

CS 414 – Multimedia Systems Design
Lecture 10 – MPEG-1 Video
and MP3 Audio) (Part 5)

Klara Nahrstedt
Spring 2008



Administrative

- MP1 deadline Friday, February 8
- Demonstrations, Friday, February 8, 5-7pm
- Please sign up for a demonstration slot
- MP2 will be posted on Friday, February 8 and deadline will be Friday, 22 (please, start early – we will have two discussion sections for MP2)



Motion Picture Expert Group (MPEG)

- General Information about MPEG
 - Began in 1988; Part of Same ISO as JPEG
- MPEG-1/Video
- MPEG/Audio – MP3
- MPEG-2
- MPEG-4
- MPEG-7



MPEG General Information

- Goal: data compression 1.5 Mbps
- MPEG defines video, audio coding and system data streams with synchronization
- MPEG information
 - Aspect ratios: 1:1 (CRT), 4:3 (NTSC), 16:9 (HDTV)
 - Refresh frequencies: 23.975, 24, 25, 29.97, 50, 59.94, 60 Hz



MPEG Image Preparation (Resolution and Dimension)

- MPEG defines exactly format
 - Three components: Luminance and two chrominance components (2:1:1)
 - Resolution of luminance comp: $X1 \leq 768$; $Y1 \leq 576$ pixels
 - Pixel precision is 8 bits for each component
- Example of Video format: 352x240 pixels, 30 fps; chrominance components: 176x120 pixels

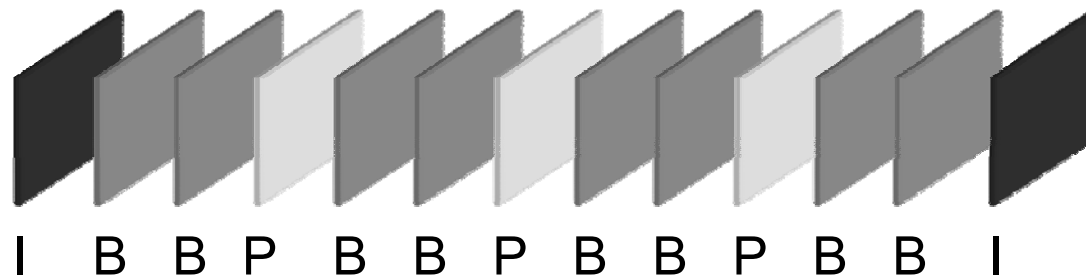


MPEG Image Preparation - Blocks

- Each image is divided into **macro-blocks**
- Macro-block : 16x16 pixels for luminance;
8x8 for each chrominance component
- Macro-blocks are useful for Motion
Estimation
- No MCUs which implies sequential non-
interleaving order of pixels values

MPEG Video Processing

- Intra frames (same as JPEG)
 - typically about 12 frames between I frames
- Predictive frames
 - encode from previous I or P reference frame
- Bi-directional frames
 - encode from previous and future I or P frames





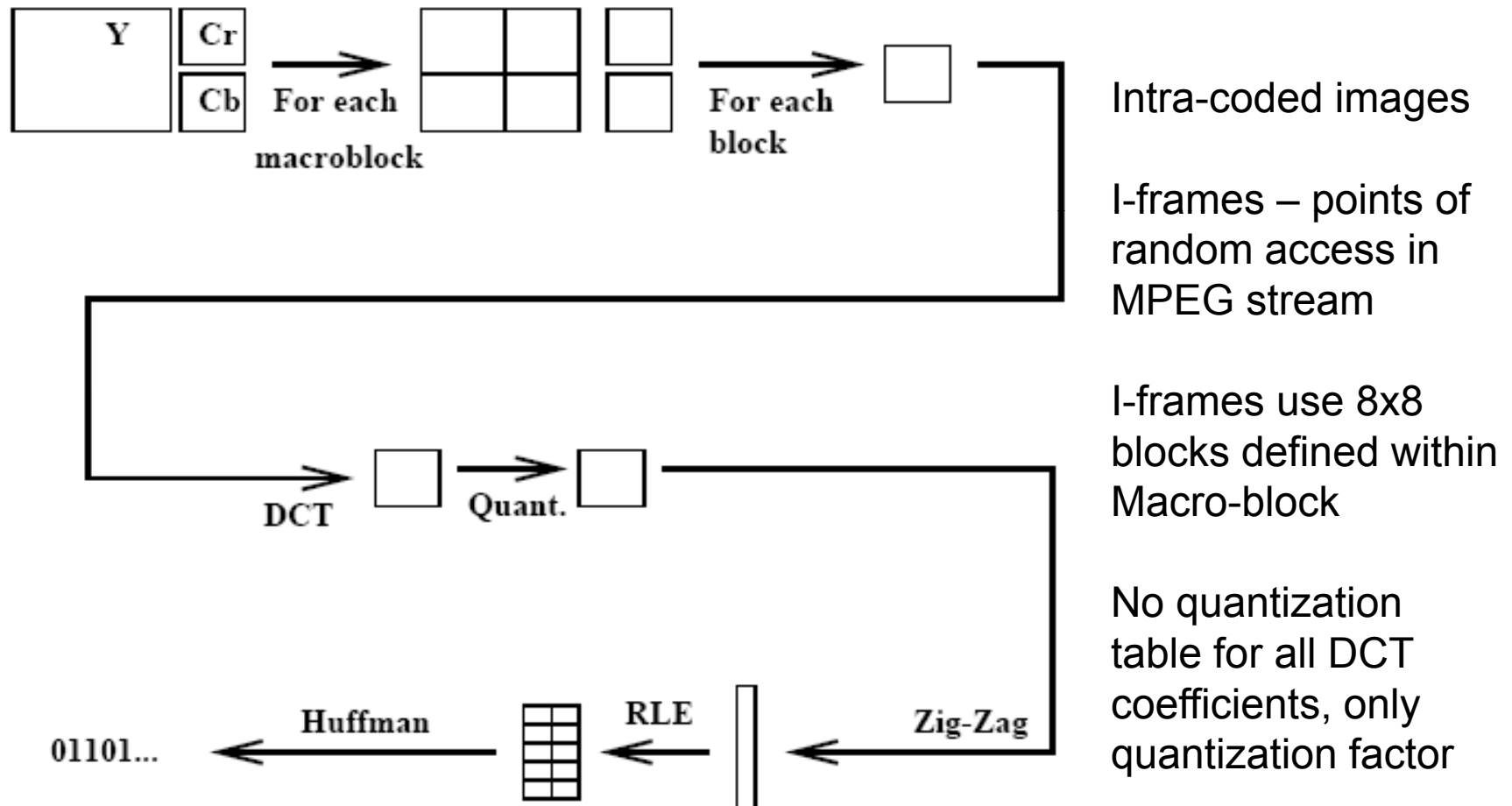
Selecting I, P, or B Frames

■ Heuristics

- change of scenes should generate I frame
- limit B and P frames between I frames
- B frames are computationally intense

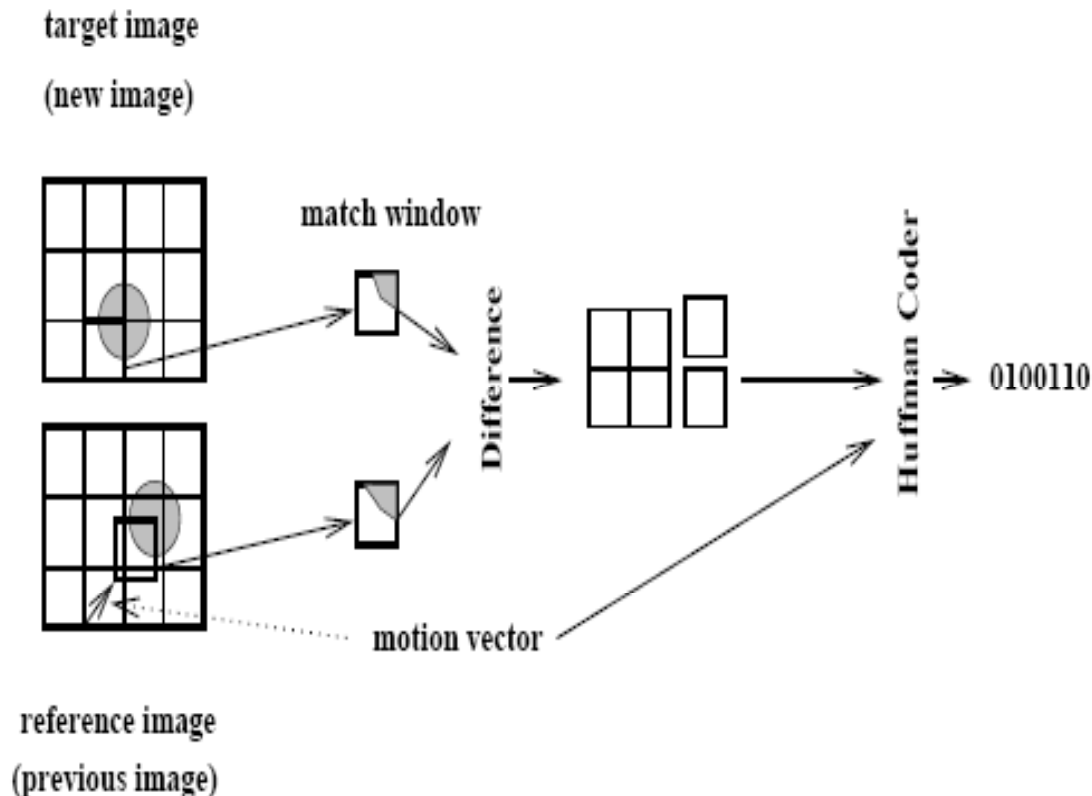
Type	Size	Compress
I	18K	7:1
P	6K	20:1
B	2.5K	50:1
Avg	4.8K	27:1

MPEG Video I-Frames




MPEG Video P-Frames

Motion Estimation Method



Predictive coded frames require information of previous I frame and or previous P frame for encoding/decoding

For **Temporary Redundancy** we determine last P or I frame that is most similar to the block under consideration



Motion Computation for P Frames

- Predictive search
- Look for match window within a given search window
 - Match window – macro-block
 - Search window – arbitrary window size depending how far away are we willing to look
- Displacement of two match windows is expressed by motion vector



Matching Methods

- SSD metric

$$SSD = \sum_{i=0}^{N-1} (x_i - y_i)^2$$

- SAD metric

$$SAD = \sum_{i=0}^{N-1} |x_i - y_i|$$

- Minimum error represents best match
 - must be below a specified threshold
 - error and perceptual similarity not always correlated

Example of Finding Minimal SSD

$$\begin{array}{ccc}
 7 & 9 & 8 \\
 5 & 4 & 6 \\
 9 & 8 & 2
 \end{array}
 \text{ versus }
 \begin{array}{ccc}
 8 & 7 & 9 \\
 7 & 5 & 4 \\
 7 & 5 & 4
 \end{array}
 \Rightarrow
 \begin{aligned}
 \text{SSD} &= (7-8)^2 + (9-7)^2 + (8-9)^2 \\
 &\quad + (5-7)^2 + (4-5)^2 + (6-4)^2 \\
 &\quad + (9-7)^2 + (8-5)^2 + (2-4)^2 \\
 &= 1 + 4 + 1 + 4 + 1 + 4 + 4 + 9 + 4 \\
 &= 32
 \end{aligned}$$

$$\begin{array}{ccc}
 7 & 9 & 8 \\
 5 & 4 & 6 \\
 9 & 8 & 2
 \end{array}
 \text{ versus }
 \begin{array}{ccc}
 8 & 7 & 10 \\
 6 & 5 & 4 \\
 10 & 7 & 1
 \end{array}
 \Rightarrow \text{SSD} = 18$$

min SSD = 18 =>
take match windows:

$$\begin{array}{ccc}
 7 & 9 & 8 \\
 5 & 4 & 6 \\
 9 & 8 & 2
 \end{array}
 \text{ and }
 \begin{array}{ccc}
 8 & 7 & 10 \\
 6 & 5 & 4 \\
 10 & 7 & 1
 \end{array}$$

Example of Comparing Minimal SSD and SAD

SSD: 7 9 8 8 7 10
 5 4 6 versus 6 5 4 -> SSD = 18
 9 8 2 10 7 1

 7 9 8 8 7 10
 5 4 6 versus 6 5 4 -> SSD = 40,017
 9 8 2 10 7 202 **Outlier**

SAD:

 7 9 8 8 7 10
 5 4 6 versus 6 5 4 -> SAD = 211
 9 8 2 10 7 202



Syntax of P Frame

Addr	Type	Quant	Motion Vector	CBP	b0	b1	...	b5
-------------	-------------	--------------	--------------------------	------------	-----------	-----------	------------	-----------

Addr: address the syntax of P frame

Type: INTRA block is specified if no good match was found

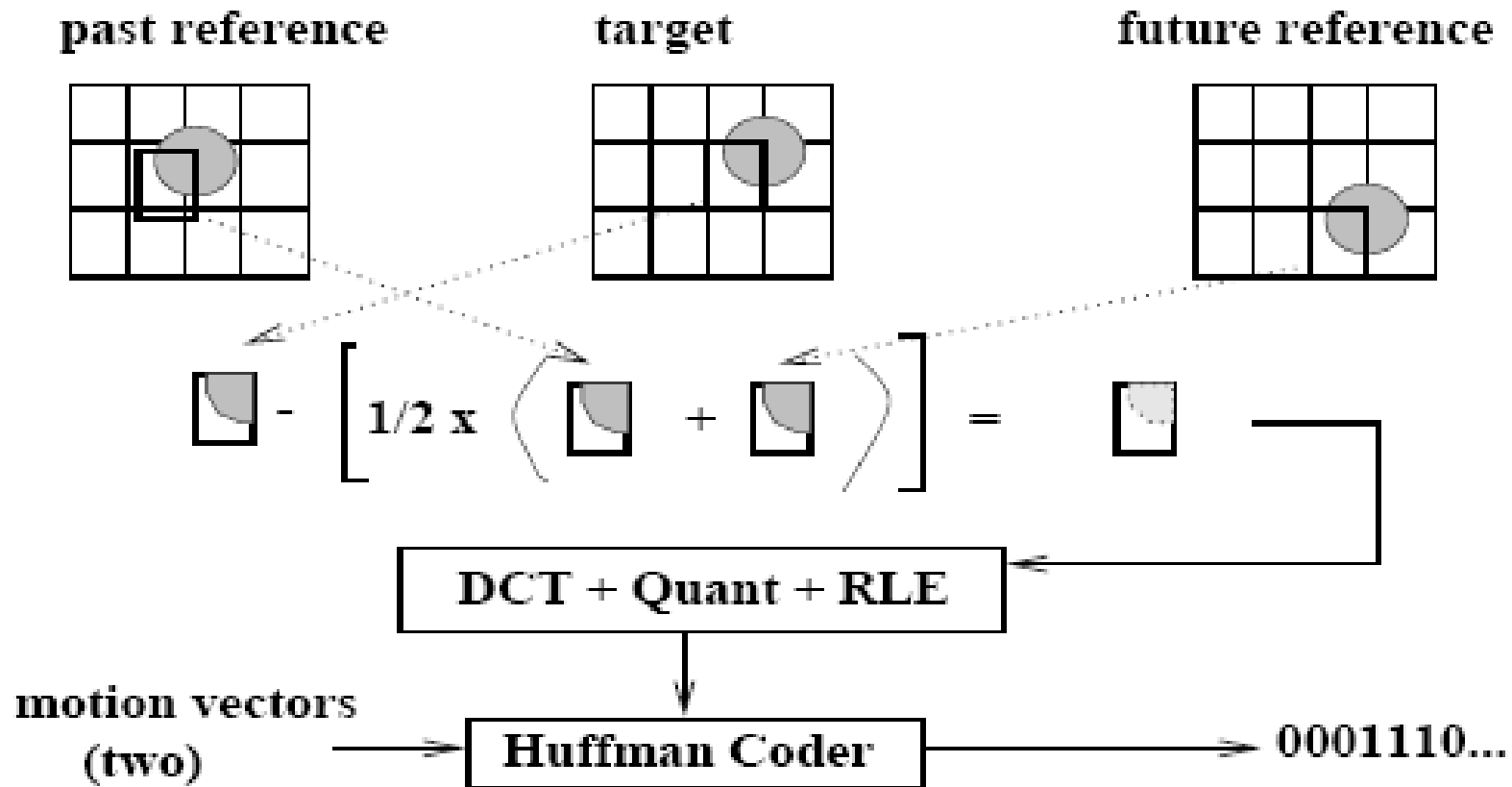
Quant: quantization value per macro-block (vary quantization to fine-tune compression)

Motion Vector: a 2D vector used for motion compensation provides offset from coordinate position in target image to coordinates in reference image

CBP(Coded Block Pattern): bit mask indicates which blocks are present

MPEG Video B Frames

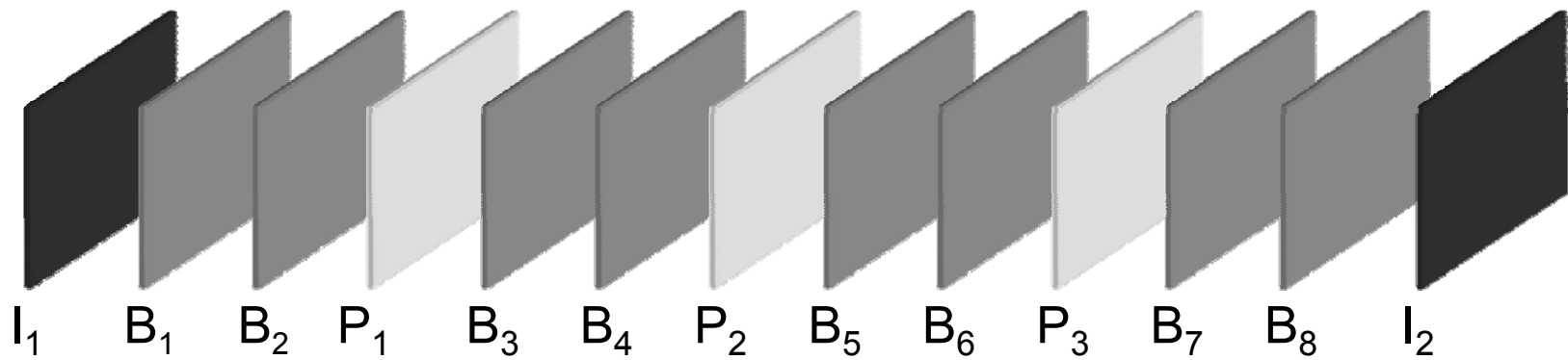
Bi-directionally Predictive-coded frames



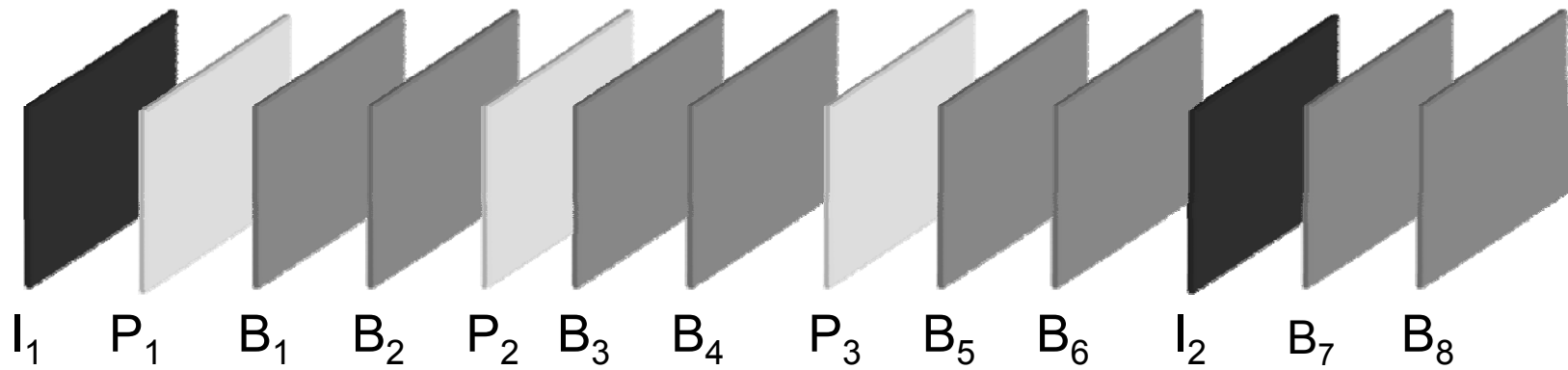


MPEG Video Decoding

Display Order



Decoding Order

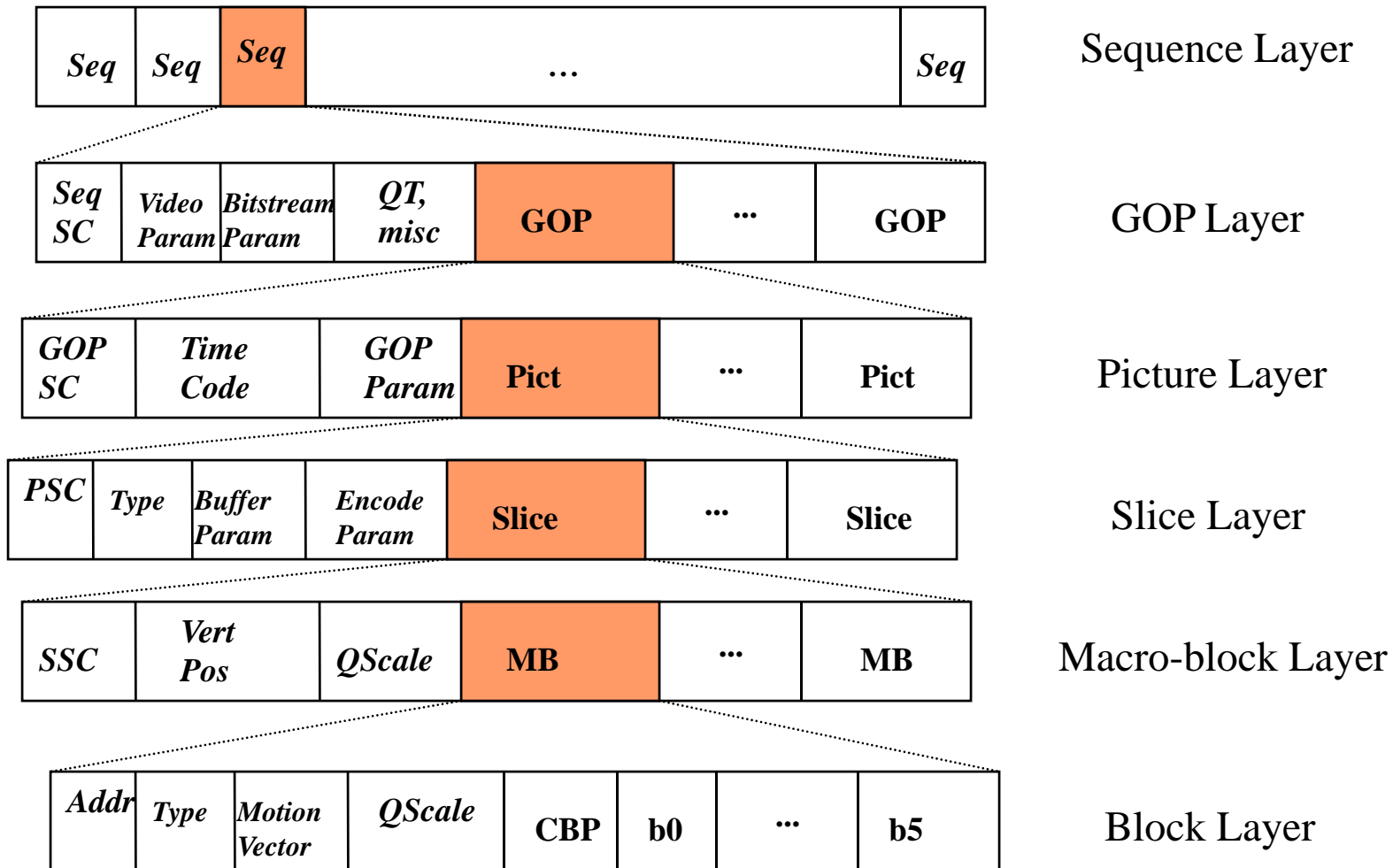




MPEG Video Quantization

- AC coefficients of B/P frames are usually large values, I frames have smaller values
 - Adjust quantization
- If data rate increases over threshold, then quantization enlarges step size (increase quantization factor Q)
- If data rate decreases below threshold, then quantization decreases Q

MPEG-1 Interchange Format



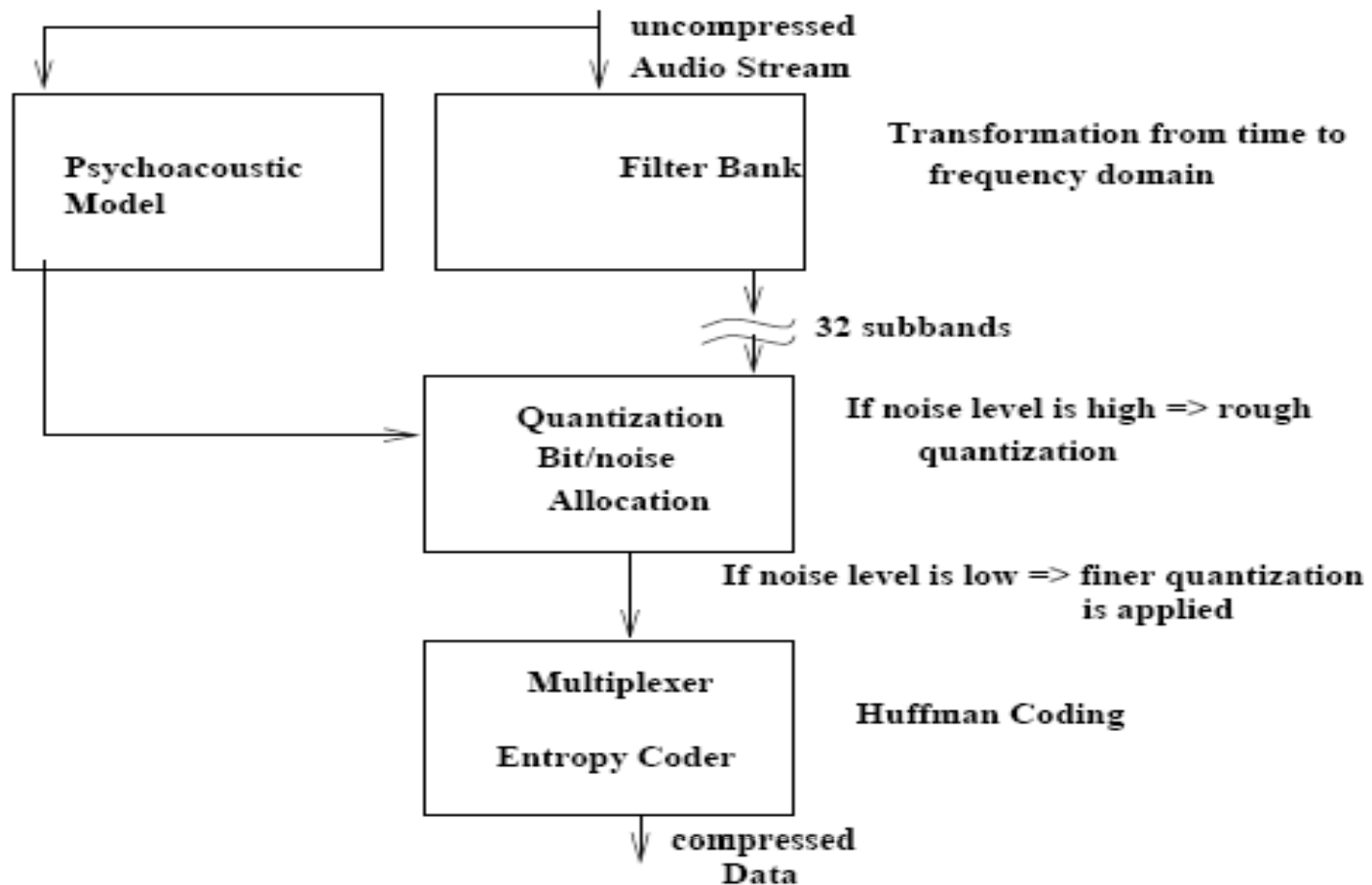


MPEG Audio Encoding

■ Characteristics

- Precision 16 bits
- Sampling frequency: 32KHz, 44.1 KHz, 48 KHz
- 3 compression layers: Layer 1, Layer 2, Layer 3 (MP3)
 - Layer 3: 32-320 kbps, target 64 kbps
 - Layer 2: 32-384 kbps, target 128 kbps
 - Layer 1: 32-448 kbps, target 192 kbps

MPEG Audio Encoding Steps




MPEG Audio Filter Bank

- Filter bank divides input into multiple sub-bands (32 equal frequency sub-bands)
- Sub-band i defined

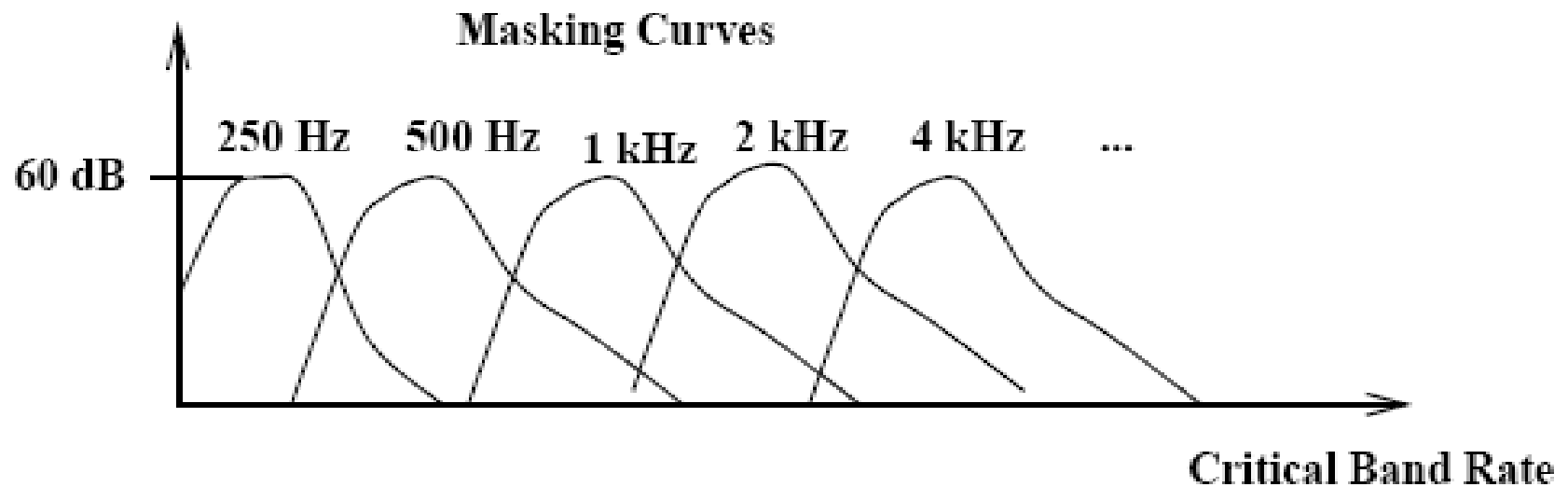
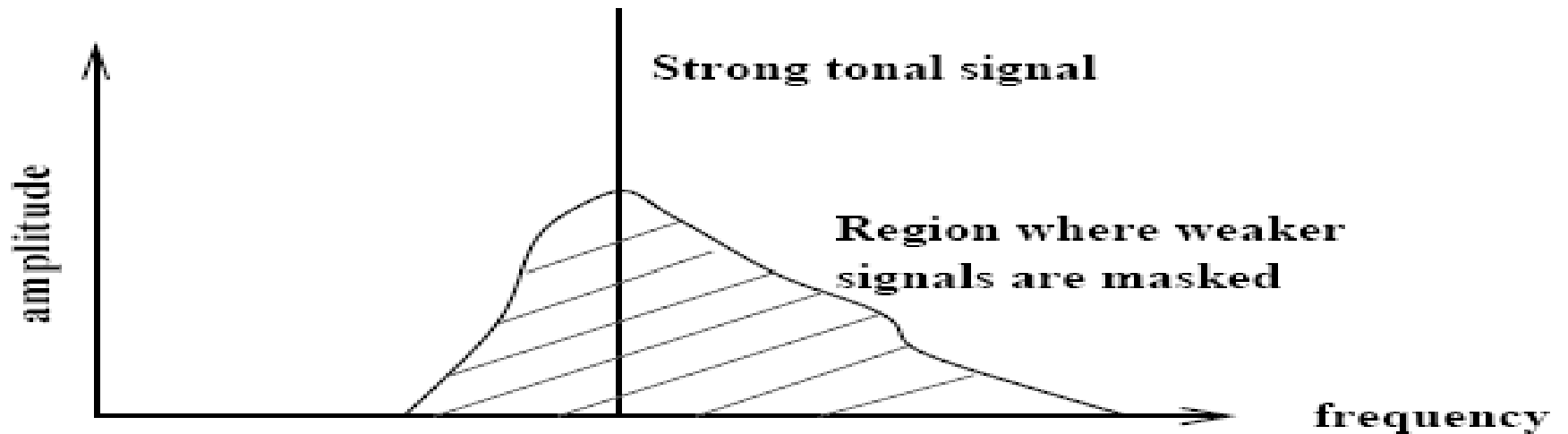
$$S_t[i] = \sum_{k=0}^7 3 \sum_{j=0}^7 \cos\left(\frac{(2i+1)(k-16)\pi}{64}\right) * (C[k+64j] * x[k+64j])$$

- $i \in [0,31]$, $S_t[i]$ - filter output sample for sub-band i at time t , $C[n]$ – one of 512 coefficients, $x[n]$ – audio input sample from 512 sample buffer



MPEG Audio Psycho-acoustic Model

- MPEG audio compresses by removing acoustically irrelevant parts of audio signals
- Takes advantage of human auditory systems inability to hear quantization noise under auditory masking
- Auditory masking: occurs when ever the presence of a strong audio signal makes a temporal or spectral neighborhood of weaker audio signals imperceptible.



MPEG/audio divides audio signal into frequency sub-bands that approximate critical bands. Then we quantize each sub-band according to the audibility of quantization noise within the band



MPEG Audio Bit Allocation

- This process determines number of code bits allocated to each sub-band based on information from the psycho-acoustic model
- Algorithm:
 1. Compute mask-to-noise ratio: $MNR = SNR - SMR$
 - Standard provides tables that give estimates for SNR resulting from quantizing to a given number of quantizer levels
 2. Get MNR for each sub-band
 3. Search for sub-band with the lowest MNR
 4. Allocate code bits to this sub-band.
 - If sub-band gets allocated more code bits than appropriate, look up new estimate of SNR and repeat step 1



MPEG Audio Comments

- Precision of 16 bits per sample is needed to get good SNR ratio
- Noise we are getting is quantization noise from the digitization process
- For each added bit, we get 6dB better SNR ratio
- Masking effect means that we can raise the noise floor around a strong sound because the noise will be masked away
- Raising noise floor is the same as using less bits and using less bits is the same as compression



Conclusion

- MPEG system data stream
 - Interchange format for audio and video streams
 - Interleave audio and video packets; insert time stamps into each “frame” of data
- Synchronization
 - SRC – system clock reference; DTS – decoding time stamp; PTS – presentation time stamp
- During encoding
 - Insert SCR values into system stream
 - Stamp each frame with PTS and DTS
 - Use encode time to approximate decode time
- During decoding
 - Initialize local decoder clock with start values
 - Compare PTS to the value of local clock
 - Periodically synchronize local clock to SCR