UIT - Secteur de la normalisation des télécommunications
ITU - Telecommunication Standardization Sector
UIT - Sector de Normalización de las Telecomunicaciones

Study Period 2001-2004

Commission d'études / Study Group / Comisión de Estudio **16**

Contribution tardive / Delayed Contribution / Contribución tardía **D.146**

Porto Seguro, 28. May - 8. June 2001

Texte disponible seulement en / Text available only in / Texto disponible solamente en **E**

Question(s): 6

SOURCE*:	Heiko Schwarz and Thomas Wiegand Image Processing Department Heinrich-Hertz-Institute Einsteinufer 37 10587 Berlin Germany	

TITLE: An Improved H.26L Coder Using Lagrangian Coder Control

---

Summary

This document describes an H.26L coder whose operational mode is rate-distortion optimized using Lagrangian techniques. A similar method has been proposed to ITU-T VCEG by the second author in Q15-D-13[1] for the application to the test model of the ITU-T Recommendation H.263, version 2. The proposal in Q15-D-13 led to the creation of a new encoder recommendation (TMN-10) and is up to now the core of the high-complexity mode of the latest TMN.
The coder generates an H.26L bit-stream compliant with the H.26L Test Model Long Term Number 7 (TML-7)[2] where the context-based adaptive binary arithmetic coding (CABAC)[3] is used for

[1] ITU-T/SG 16/Q15-D-13 can be obtained via anonymous ftp to standard.pictel.com/video-site/9804_Tam/q15d13.doc.

[2] ITU-T/SG 16/Q.6/VCEG-M-81, "H.26L Test Model Long Term Number 7 (TML-7) draft0", May 2001.

* **Contact:** Tel: +49 30 31 002 206 / 617
Fax: +49 30 392 72 00
E-mail: hschwarz@hhi.de, wiegand@hhi.de

entropy coding. Comparisons are made against the ITU-T H.26L codec version 6.2. The simulations are run for the set of test sequences and conditions as specified in the MPEG coding efficiency CfP[4]. All bit-streams have been successfully decoded and an accompanying excel document contains results on R-D performance.

For motion estimation, reference frame selection, and intra 4x4-mode decision, we use bit-rates in our experiments that would have been obtained for the UVLC, while the actual entropy coding uses CABAC. The macroblock mode decision is based on the actual CABAC bit-rates.

We have observed that for all sequences tested, the improved encoding strategy provides PSNR gains of between 0.2 and 0.3 dB or, correspondingly, bit-rate savings of about 3-6% comparing against the TML-7 encoding strategy using CABAC.

1 Motion Estimation and Mode Selection

The problem of optimum bit allocation to the motion vectors and the residual coding in any hybrid video coder is a non-separable problem requiring a high amount of computation. To circumvent this joint optimization, we split the problem of macroblock encoding into two parts: motion estimation and mode decision. For that, motion estimation for the various modes (*16x16, 16x8, 8x16, 8x8*, etc.) is conducted first, and then given these motion vectors, the overall rate-distortion costs for all macroblock modes are computed for rate-constrained mode decision. In the remainder of this section, our approaches to motion estimation and mode decision are described. In the next section, the complete algorithm is given.

1.1 Rate-Constrained Motion Estimation

For each block or macroblock the motion vector is determined by full search on integer-pixel positions followed by quarter-pixel refinement. The integer-pixel search for the 16x16 block and the most recent decoded frame is conducted over the range [-16…16]x[-16…16] pixels relative to the motion vector prediction for the regarded block. For smaller blocks or older pictures the search range is reduced according to the TML-7. The search is conducted given the predictor of the block motion vector.

We view motion-compensated prediction as a source coding problem with a fidelity criterion. For bit-allocation, we use a Lagrangian formulation wherein distortion is weighted against rate using a Lagrange multiplier. More precisely, our integer-pixel motion search as well as our sub-pixel refinement returns the motion vector that minimizes

$$J(\mathbf{m} \mid \lambda_{MOTION}) = SAD(s,\ c(\mathbf{m})) + \lambda_{MOTION} \cdot R(\mathbf{m} - \mathbf{p})$$

with $\mathbf{m} = (m_x, m_y)^T$ being the motion vector, $\mathbf{p} = (p_x, p_y)^T$ being the prediction for the motion vector, and λ_{MOTION} being the Lagrange multiplier. The rate term $R(\mathbf{m} - \mathbf{p})$ represents the motion

---

[3] ITU-T/SG 16/Q.6/VCEG-L-13, “Adaptive Codes for H.26L“, January 2001.

[4] ISO/IEC JTC 1/SC 29/WG 11, “Call For Proposals On New Tools For Video Compression Technology”, Doc. N4065**,** March 2001, Singapore

information only and is computed by a table-lookup. The rate is estimated by using the universal variable length code (UVLC) table of the TML-7. The SAD is computed as

$$SAD(s,\ c(\mathbf{m})) = \sum_{x=1,\ y=1}^{B,\ B} \left| s[x,y] - c[x-m_x, y-m_y] \right|, \qquad B = 16, 8 \text{ or } 4 .$$

with *s* being the original video signal and *c* being the coded video signal. The choice of λ_{MOTION} has a rather small impact on the result of the 16x16 block motion estimation. But the search result for smaller blocks is strongly affected by λ_{MOTION}. In our coder, we choose

$$\lambda_{MOTION} = 2\text{^}(QP/6 - 1/2),$$

where QP is the macroblock quantization parameter.

The current H.26L test model (TML-7) already defines this concept of rate-constrained motion estimation. We have only adapted the Lagrangian multiplier λ_{MOTION}. Another difference is that for motion prediction of a 16x16 block with zero vector components no bias is subtracted from the motion vector cost $J(\mathbf{m} \mid \lambda_{MOTION})$ in our approach. There is no necessity to favour the skip mode during motion estimation since it will be treated separately in the macroblock mode decision (see section 1.2).

1.2 Rate-Constrained Mode Decision

In the proposed coder, the current macroblock mode is chosen given the mode decisions made for the past macroblocks. Rate-constrained mode decision refers to the minimization of the following Lagrangian functional

$$J(s,c,MODE \mid QP, \lambda_{MODE}) = SSD(s,c,MODE \mid QP) + \lambda_{MODE} \cdot R(s,c,MODE \mid QP)$$

where QP is the macroblock quantizer, λ_{MODE} is the Lagrange multiplier for mode decision, and $MODE$ indicates a mode chosen from the set of potential prediction:

I-frame: $MODE \in \{INTRA4x4,\ INTRA16x16\}$,

P-frame: $MODE \in \begin{Bmatrix} INTRA4x4,\ INTRA16x16,\ SKIP, \\ 16x16,\ 16x8,\ 8x16,\ 8x8,\ 8x4,\ 4x8,\ 4x4 \end{Bmatrix}$,

B-frame: $MODE \in \begin{Bmatrix} INTRA4x4,\ INTRA16x16,\ BIDIRECT,\ DIRECT, \\ FWD16x16,\ FWD16x8,\ FWD8x16,\ FWD8x8,\ FWD8x4, \\ FWD4x8,\ FWD4x4,\ BAK16x16,\ BAK16x8,\ BAK8x16, \\ BAK8x8,\ BAK8x4,\ BAK4x8,\ BAK4x4 \end{Bmatrix}$.

Note that the *SKIP* mode refers to the 16x16 mode where no motion and residual information is encoded. *SSD* is the sum of the squared differences between the original block s and its reconstruction c being given as

$$SSD(s,c,MODE \mid QP) = \sum_{x=1,y=1}^{16,16} \left(s[x,y] - c[x,y,MODE \mid QP]\right)^2 ,$$

and $R(s,c,MODE \mid QP)$ is the number of bits associated with choosing $MODE$ and QP including the bits for the macroblock header, the motion, and all DCT blocks. $c[x,y,MODE \mid QP]$ represents the reconstructed luminance values corresponding to $s[x,y]$.

We choose

$$\lambda_{MODE} = 2\text{^}(QP/3 - 1),$$

where QP is the macroblock quantization parameter.

The determination of the reference frame *REF* and the associated motion vectors for the *NxM* inter modes in P-frames and the *FWD NxM* modes in B-frames is done similar to the definition in the TML-7 by minimizing

$$J(REF \mid \lambda_{MOTION}) = SAD(s, c(REF, \mathbf{m}(REF))) + \lambda_{MOTION} \cdot (R(\mathbf{m}(REF) - \mathbf{p}(REF)) + R(REF)).$$

The rate term $R(REF)$ represents the number of bits associated with choosing REF and is computed by table-lookup using UVLC. The reference frame and block sizes for the bi-directional mode are chosen as combination of the "best" forward and backward mode.

The best *INTER16x16* prediction mode is chosen according to the TML-7. For the *INTRA4x4* prediction, the mode decision for each 4x4 block is performed similar to the macroblock mode decision by minimizing

$$J(s,c,IMODE \mid QP, \lambda_{MODE}) = SSD(s,c,IMODE \mid QP) + \lambda_{MODE} \cdot R(s,c,IMODE \mid QP)$$

where QP is the macroblock quantizer, λ_{MODE} is the Lagrange multiplier for mode decision, and $IMODE$ indicates an intra prediction mode:

$$IMODE \in \{DC, HOR, VERT, DIAG, DIAG\_RL, DIAG\_LR\}.$$

SSD is the sum of the squared differences between the original 4x4 block s and its reconstruction c and $R(s,c,IMODE \mid QP)$ represents the number of bits associated with choosing *IMODE*. It includes the bits for the intra prediction mode and the DCT-coefficients for the 4x4 luminance block. The rate term is computed using the UVLC entropy coding since the state of the arithmetic coder for coding the intra prediction modes and the DCT-coefficients is not known during the process of intra mode selection (CBP is unknown for most of the 4x4 blocks at this stage).

2 The Algorithm for Rate-Constrained Encoding

The procedure to encode one macroblock s in a I-, P- or B-frame in our video codec is summarized as follows.

1. Given the last decoded frames, λ_{MODE}, λ_{MOTION}, and the macroblock quantizer QP

2. Choose intra prediction modes for the *INTRA 4x4* macroblock mode by minimizing

$$J(s,c,IMODE \mid QP,\lambda_{MODE}) = SSD(s,c,IMODE \mid QP) + \lambda_{MODE} \cdot R(s,c,IMODE \mid QP)$$

with $IMODE \in \{DC, HOR, VERT, DIAG, DIAG\_RL, DIAG\_LR\}$.

3. Determine the best $INTRA16x16$ prediction mode according to the TML-7.

4. Perform motion estimation and reference frame selection by minimizing

$$J(REF,\mathbf{m}(REF) \mid \lambda_{MOTION}) = SAD(s,c(REF,\mathbf{m}(REF))) + \lambda_{MOTION} \cdot (R(\mathbf{m}(REF) - \mathbf{p}(REF)) + R(REF))$$

for each reference frame and motion vector of a possible macroblock mode.

5. Choose the macroblock prediction mode by minimizing

$$J(s,c,MODE \mid QP,\lambda_{MODE}) = SSD(s,c,MODE \mid QP) + \lambda_{MODE} \cdot R(s,c,MODE \mid QP),$$

given QP and λ_{MODE} when varying $MODE$. $MODE$ indicates a mode out of the set of potential macroblock modes:

I-frame: $MODE \in \{INTRA4x4, INTRA16x16\}$,

P-frame: $MODE \in \begin{Bmatrix} INTRA4x4, INTRA16x16, SKIP, \\ 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4 \end{Bmatrix}$,

B-frame: $MODE \in \begin{Bmatrix} INTRA4x4, INTRA16x16, BIDIRECT, DIRECT, \\ FWD16x16, FWD16x8, FWD8x16, FWD8x8, FWD8x4, \\ FWD4x8, FWD4x4, BAK16x16, BAK16x8, BAK8x16, \\ BAK8x8, BAK8x4, BAK4x8, BAK4x4 \end{Bmatrix}$.

The computation of $J(s,c,SKIP \mid QP,\lambda_{MODE})$ and $J(s,c,DIRECT \mid QP,\lambda_{MODE})$ is simple. The costs for the other macroblock modes are computed using the intra prediction modes or motion vectors and reference frames, which have been estimated in steps 2- 4.

3 Remaining Coder Parts

All remaining coder parts are operated as described in the H.26L Test Model Long Term Number 7 (TML-7) draft0 where CABAC is used for entropy coding.

4 Experimental Results

For illustration purpose of effectiveness of the proposed encoding strategy in comparison to the TML-7 using CABAC for entropy coding, parameterized rate-distortion curves are plotted. These curves show PSNR of the luminance component versus bit-rate measured of the complete bit-stream. PSNR is measured as the arithmetic mean of the PSNR values for each frame. The bit-rate is averaged over the complete sequence. The plots have 0.5 dB grid lines on the PSNR axis.

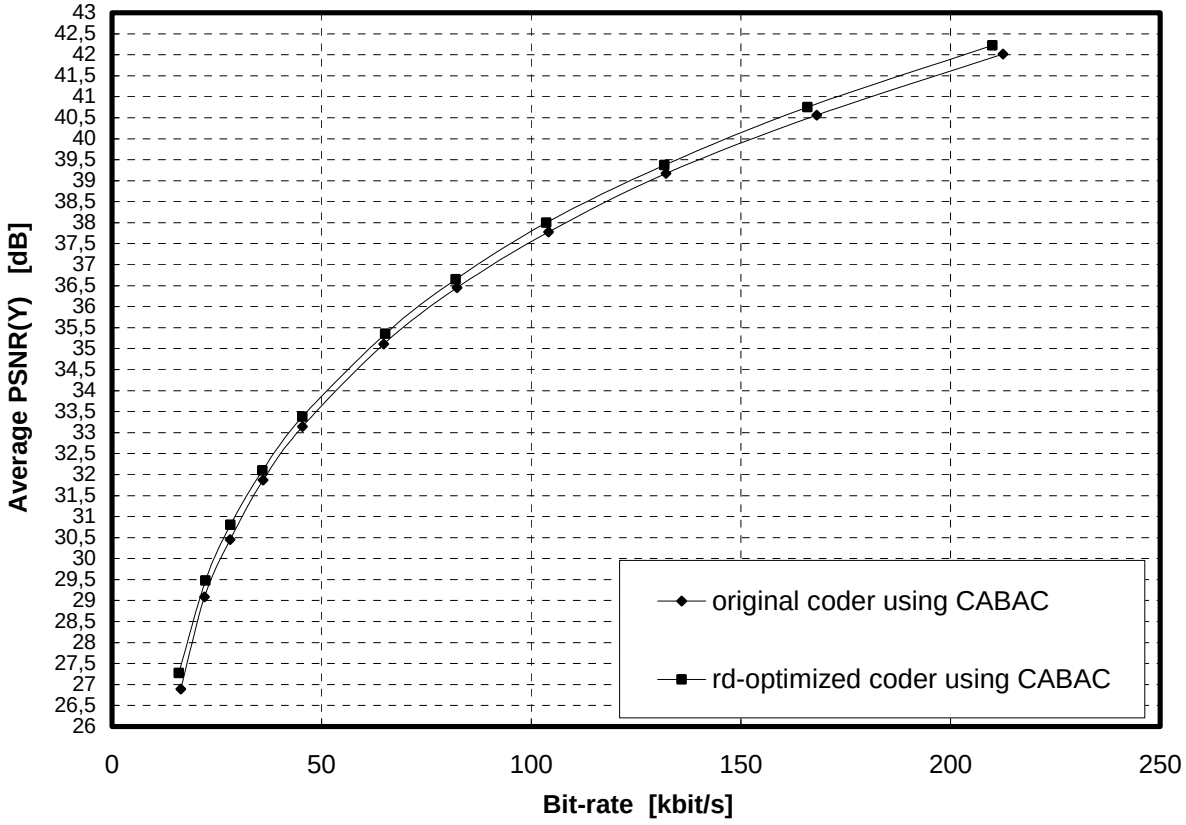
For the following sequences specified in the coding efficiency CfP rate-distortion curves are depicted:

Name	Format	Frame Rate	# of frames
Foreman	QCIF	10 fps	100
News	QCIF	10 fps	100
Container Ship	QCIF	10 fps	100
Tempete	QCIF	10 fps	87
Foreman	CIF	15 fps	150
News	CIF	15 fps	150
Container	CIF	15 fps	150
Tempete	CIF	15 fps	130
Bus	CIF	15 fps	75
Mobile & Calendar	CIF	15 fps	125
Flowers & Garden	CIF	15 fps	125

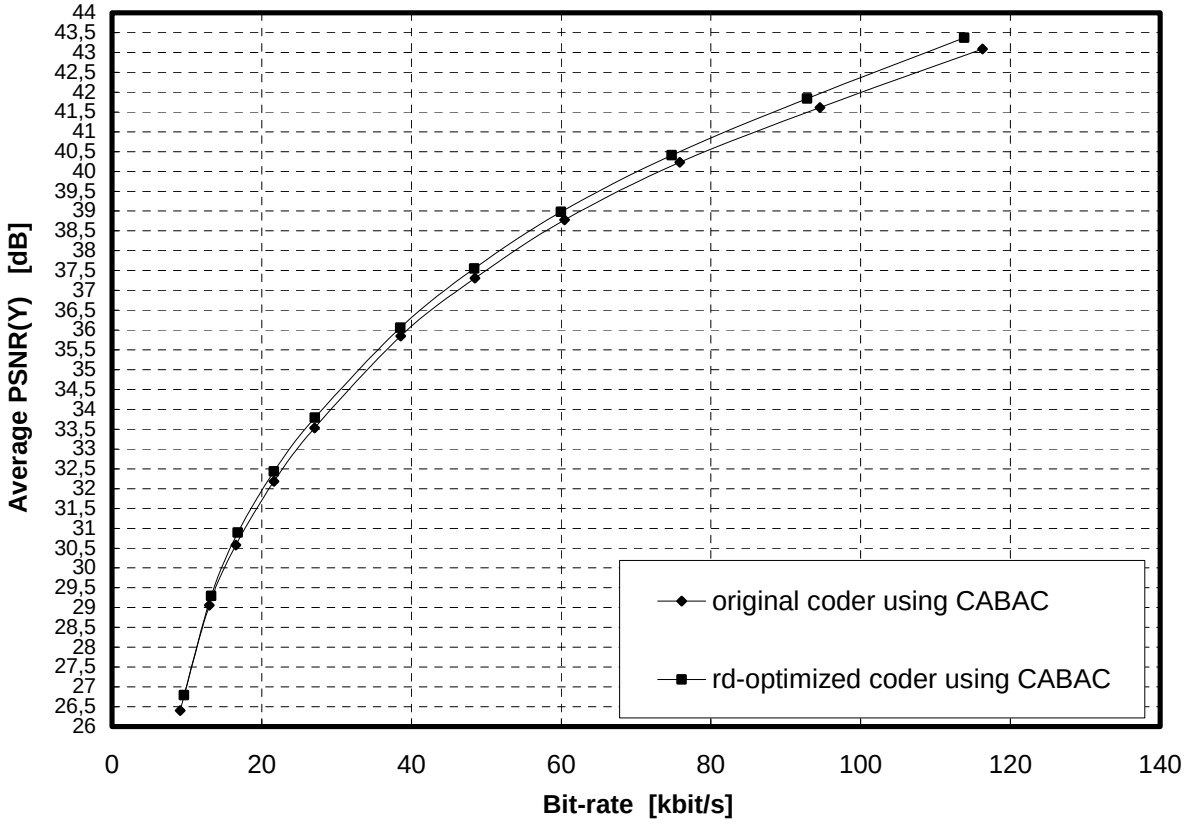
The rate-distortion curves are generated by varying the value of the macroblock quantizer QP that is fixed for a sequence. The parameters given below have been used for the original ITU-T H.26L coder version 6.2 coder as well as for the rate-distortion optimized coder.

QP_I (I-frame) and QP_P (P-frame)	31, 27, 25, 23, 21, 18, 16, 14, 12, 10, 8
QP_B (B-frame)	QP_P +2
M (distance P-P)	3 (i.e., 2 B-frames inserted)
Number of reference frames	5
Use of hadamard transform	enabled

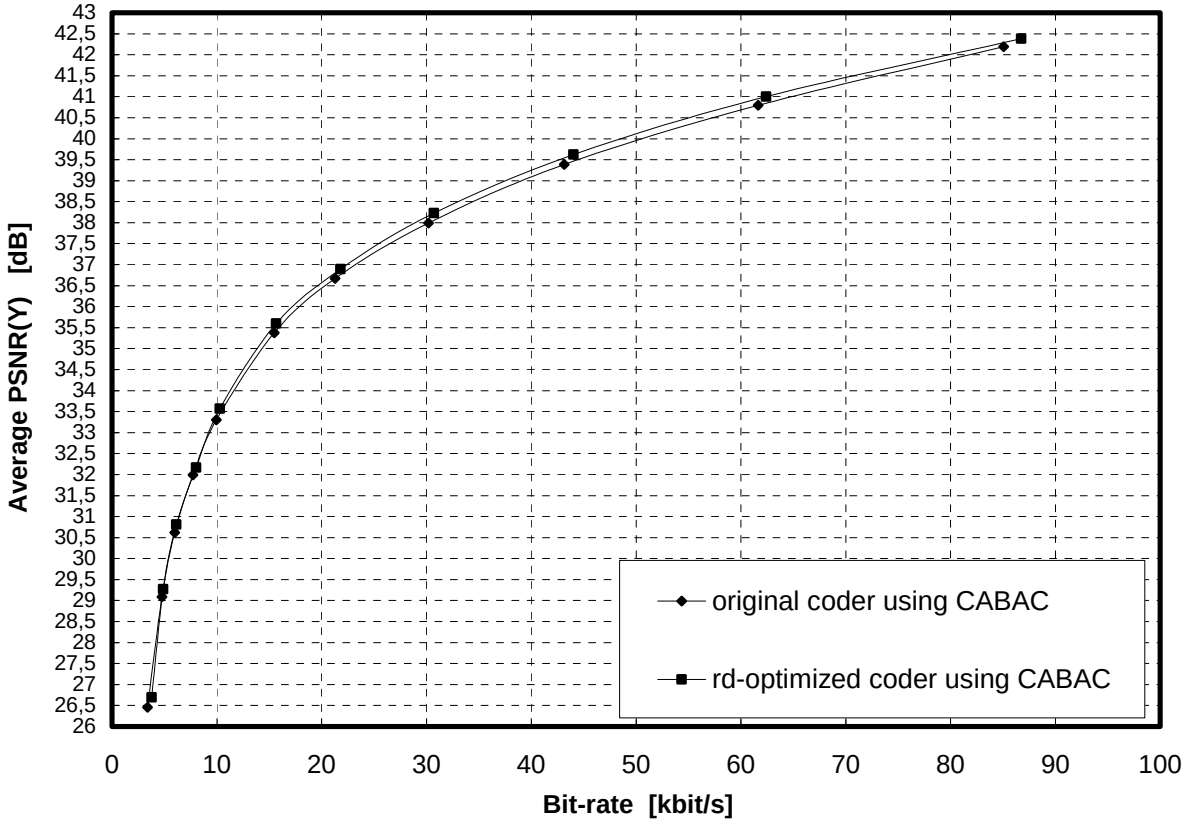
FOREMAN (QCIF, 10fps)



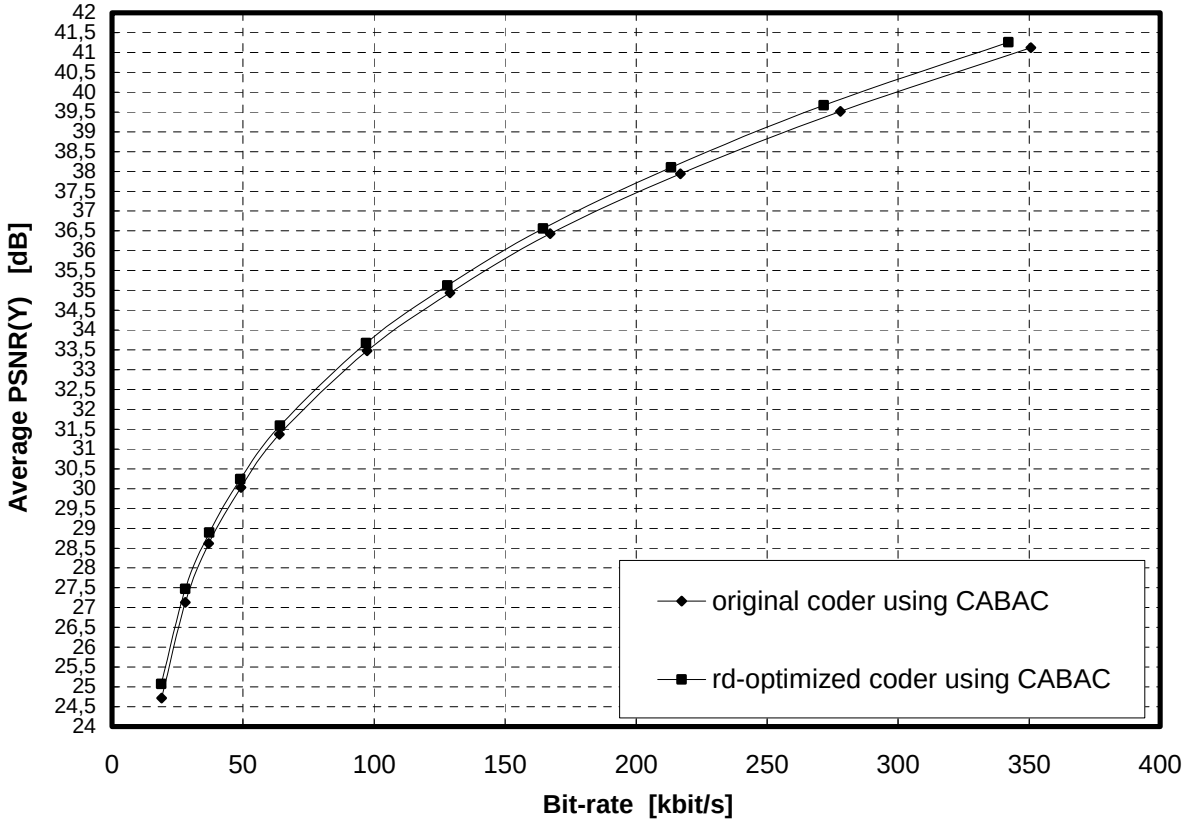
NEWS (QCIF, 10fps)



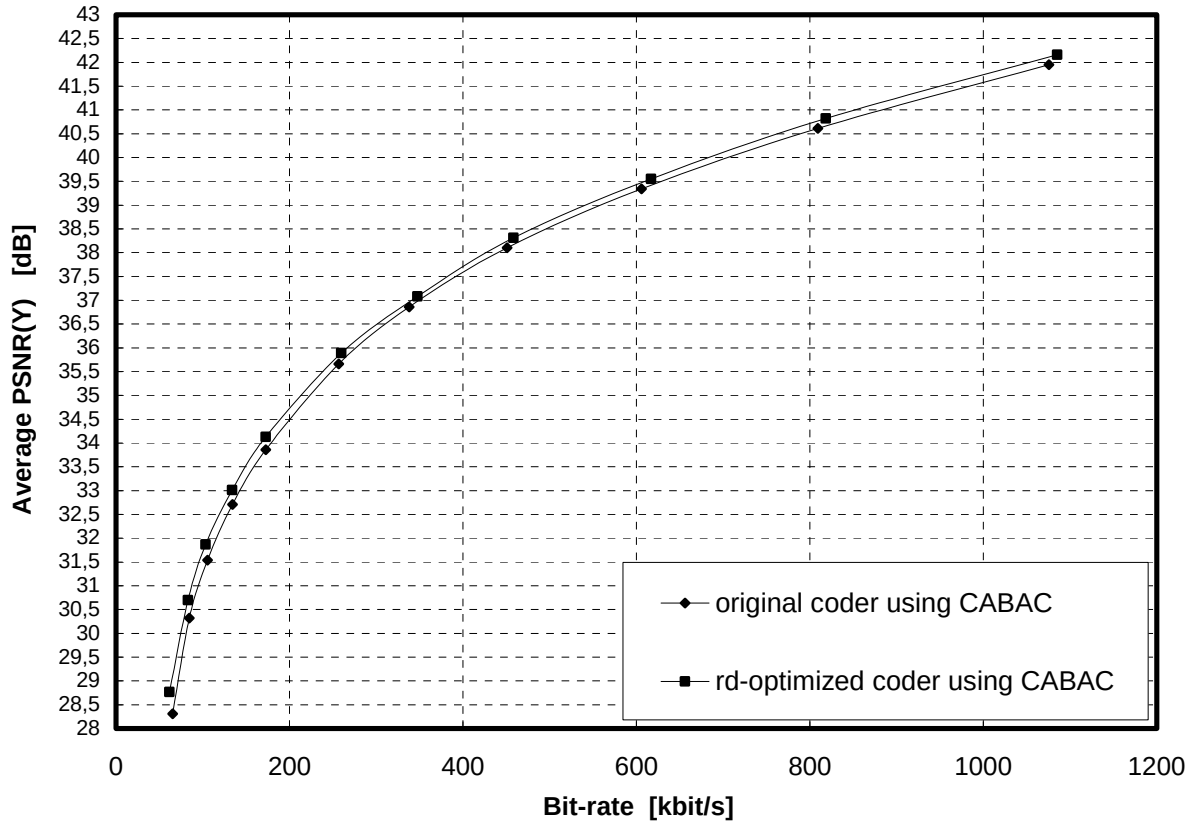
CONTAINER (QCIF, 10fps)



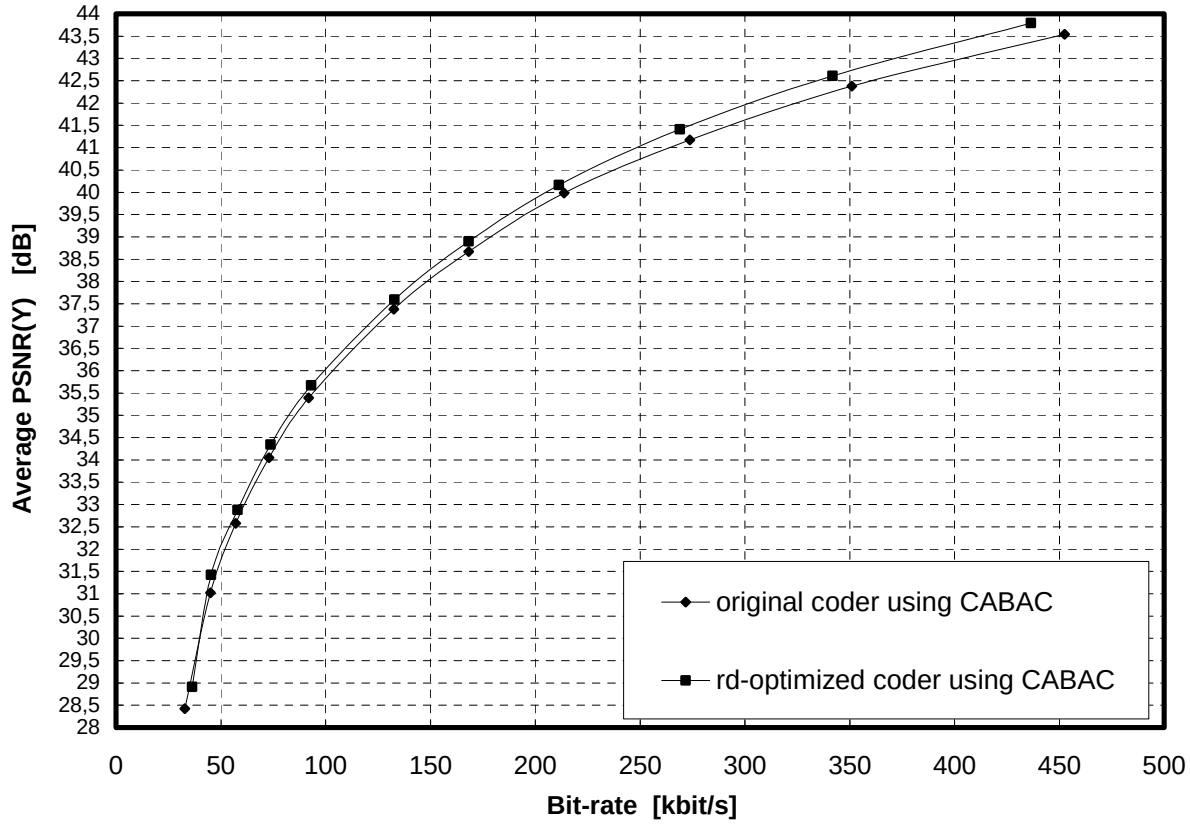
TEMPETE (QCIF, 10fps)



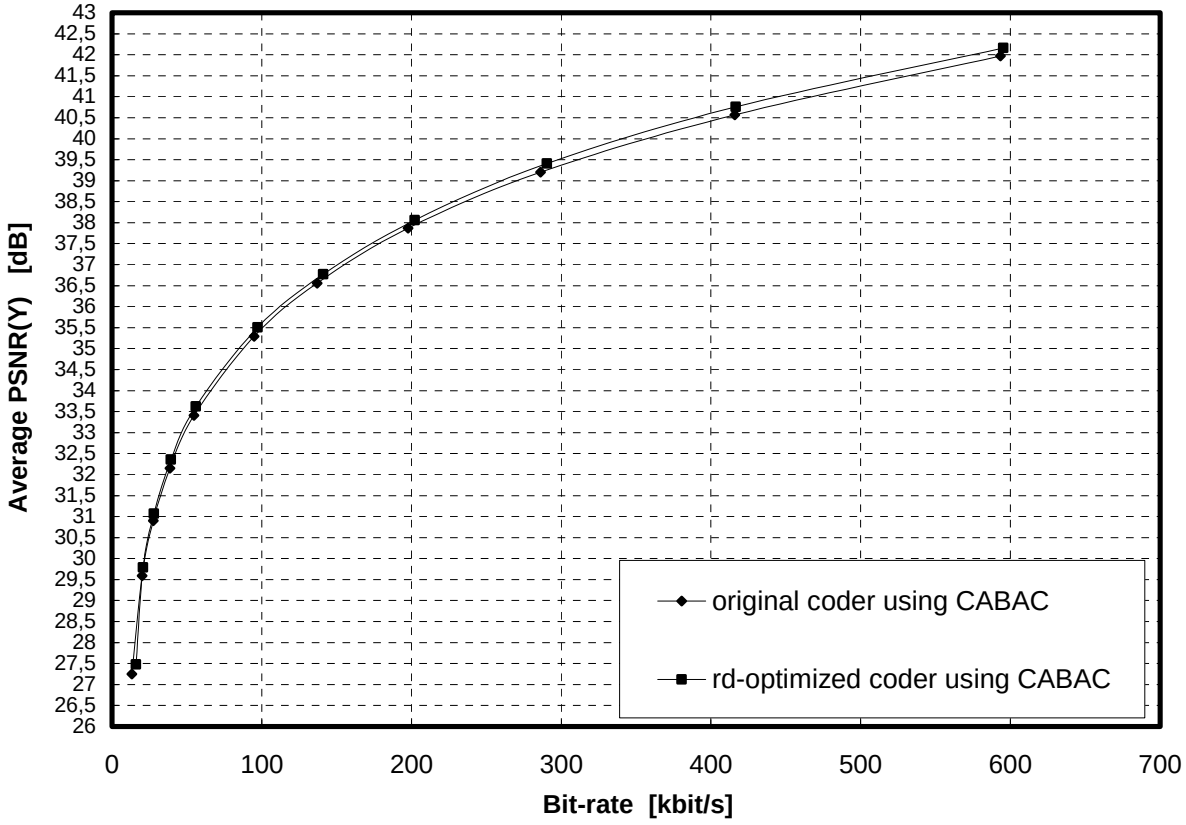
FOREMAN (CIF, 15fps)



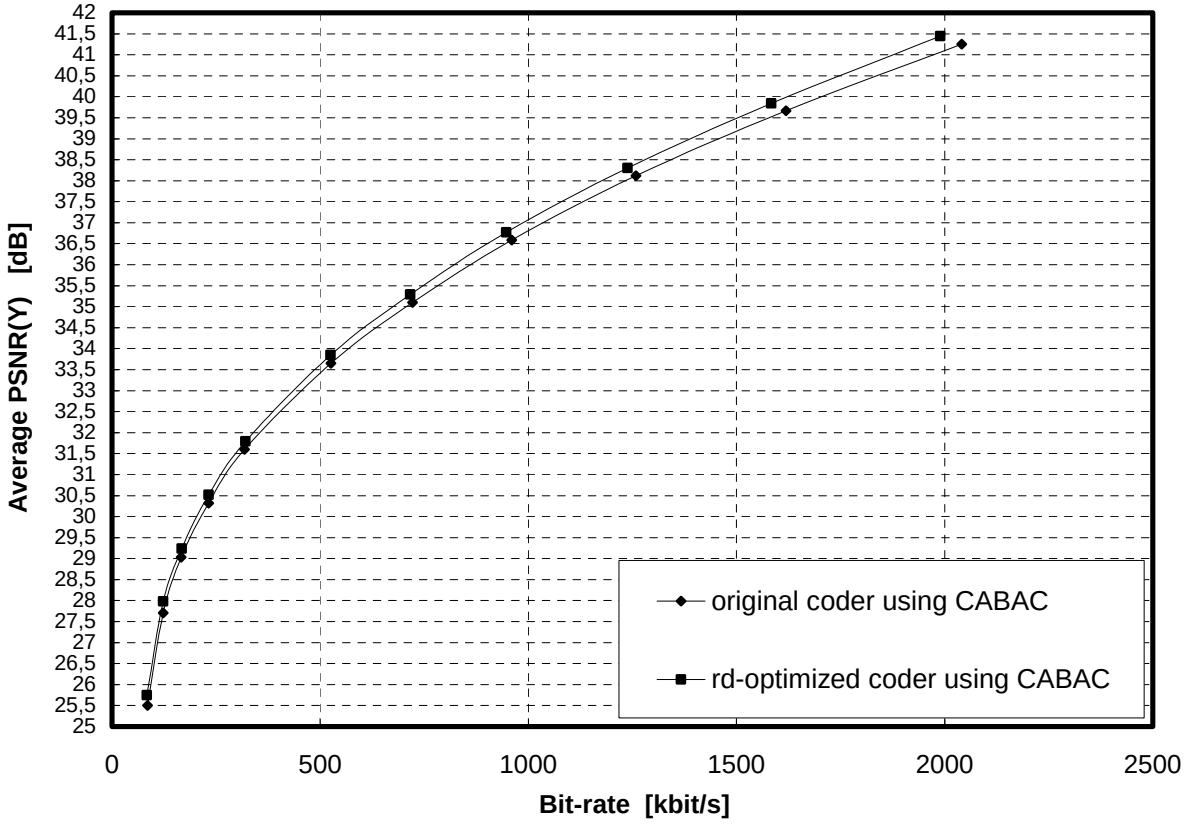
NEWS (CIF, 15fps)



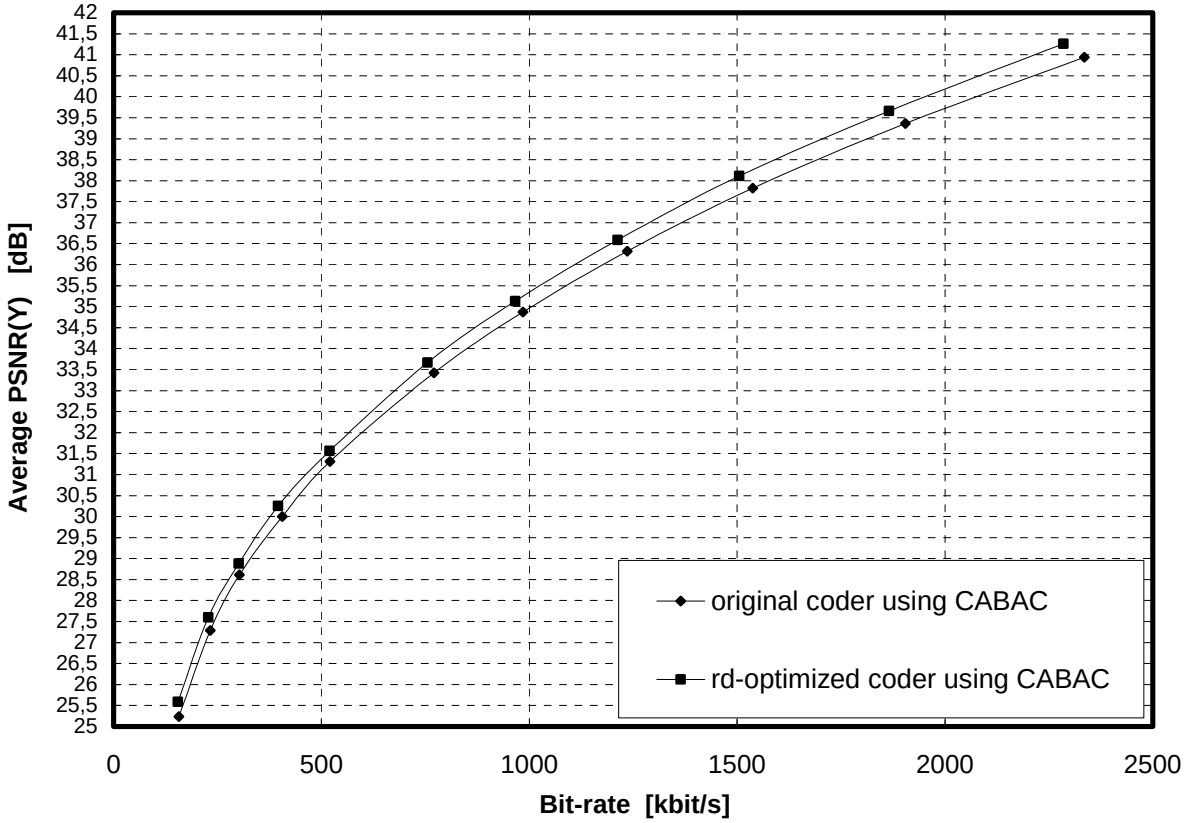
CONTAINER (CIF, 15fps)



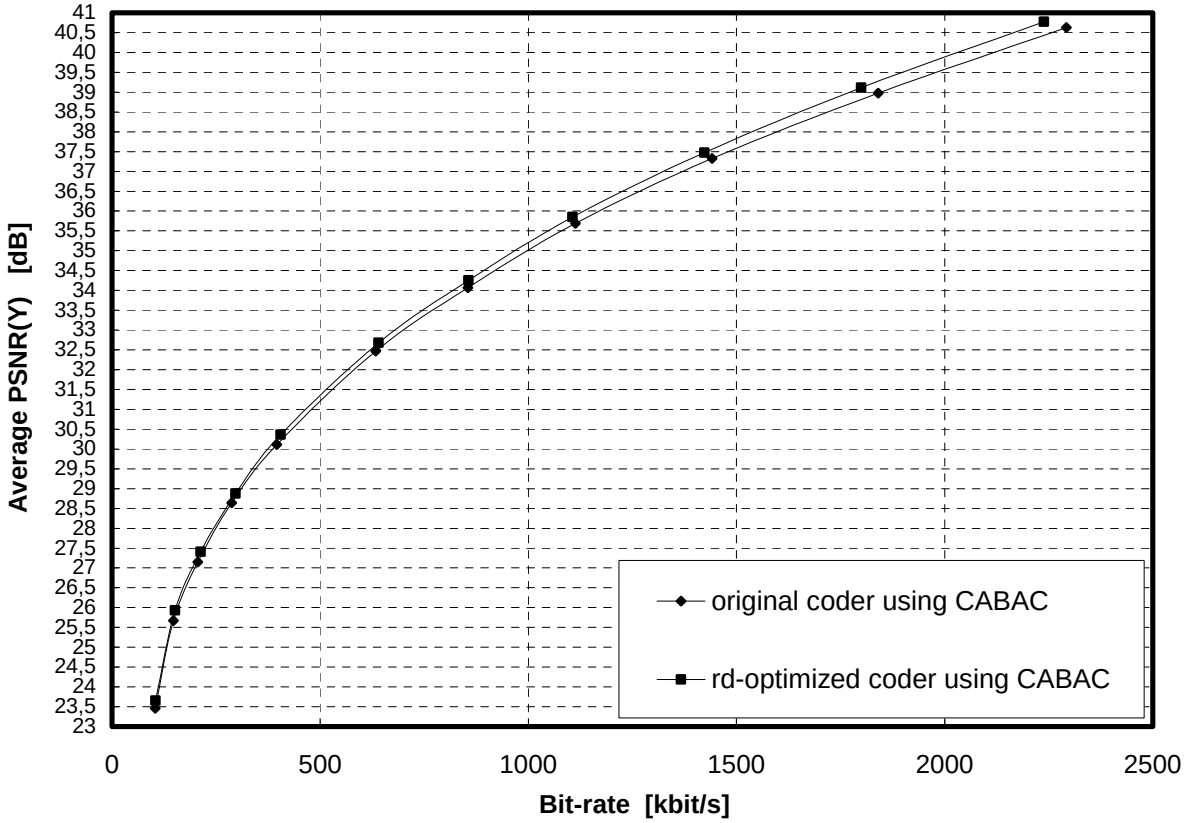
TEMPETE (CIF, 15fps)



BUS (CIF, 15fps)



MOBILE & CALENDAR (CIF, 15fps)



FLOWERS & GARDEN (CIF, 15fps)