Reference Material

Background on Digital Media
http://xiph.org/video/vid1.shtml http://xiph.org/video/vid2.shtml

Excellent Overview of Media Compression at http://people.xiph.org/~tterribe/pubs/lca2012/auckland/intro_to_video1.pdf ; https://www.xiph.org/daala/

HEVC Information http://hevc.hhi.fraunhofer.de/
http://www.atlanta-smpte.org/HEVC-Tutorial.pdf

H.264 Information http://www.itu.int/rec/T-REC-H.264

Tools : www.ffmpeg.org http://www.videolan.org/

VP9 Presentation at Google IO 2013 : http://www.youtube.com/watch?v=K6JshvblIcM

Wikipedia actually quite good on this stuff

Rate Control in H.264 : http://www.pixeltools.com/rate_control_paper.html

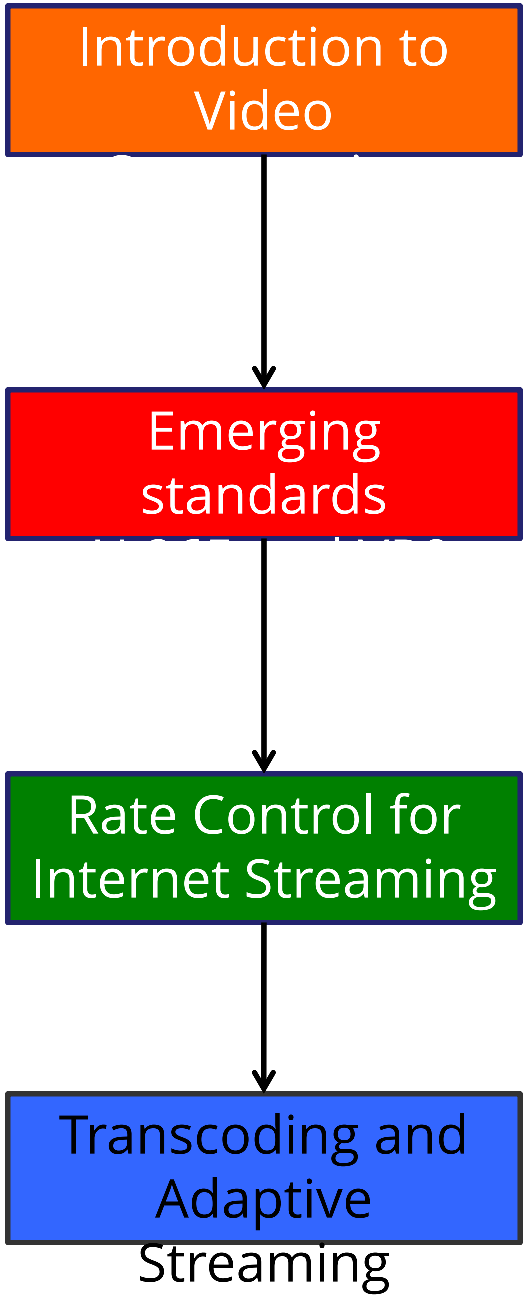
Rate Control

Motion Picture Engineering
Anil Kokaram

Content

Lab report on rate distortion
and using ffmpeg

Assignment on Adaptive Streaming



What is rate control?

- Rate control in this context refers to the process of creating an encoded media file which obeys constraints on any or all of the following

  Bitrate

  Visual Quality

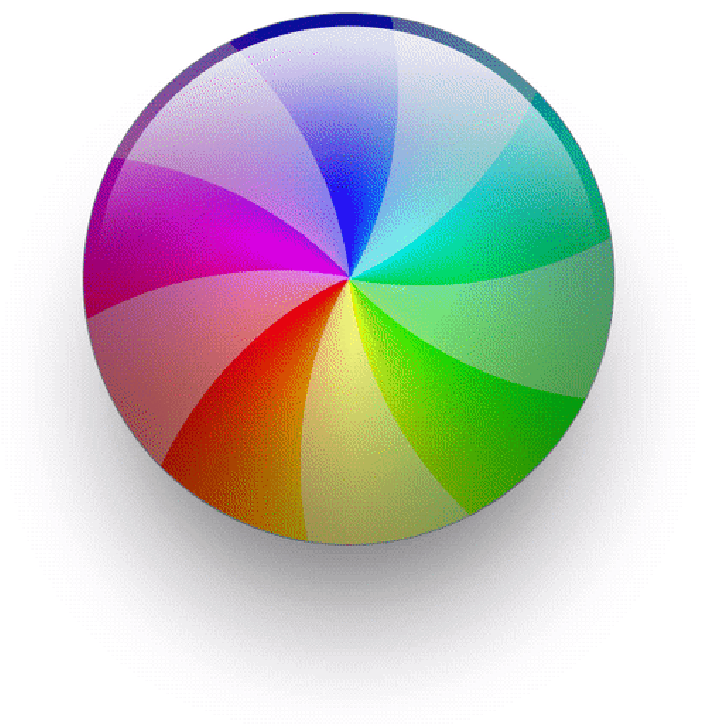
  Decoder capability

  (e.g. decoder buffer size, speed of processing)

  Energy

- Rate Control is not defined as part of the video standard. It is left up to the ingenuity of the motion picture engineer to adjust encoder parameters so that the output compressed bitstream hits the desired constraints.

- It is the reason why Zoom worked better than Teams in the early days of the pandemic, and also the reason why Teams and Zoom are now almost the same. Basically : rate control algorithms can be changed on the fly as we learn more about how to target the right constraints.



No one wants the spinning wheel

Also : we want to watch decent quality pictures

Visual Effect of different Quality/Rate conbinations

Where do the constraints come from?

- Bandwidth (bits per second) over a link is often unknown and certainly time varying
- Packet switched networks operate with unknown and time varying delays and packet losses
- Device capability varies enormously

Movie Download

Download speed proportional to size of file. Large files imply less movies on your device, and long download times.

Files encoded at a high bits/pixel mean that a lot of data has to be read from storage before decoding. Your storage access speed might not be fast enough to keep up with real time playback.

Long download times and lots of storage access = low battery life.

Streaming Movies

Files encoded at a high bits/pixel (high compression rate) mean that a lot of data has to be streamed before you can decode a picture. Hence
-- Your decoder buffers might "overflow" before you can decode a picture. So you end up skipping frames.
-- Or your decoder/cpu might not be fast enough to decode the data in time. So you end up skipping frames.
-- Or your bandwidth might be too low to keep up with your decoder (buffer underflow) and you get the YouTube "spinny wheel".

Files encoded at low bits/pel (low compression rate) can imply bad looking pictures.

Why rate control : because no free lunch

HIGH RATE (HIGH BITS/PEL) == GOOD QUALITY MEDIA (low distortion)
BIG FILES

LOW RATE (LOW BITS/PEL) == LOW QUALITY MEDIA (high distortion)
SMALL FILES

What is the best rate we can achieve for a given distortion?

What is the lowest distortion we can achieve for a given rate?

What is the best rate/distortion combination for a given decoder complexity?

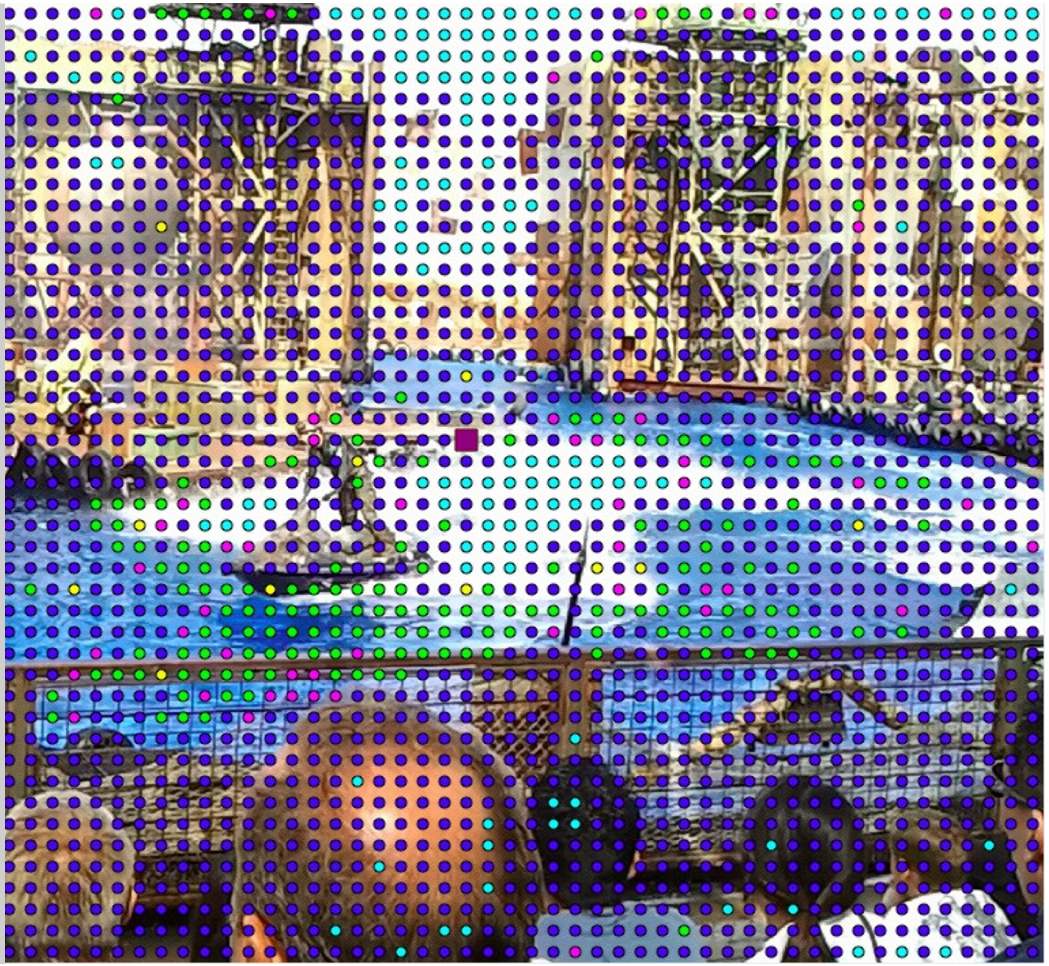
What is the best rate/distortion combination given a maximum rate ceiling?

What are our control knobs?

MODE DECISIONS
Block or Frame
Skip/Intra/P/B/Directional etc.
Block Size 16x16, 4x4, 8x8

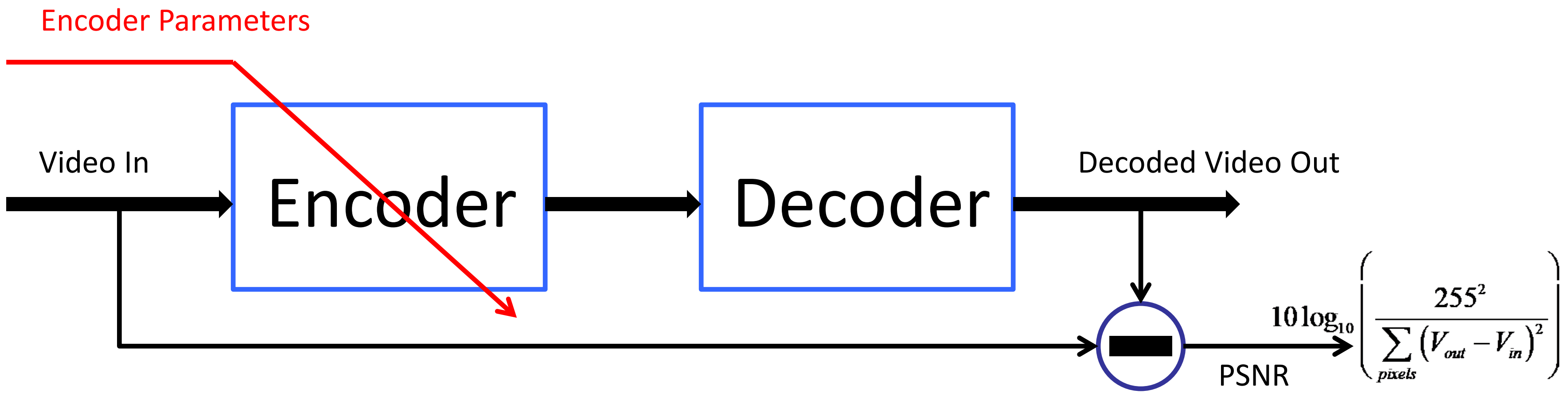
ALGORITHM
Entropy Coding :
CABAC/CAVLC
Motion Estimation : Search width/Accuracy

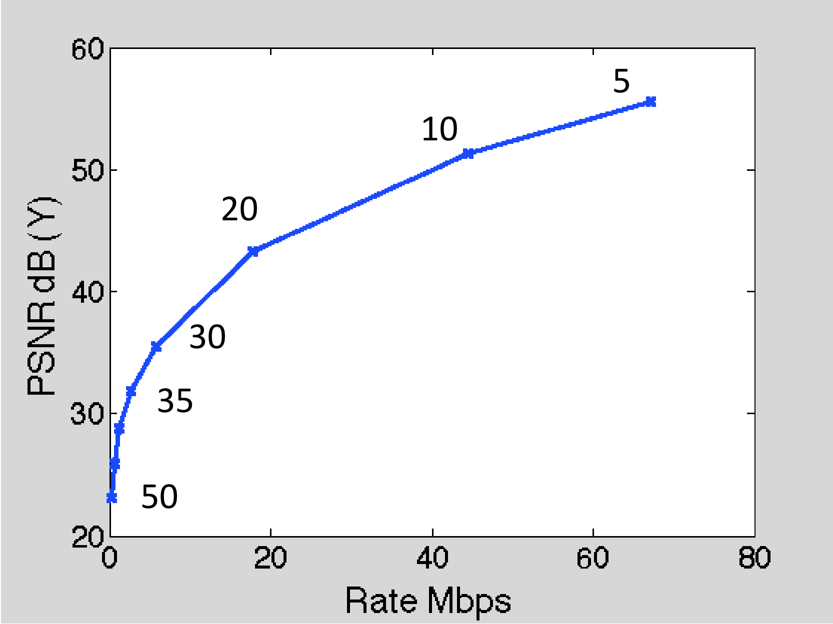
QUANTISATION OF DCT COEFFICIENTS
Quantisation Step Size
Scan direction



Intra_MB_16x16 Intra_MB_8x8 Intra_MB_4x4 P_MB_16x16 (8) P_MB_Skip

R/D Curve





RD curve expresses the tradeoff between rate and distortion (Quality) for a set of encoder parameters. Here we are changing just one parameter : the quantisatio step size Q

Rate/Distortion Theory

Elements of Information Theory, T.M. Cover and J.A. Thomas, Wiley 1001, Chapter 13 Rate Distortion Theory

A Mathematical Theory of Communication, Shannon, Bell System Tech Journal, Vol 27, pp 379—423, 1948.

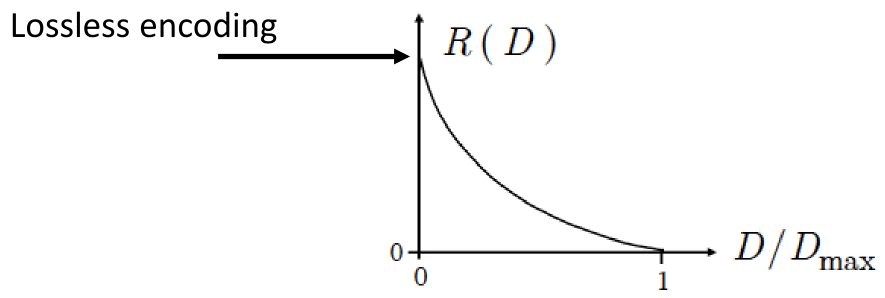
Shannon showed that is was possible to work out the MAXIMUM RATE required to achieve a TARGET DISTORTION without specifying a particular encoder mechanism.

Application of Information Theory

Shannon's 1948 paper remains important today and is VERY readable.

He discusses R/D on page 47 and uses N where we use D

Intuition



R(D) curve must never increase with increasing distortion

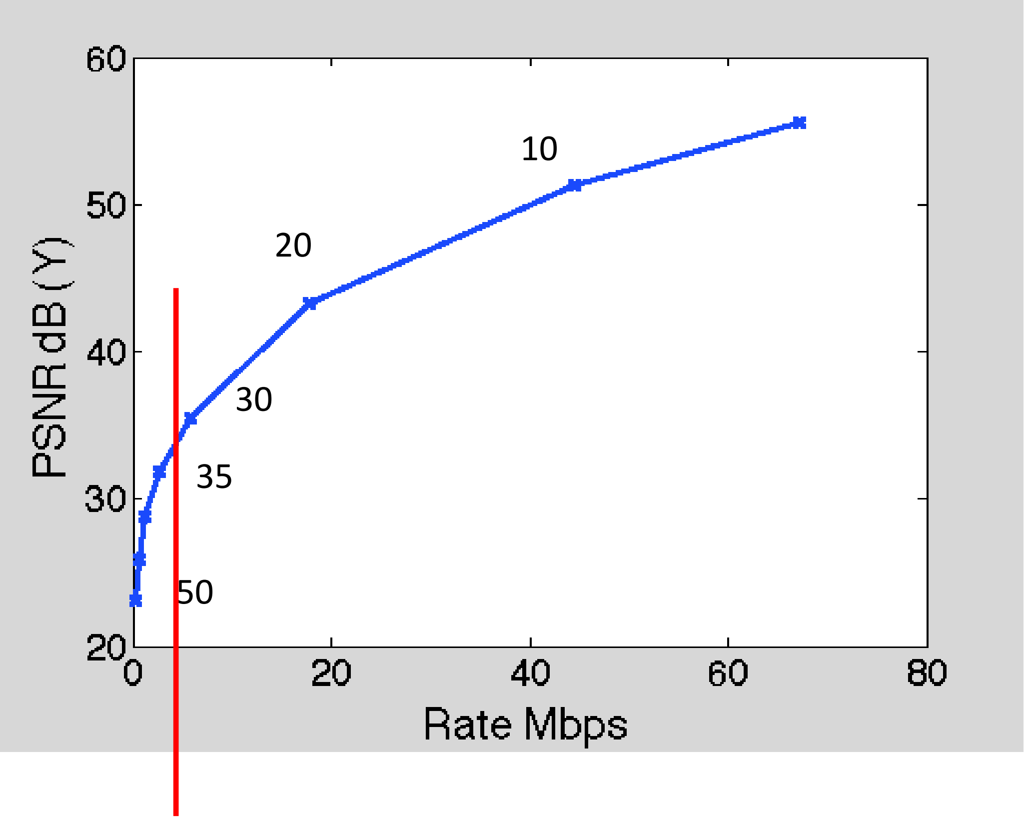
Consider quantisation of DCT coefficients : the more you quantise, the less coefficients you have to transmit so the lower your RATE and the more you distort the decoded signal.

Given a MAXIMUM required distortion, there is an LOWER BOUND on the RATE at which it is possible to achieve that distortion.

Consider quantisation of DCT coefficients to achieve a required DISTORTION : the more you quantise, the more you distort the signal, and the more your rate reduces. If you don't want to have distortion higher than a certain amount then you can't code it at a rate lower than a certain amount. You can always encode it at a high rate and achieve lower distortion.

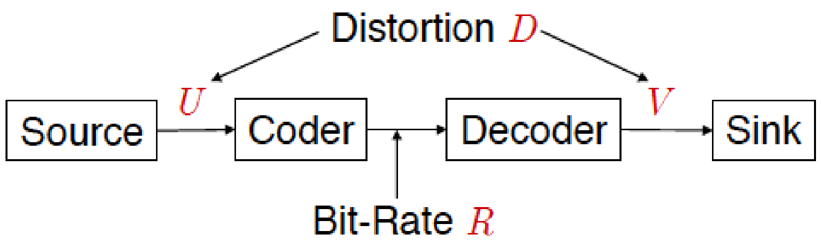
THIS WAS ONE OF SHANNON'S INSIGHTS ABOUT THE RELATIONSHIP BETWEEN RATE AND DISTORTION.

Finding the optimum tradeoff



For a max rate of 5Mbps a Q between 30 and 35 gives you the best PSNR of about 32dB

Some elements of the theory



Source Symbols $$U_k \in \{u_o, u_1, u_2, \ldots u_{M-1}\} \qquad P(u)$$

i.i.d.

Binary Source $U_k \in \{0,1\}$

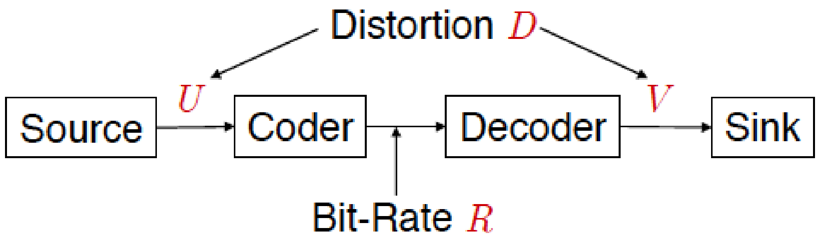
Image $U_k \in \{0,1,2,3,4,5,\ldots 255\}$

Reconstruction Symbols

$$V_k \in \{v_o, v_1, v_2, \ldots v_{M-1}\} \qquad P(v)$$

The sets U,V need not be the same.

Some elements of the theory

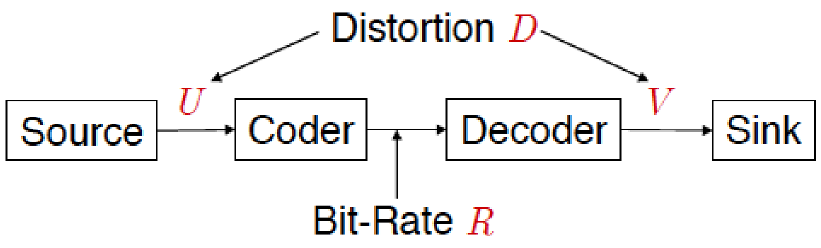


Describe statistical relationship between coder/decoder with function Q(v|u)

This is the conditional probability distribution of the symbols of the reconstruction alphabet v given an observed symbol in the source alphabet u.

Transmission system defined by the joint p $P(u, v)$

Some elements of the theory



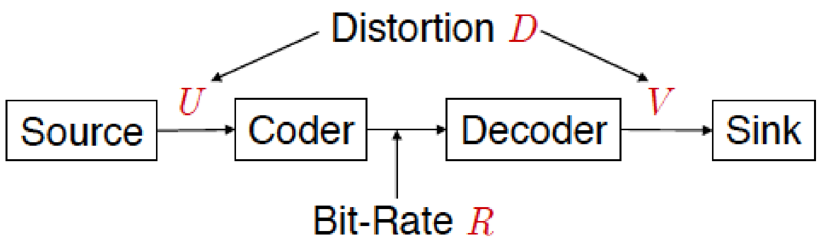
Describe distortion with a non-negative COST function

MSE $$d(u,v) = |u - v|^2$$

Hamming Distance
$$d(u,v) = 0 \text{ for } u == v$$
$$d(u,v) = 1 \text{ for } u \neq v$$

Average Distortion $$D(Q) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} \underbrace{P(u) \cdot Q(v \mid u)}_{P(u,v)} \cdot d(u,v)$$

Deriving the R(D) function



$$R(D^*) = \min_{Q:D(Q)\le D^*} \{I(Q)\}$$

The minimum rate required to achieve a maxium distortion D*

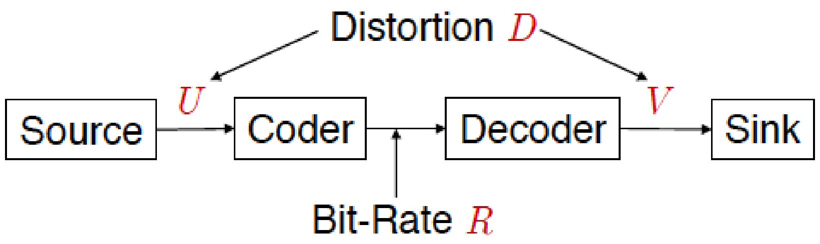
R(D*) is in bits

Minimise the Mutual Information of u,v with respect to the mapping Q between u,v in such a way as to ensure that the Average distortion is less than D*

The minimisation is conducted by searching over all possible mappings Q that satisfy the average distortion constraint.

Shannon's Mutual Information expressed as a function of Q(v|u)
I(u,v) = H(u) – H(u|v)
I is a function of the source entropy and the conditional entropy of u given v.
It is a measure of the amount of missing information in v

The R(D) function



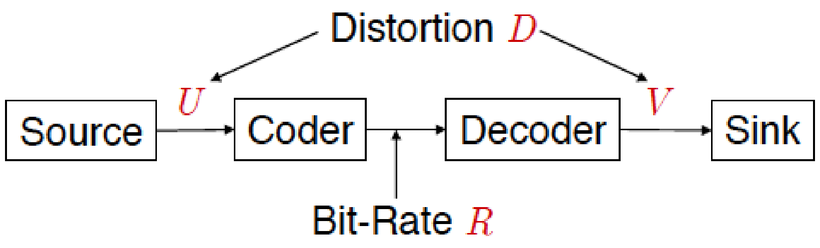
For a MEMORYLESS GAUSSIAN source p(u), and using MSE as D(u,v) you can show that

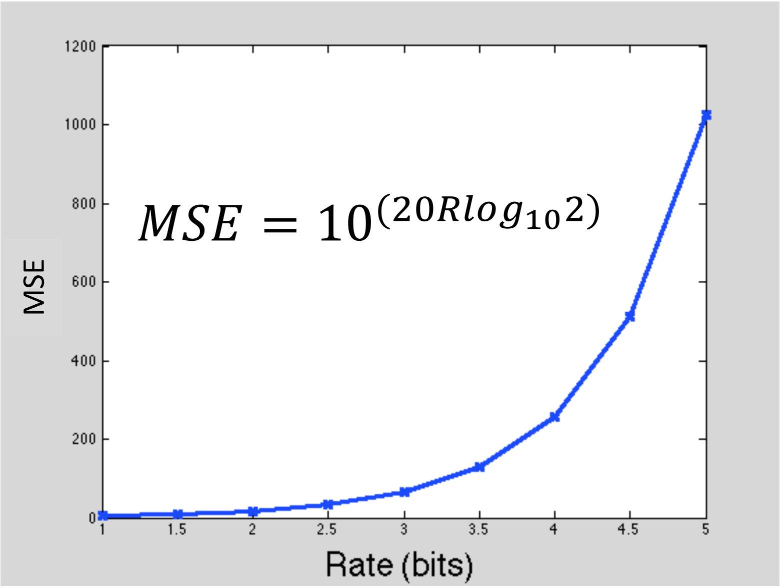
$$R(D^*) = \frac{1}{2}\log_2 \frac{\sigma^2}{D^*}; \qquad D(R^*) = \sigma^2 2^{-2R^*}$$

$$\begin{aligned} SNR = 10\log_{10}\frac{\sigma^2}{D^*} &= 10\log_{10}\frac{\sigma^2}{\sigma^2 2^{-2R^*}} \\ &= 10\log_{10} 2^{2R} \\ &= R20\log_{10} 2 \approx 6R \end{aligned}$$

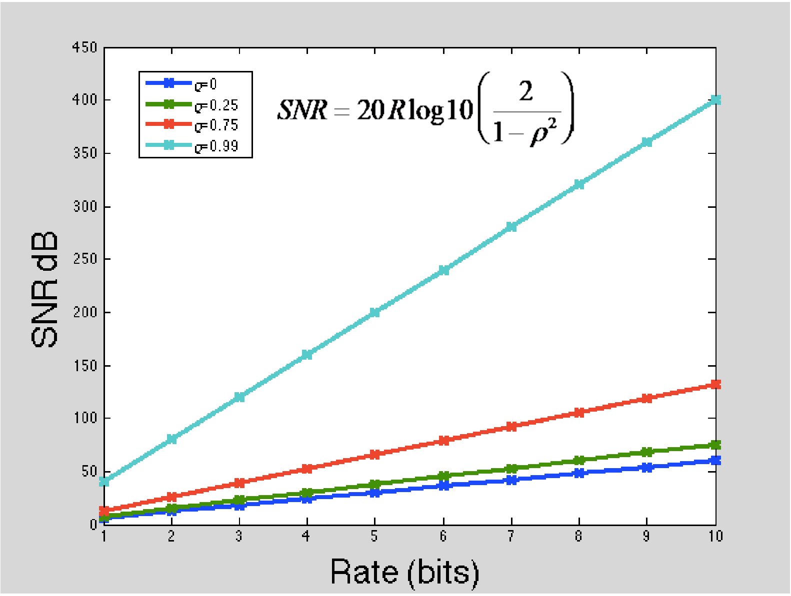
Rule of thumb 6dB Distortion roughly 1 bit in rate
R(D) for non-Gaussian sources ALWAYS below this curve

The theoretical R(D) function



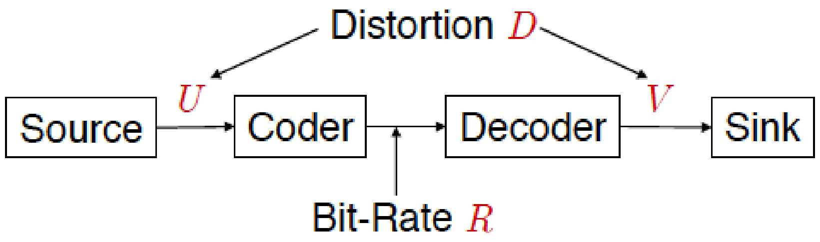


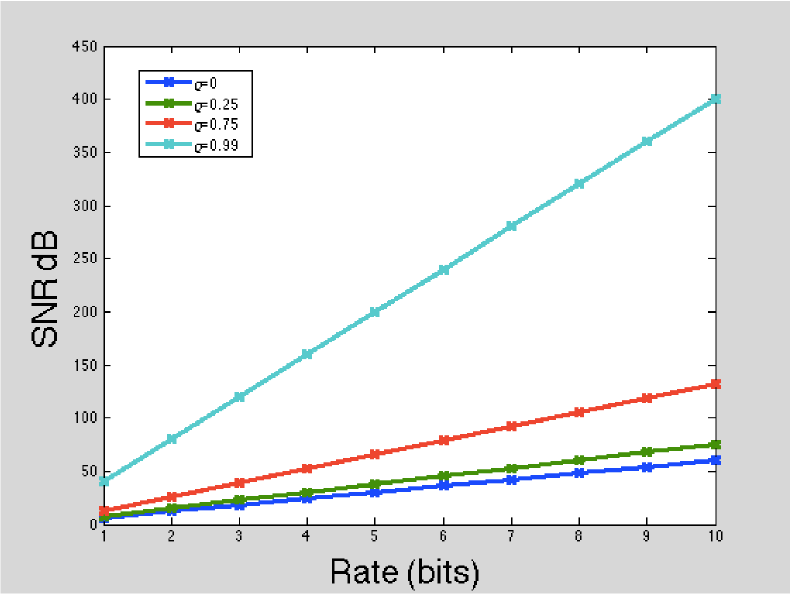
Memoryless, Gaussian Source



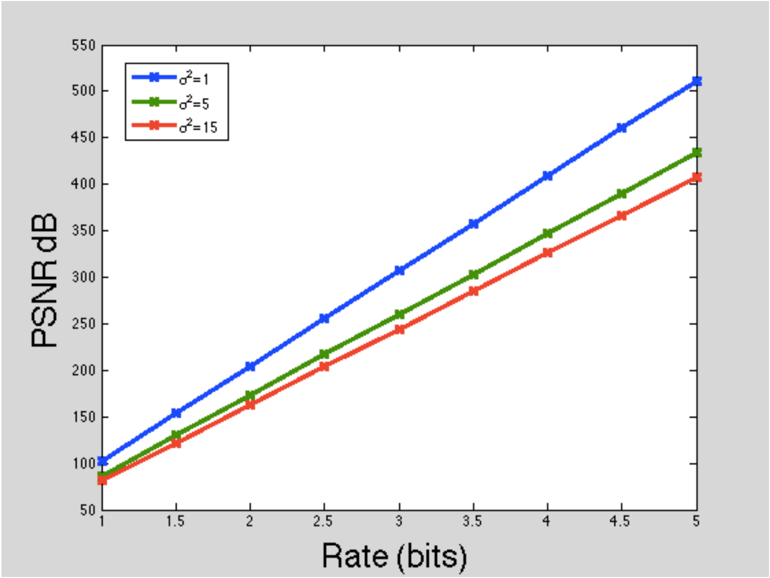
Gaussian Source, Correlation ρ

The R(D) function (PSNR different from SNR)





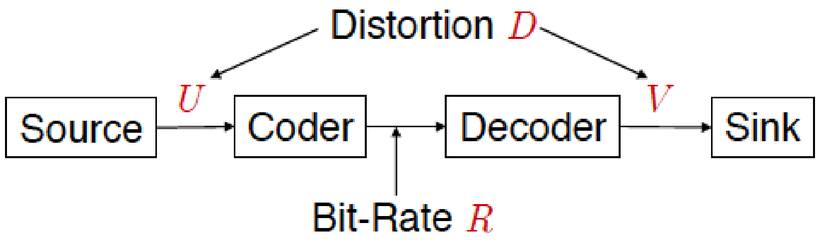
Gaussian Source, Correlation ρ

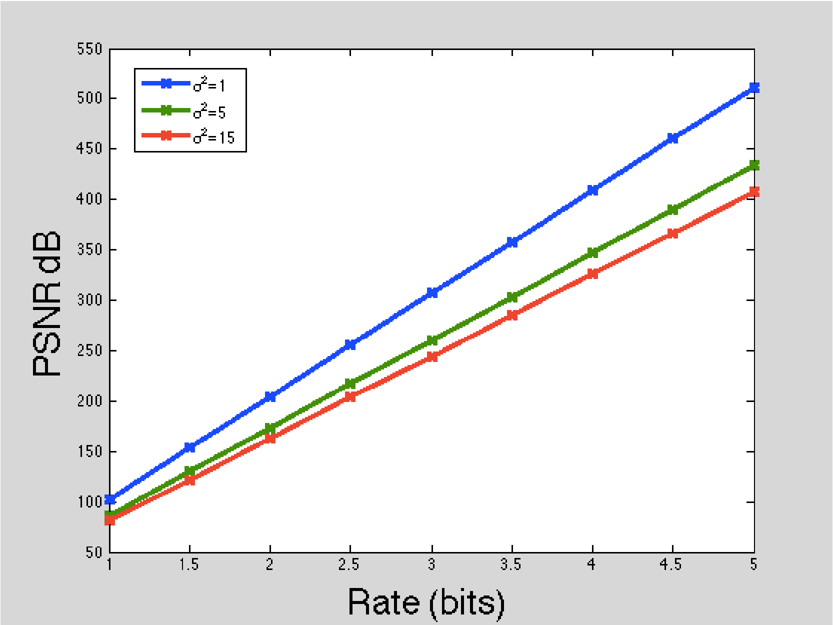


Memoryless, Gaussian Source

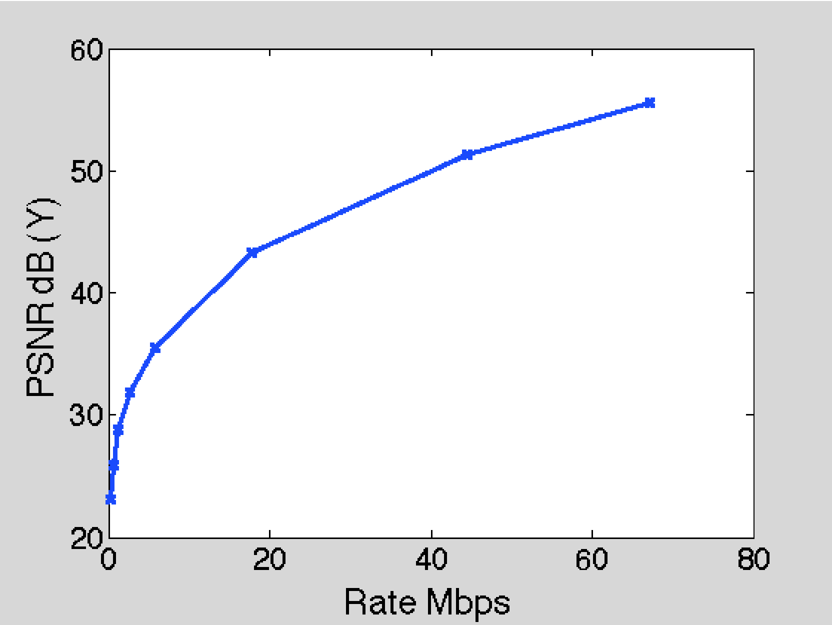
$$PSNR = 20R\log 10\left(\frac{255\times 255\times 2}{\sigma^2}\right)$$

Rate Distortion Theory gives theoretical limits. Real video is substantially more complex and contais much more redundancy than a memoryless Gaussian source. We can achieve much better R/D tradeoffs with real signals.





Memoryless Gaussian Source



Real Source (waterworld clip)

Is this useful?

The theory gives us bounds, given assumptions about the statistical nature of the source

Real sources are more complex. BUT given some on-the-fly measurements using this theory helps us predict what might happen given a particular video as we encode it.

So can be useful for control of quantisation etc

But we are still FAR away from a rate control system that we can use.

Practical rate control

TWO-PASS
(or MULTI PASS)

1. Encode the entire video file with some generic settings (quantisation, mode decision thresholds, CABAC)
2. Measure statistics of each frame encoding to get an idea how “complex” your source is.
3. Start again and use measurements from the first pass to to control the quantiser setting in the second pass by anticipating what’s going to happen next.

ONE-PASS

1. Start encoding with some generic settings.
2. Measure bits/sec used at each frame instant.
3. Use past behaviour to change the future settings of the quantiser (say). For example, if past 10 frames hit a higher rate than your target, increase the quantiser step size in the next frame. Conversely, if you are lower than your target, decrease the quantiser step size in the next frame.

Layers of Rate Control

Control rate from Block to Block

Control rate from Frame to Frame

Rate/Distortion optimisation of Mode decisions and Motion Estimation often used within the encoder.

Motion Estimation : Choose a motion vector which not only minimises the prediction error, but also yields a vector which can be encoded with minimum bits.

Mode Decision : Choose a block mode which not only minimises distortion but also yields the lowest number of bits for choosing that mode.

Minimise D over some parameter space w.r.t. the constraint $R < R^*$

Rate Control Optimisation

Rate-Distortion Optimisation for Video Compression, IEEE Signal Processing Magazine, vol 15, no 6, pp 74-50, Nov 1998

Gives an overview of methods for optimising coding decisions over each region in the image.

$$\text{Min}\{D\} \text{ , subject to } R < R_c$$
$$\text{is equivalent to}$$
$$\text{Min}\{J\} \text{ , where } J = D + \lambda R$$

This means we can solve for the best R,D combination by minimising J over a range of values for the Lagrange Multiplier.

1. Choose a value for λ
2. Vary the parameter in question (say Quantisation) to yield a range of values for R,D choose the R/D that gives the lowest J.
3. Repeat for other values of λ and pick the best when you're done.

Lagrangian optimisation

$$\mathrm{Min}\{D\}\ \text{, subject to } R < R_c$$
$$\text{is equivalent to}$$
$$\mathrm{Min}\{J\}\ \text{, where } J = D + \lambda R$$

Clearly if lambda is BIG then you'll penalise large R so your optimum for BIG lambda is to have low R and "higher" D

If lambda is SMALL then you'll penalise large D so your optimum for SMALL lambda is to have high R and "lower" D

BUT we can only find the R < Rc with the "right" lambda. That is the problem.

Basic intuition set out by Wiegand and Girod in 2001

Called RDO : Rate Distortion Optimisation

Bitrate R is a function of Δ the quantisation step size. Distortion D is also a function of Δ

Problem : Minimise $D(\Delta)$ w.r.t $\mathrm{R}(\Delta) \leq R_t$
The Lagrangian approach implies that this is the same as minimising $J(\Delta) = D(\Delta) + \lambda R(\Delta)$ w.r.t R
Problem becomes to choose Δ and λ which allows the lowest distortion for some target rate

So $\frac{dJ}{dR} = \frac{dD}{dR} + \lambda = 0$ at minimum ⇒ $\lambda = -\frac{dD(\Delta)}{dR(\Delta)}$ (inserting the dependency on quantizer step size again)

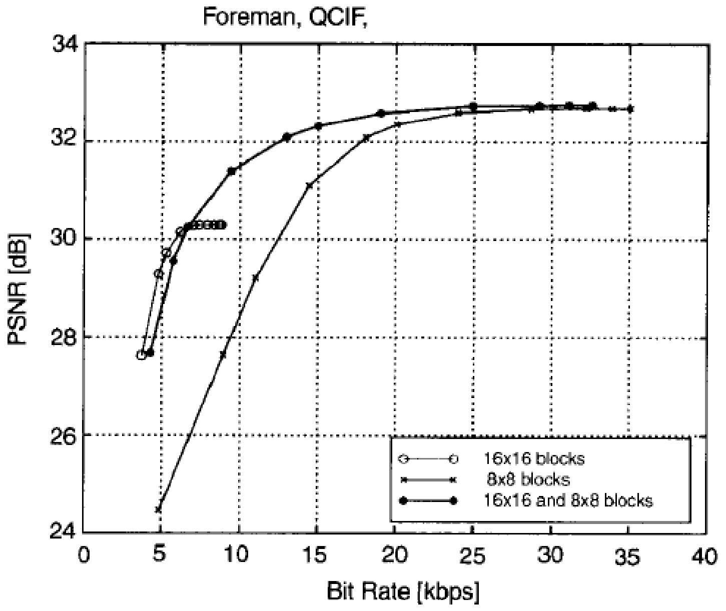
Using some approximations $R(\Delta) = alog_2(\frac{b}{D(\Delta)})$; and $D = (2\Delta)^2/12$; they show that $\lambda_{opt} \propto \Delta^2$

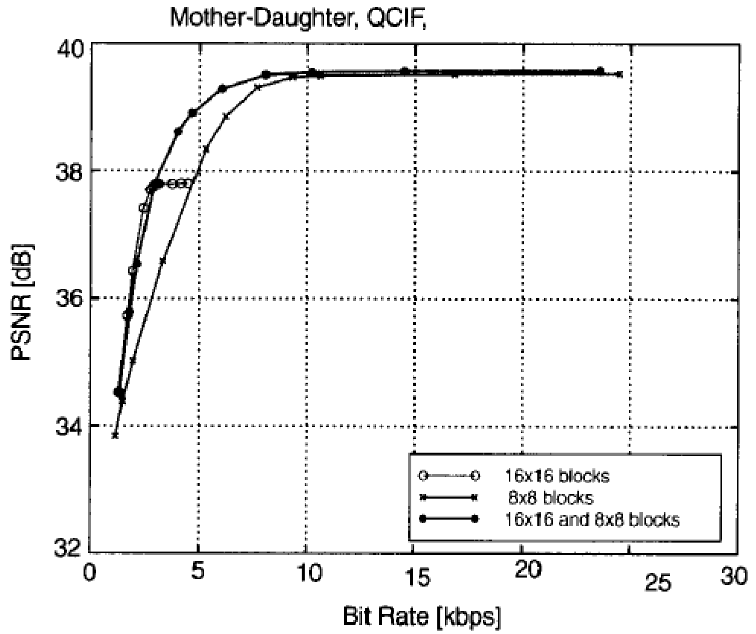
By experiment with a number of files and H.264 encoding they confirm that the optimal λ is related to Δ

$$\lambda_{opt} = 0.85\Delta^2$$

T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," Proceedings 2001 International Conference on Image Processing, Thessaloniki, Greece, 2001, pp. 542-545 vol.3

Some performance results





$$J_{\mathrm{MOTION}} = D_{\mathrm{DFD}} + \lambda_{\mathrm{MOTION}} R_{\mathrm{MOTION}}$$

Choose MV to minimise DFD wrt R<Rc
Use 1 bit for SKIP, and 2 bits for INTER, INTER+4MV

H.264 Rate Control

The best rate control algorithm proposed by Loren Merrit et al of x264/VideoLan open source community.

Not much information available about it outside the codebase itself.

They spoke about it in this:
Improved Rate Control and Motion Estimator for H.264 Encoder. Loren Merrit and Rahul Vanam, IEEE International Conference on Image Processing 2007, Vol 5, pp 309-312

Not a great paper, leaving out KEY information like the _exact_ control algorithm. I suspect deliberately.

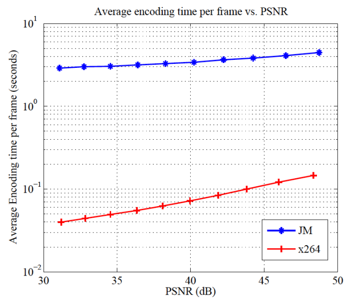
2-pass H.264

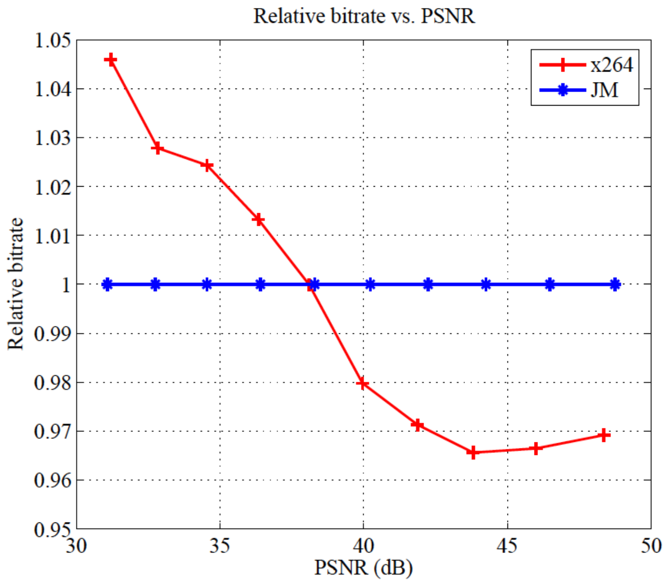
Task : Choose QP (quantisation parameter) to hit some bit rate constraint.

Given the required rate Rc bits/sec, calculate the target file size F = Rc*T bits where the video duration is T secs.

1. Run through all the frames in a first pass using constant QP (=15?), recording the bits allocated to each frame. Assign bits for frame k as b_k
2. For each P frame, calculate $g_k = a*b_k^{0.6}$.
3. Assuming g_k is the bits/frame for each P frame, scale each of those g_k to match the filesize F bits i.e. $m(g_0+g_1+g_2+\ldots g_N)$ =F, where the scale factor is m, and the number of frames is N.
4. Start encoding frame 1 with a fixed QP, measure the rate thus far
5. If second pass is "consistently off" from predicted filesize, then add an offset to all future QPs, $d = 2^{[(F/f)/6]}$
6. Continue

Well … it works



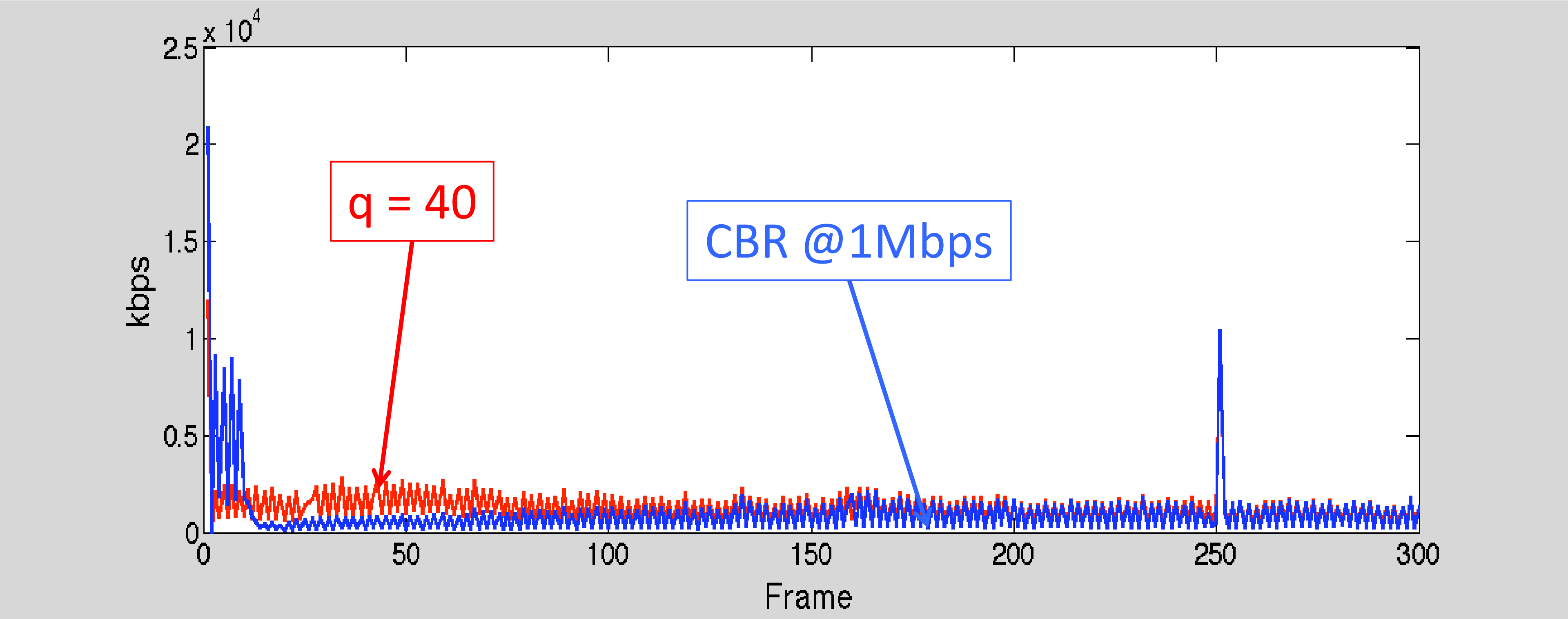


Approach	Bitrate (kb/s)	PSNR (dB)
JM-CBR	1189.4	40.810
x264-CBR	1120.7	41.569
JM-CQP	1212.9	44.035
x264-CQP	1137.2	44.033
x264-ABR	1169.9	42.815
x264-CRF	1133.9	44.176
x264-2pass	1199.2	44.472

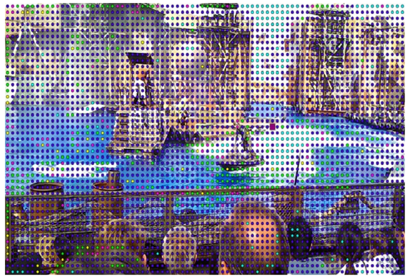
http://www.elephantsdream.org/

Aminated short @HD

H.264 Constant Q Versus CBR



You can manually choose the “right” q to get about 1Mbps. But the CBR algorithm works out what it should be adaptively.



Modern practice ….

Netflix Engineering Blogs

Per Title Encode Optimisation

NETFLIX

The Netflix Tech Blog

Monday, December 14, 2015

Per-Title Encode Optimization

We've spent years developing an approach, called per-title encoding, where we run analysis on an individual title to determine the optimal encoding recipe based on its complexity. Imagine having very involved action scenes that need more bits to encapsulate the information versus unchanging landscape scenes or animation that need less. This allows us to deliver the same or better experience while using less bandwidth, which will be particularly important in lower bandwidth countries and as we expand to places where video viewing often happens on mobile networks.

Background

In traditional terrestrial, cable or satellite TV, broadcasters have an allocated bandwidth and the program or set of programs are encoded such that the resulting video streams occupy the given fixed capacity. Statistical multiplexing is oftentimes employed by the broadcaster to efficiently distribute the bitrate among simultaneous programs. However, the total accumulated bitrate across the programs should still fit within the limited capacity. In many cases, padding is even added using null packets to guarantee strict constant bitrate for the fixed channel, thus wasting precious data rate. Furthermore, with pre-set channel allocations, less popular programs or genres may be allocated lower bitrates (and therefore, worse quality) than shows that are viewed by more people.

Links

- Netflix US & Canada Blog
- Netflix America Latina Blog
- Netflix Brasil Blog
- Netflix Benelux Blog
- Netflix DACH Blog
- Netflix France Blog
- Netflix Nordics Blog
- Netflix UK & Ireland Blog
- Netflix ISP Speed Index
- Open positions at Netflix
- Netflix Website
- Facebook Netflix Page
- Netflix UI Engineering
- RSS Feed

About the Netflix Tech Blog

Final Comments

- Rate control is NOT defined in the standards.
- Rate control is an important “special sauce” that makes a codec from one company better than another
- Rate control is crucial for enabling the Digital Video Market
- It is quite a gnarly topic. Many ad-hoc “rules that work” applied in practice
- Multi-pass encode is much better than single-pass encode for attaining a rate/distortion operating point.
- Real time video conferencing can only use single-pass (or nearly so) encoding.
- For next lecture watch this : https://www.youtube.com/watch?v=DYYZd_d4QBw (Thanks Viboothi)