Coding efficiency of the DCT and DST in hybrid video coding

Matthias Narroschke

***Abstract*—Standardized hybrid video coding algorithms, e.g. HEVC, apply intra frame prediction or motion compensated prediction and subsequent integer approximated DCT or DST transform coding. The coding efficiency of the transforms depends on the statistical moments and probability distribution of the input signals. For a Gaussian distribution, the DCT always leads to a data rate reduction. However, for Laplacian distributed prediction errors, the transforms sometimes increase the data rate. This paper presents a theoretical analysis, which explains the reason for an increase of the data rate, which is due to the generation of higher statistical moments of the coefficients by the DCT or DST in the case of Laplacian distributed input signals. The data rate can increase by up to 0.10 bit per sample for blocks with low correlation. For screen content, in about 20% of the blocks, the transform increases the data rate. By skipping the transform for these blocks, HEVC achieves a 7% data rate reduction.**

***Index Terms*—HEVC, DCT, DST, transform skip, Laplace**

I. Introduction

STANDARDIZED video coding algorithms, such as H.262/MPEG-2 [1], MPEG-4 Visual [2], and H.264/MPEG-4 AVC [3][4], as well as HEVC [5][6], apply hybrid coding. It consists of the two basic steps, prediction and subsequent prediction error coding. In order to exploit statistical dependencies in the temporal direction between consecutive frames of a video, motion compensated prediction [7][8] is applied. Each frame to be coded is partitioned into blocks. For each block, a displacement vector is estimated describing the displacement of the block between the current frame and an already decoded frame. In order to exploit statistical dependencies in the spatial direction between consecutive samples, so called intra prediction is applied block-wise[3][4][9]. The prediction error resulting from both, motion compensated prediction and intra prediction, is transform coded exploiting correlation in spatial direction. For the purpose of transform coding, primarily a discrete cosine transform [10], DCT, or integer approximations thereof [11], are applied followed by quantization of the coefficients. In the case of intra prediction and a transform block size of 4×4 samples, HEVC applies an integer approximation of the discrete sine transform [12], DST. The block partitioning, the displacement vectors, and the quantized coefficients are coded and transmitted to the receiver.

MPEG-4 Visual reduces the bit rate by 20% at the same mean squared error compared to H.262/MPEG-2 [13], achieved by a more accurate motion compensated prediction. Displacement vectors of 1/4 pel accuracy in combination with an 8-tap Wiener interpolation filter are used in MPEG-4 Visual instead of displacement vectors of 1/2-pel accuracy combined with bilinear interpolation [14]. Compared to MPEG-4 Visual, H.264/MPEG-4 AVC further reduces the data rate by 44% [13], again achieved by increased accuracy of the motion compensated prediction: Key techniques are the multi-frame prediction [15], generalized B pictures [16] and a more flexible block partitioning. HEVC applies an even more flexible block partitioning [17][18] and a more efficient displacement vector coding compared to H.264/MPEG-4 AVC and further reduces the bit rate by another 35% [13].

The result of a more accurate prediction are decreasing statistical dependencies between prediction error samples. For the case of motion compensated prediction, this is mathematically described in [8]. The power density spectrum $S_{ee}(\omega_x, \omega_y)$ of the prediction error signal e is dependent on the power density spectrum $S_{ss}(\omega_x, \omega_y)$ of the video input signal s and the power density spectrum $S_{qq}(\omega_x, \omega_y)$ of the quantization noise q:

$$S_{ee}(\omega_x, \omega_y) = 2 \cdot S_{ss}(\omega_x, \omega_y) \cdot (1 - \mathrm{Re}\{\mathcal{F}\{p(\Delta_x, \Delta_y)\}\}) + S_{qq}(\omega_x, \omega_y) \quad (1)$$

Hereby, $\mathcal{F}\{p(\Delta_x, \Delta_y)\}$ is the two-dimensional Fourier transform of the probability density function $p(\Delta_x, \Delta_y)$ of the displacement estimation errors Δ_x and Δ_y. Since the power density spectrum is the Fourier transform of the correlation function, it can be derived from equation 1 that the correlation between the prediction error samples decreases with an increased accuracy of the motion compensated prediction. In addition, it can be seen that the correlation also decreases with a decreasing correlation between the samples of the input signal to be coded. This typically happens with improvements to the cameras aquiring the input signals. To illustrate this, average correlation coefficients $\bar{\rho}_s(\tau)$ between horizontally adjacent samples of the distance τ are measured for various camera aquired HD test sequences of 1080 lines, which have been used in the international standardization activities of the ISO/IEC and the ITU-T in the year 2006 and in the year 2011. These are shown in Fig. 1. In addition to camera aquired content, computer generated screen content, such as web pages, graphics, and presentation slides, is considered more often in video coding applications [19][20].

Manuscript received January 30, 2013.

M. Narroschke is with the Panasonic R&D Center Germany, Germany.

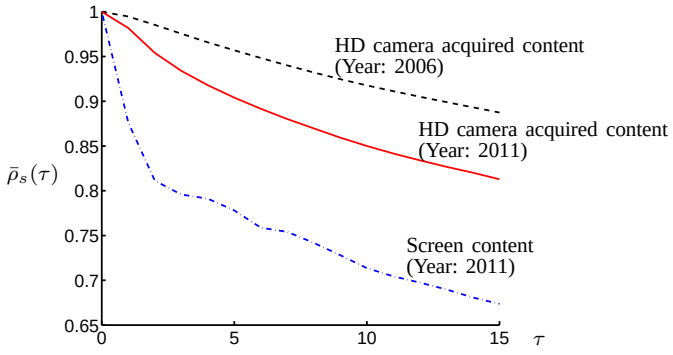

Fig. 1. Average correlation coefficients $\bar{\rho}_s(\tau)$ measured between horizontally adjacent samples of the distance τ on the basis of test sequences used in the international standardization activities of the ISO/IEC and the ITU-T. Sequences for the year 2006: Rolling tomatoes, Dinner table, Freeway, Card toss. Sequences for the year 2011: BQTerrace, Cactus, Kimono1, Park scene. Screen content sequences: Slide editing, China speed, Basketball drill text, Slide show [21].

For this type of content, the spatial correlation between the samples is typically lower than for camera aquired HD content as can also be seen in Fig. 1.

Figure 2 shows the probability distribution $P_e(e)$ of the prediction error, which is generated by the HEVC reference model HM8 [22] while encoding the test sequences used in the HEVC standardization activities of the ISO/IEC and the ITU-T [21]. It can be seen that it is close to a Laplacian distribution $P_L(e)$, which is also shown in Figure 2 for the same variance $\sigma_e^2 = 26$.

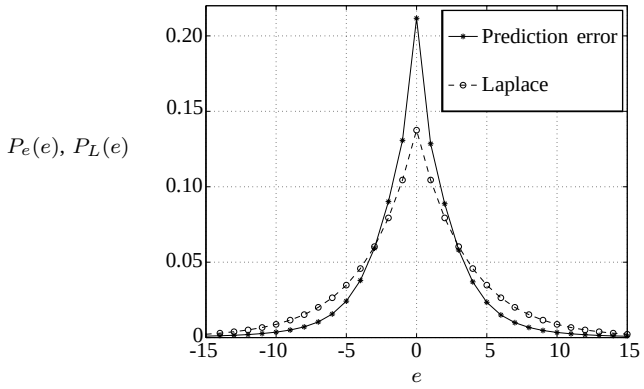

Fig. 2. Distribution of the prediction error generated by the HEVC reference model HM8 [22] using the Random Access configuraton at Main Profile and QP=22 [21], and Laplacian distribution at the same variance.

Since ongoing improvements to both, motion compensated prediction and cameras, continuously lower the correlation between prediction error samples, the general question arises if the use of the transform is efficient for the coding of prediction errors of low correlation. In [23] and [24], an adaptive prediction error coding in the spatial and frequency domain is described, which applies the DCT only for some parts of the prediction error. The samples of the remaining parts are quantized and coded without the use of the DCT. This technique had been proposed and adopted in the context of the Key Technical Area investigations of the ITU-T [25][26][27] performed prior the HEVC standardization. Bit rate reductions of around 2-5% were achieved at the same mean squared error. This adaptive prediction error coding has been proposed denoted as *Transform skip* and has been adopted into HEVC. It leads to bit rate reductions of around 7% for screen content. For camera acquired content, the bit rate reduction is sequence dependent up to 1%. Detailed experimental results for the bit rate reduction achieved by *Transform skip* are documented in the corresponding proposals to the HEVC standardization [28][29]. Further adjustments of the coding of the quantized samples to their statistics, which are not adopted into HEVC, can further increase the coding efficiency for both kinds of content [24][30][31]. The coding efficiency increase achieved by skipping the transform might be explained by the answer to the raised question.

Motivated by this question, this paper presents a generalized theoretical analysis of the efficiency of the transforms for the coding of prediction errors. For the analysis, the prediction errors are modelled as Laplacian distributed sources with various statistical dependencies between the input samples. The analysis is primarily performed on the basis of the DCT and supplemented for the DST in places. For completion, it is investigated if the coding efficiency increase achieved by the adaptive prediction error coding of HEVC can be explained by the analysis results.

Section II describes the criterion for the assessment of the efficiency of the transforms. In Section III, the efficiency of the transforms is analyzed with respect to this criterion. In Section IV, the HEVC prediction error coding is investigated with respect to the analysis results. The paper closes with the conclusion.

II. Criterion for the Coding Efficiency of the Transform

In this section, the efficiency of the transform is investigated in dependence on the source statistics of the input signal. For this investigation, a one-dimensional, time-discrete, value-continuous, and stationary signal e of zero mean is considered. The variance of this signal is σ_e^2. The probability density function of e as well as the statistical dependencies between N successive samples $e_1, \ldots, e_N$ of e are described by the joint probability density function $p_{\vec{e}}(e_1, \ldots, e_N)$. All marginal probability density functions are identical due to the stationarity of e. They are denoted as $p_e(e)$. The correlation coefficient of two successive samples of e is ρ_e.

For the assessment of the efficiency of the transform for the coding of the signal e, the data rate of a coding using the transform is compared to the data rate of a coding without using the transform at the same mean squared quantization error. In the following, the coding using the transform is also denoted by a coding in the frequency domain since it generates a spectral representation of the input signal. The coding without using the transform is also denoted by a coding in the spatial domain. The determination of the data rates is performed for both codings under the assumption of a memoryless source.

A. Data rate of a coding in the spatial domain

Fig. 3 shows the block diagram of a coding in the spatial domain.

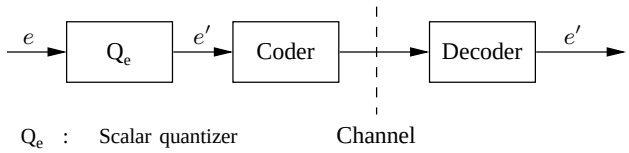

Fig. 3. Block diagram of a coding in the spatial domain.

The samples of the input signal e are quantized by the scalar quantizer $\mathrm{Q_e}$. Each sample is quantized independently from all other samples. The resulting quantized samples are e'. The variance of the quantization error $q_e = e - e'$ is $\sigma^2_{q_e}$. The quantization error is assumed to be of zero mean. The quantized samples e' are coded and transmitted via the channel. The coding is assumed to be lossless. A decoding is performed to retrieve the quantized samples e'. The minimum data rate required to code the quantized samples e' is the entropy $H(e')$. This entropy is the difference of the differential entropy $H(e)$ of the samples e and the differential entropy $H(q_e)$ of the quantization error q_e, which originates during the quantization $\mathrm{Q_e}$ [32]:

$$H(e') = H(e) - H(q_e). \tag{2}$$

The differential entropies $H(e)$ and $H(q_e)$ can be calculated by the use of the corresponding probability density functions $p_e(e)$ and $p_{q_e}(q)$:

$$H(e) = -\int_{-\infty}^{\infty} p_e(e) \cdot \log_2 p_e(e) \quad de \tag{3}$$

$$H(q_e) = -\int_{-\infty}^{\infty} p_{q_e}(q) \cdot \log_2 p_{q_e}(q) \quad dq. \tag{4}$$

As a consequence of the assumption of a memoryless source, the data rate of the coding in the spatial domain is only dependent on the probability density functions $p_e(e)$ and $p_{q_e}(q)$.

B. Data rate of a transform-coding in the frequency domain

Fig. 4 shows the block diagram of a coding in the frequency domain using a transform.

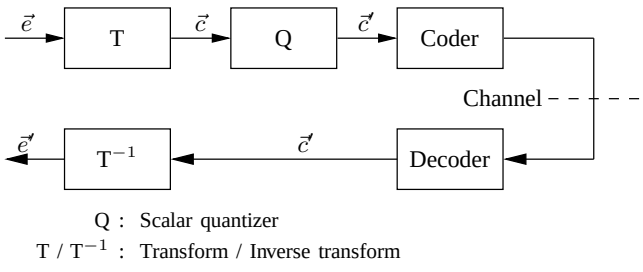

Fig. 4. Block diagram of a coding in the frequency domain with the use of a transform.

The input signal of the transform is a column vector $\vec{e} = (e_1, \ldots, e_N)^T$ of N successive samples $e_1, \ldots, e_N$ of the signal e. The result of a multiplication of the column vector $\vec{e}$ and the transform matrix $\mathbf{T}$ is a column vector of N coefficients:

$$\vec{c} = (c_1, \ldots, c_N)^T = \mathbf{T} \cdot \vec{e}. \tag{5}$$

Each coefficient c_i can be interpreted as a function of the elements of the input vector:

$$c_i = f_i(e_1, \ldots, e_N), \qquad \forall \;\; i = 1, \ldots, N. \tag{6}$$

These functions are given by the row vectors of the transform matrix $\mathbf{T}$. The N coefficients are quantized independently from each other by the use of a scalar quantizer Q. The N resulting quantized coefficients are denoted as $\vec{c}' = (c'_1, \ldots, c'_N)^T$. Due to the quantization, the N quantization errors $\vec{q}_c = (q_{c_1}, \ldots, q_{c_N})^T = \vec{c} - \vec{c}'$ are generated. They are assumed to be of zero mean. The variances are denoted as $\sigma^2_{q_{c_i}}$. The N quantized coefficients are coded losslessly and independently from each other. By decoding, the column vector $\vec{c}' = (c'_1, \ldots, c'_N)^T$ of N quantized coefficients is reconstructed. The subsequent inverse transform generates a column vector of N reconstructed samples:

$$\vec{e}' = (e'_1, \ldots, e'_N)^T = \mathbf{T}^{-1} \cdot \vec{c}'. \tag{7}$$

The difference $\vec{r} = \vec{e} - \vec{e}'$ is called reconstruction error. In the case of an orthonormal transform as the DCT and DST, the variance σ_r^2 of the reconstruction error equals the mean variance of the quantization errors that are generated by the quantization of the coefficients[32]:

$$\begin{aligned} \sigma_r^2 &= \tfrac{1}{N} E\left[(\vec{e} - \vec{e}')^T (\vec{e} - \vec{e}')\right] = \tfrac{1}{N} E\left[(\vec{c} - \vec{c}')^T (\vec{c} - \vec{c}')\right] \\ &= \tfrac{1}{N} \cdot \sum_{i=1}^{N} \sigma^2_{q_{c_i}}. \end{aligned} \tag{8}$$

Hereby, $E[\cdot]$ denotes the expected value. The minimum data rate required for the coding of the quantized coefficients is given by

$$H(c') = \frac{1}{N} \cdot \sum_{i=1}^{N} \Big(H(c_i) - H(q_{c_i}) \Big). \tag{9}$$

In the above equation 9, $H(c_i)$ is the differential entropy of the coefficient c_i and $H(q_{c_i})$ is the differential entropy of the corresponding quantization error q_{c_i}. The differential entropies $H(c_i)$ result from the probability density functions $p_{c_i}(c)$ of the coefficients:

$$H(c_i) = -\int_{-\infty}^{\infty} p_{c_i}(c) \cdot \log_2 p_{c_i}(c) \quad dc \tag{10}$$

Respectively, the differential entropies $H(q_{c_i})$ results result from the probability density functions $p_{q_{c_i}(q)}$ of the quantization errors:

$$H(q_{c_i}) = -\int_{-\infty}^{\infty} p_{q_{c_i}}(q) \cdot \log_2 p_{q_{c_i}}(q) \quad dq. \tag{11}$$

In the following, the term

$$H(c) = \frac{1}{N} \cdot \sum_{i=1}^{N} H(c_i) \tag{12}$$

is also denoted as the mean differential entropy of the coefficients.

The probability density functions $p_{c_i}(c)$ of the N coefficients are dependent on the joint probability density function of the input signal e and on the transform matrix. The joint probability density function $p_{\vec{c}}(c_1, \ldots, c_N)$ of the N coefficients, which have been generated by the transform, can be determined by the use of the joint probability density function $p_{\vec{e}}(e_1, \ldots, e_N)$ of the input signal e and the functions $f_1(e_1, \ldots, e_N)$ until $f_N(e_1, \ldots, e_N)$ which are defined by the row vectors of the transform according to equation 6 [33]:

$$\begin{aligned} p_{\vec{c}}(c_1, \ldots, c_N) = & \int_{-\infty}^{-\infty} \ldots \int_{-\infty}^{-\infty} p_{\vec{e}}(e_1, \ldots, e_N) \cdot \\ & \prod_{i=1}^{N} \delta_0 \left(c_i - f_i(e_1, \ldots, e_N) \right) \quad de_1 \ldots de_N, \\ & \text{with} \quad \delta_0(k) = \begin{cases} 1 & ; \quad k = 0 \\ 0 & ; \quad k \neq 0. \end{cases} \end{aligned} \tag{13}$$

The probability density functions $p_{c_i}(c)$ of the N coefficients c_i are the marginal probability density functions of the joint probability density function $p_{\vec{c}}(c_1, \ldots, c_N)$, which is given in the previous equation 13:

$$p_{c_i}(c_i) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p_{\vec{c}}(c_1, \ldots, c_N) \; dc_1 \; \ldots \; dc_{i-1} \; dc_{i+1} \; \ldots \; dc_N. \tag{14}$$

According to equation 13, the probability density functions of the coefficients are dependent on the joint probability density function of the input signal e. As a consequence, the data rate of the coding of the coefficients is also dependent on the joint probability density function of the input signal e.

C. Coding efficiency of the transform

For the assessment of the efficiency of the transform coding in the frequency domain, the difference ΔH between $H(e')$, see equation 2, and $H(c')$, see equation 9, is calculated:

$$\Delta H = H(e') - H(c') =$$

$$\left(H(e) - \frac{1}{N} \sum_{i=1}^{N} H(c_i) \right) + \left(\frac{1}{N} \sum_{i=1}^{N} H(q_{c_i}) - H(q_e) \right). \tag{15}$$

Based on the explanations above it is observable that ΔH and thus the efficiency of the transform is dependent on the following three terms:

1) Joint probability density function $p_{\vec{e}}(e_1, \ldots, e_N)$ of the input samples e
2) Probability density function $p_{q_e}(q)$ of the quantization error q_e
3) Probability density functions $p_{q_{c_i}}(q)$ of the quantization errors q_{c_i}.

In the following, uniform quantization of the samples e and of the coefficients c_i with a sufficiently fine quantization step size of Δ_{fine} is assumed. In this case, the quantization errors q_e and q_{c_i} are uniformly distributed. Their variances are $\sigma^2_{q_e} = \sigma^2_{q_{c_i}} = \Delta^2_{fine}/12$ [32]. In this case, the differential entropies are given by $H(q_e) = H(q_{c_i}) = -\log_2 \Delta_{fine}$. Under this assumption, the difference ΔH according to 15 simplifies to

$$\Delta H = H(e) - \frac{1}{N} \sum_{i=1}^{N} H(c_i). \tag{16}$$

Consequently, the use of a transform for the coding of the signal e leads to a reduction of the data rate as long as $\Delta H \geq 0$. In this case, the transform is efficient. For $\Delta H < 0$, the use of a transform leads to an increased data rate. In this case, the transform is not efficient.

III. Coding efficiency of the transform for Laplacian distributed sources

In this section, ΔH is calculated for Laplacian distributed sources, whereby the calculation is performed analytically in dependency on the statistical dependencies between the input samples e. In a first step, the DCT is analyzed.

A. Coding efficiency of the DCT of a block size two

To simplify calculations, a one-dimensional DCT of a block size two with the transform matrix

$$\mathbf{T}_{DCT,2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{17}$$

is considered in a first instance. This DCT generates the two coefficients c_1 and c_2. The variances $\sigma^2_{c_1}$ and $\sigma^2_{c_2}$ of these two coefficients are:

$$\begin{aligned} \sigma^2_{c_1} = E[c_1^2] = \frac{1}{2} \cdot E\left[\; (e_1 + e_2)^2 \; \right] = \sigma_e^2 \cdot (1 + \rho_e) \\ \sigma^2_{c_2} = E[c_2^2] = \frac{1}{2} \cdot E\left[\; (e_1 - e_2)^2 \; \right] = \sigma_e^2 \cdot (1 - \rho_e). \end{aligned} \tag{18}$$

Each of the two coefficients has a variance which depends on the variance σ_e^2 of the input signal and on the correlation coefficient ρ_e. The calculation of the correlation $E[c_1 \cdot c_2]$ shows that the two coefficients c_1 und c_2 are uncorrelated:

$$E[c_1 \cdot c_2] = \frac{1}{2} \cdot E\left[\; (e_1 + e_2) \cdot (e_1 - e_2) \; \right] = 0. \tag{19}$$

In the case of a one-dimensional signal e of zero mean with a Laplacian probability density function:

$$p_{e,L}(e) = \frac{1}{\sqrt{2}\cdot\sigma_e}\cdot \mathrm{e}^{-\frac{\sqrt{2}\cdot|e|}{\sigma_e}}. \tag{20}$$

The statistical dependencies between N samples of e can be fully described by the joint probability density function $p_{\vec{e},L}(e_1,\ldots,e_N)$. In the following, three joint probability density functions $p_{\vec{e},L1}$, $p_{\vec{e},L2}$, and $p_{\vec{e},L3}$ with Laplacian marginal probability density functions are considered. In the Literature, there are only a few mathematical descriptions of such joint probability density functions. In [34], the following description is given for the case of $N = 2$:

$$p_{\vec{e},L1}(e_1,e_2,\rho) = \frac{K_0\left(\frac{2\cdot(e_1^2-2\cdot\rho\cdot e_1\cdot e_2+e_2^2)}{\sigma_e^2\cdot(1-\rho^2)}\right)}{\pi\cdot\sigma_e^2\cdot(1-\rho^2)}, \tag{21}$$

with

$$K_0(k) = \frac{1}{2}\cdot\int_0^\infty \frac{1}{t}\cdot \mathrm{e}^{-t-\frac{k^2}{4t}}\ dt,\quad k>0. \tag{22}$$

Hereby, $K_0(k)$ is the modified Bessel function of the third kind [34]. In this description, the parameter ρ equals the correlation coefficient $\rho = \rho_e$. In the case of $\rho_e = 0$, the joint probability density function according to equation 21 is not equal to the joint probability density function of a memoryless Laplacian distributed source having no statistical dependencies between the samples:

$$p_{\vec{e},L2}(e_1,e_2) = p_{e,L}(e_1)\cdot p_{e,L}(e_2) = \frac{1}{2\cdot\sigma_e^2}\cdot \mathrm{e}^{-\frac{\sqrt{2}}{\sigma_e}\cdot\left(|e_1|+|e_2|\right)}. \tag{23}$$

However, for the joint probability density function according to equation 23, the correlation coefficient is also zero: $\rho_e = 0$. Fig. 5 illustrates the two joint probability density functions $p_{\vec{e},L1}(e_1,e_2,\rho = \rho_e = 0)$ and $p_{\vec{e},L2}(e_1,e_2)$ according to equations 21 and 23 respectively.

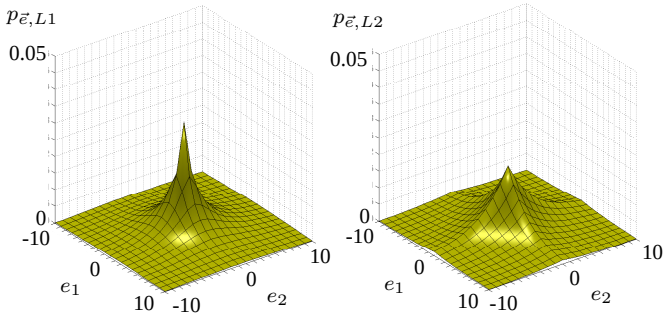

Fig. 5. Illustration of the joint probability density functions $p_{\vec{e},L1}(e_1,e_2,\rho=0)$ according to equation 21 and $p_{\vec{e},L2}(e_1,e_2)$ according to equation 23 for the case of uncorrelation and a variance of $\sigma_e^2 = 26$.

Both joint probability density functions have the same Laplacian marginal distributions and the same correlation coefficient $\rho_e = 0$. They differ in the joint moments of higher order.

Since there are no other joint probability density functions known from the literature, which have Laplacian marginal probability density functions, linear combinations of the two joint probability density functions according to equations 21 and 23 are considered in addition:

$$p_{\vec{e},L3}(e_1,e_2,\rho,a) =$$

$$a\cdot p_{\vec{e},L2}(e_1,e_2) + (1-a)\cdot p_{\vec{e},L1}(e_1,e_2,\rho), \tag{24}$$

$$\text{with}\quad 0\le a\le 1.$$

The marginal probability density functions of these linear combinations are Laplacian. The correlation coefficient of a linear combination can be calculated by

$$\rho_e = \frac{1}{\sigma_e^2}\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e_1\cdot e_2\cdot p_{\vec{e},L3}(e_1,e_2,a,\rho)\ \ de_1\ de_2. \tag{25}$$

Fig. 6 shows two joint probability density functions with the same example correlation coefficient $\rho_e = 0.61$.

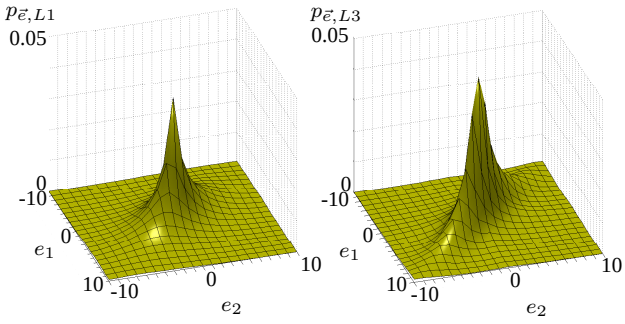

Fig. 6. Illustration of the joint probability density functions $p_{\vec{e},L1}(e_1,e_2,\rho=0.61)$ according to equation 21 and $p_{\vec{e},L3}(e_1,e_2,a=0.35,\rho=0.94)$ according to equation 24, which have the same correlation coefficient $\rho_e = 0.61$. The variance is $\sigma_e^2 = 26$.

Fig. 7 illustrates $\Delta H_{Laplace}$ for the three joint probability density functions according to equations 21, 23, and 24 as a function of the correlation coefficient ρ_e. It can be seen that $\Delta H_{Laplace}$ is always positive for the case of the joint probability density function $p_{\vec{e},L1}(e_1,e_2,\rho)$. Consequently, the DCT is efficient in this case. In the case of the joint probability density function $p_{\vec{e},L2}(e_1,e_2)$, $\Delta H_{Laplace}$ is negative. This shows that in the case of statistically independent samples with a Laplacian probability density function, the DCT is not efficient. In the case of the linear combination $p_{\vec{e},L3}(e_1,e_2,\rho,a)$, the data rate is either reduced or increased by the use of the DCT, dependent on the correlation coefficient ρ_e. It can be summarized that the DCT-coding in the frequency domain can increase the data rate if the correlation is low. In these situations, the DCT is not efficient.

From Fig. 7, it can be drawn that dependent on the joint Laplacian probability density function the efficiency of the DCT can vary at the same correlation. The joint probability density functions with Laplacian marginal probability density functions differ at the same correlation due to different joint moments of higher order. Thus, the efficiency of the DCT has to be dependent on these joint moments of higher order.

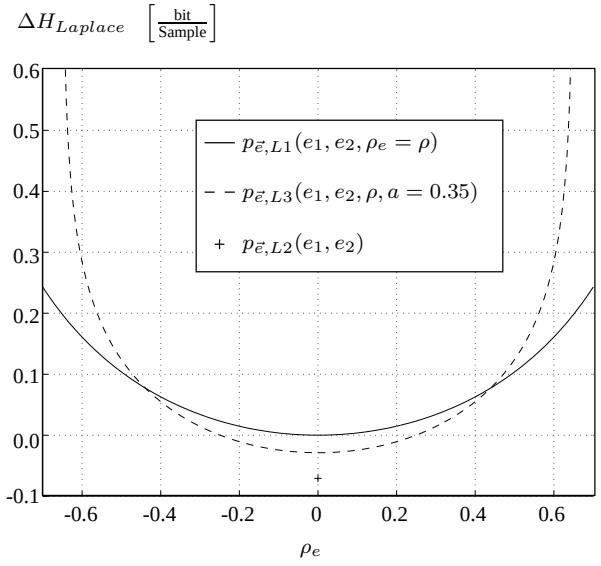

Fig. 7. Illustration of the difference $\Delta H_{Laplace} = H(e') - H(c')$ of the data rates of two codings under the assuption of memoryless sources as a function of the correlation coefficient ρ_e. $H(e')$ is the minimum data rate required to code the source e with a Laplacian probability density function. $H(c')$ is the minimum data rate required to code the coefficients generated by the application of a DCT of a block size two to the source e.

In order to understand this, the joint moment of higher order

$$E\left[c_1^2 \cdot c_2^2\right] = \frac{1}{4} \cdot E\left[(e_1+e_2)^2 \cdot (e_1-e_2)^2\right] = \frac{1}{2} \cdot E\left[e^4\right] - \frac{1}{2} \cdot E\left[e_1^2 \cdot e_2^2\right] \tag{26}$$

of the coefficients c_1 and c_2 is considered. In the case that the joint moment of equation 26 and the product of the single moments

$$E\left[c_1^2\right] \cdot E\left[c_2^2\right] = \frac{1}{4} \cdot E\left[(e_1+e_2)^2\right] \cdot E\left[(e_1-e_2)^2\right] = \sigma_e^4 - E^2\left[e_1 \cdot e_2\right] \tag{27}$$

differ, the coefficients c_1 and c_2 are statistically dependent. The difference is

$$E\left[c_1^2 \cdot c_2^2\right] - E\left[c_1^2\right] \cdot E\left[c_2^2\right] = \frac{1}{2} \cdot E\left[e^4\right] - \sigma_e^4 - \frac{1}{2} \cdot E\left[e_1^2 \cdot e_2^2\right] + E^2\left[e_1 \cdot e_2\right]. \tag{28}$$

The term $E[e^4]$ depends on the probability density function of the signal e. For sources with a Laplacian probability density function, the term can be calculated to $E[e^4] = 6 \cdot \sigma_e^4$. Thus, the difference according to equation 28 results in

$$E\left[c_1^2 \cdot c_2^2\right] - E\left[c_1^2\right] \cdot E\left[c_2^2\right] = 2\sigma_e^4 - \frac{1}{2} \cdot E\left[e_1^2 \cdot e_2^2\right] + E^2\left[e_1 \cdot e_2\right]. \tag{29}$$

In the following, a source with the specific joint probability density function $p_{\vec{e},L2}(e_1, e_2)$ is considered, whose samples are statistically independent and of zero mean. For this source, it is $E\left[e_1^2 \cdot e_2^2\right] = \sigma_e^4$ and $E\left[e_1 \cdot e_2\right] = 0$. Equation 29 results in

$$E\left[c_1^2 \cdot c_2^2\right] - E\left[c_1^2\right] \cdot E\left[c_2^2\right] \;=\; \frac{3}{2} \cdot \sigma_e^4 \;\neq\; 0. \tag{30}$$

Consequently, in the case of statistically independent samples e with a Laplacian probability density function, the DCT generates coefficients which are uncorrelated according to equation 19 but statistically dependent according to equation 30.

According to the characteristic function, the generated statistical dependencies between the coefficients c_1 and c_2 influence the probability density functions of the coefficients. The calculation of the probability density functions $p_{c_1}(c)$ and $p_{c_2}(c)$ by the use of the equations 13 and 14 results in

$$p_{c_1}(c) = p_{c_2}(c) = \frac{\sigma_e + 2 \cdot \mid c \mid}{2 \cdot \sigma_e^2} \cdot \mathrm{e}^{\frac{-2 \cdot |c|}{\sigma_e}}. \tag{31}$$

Both coefficients c_1 and c_2 have the same probability density function. However, equation 31 is not a Laplacian probability density function. For the considered case of $\rho_e = 0$ the variance of each coefficient c_1 and c_2 equals the variance of the input signal e according to equation 18: $\sigma_{c_1}^2 = \sigma_{c_2}^2 = \sigma_e^2$. In Fig. 8 the probability density function p_e of the samples e as well as the probability density functions p_{c_1} and p_{c_2} of the coefficients c_1 and c_2 are shown for an example variance of $\sigma_e^2 = 26$.

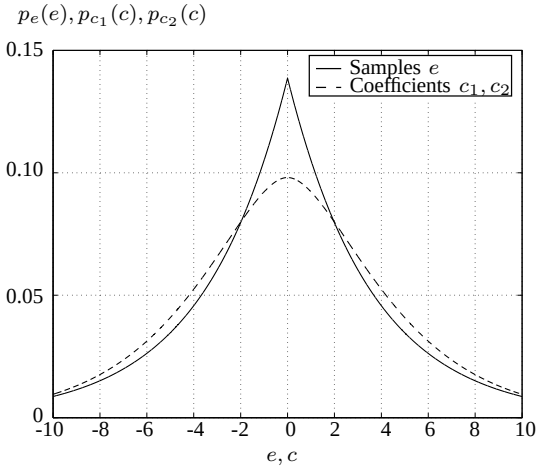

Fig. 8. Illustration of three probability density functions. $p_e(e)$ is a Laplacian probability density function of statistically independent samples. $p_{c_1}(c) = p_{c_2}(c)$ are the probability density functions of the coefficients c_1 and c_2 which are generated by the use of a one-dimensional DCT of a block size two. Variance: $\sigma_e^2 = 26$.

The probability density functions $p_e(e)$ according to equation 20 and $p_{c_1}(c) = p_{c_2}(c)$ according to equation 31 lead to different differential entropies $H(e)$ und $H(c)$. The difference can be calculated to $\Delta H \approx -0.07$ bit per sample. This corresponds to the value shown in Fig. 7.

B. Coding efficiency of the DCT of a block size four

Memoryless Laplacian distributed sources are considered as input to the DCT. For a block size of four, the DCT with the transform matrix

$$\mathbf{T}_{DCT,4} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ \cos\frac{\pi}{8} & \cos\frac{3\pi}{8} & -\cos\frac{3\pi}{8} & -\cos\frac{\pi}{8} \\ 1/\sqrt{2} & -1/\sqrt{2} & -1/\sqrt{2} & 1/\sqrt{2} \\ \cos\frac{3\pi}{8} & -\cos\frac{\pi}{8} & \cos\frac{\pi}{8} & -\cos\frac{3\pi}{8} \end{bmatrix} \tag{32}$$

generates the four coefficients c_i, with $i = 1, \ldots, 4$. The corresponding probability density functions $p_{c_i}(c)$ are calculated by the use of equations 13 and 14 to:

$$p_{c_1}(c) = p_{c_3}(c) = e^{\frac{-2\sqrt{2}|c|}{\sigma_e}} \left(\frac{5\sqrt{2}\sigma_e + 12|c|}{16 \cdot \sigma_e^2} + \frac{(3\sigma_e^2 + 6\sqrt{2}|c|\sigma_e + 4c^2)}{6\sigma_e^4} \right)$$

and

$$p_{c_2}(c) = p_{c_4}(c) =$$

$$e^{\frac{-2\sqrt{2}|c|}{A\sigma_e}} \left(\frac{\sqrt{2}(A^3 + 2BA^2)}{C} - \frac{\sqrt{2}A^2B^2\sigma_e + 2A^2B|c| - 2A^3|c|}{D} \right) +$$

$$e^{\frac{-2\sqrt{2}|c|}{B\sigma_e}} \left(\frac{\sqrt{2}(-B^3 - 2AB^2)}{C} + \frac{\sqrt{2}A^2B^2\sigma_e + 2AB^2|c| - 2B^3|c|}{D} \right),$$

$$\text{with} \quad A = \sqrt{2}\cos\tfrac{\pi}{8}, \quad B = \sqrt{2}\cos\tfrac{3\pi}{8},$$

$$C = 2\sigma_e \left(A^4 + 2BA^3 - 2AB^3 - B^4 \right)$$

$$D = \sigma_e^2 \left(-BA^4 + 2B^3A^2 - B^5 + A^5 - 2A^3B^2 + AB^4 \right). \tag{33}$$

All probability density functions $p_{c_i}(c)$ are not Laplacian. In Fig. 9 the probability density function p_e of the samples e as well as the probability density functions p_{c_i}, with $i = 1, \ldots, 4$, of the coefficients c_i are shown for an example variance of $\sigma_e^2 = 26$.

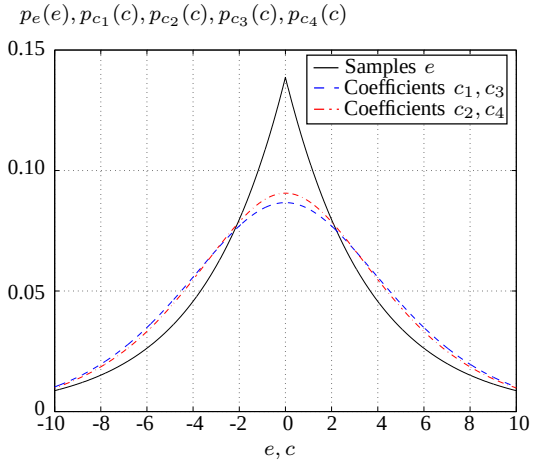

Fig. 9. Illustration of five probability density functions. $p_e(e)$ is a Laplacian probability density function of statistically independent samples. $p_{c_1}(c) = p_{c_3}(c)$ and $p_{c_2}(c) = p_{c_4}(c)$ are the probability density functions of the coefficients c_1, c_2, c_3, and c_4 which are generated by the use of a one-dimensional DCT of the block size four. Variance: $\sigma_e^2 = 26$.

The probability density functions $p_e(e)$ according to equation 20 and $p_{c_i}(c)$, with $i = 1, \ldots 4$, according to equation 33 lead to different differential entropies $H(e)$ and $H(c)$. The difference is $\Delta H \approx -0.09$ bit per sample.

C. Coding efficiency of the DST of a block size two

Memoryless Laplacian distributed sources are considered as input to the DST. For a block size of two, the DST with the transform matrix

$$\mathbf{T}_{DST,2} = \frac{2}{\sqrt{5}} \begin{bmatrix} \sin\frac{\pi}{5} & \sin\frac{3\pi}{5} \\ \sin\frac{3\pi}{5} & -\sin\frac{\pi}{5} \end{bmatrix} \tag{34}$$

generates the two coefficients c_1 and c_2, which are uncorrelated:

$$E\left[c_1 \cdot c_2\right] =$$

$$\tfrac{4}{5} E\left[\left(\sin\tfrac{\pi}{5} \cdot e_1 + \sin\tfrac{3\pi}{5} \cdot e_2 \right) \cdot \left(\sin\tfrac{3\pi}{5} \cdot e_1 - \sin\tfrac{\pi}{5} \cdot e_2 \right) \right] = 0 \tag{35}$$

The corresponding probability density functions $p_{c_i}(c)$, which are not Laplacian, are calculated by the use of equations 13 and 14 to:

$$p_{c_1}(c) = p_{c_2}(c) = \sum_{i=1}^{2} \frac{1}{2\sqrt{2} \cdot \sigma_e \left(A + (-1)^i \cdot B \right)} \left(e^{\frac{-\sqrt{2} \cdot |c|}{A \cdot \sigma_e}} + (-1)^i \cdot e^{\frac{-\sqrt{2} \cdot |c|}{B \cdot \sigma_e}} \right) \tag{36}$$

$$\text{with} \quad A = \tfrac{2}{\sqrt{5}} \sin\tfrac{\pi}{5}, \quad B = \tfrac{2}{\sqrt{5}} \sin\tfrac{3\pi}{5}$$

The probability density functions $p_e(e)$ according to equation 20 and $p_{c_i}(c)$, with $i = 1, \ldots 2$, according to equation 36 lead to different differential entropies $H(e)$ and $H(c)$. The difference is $\Delta H \approx -0.06$ bit per sample, which is similar to the difference of $\Delta H \approx -0.07$ bit per sample in the case of the DCT. As the probability density functions of the samples and coefficients differ and as the coefficients are uncorrelated, statistical dependencies of higher order are also generated by the DST according to the characteristic function. These cause the data rate increase.

D. Coding efficiency of the DCT and DST of large block sizes

According to equation 6, each coefficient, which is generated by the DCT or DST can be interpreted as a function of the elements of the input vector. With a further increase of the block size of the DCT or DST, the number of elements in the input vector also increases further. According to the central limit theorem, the probability density function of each coefficient becomes approximately a Gaussian probability density function in the case that the number of elements in the input vector and thus the block size of the transform increases to infinity [33]. In this specific case, the differential entropy of the input signal with a Laplacian probability density function is $H(e) = \frac{1}{2}\log_2(2 \cdot \mathrm{e}^2 \cdot \sigma_e^2)$ and the differential entropy of each coefficient c_i is $H(c_i) = \frac{1}{2}\log_2(2 \cdot \pi \cdot \mathrm{e} \cdot \sigma_e^2)$. Hereby, e is the Euler's constant.

The difference ΔH_∞ according to equation 16 can be calculated to

$$\Delta H_\infty = H(e) - H(c_i) = \frac{1}{2}\log_2(2\mathrm{e}^2\sigma_e^2) - \frac{1}{2}\log_2(2\pi\mathrm{e}\sigma_e^2)$$

$$\frac{1}{2}\log_2\left(\frac{\pi}{\mathrm{e}}\right) \approx -0.104 \quad \frac{\text{bit}}{\text{Sample}}. \tag{37}$$

Although Laplacian distributed sources are in the focus of this paper, Gaussian distributed sources are considered for comparison as they are often used in the literature to assess the efficiency of transforms.

E. Coding efficiency of the DCT for Gaussian distributed sources

For comparison, samples e with a Gaussian probability density function are considered. The Gaussian probability density function has the specific feature that with the specification of the correlation coefficient ρ_e all statistical dependencies of higher order are also specified [33]. Therefore, the two-dimensional joint probability density function of two successive samples can be defined by the single parameter ρ_e. It is given by [33]

$$p_{\vec{e},G}(e_1,e_2) = \frac{1}{2\cdot\pi\cdot\sigma_e^2\cdot\sqrt{1-\rho_e^2}}\cdot \mathrm{e}^{\frac{-e_1^2-e_2^2+2\cdot e_1\cdot e_2\cdot\rho_e}{2\cdot\sigma_e^2\cdot(1-\rho_e^2)}}. \quad (38)$$

Consequently, for the assessment of the efficiency of the DCT it is sufficient to consider the dependency on the correlation coefficient ρ_e. The probability density functions of the coefficients, which are given by the equations 13 and 14 are also Gaussian since every linear combination of Gaussion distributed signals is also Gaussian distributed [33]. However, each coefficient has a different variance, see equation 18. According to [32], the differential entropy of a source of a Gaussian probability densitiy function is

$$H(e) = \frac{1}{2}\log_2(2\cdot\pi\cdot\mathrm{e}\cdot\sigma_e^2). \quad (39)$$

With equation 39, the difference H_{Gauss} becomes:

$$\begin{aligned}
&\Delta H_{Gauss}\\
&= \frac{1}{2}\cdot\log_2(2\cdot\pi\cdot\mathrm{e}\cdot\sigma_e^2) - \frac{1}{2}\cdot\sum_{i=1}^{2}\frac{1}{2}\cdot\log_2(2\cdot\pi\cdot\mathrm{e}\cdot\sigma_{c_i}^2)\\
&= \underbrace{\frac{1}{6.02}\cdot 10\cdot\log_{10}\frac{\sigma_e^2}{\sqrt{\prod_{i=1}^{2}\sigma_{c_i}^2}}}_{G_{TC}} = \frac{1}{6.02}\cdot 10\cdot\log_{10}\frac{1}{\sqrt{1-\rho_e^2}}.
\end{aligned} \quad (40)$$

In the literature, the term G_{TC} is denoted as the transform coding gain [32]. In Fig. 10, ΔH_{Gauss} is illustrated as a function of the correlation coefficient ρ_e. It can be seen that $\Delta H_{Gauss} \geq 0$ for all possible values of the correlation coefficient ρ_e. Consequently, the data rate of the coefficients is always lower than the data rate of the samples themselves. Therefore, the DCT is always efficient for the coding of Gaussian distributed sources.

In the specific case of $\rho_e = 0$, the difference $\Delta H_{Gauss} = 0$ and thus the data rate of the coefficients equals the data rate of the samples. The coefficients are also uncorrelated according to equation 19. In the specific case of a Gaussian probability density function, the uncorrelated coefficients are also statistically independent [33].

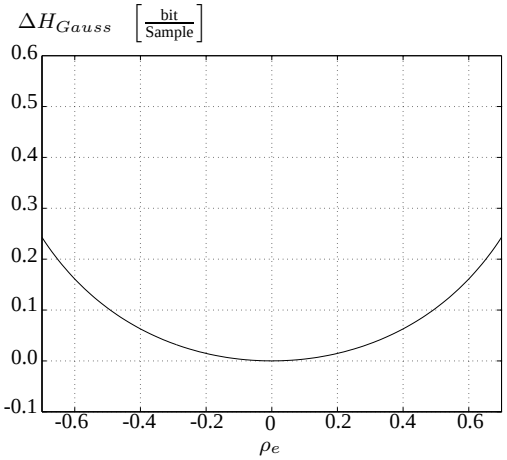

Fig. 10. Illustration of the difference $\Delta H_{Gauss} = H(e') - H(c')$ of the data rates of two codings under the assuption of memoryless sources as a function of the correlation coefficient ρ_e. $H(e')$ is the minimum data rate required to code the source e with a Gaussian probability density function. $H(c')$ is the minimum data rate required to code the coefficients generated by the application of a DCT of a block size two to the source e.

IV. Investigation of the HEVC prediction error coding

The prediction error generated by the HEVC reference model HM8 [22] in the context of Fig. 2 is split into two sources S_1 and S_2, whereby S_1 contains samples of low and S_2 samples of high statistical dependencies. According to Fig. 7, ΔH is negative for S_1 and positive for S_2. In accordance with HEVC, the prediction error is split blockwise using blocks of 4×4 samples. The assignment of an individual block to a source is performed using a locally measured ΔH_{block}, which is determined out of the data rate of its 16 samples e_i, with $i = 1,\ldots,16$, and the data rate of its 16 coefficients c_i. The 16 coefficients are generated by the application of a one-dimensional DCT according to equation 17 to the samples of the block. As the behaviour of the DCT and the DST is similar, these calculations use only the DCT for simplicity reasons, although HEVC applies both depending on whether intra or inter prediction is used. To estimate the data rates, the distributions of the samples and of the coefficients are approximated by step functions of the same step size. The approximated distribution of the samples is denoted as $P_{\Delta,e}(e)$, the approximated distributions of the coefficients are denoted as $P_{\Delta,c_1}(c)$ and $P_{\Delta,c_2}(c)$. The step size is assumed to be small such that the quantization error is uniformly distributed and the mean squared error according to equation 8 is the same in the spatial and in the frequency domain. A self information can be determined out of the approximated distributions assuming a memoryless source. The difference ΔH_{block} results in:

$$\begin{aligned}
\Delta H_{block} &= -\sum_{i=1}^{16}\log_2 P_{\Delta,e}(e_i)\\
&\quad + \sum_{i=1}^{8}\left(\log_2 P_{\Delta,c_1}(c_{1,i}) + \log_2 P_{\Delta,c_2}(c_{2,i})\right)
\end{aligned} \quad (41)$$

All blocks with $\Delta H_{block} < 0$ are assigned to S_1, all remaining blocks with $\Delta H_{block} \geq 0$ to S_2. The properties

of the resulting two sources are summarized in Table I. Fig. 11 shows the corresponding joint distributions $P_{\vec{e},S_1}(e_1,e_2)$ and $P_{\vec{e},S_2}(e_1,e_2)$. For S_1 holds $\Delta H_{S_1} < 0$. The joint distribution $P_{\vec{e},S_1}$ resembles the joint probability density function $p_{\vec{e},L2}(e_1,e_2)$ of a memoryless Laplacian distributed source of Fig. 5. For S_2 holds $\Delta H_{S_2} \geq 0$. The joint distribution $P_{\vec{e},S_2}$ resembles the joint probability density function $p_{\vec{e},L1}(e_1,e_2,\rho_e = 0.61)$ of Fig. 6. For the case that both sources S_1 and S_2 are coded in the frequency domain, the differential entropy is measured to $H_{S_1+S_2}(c) = 3.70$ bit per sample.

TABLE I
PROPERTIES OF THE SOURCES S_1 AND S_2 OF THE HEVC PREDICTION ERROR FOR THE TEST SEQUENCES USED IN THE HEVC STANDARDIZATION [21].

Source S_1			Source S_2		
$H_{S_1}(e)$	=	4.21 bit/sample	$H_{S_2}(e)$	=	3.74 bit/sample
$H_{S_1}(c)$	=	4.30 bit/sample	$H_{S_2}(c)$	=	3.58 bit/sample
ΔH_{S_1}	=	−0.09 bit/sample	ΔH_{S_2}	=	0.18 bit/sample
$P(S_1)$	=	0.12	$P(S_2)$	=	0.88
ρ_{e,S_1}	=	0.22	ρ_{e,S_2}	=	0.61
σ^2_{e,S_1}	=	48.97	σ^2_{e,S_2}	=	23.35

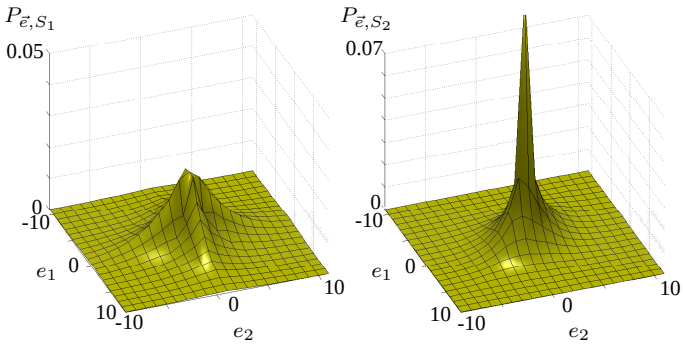

Fig. 11. Illustration of the measured joint distributions $P_{\vec{e},S_1}(e_1,e_2)$ of the source S_1 and $P_{\vec{e},S_2}(e_1,e_2)$ of the source S_2 of the prediction error signal for the sequences used in the HEVC standardization [21].

The adaptive coding of S_1 in the spatial domain and of S_2 in the frequency domain could lead to the data rate reduction

$$H_{S_1+S_2}(c) - \left(H_{Dom} + \sum_{i=1} H_{S_i}(e) \cdot P(S_i) \right) = 0.015 \frac{\text{bit}}{\text{Sample}}, \tag{42}$$

which corresponds to 0.4%, if the selection of the domain is coded under the assumption of a memoryless source:

$$H_{Dom} = -\sum_{i=1}^{2} P(S_i) \cdot \log_2 P(S_i) \quad \frac{\text{bit}}{4 \times 4 \text{ Samples}}. \tag{43}$$

As the data rate reduction by the adaptive prediction error coding of HEVC is particularly high for screen content, it is considered separately in the following. Table II summarizes the corresponding properties of the two sources S_1 and S_2. Compared to Table I, it can be recognized that the probability and the variance of the source S_1, for which the DCT is inefficient, are both larger. For the case that both sources S_1 and S_2 are coded in the frequency domain, the differential entropy is measured to $H_{S_1+S_2}(c) = 2.42$ bit per sample.

TABLE II
PROPERTIES OF THE SOURCES S_1 AND S_2 OF THE HEVC PREDICTION ERROR FOR SCREEN CONTENT TEST SEQUENCES USED IN THE HEVC STANDARDIZATION [21].

Source S_1			Source S_2		
$H_{S_1}(e)$	=	3.24 bit/sample	$H_{S_2}(e)$	=	2.59 bit/sample
$H_{S_1}(c)$	=	3.34 bit/sample	$H_{S_2}(c)$	=	2.04 bit/sample
ΔH_{S_1}	=	−0.10 bit/sample	ΔH_{S_2}	=	0.55 bit/sample
$P(S_1)$	=	0.20	$P(S_2)$	=	0.80
ρ_{e,S_1}	=	0.22	ρ_{e,S_2}	=	0.67
σ^2_{e,S_1}	=	69.03	σ^2_{e,S_2}	=	23.60

For these screen content sequences, the adaptive coding of S_1 in the spatial domain and of S_2 in the frequency domain could lead to the data rate reduction of 0.09 bit per sample, corresponding to 3.7%.

The measured data rate reduction, which HEVC achieves by its adaptive prediction error coding, is with 7% larger than the calculated one. The difference is caused by the approximations used in the calculations.

V. CONCLUSION

This paper presents a theoretical analysis of the efficiency of the DCT and DST for the coding of prediction errors in hybrid video coding. The prediction error is modelled as a Laplacian distributed source. The transform is considered being efficient if the data rate required to code the coefficients is lower than the data rate required to code the input samples themselves at the same mean squared reconstruction error. For the coding of the input samples as well as for the coding of the coefficients, memoryless sources and fine quantization are assumed. The coding efficiency of the transforms depends on the statistical moments and on the probability distribution of the input signal. For a Gaussian distributed input signal, the DCT always leads to a data rate reduction. However, for a Laplacian distributed input signal, the transforms sometimes increase the data rate. It is shown that the inrease of the data rate is caused by the generation of higher statistical moments of the coefficients by the DCT or DST. For a memoryless Laplacian distributed input signal and a block size of two, the data rate increase is calculated for a one-dimensional DCT to 0.07 and for a DST to 0.06 bit per sample. A one-dimensional DCT of the block size four leads to a data rate increase of 0.09 bit per sample. For both transforms, the data rate increase converges to 0.10 bit per sample for larger block sizes. It is shown that the prediction error generated by the HEVC reference model can be split block-wise into two sources, whereas for one source, the transform leads to a data rate reduction, but for the second source, it leads to a data rate inrease. For screen content, the transform inreases the data rate by around 0.10 bit per sample for approximately 20% of the blocks with low correlation. This results in a 7% data rate reduction, which HEVC achieves by skipping the transform.

ACKNOWLEDGMENT

This work was primarily done at the University of Hannover. The author would like to thank Prof. Dr.-Ing. Hans Georg Musmann for his valuable comments.

REFERENCES

[1] ITU-T and ISO/IEC, "Generic Coding of Moving Pictures and Associated Audio Information - Part 2: Video," ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2), 1994.

[2] ISO/IEC, "Coding of Audio-Visual Objects - Part 2: Visual," ISO/IEC 14496-2 (MPEG-4 Visual version 1), 1999.

[3] ITU-T and ISO/IEC, "Advanced Video Coding for Generic Audiovisual Services," ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Version 8, Jul. 2007.

[4] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer und T. Wedi, "Video coding with H.264/AVC: Tools, Performance, and Complexity," *IEEE Circuits and Systems Magazine*, vol. 4, no. 1, pp. 7–28, Apr. 2004.

[5] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, T. Wiegand, "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Consent)," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-L1003, 12th Meeting: Geneva, CH, Jan. 2013.

[6] G. J. Sullivan, J. R. Ohm, W.-J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1649–1668, Dec. 2012.

[7] H. G. Musmann, P. Pirsch, and H. J. Grallert, "Adavances in Picture Coding," *Proceeding of IEEE*, vol. 73, no. 4, pp. 523–548, Apr. 1985.

[8] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 7, Jan. 1987.

[9] J. Lainema, F. Bossen, W.-J. Han, J. Min, K. Ugur, "Intra Coding of the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1792–1801, Dec. 2012.

[10] N. Ahmed, T. Natarajan und K.R. Rao, "Discrete Cosine Transform," *IEEE Transactions on Computers*, pp. 90–93, Aug. 1974.

[11] H. S. Malvar, A. Hallapuro, M. Karczewicz, L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 598 – 603, Jul. 2003.

[12] A. Saxena, F. Fernandes, "CE7: Mode-dependent DCT/DST without 4*4 full matrix multiplication for intra prediction," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-E125, 5th Meeting: Geneva, CH, Mar. 2011.

[13] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1669–1684, Dec. 2012.

[14] T. Wedi, H. G. Musmann, "Motion and Aliasing Compensated Prediction for Hybrid Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, Jul. 2003.

[15] T. Wiegand, B. Girod, "Multi-frame motion-compensated prediction for video transmission," Kluwer Academic Publishers, 2001.

[16] M. Flierl, B. Girod, "Generalized B Pictures and the Draft H.264/AVC Video Compression Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 587–597, Jul. 2003.

[17] I.-K. Kim, J. Min, T. Lee, W.-J. Han; J. Park, "Comparison of the Coding Efficiency of Video Coding Standards including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1697–1706, Dec. 2012.

[18] P. Helle, S. Oudin, B. Bross, D. Marpe, M. O. Bici, K. Ugur, J. Jung, G. Clare, T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1720–1731, Dec. 2012.

[19] C. Lan, F. Wu, G. Shi, "Compress Compound Images in H.264/MPEG-4 AVC by Fully Exploiting Spatial Correlation," Proc. of IEEE International Symposium on Circuits and Systems, pp. 2818–2821, 2009.

[20] C. Lan, X. Peng, J. Xu, F. Wu, "Intra and inter coding tools for screen contents," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-E145, m19662, 5rd Meeting: Geneva, CH, Mar. 2011.

[21] F. Bossen, "HM 8 Common Test Conditions and Software Reference Configurations," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-J1100, 10th Meeting: Stockholm, SE, Jul. 2012.

[22] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, "HM8," http://hevc.kw.bbc.co.uk/trac/browser/tags/HM-8.0, Aug. 2012.

[23] M. Narroschke, "Extending H.264/AVC by an adaptive coding of the prediction error," Proc. of Picture Coding Symposium, Beijing, China, Apr. 2006.

[24] ——, "Adaptive Prädiktionsfehlercodierung fuer die Hybridcodierung von Videosignalen," Fortschritt-Berichte VDI, Reihe 10, Nr. 786, ISBN 978-3-18-378610-7, 2008.

[25] M. Narroschke, H. G. Musmann, "Adaptive prediction error coding in spatial and frequency domain for H.264/AVC," ITU-T SG16/Q.6, Dokument VCEG-AB06, Bangkok, Thailand, Jan. 2006.

[26] ——, "Adaptive prediction error coding in spatial and frequency domain with a fixed scan in the spatial domain," ITU-T SG16/Q.6, Dokument VCEG-AD07, Hangzhou, China, Oct. 2006.

[27] M. Narroschke, "Experimental results for adaptive prediction error coding in spatial and frequency domain for common test conditions," ITU-T SG16/Q.6, Dokument VCEG-AE15, Marrakech, Marocco, Jan. 2007.

[28] C. Lan, J. Xu, G. J. Sullivan, F. Wu, "Intra transform skipping," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-I0408, 9th Meeting: Geneva, CH, Apr. 2012.

[29] X. Peng, C. Lan, J. Xu, G. J. Sullivan, "Inter transform skipping," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-J0237, 10th Meeting: Stockholm, SE, Jul. 2012.

[30] K. Andersson, J. Ström, "Fix for residual coding of transform skipped blocks," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-K0245, 11th Meeting: Shanghai, CN, Oct. 2012.

[31] D. He, J. Wang, G. Martin-Cocher, "Rotation of Residual block for Transform Skipping in HM8," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Document JCTVC-K0237, 11th Meeting: Shanghai, CN, Oct. 2012.

[32] N.S. Jayant and P. Noll, "Digital Coding of Waveforms," Prentice-Hall, 1984.

[33] A. Papoulis and S. U. Pillai, "Probability, Random Variables and Stochastic Processes," McGraw-Hill, 4th edition, 2002.

[34] S. Kotz, T. J. Kozubowski und K. Podgórski, "An asymmetric multivariate Laplace Distribution," Technical Report No. 367, Department of Statistics and Applied Probability, University of California at Santa Barbara, 2000.

Matthias Narroschke Matthias Narroschke received his Dipl.-Ing. degree in electrical engineering in 2001 (with highest honors) and the Dr.-Ing. degree in 2008, both from the University of Hannover. From 2001 until 2007 he worked as a research engineer and teaching assistant at the Institut fuer Informationsverarbeitung of the University of Hannover. The focus of his research was the coding of the prediction error in the context of hybrid video coding. Furthermore, his research included image processing and internet streaming. In 2003, he became the Ober-Ingenieur of this institute. In 2007, he joined the Panasonic R&D Center Germany in Langen where he is currently principle engineer. His responsibilities include video coding standardization activities and the development of future video coding schemes. In 2008, he became a guest lecturer at the University of Hannover.

He received the Robert-Bosch-Prize for the best Dipl.-Ing. degree in electrical engineering in 2001. Dr. Narroschke is an active contributer to the Motion Picture Experts Group (MPEG) of ISO/IEC SC29 and to the Video coding Experts Group (VCEG) of the ITU. He has filed several international patents in the area of video coding, most of them in cooperation with Panasonic.