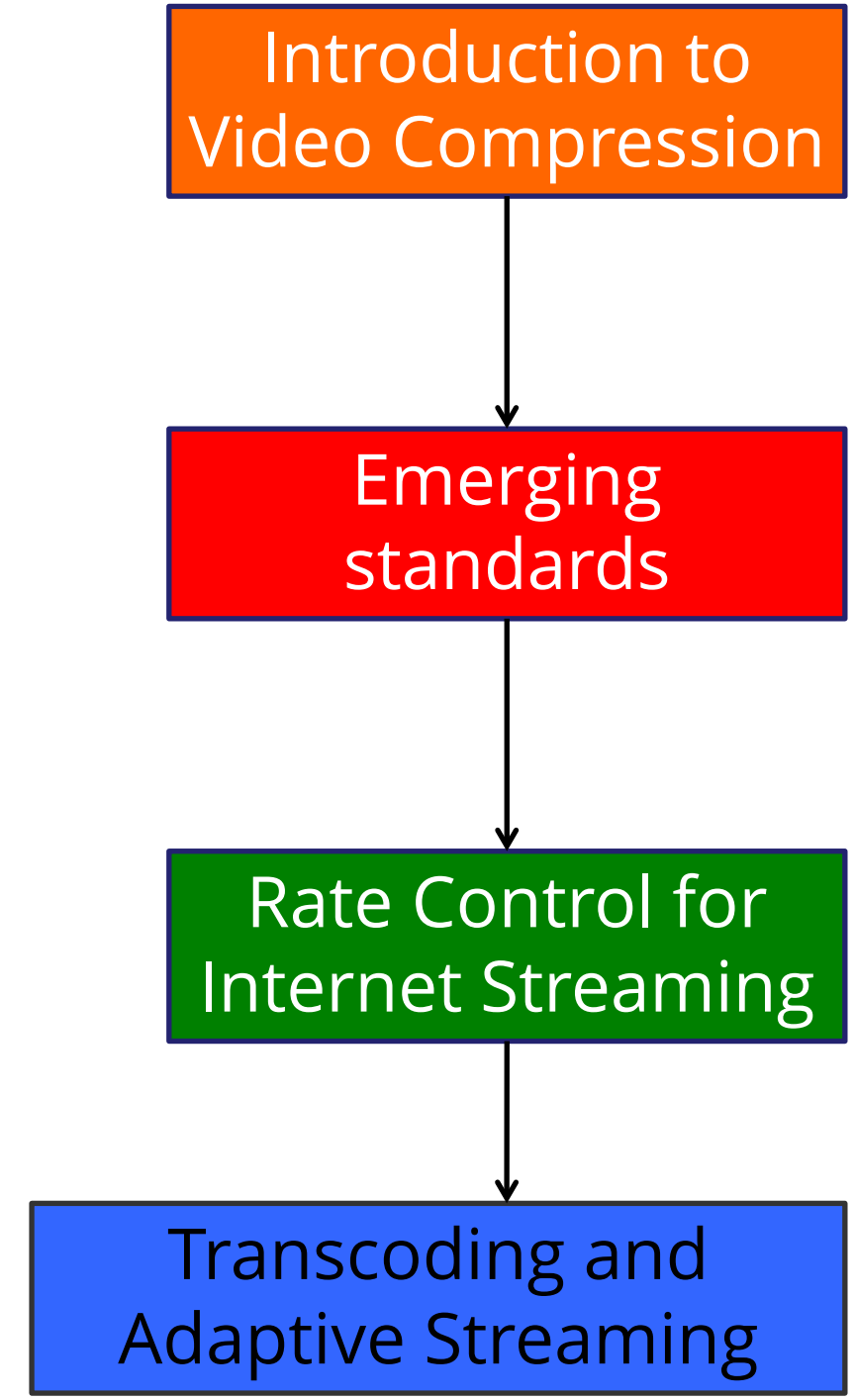
Motion Picture Engineering

Video Compression and Transcoding

Content


Introduction to Video Compression
Emerging standards
Rate Control for Internet Streaming
Transcoding and Adaptive Streaming

The entropy of a random variable X with a probability mass function $p(x)$ is defined by

$$H(X) = -\sum_x p(x) \log_2 p(x). \tag{1.1}$$

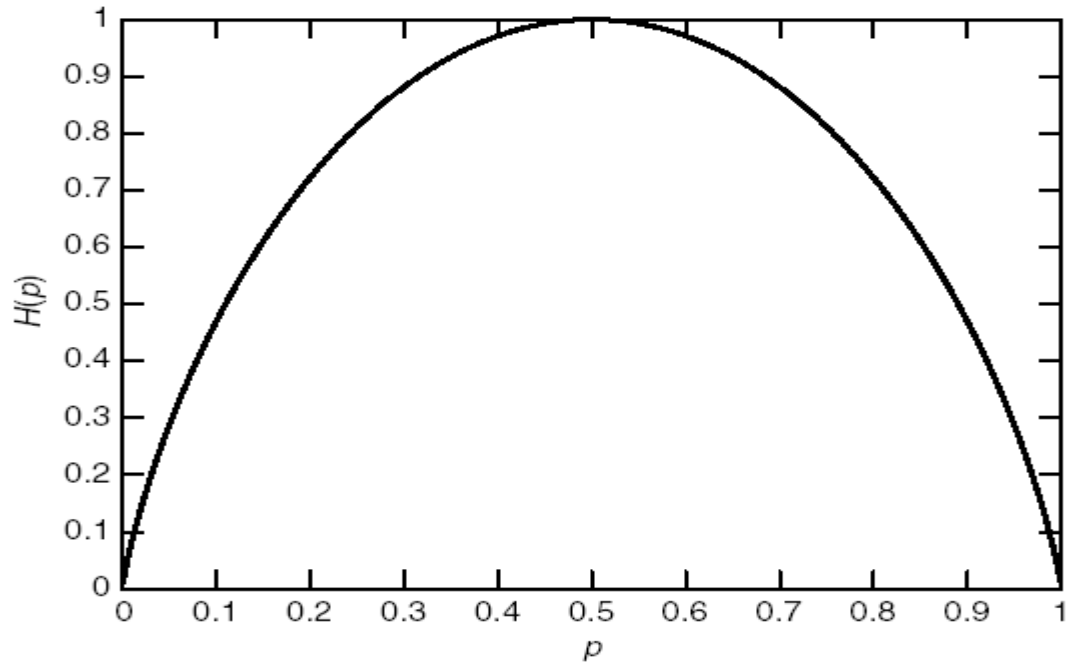
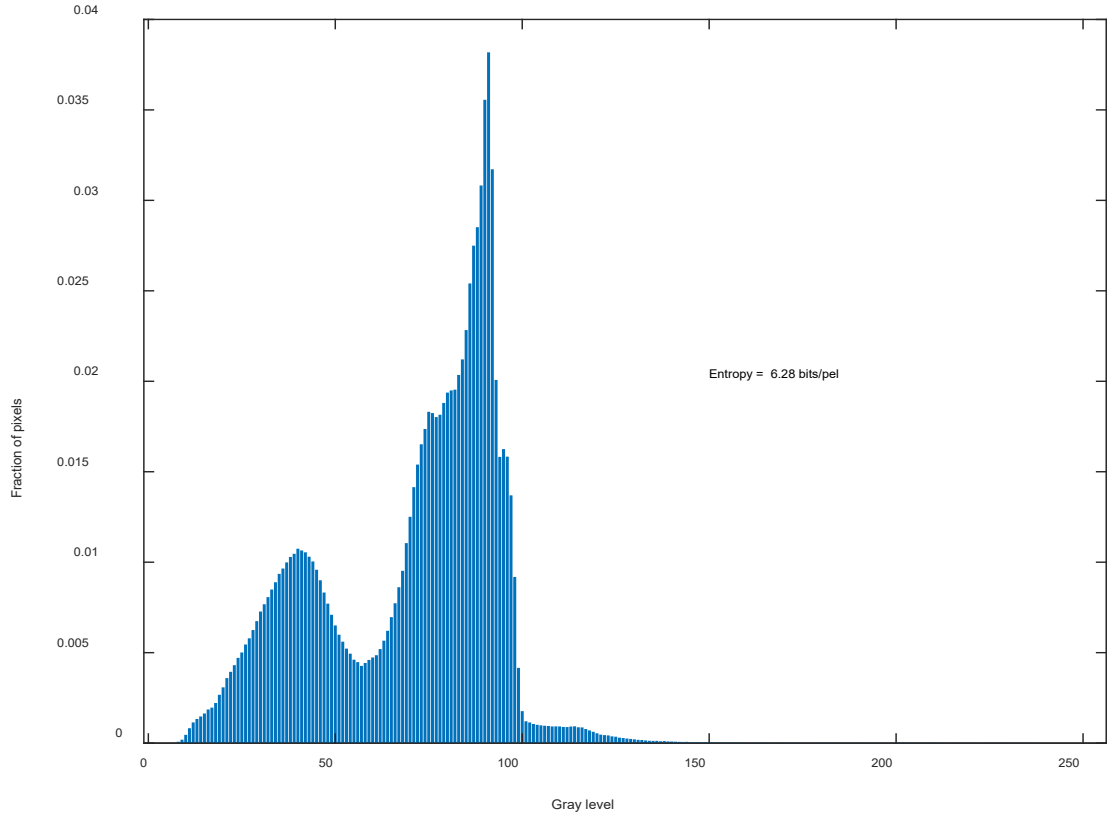


FIGURE 2.1. $H(p)$ vs. p.

Shannon's Information Limit : Entropy Measured in Bits/pel

Information content in the raw 8bit image

Raw 8bit image 1080p

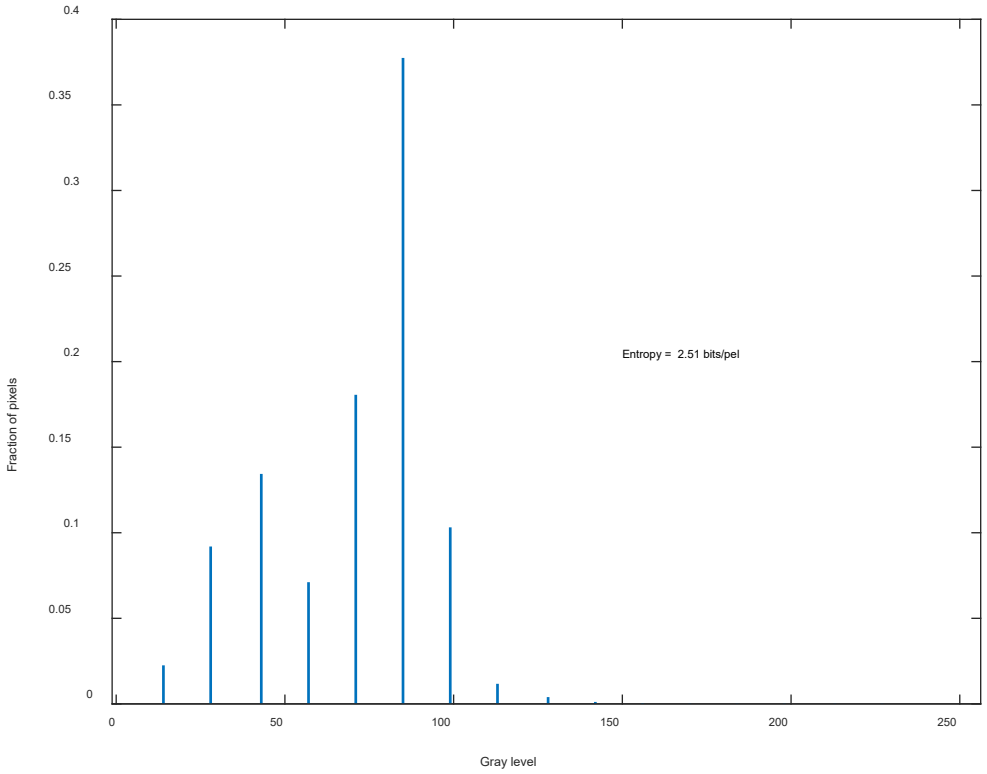


Shannon's theory tells us that 6.28 bits/pel is the best we can do

Compression by quantising the image itself

Throw away bits in the raw image to achieve compression. But it comes at the cost of reduced image quality

Raw 8bit image 1080p Quantised using Qstep = 15
i.e. gray values rounded off to $0{:}\,15{:}\,255$
PSNR doesn't seem that bad … but banding "perceptually" really bad
More evidence that PSNR isn't that great



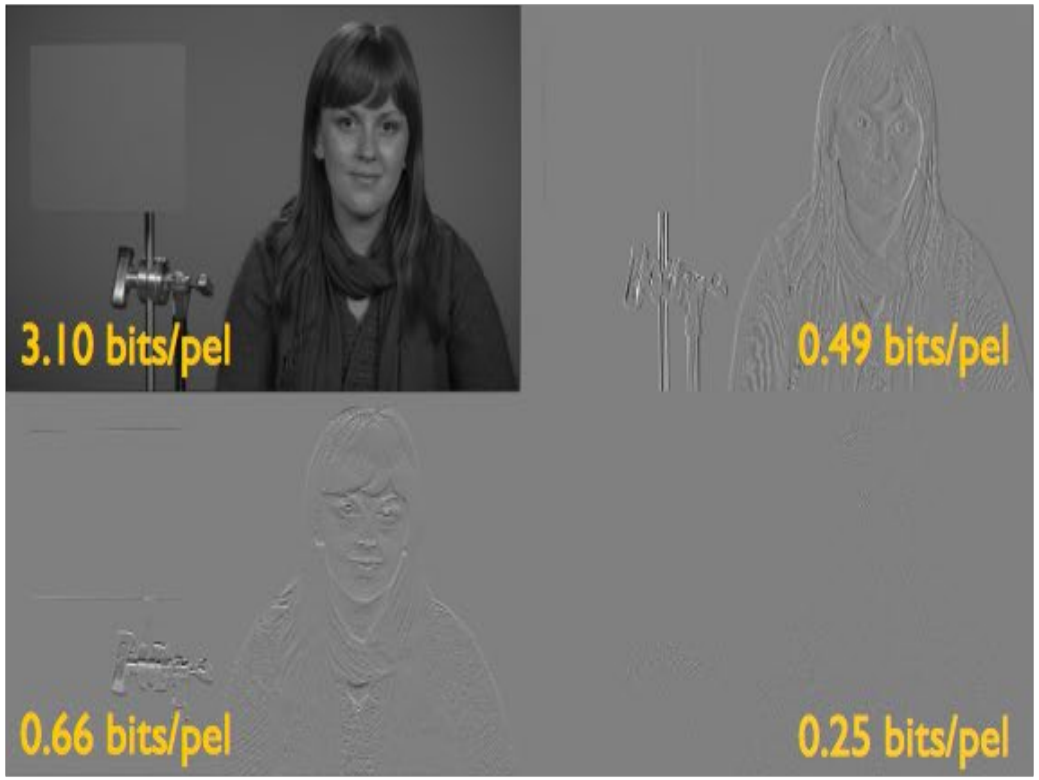
Shannon's theory now tells us that 2.51 bits/pel is the best we can do. We have achieved compression!

We can do better using image transforms before quantising

Q=15

Quantised using Qstep = 15 IN THE TRANSFORMED DOMAIN!
i.e. gray values rounded off to 0: 15: 255
Perceptual quality almost lossless and at 1.12 bits/pel its way better than quantising in the raw image space. Factor of 2 better, with better picture quality !

Q=15

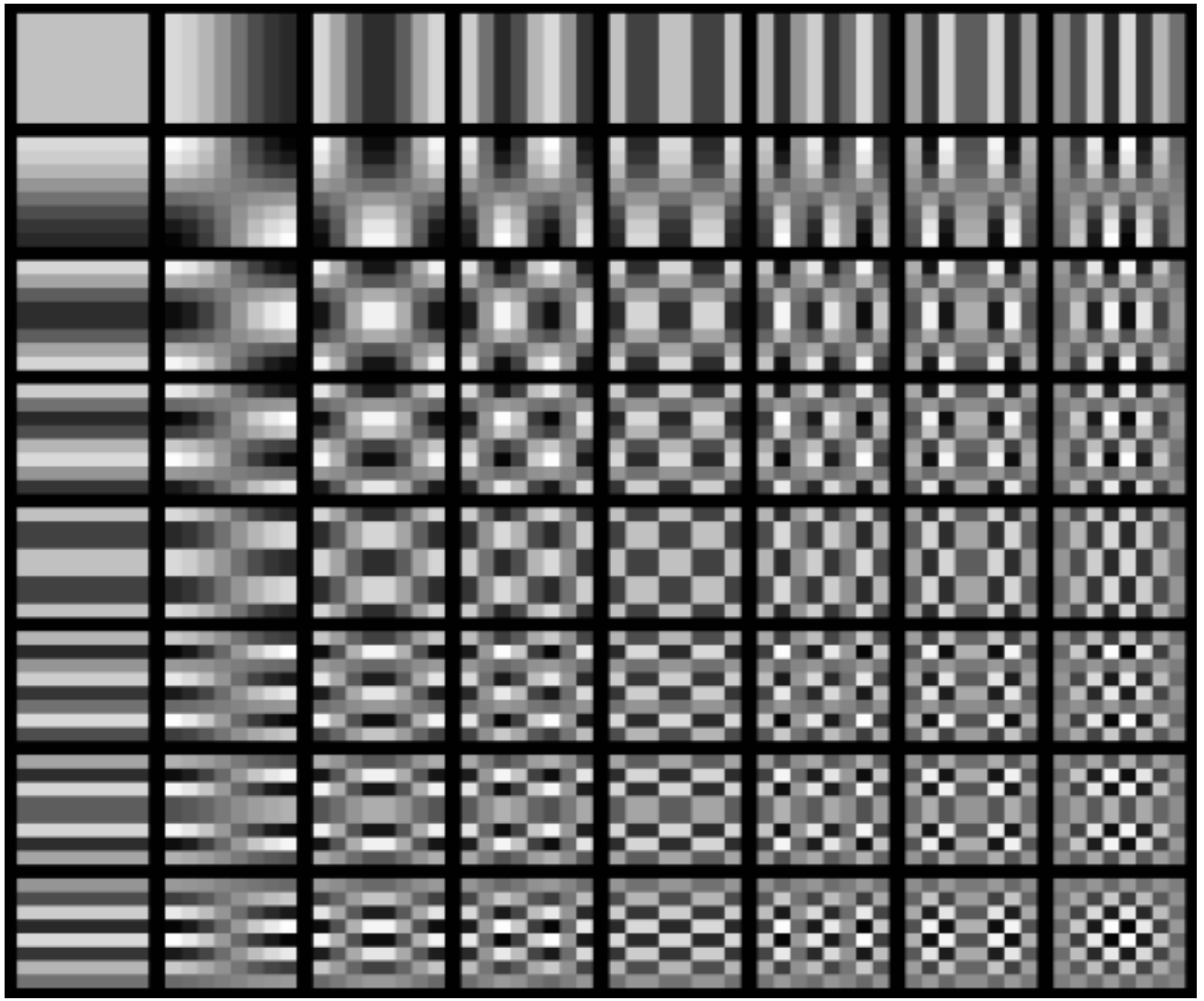


1.12 bits/pel

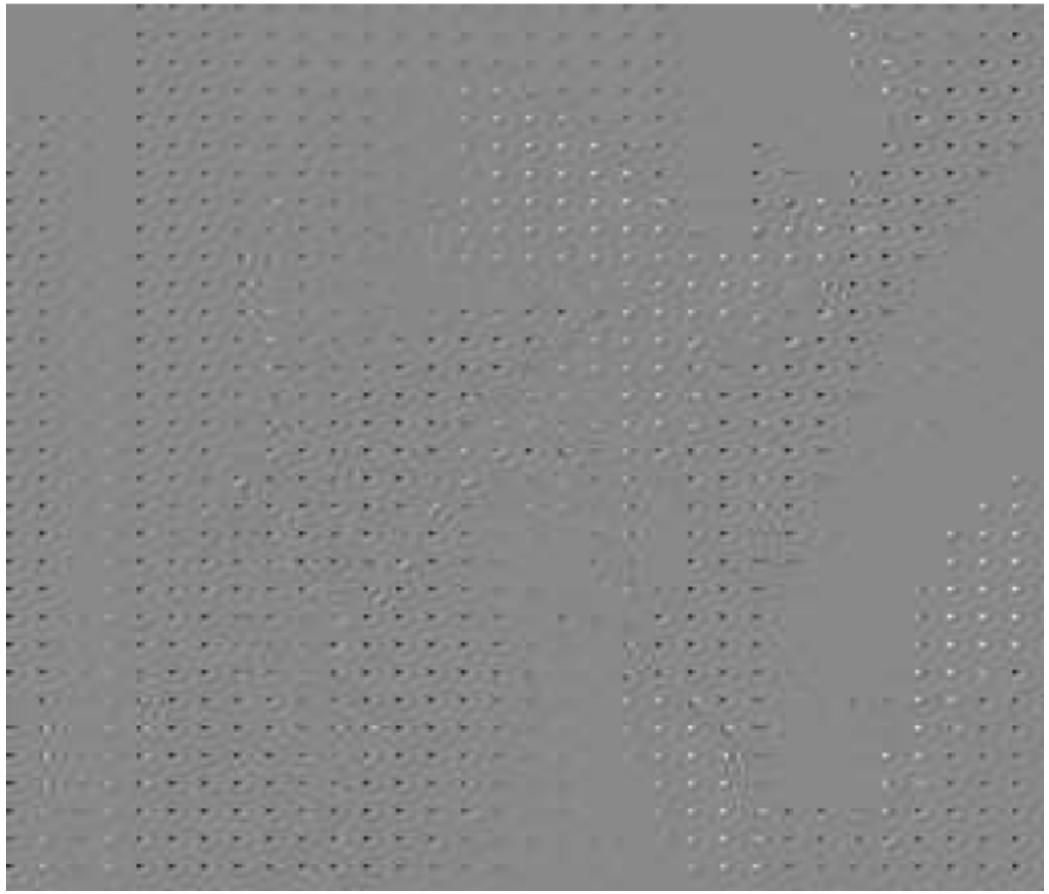
Haar Transform

DCT Basis Functions for an 8x8 Block

All modern compression schemes use the DCT (Discrete Cosine Transform)



Still Picture Coding Established the Idea with JPEG



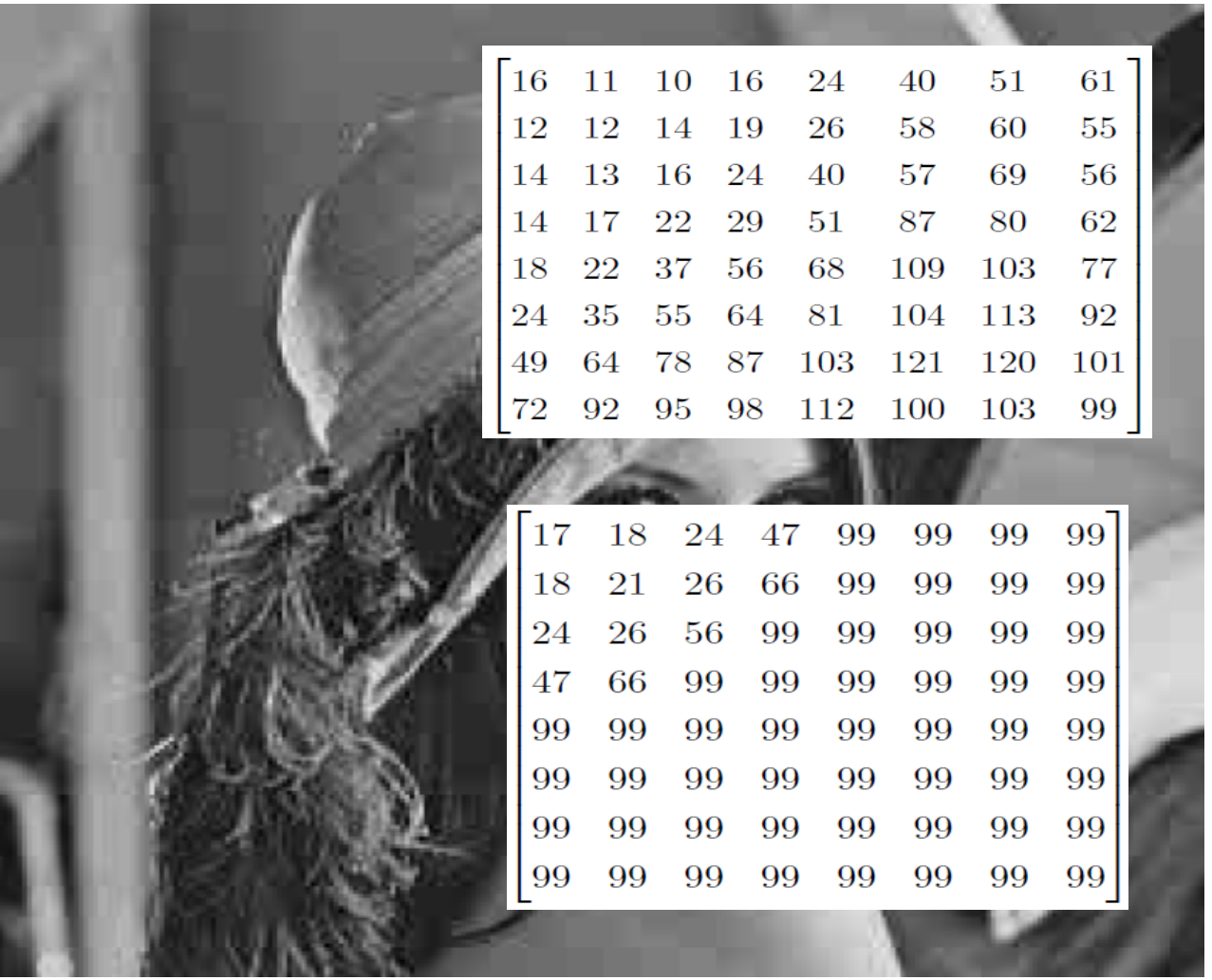
1.3 Bits/pel after rounding to nearest “15”

Tradeoff between Quantisation and Reconstruction Error

Rounding to nearest 15 = 1.3 bits/pel

Rounding to nearest 30 = 0.81 bits/pel

Tradeoff between Quantisation and Reconstruction Error



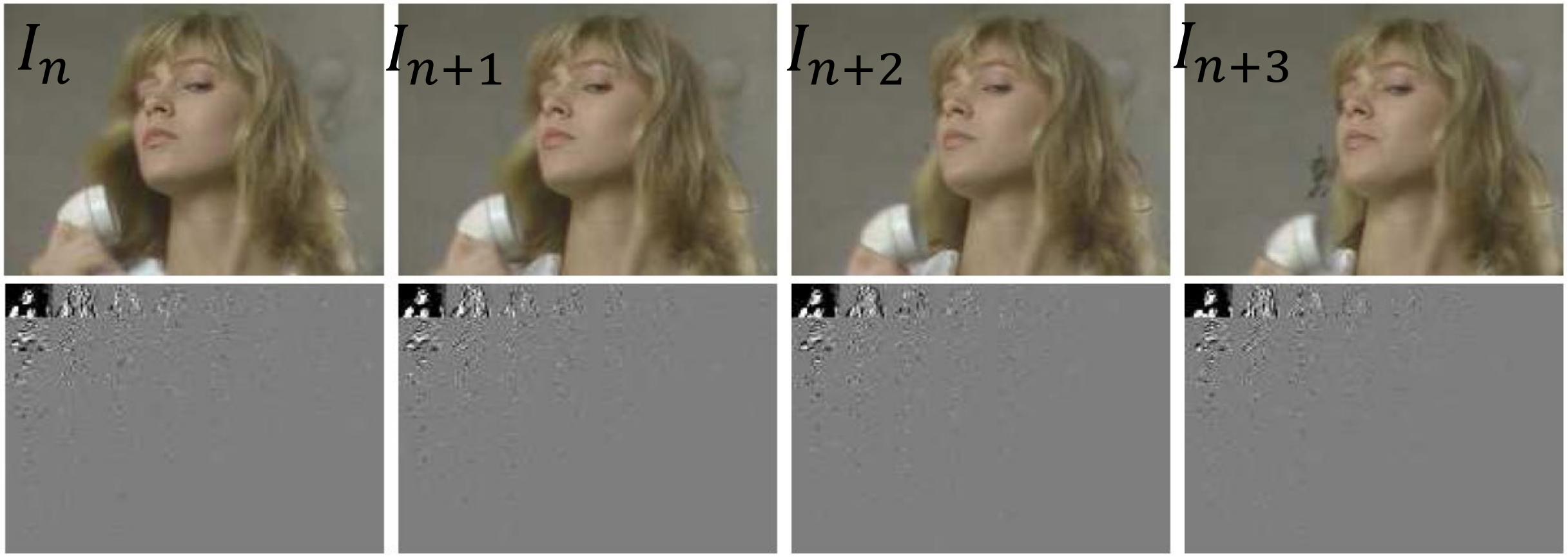
$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120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

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

Rounding to nearest 30 = 0.81 bits/pel

JPEG Quantisation Matrix = 0.85 bits/pe

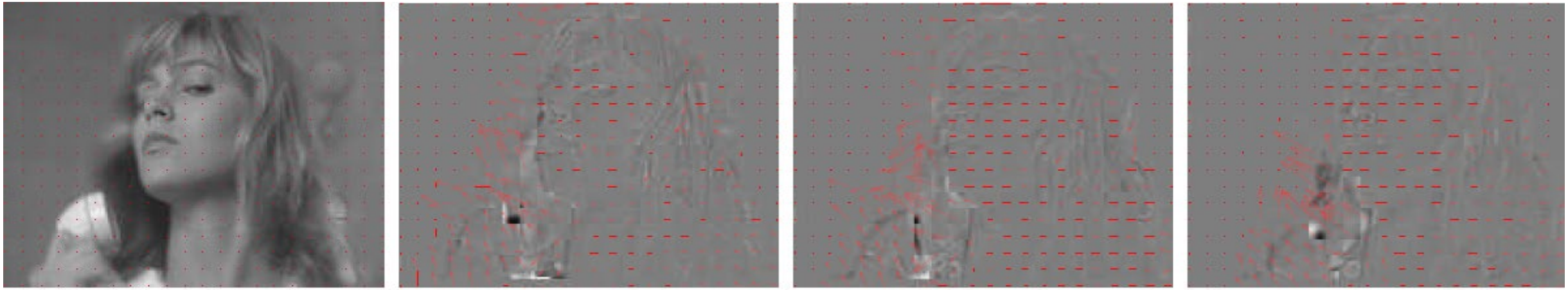
Compressing Video



Transforming the Picture sequence into a Difference Sequence

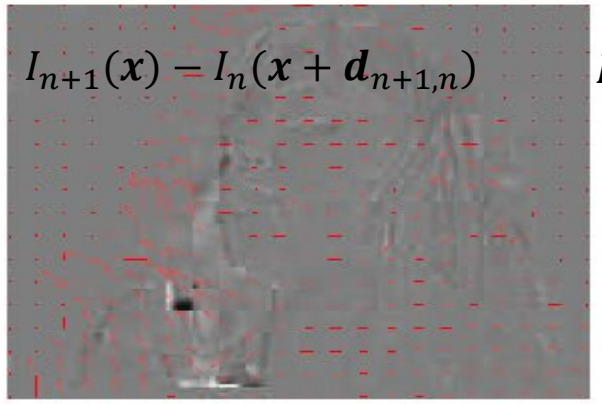
I_n $I_{n+1} - I_n$ $I_{n+2} - I_{n+1}$ $I_{n+3} - I_{n+2}$

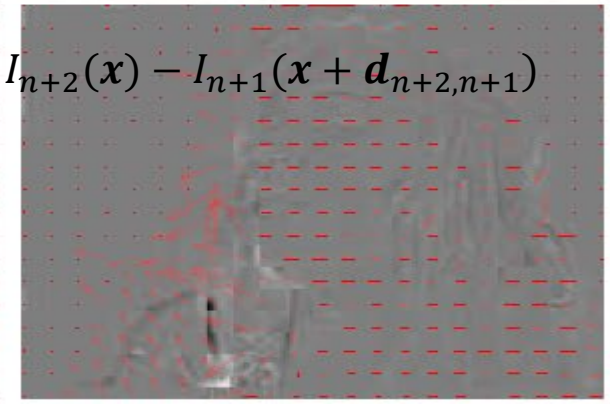
Motion Compensated Difference Sequence

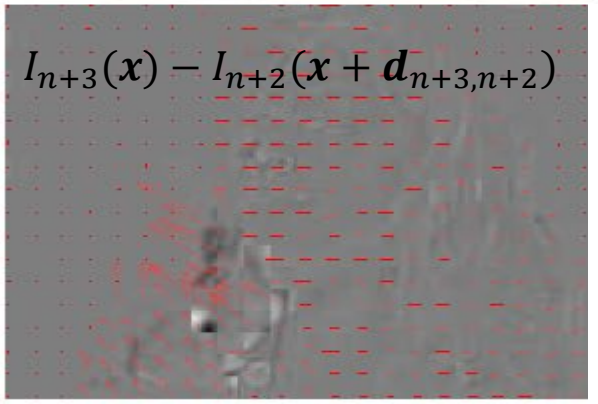


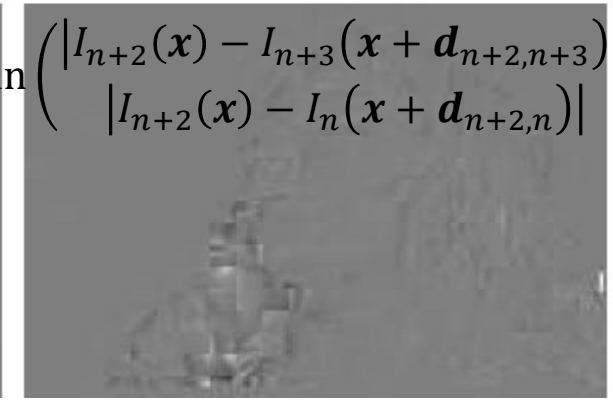
I_n $I_{n+1}(\boldsymbol{x}) - I_n(\boldsymbol{x} + \boldsymbol{d}_{n+1,n})$ $I_{n+2}(\boldsymbol{x}) - I_{n+1}(\boldsymbol{x} + \boldsymbol{d}_{n+2,n+1})$ $I_{n+3}(\boldsymbol{x}) - I_{n+2}(\boldsymbol{x} + \boldsymbol{d}_{n+3,n+2})$

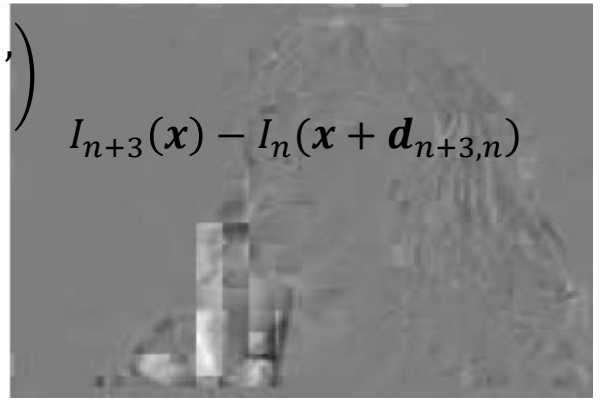
MC Difference Sequence : Smarter Prediction



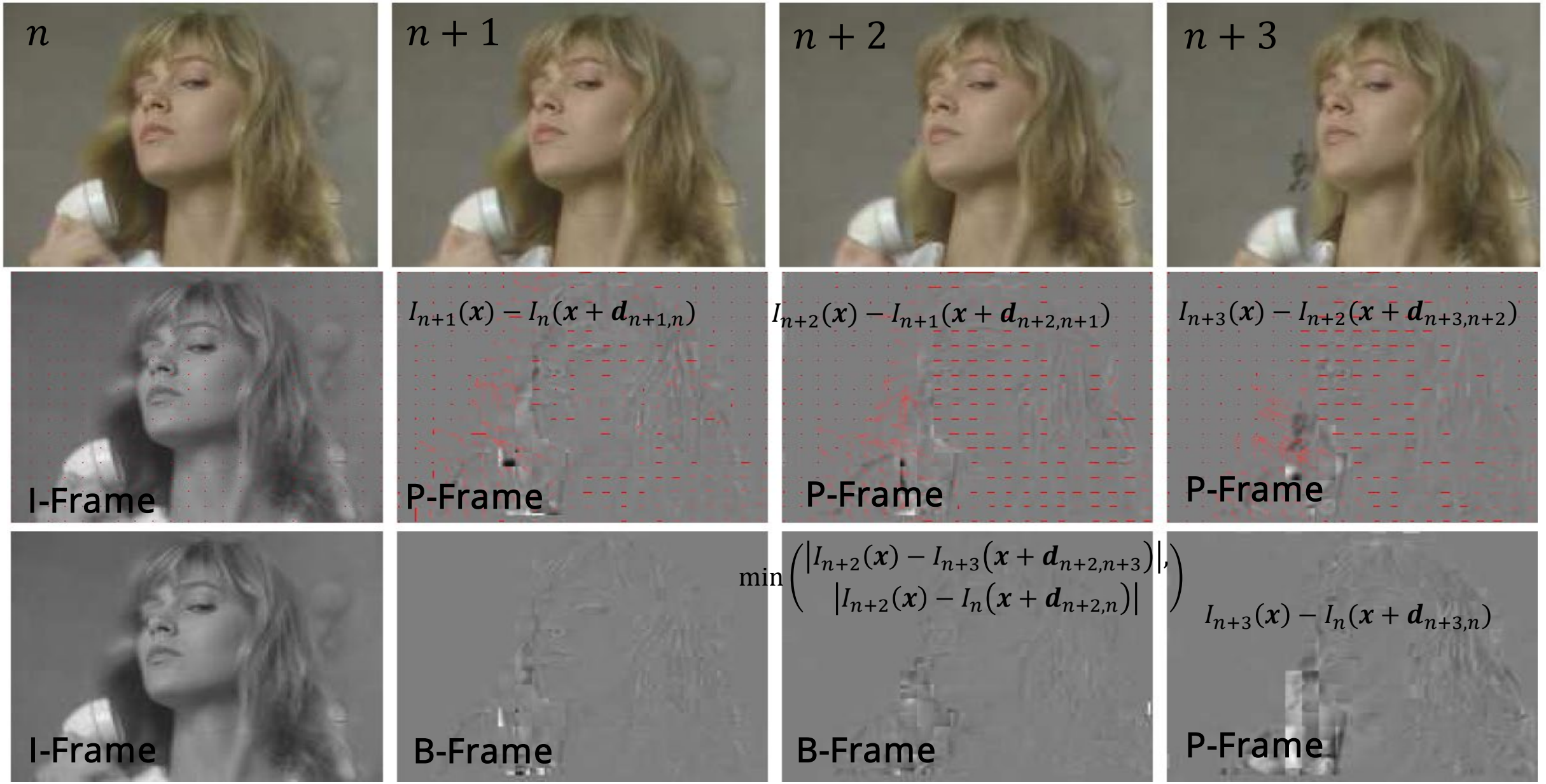




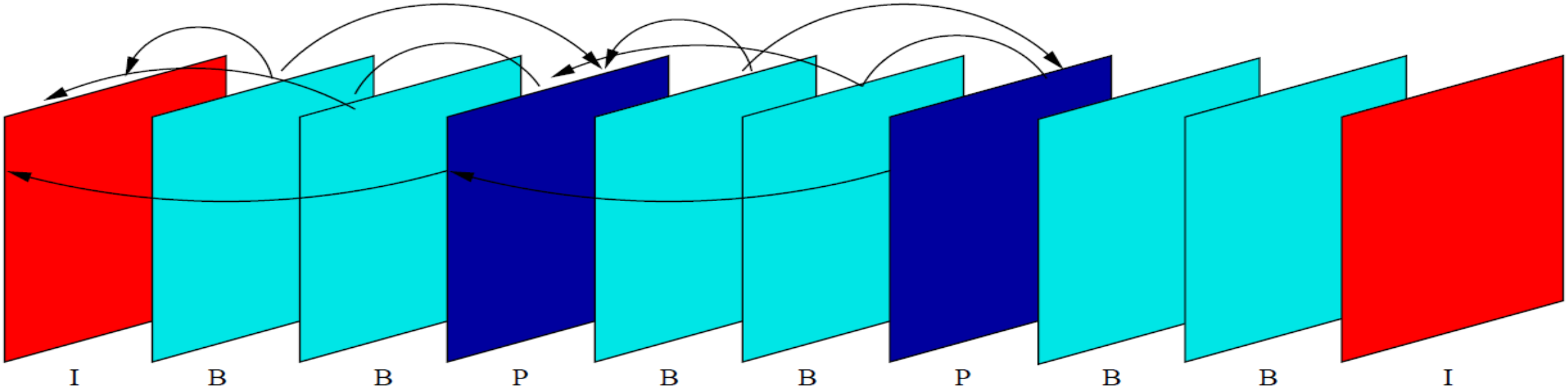




MC Difference Sequence : Smarter Prediction



Group Of Pictures (GOP) : Preventing error propagation, enabling fast(er) seek/edit



Comparing IPPP, IBBP etc

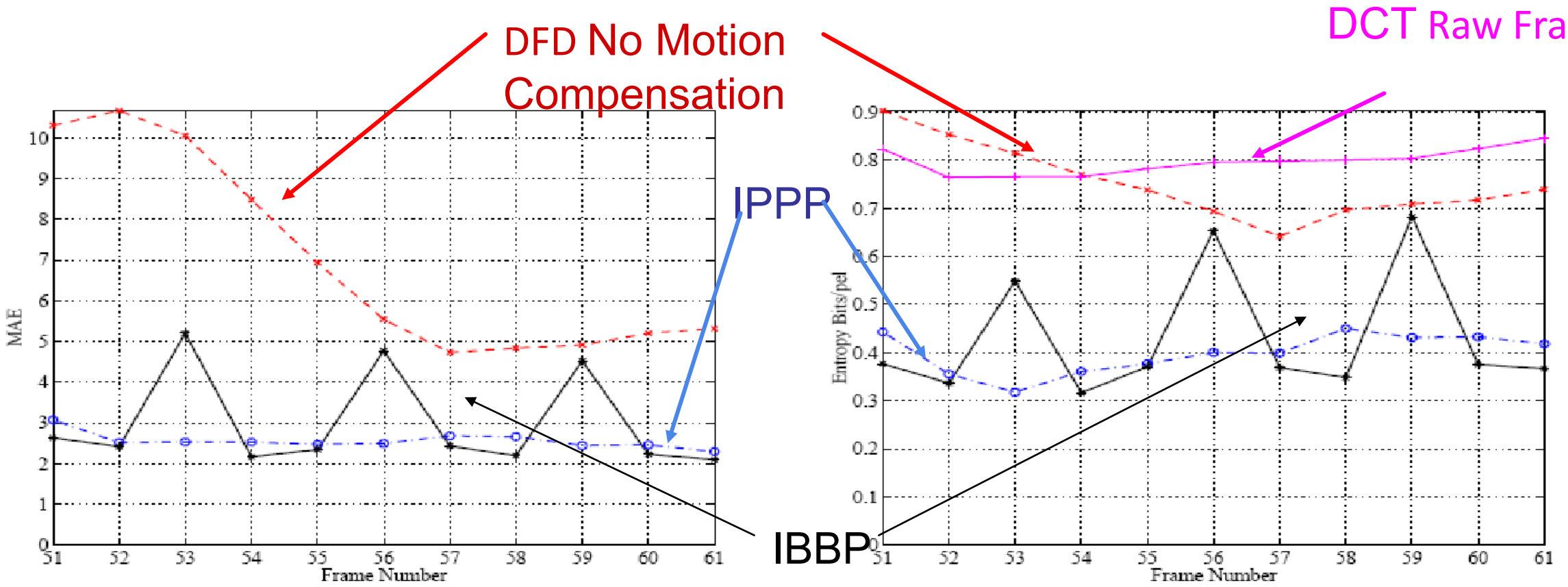


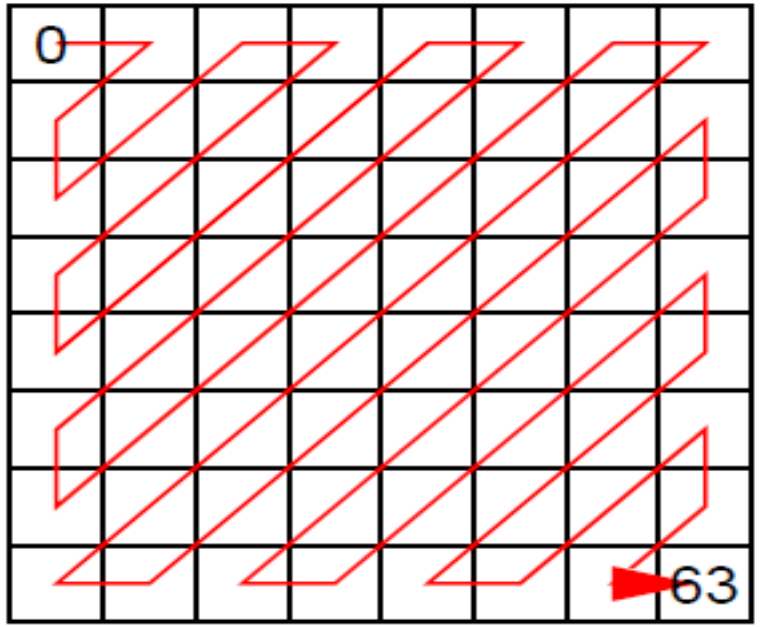
Figure 6: Comparison of MAE (left) and Entropy (right) for frames 51-61 of Suzie.

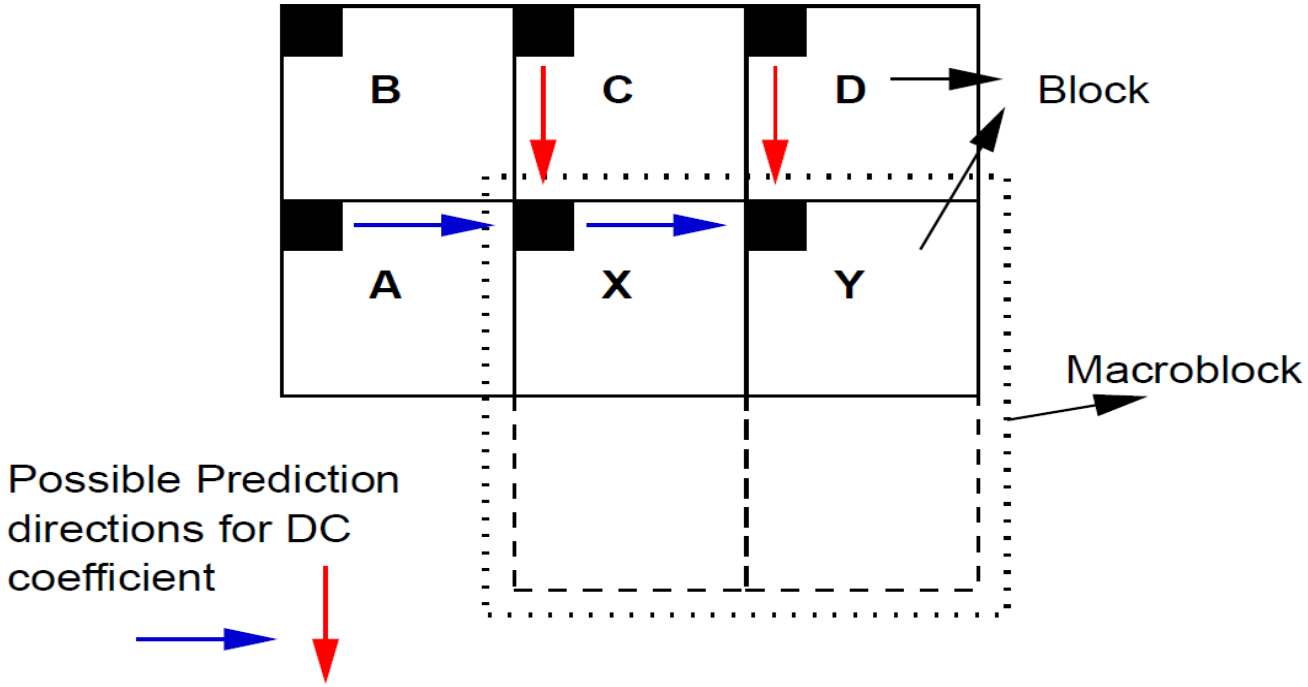
Motion Compensated Prediction

Always on a block basis (*MACROBLOCK/SUBBLOCK)*

	MPEG2	**MPEG4**	**H.264**
16x16	YES	YES	YES
16x8	Interlace Only	Interlace Only	YES and 8x16
8x8	NO	YES	YES
4x4	NO	NO	YES, 8x4, 4x8
Accuracy	1/2	½ and 1/4	¼ only
Reference Pictures	Intra, Pred	Intra, Pred	Intra, Pred, Multiple Others
Motion Estimation	Block only	Block and Global and Unrestricted Size with MVPred	Block only, Unrestricted, And MV Pred

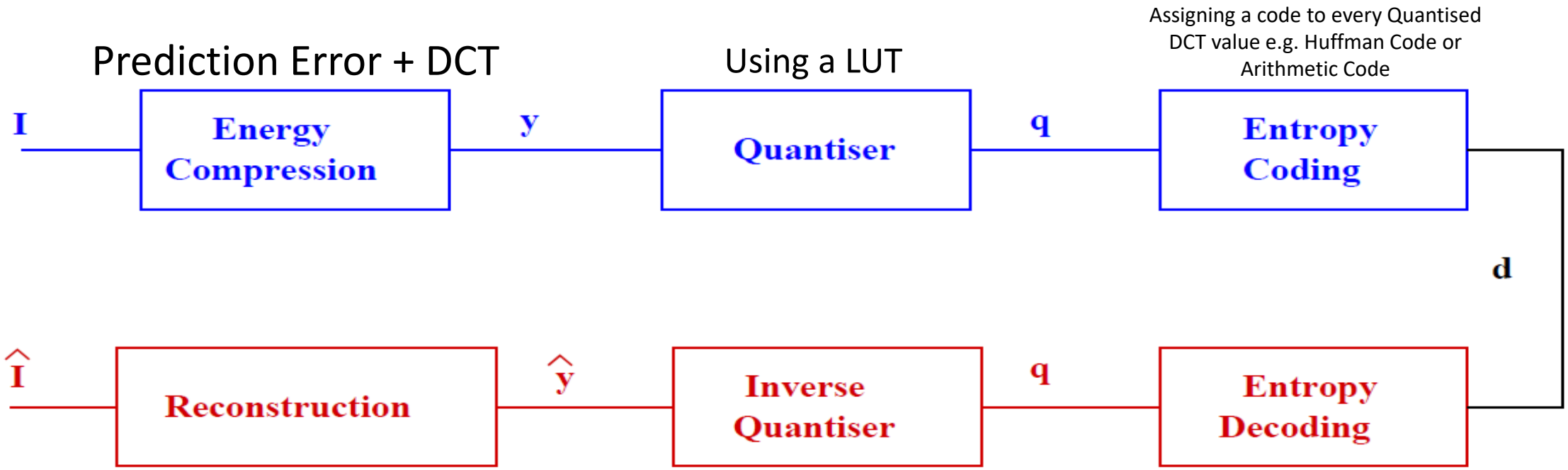
The DCT coefficients in a block are scanned to create a linear bitstream sequence





- Macroblocks contain smaller blocks. In H.264 the largest block size is a macoblock of 16×16 pixels
- DC Coefficient encoded separately using differential encoding
- Rest of coeffs encoded using run,length huffman codes (JPEG, MPEG2,4) But H.264 uses CABAC or CAVLC

Basic DSP Elements of Media Compression



Rate-Distortion Curves

Summarise the performance of an encoder over a range of parameter values

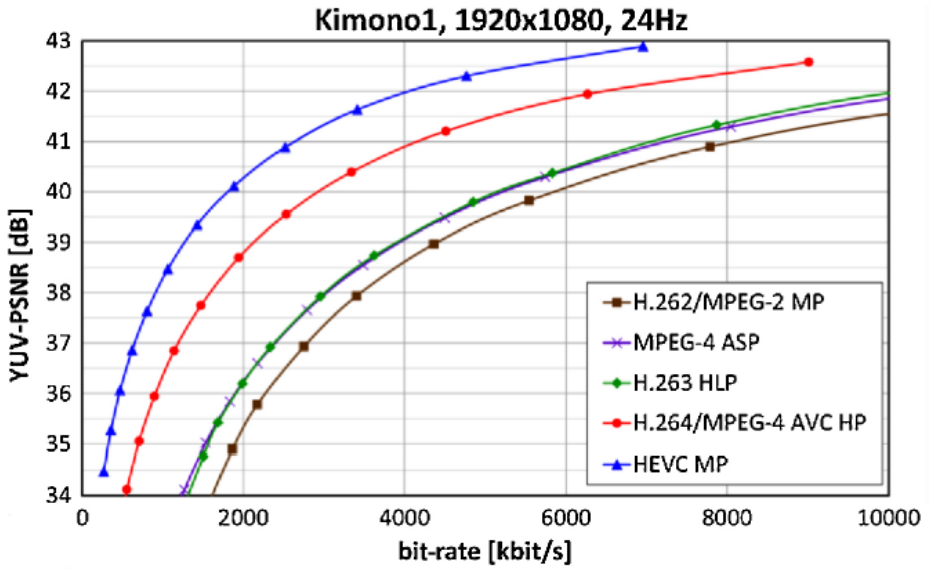
VERY IMPORTANT PERFORMANCE INDICATOR

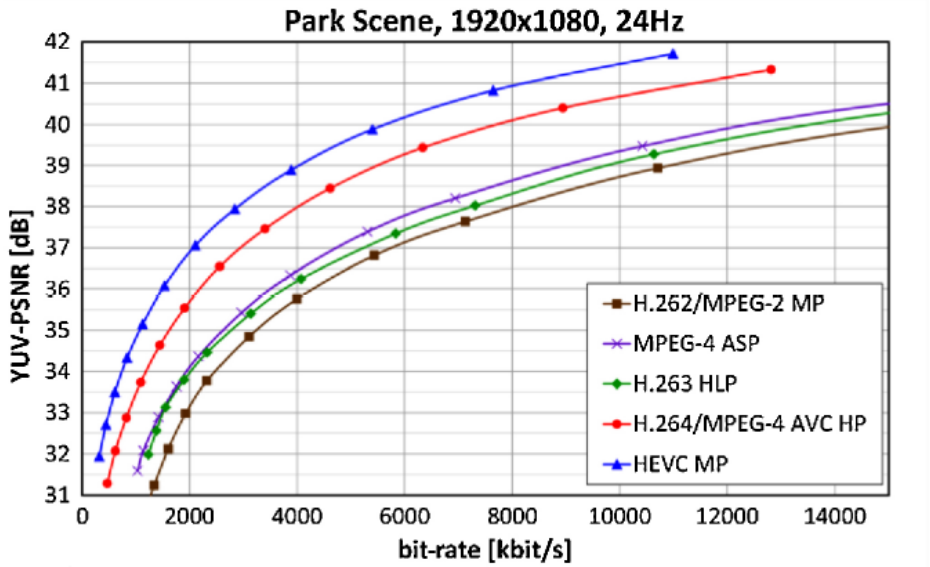
For a particular set of parameter settings, plots Bitrate and Image Quality Achieved

As bitrate increases, so does quality.

These curves allow fair comparison of codecs AND can be used to select appropriate settings for transcoding video content.

YOUR FIRST LAB IS ABOUT THIS





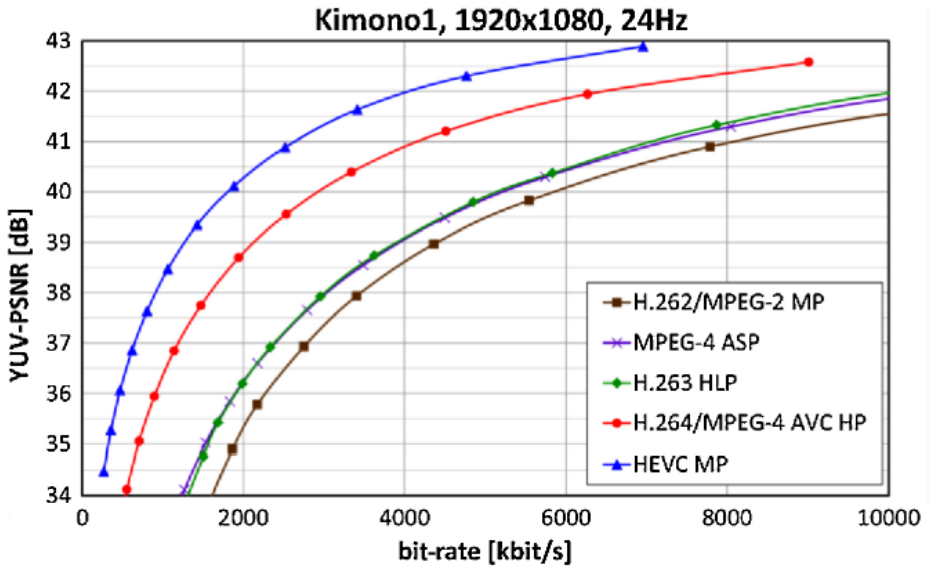
Rate-Distortion Curves

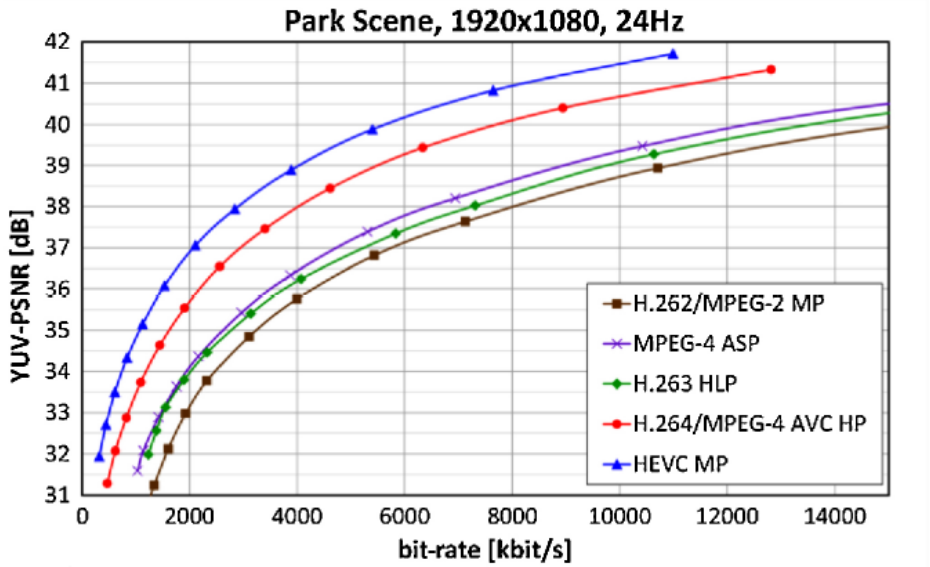
Each point on these curves represents the encoding of the video data using a particular encoder parameter setting.

Typical to use QP (Quantiser) as the parameter, but in some encoders CRF (Constant rate factor) is the main parameter

The measurement is Rate (in bits/sec) of the encoded file and PSNR (or some quality measurement) of that file Relative to the input

YOUR FIRST LAB IS ABOUT THIS





Standards

- Visual compression standards define the syntax of a bitstream i.e. how to decode a compressed bitstream.
- It does not define how to create a bitstream
- For example the H.264 standard

  Defines the codes used for different DCT values
  The code that defines the start of an I, P, B frame
  The order in which the bits are put together
  The code that marks the start of motion information
  and so on
- But it does not define HOW to do motion estimation, or HOW to generate a DCT coefficient
- Cleary it helps to have a bit of software which shows people how to all of these things and that is called a REFERENCE DECODER there is usually also a REFERENCE ENCODER. The standards bodies coordinate the creation and testing of these REFERENCE MODELS for the standard. These are very important for engineers to have faith in the standards and can build “Standards compliant” hardware or software modules.

Standards specify the BITSTREAM SYNTAX, they do NOT specify coding methods

This allows competition in the technology for encoding and decoding while maintaining an open digital video market.

For example, you might use Intel's encoder to create an H.264 video file but Sony's decoder can still decode it.

In some sense then, a video standard defines how the decoder can interpret the bitstream but not how that bitstream is made.

This is a crucial point.

History

JPEG **1992**, MPEG-1 1992 [Moving Picture Experts Group]

MPEG-2 **1996-1998** 

MPEG-4 (Part II) **1999-2002** [Initially about very Low Bit Rate Coding i.e. video phones, hence Video Objects …. but then not much savings over MPEG2]

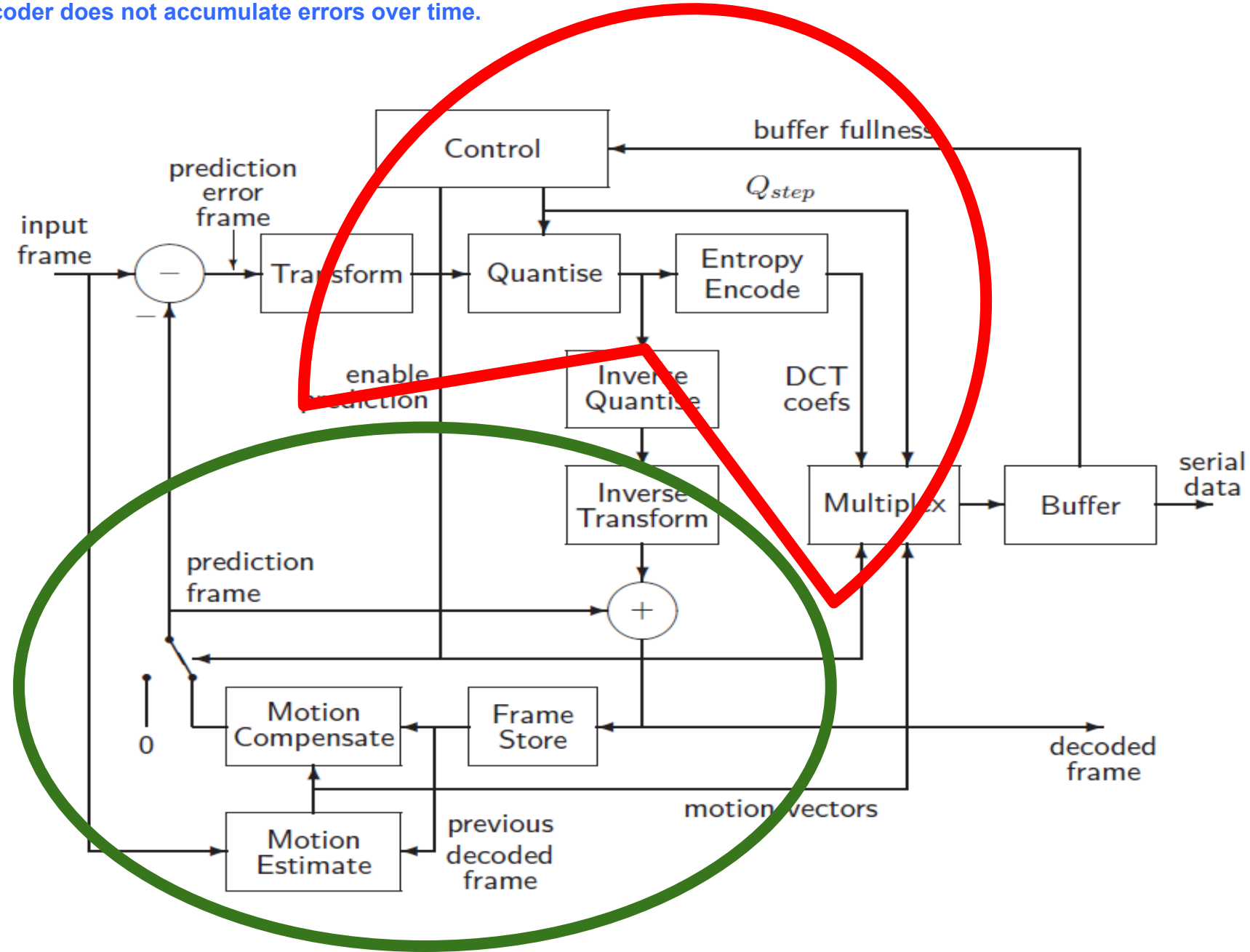
H.264 (MPEG4 Part 10) **2003** [AVC in MPEG4 standard, 1.5 to x2 better than MPEG2)

VP9, HEVC **2013** x2 better than H.264 ?

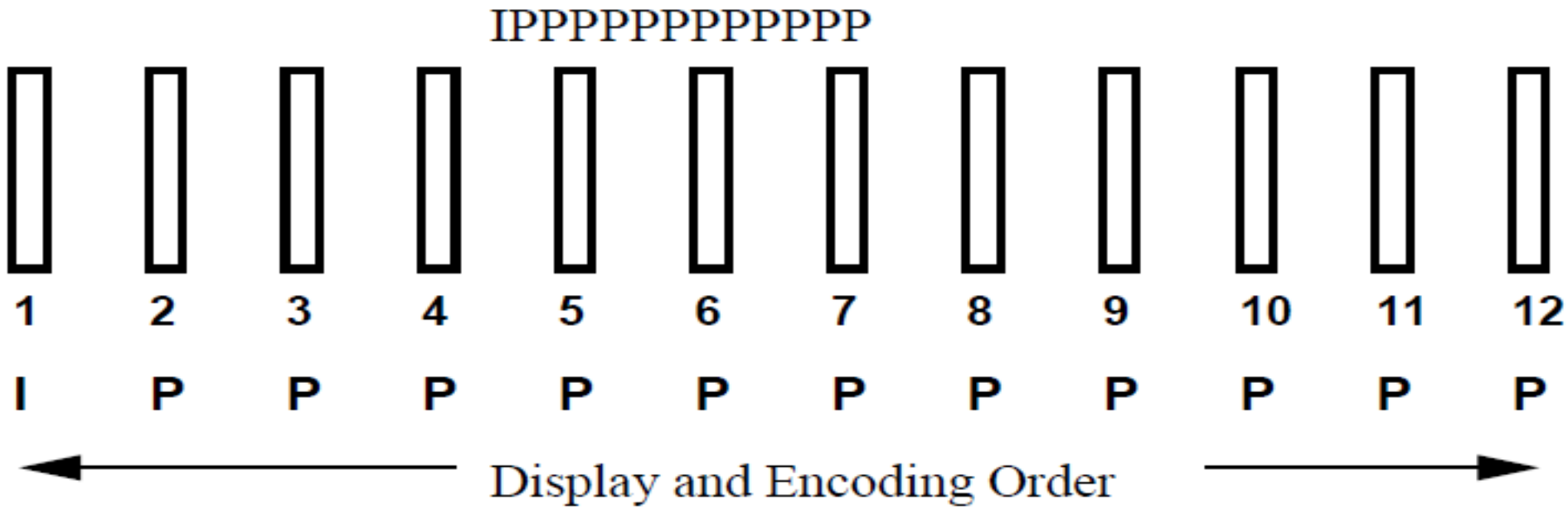
MPEG-LA Has nothing to do with Los Angeles or MPEG itself. It is the consortium that handles the centralised licensing for MPEG suite of technology. It is based in Denver.

All Video Encoders are Hybrid Encoders

They contain a decoder inside the encoder so that prediction is performed using the decoded (imperfect) frame instead of the (input) original. This ensures that the decoder does not accumulate errors over time.

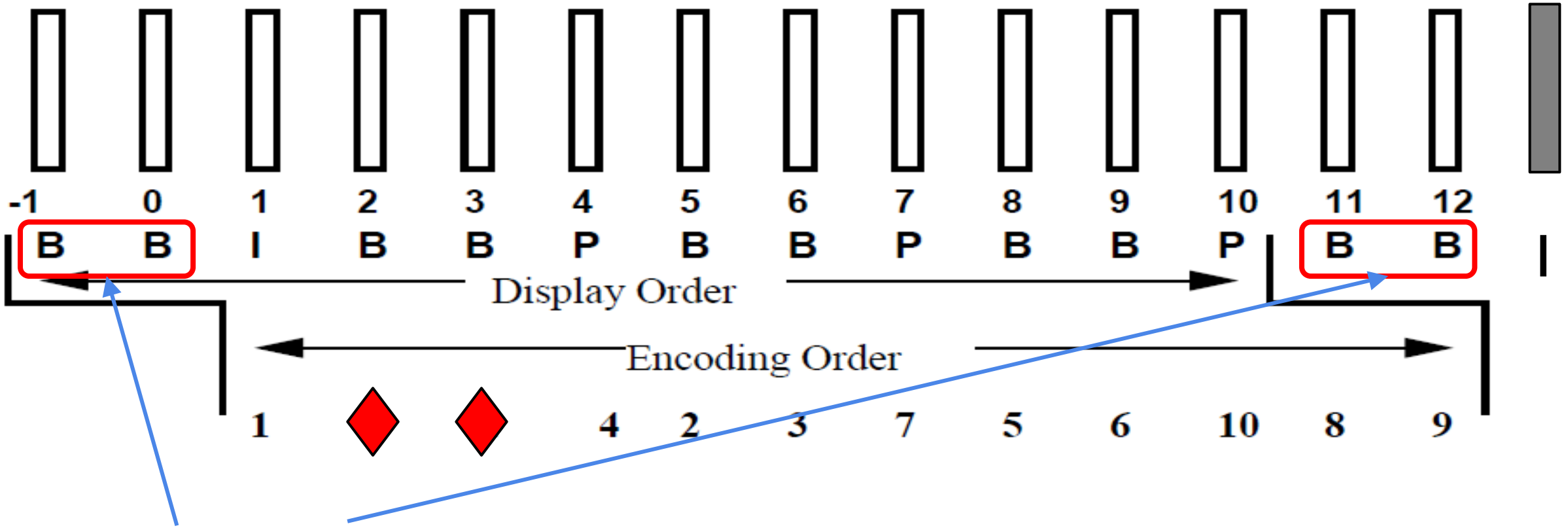


Subtle Tricks 1: Encoding/Decoding Order and Closed/Open GOPs



PTS Stands for “Presentation Time Stamp” it is the Time Stamp at which the frame is to be displayed
DTS Stands for “Decoding Time Stamp” it is the time at which the frame is decoded. You would expect frames to arrive in decode order which is not the same as the presentation order.

Subtle Tricks 1: Closed/Open GOPs



In a CLOSED GOP these B frames are NOT allowed to use the I frame in the next GOP for prediction

Lossless versus Lossy

What is says on the tin

Raw Data Rates : 20 MBytes/sec SD, 120 MB/sec HD

Expect for images/video lossless compression achieves about 3:1 compaction

Lossy Compression 45:1 (DVD), 100:1 (BluRay), 320:1 (YT 500Kb/sec SD?)

Some Digital Video Distribution Jargon

Top Floor

Lossless Compression : TIF, Cineon

10's MBytes/sec

Mezzanine Floor

Mezzanine Lossy Formats : Apple ProRes, Avid's DNxHD, Red's RedCode, GoPro Cineform, BBC's DiracPRO

50 Mb/sec

Ingest@ NETFLIX YouTube

1-6 Mb/sec up to 1080p

Bottom Floor

Distribution = **Lossy Formats** H.264, MP4, MP2

Distribution@ NETFLIX YouTube

Issues to be addressed in the design of standards

Multiplexing

Media File/Transport Formatting

Sequencing

Error Resilience

Rate Control

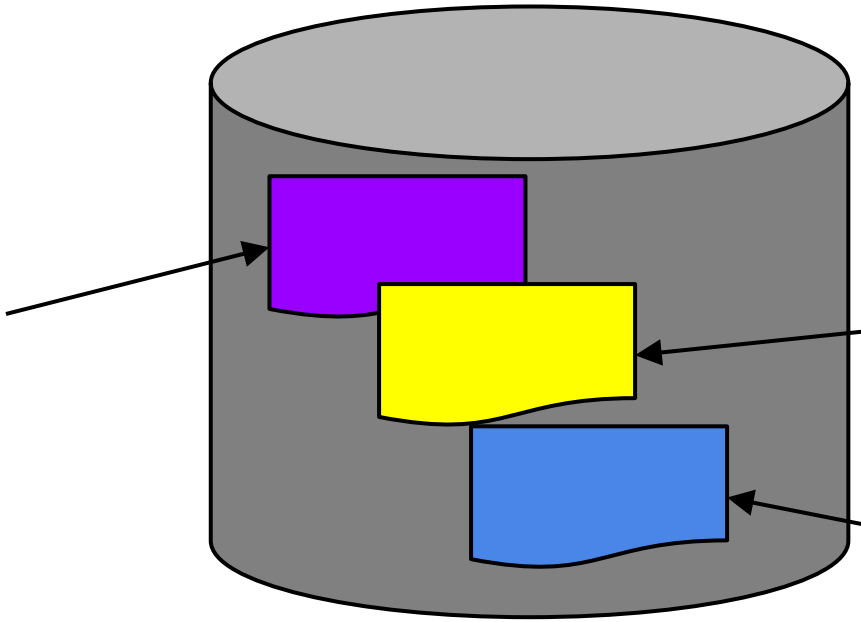
Scalability/Mutiplatform

Subtle Tricks 2 : Containers and Essenses

Containers solve the problem of multiplexing and digital media storage/transmission.

They CONTAIN the audio, video and metadata BITSTREAMS

Easy to confuse term MPEG4 with container OR Essence



.mp4 File defined by the MPEG4-Part 14 document

h.264 Elementary Video Bitstream (or Video Essence) defined by the MPEG4-Pt 10 AVC document

h.264 Elementary Audio Bitstream (AAC Encoding) definied by the MPEG4-Pt 2 AAC Document

MetaData defined by the Part 14 document

Subtle Tricks 2 : Containers and Essenses

Container	File Extension	Video Essense	Audio Essense
MPEG4-14	.mp4	h.264	aac, pcm
MPEG2-1 (Transport Stream)	.m2ts	mpeg2-video	aac, pcm
Apple QuickTime	.mov	prores, h264	pcm, aac
Flash	.flv	h.263	.mp3
Matroska	.mkv	h.264, mpeg2, etc	aac, ac3 etc

Profiles and Levels

- MPEG supports a wide variety of scenarios
  - eg. high quality tv broadcast, low bit rate internet streaming etc
  - decoders can have varying degrees of complexity + plus a decoder for internet streaming should not have to support decoding of digital tv signals.

- MPEG defines Profiles and Level for streams
  - Profiles define the required decoder complexity (feature set) to decode the stream
  - Levels define the maximum allowed resolution frame rate and bit rate.

Subtle Tricks 3: Skipped Macroblocks

If your prediction of the current macroblock is perfect why bother to code/send it? Just use the prediction that the decoder already has.

In all codecs MPEG2-H.264 you can mark a macroblock as “skipped” for this reason.

In MPEG2 this usuallly implies a Zero motion vector

In H.264 a skipped macroblock uses a predicted motion vector from the median of the vectors around as the motion compensated vector.

Notion of MODE decisions at a block level : Intra/Predicted/Skipped

Subtle Tricks 4: Macroblock types

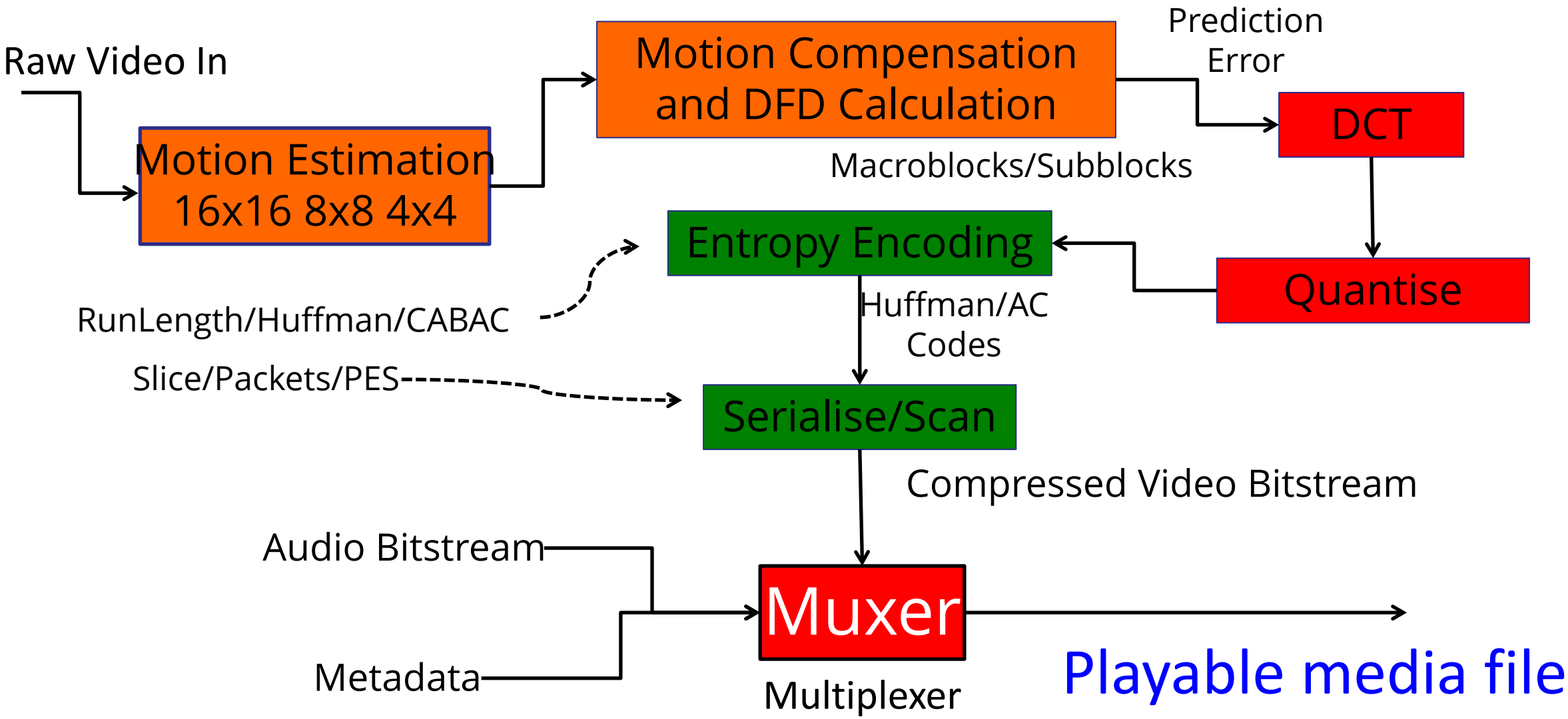
Sometimes you just can't get a good prediction from the previous frame. Possibly the motion is just that tricky or a completely new object has entered the scene.

H.264 therefore allows Macroblocks to switch between Intra and Predicted (or Bi-Predicted) in P and B frames.

The rate controller decides which type a Macroblock can be.

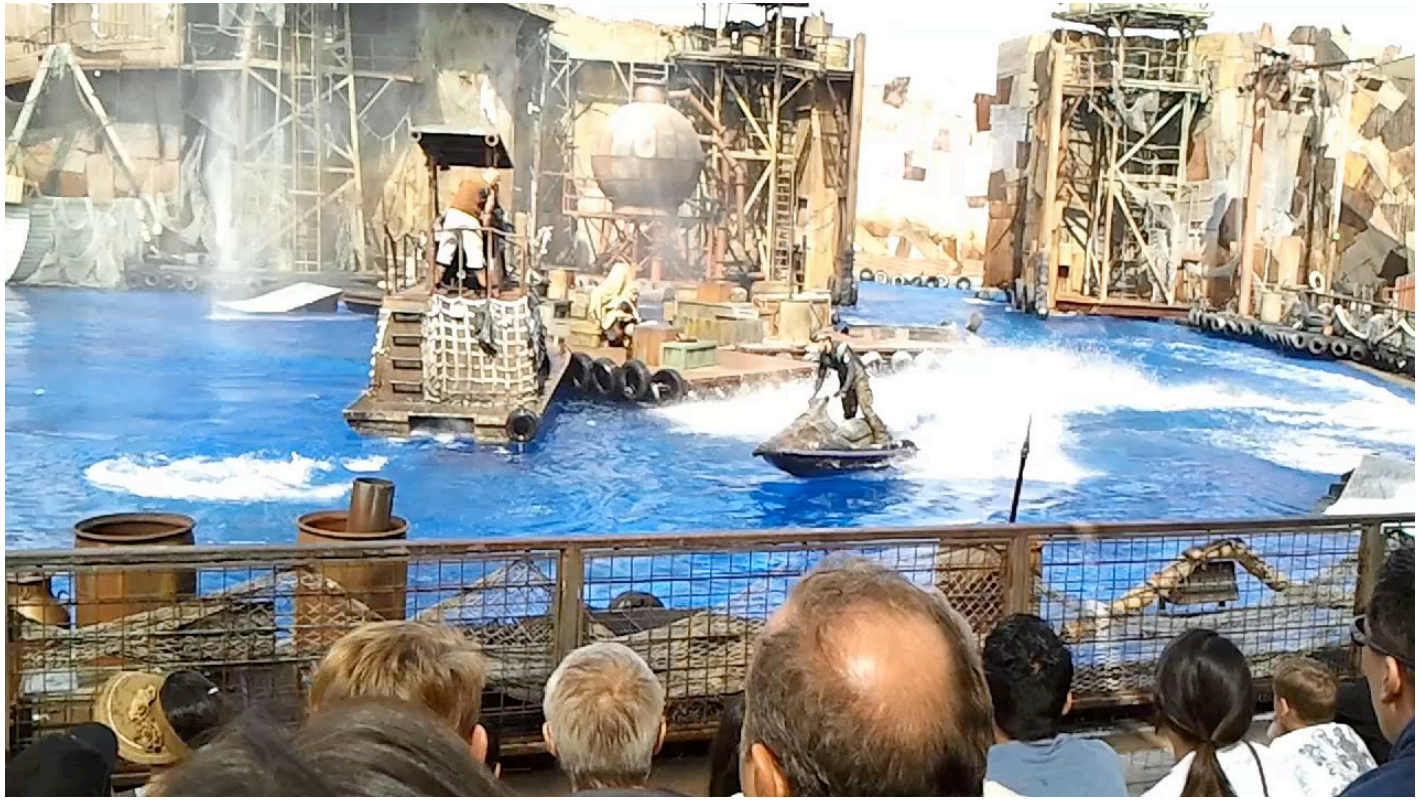
This means that at a scene cut you could potentially have a P-frame that consists of many I-Macroblocks!

A video encoder summary….



Examine .MP4 container with H.264 essense

Video recorded and packaged by a Samsung Galaxy SII
(Using www.codecvisa.com for easy to use diagnostic, could also use ffmpeg)



Starting from 1st frame : I, P, P, B, P, B, B, P, P, P, P, P, P, B, B, P, B, B, P, B, B ....

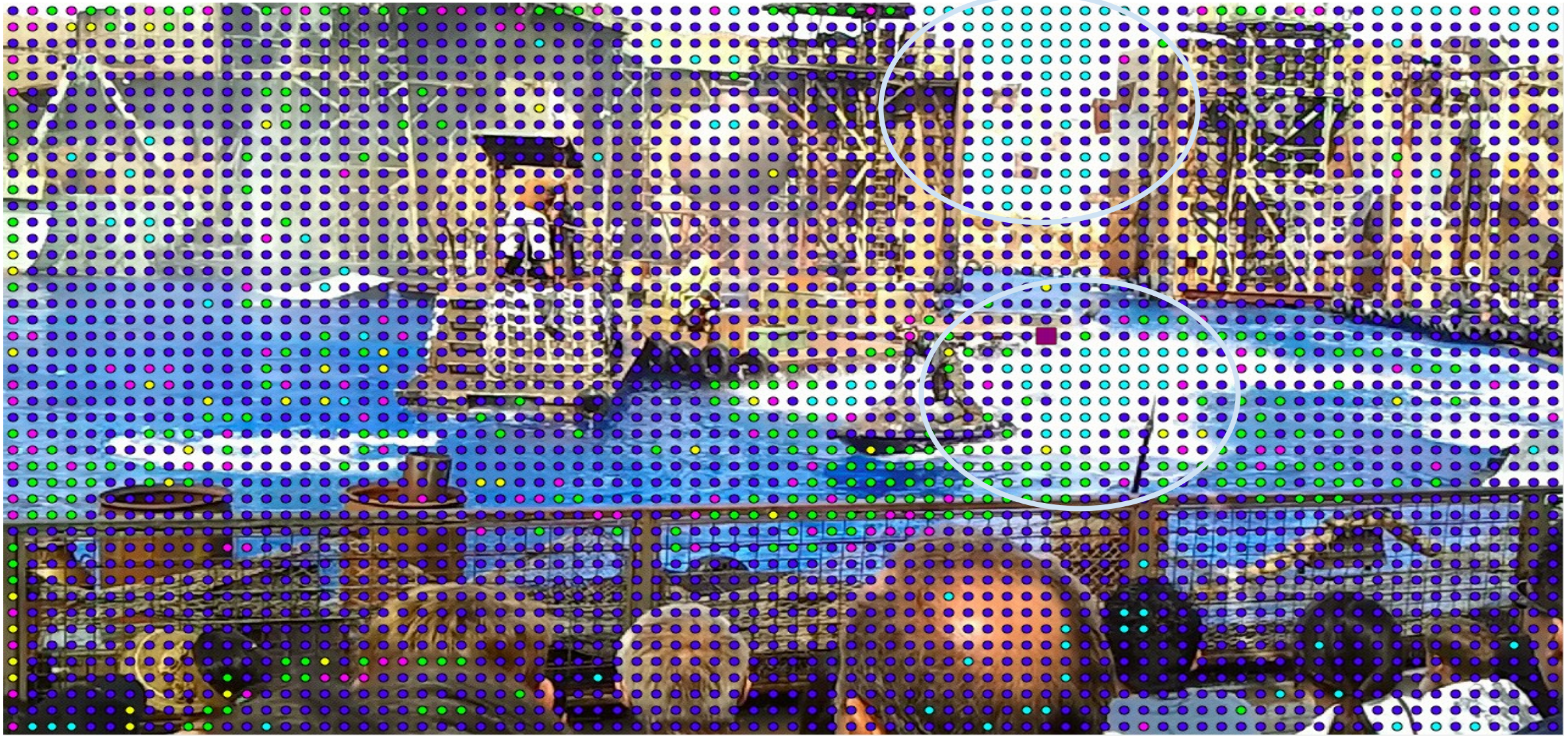
I, P, P, B, P, B, B, P, P, P, P, P, P, B

Frame 0, Intra-Frame (PTS:0, DTS:0) : 1.3 Mbits

Intra_MB_16x16 Intra_MB_8x8 Intra_MB_4x4

I, P, P, B, P, B, B, P, P, P, P, P, P, B

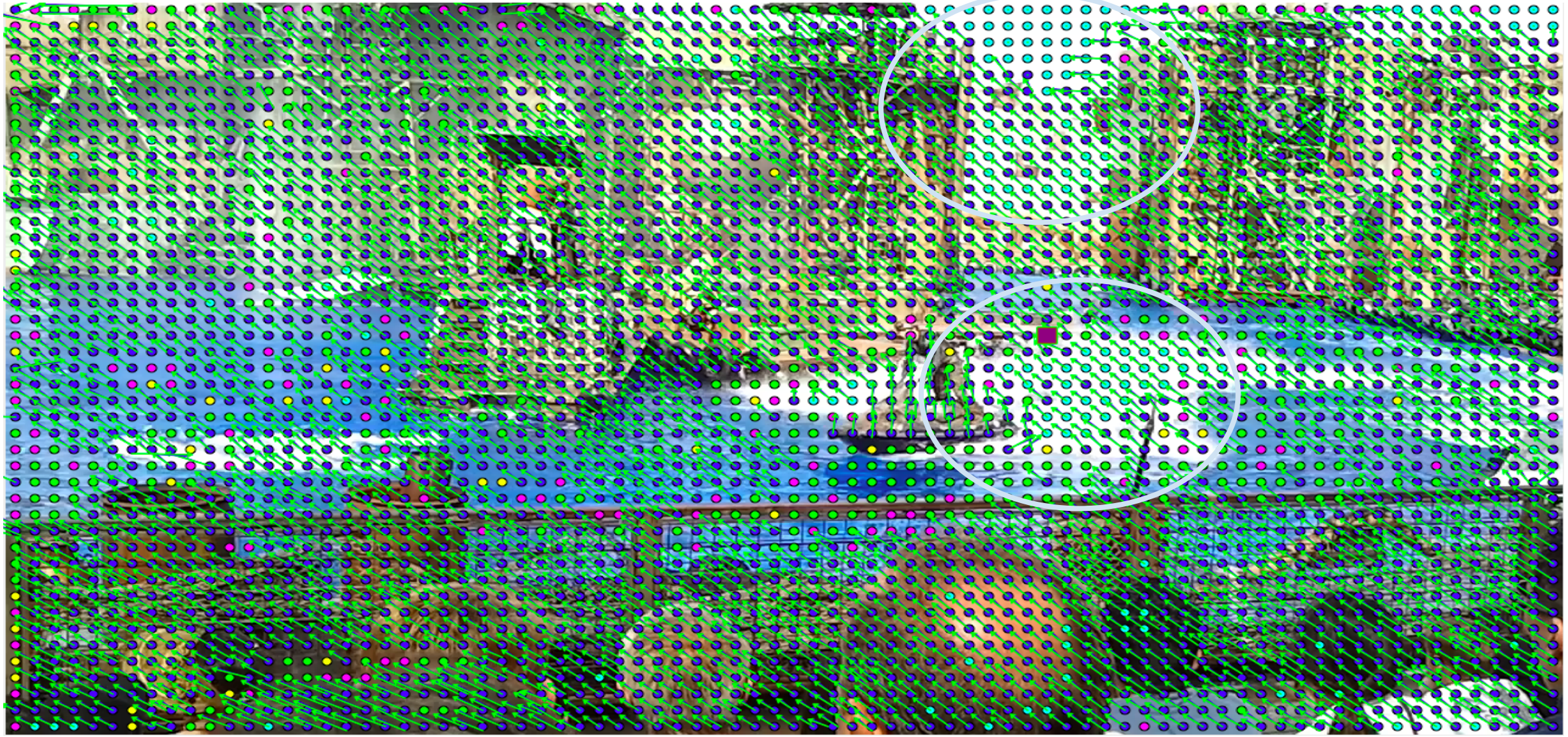
Frame 1, P-Frame (PTS:1, DTS:1) : 0.6 MBits



Intra_MB_16x16 Intra_MB_8x8 Intra_MB_4x4 P_MB_16x16 (8) P_MB_Skip

I, P, P, B, P, B, B, P, P, P, P, P, P, B

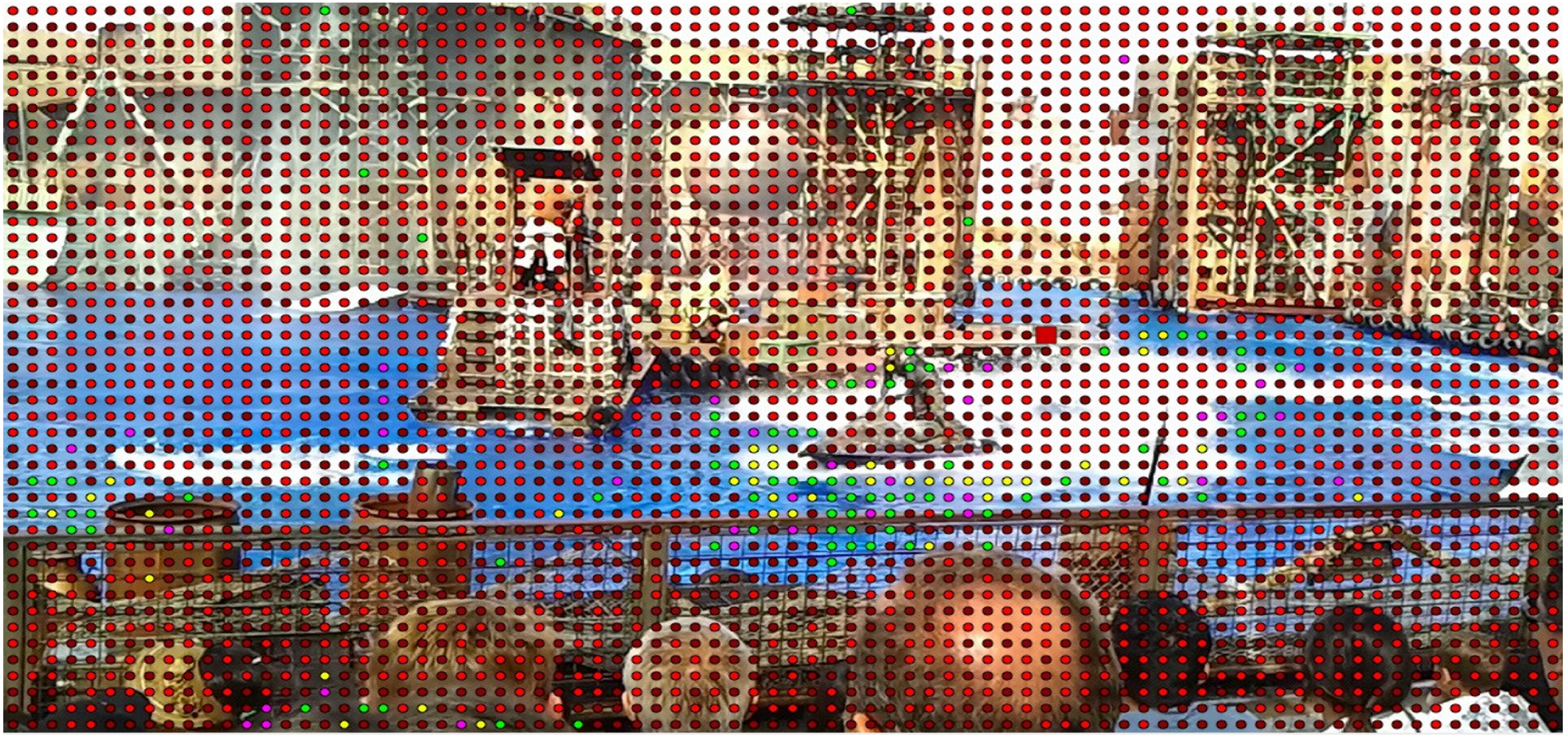
Frame 1 Motion, P-Frame (PTS:1, DTS:1)



Intra_MB_16x16 Intra_MB_8x8 Intra_MB_4x4 P_MB_16x16 (8) P_MB_Skip

I, P, P, B, P, B, B, P, P, P, P, P, P, B

Frame 3, B-Frame (PTS:3, DTS:4) : 0.2 Mbits



Intra_MB_16x16 Intra_MB_8x8 Intra_MB_4x4 B_MB_16x16 (8) B_MB_Skip

Final Comments

Standardisation brings complexity?

Computational elements explain only part of the nature of the standards

Using codecs requires an awareness of all the layers

Next : VP9/HEVC