# USING AN EXPONENTIAL POWER MODEL FOR WYNER ZIV VIDEO CODING

*Thomas Maugey[1], Jérôme Gauthier[1], Béatrice Pesquet-Popescu[1] and Christine Guillemot[2]*

[1] Telecom ParisTech,
Signal and Image Processing Department,
46 rue Barrault, 75634 Paris Cedex 13, France
{ maugey,gauthier,pesquet}@telecom-paristech.fr

[2] IRISA
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France
Christine.Guillemot@inria.fr

## ABSTRACT

The Laplacian model is the standard distribution for correlation noise estimation at the turbodecoder in Wyner-Ziv coding schemes. In practice, this hypothesis is not always satisfied and, regularly, the estimated model sensibly differs from the error distribution. In this work, we prove that using a model better fitted to the true distribution improves the performances, and we thus propose to use the more general exponential power distribution ($\mathcal{EPD}$) which has never been tested in a distributed video coding context. Gains in rate-distortion over the Laplacian model are illustrated by results on several video sequences, showing that the $\mathcal{EPD}$ model outperforms the Laplacian one in off-line (oracle) as well as in on-line (practical implementation) modes. These results also indicate that, in some cases, the on-line $\mathcal{EPD}$ model reduces the bitrate even over the off-line Laplacian model.

*Index Terms*— Wyner Ziv, distributed video coding, Laplacian model, Generalized Gaussian, side information

## 1. INTRODUCTION

Distributed video coding (DVC) is a quite recent paradigm with numerous applications. This information theoretic framework was shown to enable video transmission techniques with a reduced complexity at the encoder without affecting the global rate distortion performances. This is an important feature for low power systems such as cell phones, video surveillance or sensor networks. DVC is based on a fundamental result of information theory from Slepian and Wolf [1], later extended by Wyner and Ziv [2]. They proved that two correlated sources can be independently encoded without affecting the transmission performances, provided that they are jointly decoded. In the particular case of video transmission, the study of frame correlation can be avoided at the encoder side, without reducing the reconstructed visual quality or increasing the transmission rate.

Two main solutions have been proposed by now for DVC: the PRISM [3] and the Stanford one [4]. In our work, we adopt the second approach, which consists in separating the video sequence into two sets of frames: the key frames (KFs) and the Wyner-Ziv frames (WZFs). The KFs are transmitted using an intra codec (H.264 in this scheme) and are used at the decoder to compute an estimation of the WZFs. This estimation is called side information (SI) and is commonly generated through temporal interpolation methods [5]. At the encoder side, the WZFs are DCT transformed, quantized and then turbo encoded. At the decoder, the SI is corrected with the parity bits sent by the WZ encoder. Finally, an inverse DCT is applied on the resulting coefficients. This channel encoding/decoding process is performed under the assumption that the estimation error can be considered as a channel error. The turbo decoding process

needs a model for the correlation noise between the WZF and its SI and the performances of the codec highly depend on the quality of this model. In the literature [6, 7], this correlation noise is usually estimated with a Laplacian model. The parameters of the Laplacian can be estimated "off-line" [6], meaning that the true error between the WZF and its SI is computed and used for parameter estimation. A more realistic solution [8] estimates the parameters "on-line". The error between the WZF and its SI is approximated by a residual, which is the difference between the two motion compensated KFs used for the SI construction.

The Laplacian hypothesis is not always satisfied, as illustrated by the examples in Fig. 1. These phenomena, observed under different conditions, occur very often in practice. To better fit to the shape of a wider range of distributions, a more general model needs to be considered. The Exponential Power Distribution ($\mathcal{EPD}$), sometimes named "Generalized Gaussian", covers a wide range of classical distributions including the Gaussian and the Laplacian ones. The $\mathcal{EPD}$ model was shown to suit well the distribution of wavelet coefficients of signals and images [9]. It later proved to be useful for applications in image denoising [10] or coding [11]. This distribution has also been proposed for modelling DCT coefficients of natural images [12]. So far the $\mathcal{EPD}$ has never been considered in the DVC framework and more precisely at the turbodecoding step to estimate the distribution of the difference between a SI and the corresponding original WZF.

A change of model raises the question to know whether a distribution better fitted to the error leads to an improvement in the rate-distortion (RD) performances. Consequently, after recalling, in Sec. 2, the estimation methods of the Laplacian parameter in the context of state-of-the-art DVC and after describing two classical methods to estimate the parameters of an $\mathcal{EPD}$, we propose in Sec. 3 an original approach for validating if the distance, measured by different distortion metrics, between the estimated distribution and the error, is correlated with the RD coding performances. Finally, in Sec. 4, we compare the coding results obtained under a Laplacian or an $\mathcal{EPD}$ model.

## 2. PARAMETER ESTIMATION

### 2.1. Notations

We first introduce some notations. Let $X$ denote the original WZ frame, while $I_b$ and $I_f$ correspond to the backward and forward decoded reference frames. At the decoder, the side information is denoted by $Y$, and we introduce the residual, $R$, defined as the difference between the forward and backward motion compensated frames. If $\mathbf{s} = (x, y)$ is a given pixel and the backward and the forward motion vector fields are denoted by $MV_b$ and $MV_f$, then $Y$
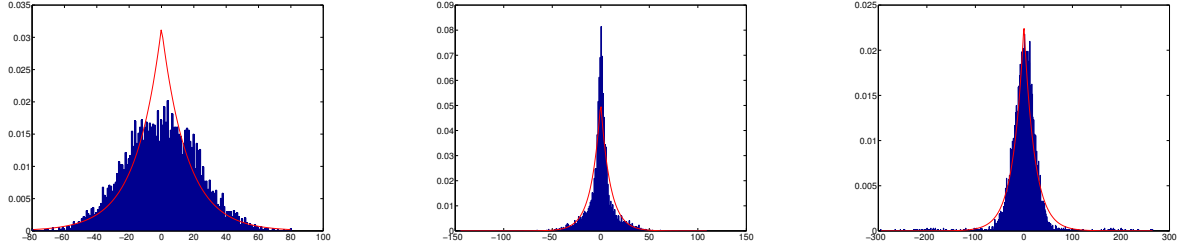
**Fig. 1**. Examples of error distributions and the associated Laplacian model, for different bands and different sequences.

and $R$ can be written as:

$$Y(\mathbf{s}) = \frac{I_b(\mathbf{s} + MV_b(\mathbf{s})) + I_f(\mathbf{s} + MV_f(\mathbf{s}))}{2}, \quad (1)$$

$$R(\mathbf{s}) = \frac{I_b(\mathbf{s} + MV_b(\mathbf{s})) - I_f(\mathbf{s} + MV_f(\mathbf{s}))}{2}. \quad (2)$$

$X$, $Y$ and $R$ are then transformed with a $4 \times 4$ integer DCT, and the $i^{th}$ coefficients in the $k \in [1, .., 16]$ band are denoted respectively by $x_{k,i}$, $y_{k,i}$ and $r_{k,i}$. At the decoder side, the turbo decoder needs an approximation of the probability $p_{x|y}(x)$.

In early papers on DVC [6], the coefficients were estimated off-line, meaning that it was assumed that the coefficients were considered as known at the decoder. In practice, this is unrealistic, since the estimation error $e_{k,i} = x_{k,i} - y_{k,i}$ is computed at the decoder. An on-line solution has been proposed by Brites et al. [7, 8]: the errors $e_{k,i}$ are replaced by the residual coefficients $r_{k,i}$.

As emphasized earlier, the Laplacian pdf is commonly used in the literature to model the correlation between $X$ and $Y$. The underlying assumption is that the correlation depends on the band, but remains stable inside a band. In the two following sections we will explain how to estimate the parameters of the Laplacian and Generalized Gaussian distributions from the error coefficients $\boldsymbol{\xi}_k = (\xi_{k,i})_{1 \leq i \leq N}$ in a given subband (corresponding to $e_{k,i}$ in off-line mode and $r_{k,i}$ in on-line mode).

### 2.2. Laplacian model

Let us first recall the Laplacian probability distribution function (pdf):

$$f_{lap}(x) = \frac{1}{2\alpha} e^{-\frac{|x|}{\alpha}}. \quad (3)$$

Given a set of zero-mean observations $\boldsymbol{\xi}_k$, the parameter $\alpha$ is estimated by first computing an estimation of the second order moment $\mu_2$ of the vector (under the zero-mean assumption, $\mu_2$ is actually the variance). The estimation $\widehat{\alpha}$ is then obtained as: $\widehat{\alpha} = \sqrt{\mu_2/2}$.

As proposed in [7], the on-line estimation of $\alpha$ can be performed at a finer level of precision which consists in calculating another $\alpha$ value for some deficient coefficient inside a band. This approach is used for experimental results in this work.

### 2.3. Exponential power distribution model

The pdf of an $\mathcal{EPD}$ with zero mean and parameters $\alpha \in \mathbb{R}_+^*$ and $\beta \in \mathbb{R}_+^*$ reads

$$f_{gg}(x) = \frac{\beta}{2\alpha \Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x|}{\alpha}\right)^{\beta}},$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the classical "gamma" function. Several methods are available to compute the parameters $(\alpha, \beta)$ of the $\mathcal{EPD}$, among them, we choose to work with the maximum likelihood estimation and the moment estimation.

The idea of the so called "moment method" is to compute the moments of order 2 and 4, $\mu_2$ and $\mu_4$, of the observations $\boldsymbol{\xi}_k$. It can then be shown that: $\frac{\mu_4}{\mu_2^2} = \frac{\Gamma\left(\frac{5}{\beta}\right)\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{3}{\beta}\right)^2} = g(\beta)$ and thus the estimate $(\widehat{\alpha}, \widehat{\beta})$ of $(\alpha, \beta)$ is given by

$$\widehat{\beta} = g^{-1}(\kappa) \quad \text{and} \quad \widehat{\alpha} = \sqrt{\frac{\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{3}{\beta}\right)} \mu_2}. \quad (4)$$

The second method relies on the assumption that the samples in $\boldsymbol{\xi}_k$ are independent and identically distributed, the computation of the maximum of likelihood leads to an estimation of $(\alpha, \beta)$. It can be shown that $\widehat{\beta}$ can be computed as the argmin of the following function:

$$h(\beta) = \frac{1}{\beta} - \ln\left(\frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)}\right) + \frac{1}{\beta} \ln\left(\frac{\beta}{N} \sum_{i=1}^{N} |\xi_{k,i}|^\beta\right)$$

and $\widehat{\alpha}$ is given by: $\widehat{\alpha} = \left(\frac{\widehat{\beta}}{N} \sum_{i=1}^{N} |\xi_{k,i}|^{\widehat{\beta}}\right)^{\frac{1}{\beta}}$.

## 3. APPROACH VALIDATION

As seen in Fig. 1, the Laplacian model is not always fitting the error distribution. In this section, we thus study wether the decoding performances are improved by using a model which better fits the actual noise distribution.

For a band $b$, the error lies between a minimum value, $min$, and a maximum value, $max$. In this range, a model is estimated at the decoder, the obtained function is denoted by $f$. Let $H_b$ be the distribution of the error ( *i.e.*, the histogram of error values). To evaluate the discrepancy between $H_b$ and $f$, many classical measures can be considered. In this work, we have chosen the following family of functions:

$$d^a(f, H_b) = \sum_{n=min}^{max} |f(n) - H_b(n)|^a,$$

where $a \in \mathbb{R}_+^*$. For each band $b$ of a given frame, two models are estimated, $f_1$ and $f_2$. The decoding of this band is performed and the obtained rate is denoted by $r_b^1$ if $f_1$ has been used for the

turbo decoding (respectively $r_b^2$ if $f_2$ has been used for the turbo decoding). We recall that this rate corresponds to the number of bits required to reach a bit error probability lower than $10^{-3}$. Let $a$ be in $\mathbb{R}_+^*$ and let us introduce the following Hypothesis, $Hyp$:

$$\text{For each band, } \forall (i,j) \in [1,2]^2, i \neq j,$$
$$d^a(f_i, H_b) \leq d^a(f_j, H_b) \Leftrightarrow r_b^i \leq r_b^j$$

Minimizing the distance between $H_b$ and $f$ is justified only if $Hyp$ is true. For four CIF test sequences, we test for every band of every frame if $Hyp$ is verified. For the experiments, $f_1$ and $f_2$ correspond to a Laplacian and an $\mathcal{EPD}$ distributions. The obtained results are presented in Tab. 1 for $a \in \{2, 1, \frac{1}{2}, \frac{1}{3}\}$, corresponding to the most representative values among the experimental set.

|  | $d^2$ | $d^1$ | $d^{\frac{1}{2}}$ | $d^{\frac{1}{3}}$ |
|---|---|---|---|---|
| waterfall | 97 | 97 | 97 | 97 |
| foreman | 94 | 91 | 91 | 97 |
| football | 82 | 94 | 94 | 82 |
| mobile | 94 | 85 | 88 | 88 |

**Table 1**. % of measures where $Hyp$ was verified.

The obtained statistics show that there is a strong correlation between the distances $d^a$ and the measured rates. In other words, attempting to fit well the histogram is justified by the fact that it is likely to improve the performances. Based on this idea, in the next section we test the performances of the $\mathcal{EPD}$ distribution.

## 4. EXPERIMENTAL RESULTS

In the previous section we proved that fitting well to the error distribution can improve the coding performances. In this section we test the coding efficiency of using an $\mathcal{EPD}$ instead of the classical Laplacian model employed in the literature.

### 4.1. Experimental setting

The presented experimental results were obtained with the DVC scheme described in the introduction. Tests were run on two CIF video sequences: "City" and "Football" ($352 \times 288$, 30Hz) and one QCIF sequence: Foreman ($176 \times 144$, 15Hz). The 100 first frames (50 KFs, and 50 WZFs) of each sequence are coded, and for each coding configuration, the average rate (in kbps) has been measured. To cover a wide range of rates, the methods have been tested at four quantization levels (Q-Index for the WZFs | Q-Step for H.264 intra coding of the KFs) chosen as follows: 1|42, 4|34, 6|31 and 8|28. Tests are run both for the Laplacian and the $\mathcal{EPD}$ models, with the on-line and off-line coefficient estimation modes. For the $\mathcal{EPD}$ model, the maximum likelihood (ML) and moment (Mom) estimation methods are both employed for "on/off-line" parameter prediction. Results are shown in Tab. 2, presenting the average rate gain (in %). These gains are estimated with the Bjontegaard metric [13]. Additional results are shown in Tabs. 3 and 4, presenting the bitrates obtained by different methods for the four quantization levels on the CIF Football sequence and QCIF Foreman sequence. Finally, Fig. 2 presents the RD results of the different models for the CIF Football sequence. The following notations are used in these tables: "Lapl" stands for Laplacian method and "On", resp. "Off" mean on-line and off-line estimation modes.

### 4.2. Comparison in the off-line setting

We first compare the results of the different methods in the off-line mode. The corresponding results on the test sequences can be read from the first two lines of Tab. 2, 3 and 4 and from the red plots of Fig. 2. We see that on both videos the $\mathcal{EPD}$ model (in ML or Mom case) needs a smaller bitrate than the Laplacian model, with average bitrate gains up to 3.73% for Football (CIF) and 1.78% for Foreman (QCIF). At high bitrate for these two sequences, the transmission rate can be reduced by 194kbps with a CIF video and 44kbps with a QCIF sequence. Another interesting conclusion is that the maximum likelihood estimation performs systematically better than the moment method.

### 4.3. Comparison in the on-line scenario

A second comparison is performed in the on-line mode. The results in this case are reported in the third and fourth lines of Tab. 2. Black plots in Fig. 2 also present the on-line mode results for the Football sequence. Once again, the $\mathcal{EPD}$ model outperforms the Laplacian model. Yet, it is interesting to note that unlike the off-line setting, the moment method yields better results than the ML, meaning that the moment estimation method seems more robust. The bitrate gain reaches 4.3% for the Football (CIF) sequence and 1.88% for the Foreman video (QCIF). In Tabs. 3 and 4 we see that in the on-line mode the $\mathcal{EPD}$ method reduces the transmission rate by 128kbps for a CIF video and by 46kbps for a QCIF sequence when compared with the Laplacian method. This realistic scheme also outperforms H.264 intra coding (7% of rate saving, and 0.35dB of quality improvement for the Football sequence).

### 4.4. Comparison between the off-line and on-line settings

Finally, we compare the results obtained in the off-line and on-line settings. Considering the fifth and sixth lines of Tab. 2, it is worth noting that the loss incurred by switching from off-line to on-line is slightly higher with the Laplacian model.

The last considered case is the comparison between Laplacian off-line and $\mathcal{EPD}$ on-line, with results reported in the last line of Tab. 2 and in Fig. 2. It is interesting to note that the on-line results obtained with $\mathcal{EPD}$ are better than the off-line results with the Laplacian model for the Football and Foreman sequence. In other words, it means that the $\mathcal{EPD}$ model with parameters computed without knowledge of the original WZ performs better than the Laplacian model with parameters estimated with this knowledge. For the City sequence, these rates are close (0.44%) when considering the whole bitrate range. Note that for this last sequence with high bitrates (1600kbps to 4000kbps), we observe that the $\mathcal{EPD}$ on-line performs slightly better (0.75% gain in bitrate) than the Laplacian off-line.

| Method 1 | Method 2 | City | Football | Foreman |
|---|---|---|---|---|
| Lapl Off | $\mathcal{EPD}$ Off ML | -0.96 | -3.73 | -1.78 |
| Lapl Off | $\mathcal{EPD}$ Off Mom | 1.21 | -3.61 | -1.52 |
| Lapl On | $\mathcal{EPD}$ On ML | 0.36 | -3.29 | -0.90 |
| Lapl On | $\mathcal{EPD}$ On Mom | -1.3 | -4.30 | -1.88 |
| Lapl Off | Lapl On | 1.73 | 2.67 | 1.53 |
| $\mathcal{EPD}$ Off ML | $\mathcal{EPD}$ On Mom | 1.4 | 2.10 | 1.39 |
| Lapl Off | $\mathcal{EPD}$ On Mom | 0.44 | -1.64 | -0.38 |

**Table 2**. Rate gains (%) by method 2 over method 1 on City, Football (CIF, 30Hz) and Foreman (QCIF, 15Hz) sequences.
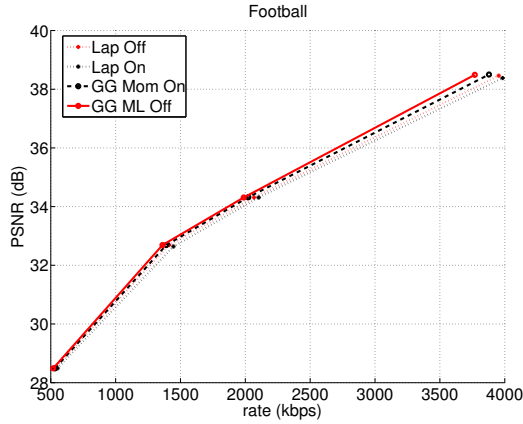
**Fig. 2**. Rate-distortion performances for "Football" sequence, CIF, 100 frames, 30 fps.

### 4.5. Discussion

Knowing that a better fitted distribution enables an improvement of the RD permormances, the purpose of these tests is to measure the reliability of the $\mathcal{EPD}$ model. Experimental results have shown that the $\mathcal{EPD}$ model is finer than the Laplacian one, yielding bitrate improvements on the considered test sequences. Improvements may of course vary from one video to another depending on how close the residual distribution is to a Laplacian one. We also want to emphasize that the gains obtained in this paper can be compared to those offered by other works involving refinements of the noise model [7, 8]. Moreover, another purpose of this work was to propose a realistic model, in the sense that it does not need the knowledge of the original WZ frame. This is precisely what is shown in Sec. 4.3 and 4.4. Indeed, we proposed an efficient on-line solution, which even outperforms the off-line standard technique in some cases.

| PSNR (in dB) | 28.49 | 32.64 | 34.31 | 38.38 |
|---|---|---|---|---|
| Lapl OFF | 531 | 1402 | 2066 | 3916 |
| $\mathcal{EPD}$ OFF MV | 519 | 1351 | 1988 | 3722 |
| $\Delta$rate (kbps) | **-12** | **-51** | **-78** | **-194** |
| Lapl ON | 552 | 1448 | 2103 | 3953 |
| $\mathcal{EPD}$ ON Mom | 532 | 1380 | 2019 | 3825 |
| $\Delta$rate (kbps) | **-20** | **-68** | **-84** | **-128** |

**Table 3**. Rate results (kbps) on the Football sequence (CIF, 30Hz) for different values of average PSNR.

### 5. CONCLUSION

In this paper, we have proposed a refinement of the correlation model in DVC, by replacing the Laplacian distribution with an $\mathcal{EPD}$, never used in the WZ context. We proved that fitting well to the correlation noise was directly correlated with a decreasing of the coding rate. Experiments have confirmed that the more general $\mathcal{EPD}$ was a finer model which improves the global rate-distortion performances.

| PSNR (in dB) | 31.36 | 34.4 | 36.44 | 39.94 |
|---|---|---|---|---|
| Lapl OFF | 225 | 424 | 624 | 1055 |
| $\mathcal{EPD}$ OFF MV | 224 | 421 | 611 | 1009 |
| $\Delta$rate (kbps) | **-1** | **-3** | **-13** | **-44** |
| Lapl ON | 227 | 432 | 632 | 1080 |
| $\mathcal{EPD}$ ON Mom | 226 | 425 | 622 | 1034 |
| $\Delta$rate (kbps) | **-1** | **-7** | **-10** | **-46** |

**Table 4**. Rate results (kbps) on the Foreman sequence (QCIF, 15Hz) for different values of average PSNR.

### 6. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[3] R. Puri and K. Ramchandran, "PRISM: A video coding architecture based on distributed compression principles," Tech. Rep. UCB/ERL M03/6, EECS Department, University of California, Berkeley, 2003.

[4] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[5] M. Cagnazzo, T. Maugey, and B. Pesquet-Popescu, "A differential motion estimation method for image interpolation in distributed video coding," in *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, Taipei, Taiwan, Apr. 2009, pp. 1861–1864.

[6] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals, Systems and Computers*, Nov. 2002, vol. 1, pp. 240–244.

[7] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner–Ziv video coding," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 18, no. 9, pp. 1177–1190, Sep. 2008.

[8] C. Brites, J. Ascenso, and F. Pereira, "Studying temporal correlation noise modeling for pixel based Wyner-Ziv video coding," in *Proc. Int. Conf. on Image Processing*, Atlanta, GA, USA, Oct. 2006, pp. 273–276.

[9] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. on Pattern Anal. and Match. Int.*, vol. 11, no. 7, pp. 674–693, July 1989.

[10] A. Antoniadis, D. Leporini, and J.-C. Pesquet, "Wavelet thresholding for some classes of non-Gaussian noise," *Statistica neerlandica*, vol. 56, no. 4, pp. 434–453, 2002.

[11] R.W. Buccigrossi and E.P. Simoncelli, "Progressive wavelet image coding based on a conditional probability model," in *Proc. Int. Conf. on Acoust., Speech and Sig. Proc.*, Munich, Ger., Apr. 1997, vol. 4, pp. 2957–2960.

[12] F. Müller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electronics Letters*, vol. 29, no. 22, pp. 1935–1936, Oct. 1993.

[13] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," Tech. Rep., 13th VCEG-M33 Meeting, Austin, TX, USA, Apr. 2001.