UNIT - II

Audio & Video compression

Audio compression:-

* The digitization process in known as pulse code modulation.

* This involves sampling the (analog) audio signal /waveform at a minimum rate which is twice that of the maximum frequency component that makes up the signal.

* Alternatively if the (frequency) bandwidth of the communications channel to be used is less than that of the signal,

* Then the sampling rate is determined by the bandwidth of the communication channel.

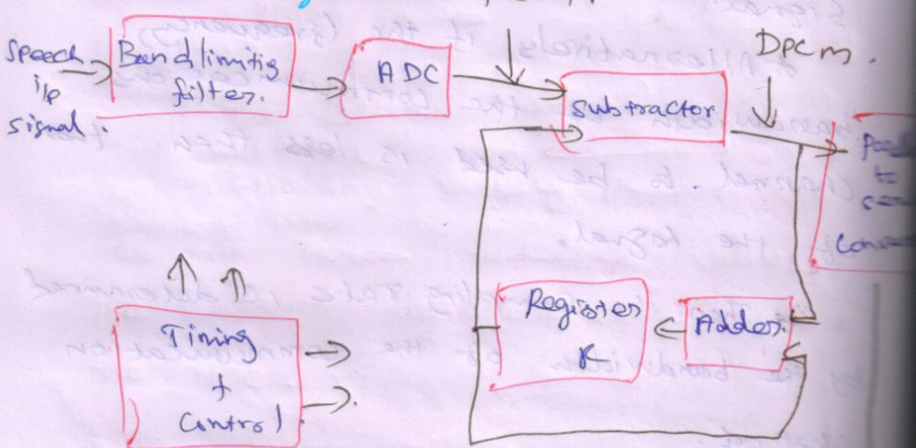
* The latter is then known as a Band limited signal.

* A speech signal the maximum frequency components is 10 kHz and hence the minimum sampling rate is 20 kbps.

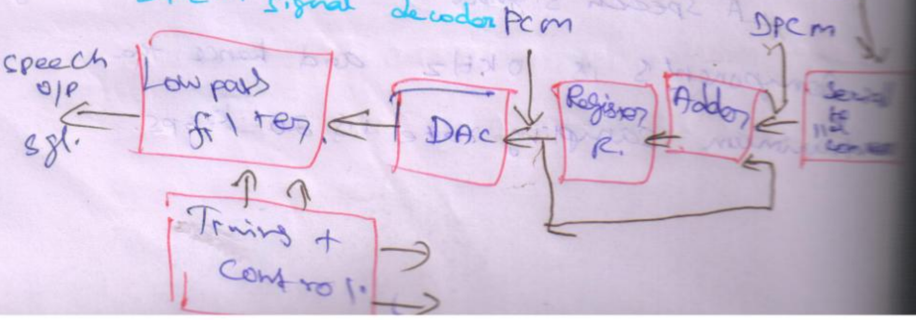
Differential pulse code modulation.

* DPCM is a derivative of stan PCM and exploits the fact that, fo most audio signals, the range of the differences in amplitude b/w successi Samples of the audio waveform is less than the range of the actual Sample amplitudes.

DPCM Signal encoder PCM.



DPCM signal decoder PCM



* Hence if only the digitized difference signal is used to encode the waveform then fewer bits are required than for a comparable PCM signal with the same sampling rate.

* As we can deduce from the circuit shown in figure. the output of the ADC is used directly and hence the accuracy of each computed difference signal - also known as the residual (signal) - is determined by the accuracy of the previous signal / value held in the register.

* This means therefore that with a basic DPCM scheme, the previous value held in the register is only an approximation.

* Hence more sophisticated techniques have been developed for estimating - also known as predicting.

* A more accurate version of the previous signal.

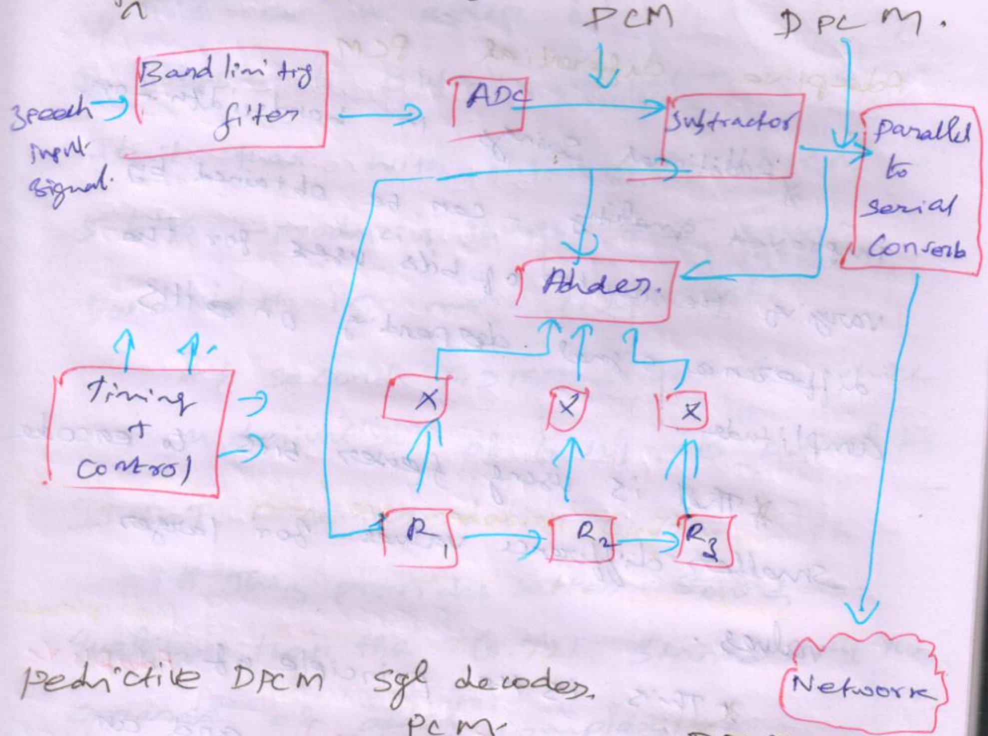
* To acheire this, these predict the previous signal by using not only the estimate of the current signal but also varying proportions of a number of the immediately preceding estimated sig

* The proportions used are determin by what are known as Predictor Co-efficients. and the principle is sho in figure.

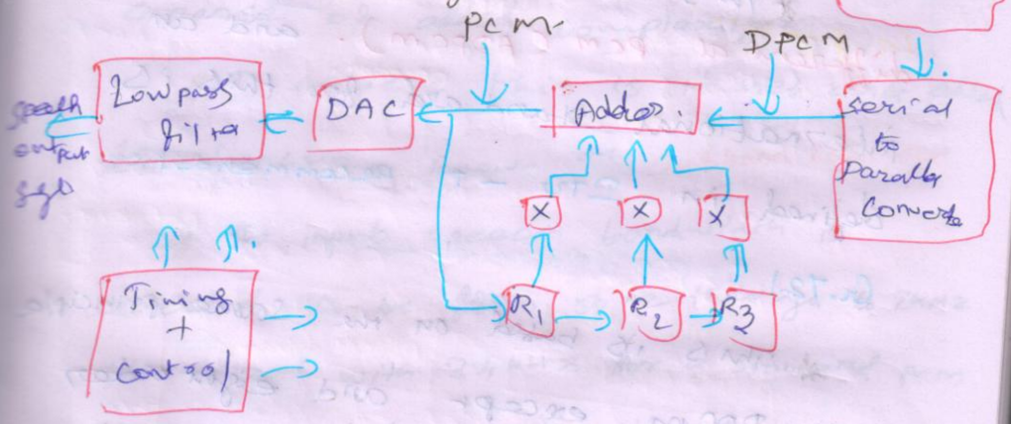
* The difference signal is compute by subtracting varying proportions of the last three predicted values from the current digitized value output by the ADC.

* For example if the three predictor coefficients have the valu $C_1 = 0.5$ and $C_2 = C_3 = 0.25$. then the contents of register R_1 would be shifted right by 1 bit. and register R_2 and R_3 by 2bits.

predictive DPCM sgl encoder.



predictive DPCM sgl decoder.



c_1, c_2, c_3 = predictor co-efficients.

Adaptive differential PCM.

* Additional savings in bandwidth - or improved quality - can be obtained by varying the number of bits uses for the difference signal depending on its amplitude.

* That is using fewer bits to enc[illegible] smaller difference values for larger values.

* This is the principle of Adapti[illegible] differential PCM (ADPCM). and an international standard for this is defined in ITU - T Recommendatio[illegible] G.721

* This is based on the same pr[illegible] as DPCM except and eight ord[illegible] predictor is used and the num[illegible] of bits used to quantize each di[illegible] value is varried.

* This can be either 6 bits - producing 32 k bps - to obtain a better quality output than with third-order DPCM or 5 bits - producing 16kpbs - if lower bandwidth is more important.

* A second ADPCM standard, which is a derivative of G.721. is defined in ITU-T Recommendation G.722.

* This provides better sound quality than the G.721 standard at the expense of added complexity.

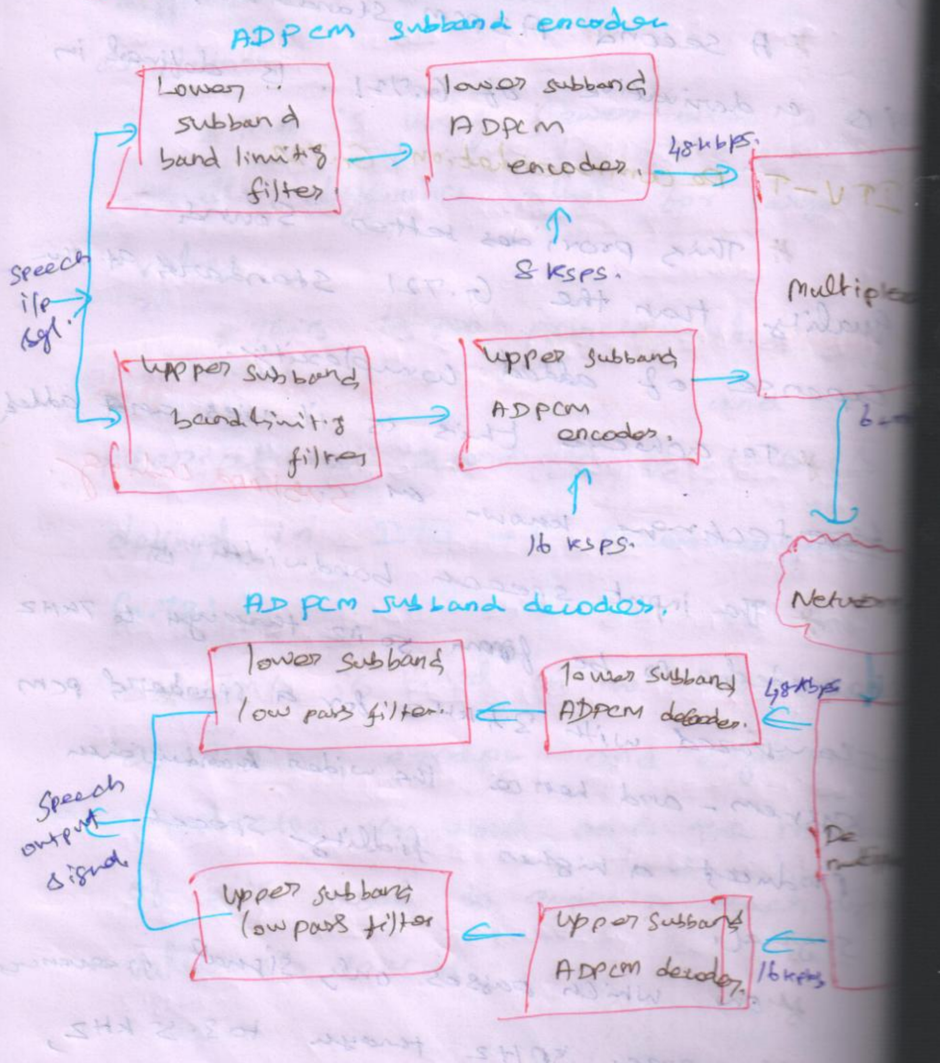
* To achieve this is it uses an added ~~the~~ technique known as subband coding.

* The input speech bandwidth is extended to be from 50 Hz through to 7kHz - compared with 3.4 kHz for a standard PCM system - and hence the wider bandwidth produces a higher fidelity speech signal.

* One which passes only signal frequencies in the range 50 Hz through to 3.5 kHz, and the other only frequencies. in the range

3.5 kHz through to 7kHz.

* By doing this the input signal (Speech) is effectively divided into two separate equal bandwidth signals, the first known as the lower subband signal & the second the upper subband signal.



* A third standard based on ADPCM is also available.

* This is defined in ITU-T Recommendation G.726.

* This also uses subband coding but with a speech bandwidth of 2.4 kHz.

* The operating bit rate can be 40, 32, 24 or 16 kbps.

Adaptive predictive coding.

* Even higher levels of compression - but at higher levels of complexity - can be obtained by also making the predictor coefficient adaptive.

* This is the principle of adaptive predictive coding (APC) and with this the predictor coefficients continously change.

* In practice, the optimum set of predictor coefficients continuously vary.

period:

This is the duration of the signal.

Loudness:

This is determined by the amo[unt] of energy in the signal.

* In addition the origins of the s[ound] are important. These are known as vocal tract excitation parameters and classified as:

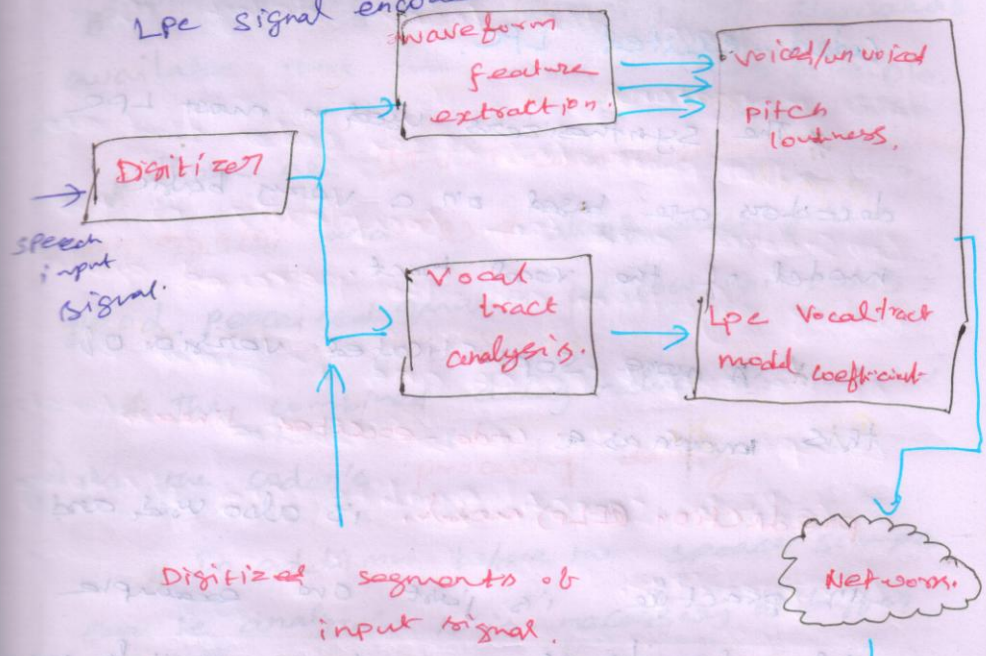
Voiced sounds:

These are generated through [the] vocal chords and examples includes [the] sounds relating to the letters m, v, [...]

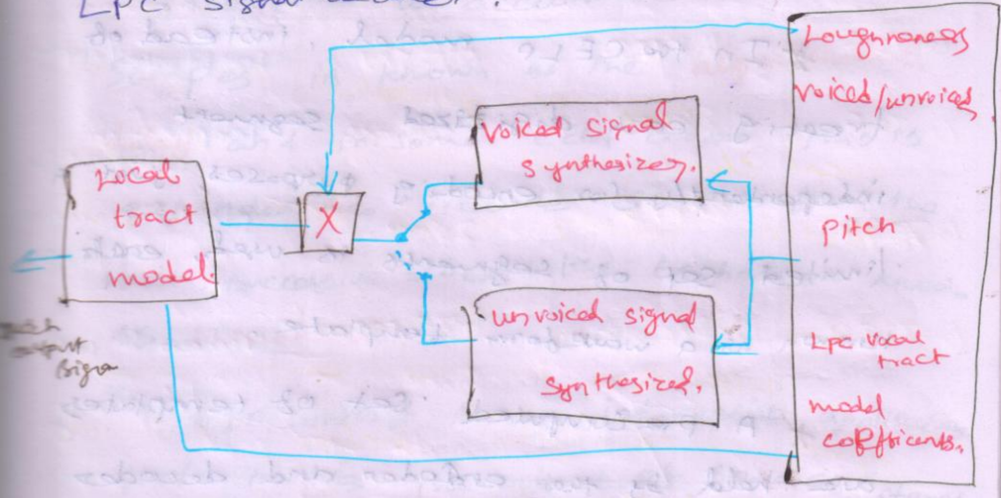
Unvoiced sounds:

With these the vocal chords are open and examples include the sou[nds] relating to the letters f and s.

LPC signal encoder.



LPC signal decoder.



Coded - excited LPC :-

* The synthesizers used in most LPC decoders are based on a very basic model of the vocal tract.

* A more sophisticated version of this known as a code-excited linear prediction (CELP) model, is also used in practice. is just one example of a family of vocal tract models as enchanced excitation (LPC) mo

* In the CELP model, instead of treating each digitized segment independently for encoding purposes, just limited set of segments is used, each known as a waveform template.

* A precomputed set of templates are held by the encoder and decoder in what is known as a template code

* There are now four international standards available that are based on this principle.

* These are ITU-T recommendations G.728, 729, 729(A) and 723.1 all of which give a good perceived quality at low bit rates.

* The combined delay value is known as the coder's processing delay.

* In addition before the speech samples can be analyzed, it is necessary to buffer - store in memory - the block of samples.

* The time to accumulate the block of samples is known as the algorithmic delay and in some CELP coders this is extended to include samples from the next successive block, the technique known as lookahead.

* In contrast in an interactive application that involves the output of speech stored in a file, for example, a delay of several seconds before the speech start to the output is often acceptable

and hence the coders delay is less important.

* other parameters of the coder that are considered are the complexity of the coding algorithm and the perceived quali[ty] of the output speech and in general a compromise has to be reached between a coder's speech quality and its del[ay] complexity.

Standard	Bitrate	total coder delay	example application domain
G.728	16 KBPS.	0.625 ms	Low bit rate telep[hony]
G.729	8 KBPS	25 ms	Telephony in cellular [illegible]
G.729(A)	8 KBPS	25 ms	Digital simul[taneous] Voi[ce]
G.723.1	5.3/6.3 Kbps.	67.5 ms	Video & inte[rnet] telephony

video compression - principle.

* Video (with sound) features in a number of multimedia applications:

Interpersonal

video telephony and video conferencing

Interactive

access to stored video in various forms.

Entertainment :-

digital television and movie / video - on-demand.

* In the context of compression, since video is simply a sequence of digitized pictures, video is also referred to as moving pictures and the terms "frames" and "picture" are used interchangabley

* In principle one approach to compressing a video source is to apply the JPEG algorithm described earlier ~~to~~ to each frame independently.

* This approach is known as moving JPEG or MJPEG.

H.261.

* H.261 video compression standard has been defined by the ITU-T for the provision of video telephony and video conferencing services over an integrated service digital network.

* Hence as we described earlier in it is assumed that the network offers transmission channels of multiples of 64 kbps

* The standard is also known therefore as p×64. Where p can be 1 through 30.

* The digitization format used is either the common intermediate format or the quarter CIF.

* Normally the CIF used for videoconferencing and the QCIF for video telephony, both of which in the section.

CIF : $Y = 352 \times 288$

QCIF : $Y = 176 \times 144$

$C_b = C_r = 176 \times 144$

$C_b = C_r = 88 \times 72.$

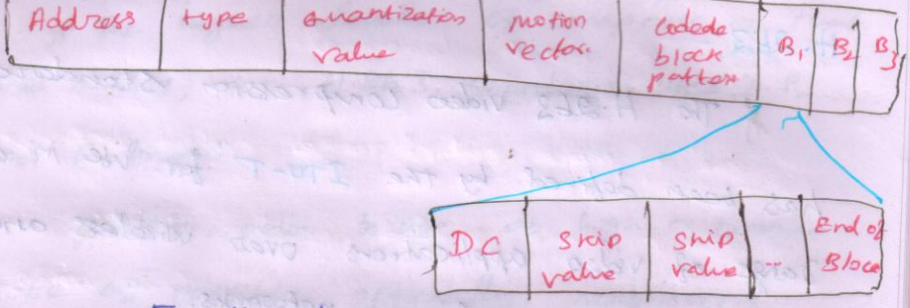


Fig. Macroblock format

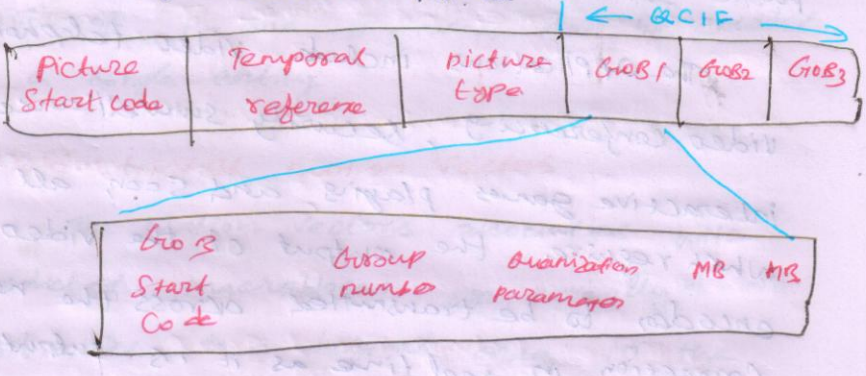


Fig:- Frame picture format.

* The start of each new (encoded) video frame/picture is indicated by the picture start code.

* This is followed by a temporal reference field which is a time stamp to enable the decoder to synchronize each video block with and associated audio block containing the same time stamp.

H.263:-

* The H.263 video compression standard has been defined by the ITU-T for use in a range of video applications over wireless and public switched telephone networks.

* The applications include video telephony video conferencing, security surveillance, interactive games playing, and so on, all of which require the output of the video encoder to be transmitted across the network connection in real time as it is output by the encoder.

Digitization format:

* The various digitization format associated with digital video in the H.263 standard, the two mandatory formats are the QCIF and the Sub-QCIF.

QCIF: $Y = 176 \times 144$

S-QCIF: $Y = 128 \times 96$

$C_b = C_r = 88 \times 72$

$C_b = C_r = 64 \times 48$.

Frame types

* The higher levels of compression that are needed, the H.263. Standard uses I, P- and B-frames.

* Also in order to use as high a frame rate as possible, optionally, neighboring pairs of P- and B frames can be encoded as a single entity.

Unrestricted motion vectors

* The motion vectors associated with predicted macroblocks are normally restricted to a defined area in the reference frame around the location in the target frame of the macroblock being encoded.

* In practice, with the small digitized frame formats that are used with the H.263 standard, this has been to give a significant improvement in the level of compression obtained.

Error tracking:-

* With real time applications such as video telephony, a two way communication channel is required for the exchange of the compressed

audio & video information generated by the codec in each terminal.

one or more out of range motion vector/s

one or more invalid variable-length code words.

one or more out of range DCT coefficients

An excessive number of coefficients within a macroblock.

MPEG-1

* Motion Pictures Expert group is defined in a series of documents which are all sub sets of ISO Recommendation 11172. The video resolution is based on the source intermediate digitization Format (SIF) with a resolution of upto 352×288 pixels

* The Standard is intended for the storage of VHS-quality audio and video on CD-Rom at bit rates up to 1.5Mbps

Normally however, higher bitrates of multiples of this are more common in order to provide faster access to the stored material.

NTSC : $Y = 352 \times 240$ $C_b = C_r = 176 \times 120$

PAL : $Y = 352 \times 288$ $C_b = C_r = 176 \times 144$.

* The standard follows the use of I-frames only I-and P-frames only or I-P-and B-frames, the latter being the most common.

* No-D-frames are supported in any of the MPEG standard and hence in the case of MPEG-1, I-frames must be used for the various random-access function associated with VCR's.

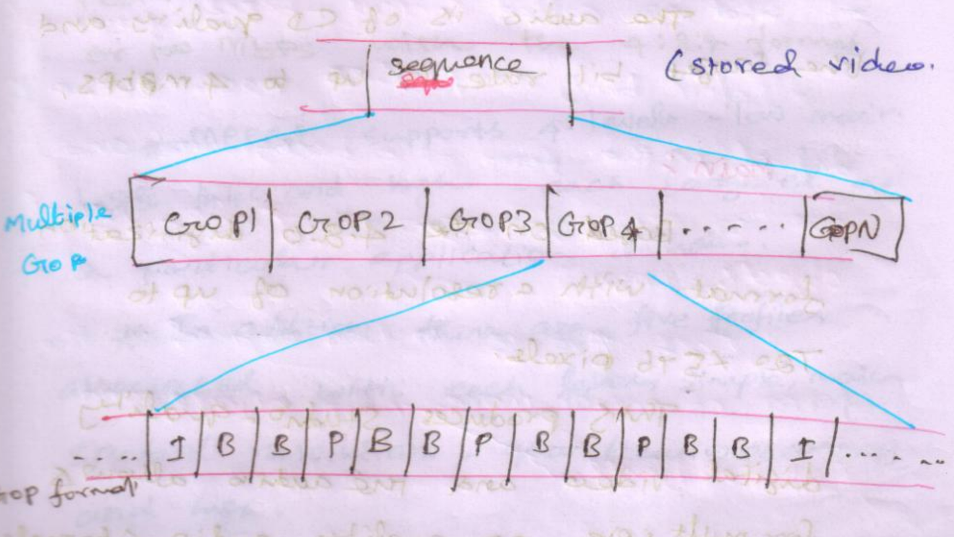


Fig. MPEG-1 video bit stream

MPEG-2 :-

* This is defined in a series of documents which are all subsets of ISO Recommendation 13818.

* It is intended for the recording and transmission of studio-quality audio and video

* The standard covers four levels of video resolution.

Low :- Based on the SIF digitization format with a resolution of up to 352 × 288 pixels. It is compatible with the MPEG-1 standard and produces VHS-quality video.

The audio is of CD quality and the target bit rate is up to 4 mbps.

Main :-

Based on the 4:2:0 digitization format with a resolution of up to 720 × 576 pixels.

This produces studio-quality digital video and the audio allows for multiple CD-quality audio channels

The target bit rate is upto 15 Mbps or 20 Mbps with the 4:2:2 digitization format.

High 1440:-
Based on the 4:2:0 digitization format with a resolution of 1440 x 1152 pixels.

It is intended for high difinition television (HDTV) at bit rates up to 60 Mbps or 80 Mbps with the 4:2:2 format.

High :-
Based on the 4:2:0 digitization format with a resolution of 1920 x 1152 pixels.

It is intended for wide screen HDTV at a bit rate of up to 80 Mbps or 100 Mbps with the 4:2:2 format.

* MPEG2 supports 4 levels - low main high 1440 and high - each targeted at a particular application domain.

* In addition there are five profiles associated with each level, simple, main spatial resolution, quantization accuracy and high.

MPEG-4

* Initially this standard was concerned with a similar range of applications to those of H.263, each running over very low bit rate.

Channel ranging from 4.8 to 64 kbps

* Later its scope was expanded to embrace a wide range of interactive multimedia applications over the internet and the various types of entertainment of network.

* The main difference between MPEG-4 and their other standards we have considered is that MPEG-4 has a number of what are called content-based functionalities.

* Before being compressed each scene is defined in the form of a background and one or more foreground Audio-visual objects (AVOs)

* Each AVO is in turn defined in the form of one or more video objects and Audio objects.