CISC 7610 Lecture 3
Multimedia data and data formats

Topics:

Perceptual limits of multimedia data

JPEG encoding of images

MPEG encoding of audio

MPEG and H.264 encoding of video

Multimedia data: Perceptual limits

- Hearing
  - 20 – 20,000 Hz frequency range
  - 120 dB dynamic range of loudness (1,000,000 : 1)
- Vision
  - 30 frames per second
  - 300 dots per inch at one foot viewing distance
    - 226M highest-res pixels to cover field of view (Deering, 1998)
  - 140 dB dynamic range (10,000,000 : 1) over time

Audio can be fully captured

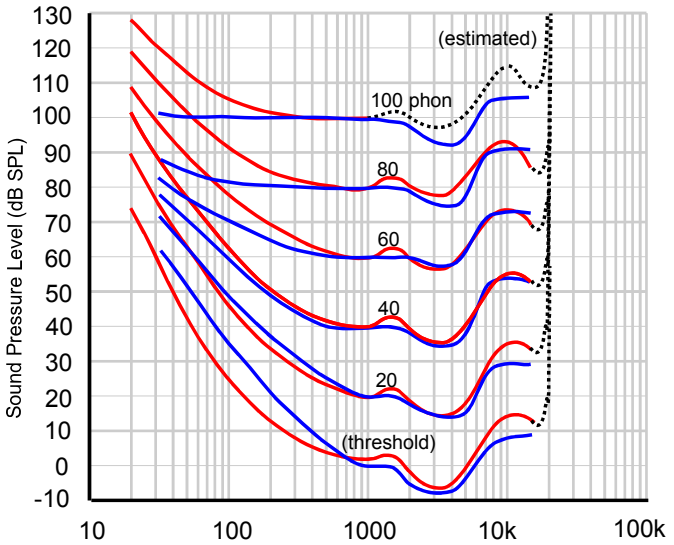
- Hearing
  - 20 – 20,000 Hz frequency range
  - 140 dB dynamic range of loudness (10,000,000 : 1)
- CD quality recording
  - Sampling rate: 44,100 Hz → max freq 22,050 Hz
  - Bit depth: 16 bits → dynamic range 96 dB
  - Bit rate: 44,100 samps/s x 16 bits x 2 chan. = 1.4 Mb/s
- Typical MP3 compressed recording
  - Bit rate: 128 kb/s (11x compression)

Video streams can still be improved

- Vision
  - 30 frames per second
  - 300 dots per inch at one foot viewing distance
    - 226M highest-res pixels to cover field of view
  - 140 dB brightness dynamic range (10,000,000 : 1) over time
- HD broadcast quality video = 1.5 Gb/s
  - 1920 x 1080 pixels/frame x 30 frames/s x 24 bits/pixel
- Typical H.264 compressed recording = 30 Mb/s (>50x)
  - 25 GB Blu-ray disc holds 2 hours, including audio

Compression lets us store and process these data efficiently

- Remove data that are redundant or irrelevant
- Redundant: implicit in remaining data
  - Can be fully reconstructed (lossless compression)
- Irrelevant: unique but unnecessary
  - For example: imperceptible to humans (lossy compression)

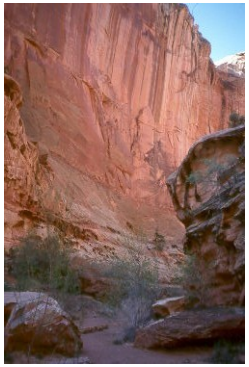


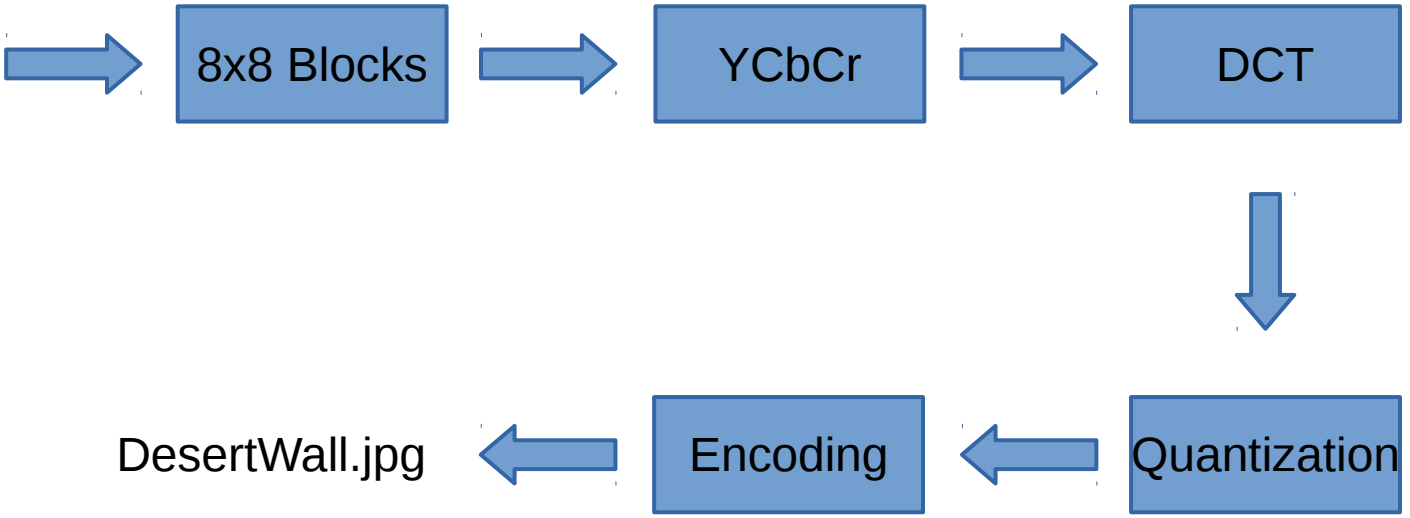
Equal-loudness contours (red) (from ISO 226:2003 revision)
Fletcher–Munson curves shown (blue) for comparison

Considerations for compression

- Latency: Amount of preceding signal that needs to be observed to compress a given sample
  - Important for real-time applications
- Locality: Amount of decoded signal that would be affected by changing one bit in encoded signal
  - Important for error robustness
- Generality: Variety of signals that can be encoded (efficiently) by a given encoder-decoder
- Decoder complexity vs encoder complexity

Image compression: JPEG encoding



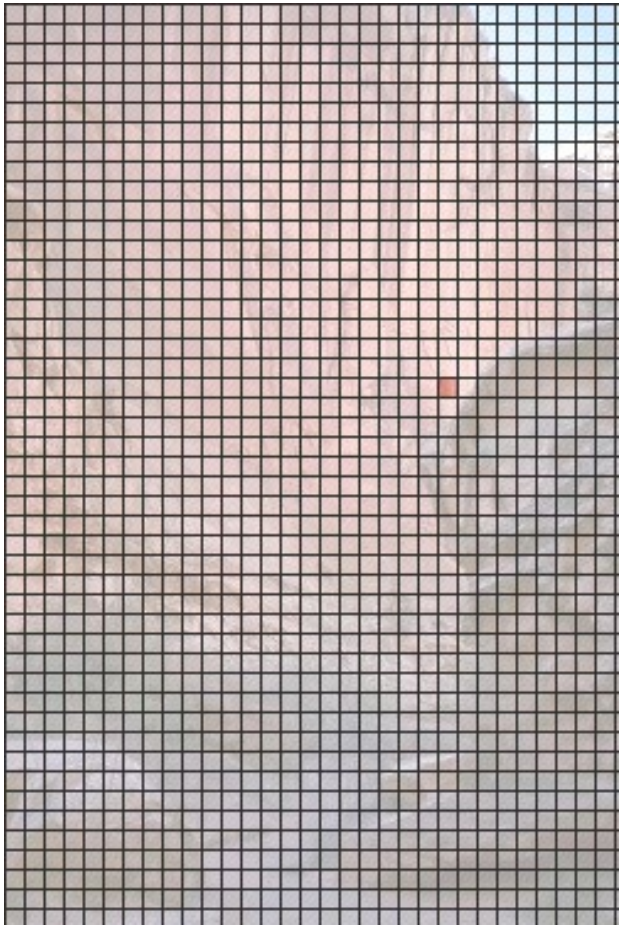


Images from: http://www.ams.org/samplings/feature-column/fcarc-image-compression

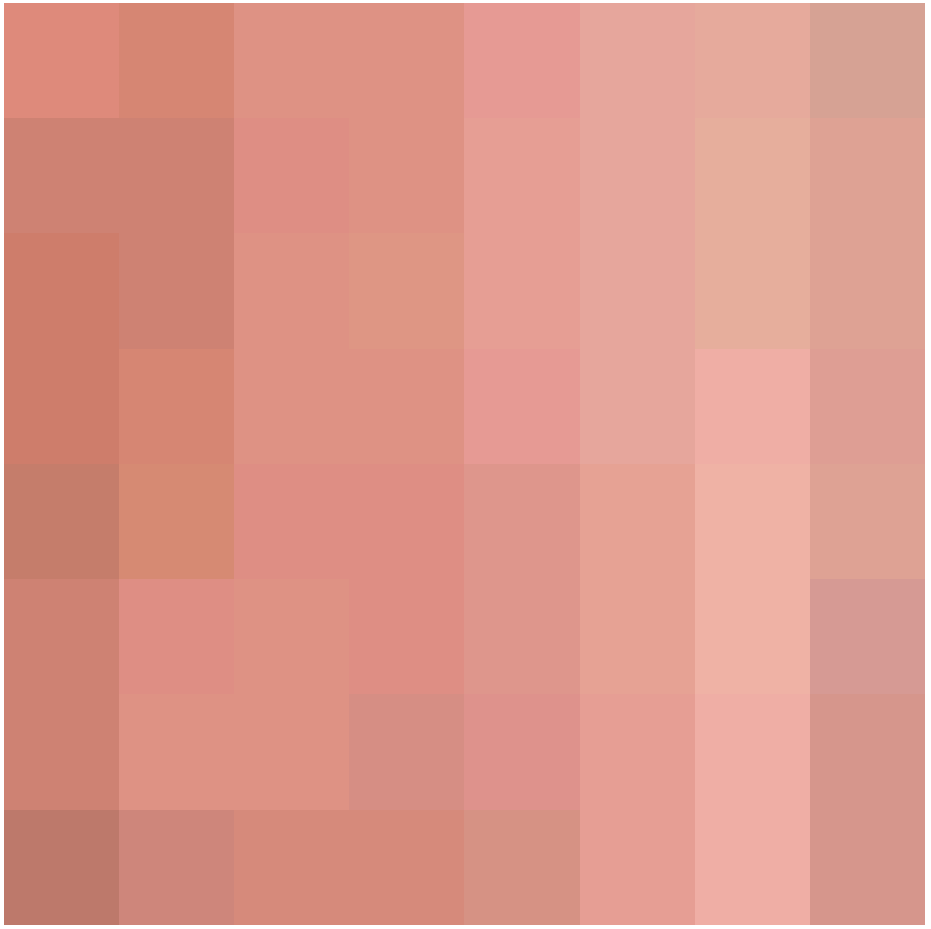
Break into 8x8 pixel blocks

Break into 8x8 pixel blocks

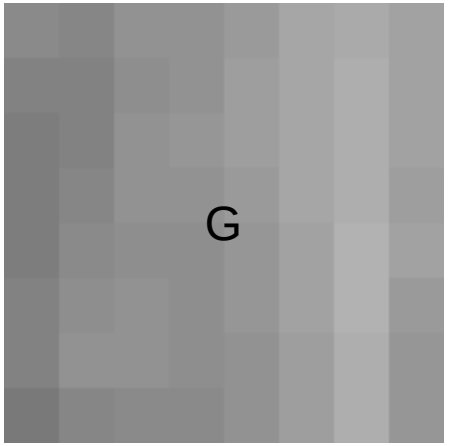
Break into 8x8 pixel blocks

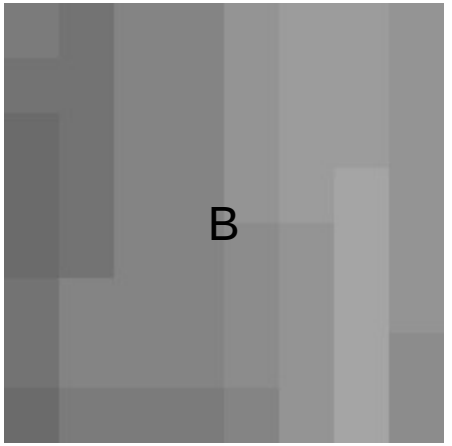


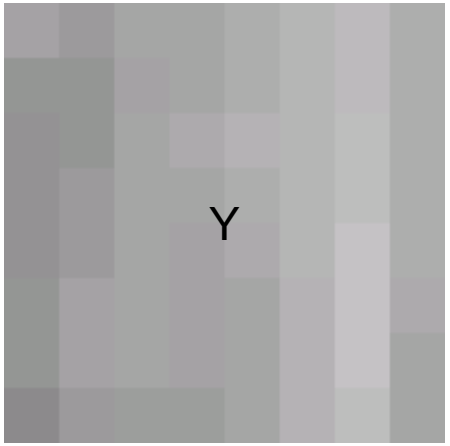
Break into 8x8 pixel blocks

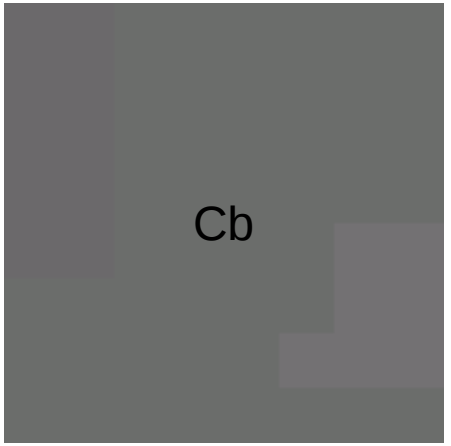


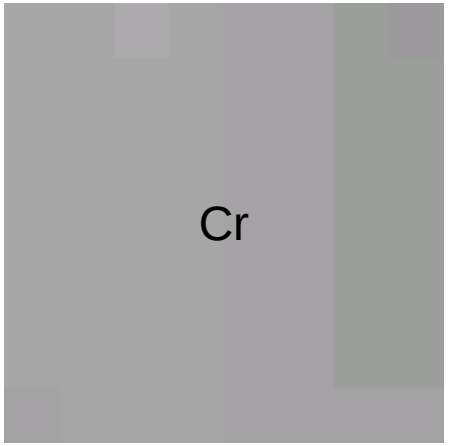
Transform from RGB to YCbCr











Transform from RGB to YCbCr

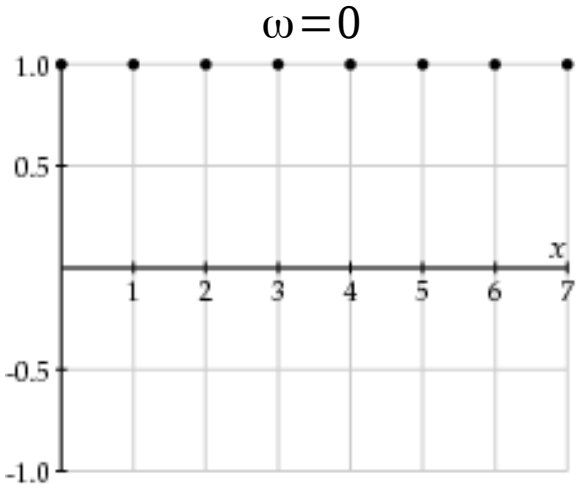
$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.29900 & 0.58700 & 0.11400 \\ -0.16874 & -0.33126 & 0.50000 \\ 0.50000 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

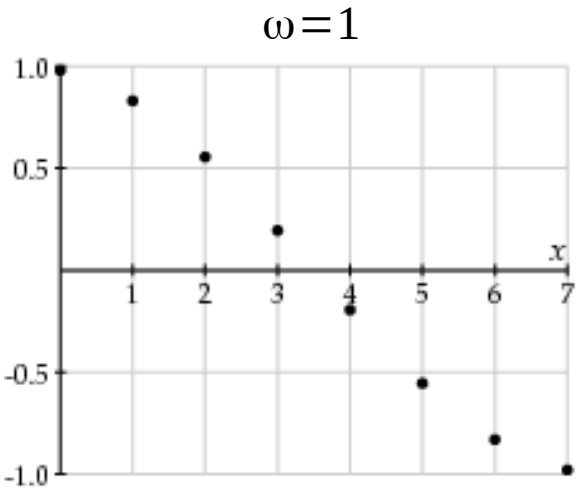
Discrete cosine transform (DCT)

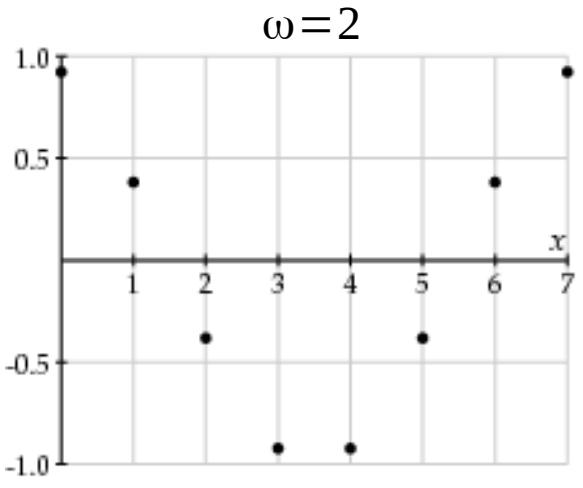
Forward: $$F_\omega = \frac{1}{2} C_\omega \sum_{x=0}^{7} f_x \cos\left(\frac{\pi(2x+1)\omega}{16}\right)$$

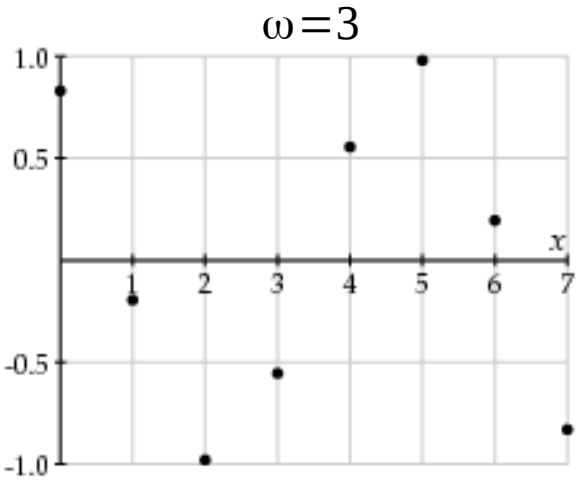
Inverse: $$f_x = \frac{1}{2} \sum_{\omega=0}^{7} C_\omega F_\omega \cos\left(\frac{\pi(2x+1)\omega}{16}\right)$$

First few bases





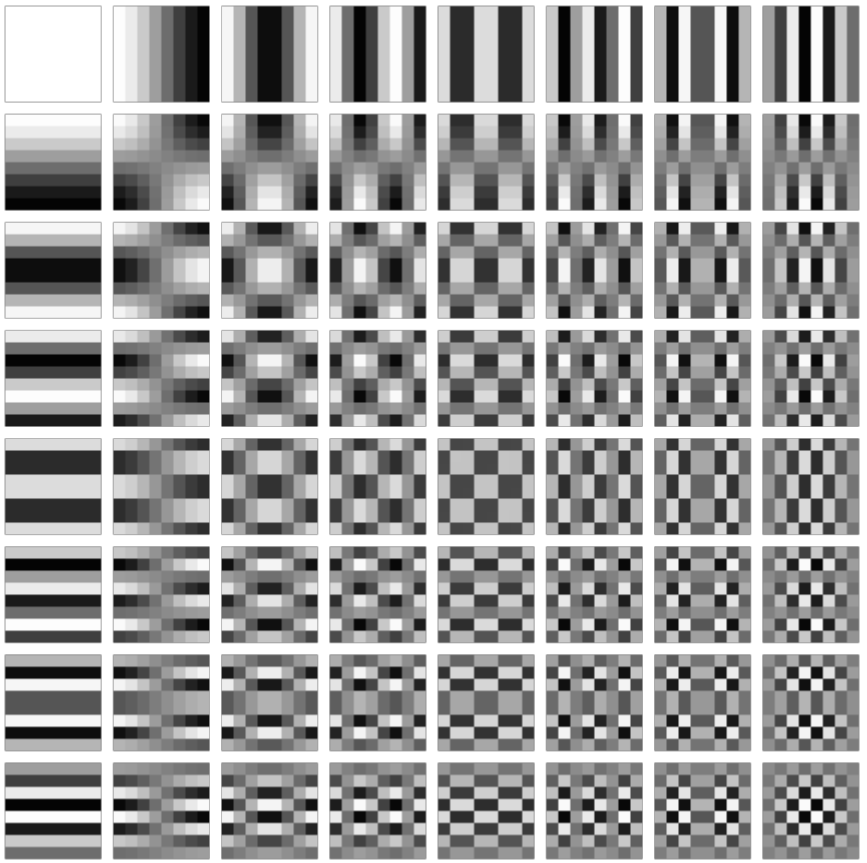




2D DCT is a 1D DCT in x and y

Forward: $$F_{u,v}=\frac{1}{4}C_u C_v \sum_{x=0}^{7}\sum_{y=0}^{7} f_{x,y}\cos\left(\frac{\pi(2x+1)u}{16}\right)\cos\left(\frac{\pi(2y+1)v}{16}\right)$$

Bases (cos(...)):



Quantize 2D DCT coefficients

- Global quality parameter
- Frequency-dependent "sensitivity" parameter

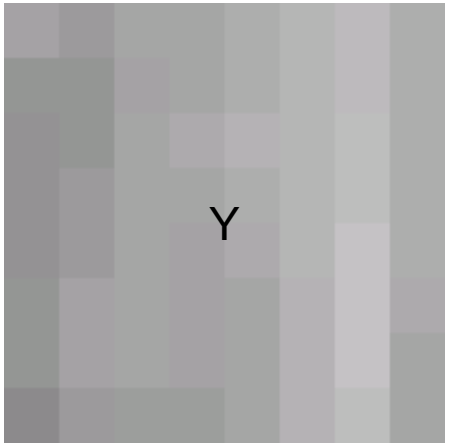
$$F'_{u,v} = round\left(\frac{F_{u,v}}{\alpha Q_{u,v}}\right)$$

- Bigger Q or α means fewer quantized values

$$Q_l = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

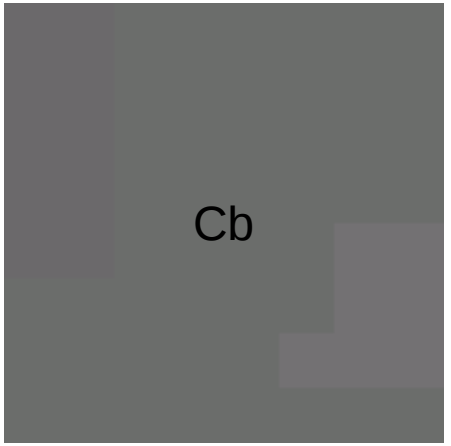
$$Q_c = \begin{bmatrix} 17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\ 18 & 21 & 26 & 66 & 99 & 99 & 99 & 99 \\ 24 & 26 & 56 & 99 & 99 & 99 & 99 & 99 \\ 47 & 66 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \end{bmatrix}$$

Quantized DCT coefficients for each channel



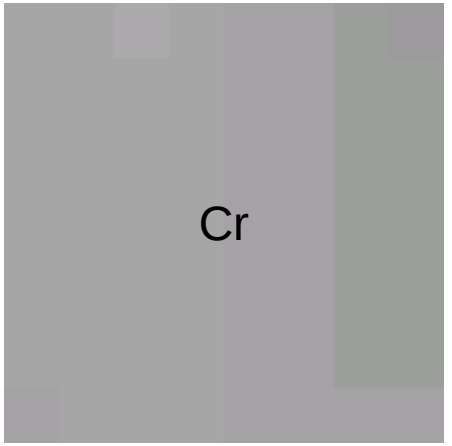
20	-7	-1	1	-2	1	0	0
1	0	0	0	1	0	0	0
-1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Y



-9	-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	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

Cb

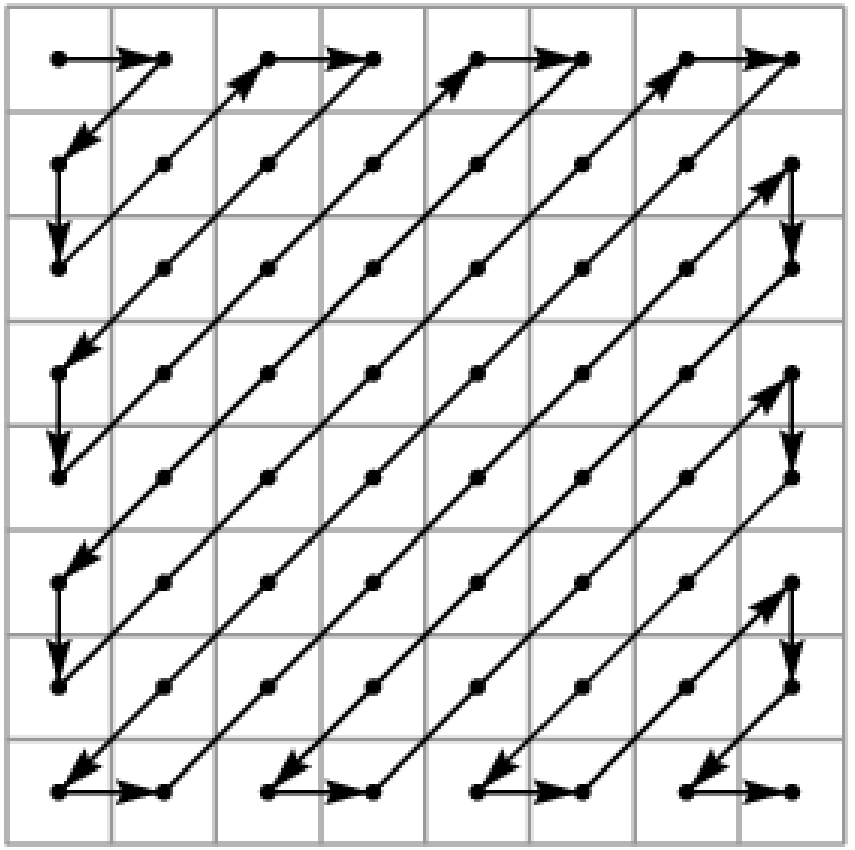


17	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

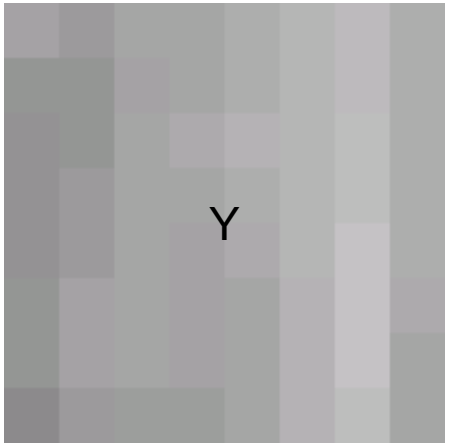
Cr

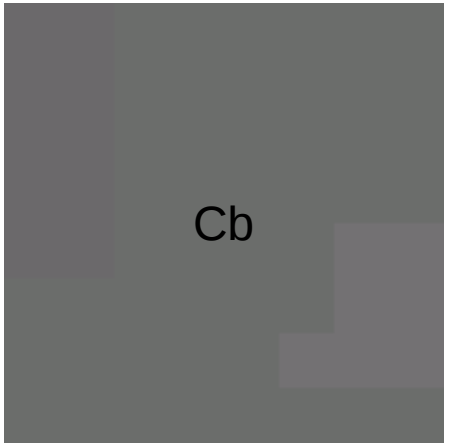
Quantized coefficients are losslessly encoded

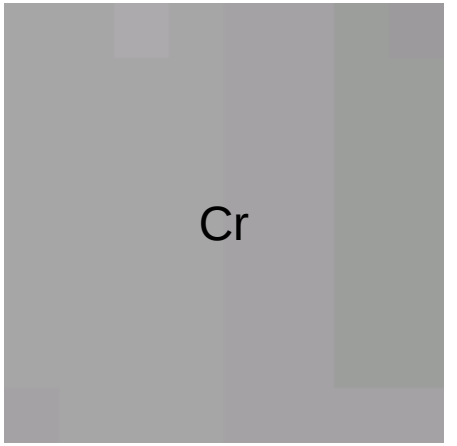
- Zig-zag pattern turns 2D matrix into 1D sequence
- Top left (DC) is encoded relative to previous block
- Rest of sequence is run-length encoded
- Each non-zero entry is coded as its value and the number of preceding zeros
- These symbols are Huffman coded

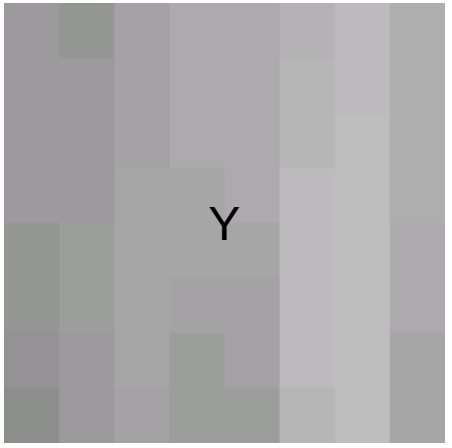


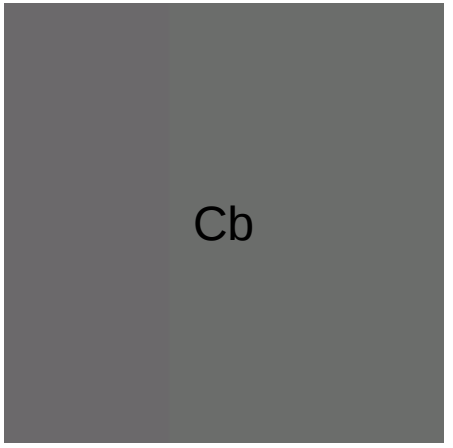
Reconstructed YCbCr channels

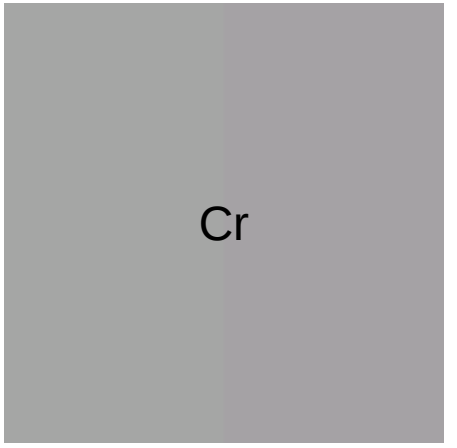




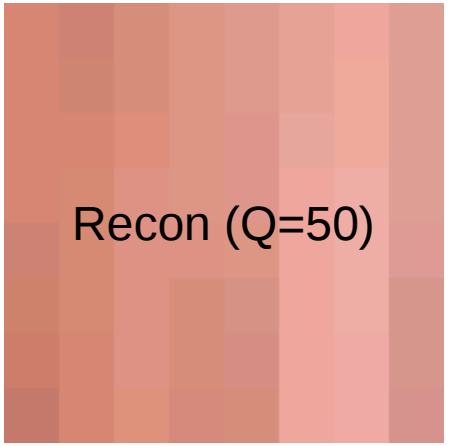


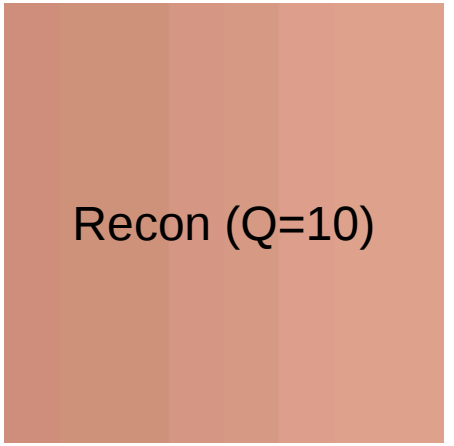




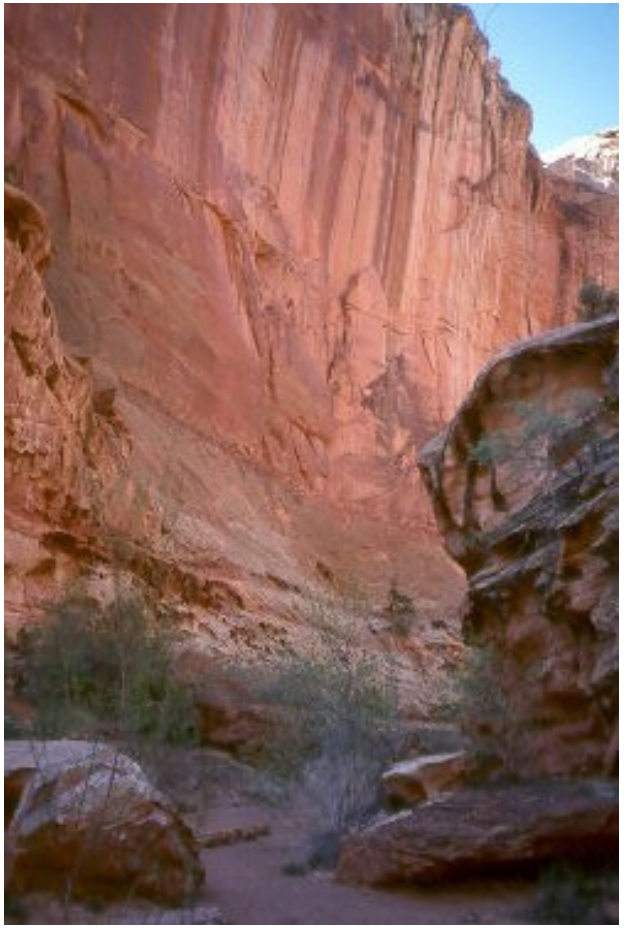


Reconstructed blocks

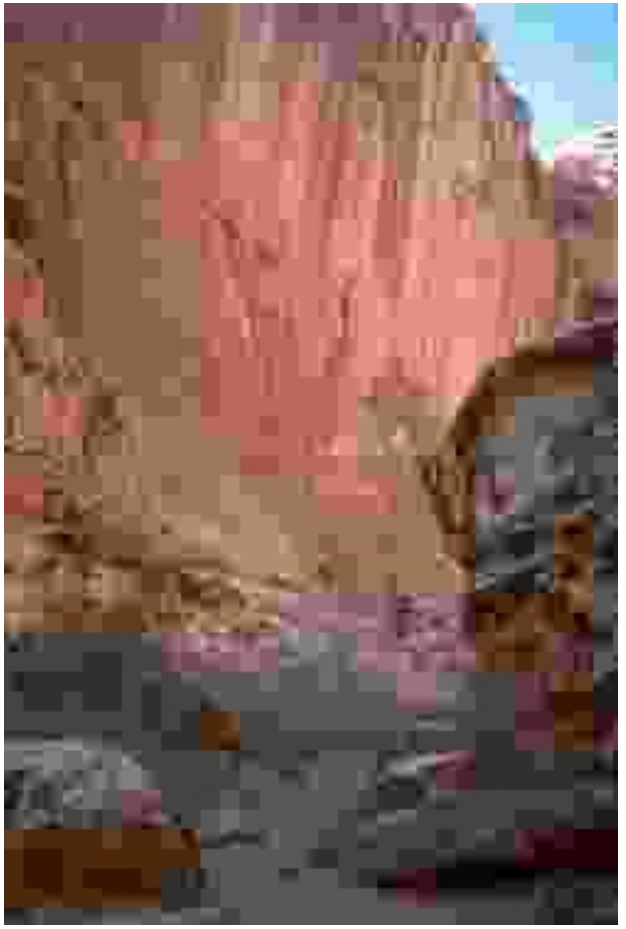




Reconstructed images



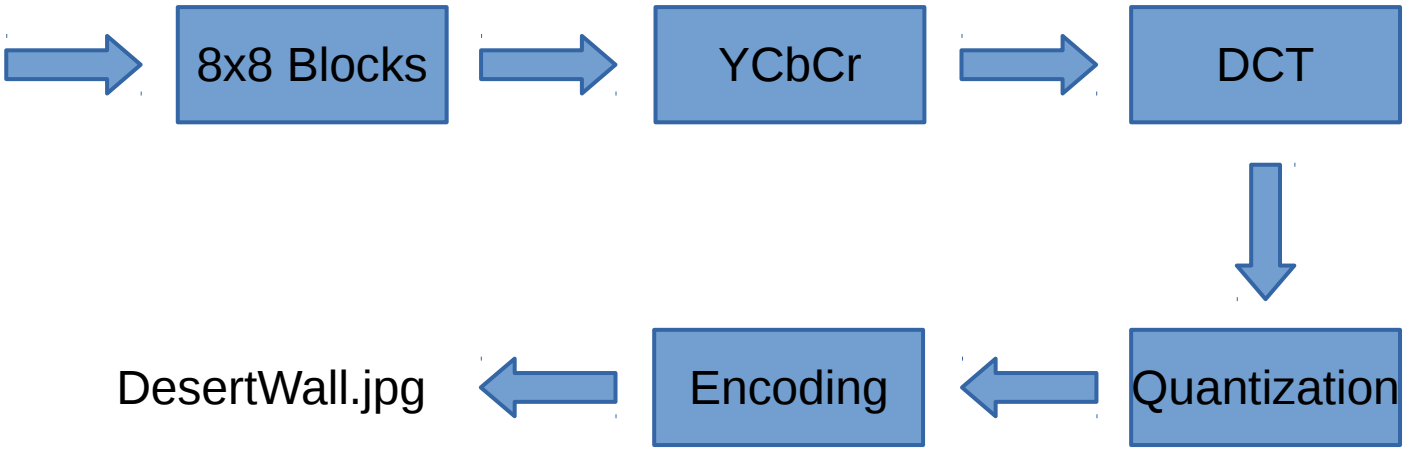
Q=50
size=32kB



Q=10
Size=1.7kB

Image compression: JPEG encoding

- For more information, see:
  - Austin, D. “Image Compression: Seeing What’s Not There.” AMS feature column, Jan 2008.
  - Wallace, Gregory K. "The JPEG still picture compression standard." IEEE Transactions on Consumer Electronics, 38.1, 1992



Audio coding: two types

- Based on properties of sound production
  - Properties of the voice or other sound source
  - Control a sound synthesizer accurately “enough”
  - May only work for source of interest
- Based on properties of sound reception
  - Properties of the ear
  - Represent sound accurately “enough”
  - Works for any sound

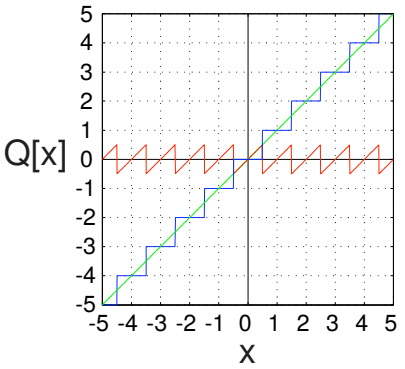
Images from: https://www.ee.columbia.edu/~dpwe/e6820/lectures/L07-coding.pdf

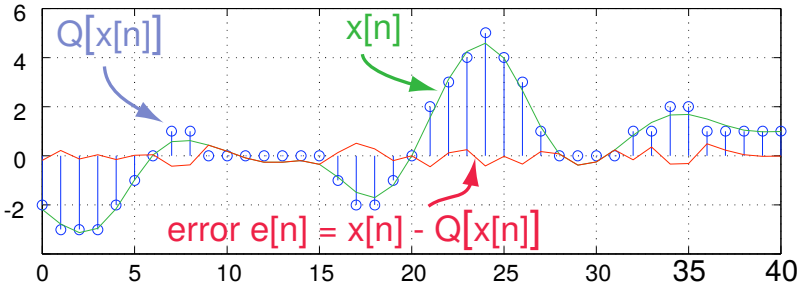
Perceptual audio coding: MPEG Audio

- We saw with JPEG that quantization is key to coding efficiency
  - Approximate the true image with one that is close, but shorter to encode
- Perceptual audio codecs approximate the true sound with one that is close, but shorter to encode
- Two signals that are imperceptibly different should have the same quantized representation

Quantization in audio

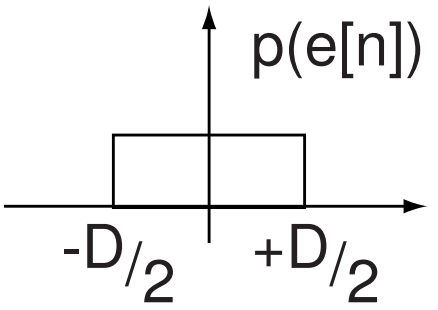
- Represent waveform with discrete levels





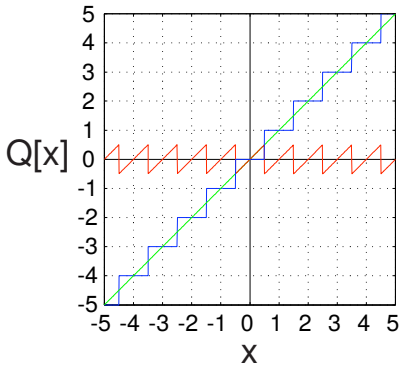
- Equivalent to adding an error of uniform noise

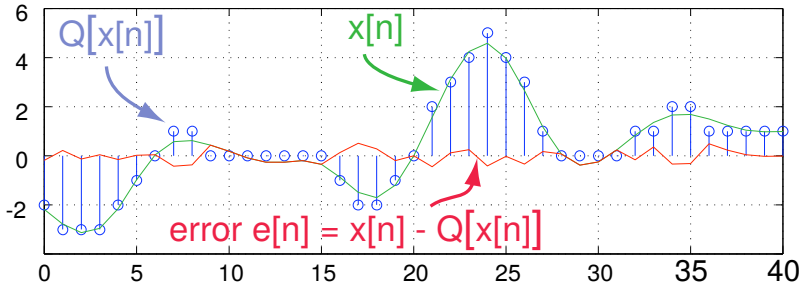
$$Q[x[n]] = \text{round}(x[n]/D)$$



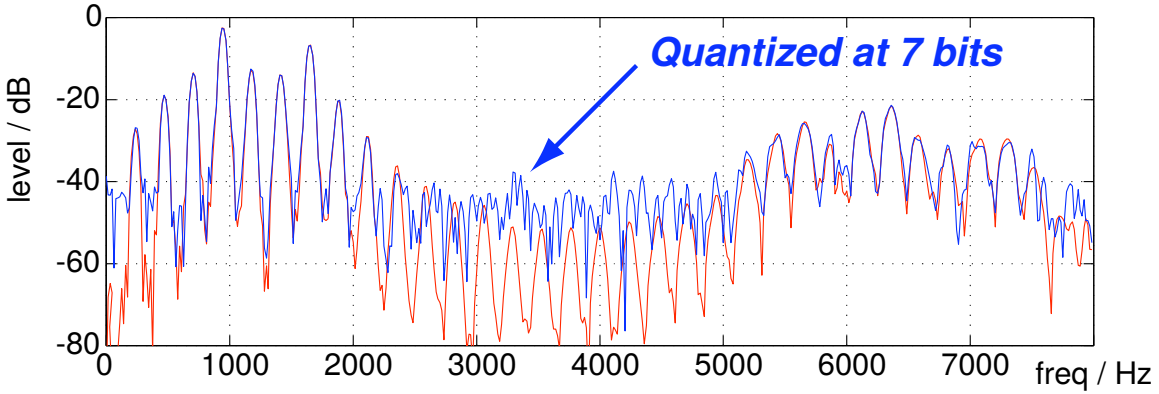
Quantization in audio

- Represent waveform with discrete levels



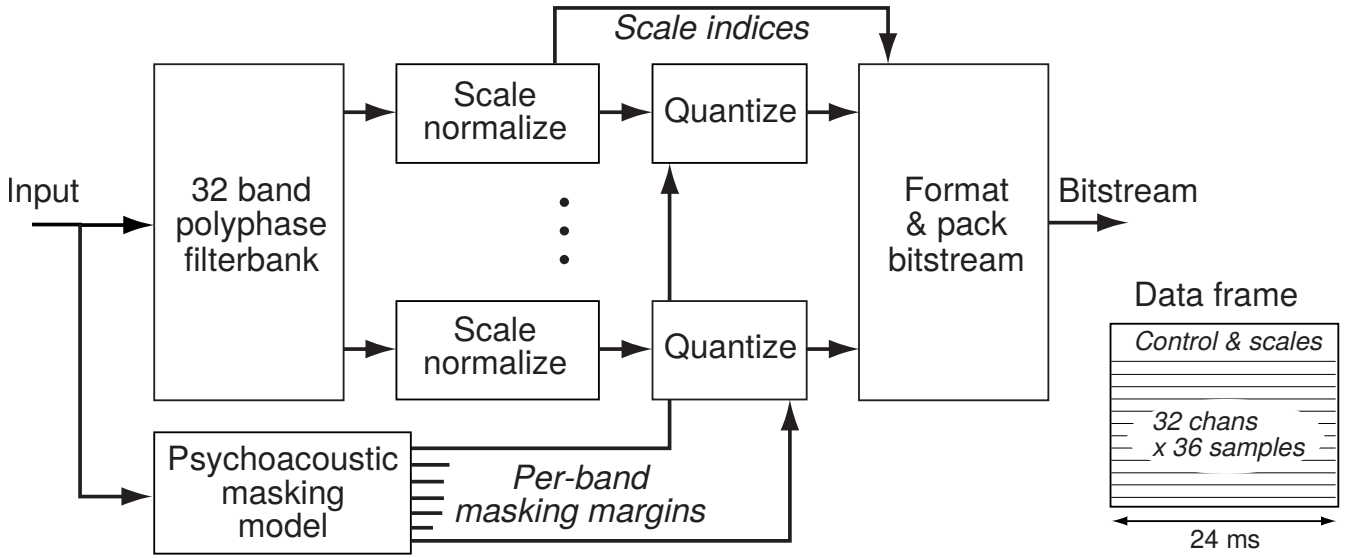


- Equivalent to adding an error of uniform noise



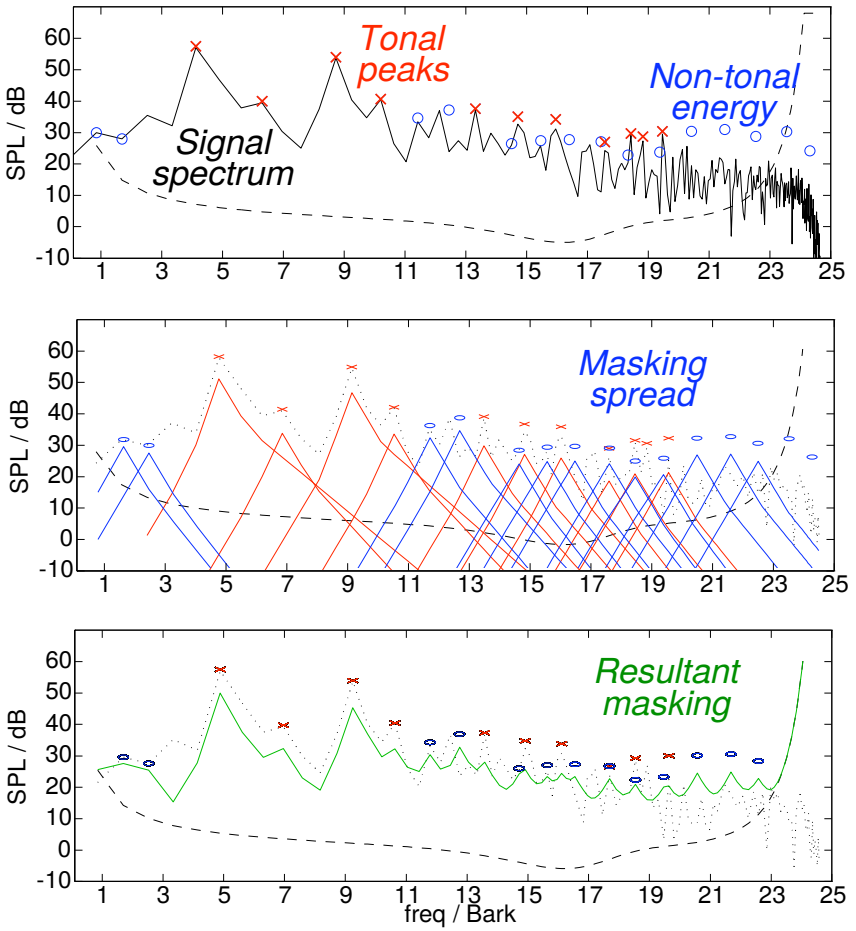
MPEG Audio basic idea

- Break audio into different frequencies
- Quantize each frequency as much as possible
  - While remaining imperceptible
- "Hide" quantization behind louder signals
  - Need psychoacoustic model of "hiding"



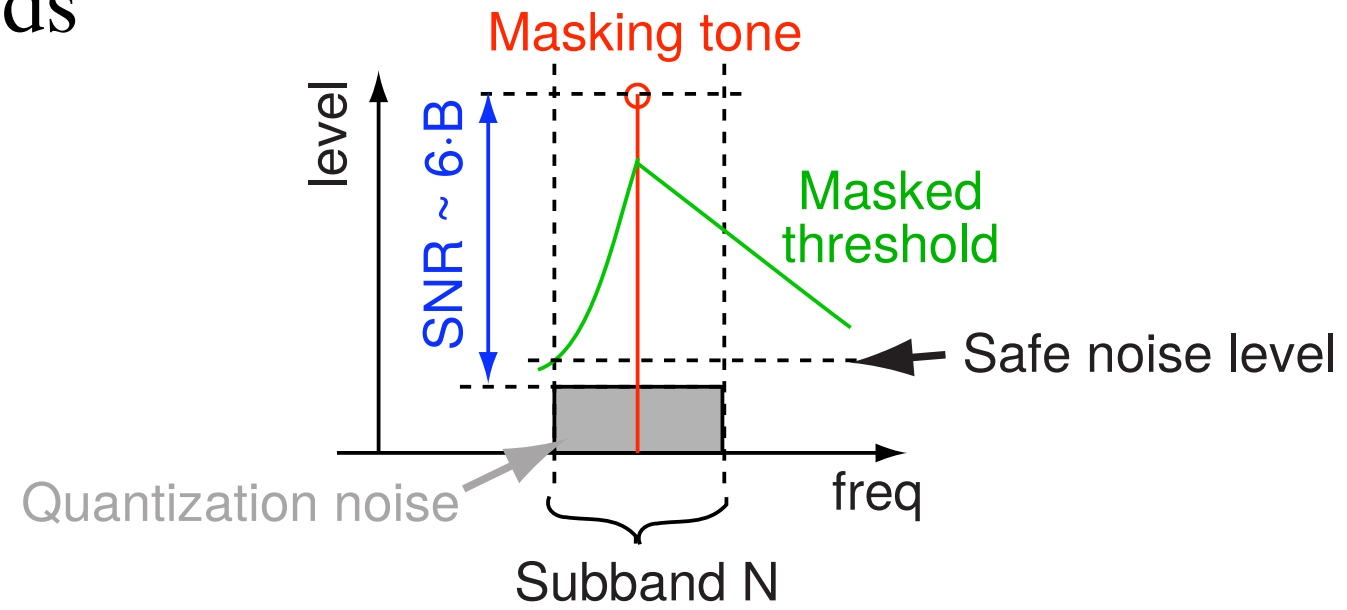
MPEG psychoacoustic model

- Model masking of sounds by tonal & noisy sounds
- Difficulties
  - Masking is non-linear in frequency and intensity
  - Complex-dynamic sounds not well understood



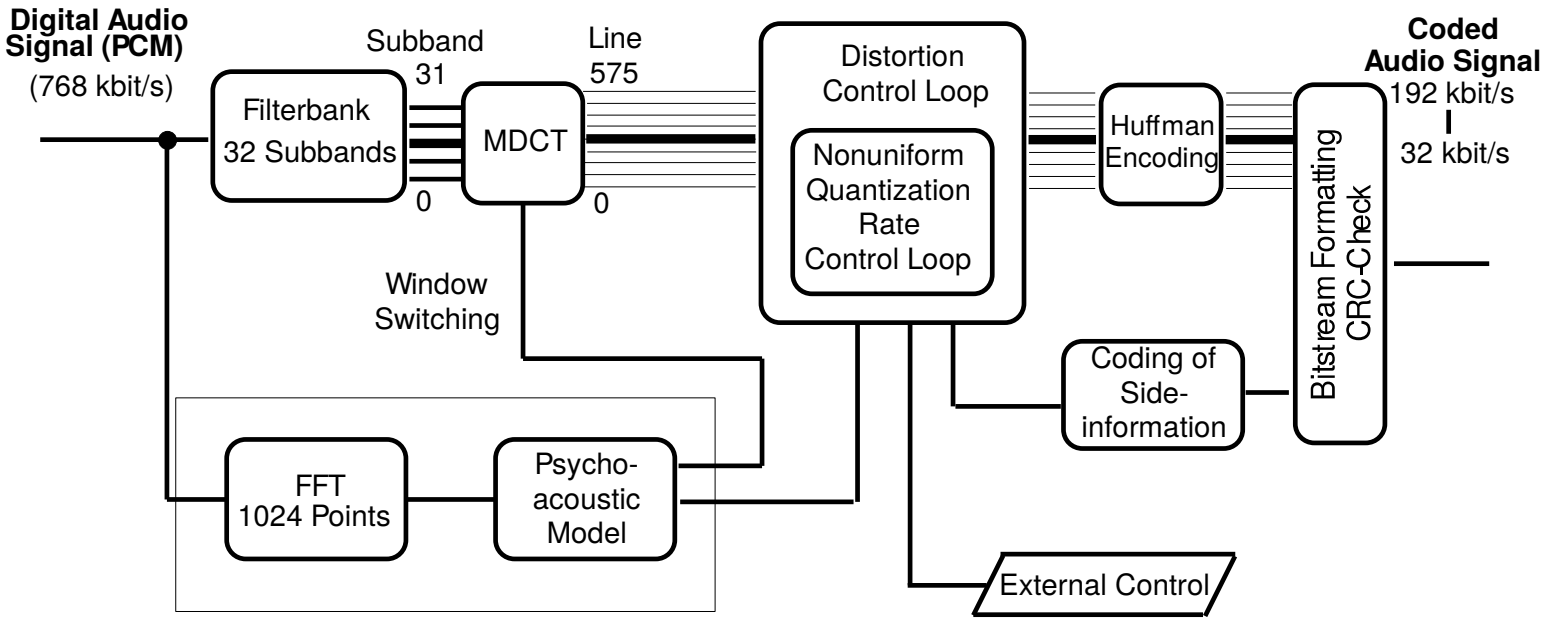
MPEG bit allocation

- Result of psychoacoustic model is maximum tolerable noise per subband
- Safe noise level → required signal-to-noise ratio → required number of bits
- For fixed bit-rates, may not be able to satisfy all subbands



MPEG Audio Layer III (MP3)

- Do an additional frequency analysis on subbands of layers I and II
  - Adapts to content (better time or frequency resolution)



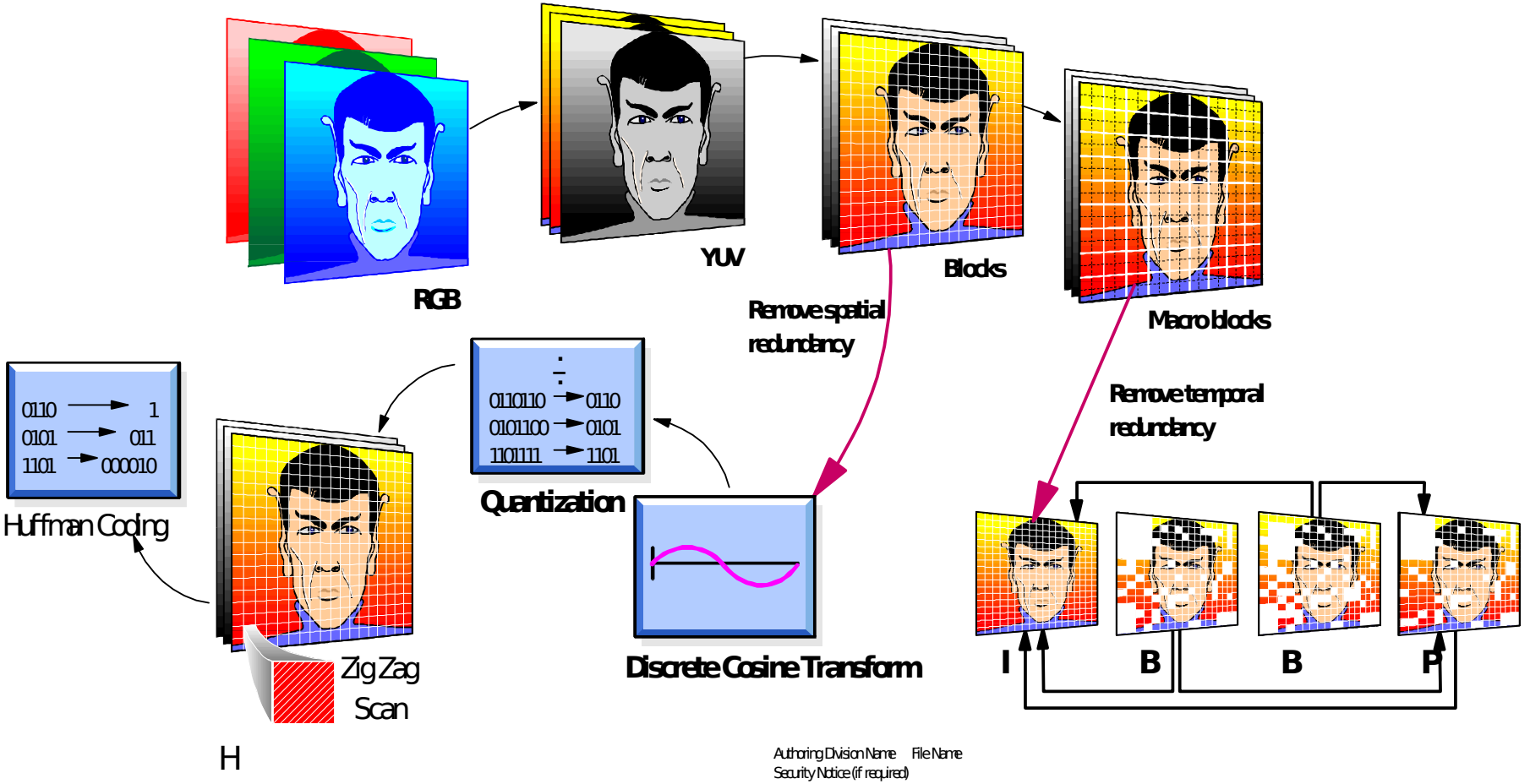
Perceptual audio coding

- For more information see:
  - Pan, D. "A tutorial on MPEG/audio compression." IEEE multimedia 2, pp60-74, 1995.
  - Painter, T., and A. Spanias. "Perceptual coding of digital audio." Proceedings of the IEEE, 88.4, pp451-515, 2000.

Video coding: MPEG2 and H.264

- Take advantage of redundancy in space and time
  - Predict pieces of frames from other frames
- Take advantage of irrelevance using quantization like JPEG
- Final lossless coding to squeeze out last bits
- Video codecs rely more heavily on “analysis-by-synthesis” than audio and image codecs

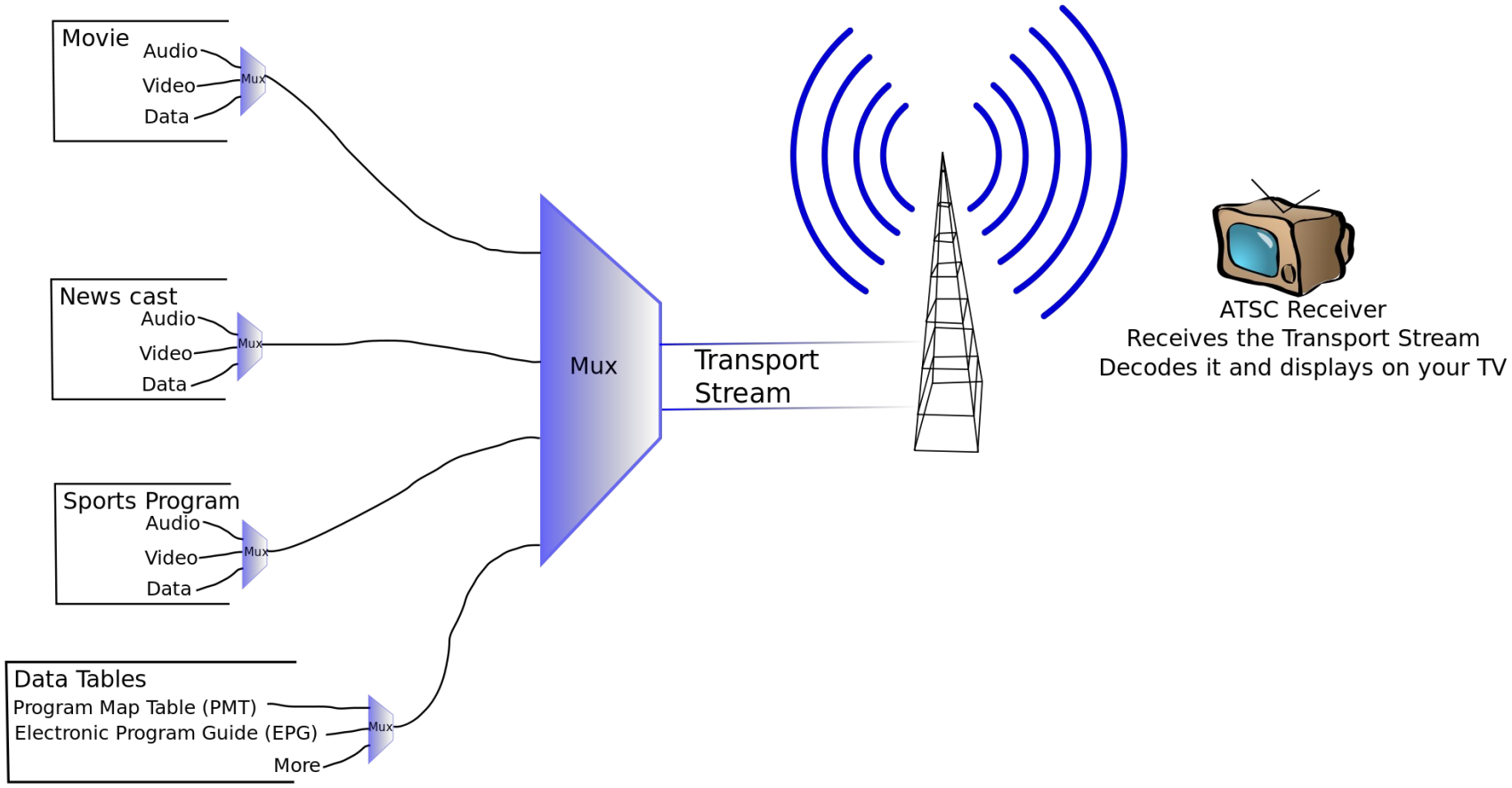
MPEG2 video encoding basics



Used in DVD, digital (HD) TV

Image from: http://www.keysight.com/upload/cmc_upload/All/6C06MPEGTUTORIAL1.pdf

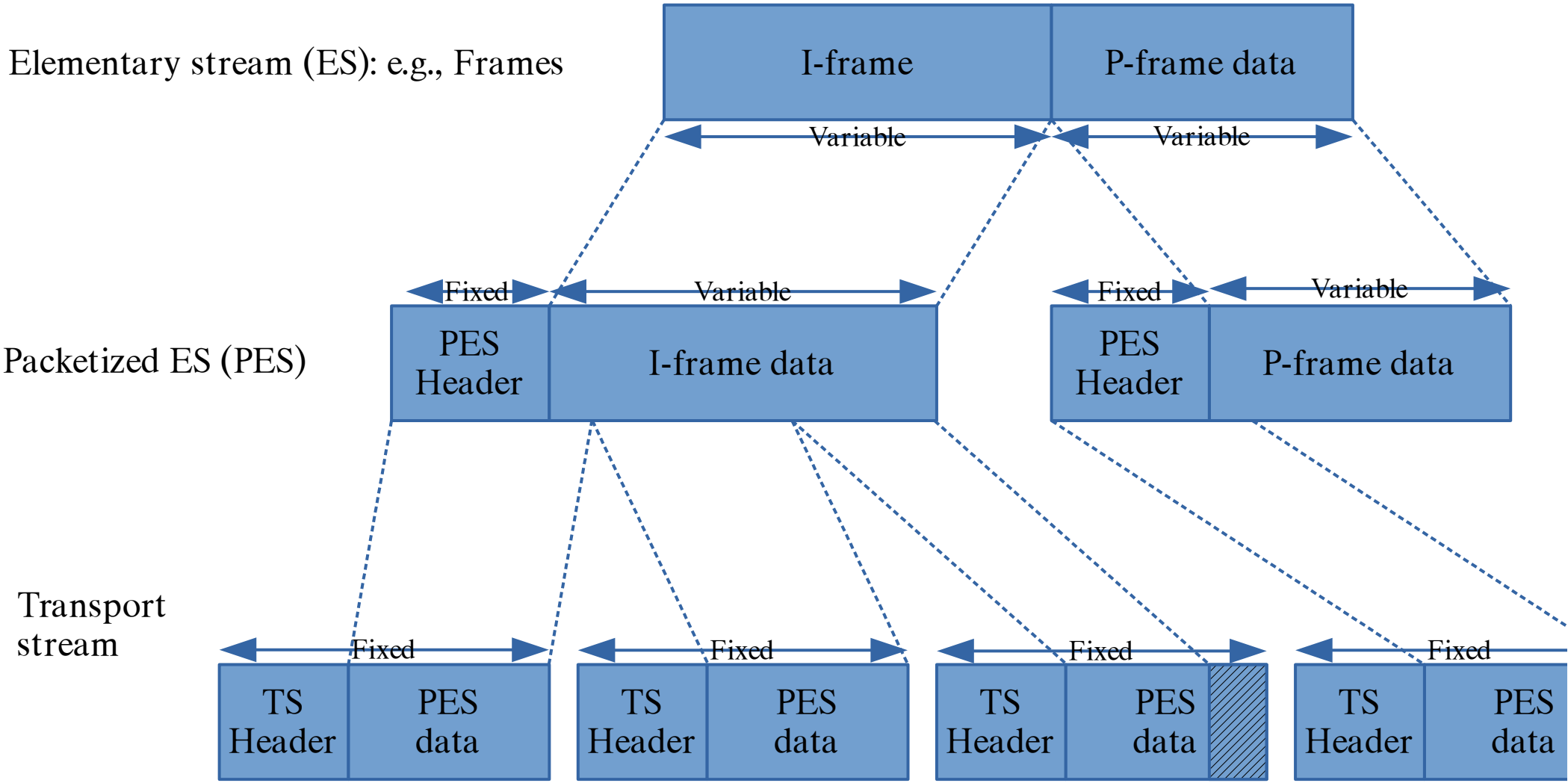
MPEG2 data is composed of streams



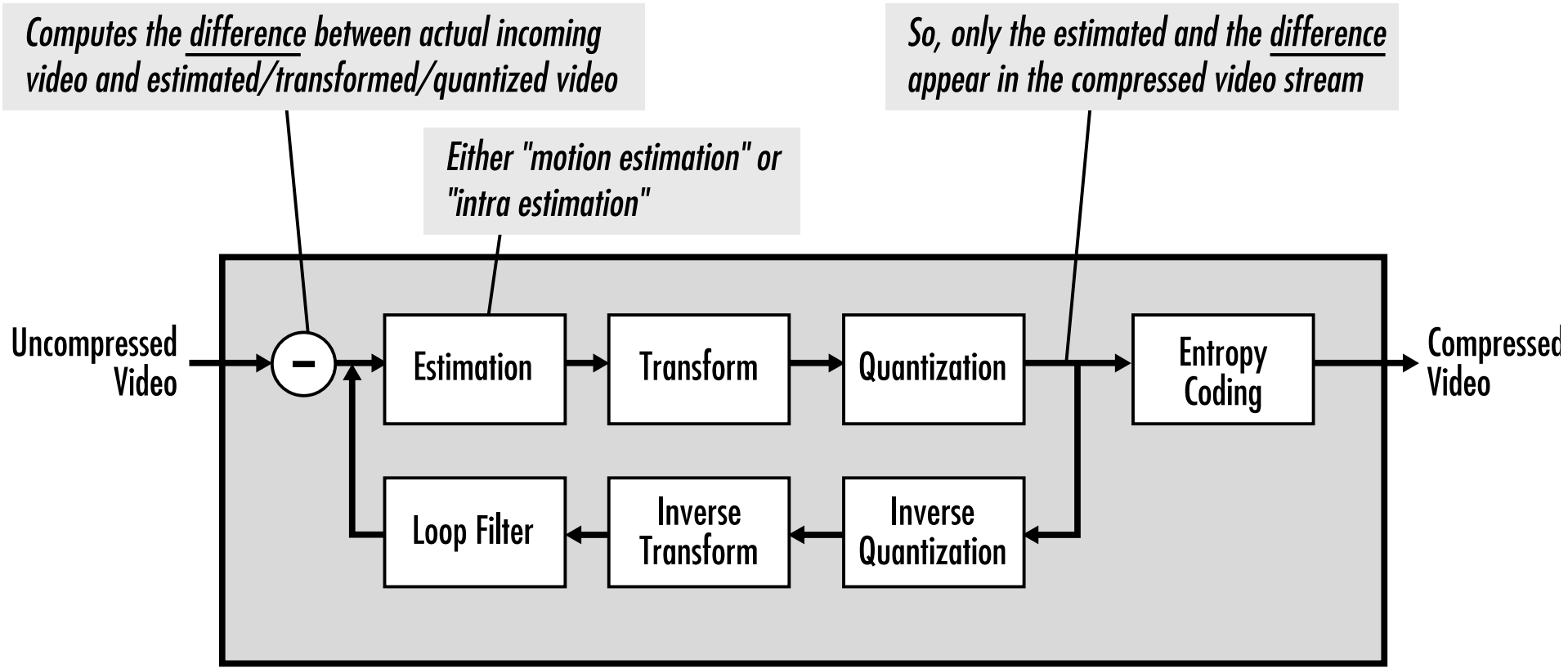
MPEG2 data is composed of streams

- Streams for different types of data
  - Audio, video, program information
- Streams are multiplexed together
  - Well-defined bistream “syntax” for keeping track of what came from where
- Different conceptual streams for coding, broadcast, disc-based storage
- Enforce synchronization through different clocks
- Information to make the stream seek-able

MPEG2 data is composed of streams



H.264 analysis-by-synthesis



H.264 Advanced video coding

- Also known as MPEG4 part 10
- Almost same stream format as MPEG2
  - But about 2x compression advantage
- Used in Blu-ray, YouTube
  - Widespread browser support
- Improvements over MPEG2
  - Better support for periodic movement
  - Better support for small motion blocks
  - Better lossless coding (arithmetic coding)

Video coding

- For more information:
  - Hewlett-Packard. "MPEG-2: The Basics of How It Works." 1999.
  - Navarro, A. “Introduction to digital television.” Slideshare, 2012 http://www.slideshare.net/chetanrao2012/introdgitaltv
  - Wiegand, Thomas, and Gary J. Sullivan. "The H. 264/AVC video coding standard." IEEE Signal Processing Magazine, 24.2, pp148-153, 2007.

Summary

- Compression lets us store data efficiently
- Remove data that are redundant or irrelevant
- Redundant: implicit in remaining data
  - Can be fully reconstructed (lossless compression)
- Irrelevant: unique but unnecessary
  - For example: imperceptible to humans (lossy compression)