

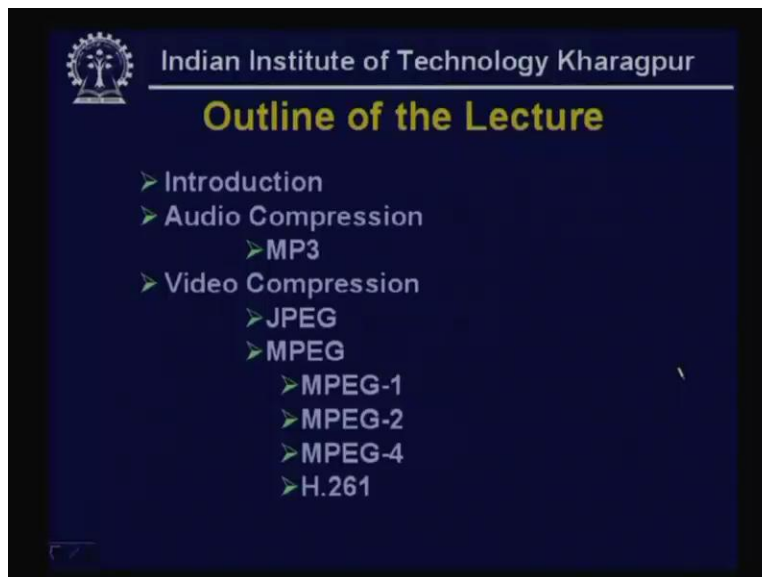
**Data Communication**  
**Prof. A. Pal**  
**Department of Computer Science & Engineering**  
**Indian Institute of Technology, Kharagpur**  
**Lecture - 37**  
**Audio and Video Compression**

Hello and welcome to today's lecture on audio and video compression. In the last lecture we have discussed in detail the bandwidth and SAS factors of different multimedia signal. We have also discussed the network performance parameters of different networks. It is quite evident from the discussion of the last lecture that sending multimedia signal through internet is not possible without compression. So in today's lecture I shall try to give an overview of the compression techniques.

Fortunately in the last couple of decades rigorous research in the area of compression technology has led to the emergence of sophisticated compression techniques leading to massive compression. This has made possible transmission of multimedia signals through internet. Hence discussion of the compression technique is very much essential but the subject is very vast it is not really possible to cover everything in one lecture. A full course of forty lectures can be taken on compression techniques. So in this lecture I shall essentially try to give an overview of audio and video compression.

Here is the outline of the lecture.

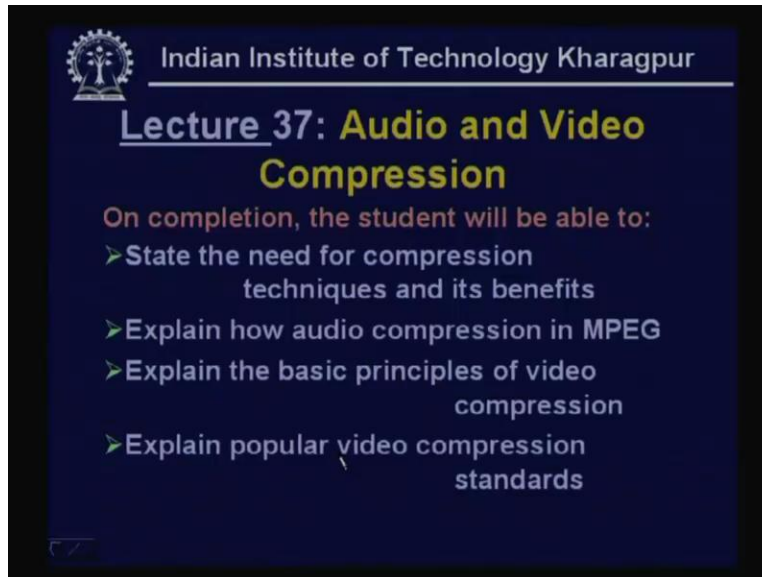
(Refer Slide Time: 02:28)



I shall give a brief introduction and then discuss the audio compression technique particularly what is being used in MPEG-3 that is MPEG version 3 then I shall discuss the video compression techniques; essentially there are two approaches JPEG and MPEG. MPEG has different versions like MPEG-1 MPEG-2 and MPEG-4 I shall give an

overview of these three versions and also the H.261 which is essentially used for low bit rate compression and I shall give an overview of that also.

(Refer Slide Time: 03:16)



Indian Institute of Technology Kharagpur

## Lecture 37: Audio and Video Compression

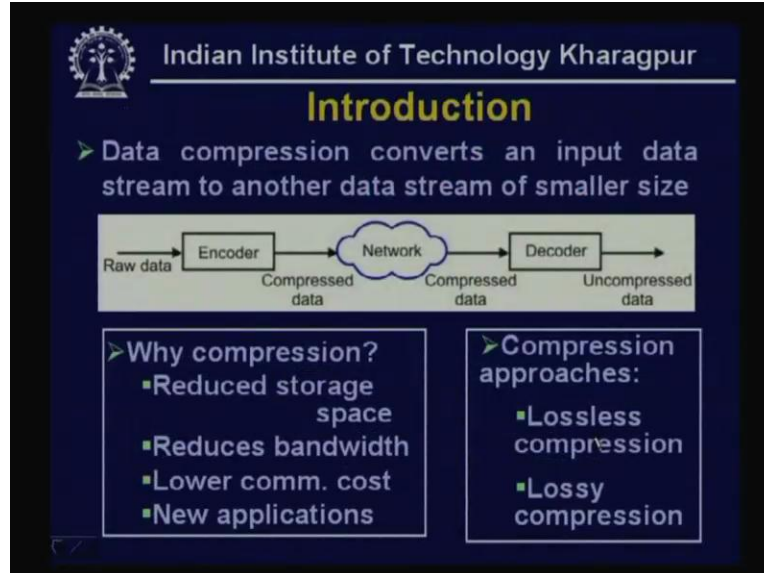
On completion, the student will be able to:

- State the need for compression techniques and its benefits
- Explain how audio compression in MPEG
- Explain the basic principles of video compression
- Explain popular video compression standards

On completion the student will be able to state the need for compression, why compression is needed and its benefits. They will be able to explain how audio compression in MPEG is performed, they will be able to explain the basic principles of video compression and they will be able to explain the popular video compression standard **as I have mentioned just now**.

Now first let us consider the definition of compression. What do we really mean by compression? By compression it essentially means that the compression converts an input data stream to another data stream of smaller size. So essentially it reduces the size of the data.

(Refer Slide Time: 03:45)



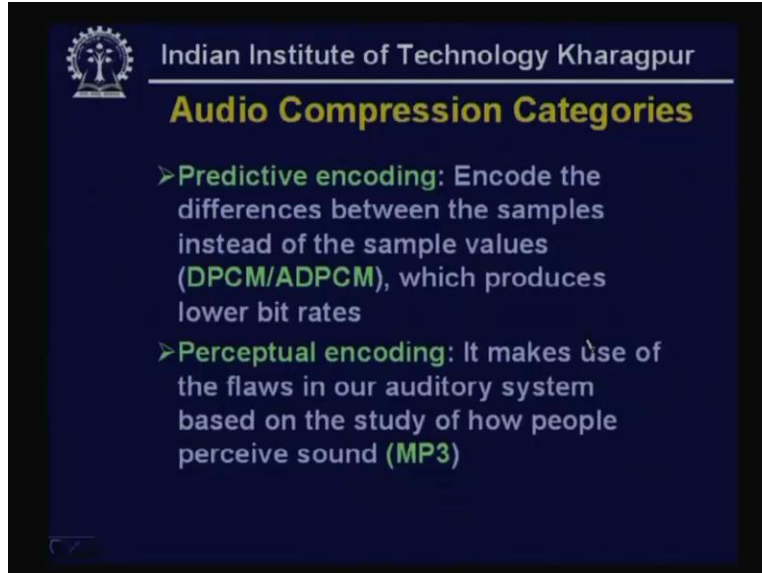
For example here we apply the raw data to a device known as encoder, the device which performs the compression is known as encoder and after this encoder performs the compression the compressed data is sent through the network and the compressed data is received at the other end by the decoder which converts back the compressed data to uncompressed data. So the need for compression is quite clear. First of all it reduces the storage space the amount of storage that we will require in uncompressed data is significantly larger than the compressed data. Similarly, it will reduce bandwidth for transmission through a network because of lesser rate bandwidth of the converted signal it will be possible to send through a network of smaller bandwidth or communication link of lower bandwidth so it will lead to lower communication cost. As you know the higher the bandwidth of the link more is the cost so lower bandwidth requirement reduces the cost of communication.

Moreover, it leads to emergence of new and new applications. The compression approaches can be broadly divided into two types; first one is known as lossless compression. That means after compression the data is reduced in size but it is done in such a way that at the other end the original data can be recovered so data in original form can be recovered without loss of any information.

On the other hand, there are some techniques in fact which are more common which are known as lossy compression. In such cases the uncompressed data may not be exactly same as the original raw data, the reason for that is our eyes and ears do not really recognize the differences, as a consequence we can afford to lose some information so even the uncompressed data in the lossy compression gives you good quality production although some information is lost. These are the two basic approaches we shall consider particularly the lossy compression techniques.

First I shall focus on audio compression.

(Refer Slide Time: 07:08)

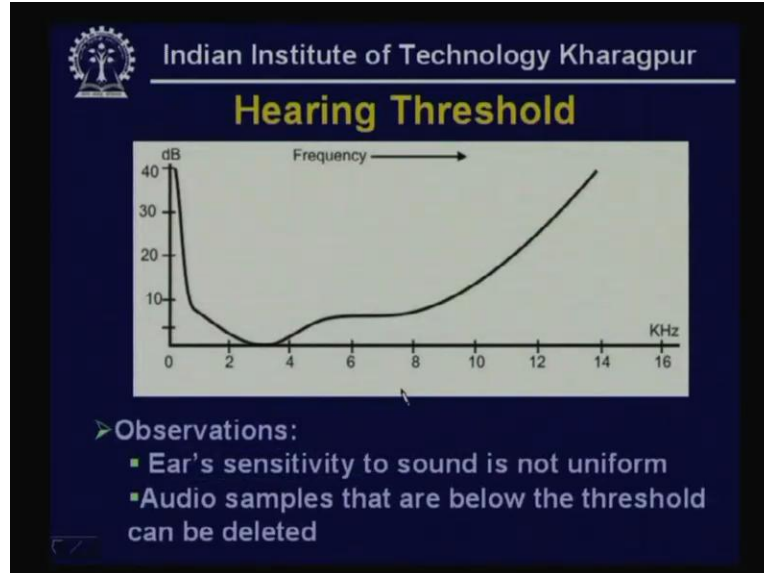


Audio compression techniques can be broadly divided into two types; one is known as predictive encoding another is known as perceptual encoding. Predictive encoding encodes essentially the differences between the samples instead of the absolute sample values that are encoded.

We have already discussed in detail the differential pulse code modulation or adaptive differential pulse code modulation where the differences between sample values are sent instead of the absolute values and this leads to reduction of lower bit rates so this is essentially the basic idea behind predictive encoding. However, the amount of compression that you can achieve by using predictive encoding is not very high that's why the perceptual encoding is more common.

You will be surprised to know that it makes use of the flaws of our auditory system based on the study of how people perceive sound. So it essentially exploits the flaws and limitations of our **ears to do** not hear everything equally. It is based on the science of understanding how people perceive sound and based on that the encoding is performed. Let us see the hearing threshold of a normal person. As you can see this is the frequency range 0 Hz to 16 KHz.

(Refer Slide Time: 08:54)



Over this range as you can see 2 KHz to 4 KHz our ear is most sensitive to this region and attenuation increases as we go towards lower frequency as well as we go towards higher frequency. Our ear is very insensitive to high frequencies like 14 KHz or 16 KHz or 20 KHz however the situation is different for dogs.

Dogs can even hear ultra sound, such a higher frequency signal but for human beings our ear is not really sensitive to these higher frequencies. This is how it is expressed in decibel. This is 40 db and this is 0 db so our observation is, ear sensitivity to sound is not uniform, it won't hear all the frequencies equally, so audio samples that are below the threshold can be deleted. So what is the point of sending signals with amplitude in this range of higher frequency? In any case the ear will not hear them so the basic idea is it is better to discard and save bandwidth of the link. This is one approach. Another is, some sounds can mask other sounds. It has been observed that some sounds can mask other sounds.

(Refer Slide Time: 10:41)

Indian Institute of Technology Kharagpur

## Lossy Audio Compression

- Some sound can mask other sounds
- **Frequency Masking:** A loud sound in a frequency range can partially or fully masks another sound in the nearby frequency range
- **Temporal masking:** A loud sound can numb our ears for a short duration even after the sound has stopped

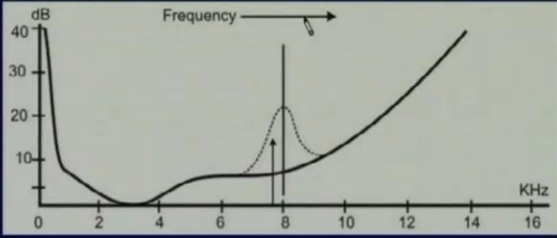
There are two types of masking. One is known as frequency masking. It has been observed that a loud sound in a frequency range can partially or fully mask another sound in the nearby frequency range. For example, here it is shown (Refer Slide Time: 11:10) if there is loud sound in this frequency say 8 KHz then how the threshold is increases shown by this dotted line. So, if there are sounds with amplitude in this range it will not be audible to the ear. Hence, the threshold is being raised by the presence of a loud signal of a particular frequency, this is known as frequency masking.

(Refer Slide Time: 11:05)

Indian Institute of Technology Kharagpur

## Frequency masking

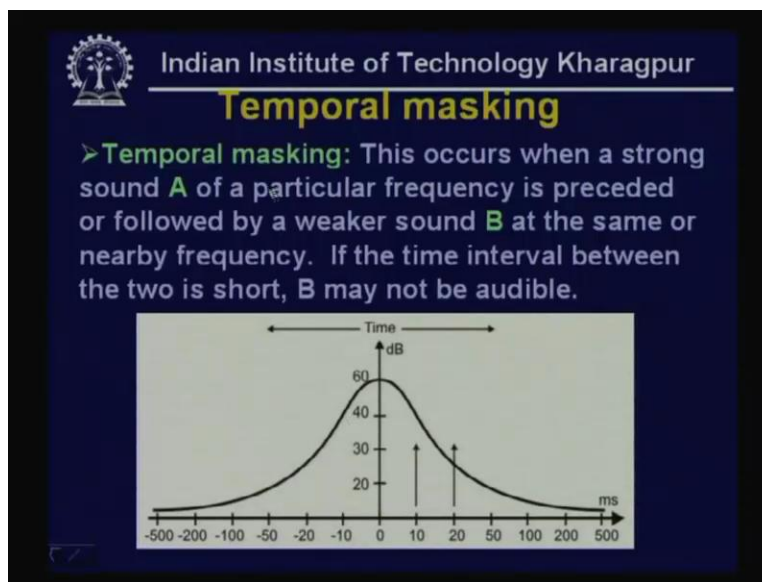
- **Frequency masking:** This occurs when a sound that we can normally hear is masked by another sound with a nearby frequency.



Another type of masking is possible which is known as temporal masking. In temporal masking a loud sound can numb our ears for a short duration even after the sound has stopped. So, if there is a loud sound at particular time instant then some sounds around that time will not be audible to the ear. for example, whenever a train passes in the nearby region by blowing whistle or a plane crosses over your head then because of the large background sound we cannot talk to each other or when if you are taking a class it is difficult for the teacher to make it audible to the students. This we have observed in our day to day life and we shall see how they can be exploited to achieve compression.

Here for example the temporal masking is explained in detail.

(Refer Slide Time: 12:40)




There is a strong sound A that has occurred at time instant 0 of a particular frequency preceded or followed by a weaker sound B at the same or nearby frequency and if the time interval between the two is short B may not be audible. For example, this signal (Refer Slide Time: 13:09) will not be audible because it is within 10 milliseconds and it is about 32 db. On the other hand the threshold has been raised to 60 db. You can see threshold is suddenly raised and it falls gradually.

However, after 20 milliseconds it can be heard because it comes above the threshold value. So depending on where the sound is taking place with respect to the loud sound it may be audible or it may not be audible so a particular person may or may not hear depending on when it is occurring. So the sum of the signals in this range which are present can be removed from the signal for transmission. This is how compression can be done.

What is being done is the critical bands are determined according to the sound perception of the ear. So the entire range is divided as 27 bands here.

(Refer Slide Time: 14:01)



Indian Institute of Technology Kharagpur

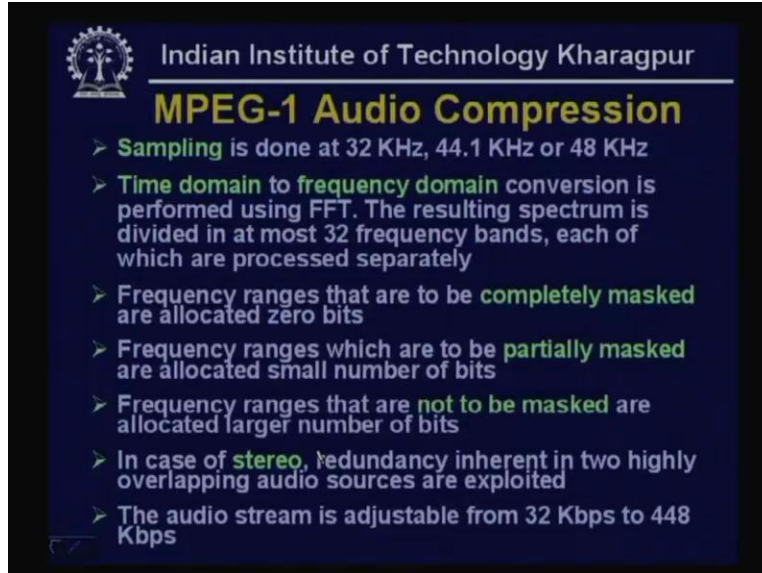
Critical Bands are determined according to the sound perception of the ear (Bark)

band	range	band	range	band	range
0	0-50	9	800-940	18	3280-3840
1	50-95	10	940-1125	19	3840-4690
2	95-140	11	1125-1265	20	4690-5440
3	140-235	12	1265-1500	21	5440-6375
4	235-330	13	1500-1735	22	6375-7690
5	330-420	14	1735-1970	23	7690-9375
6	420-560	15	1970-2340	24	9375-11625
7	560-660	16	2340-2720	25	11625-15375
8	660-800	17	2720-3280	26	15375-20250

Of course the range is expressed in a by using a unique known as bark by the name of by the name **HG bark bouson** a German Scientist and in his name it is being expressed. So, for frequencies below 500 Hz 1 bark is equal to  $8/100$ . On other hand for frequencies with above 500 Hz then 1 bark is equal to  $9 \text{ plus } 4/4 \log 8/1000$ . So, at higher frequencies the frequency range is more in a particular band. Thus, in band 0 it is only 0 to 50 Hz, in band 1 it is 50 to 95 Hz. On the other hand, in band 10 it is 940 to 1125 Hz, in band 19 it is 3840 to 4690. On the other hand, in band 26 it is 15375 to 20250 KHz. Therefore, the range is much larger in higher bands than the lower bands. Essentially because of the sensitivity of the ear is lower in higher frequency bands this is how it is being divided.



(Refer Slide Time: 15:44)



Indian Institute of Technology Kharagpur

## MPEG-1 Audio Compression

- Sampling is done at 32 KHz, 44.1 KHz or 48 KHz
- Time domain to frequency domain conversion is performed using FFT. The resulting spectrum is divided in at most 32 frequency bands, each of which are processed separately
- Frequency ranges that are to be completely masked are allocated zero bits
- Frequency ranges which are to be partially masked are allocated small number of bits
- Frequency ranges that are not to be masked are allocated larger number of bits
- In case of stereo, redundancy inherent in two highly overlapping audio sources are exploited
- The audio stream is adjustable from 32 Kbps to 448 Kbps


Now let us see how MPEG-1 performs the audio compression based on the techniques that I have discussed. First of all sampling is done at 32 KHz, 44.1 KHz or 48 KHz. of course the 44.1 KHz is very common for CD quality audio. Then the signal is converted from time domain to frequency domain using first Fourier transform. The resulting spectrum is divided in at most 32 frequency bands each of which is processed separately. Processed separately means as we have seen each of the different frequency bands are removed or reduced in a different way or said to be sent with a different resolution.

For example, frequency ranges that are to be completely masked are allocated zero bits, frequency ranges which are to be partially masked are allocated smaller number of bits. On the other hand, frequency ranges that are not to be masked are allocated large number of bits. So, higher precision is allocated to frequency ranges to which our ear is more sensitive and lower frequency to less sensitive and zero bits which are not at all audible to the ear.

Of course, in addition to that in case of stereo redundancy inherent in two highly overlapping audio sources are exploited. So whenever we are recording stereophonic sound two channels have lots of commonality or redundancy that can be exploited and that redundancy inherent in two overlapping audio sources are exploited for further compression. By doing MPEG-1 the audio stream is adjustable from 32 Kbps to 448 Kbps depending on the bandwidth of the source and different frequency components present in the signal.

And as we can see whenever it is uncompressed voice 8 KHz sampling frequency and a resolution of 8 bits per sample gives you 64 Kbps.

(Refer Slide Time: 18:00)



Indian Institute of Technology Kharagpur

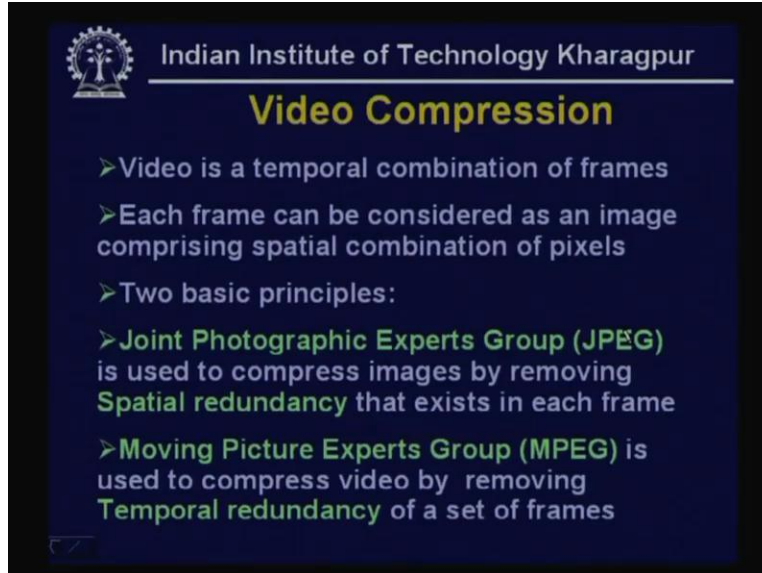
### Bandwidth Reduction

Audio Quality	NC	Sampling Frequency	Resolution (bits)	Bandwidth
Voice (UC)	1	8 kHz	8	64 Kbps
Voice (C)	1	8 kHz	8	4-32 Kbps
CD (UC)	2	44.1 kHz	16	1.411 Mbps
CD (C)	2	44.1 kHz	16	64-192 Kbps

On the other hand, whenever voice is compressed with the sampling frequency and the resolution can lead to 4 to 32 Kbps so 8:1 or 2:1 compression is possible. For CD quality audio if it is uncompressed, as we know, because of two channels and sampling at 44.1 KHz with resolution of 16 bits per sample it gives you 1.411 Mbps and whenever you compress it, it gives you 64 to 192 Kbps which is the amount of compression that is possible for audio signal.

Now let us focus on video compression. Video is essentially a temporal combination of different frames and each frame can be considered as an image. Image means it can be considered as a still image which comprises spatial combination of pixels.

(Refer Slide Time: 19:03)



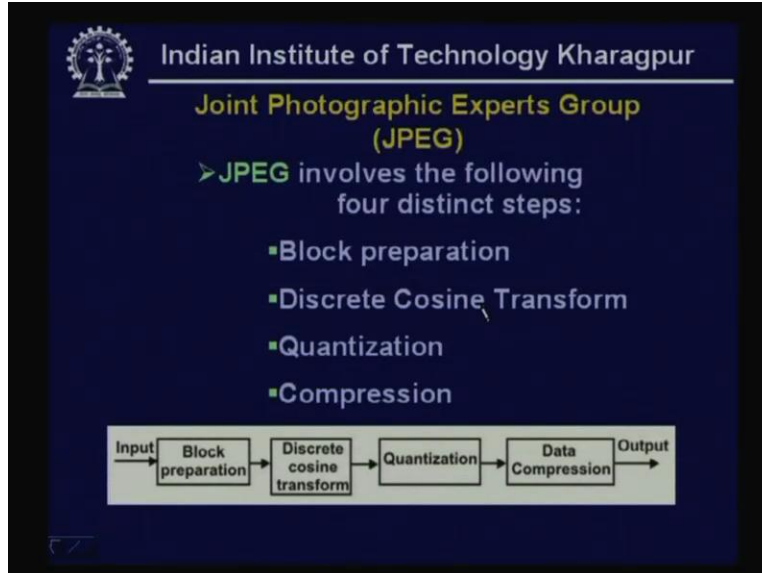
Indian Institute of Technology Kharagpur

## Video Compression

- Video is a temporal combination of frames
- Each frame can be considered as an image comprising spatial combination of pixels
- Two basic principles:
  - **Joint Photographic Experts Group (JPEG)** is used to compress images by removing **Spatial redundancy** that exists in each frame
  - **Moving Picture Experts Group (MPEG)** is used to compress video by removing **Temporal redundancy** of a set of frames

Now two basic principles are used. one is known as joint photographic experts group standard that is JPEG standard that is primarily used to compress images by removing spatial redundancy that exist in each frame. So each frame is considered as a still picture and spatial redundancy present in it is reduced with the help of JPEG. On the other hand, MPEG which is Moving Picture Experts Group standard is used to compress video by removing the temporal redundancy of a set of frames because the differences between two adjacent frames can be very very small so that is being exploited in temporal redundancy. These are the two techniques used. First let us focus on JPEG. JPEG involves the following four distinct steps.

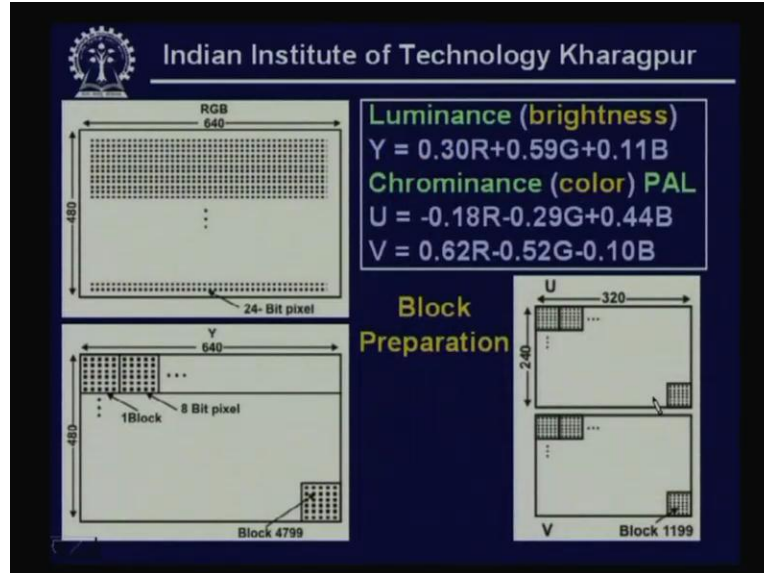
(Refer Slide Time: 20:22)



First one is block preparation and second one is discrete cosine transform, third is quantization fourth is compression. These are performed one after another to generate the compressed output. The raw data is received here then block preparation, discrete cosine transform, quantization and data compression is performed to generate the output signal.

First let us consider the block preparation. As you know after the video signal is digitized it is converted into an array of pixels for example for CD quality or VCR quality video with the typical number of pixels in horizontal direction 640 and in the vertical direction 480 so you have got 640/480 pixels and each of the pixel has got RGB components Red, Green and Blue components each is represented by 8 bits. This leads to twenty four bits per pixel. This is how still image is digitized.

(Refer Slide Time: 20:52)



However, before performing any compression approach, before performing any operation it is converted into luminance and chrominance component. The reasons for translating RGB to luminance and chrominance is our eyes are more sensitive to luminance than chrominance so chrominance can be sent with lesser resolution. Secondly by converting to luminance and chrominance components it allows more compression than it is present in RGB that's why the video digital signal has converted into Y U V components.

Of course each pixel has got again, in place of RGB it has got Y U and V components. The formula that is being used for conversion is shown here (Refer Slide Time: 22:47). Whenever you will be playing at the other end the receiving end the inverse conversion has to be done to get back the RGB. This conversion to luminance and chrominance has got another benefit. It makes it compatible with black and white technology. For example, if the luminance component is sent to a black and white receiver we shall get that black and white picture corresponding to a color picture.

Now let us see how the block preparation is done. Each pixel is represented by Y U V and it is divided into 8/8 pixels. In each block it has got 8/8 pixel and each block is processed separately so that the computation required for compression is minimized, that is the basic objective of this block preparation. So the Y component which is very important and to which our eyes are more sensitive as you can see the number of pixels remains unaltered so it has got 4800 blocks. So after the block preparation the Y component has got 4800 blocks and each block can be processed separately.

On the other hand, the U and V blocks (Refer Slide Time: 24:33) they are average, I mean 4 pixels are average to generate 1 pixel and U and V both are represented by 320 by 240 after averaging out. So, after averaging out we get 320/240 so this significantly reduces the size and in turn the number of blocks to be processed is reduced to 1200 for each.

So 1200, 1200, 1200 for U and 1200 for V and 4800 for Y so altogether 7200 blocks are to be processed. Let's see what kind of processing is done and see how each block is processed. Each block of 64 pixels that is 8/8 goes through a transformation called discrete cosine transform.

(Refer Slide Time: 25:24)

Indian Institute of Technology Kharagpur

## Discrete Cosine Transform

➤ Each block of 64 pixels goes through a transformation called DCT

Case-1

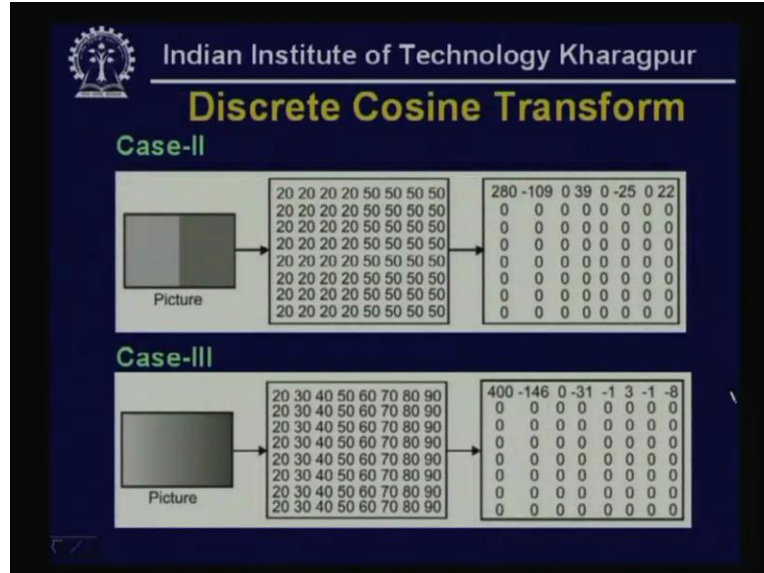
Picture	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	30	30	30	30	30	30	30	30	30
	240	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0

This discrete cosine transform is highly mathematical. I shall give three examples to show how it helps to minimize the information or how the numbers reduce.

Ultimately we have got a block where there are 8 into 8 is equal to 64 numbers. Now as it is it is necessary to send 64 numbers each of 8 bit. However, suppose there is a block where there is uniform intensity there is no change. By using discrete Fourier transform it is converted into a frame as you can see here. We have got only one DC component and all the AC components which essentially represent changes with respect to this 00. And AC components are 0 because with respect to that there is no difference in any other pixel. So essentially it has got one DC component and all DC components are 0. So you can see these zeros can ultimately send in a very compact form as we shall see later.

Then let us consider the second case where you have got two different tones that means two different intensities, 1 and 2. And as you can see intensity level is represented by 20 20 20 and the other half is 50 50 50 50.

(Refer Slide Time: 27:02)



So after discrete cosine transform as you can see it has got one DC component and few AC components and a large number of zero values. So the number of zeros is significantly increased and you have got very few AC components and DC components.

Now, if the intensity changes uniformly as you can see here it is 20 30 40 50 60 70 80 90 then also after discrete cosine transform it generates numbers like 400 which is essentially average value with a multiplication factor and other AC values like 146 and so on so these are the AC components and here also you see you have got large number of zeros. Of course in real life picture you will not have a uniform tone or just two tones or a gradually increasing tone but there will be some variation of course the number of AC components will be larger than this but definitely it will have a large number of zeros. So this is how the discrete cosine transform is performed and it helps to get large number of zeros in different pixel points.

(Refer Slide Time: 28:49)

**Quantization**

Indian Institute of Technology Kharagpur

1	1	2	4	8	16	32	64
1	1	2	4	8	16	32	64
2	2	4	8	16	32	64	
4	4	4	8	16	32	64	
8	8	8	8	16	32	64	
16	16	16	16	16	32	64	
32	32	32	32	32	32	64	
64	64	64	64	64	64	64	

160	95	60	23	6	4	0	0
90	85	40	15	5	3	0	0
56	48	32	12	5	2	0	0
15	11	10	7	4	3	0	0
5	3	2	1	1	1	0	0
3	2	2	1	1	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

160	95	30	6	1	0	0	0
90	85	20	4	1	0	0	0
28	24	16	3	1	0	0	0
4	3	3	2	1	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Now another step is performed which is known as quantization which further increases the number of zeros. As we have seen this DC component is the most important component and along that other components are important. However, the pixel components on this corner (Refer Slide Time: 29:12) are not really that much important so they are divided by different numbers and then the fraction part is removed.

For example, these four pixel values are divided by 1 and the values in this row and this column are divided by 2. Then this row and this column the values are divided by 4 and in this way the last row and last column is divided by 64 and as a consequence for example if this is the original DCT Discrete Cosine Transform coefficients after performing quantization you can see the number of zeros has increased. For example in this range we have more number of zeros compared to this one (Refer Slide Time: 29:58) so this helps you to increase the number of zeros after quantization.

And in fact the MPEG is lossy because of these quantization steps. The prior steps namely DCT and block preparation are not essentially lossy. But this is where some information loss takes place because of quantization. However, our eyes will not be able to detect the differences. After we have got the blocks each block is scanned in this zig zag fashion, this one, this one, this one, then this one, this one, this one, (Refer Slide Time: 30:40) in this way.



(Refer Slide Time: 30:32)

Indian Institute of Technology Kharagpur

➤ A zigzag scanning pattern is used to concentrate all the 0's together. The runs of 0's can be replaced by a single count (say, 38 0's)

**Run-length Encoding**

160	95	30	6	1	0	0	0
90	85	20	4	1	0	0	0
28	24	16	3	1	0	0	0
4	3	3	2	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

21000000000000000000000000000003 => 21A0263

You may be asking what is the purpose of this. The purpose is if you scan in this manner then the runs of zeros will be lesser than if you scan in this zig zag fashion.

For example, if you scan in horizontal manner the number of zeros in consecutive locations will be 24 plus 7 is equal to 31. On the other hand if you scan in this zig zag manner then as you can see from here all are zeros. So you have got 38 zeros in a row that is called runs of zeros. Essentially it helps you to get more runs of zeros. Then these runs of zeros can be sent in a compact form as it is shown here.

For example, this 26 here there are 26 zeros and these 26 zeros can be sent by this part of the information. So a block is now converted into fewer numbers 2 1 a 0 26 3 so initially we had 64 numbers now it has got converted into fewer numbers and that can be transmitted to the other end where coding can be done to get back the original signal. This is the basic concept behind the JPEG. Now let us see what is performed by MPEG.

(Refer Slide Time: 32:19)

Indian Institute of Technology Kharagpur

## MPEG-1

- The first standard to be finalized for video compression was MPEG-1 for interactive video on CDs and for digital audio broadcasting

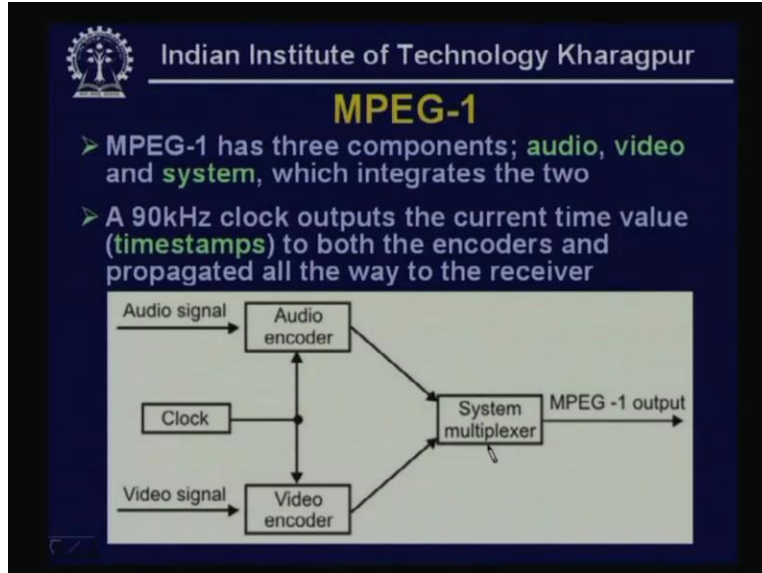
VCR	640	480	24	25	368.64 Mbps	MPEG-1 1.5Mbps
-----	-----	-----	----	----	-------------	-------------------

- It is likely to dominate the encoding of CD-ROM based movies
- It can be transmitted over twisted-pair for modest distances

MPEG-1 was the first standard to be finalized for video compression for interactive video on CDs and for digital audio broadcasting. With the help of MPEG-1 a VCR quality video having 640/480 pixels with 25 frames per second and 24 bits per pixel gives you 368.64 Mbps in uncompressed form, this is in uncompressed form. And after performing MPEG-1 compression it produces 1.5 Mbps so there is tremendous compression ratio, significant reduction in the size which can be sent through many networks. And this 368.64 is very difficult to send through many networks, it cannot be sent. And this MPEG-1 is likely to dominate the encoding of CD-ROM based movies because it gives you quite good quality performance and another advantage is this 1.5 Mbps can be transmitted over twisted-pair for modest distances. For example, it can be transmitted through ADSL network quite efficiently so by using ADSL twisted-pair of network it can be sent over a distance of 18000 ft or roughly 5 km. because of this compression it is possible to send MPEG video through ADSL.

MPEG has three components; audio, video and system. We have already discussed how the audio compression is done.

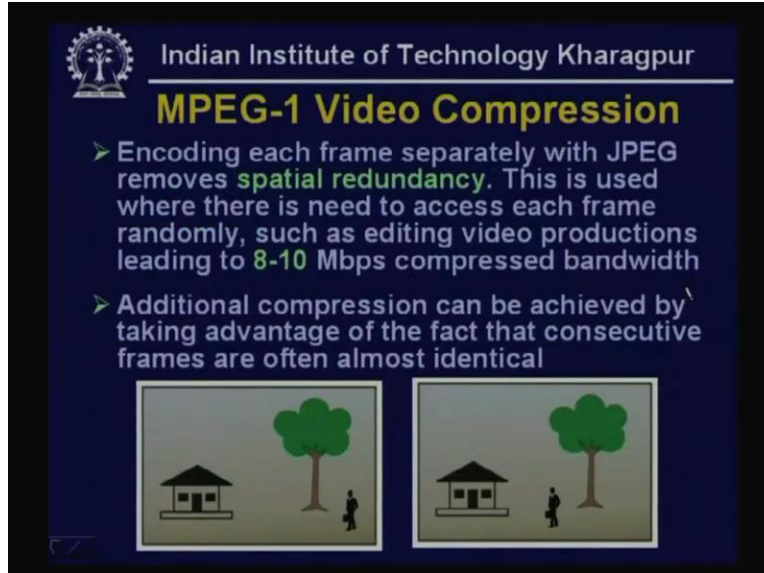
(Refer Slide Time: 34:08)



Audio signal is applied to the audio encoder which does the compression independent of the video encoder. So video signal is applied to the video encoder and audio signal is applied to the audio encoder after the sampling. Now a clock is used which operates at 90 KHz to provide the information in the form of time stamp. So this time stamping is performed and this time stamped audio and video signals are multiplexed to generate MPEG-1 output which is propagated all the way to the receiver. So this time stamping helps you to do the synchronization of audio and video signal. As we have seen one of the important SAS factor is lip synchronization, one of the important requirement is lip synchronization. So to facilitate lip synchronization this kind of time stamping is necessary and this time stamping is also necessary for streaming as we shall see in the next lecture.

Let us see how MPEG video compression is performed. Encoding each frame separately with JPEG removes spatial redundancy. We have already seen how JPEG can be used to remove spatial redundancy. However, just JPEG encoded frames can be sent without further compression. This can be done in situations where each frame is accessed randomly such as when you are editing for video production. However, in such a case after compression you will get 8 to 10 Mbps compressed bandwidth and not very high compression.


(Refer Slide Time: 35:31)



Indian Institute of Technology Kharagpur

## MPEG-1 Video Compression

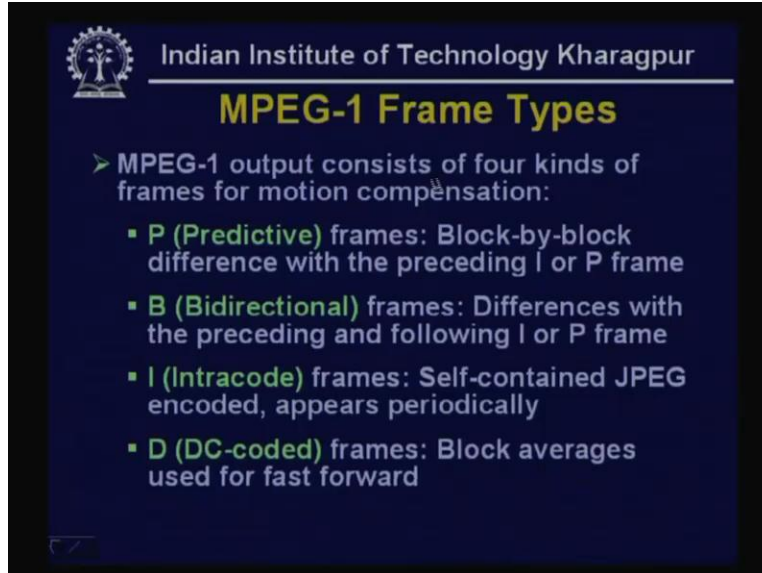
- Encoding each frame separately with JPEG removes **spatial redundancy**. This is used where there is need to access each frame randomly, such as editing video productions leading to **8-10 Mbps** compressed bandwidth
- Additional compression can be achieved by taking advantage of the fact that consecutive frames are often almost identical



However, additional compression can be achieved by taking advantage of the fact that consecutive frames are often almost identical that is temporal redundancy. For example, two frames are shown here (Refer Slide Time: 36:38) this is frame one and this is frame n. As you can see here in this frame the background is identical, this house, this tree this all are same except the blocks in which this person is there and the blocks in which this person is here. So in this region and in this region only in these two regions there are differences. So, if this frame is used to reconstruct this frame then only information of this block and this block is sufficient to encode this frame and also at the receiving end to regenerate this frame from this frame. This is the basic idea behind MPEG.

To do that MPEG-1 output consists of four kinds of frames for motion compensation. Essentially the difference between two frames are to be compensated which is known as motion compensation. This is done by sending four different types of frames.

(Refer Slide Time: 37:28)

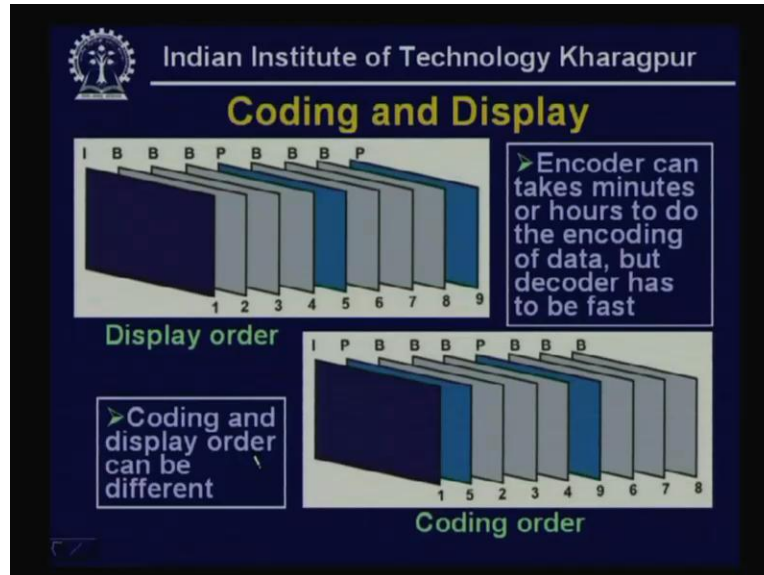


First one is I type of frame and this I type of frames are essentially intracode frame, these are self contained JPEG encoded which appears periodically. So they are essentially JPEG encoded and contain all the information of the JPEG encoding and they can be decoded independently. On the other hand, the P type the predictive frames use the block by block difference with the preceding I or P frames. So a P type of frame which is predictive frame cannot be decoded independently which are not self contained so preceding I and P frames will be required to encode P frames and also regenerate or decode the P frames.

Another type of frames is sent known as bidirectional frames. In bidirectional frames differences with the preceding and also the following I and P frames are used as references. These are the three basic frames used. However, this D type of frames is essentially block averages which are essentially used for fast forward but not really used for compression.

Now let's us see how I, P and D frames are used. Here for example this is the order in which different frames are to be displayed. This is the I frame, 2, 3 and 4 are B frames (Refer Slide Time: 39:22) that means 2 depends on 1 as well as on 5, 3 depends on 1 as well as on 5, 4 also depends on 1 as well as on 5. On the other hand, 5 depends only 1 the previous 1 and not the succeeding following 1, 6 depends on 1 as well as 9, 7 depends on 1 as well as 9, 8 depends on 1 as well as 9 on the other hand 9 depends only on 1 and may be on 5.

(Refer Slide Time: 39:09)



Now, because of this dependency what is being done is the encoder does not send the frames in this order the way they are to be displayed. What is being done is first 1 is sent then 5 is sent because as you can see 2, 3 and 4 you will require the information of 1 as well as 5 so it is necessary to receive the frame 5 before 2, 3 and 4 can be reconstructed at the receiving end that's why 1 is sent then 5 is sent then 2, 3, 4 similarly 9 is sent before 6, 7 and 8. So you can see the display order and the coding order is different. So the encoder generates the frames in this form which is being received by the decoder and the decoder after doing necessary processing will again display in a natural sequence like 1, 2, 3, 4 etc.

So one observation is coding and display order can be different. Second is encoder takes minutes or hours to do the encoding of the data but decoder has to be fast. So it has been observed that encoder algorithm is quite complex so it may take minutes or hours for encoding purpose. However, at the receiving end the decoder has to do on the fly processing and display the images that's why the decoder has to be pretty fast. So there is a difference in the complexity of encoder and decoder. Now let us consider the MPEG-2.

(Refer Slide Time: 42:01)

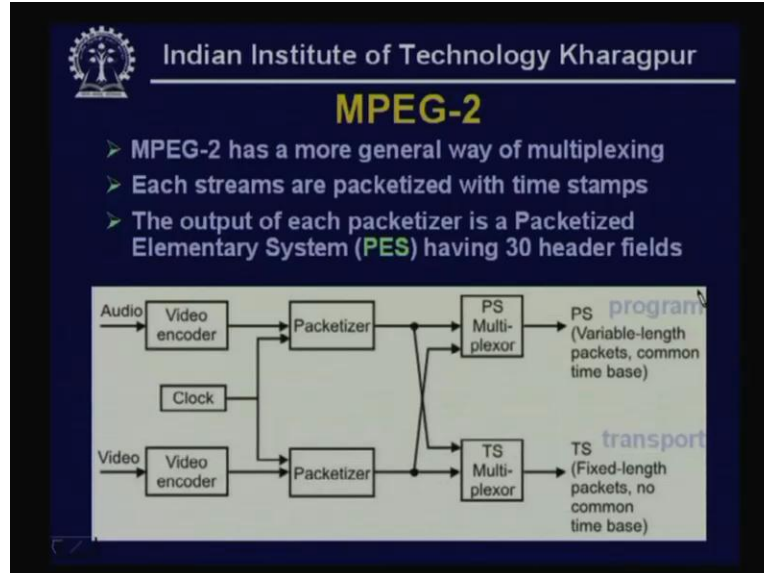
Indian Institute of Technology Kharagpur

- Although MPEG-2 is similar to MPEG-1, it was developed for digital television
- Differences: **MPEG-2**
  - D frames are not supported.
  - DCT is **10X10** instead of 8X8 to provide better quality
  - Supports four resolutions
  - Supports five profiles (for different applications)

HDTV	1920	1080	24	60	2986 Mbps	MPEG-2 25-34Mbps
TV	720	576	24	25	498 Mbps	MPEG-2 3-6Mbps

Although MPEG-2 is similar to MPEG-1 it was developed for digital television. Of course in this case D frames are not supported because usually in television you don't really do fast forward or anything of that kind so D types of frames are not really required. And to have greater quality DCT is 10 bit by 10 bit instead of 8/8 that is used in MPEG-1 to provide better quality and it supports four different resolutions, two of them are shown here HDTV and TV, another two are there and it supports five different profiles for different applications. So we can see this MPEG-2 provides different options, it is somewhat like a shopping list we choose from, for a particular profile whether this is the resolution and so on so you can make choices accordingly depending upon your application. And MPEG-2 has a more general way of multiplexing. Here each stream is packetized with time stamps as it is done in case of MPEG-1 and the output of each packetizer is a Packetized Elementary System having 32 header field.

(Refer Slide Time: 43:14)



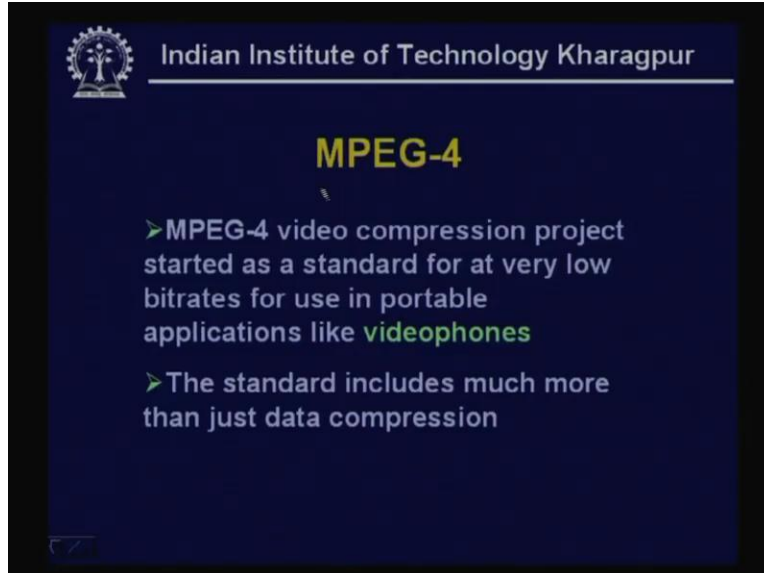
So each packetizer produces a Packetized Elementary System which has got 30 header fields and that 30 header fields are essentially gives the time stamping, error detection and various other fields are there which are used then these are multiplexed. One is PS multiplexer which is essentially a program multiplexer and another is a transport multiplexer. This program multiplexer is essentially variable length packets and a common time base. As we have seen after compression it can generate variable rate data, that's why this PS is variable length packets however a common time based is used.

On the other hand, this transport sequence is fixed length packets with no common time base so these two are sent after the compression.

Coming to the MPEG-4, MPEG-4 video compression project started as a standard for very low bit rates like for using portable applications like videophones. So to support very low bit rates this MPEG-4 is being used and this standard includes much more than just data compression.

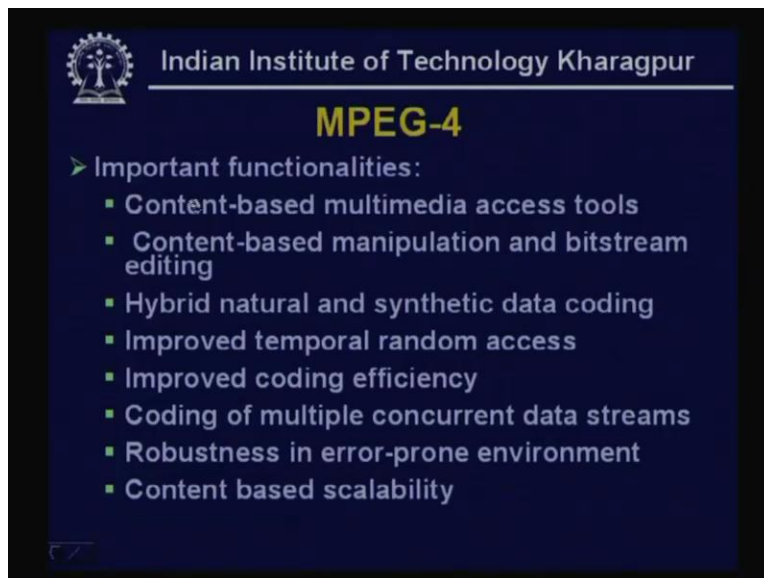


(Refer Slide Time: 44:34)



We have seen in MPEG-1 and MPEG-2 essentially the techniques are concerned with compression. But apart from compression as we can see MPEG-4 has got many important functionalities.

(Refer Slide Time: 45:14)



Some of the important functionalities are mentioned here.

- content based multimedia access tools
- content based manipulation and bit stream editing
- hybrid natural and synthetic data coding

- improved temporal random access
- improved coding efficiency
- coding of multiple concurrent data streams
- robustness in error-prone environment and
- content based scalability

So the main feature as we can see it has got many facilities for content based processing. you must have noticed that nowadays when you are watching relay of cricket matches sometimes whenever a batsman loses out or a bowler receives a catch that part can be replayed very quickly and that is precisely because of this content based multimedia access tools. Or whenever a commentator comments about a catch or wrong fielding etc then this situation can be very easily identified and accessed with the help of this kind of processing capability which MPEG-4 allows. Moreover, it allows some kind of data streaming so multilingual transmission is also allowed with the help of MPEG-4.

Now let us focus on the last standard that is your H.261. This H.261 was developed as a standard for digital telephony for ISDN services.

(Refer Slide Time: 46:52)

Indian Institute of Technology Kharagpur

## H.261

- H.261 was developed as a standard for digital telephony for ISDN services
- H.261 limits the images to just two sizes; the Common Intermediate Format (CIF) and quarter CIF (QCIF)

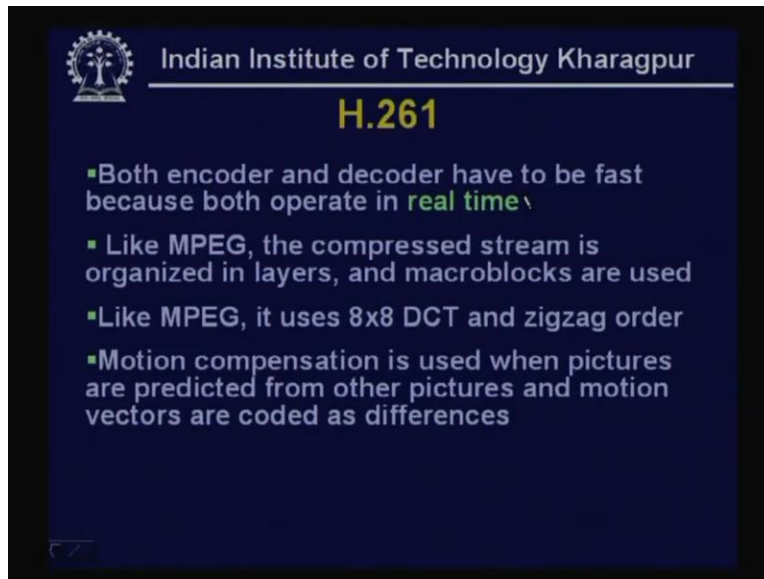
CIF	352	288	8	15	24.33 Mbps	H.261 112 Kbps
QCIF	176	144	8	10	4.0 Mbps	MPEG-4 <64 Kbps

As we have seen in case of ISDN services the rates are multiples of 64 Kbps, that's why after compression it is essentially necessary to transmit at a rate lower than 64 kilo bits or some multiples of it. So H.261 limits the images to just two sizes; one is known as Common Intermediate Format CIF or Quarter Common Intermediate Format QCIF.

So, using CIF the frame size is 352/288 and each pixel is represented by 8 bit and there are 15 frames per second so that gives you 24.33 Mbps without compression and by using H.261 you get 112 Kbps. And in case of QCIF still smaller frame size is used and the

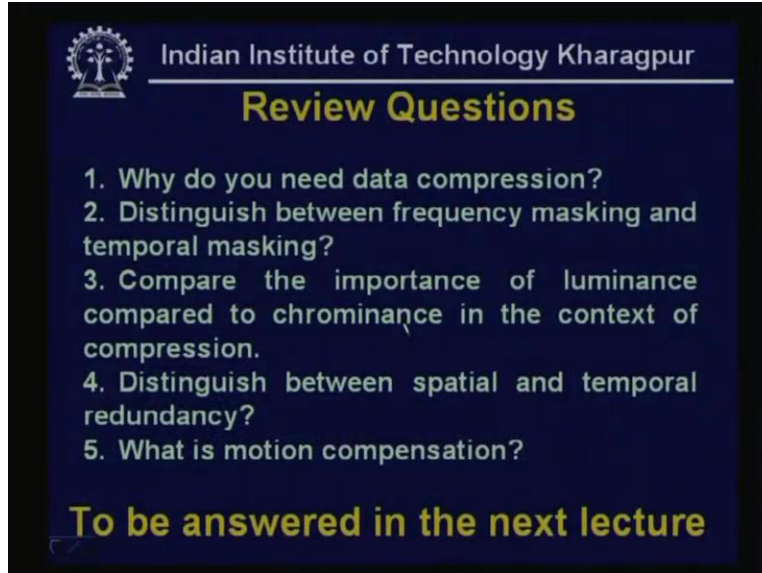
number of frames per second is also reduced from 15 to 10. Of course the number of bits used for each pixel is 8 and the uncompressed signal has got 4.0 Mbps and after compression by using MPEG-4 you can also use H.261 it gives you less than 64 Kbps. This is primarily used for transmission through ISDN network. Another requirement of H.261 is both encoder and decoder have to be fast because both operate in real-time.

(Refer Slide Time: 48:43)



So, in case of MPEG-1 or MPEG-2 or MPEG-4 we have seen that encoder can be very slow because it is very complex and on the other hand decoder has to be fast. But in case of H.261 both encoder and decoder has to be very fast because it is used for interactive video transmission for example video conferencing. And like MPEG the compressed stream is organized in layers and macro blocks and like MPEG it uses 8/8 DCT and zig zag order as we have seen for the purpose of compression and motion compensation is used when pictures are predicted from other pictures and motion vectors are coded as differences. This part is similar to MPEG. The way the motion compensation is done in MPEG it is done in the similar manner in H.261. So we have discussed different audio and video compression techniques in this lecture. Let us see the review questions.

(Refer Slide Time: 50:11)



Indian Institute of Technology Kharagpur

### Review Questions

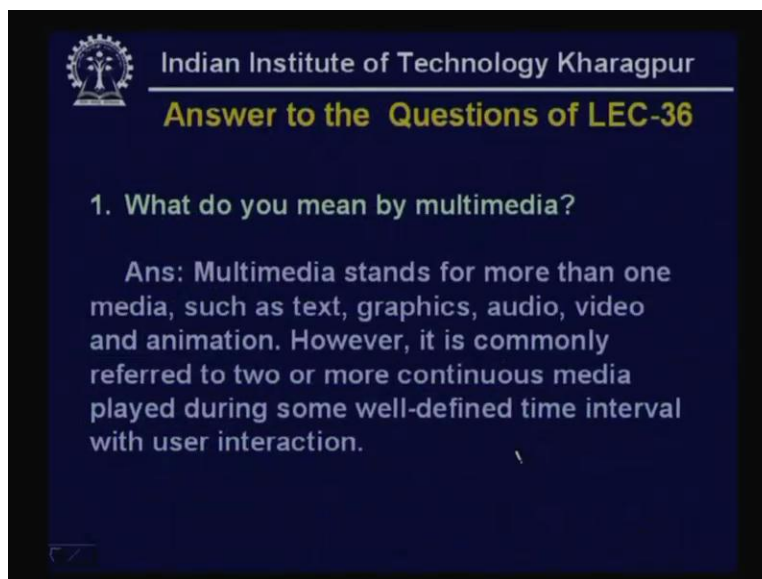
1. Why do you need data compression?
2. Distinguish between frequency masking and temporal masking?
3. Compare the importance of luminance compared to chrominance in the context of compression.
4. Distinguish between spatial and temporal redundancy?
5. What is motion compensation?

**To be answered in the next lecture**

- 1) Why do you need data compression?
- 2) Distinguish between frequency masking and temporal masking?
- 3) Compare the importance of luminance compared to chrominance in the context of compression obviously in the context of video compression.
- 4) Distinguish between spatial and temporal redundancy?
- 5) What is motion compensation?

Now it is time to give the answers of question of lecture minus 36.

(Refer Slide Time: 50:48)



Indian Institute of Technology Kharagpur

### Answer to the Questions of LEC-36

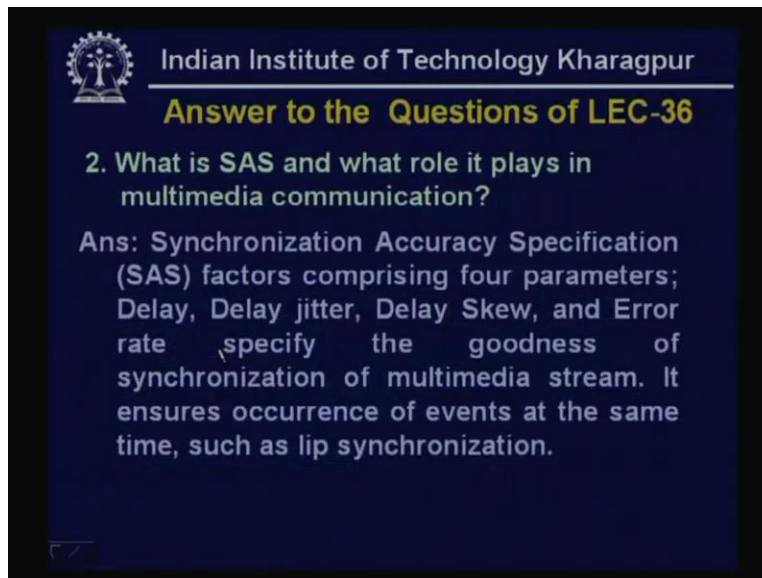
1. What do you mean by multimedia?

Ans: Multimedia stands for more than one media, such as text, graphics, audio, video and animation. However, it is commonly referred to two or more continuous media played during some well-defined time interval with user interaction.

1) What do you mean by multimedia?

As we know multimedia stands for more than one media. So, by definition multimedia means more than one media such as text, graphics, audio, video and animation. For example, while delivering this lecture I have used text and also used graphics and animation so we can say I have used multimedia facilitations and also audio/video. However, it is commonly referred to two or more continuous media such as audio and video played during some well-defined time interval using user interaction. So commonly this is how the multimedia is defined.

(Refer Slide Time: 52:05)

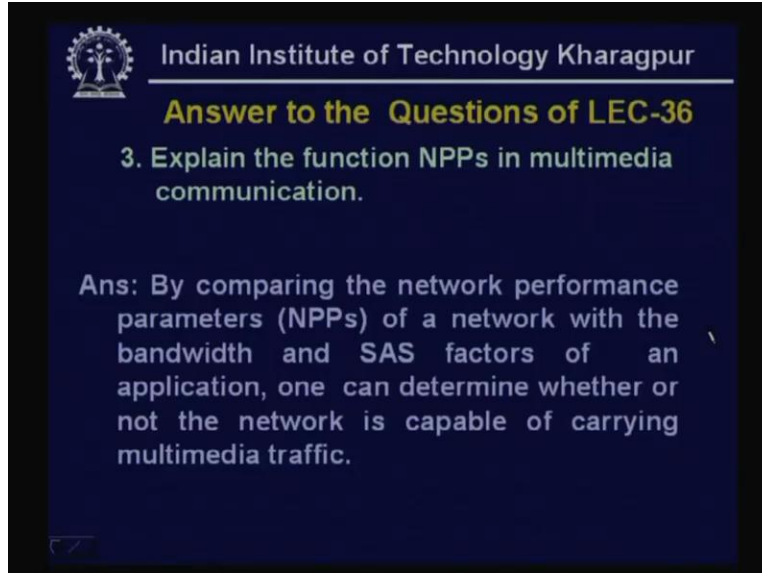


The slide is a dark blue rectangle with white and yellow text. At the top left is the IIT Kharagpur logo. To its right, the text reads 'Indian Institute of Technology Kharagpur' in white, followed by 'Answer to the Questions of LEC-36' in yellow. Below this, the question '2. What is SAS and what role it plays in multimedia communication?' is written in white. The answer follows: 'Ans: Synchronization Accuracy Specification (SAS) factors comprising four parameters; Delay, Delay jitter, Delay Skew, and Error rate specify the goodness of synchronization of multimedia stream. It ensures occurrence of events at the same time, such as lip synchronization.'

2) What is SAS synchronization accuracy specification and what role it plays in multimedia communication?

As we have seen the multimedia signals require bandwidth as well as well as SAS factors to specify the goodness, transmission and particularly the SAS factors comprising four parameters like delay, delay jitter, delay skew and error rate as we have already discussed in detail. They specify the goodness of synchronization of multimedia stream in addition to the bandwidth which is required. It ensures occurrence of events at the same time, such as lip synchronization. This synchronization accuracy specification ensures that the entire multimedia signal takes place in precise time instance in a synchronized manner. That is the role of SAS factors.

(Refer Slide Time: 53:13)



The slide features the IIT Kharagpur logo in the top left corner. The text is as follows:

Indian Institute of Technology Kharagpur

---

**Answer to the Questions of LEC-36**

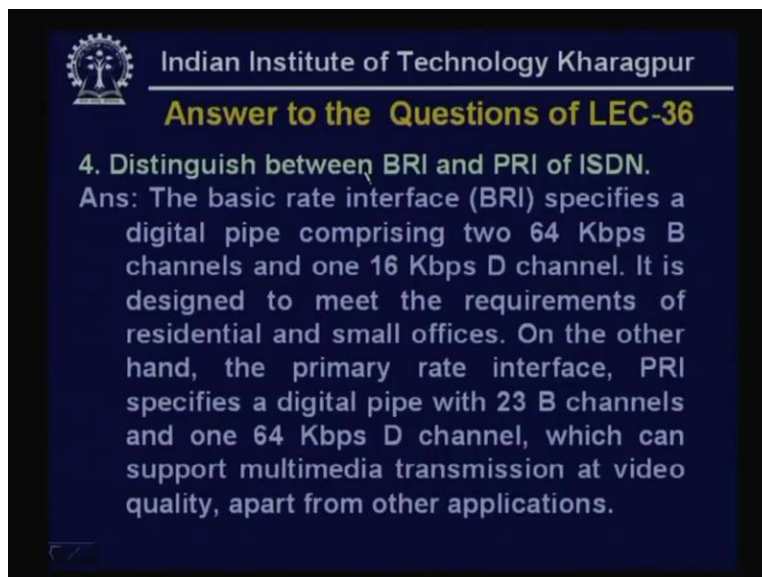
3. Explain the function NPPs in multimedia communication.

Ans: By comparing the network performance parameters (NPPs) of a network with the bandwidth and SAS factors of an application, one can determine whether or not the network is capable of carrying multimedia traffic.

3) Explain the function of NPPs in multimedia communication.

For multimedia transmission the requirement is specified by SAS factors and bandwidth. On other hand, the network performance is specified in terms of network parameters and these two are to be compared the network performance parameters and SAS factors as well as bandwidth requirement of an application. By comparing them one can determine whether or not the network is capable of carrying multimedia traffic. So this network performance parameter is very essential to decide whether a network is suitable for multimedia communication or not.

(Refer Slide Time: 54:11)



The slide features the IIT Kharagpur logo in the top left corner. The text is as follows:

Indian Institute of Technology Kharagpur

---

**Answer to the Questions of LEC-36**

4. Distinguish between BRI and PRI of ISDN.

Ans: The basic rate interface (BRI) specifies a digital pipe comprising two 64 Kbps B channels and one 16 Kbps D channel. It is designed to meet the requirements of residential and small offices. On the other hand, the primary rate interface, PRI specifies a digital pipe with 23 B channels and one 64 Kbps D channel, which can support multimedia transmission at video quality, apart from other applications.

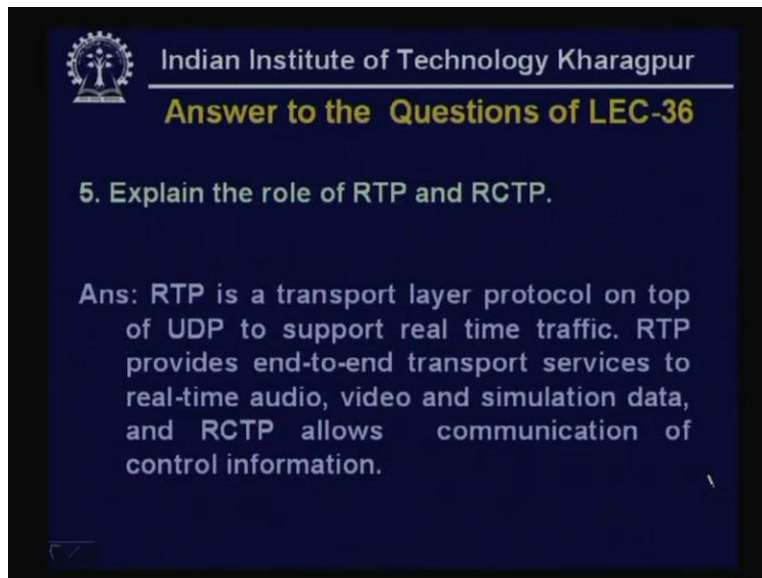
4) Distinguish between BRI and PRI of ISDN.

We have seen that ISDN uses two important interfaces; one is known as basic rate interface BRI that specifies a digital pipe comprising two 64 Kbps B channels and 116 kilobits per channel D so this is of lower bandwidth and it is designed to meet the requirements of residential as well as small offices.

On the other hand, the primary rate interface PRI specifies a digital pipe with 23B channels each of 64 Kbps and 1D channel that also is of 64 Kbps. So obviously with the help of this you can achieve much higher bandwidth that can support multimedia transmission at video quality. Thus, obviously a PRI interface cannot support multimedia transmission and video quality.

And of course whenever you have got PRI interface apart from video quality multimedia transmission you can use it for various other applications such as data communication.

(Refer Slide Time: 55:35)



5) Explain the role of RTP and RCTP.

RTP is a transport layer protocol on top of UDP to support real-time traffic. UDP is a connectionless unreliable protocol so it cannot really support multimedia communication so, to facilitate that a RTP UDP RTP UDP should be used. That means UDP should be used in conjunction with RTP to support real-time traffic.

Thus, RTP provides end-to-end transport services to real-time audio, video and simulation data. Of course for control information you require another protocol that is your RCTP that allows communication of control information.

With this we come to the end of today's lecture and we have discussed how audio and video compression can be performed. In the next lecture we shall discuss the different applications possible because of compression.

Thank you.