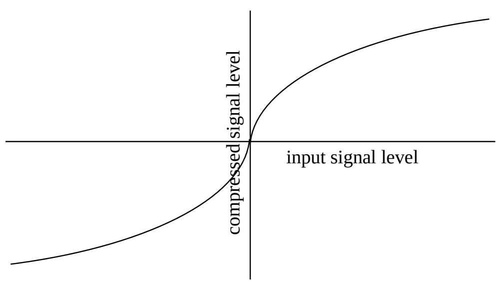


Fig 2.9 Companding

There are 2 companding schemes to describe the curve above:

1. **μ-Law Companding** (also called log-PCM)

   This is used in North America and Japan. It uses a logarithmic compression curve which is ideal in the sense that quantization intervals and hence quantization noise is directly proportional to signal level (and so a constant SQR).

2. **A- Law Companding**
   This is the ITU-T standard. It is used in Europe and most of the rest of the world. It is very similar to the μ-Law coding. It is represented by straight line segments to facilitate digital companding.

Originally the non linear function was obtained using non linear devices such as special diodes. These days in a PCM system the A to D and D to A converters (ADC and DAC) include a companding function.

Pulse Amplitude Modulation:

- In fact the pulses in a PAM signal may of Flat-top type or natural type or ideal type.
- The Flat-top PAM is most popular and is widely used. The reason for using Flat-top PAM is that during the transmission, the noise interferes with the top of the transmitted pulses and this noise can be easily removed if the PAM pulse as Flat-top.
- In natural samples PAM signal, the pulse has varying top in accordance with the signal variation. Such type of pulse is received at the receiver, it is always contaminated by noise. Then it becomes quite difficult to determine the shape of the top of the pulse and thus amplitude detection of the pulse is not exact.

Generation of PAM:

There are two operations involved in the generation of PAM signal

1. Instantaneous sampling of the message signal m(t) every Ts seconds, where the sampling rate fs = 1/Ts is chosen in accordance with the sampling theorem.
2. Lengthening the duration of each sample so obtained to some constant value T.

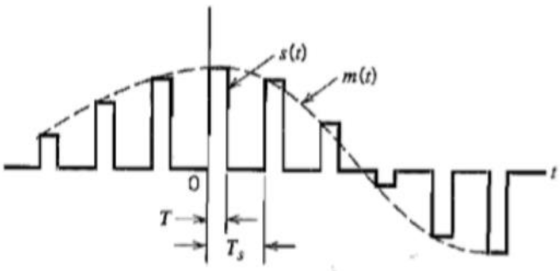


Fig 2.10 PAM Signal

Sample and Hold Circuit for Generating Flat-top sampled PAM

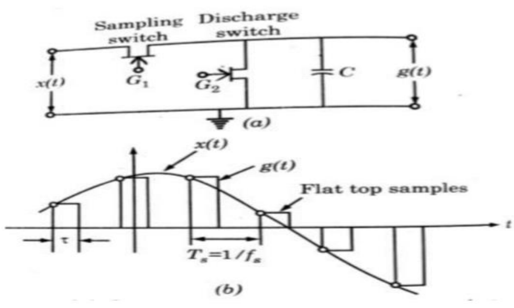


Fig 2.11 Sample and hold circuit for PAM Signal generation

- The sample and hold circuit consists of two Field Effect Transistor switches and a capacitor.

- The sampling switch is closed for a short duration by a short pulse applied to the gate G1 of the transistor. During this period, the capacitor C is quickly charged up to a voltage equal to the instantaneous sample value of the incoming signal.
- Now, the sampling switch is opened and the capacitor holds the charge. The discharge switch is then closed by a pulse applied to gate G2 of the other transistor. Due to this, the capacitor is discharged to zero volts. The discharges switch is then opened and thus capacitor has no voltage. Hence the output of the sample and hold circuit consists of a sequence of flat-top samples as shown in figure.

Mathematical Representation of PAM

We may express the PAM signal as

$$s(t) = \sum_{n=-\infty\infty}^{\infty} m(nT_s)h(t - nT_s)$$

where Ts = sampling period
m(nTs) = sample value of m(t) obtained at t = nTs
h(t) = standard rectangular pulse of unit amplitude and duration T and it is defined as The spectrum of flat-top PAM signal is

$$h(t) = \begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, t & = 0, t = T \\ 0, & otherwise \end{cases}$$

The spectrum of flat-top PAM signal is

$$S(f) = f_s \sum_{k=-\infty}^{\infty} M(f - kf_s)H(f)$$

Transmission Bandwidth of PAM:

In PAM signal the pulse duration τ is assumed to be very small compared to time period Ts between the two samples i.e τ< Ts

- If the maximum frequency in the modulating signal x(t) is fm then sampling frequency fs is given by fs>=2fm Or 1/Ts >= 2fm or Ts <= 1/2fm Therefore, τ < < Ts <= 1/2fm
- If ON and OFF time of PAM pulse is equal then maximum frequency of PAM pulse will be fmax = 1/ τ+ τ = 1/2 τ

Therefore,
transmission bandwidth >=fmax
But fmax=1/2 τ

B.W>=1/2 τ

B.W>=1/2 τ >>fm

Demodulation of PAM:

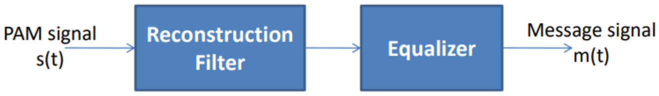


Fig 2.12 PAM Reconstruction

- The distortion caused using PAM to transmit an analog information bearing signal is referred to as the aperture effect. This distortion may be corrected by connecting an equalizer in cascade with the low-pass reconstruction filter as shown in fig.
- The equalizer has the effect of decreasing the in-band loss of the reconstruction filter as the frequency increases in such a manner as to compensate for the aperture effect.

Ideally, the magnitude response of the equalizer is given by

$$\frac{1}{|H(f)|} = \frac{1}{T\mathrm{sinc}(fT)} = \frac{\pi f}{\sin(\pi f T)}$$

The amount of equalization needed in practice is usually small.

Advantages of PAM :

• It is the simple and simple process for modulation and demodulation
• Transmitter and receiver circuits are simple and easy to construct.

CLASSIFICATION OF LINECODES

Line coding:

Line coding, refers to the process of representing the bit stream (**1**‟s and **0**‟s) in the form of voltage or current variations optimally tuned for the specific properties of the physical channel being used.

The selection of a proper line code can help in so many ways: One possibility is to aid in clock recovery at the receiver. A clock signal is recovered by observing transitions in the received bit sequence, and if enough transitions exist, a good recovery of the clock is guaranteed, and the signal is said to be **self-clocking**.

Another advantage is to get rid of DC shifts. The DC component in a line code is called the *bias* or the *DC coefficient*. Unfortunately, most long-distance communication channels cannot transport a DC component. This is why most line codes try to eliminate the DC component before being transmitted on the channel. Such codes are called *DC balanced*, *zero-DC*, *zero-bias*, or *DC equalized*. Some common types of line encoding in common-use nowadays are unipolar, polar, bipolar, Manchester, MLT-3 and Duobinary encoding. These codes are explained here:

Unipolar (Unipolar **NRZ** and Unipolar **RZ**):

Unipolar is the simplest line coding scheme possible. It has the advantage of being compatible with TTL logic. Unipolar coding uses a positive rectangular pulse $p(t)$ to represent binary **1**, and the absence of a pulse (i.e., zero voltage) to represent a binary **0**. Two possibilities for the pulse $p(t)$ exist3: Non-Return-to-Zero (NRZ) rectangular pulse and Return-to-Zero (RZ)

rectangular pulse. The difference between Unipolar NRZ and Unipolar RZ codes is that the rectangular pulse in NRZ stays at a positive value (e.g., +5V) for the full duration of the logic **1** bit, while the pule in RZ drops from +5V to 0V in the middle of the bit time.
A drawback of unipolar (RZ and NRZ) is that its average value is not zero, which means it creates a significant DC-component at the receiver (see the impulse at zero frequency in the corresponding power spectral density (PSD) of this line code.

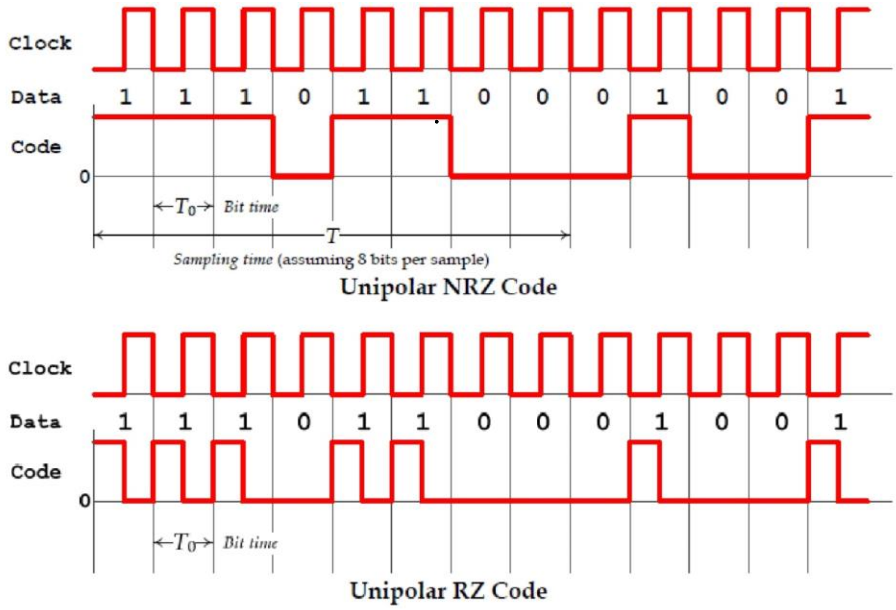


Fig 2.13 Unipolar Waveform

The ***disadvantage*** of unipolar RZ pared to unipolar NRZ is that each rectangular pulse in RZ is only half the length of NRZ pulse. This means that unipolar RZ requires twice the bandwidth of the NRZ code.

<u>Polar</u> <u>(Polar **NRZ** and Polar **RZ**):</u>

In Polar NRZ line coding binary **1**"s are represented by a pulse $p(t)$ and binary **0**"s are represented by the negative of this pulse $-p(t)$ (e.g., -5V). Polar (NRZ and RZ) signals .Using the assumption that in a regular bit stream a logic **0** is just as likely as a logic **1**,polar signals (whether RZ or NRZ) have the advantage that the resulting Dc component is very close to zero.

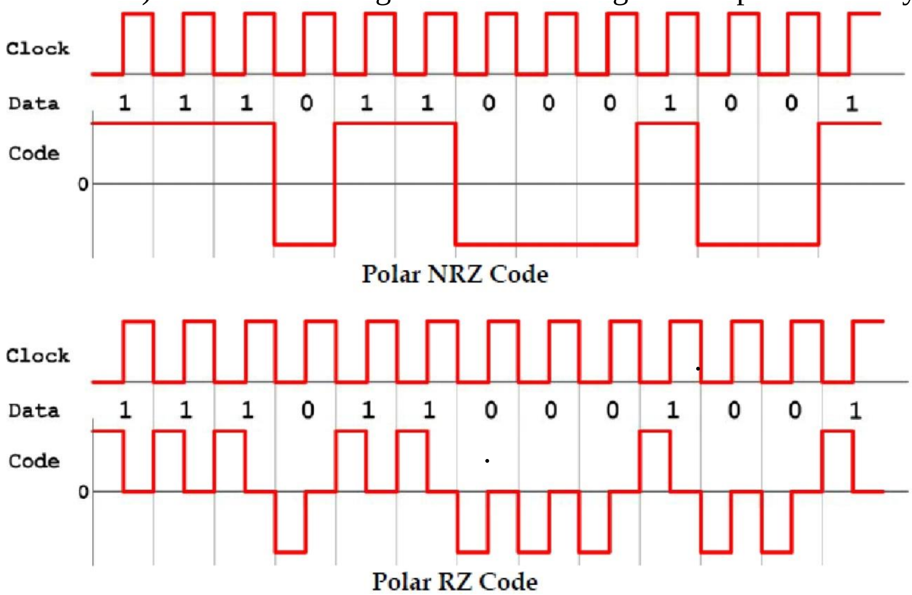


Fig 2.14 Polar waveform

The RMS value of polar signals is bigger than unipolar signals, which means that polar signals have more power than unipolar signals, and hence have better SNR at the receiver. Actually, polar NRZ signals have more power pared to polar RZ signals. The drawback of polar NRZ, however, is that it lacks clock information especially when a long sequence of **0**"s or **1**"s is transmitted.

Non-Return -to-Zero, Inverted (NRZI):

NRZI is a variant of Polar NRZ. In NRZI there are two possible pulses, $p(t)$ and $-p(t)$. A transition from one pulse to the other happens if the bit being transmitted is a logic **1**, and no transition happens if the bit being transmitted is a logic **0**.

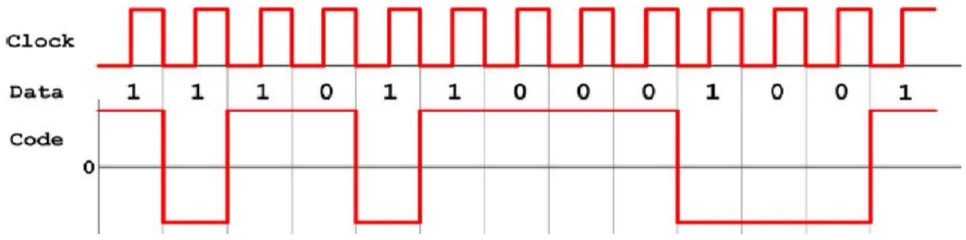


Fig 2.15 NRZI waveform

This is the code used on pact discs (CD), USB ports, and on fiber-based Fast Ethernet at 100-Mbit/s .

Manchester encoding:

In Manchester code each bit of data is signified by at least one transition. Manchester encoding is therefore considered to be self-clocking, which means that accurate clock recovery from a data stream is possible.

In addition, the DC component of the encoded signal is zero. Although transitions allow the signal to be self-clocking, it carries significant overhead as there is a need for essentially twice the bandwidth of a simple NRZ or NRZI encoding

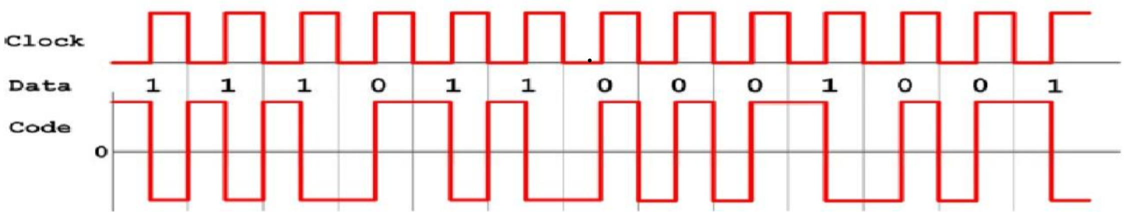


Fig 2.16 Manchester coding

POWER SPECTRA OF LINE CODES:

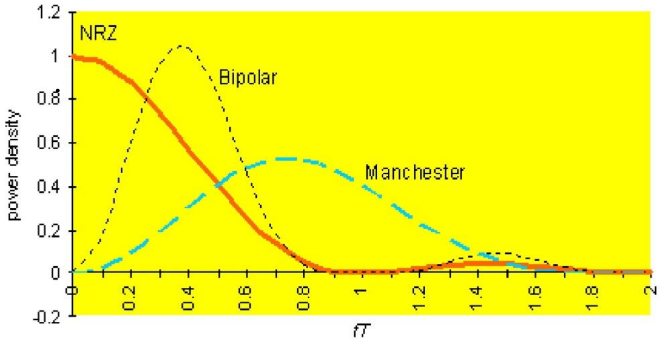


Fig 2.17 Power spectra of Line codes

- Unipolar most of signal power is centered on origin and there is waste of power due to DC component that is present.
- Polar format most of signal power is centered on origin and they are simple to implement.
- Bipolar format does not have DC component and does not demand more bandwidth, but power requirement is double than other formats.
- Manchester format does not have DC component but provides proper clocking.

Pulse Code Modulation

Pulse Code Modulation (PCM) is an extension of PAM wherein each analogue sample value is quantized into a discrete value for representation as a digital code word.

Thus, as shown below, a PAM system can be converted into a PCM system by adding a suitable analogue-to-digital (A/D) converter at the source and a digital-to-analogue (D/A) converter at the destination.

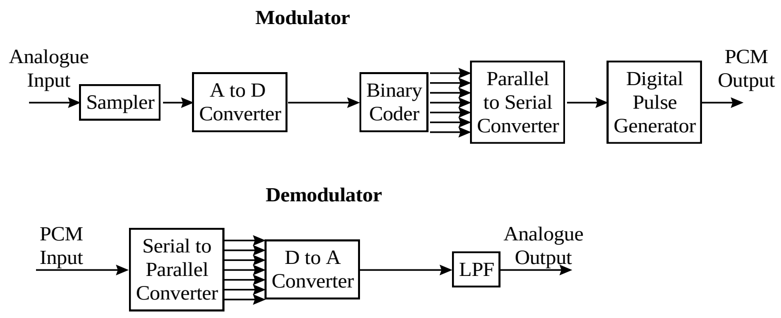


Fig 2.18 Pulse code Modulation

PCM is a true digital process as compared to PAM. In PCM the speech signal is converted from analogue to digital form.

PCM is standardised for telephony by the ITU-T (International Telecommunications Union - Telecoms, a branch of the UN), in a series of recommendations called the G series. For example the ITU-T recommendations for out-of-band signal rejection in PCM voice coders require that 14 dB of attenuation is provided at 4 kHz. Also, the ITU-T transmission quality specification for telephony terminals require that the frequency response of the handset microphone has a sharp roll-off from 3.4 kHz.

In quantization the levels are assigned a binary codeword. All sample values falling between two quantization levels are considered to be located at the centre of the quantization interval. In this manner the quantization process introduces a certain amount of error or distortion into the signal samples. This error known as quantization noise, is minimised by establishing a large number of small quantization intervals. Of course, as the number of quantization intervals increase, so must the number or bits increase to uniquely identify the quantization intervals. For example, if an analogue voltage level is to be converted to a digital system with 8 discrete levels or quantization steps three bits are required. In the ITU-T version there are 256 quantization steps, 128 positive and 128 negative, requiring 8 bits. A positive level is represented by having bit 8 (MSB) at 0, and for a negative level the MSB is 1.

Differential pulse coding schemes

PCM transmits the absolute value of the signal for each frame. Instead we can transmit information about the difference between each sample. The two main differential coding schemes are:

- Delta Modulation
- Differential PCM and Adaptive Differential PCM (ADPCM)

Delta Modulation

Delta modulation converts an analogue signal, normally voice, into a digital signal.

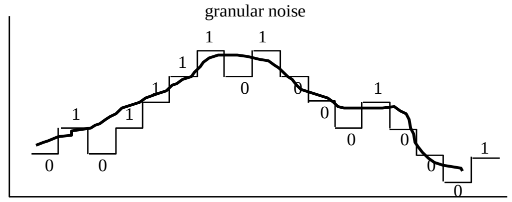


Fig 2.19 Delta Modulation Waveform

The analogue signal is sampled as in the PCM process. Then the sample is compared with the previous sample. The result of the comparison is quantified using a one bit coder. If the sample is greater than the previous sample a 1 is generated. Otherwise a 0 is generated. The advantage of delta modulation over PCM is its simplicity and lower cost. But the noise performance is not as good as PCM.

To reconstruct the original from the quantization, if a 1 is received the signal is increased by a step of size q, if a 0 is received the output is reduced by the same size step. Slope overload occurs when the encoded waveform is more than a step size away from the input signal. This condition happens when the rate of change of the input exceeds the maximum change that can be generated by the output. Overload will occur if:

$$dx(t)/dt > q/T = q * f_s$$

where: x(t) = input signal, q = step size, T = period between samples, f_s = sampling frequency

Assume that the input signal has maximum amplitude A and maximum frequency F. The most rapidly changing input is provided by $x(t) = A * \sin(2 * \pi * F * t)$.

For this $dx(t)/dt = 2 * \pi * F * A * \sin(2 * \pi * F * t)$.

This slope has a maximum value of $2 * \pi * F * A$

Overload occurs if $2 * \pi * F * A > q * f_s$

To prevent overload we require $q * f_s > 2 * \pi * F * A$

Example A = 2 V, F = 3.4 kHz, and the signal is sampled 1,000,000 times per second, requires q > 2 * 3.14 * 3,400 * 2 /1,000,000 V > 42.7 mV

Granular noise occurs if the slope changes more slowly than the step size. The reconstructed signal oscillates by 1 step size in every sample. It can be reduced by decreasing the step size. This requires that the sample rate be increased. Delta Modulation requires a sampling rate much higher than twice the bandwidth. It requires oversampling in order to obtain an accurate prediction of the next input, since each encoded sample contains a relatively small amount of information. Delta Modulation requires higher sampling rates than PCM.

Differential PCM (DPCM) and ADPCM

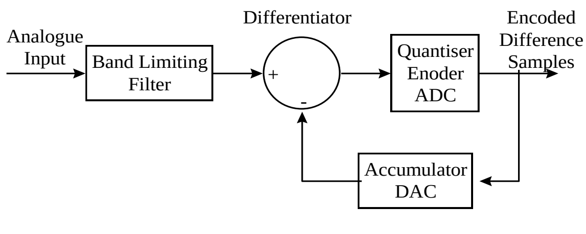


Fig 2.20 DPCM & ADPCM

DPCM is also designed to take advantage of the redundancies in a typical speech waveform. In DPCM the differences between samples are quantized with fewer bits that would be used for quantizing an individual amplitude sample. The sampling rate is often the same as for a comparable PCM system, unlike Delta Modulation.

Adaptive Differential Pulse Code Modulation ADPCM is standardised by ITU-T recommendations G.721 and G.726. The method uses 32,000 bits/s per voice channel, as compared to standard PCM's 64,000 bits/s. Four bits are used to describe each sample, which represents the difference between two adjacent samples. Sampling is 8,000 times a second. It makes it possible to reduce the bit flow by half while maintaining an acceptable quality. While the use of ADPCM (rather than PCM) is imperceptible to humans, it can significantly reduce the throughput of high-speed modems and fax transmissions.

The principle of ADPCM is to use our knowledge of the signal in the past time to predict the signal one sample period later, in the future. The predicted signal is then compared with the actual signal. The difference between these is the signal which is sent to line - it is the error in the prediction. However this is not done by making comparisons on the incoming audio signal - the comparisons are done after PCM coding.

To implement ADPCM the original (audio) signal is sampled as for PCM to produce a code word. This code word is manipulated to produce the predicted code word for the next sample. The new predicted code word is compared with the code word of the second sample. The result of this comparison is sent to line. Therefore we need to perform PCM before ADPCM.

The ADPCM word represents the prediction error of the signal, and has no significance itself. Instead the decoder must be able to predict the voltage of the recovered signal from the previous samples received, and then determine the actual value of the recovered signal from this prediction and the error signal, and then to reconstruct the original waveform.

ADPCM is sometimes used by telecom operators to fit two speech channels onto a single 64 kbit/s link. This was very common for transatlantic phone calls via satellite up until a few years ago. Now, nearly all calls use fibre optic channels at 64 kbit/s.

<u>Adaptive Delta Modulation:</u>

The performance of a delta modulator can be improved significantly by making the step size of the modulator assume a time-varying form. In particular, during a steep segment of the input signal the step size is increased. Conversely, when the input signal is varying slowly, the step size is reduced. In this way, the size is adapted to the level of the input signal. The resulting method is called adaptive delta modulation (ADM). There are several types of ADM, depending on the type of scheme used for adjusting the step size. In this ADM, a discrete set of values is provided for the step size.

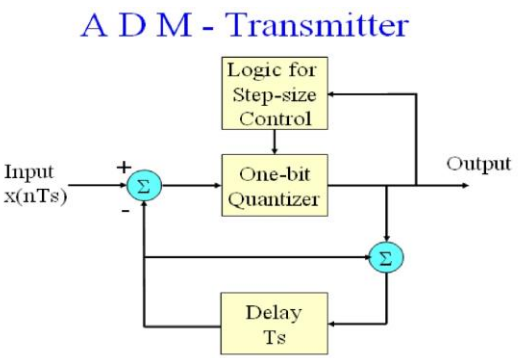


Fig 2.21 Block Diagram of ADM Transmitter

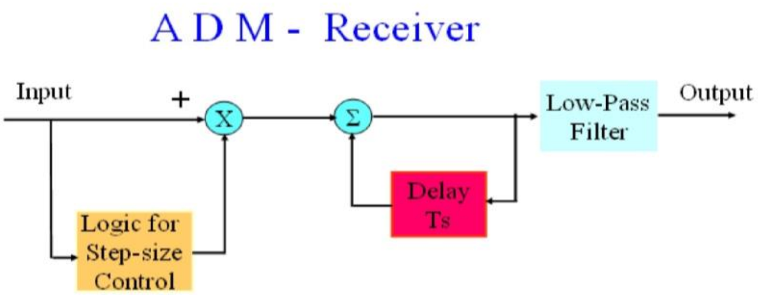


Fig 2.22 Block Diagram of ADM Receiver

Prediction filtering & Linear Predictive Coding.

The speech signal is filtered to no more than one half the system sampling frequency and then A/D conversion is performed. The speech is processed on a frame by frame basis where the analysis frame length can be variable. For each frame a pitch period estimation is made along with a voicing decision. A linear predictive coefficient analysis is performed to obtain an inverse model of the speech spectrum A (z). In addition a gain parameter G, representing some function of the speech energy is computed. An encoding procedure is then applied for transforming the analyzed parameters into an efficient set of transmission parameters with the goal of minimizing the degradation in the synthesized speech for a specified number of bits. Knowing the transmission frame rate and the number of bits used for each transmission parameters, one can compute a noise-free channel transmission bit rate.

At the receiver, the transmitted parameters are decoded into quantized versions of the coeifficent analysis and pitch estimation parameters. An excitation signal for synthesis is then constructed from the transmitted pitch and voicing parameters. The excitation signal then drives a synthesis filter 1/A (z) corresponding to the analysis model A (z). The digital samples s^(n) are then passed through an D/A converter and low pass filtered to generate the synthetic speech s(t). Either before or after synthesis, the gain is used to match the synthetic speech energy to the actual speech energy. The digital samples are the converted to an analog signal and passed through a filter similar to the one at the input of the system.

Linear predictive coding (LPC) of speech

The linear predictive coding (LPC) method for speech analysis and synthesis is based on modeling the Vocal tract as a linear All-Pole (IIR) filter having the system transfer function:

$$H(z) = \frac{G}{1 - \sum a_k z^{-k}}$$

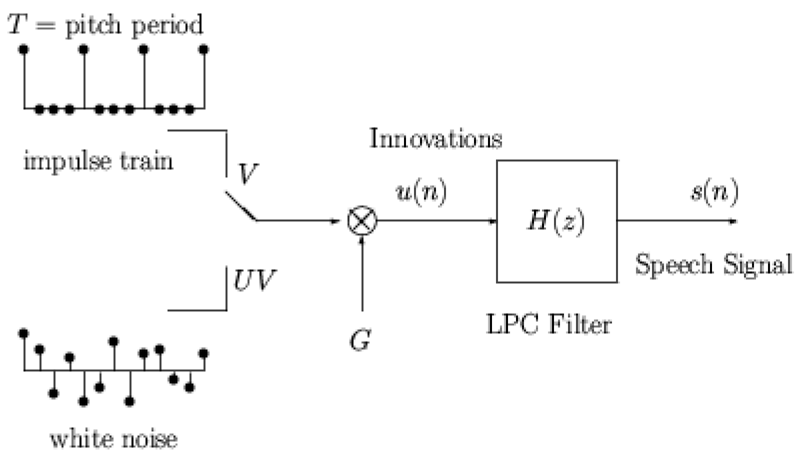


Fig 2.23 Simple speech production

Where p is the number of poles, G is the filter Gain, and a[k] are the parameters that determine the poles. There are two mutually exclusive ways excitation functions to model voiced and unvoiced speech sounds. For a short time-basis analysis, voiced speech is considered periodic with a fundamental frequency of Fo, and a pitch period of 1/Fo, which depends on the speaker. Hence, Voiced speech is generated by exciting the all pole filter model by a periodic impulse train. On the other hand, unvoiced sounds are generated by exciting the all-pole filter by the output of a random noise generator. The fundamental difference between these two types of speech sounds comes from the way they are produced. The vibrations of the vocal cords produce voiced sounds. The rate at which the vocal cords vibrate dictates the pitch of the sound. On the other hand, unvoiced sounds do not rely on the vibration of the vocal cords. The unvoiced sounds are created by the constriction of the vocal tract. The vocal cords remain open and the constrictions of the vocal tract force air out to produce the unvoiced sounds

Given a short segment of a speech signal, lets say about 20 ms or 160 samples at a sampling rate 8 KHz, the speech encoder at the transmitter must determine the proper excitation function, the pitch period for voiced speech, the gain, and the coefficients ap[k]. The block diagram below describes the encoder/decoder for the Linear Predictive Coding. The parameters of the model are determined adaptively from the data and modeled into a binary sequence and transmitted to the receiver. At the receiver point, the speech signal is the synthesized from the model and excitation signal.

The parameters of the all-pole filter model are determined from the speech samples by means of linear prediction. To be specific the output of the Linear Prediction filter is

$$\hat{s}(n) = -\sum_{k=1}^{p} a_p(k)s(n-k)$$

and the corresponding error between the observed sample S(n) and the predicted value $\hat{s}(n)$ is

$$e(n) = s(n) - \hat{s}(n)$$

by minimizing the sum of the squared error we can determine the pole parameters $a_p(k)$ of the model. The result of differentiating the sum above with respect to each of the parameters and equation the result to zero, is a sep of p linear equations

$$\sum_{k=1}^{p} a_p(k)r_{ss}(m-k) = -r_{ss(m)} \quad \text{where m=1,2,....p}$$

where $r_{ss(m)}$ represent the autocorrelation of the sequence $s(n)$ defined as

$$r_{ss(m)} = \sum_{n=0}^{N} s(n)s(n+m)$$

the equation above can be expressed in matrix form as

$$R_{ss}a = -r_{ss(m)}$$

where $R_{ss}a$ is a pxp autocorrelation matrix, r_{ss} is a px1 autocorrelation vector, and a is a px1 vector of model parameters.

The gain parameter of the filter can be obtained by the input-output relationship as follow

$$s(n) = -\sum_{k=1}^{p} a_p(k)s(n-k) + Gx(n)$$

where X(n) represent the input sequence.

We can further manipulate this equation and in terms of the error sequence we have

$$Gx(n) = s(n) + \sum_{k=1}^{p} a_p(k)s(n-k) = e(n)$$

then

$$G^2 \sum_{n=0}^{N-1} x^2(n) = \sum_{n=0}^{N-1} e^2(n)$$

if the input excitation is normalized to unit energy by design, then

$$G^2 \sum_{n=0}^{N-1} x^2(n) = \sum_{n=0}^{N-1} e^2(n) = r_{ss}(0) + \sum_{k=1}^{p} a_p(k) r_{ss}(k)$$

where G^2 is set equal to the residual energy resulting from the least square optimization .

once the LPC coefficients are computed, we can determine weather the input speech frame is voiced, and if it is indeed voiced sound, then what is the pitch. We can determine the pitch by computing the following sequence in matlab. $r_e(n) = \sum_{k=1}^{p} r_a(k) r_{ss}(n-k)$

where $r_a(k)$ is defined as follow

$$r_a(n) = \sum_{k=1}^{p} a_a(k) a_p(i+k)$$

which is defined as the autocorrelation sequence of the prediction coefficients. The pitch id detected by finding the peak of the normalized sequence $\frac{r_e(n)}{r_e(0)}$ In the time interval corresponds to 3 to 15 ms in the 20ms sampling frame. If the value of this peak is at least 0.25, the frame of speech is considered voiced with a pitch period equal to the value of $n = N_p$, where $\frac{r_e(N_p)}{r_e(0)}$ is a maximum value.

If the peak value is less than 0.25, the frame speech is considered unvoiced and the pitch would equal to zero.

The value of the LPC coefficients, the pitch period, and the type of excitation are then transmitted to the receiver. The decoder synthesizes the speech signal by passing the proper excitation through the all-pole filter model of the vocal tract.

Typically the pitch period requires 6 bits, the gain parameters are represented in 5 bits after the dynamic range is compressed logarithmically, and the prediction coefficients require 8-10 bits normally for accuracy reasons. This is very important in LPC because any small changes in the prediction coefficients result in large change in the pole positions of the filter model, which cause instability in the model. This is overcome by using the PARACOR method.

Is speech frame Voiced or Unvoiced ?

Once the LPC coefficients are competed, we can determine weather the input speech frame is voiced, and if so, what the pitch is.

If the speech frame is decided to be voiced, an impulse train is employed to represent it, with nonzero taps occurring every pitch period. A pitch-detecting algorithm is used in order to determine to correct pitch period / frequency. The autocorrelation function is used to estimate the pitch period as. However, if the frame is unvoiced, then white noise is used to represent it and a pitch period of T=0 is transmitted. Therefore, either white noise or impulse train becomes the excitation of the LPC synthesis filter

Two types of LPC vocoders were implemented in MATLAB

Plain LPC Vocoder diagram is shown below :

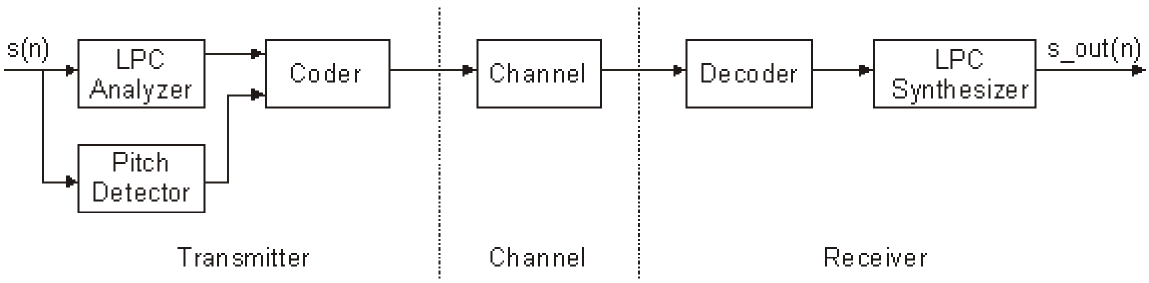


Fig 2.24 LPC Vocoder

Time Division Multiplexing (TDM) - Principle

When sending samples of a signal instead of the signal itself there is time available between each of the samples. Samples from other analogue signals can be put into this space.

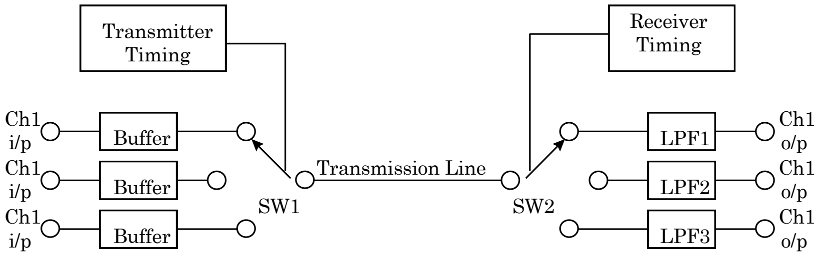


Fig 2.25 Time Division Multiplexing

The process of splitting up the time into slots and putting different signals into the time slots is known as Time Division Multiplexing (TDM). A basic real TDM system interleaves 32 signals and uses electronic switches. This is a diagram of a **3 channel PAM-TDM system.**

This diagram shows the waveforms produced during the operation of the PAM-TDM system

The switches connect the transmitter and the receiver to each of the channels in turn for a specific interval of time. In effect each channel is sampled and the sample is transmitted

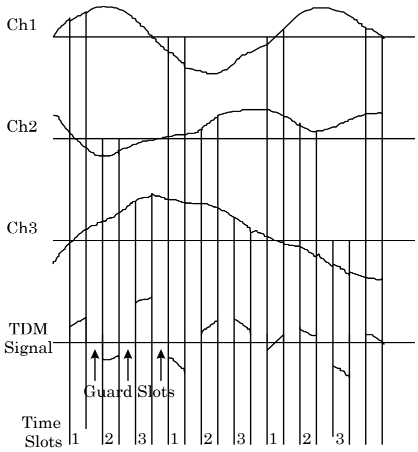


Fig 2.26 TDM Waveform

When the switches are in the channel 1 position, channel 1 forms a PAM channel with an LPF for reconstruction, and so on for channels 2 and 3. The result is that the amplitudes samples from each channel share the line sequentially, becoming interleaved to form a complex PAM wave, as shown above.

A major problems in any TDM system is the **synchronisation** of the transmitter and receiver timing circuits. The transmitter and receiver must switch at the same time and frequency. Also SW1 must be in the channel 1 position when SW2 is in the channel 1 position, so that the switches must be synchronised in position also.

In a system that uses analogue modulation (PAM) the time slots are separated by guard slots to prevent crosstalk between channels.

Frequency Division Multiplexing

- Frequency Division Multiplexing (FDM) Frequency-division multiplexing is a form of signal multiplexing which involves assigning non-overlapping frequency ranges to different signals or to each "user of a medium.
- FDM achieves the combining of several signals into one medium by sending signals in several distinct frequency ranges over a single medium.
- Frequency division multiplexing involves translation of the speech signal from the frequency band 300-3400 Hz to a higher frequency band. Each channel is translated to a different hand and then all the channels are combined to form a frequency division multiplexed signal.

In FDM, the speech channels are stacked at intervals of 4 kHz to provide a guard band between adjacent channels.

FDM can be applied when the bandwidth of a link (in hertz) is greater than the combined bandwidths of the signals to be transmitted.

A demultiplexer applies a set of filters that each extract a small range of frequencies near one of the carrier frequencies.

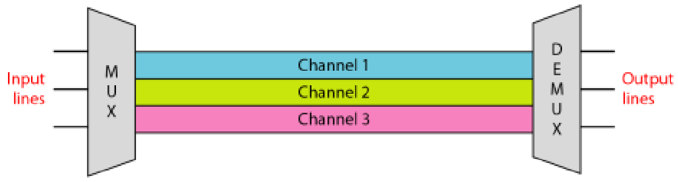


Fig 2.27 FDM

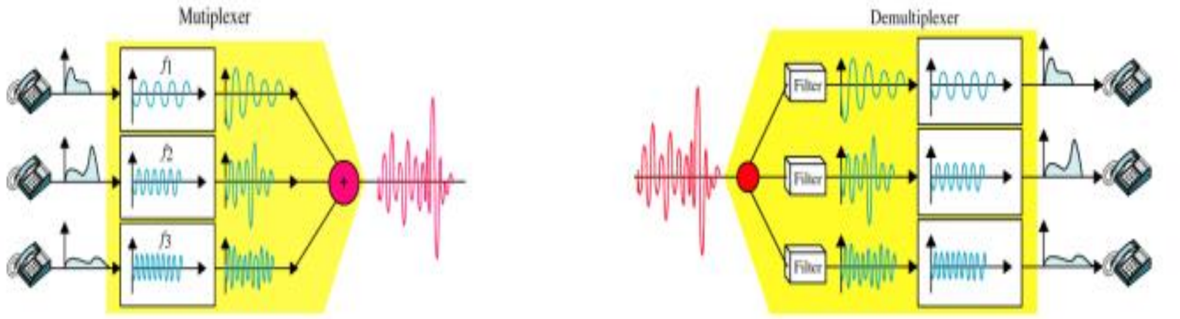


Fig 2.28 FDM Process

Advantage of FDM:

1. The senders can send signals continuously.
2. FDM support full duplex information flow
3. Works for analog signals too
4. Noise problem for analog communication has lesser effect
5. AM and FM radio broadcasting and Television broadcasting

Disadvantage of FDM:

1. Separate frequency for each possible communication
2. Inflexible, one channel idle and other one busy
3. The initial cost is high
4. A problem for one user can sometimes affect others
5. Each user requires a precise carrier frequency.