PERCEPTUALLY OPTIMIZED QP ADAPTATION AND ASSOCIATED DISTORTION MEASURE

Input document JVET-H0047 (AHG10)

S. Bosse
C. Helmrich
H. Schwarz
D. Marpe
T. Wiegand

Fraunhofer HHI

Introduction

- Current: Constant QP per picture
- Known to lead to **perceptually suboptimal** results especially at low-medium bit-rates

→ Visible distortion in "smooth" regions with low visualy activity

→ Bits in "highly textured" regions with high visual activity could be saved to improve "smooth" regions

→ Solution: block-wise perceptual **QP adaptation** (QPA) based on visual activity

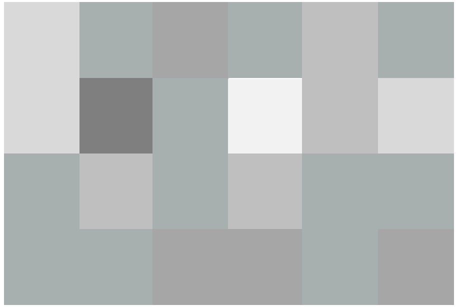
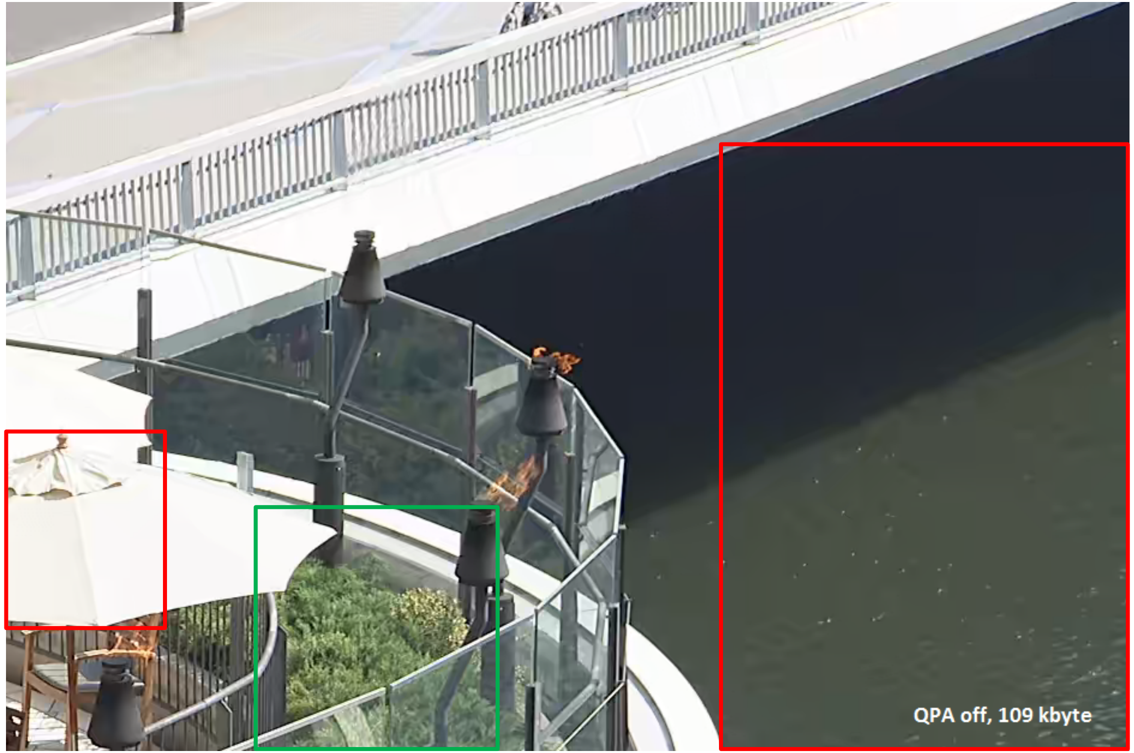


Illustration of Subjective Effect of local QP Adaptation



Bit allocation for subjective improvements:

Bits in highly textured regions

are moved to

smoother regions

Illustration of Subjective Effect of local QP Adaptation

Bit allocation for subjective improvements:

Bits in highly textured regions

are moved to

smoother regions

Adaptive QP and PSNR

- Problem: subjective quality improvements due to QPA not reflected in PSNR measure
- PSNR decreases → BD rate increase

	All Intra (10 bit)		
	PSNR		
	Y	U	V
Class A1	15.23%	6.15%	9.99%
Class A2	8.52%	0.86%	4.46%
Class B	6.89%	9.02%	10.88%
Class C	14.15%	13.97%	5.74%
Class D	6.34%	7.50%	6.25%
Class E	14.95%	9.06%	5.93%
Class F (optional)	19.57%	23.22%	22.30%
Overall (excl. F)	10.68%	7.76%	7.41%
Overall (incl. F)	11.95%	9.97%	9.54%
UHD + HD	9.96%	5.62%	8.63%
UHD only	11.88%	3.50%	7.22%

Impact on Lagrangian Bit Allocation (I)

- Current JEM bit allocation uses Lagrangian cost function to determine coding parameters $\boldsymbol{p_k}$

$$\min_{\{\boldsymbol{p_k}\}} D_{pic}(\{\boldsymbol{p_k}\}) + \lambda \cdot R_{pic}(\{\boldsymbol{p_k}\})$$

- For block-additive non-negative distortion measure like sum of squared errors $D_k = D_k^{SSE}$

$$\min_{\{\boldsymbol{p_k}\}} \sum\nolimits_{\forall k} D_k(\boldsymbol{p_k}) + \lambda \cdot \sum\nolimits_{\forall k} R_k(\boldsymbol{p_k})$$

- Allowing a per block minimization

$$\min_{\boldsymbol{p_k}} D_k^{SSE}(\boldsymbol{p_k}) + \lambda_{pic} \cdot R_k(\boldsymbol{p_k})$$

Impact on Lagrangian Bit Allocation (II)

- Weighted SSE distortion measure $D_k^{wSSE} = w_k \cdot D_k^{SSE}$ yields

$$\min_{\boldsymbol{p_k}} \; w_k \cdot D_k^{SSE}(\boldsymbol{p_k}) + \lambda \cdot R_k(\boldsymbol{p_k})$$

- Minimization can be written using a block-wise adjusted Lagrange multiplier

$$\min_{\boldsymbol{p_k}} \; D_k^{SSE}(\boldsymbol{p_k}) + \lambda_k \cdot R_k(\boldsymbol{p_k}) \; ,$$

$$\lambda_k = \frac{\lambda}{w_k}$$

- Lagrange multiplier is coupled to QP_k

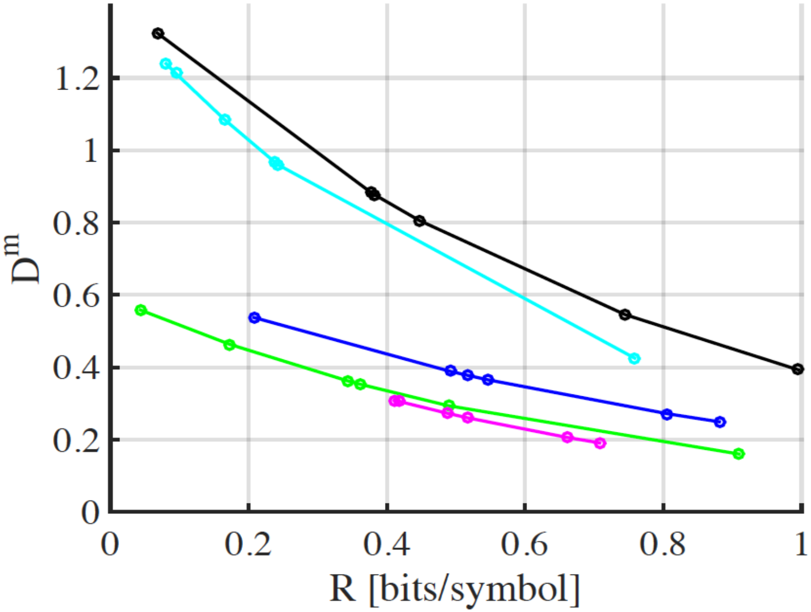
$$\lambda_k \propto \Delta_k^2 \propto 2^{\frac{\mathrm{QP}_k}{6}}$$

- Resulting in coupling of w_k and QP_k

$$\mathrm{QP}_k = \mathrm{QP} - 3\log_2 \frac{\lambda}{\lambda_k} = \mathrm{QP} - 3\log_2 w_k$$

Illustration of Impact on Lagrangian Bit Allocation (I)

Randomly pick 5 different functions with 6 accessibe R-D points



Constant QP: red circles
Adaptive QP: green circles

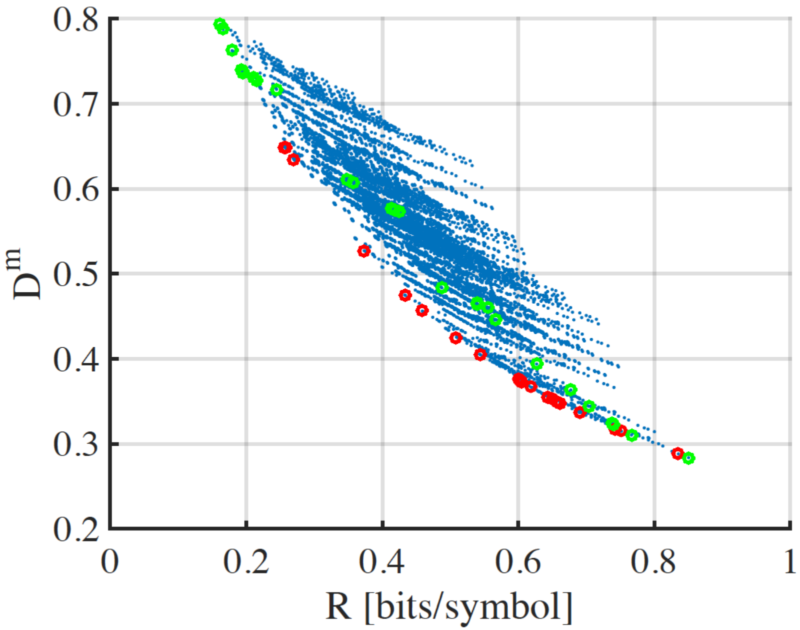
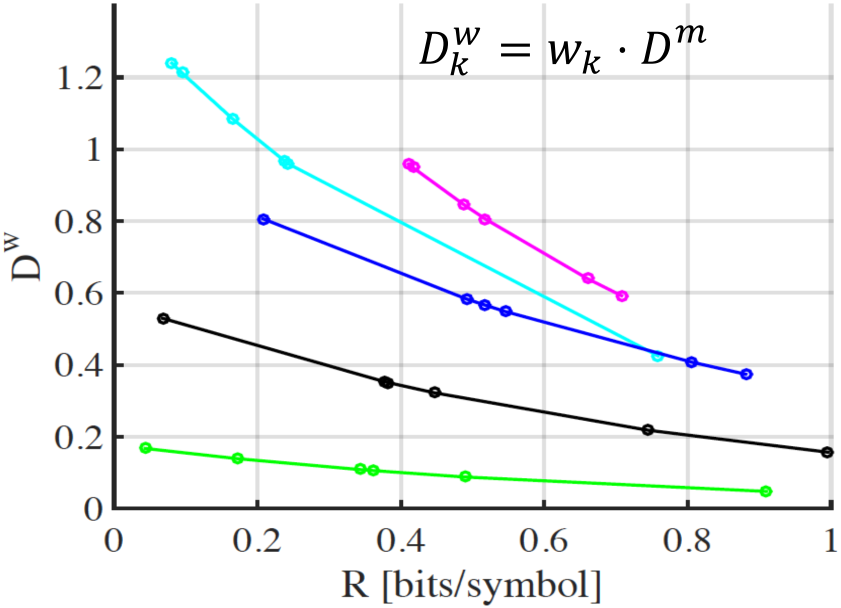
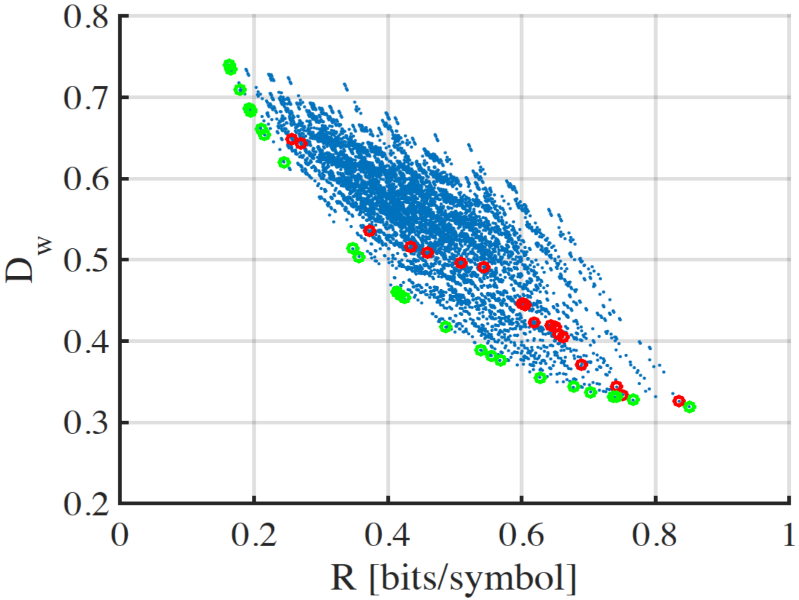


Illustration of Impact on Lagrangian Bit Allocation (II)

Random weights w_k
Weighted disortion rate functions

Constant QP: red circles
Adaptive QP: green circles





Weighted PSNR – WPSNR per picture

- Sum of squared errors (SSE) per picture

$$D_{pic}^{SSE} = \sum_k D_k^{SSE} = \sum_{(x,y)} (s[x,y] - s'[x,y])^2$$

- PSNR per picture
  $W \cdot H$ - picture size
  B - bit depth

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{W \cdot H \cdot 255^2 \cdot 2^{2B-16}}{D_{pic}^{SSE}}\right)$$

- Weighted SSE per picture
  w_k weight per block B_k

$$D_{pic}^{wSSE} = \sum_k D_k^{wSSE} = \sum_k w_k \cdot D_k^{SSE} = \\ = \sum_k w_k \sum_{(x,y)\in B_k} (s[x,y] - s'[x,y])^2$$

- Weighted PSNR per picture

$$\text{WPSNR} = 10 \cdot \log_{10}\left(\frac{W \cdot H \cdot 255^2 \cdot 2^{2B-16}}{D_{pic}^{wSSE}}\right)$$

BD-WPSNR

- BD-gains with QP adaptation vs QP for PSNR and WPSNR
- Loss of adaptive QP in BD-PSNR becomes gain in BD-WPSNR
- Per sequence

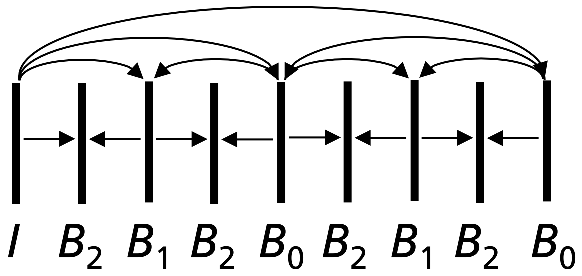
$$\mathrm{PSNR}_{\mathrm{seq}} = \frac{1}{N}\sum_{n} \mathrm{PSNR}[n]$$

$$\mathrm{WPSNR}_{\mathrm{seq}} = \frac{1}{N}\sum_{n} \mathrm{WPSNR}[n]$$

	All Intra (10 bit)					
	PSNR			WPSNR		
	Y	U	V	Y	U	V
Class A1	15.23%	6.15%	9.99%	-9.46%	-18.78%	-20.06%
Class A2	8.52%	0.86%	4.46%	-3.26%	-10.78%	-7.99%
Class B	6.89%	9.02%	10.88%	-5.12%	-4.07%	-1.93%
Class C	14.15%	13.97%	5.74%	-8.48%	-12.19%	-15.41%
Class D	6.34%	7.50%	6.25%	-5.61%	-3.24%	-4.10%
Class E	14.95%	9.06%	5.93%	-5.64%	-13.33%	-13.96%
Class F (optional)	19.57%	23.22%	22.30%	-14.17%	-19.88%	-23.00%
Overall (excl. F)	10.68%	7.76%	7.41%	-6.24%	-10.01%	-10.07%
Overall (incl. F)	11.95%	9.97%	9.54%	-7.37%	-11.42%	-11.92%
UHD + HD	9.96%	5.62%	8.63%	-5.88%	-10.66%	-9.37%
UHD only	11.88%	3.50%	7.22%	-6.36%	-14.78%	-14.02%

WPSNR for hierarchical coding structures (e.g. hier. B pics)

- Hierarchical coding structures:
  QP_n is adjusted for hierarchy level n
- Picture weighting factors $f_n = 2^{-\frac{\mathrm{QP}_n}{3}} = \lambda_n^{-1}$



- Within each pic:
  adaptive QP for subjective optimization
- Alternative for WPSNR for sequences

$$\mathrm{WPSNR}_{\mathrm{seq}} = 10 \cdot \log_{10}\left(W \cdot H \cdot 255^2 \cdot 2^{2B-16} \frac{\sum_n f_n}{\sum_n f_n D_{pic}^{wSSE}[n]} \right)$$

Comparison of WPSNR measures for hierarchical B pictures for adaptive QP bit allocation relative to constant QP/picture

	Random Access (10 bit)								
	PSNR			WPSNR			frame-weighted WPSNR		
	Y	U	V	Y	U	V	Y	U	V
Class A1	13.40%	7.85%	12.35%	-3.07%	-14.52%	-15.13%	-4.79%	-15.57%	-16.91%
Class A2	10.57%	3.57%	5.64%	-0.30%	-7.22%	-6.85%	-0.82%	-9.89%	-7.65%
Class B	11.20%	12.03%	14.78%	-2.42%	-4.21%	-1.23%	-3.87%	-5.70%	-1.89%
Class C	18.04%	16.25%	8.60%	-3.96%	-12.47%	-15.41%	-5.59%	-12.42%	-15.49%
Class D	6.27%	6.34%	3.58%	-6.54%	-5.95%	-6.37%	-7.28%	-5.87%	-7.03%
Class E									
Class F (optional)	23.27%	29.18%	25.12%	-6.66%	-18.58%	-21.66%	-12.42%	-21.01%	-22.31%
Overall (excl. F)	11.86%	9.34%	9.27%	-3.22%	-8.65%	-8.63%	-4.44%	-9.69%	-9.42%
Overall (incl. F)	13.69%	12.52%	11.80%	-3.77%	-10.24%	-10.72%	-5.72%	-11.50%	-11.48%
UHD + HD	11.68%	8.14%	11.22%	-1.97%	-8.31%	-7.24%	-3.22%	-10.03%	-8.29%
UHD only	11.98%	5.71%	8.99%	-1.69%	-10.87%	-10.99%	-2.81%	-12.73%	-12.28%

Suggestion for new JVET Experimental Common Conditions

- **Suggestion**: Use **QPA / WPSNR** instead of **constant QP / PSNR**
  - Perceptually optimized **QPA** using weights w_k prescribed by common conditions
  - **WPSNR** when matched to **QPA** (and QP cascading) can be used like PSNR for codec evaluation (even for small improvements)

- **Benefits**: investigate coding tools in a more realistic encoder setting
  - Overall visual quality notably improves (for corresponding sequences)
  - Codec design driven towards higher efficiency at higher subjective quality
  - Block-wise RD-optimized encoding processes unchanged (only $\lambda_k = \lambda / w_k$)

Modified Encoding Algorithm with Adaptive QP

1. Calculate weighting factors w_k for the blocks B_k (such as CTUs) of a video picture
2. Choose a picture QP and a corresponding picture Lagrange parameter λ
3. For each block B_k (in coding order):
   a) Set the block quantization parameter QP_k and the Lagrange multiplier λ_k

   $$\mathrm{QP}_k = \mathrm{QP} - 3\log_2 w_k \qquad \text{and} \qquad \lambda_k = \frac{\lambda}{w_k}$$

   b) Choose the coding parameters $\boldsymbol{p}_k$ for the block B_k

   $$\min_{\boldsymbol{p}_k} D_k^{SSE}(\mathrm{QP}_k, \boldsymbol{p}_k) + \lambda_k \cdot R_k(\mathrm{QP}_k, \boldsymbol{p}_k)$$

Evaluation: Bjøntegaard Delta-(W)PSNR

- Bjøntegaard delta-gains of **HEVC + JEM tools** over **HEVC**
- Coding efficiency improvements due to tools reflected by both

  BD-PSNR/constant QP

  and

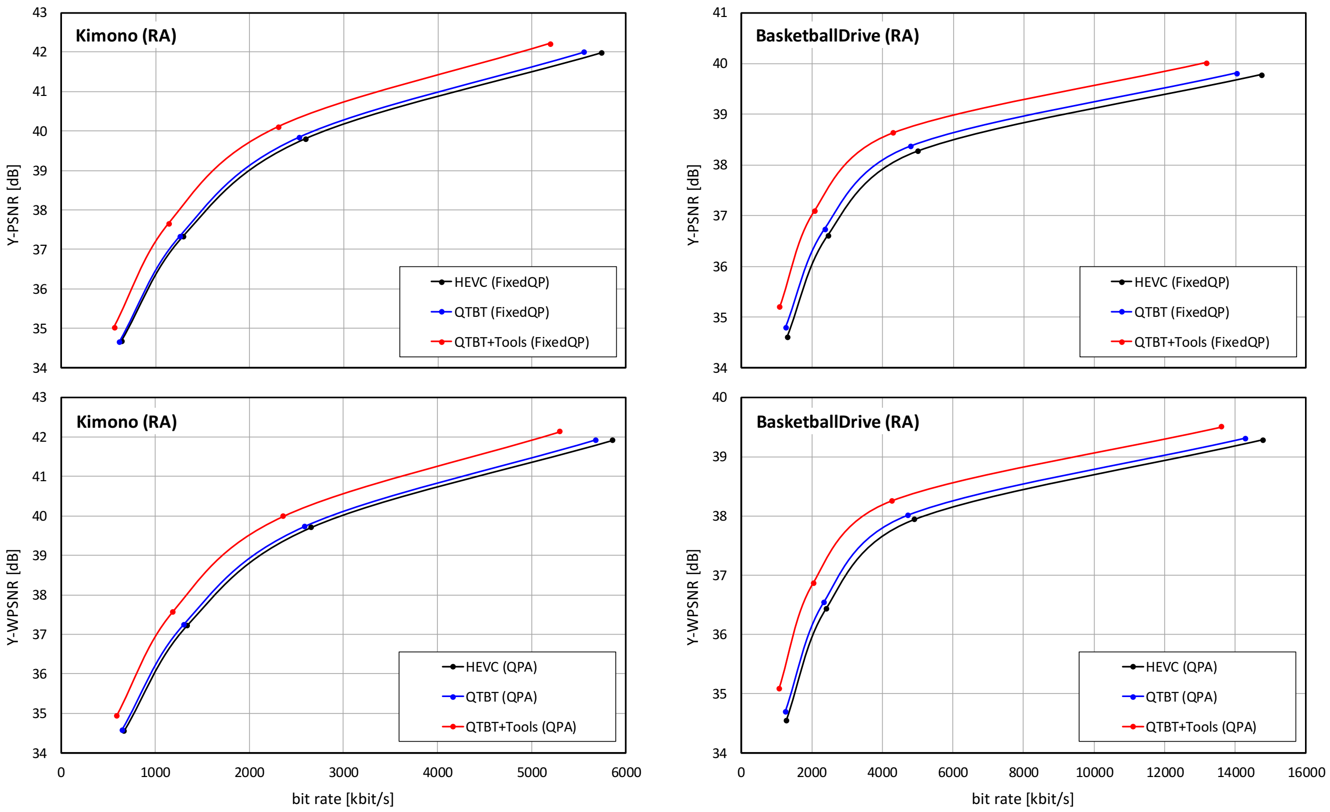
  BD-WPSNR/QPA

All Intra (10 bit)	PSNR / Fixed QP			WPSNR / QPA			frame-weighted WPSNR / QPA		
	Y	U	V	Y	U	V	Y	U	V
Class A1	-19.67%	-20.52%	-22.42%	-22.42%	-21.88%	-27.37%	-22.41%	-21.87%	-27.35%
Class A2	-22.81%	-23.71%	-20.36%	-25.06%	-22.44%	-21.38%	-25.06%	-22.45%	-21.38%
Class B	-17.13%	-17.39%	-14.86%	-16.74%	-16.45%	-14.99%	-16.74%	-16.41%	-14.96%
Class C	-17.93%	-17.96%	-21.41%	-17.99%	-18.28%	-20.28%	-17.98%	-18.27%	-20.25%
Class D	-14.90%	-13.70%	-17.50%	-14.85%	-12.46%	-16.15%	-14.86%	-12.49%	-16.18%
Class E	-21.51%	-20.24%	-22.46%	-21.72%	-21.69%	-22.46%	-21.72%	-21.69%	-22.45%
Class F (optional)	-18.52%	-22.25%	-24.37%	-19.27%	-18.98%	-20.90%	-18.46%	-18.48%	-20.15%
Overall (excl. F)	-18.81%	-18.80%	-19.52%	-19.59%	-18.65%	-20.13%	-19.59%	-18.64%	-20.12%
Overall (incl. F)	-18.77%	-19.29%	-20.21%	-19.54%	-18.70%	-20.24%	-19.43%	-18.62%	-20.12%
UHD + HD	-19.66%	-20.30%	-18.88%	-21.05%	-19.96%	-20.76%	-21.05%	-19.95%	-20.75%
UHD only	-21.24%	-22.11%	-21.39%	-23.74%	-22.16%	-24.37%	-23.73%	-22.16%	-24.37%

Random Access (10 bit)	PSNR / Fixed QP			WPSNR / QPA			frame-weighted WPSNR / QPA		
	Y	U	V	Y	U	V	Y	U	V
Class A1	-25.78%	-29.78%	-32.10%	-26.27%	-26.99%	-23.63%	-25.95%	-25.18%	-23.91%
Class A2	-34.43%	-37.14%	-31.95%	-36.35%	-37.11%	-33.34%	-34.54%	-33.36%	-31.03%
Class B	-24.59%	-30.24%	-24.71%	-23.80%	-28.07%	-22.92%	-22.56%	-23.79%	-21.35%
Class C	-23.95%	-26.40%	-29.11%	-24.06%	-25.25%	-26.47%	-23.97%	-23.96%	-25.56%
Class D	-24.13%	-23.63%	-25.93%	-24.13%	-22.50%	-23.37%	-23.77%	-22.51%	-23.57%
Class E									
Class F (optional)	-18.38%	-24.96%	-26.45%	-18.09%	-11.72%	-15.18%	-18.55%	-16.10%	-19.29%
Overall (excl. F)	-26.48%	-29.48%	-28.57%	-26.77%	-27.99%	-25.80%	-25.99%	-25.67%	-24.91%
Overall (incl. F)	-25.18%	-28.75%	-28.23%	-25.38%	-25.39%	-24.10%	-24.80%	-24.13%	-24.01%
UHD + HD	-27.98%	-32.22%	-29.21%	-28.42%	-30.52%	-26.35%	-27.29%	-27.16%	-25.12%
UHD only	-30.10%	-33.46%	-32.02%	-31.31%	-32.05%	-28.49%	-30.25%	-29.27%	-27.47%

Evaluation: Rate-(W)PSNR Curves (random access, 10 bit)

- **Constant QP** measured by luma-PSNR
- **QPA** measured by luma-WPSNR



JVET-H0047: Perceptually Optimized QP Adaptation and Assocoated Distortion Measure

Fraunhofer HHI

Calculation of weights w_k

- High-pass filter

$$h(x,y) = 4 \cdot s(x,y) - s(x-1,y) - s(x+1,y) - s(x,y-1) - s(x,y+1)$$

- For each block B_k, compute local activity

$$a_k = \max\left(1;\ \left(\frac{1}{|B_k^*|}\sum_{(x,y)\in B_k^*} |h(x,y)|\right)^2\right)$$

B_k^* : inner part of block B_k without outer 2 lines/rows for filter support
$|B_k^*|$: number of samples in B_k^*

- Average of a_k: $a_{pic} = \frac{1}{K_{pic}}\sum_{k\in K_{pic}} a_k$ or $a_{pic} = \frac{1}{K_{set}}\sum_{k\in K_{set}} a_k$ for the set of sequences

- Block-wise perceptual sensitivity weight

$$w_k = \left(\frac{a_{pic}}{a_k}\right)^{\beta} \quad \text{with} \quad \beta = \frac{2}{3}$$

β has been optimized using subjective testing

Block size B_k

- Block size $B_k = W_{B_k} \cdot H_{B_k}$ chosen proportionally to picture width W and height H:
- $W_{B_k} = H_{B_k} = \min\left(B_{\max}; 4 \cdot \mathrm{round}\left(16 \cdot \sqrt{\frac{W \cdot H}{1920 \cdot 1080}}\right)\right)$, with upper limit $B_{\max} = 128$
- Each dimension is a multiple of 4 to enable software optimizations like SIMD
- Sizes 64×64 in case of full-HD (1920×1080) and 128×128 in case of UHD or 4K
- Resort to PSNR measure for very small pictures with fewer than 2025 samples
- We only observed a small dependency on block size

Application to LIVE database

- Spearman rank-order correlation with mean opinion scores (MOS) of LIVE database:

Distortion type	PSNR*	SSIM*	MS-SSIM*	FSIM*	VSI*	HaarPSI*	WPSNR (K_{set})
JPEG 2000	**0.8954**	0.9614	0.9627	0.9724	0.9604	0.9684	**0.9562**
JPEG	**0.8809**	0.9764	0.9815	0.9840	0.9761	0.9832	**0.9685**

- Algorithmic complexity of WPSNR lower than that of MS-SSIM

* data taken from [R. Reisenhofer et al., "A Haar Wavelet-Based Perceptual Similarity Index...," 2017. www.haarpsi.org]

Summary and Suggestion

- **WPSNR**: generalization of PSNR, called **perceptually weighted PSNR**
- **WPSNR** is matched to **block-wise QP adaptation (QPA)**
- Combination of **WPSNR** and **QPA** yields:
  - **Perceptual QPA** leads to visual improvements and also to **BD-WPSNR** gains
  - JEM shows similar gains for BD-PSNR and BD-WPSNR
- **Suggestion to change JVET experimental conditions**:

Constant QP / pic
PSNR

→

Perceptual QPA
WPSNR