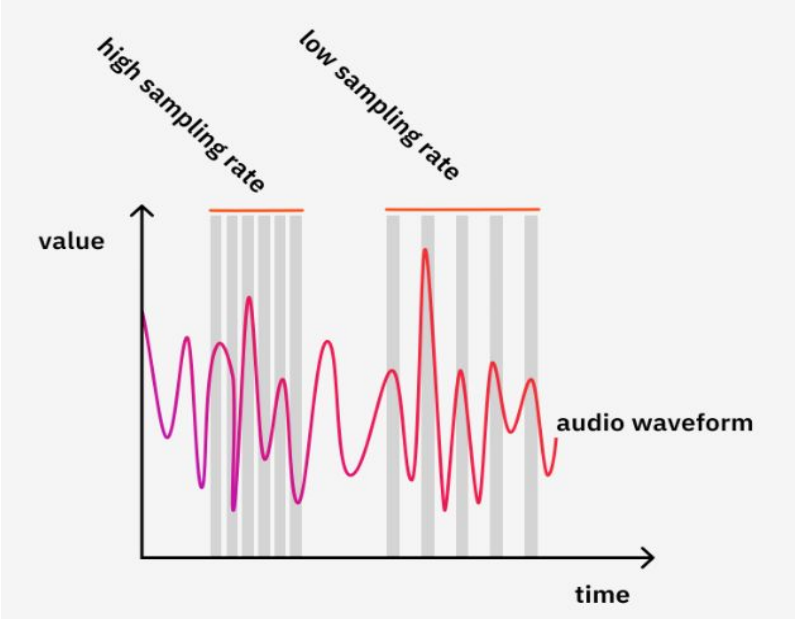
FFMPEG Tutorial

Preliminary terms:

- Audio
  - Sampling rate
  - Bitrate
  - Channels
- Image
  - Resolution
  - Bit-depth
  - Transparency
- Video
  - Resolution
  - Framerate
  - Codec

Audio - Sampling rate

- “The sampling rate is the factor that shows how many times we measure/scan/sample the input data stream.”



Audio - Bitrate

- “Bitrate represents the amount of data per second that results from our transcoding/quantization process.”

- For example, If it is 1411 kbit/s, that means that for every second of audio data, about 1411 kbit of output data will be produced.

Audio - Channels

- “Channels are simply separate recordings or streams or audio signals within an audio recording”
- Hear this binaural recording from Big Buck Bunny
- There could be more channels as well:
    - 2.1 audio, has 3 channels, 2 for stereo and one for low frequency effects aka bass

Image - Resolution

- “The resolution of an image determines how many columns and rows of pixels are in an image.”
- Video resolution is also image’s width times the height

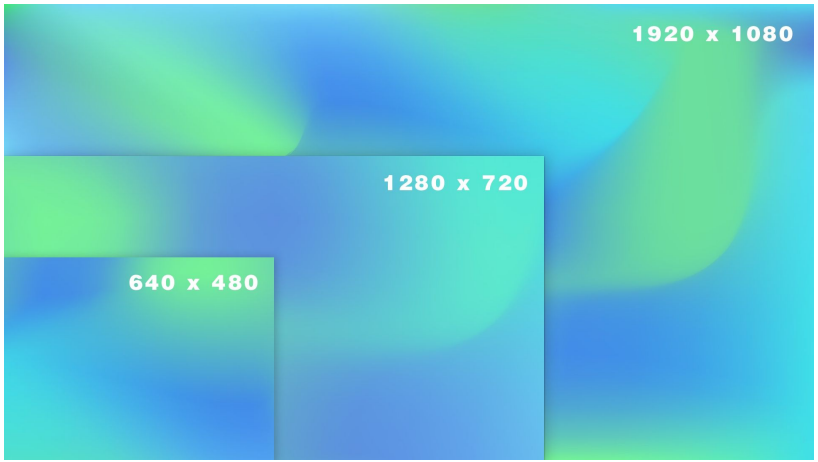


Image - Bit depth

- “Bit-depth represents the number of bits used for storing a single pixel's color value.”

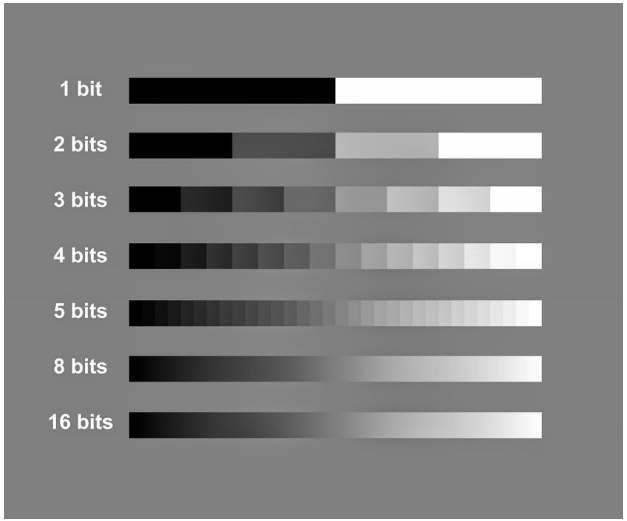
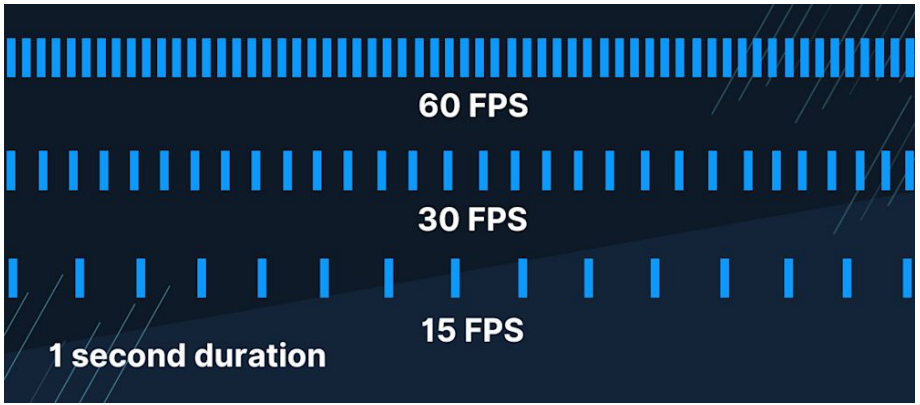


Image - Transparency

- “The alpha channel (transparency) determines how transparent a single pixel is ”
- Bit depth controls the transparency values, a single bit alpha channel can encode a pixel to be either transparent or non-transparent

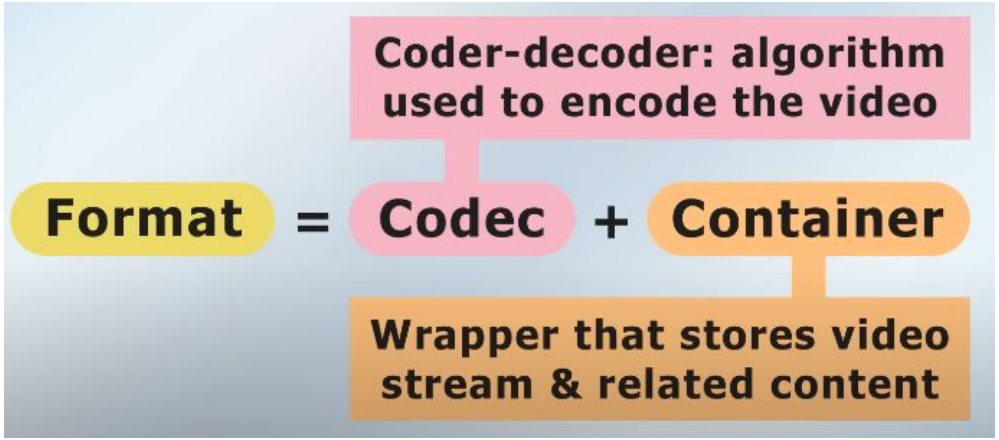
Video - framerate

- “Framerate defines how many images are shown in a second ”



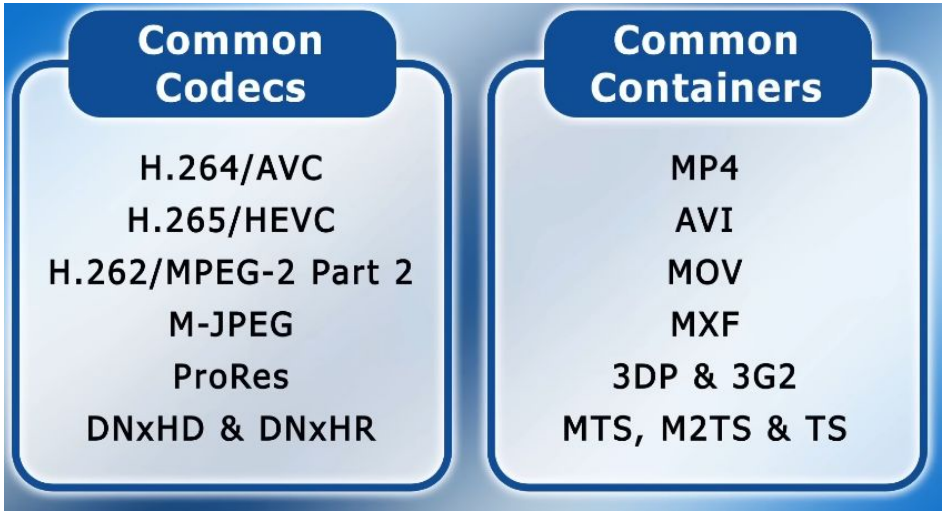
Video - codec

- “Codec defines the format how images are compressed into a video file
- Video format is not the codec!



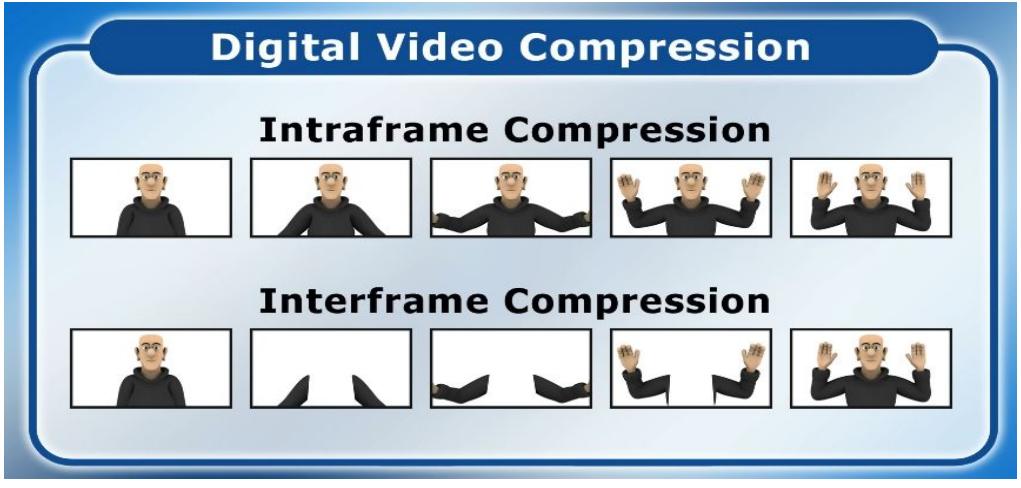
Cont …

- By looking at video container, we can not tell which codec was used to compress it



How codec work at a high level?

- The basic idea behind codec is to reduce data size by exploiting redundancies in video/images
- Intraframe compression compresses each frame individually (e.g M-JPEG, ProRes,DNxHD)
    - advantage: each frame is ready-to-fetch, high quality video
- Interframe compression stores some frame as main frames (I-frames) and next frames as delta frames, which contain incremental changes (e.g H.264/AVC, H.265/HEVC)
    - Advantage: produces smaller video files

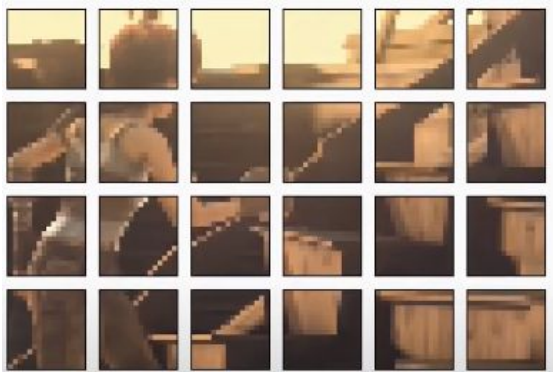


How efficient codecs are!

- Let’s say we want to compress a video 24fps video that is 1920*1080 (RGB) with a bit-depth of 8 bits/pixel
  - b_raw = 1920*1080*8*3*24
  - b_raw = 1194 Mbps
- let's say a compressed version of video transmits 6Mbps, then the compression factor f would be:
  - f = b_compressed / b_raw
  - f = 6Mbps / 1194 Mbps
  - f = 1 / 200
- With new video codec, we can achieve even a compressed video at 2-3 Mbps

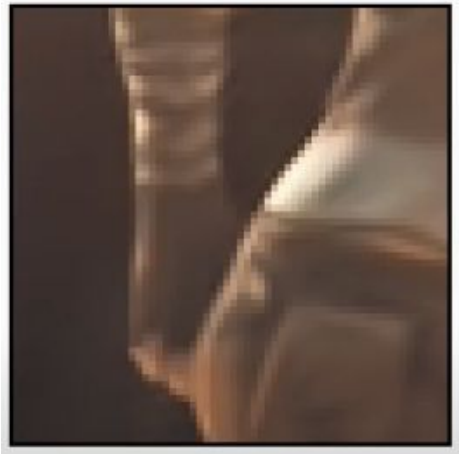
Overview of how codec works

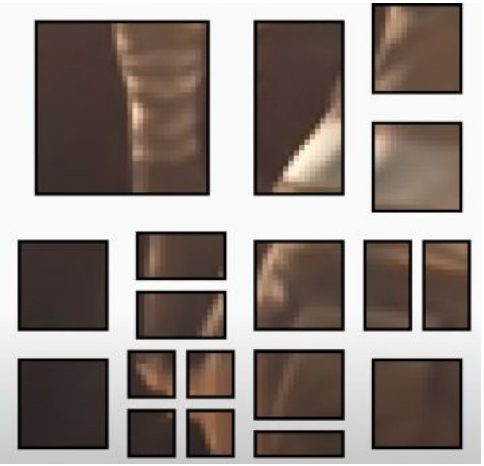
- Each codec begin by splitting the frame into n*n blocks
- Each block is then processed separately
- In modern codecs, the block size can go as large as 128*128
- Different codecs divide the blocks differently to utilize more threads to process quickly e.g tiling



Overview of how codec works

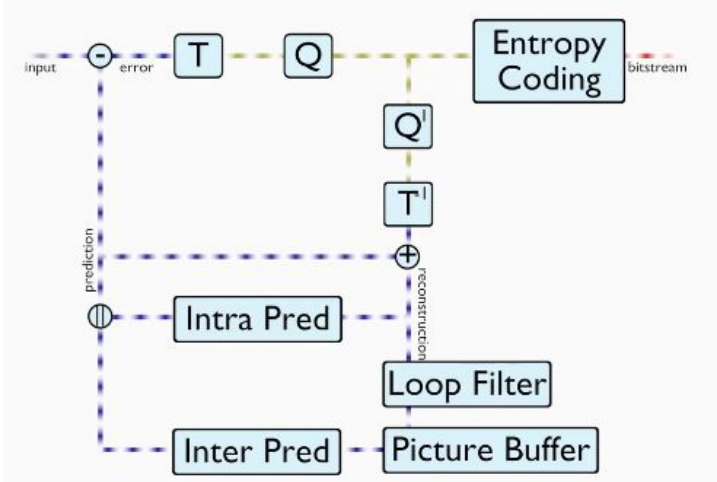
- In the next step, each n*n block is further split into a recursive-tree like structure
- Asymmetric splits also possible during the splitting process
- The smallest blocks now will go through an encoding loop





Overview of how codec works

- First, the difference between prediction and original block is calculated
- After that, a DCT/ DST is applied to convert to frequency domain
  - redundancy is removed more in frequency domain due to de-correlation property of transforms
- The transformed image is then quantized (converting floating values to one specific value)
  - Here most of information in video is lost. Higher quantization effects leads to lower quality decoded video
- The loop filtering step improves the overall quality of image e.g smoothing filters etc



Overview of DASH

- DASH is an adaptive bitrate streaming technique that enables high quality streaming of media content over the internet