Performance Evaluation of the HEVC 4x4 Integer Transforms

HARIZE Saliha
Department of Electronics, Badji Mokhtar University,
BP 12, Sidi Ammar, Annaba, Algeria.
shrz.dj@gmail.com

SEMIRA Hichem
Department of Electronics, Badji Mokhtar University,
BP 12, Sidi Ammar, Annaba, Algeria.
hichem.semira@univ-annaba.org

BENOUARET Mohamed
Department of Electronics, Badji Mokhtar University,
BP 12, Sidi Ammar, Annaba, Algeria.
mohamed.benouaret@gmail.com

***Abstract*— The state of the art High Efficiency Video Coding (HEVC) has adopted an integer DCT-II transform for square blocks of size 4x4, 8x8, 16x16 and 32x32. Alternatively, an integer DST-VI is specified for 4x4 intra-predicted luminance residuals. The DCT-II transforms show partials butterfly structures which reduce the computational complexity. But, the DST-VI lacks this propriety and presents, therefore, a serious challenge. This paper presents a performance evaluation of the HEVC main profile when using the DCT-II instead of the DST-VI. The tests carried using the HEVC test model HM-16.3 show that the use of the DCT-II yields an insignificant Bjøntegaard Delta psnr mean decrease of 0.02 dB and a slight increase in the bitrate of 0.5%. The paper ends with a discussion on the computational complexity of the DCT-II and the DST-VI transforms.**

Keywords*— *Discrete cosine transform (DCT); Discrete sine transform (DST); HEVC; B-D PSNR; B-D rate

I. INTRODUCTION

High Efficiency Video Coding (HEVC) is a video compression standard, successor to H.264/MPEG-4 AVC (Advanced Video Coding). It was jointly developed by the ISO/IEC JTC 1/SC 29/WG 11 Moving Picture Experts Group (MPEG) and ITU-T SG16/Q.6 Video Coding Experts Group (VCEG) as ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T H.265[1], [2], [3]. The main objective of the HEVC is either to double the data compression ratio compared to its predecessor H.264/MPEG-4 AVC at the same quality or, to provide considerably improved video quality at the same bit rate. It can support 8K UHD and resolutions up to 8192 × 4320.

The first version of the standard was approved and published in April 13, 2013. The second one was approved in 2014 and published in early 2015. It includes format range extensions (RExt) (supporting higher bit depths and the monochrome, 4:2:2, and 4:4:4 chroma sampling formats), scalable coding extensions (SHVC), and multi-view extensions (MV-HEVC) [4]. Additional 3D-HEVC extensions for 3D video were completed this year [5]. Further screen content coding (SCC) extensions is expected in early 2016 for video containing rendered graphics, text, or animation as well as camera-captured video scenes.

HEVC specifies four transform units (TUs) sizes of 4×4, 8×8, 16×16, and 32×32 to code the prediction residuals. TUs use integer basis functions derived from the famous discrete cosine transform (DCT-II). In addition 4×4 intra coded luminance blocks are transformed using an integer discrete sine transform (DST). This is expected to provide a 1% bit rate reduction but was restricted to 4×4 luma transform blocks due to marginal benefits for the other transform cases [1]. The suitability of the DST-VI for one dimensional intra prediction was derived by Han *et al.* [6] and, almost at the same time, by Yeo *et al.* [7]. This DST is called DST-VI in [8] and DST-VII in [9].

HEVC uses 35 directional modes for intra prediction that can be divided into 3 categories: a DC mode, a planar mode and 33 angular modes as shown in figure 1.

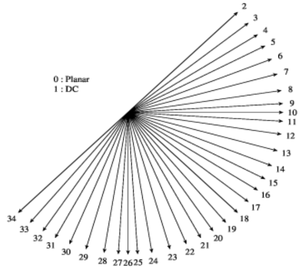

Fig.1 Intra prediction modes

Saxena *et al.* [9] showed that a combination of DCT-II and DST-VII is the optimal transform for all the oblique modes in unified intra prediction in the HEVC standardization. The DCT/DST transform scheme described in this paper was also adopted in the HEVC standardization for block size 4×4 for

Intra Luma blocks. In [10], a simplification of the DCT-II/DST-VI was proposed and was adopted in the HEVC to remove the mode-dependency of whether to select between DCT-II or DST-VI as the horizontal (and respectively vertical) transform, and the DST-VI transform was always used for the 4×4 Intra Luma blocks. Nevertheless, the computational complexity of the DST-VI is still a research subject, and for this reason, there is no DST-VI transforms for blocks bigger than 4×4.

In this paper, the quality and compression losses which occur when using the DCT-II only are evaluated and discussed.

The rest of the paper is structured as follows : In section II, the real and integer DCT-II and the DST-VI are presented and their main characteristics are given. Section III presents the simulation results of the various tests conducted on the HM16.3 HEVC reference model. The paper ends with a conclusion where the test results are thoroughly analyzed and discussed.

II. THEORETICAL BACKGROUND

For a N samples sequence, the real DCT-II, DST-II and DST-VI are given by the following formulas:

$$DCT_{II}:\left[C^{II}_{N}\right]mn = \sqrt{\frac{2}{N}}\; k_m \cos\left(\frac{m(2n+1)\pi}{2N}\right) \; m,n = 0,1,\dots,N-1 \qquad (1)$$

$$k_m = \begin{cases} 1/\sqrt{2}\,, if\ m = 0 \\ 1,\ \ otherwise \end{cases}$$

$$DST_{VI}:\left[S^{\ VI}_{N-1}\right]mn = \frac{2}{\sqrt{2N-1}} \sin\left(\frac{m(2n-1)\pi}{2N-1}\right) \quad m,n = 1,\dots,N \qquad (2)$$

For $N = 4$, the real DCT-II and DST-VI cores are:

$$Real\ (DCT_{II}) = \begin{bmatrix} 0.5000 & 0.5000 & 0.5000 & 0.5000 \\ 0.6533 & 0.2706 & -0.2706 & -0.6533 \\ 0.5000 & -0.5000 & -0.5000 & 0.5000 \\ 0.2706 & -0.6533 & 0.6533 & -0.2706 \end{bmatrix} \qquad (3)$$

$$Real\ (DST_{VI}) = \begin{bmatrix} 0.2280 & 0.4285 & 0.5774 & 0.6565 \\ 0.5774 & 0.5774 & 0 & -0.5774 \\ 0.6565 & -0.2280 & -0.5774 & 0.4285 \\ 0.4285 & -0.6565 & 0.5774 & -0.2280 \end{bmatrix} \qquad (4)$$

The integer core transforms used in the HEVC are obtained by scaling the real transform coefficients by 128 (7 bits) and rounding to the next integer. In general, an integer transform matrix A is equal to round$(64\sqrt{N}.realA)$ where N is the transform dimension. Applying this relation for $N = 4$, the following integer transform cores are obtained:

$$Integer\ (DCT_{II}) = \begin{bmatrix} 64 & 64 & 64 & 64 \\ 83 & 36 & -36 & -83 \\ 64 & -64 & -64 & 64 \\ 36 & -83 & 83 & -36 \end{bmatrix} \qquad (5)$$

$$Integer\ \ (DST_{VI}) = \begin{bmatrix} 29 & 55 & 74 & 84 \\ 74 & 74 & 0 & -74 \\ 84 & -29 & -74 & 55 \\ 55 & -84 & 74 & -29 \end{bmatrix} \qquad (6)$$

In the literature, a number of criteria are used to evaluate the efficiency of such transforms:

* Orthogonality: which exhibits the highly desirable property of the transform core as the inverse is its transpose and the quantization distortion can be estimated in the transform domain. An orthogonal transform preserves the energy and does not amplify the noise. A 2D block X is transformed using a separable approach by applying the transform to each column, then to each row.

$$Y_{N\times N} = A_{N\times N} X_{N\times N} A^{T}_{N\times N} \qquad (7)$$

A transform matrix A is orthogonal if $A \times A^T = I$, I being the identity matrix and A^T its transpose.

The orthogonality tests reveal that both the DCT-II and the DST-VI present a very small norm deviation and that the last one shows also a tiny deviation from perfect orthogonality.

$$DCT_{II} \times DCT_{II}^{T} = \begin{bmatrix} 16384 & 0 & 0 & 0 \\ 0 & 16370 & 0 & 0 \\ 0 & 0 & 16384 & 0 \\ 0 & 0 & 0 & 16370 \end{bmatrix} \qquad (8)$$

$$DST_{VI} \times DST_{VI}^{T} = \begin{bmatrix} 16398 & 0 & -15 & 15 \\ 0 & 16428 & 0 & 0 \\ -15 & 0 & 16398 & -15 \\ 15 & 0 & -15 & 16398 \end{bmatrix} \qquad (9)$$

* Coding gain: It has been established that many real signals do behave statistically like a first order stationary Gauss-Markov source [11]. In the context of one dimension, row or column:

$$x_k = x_{k-1} + \varepsilon_k \qquad (10)$$

Where ρ is the correlation parameter $0 \le \rho \le 1$ and ε_k is the white noise process, with zero mean and a variance of $1-\rho^2$

The input x is defined by a covariance matrix R_x , whose elements are:

$$R_{x_{i,j=}}\rho^{|i-j|} \qquad (11)$$

The matrix R_x is symmetric and has a Tœplitz structure.

The autocorrelation (or covariance) matrix R_y of the transformed vector source is obtained as:

$$R_y = AR_xA^T \qquad (12)$$

The formula for the coding gain is [11]:

$$C_g = \frac{\frac{1}{N}\sum_{i=0}^{N-1}\sigma_{xi}^2}{\left(\prod_{i=0}^{N-1}\sigma_{xi}^2\,\|f_i\|^2\right)^{1/N}} \qquad (13)$$

Where σ_{xi}^2 is the variance of the ith transform coefficient, being the diagonal entry of the matrix R_y and $\|f_i\|^2$ is the 2-norm of ith basis function of the transform matrix:

$$C_{g_{dB}} = 10log_{10}\left(\frac{\frac{1}{N}\sum_{i=0}^{N-1}\sigma_{xi}^2}{\left(\prod_{i=0}^{N-1}\sigma_{xi}^2\|f_i\|^2\right)^{1/N}}\right) \qquad (14)$$

Transforms with higher coding gains C_g pack more energy into less number of coefficients. Thus, leading to a higher compression rate and consequently a lower bitrate. As an example, the optimal Karhunen–Loeve (KLT) and the real DCT-II (8×8) transforms have coding gain of 8.8462 and 8.8259 respectively for a correlation coefficient $\rho = 0.95$.

Table 1 shows the coding gains of the DCT-II, the DST-VI and the combination DCT-II/DST-VI.

Table 1. Coding gains of the DCT-II, the DST-VI and their combination

ρ	0.9	0.95	0.99
C_g(DCT-II/DCT-II) $_{dB}$	5.3869	7.5700	12.7558
C_g(DST-VI/DST-VI) $_{dB}$	4.1422	5.2287	6.6227
C_g(DCT-II/DST-VI) $_{dB}$	6.1598	8.3833	13.6050

The results reveal that the combination DCT-II/DST-VI outperforms the DCT-II/DCT-II. An analysis of the outcome show that there is a mean coding gain increase of 0.6172 dB corresponding to 19.71% when using the DCT-II/DST-VI.

The use of the DST-VI is motivated by the fact that intra prediction is based on the top and left neighbors as shown in figure 2. The prediction accuracy is higher for the pixels located near the top/left neighbors than for those away from it. Therefore, the residual of pixels which are away from the top/left neighbors are usually larger than those near the neighbors. In this case, the DST transform is more suitable to code such kind of residuals. Indeed, the DST basis functions and at the opposite of the DCT-II, start with the lowest frequency and increase to the right and to the bottom.

In the next section, the effect of replacing the DST-VI by the DCT-II on the performance is evaluated in terms of the average Bjøntegaard-Delta rate (BD-rate) and Delta-PSNR (BD-PSNR) [12].

III. SIMULATION RESULTS

Two sets of tests are needed to compare the HEVC performance when using the DCT-II instead of the DST-VI.

These simulations were conducted on 6 standard video sequences [13] in the YUV color space with a slight deviation from the common HM test conditions (CTC) and software reference configurations [14]. Table 2 presents the characteristics of the video sequences. The HM 16.3 version of the HEVC test model was used with three different configurations of the Main profile: All_Intra, Lowdelay_B and Random_access. The Main profile allows for a bit depth of 8-bits per sample with 4:2:0 chroma sampling, which is the most common type of video used with consumer devices.

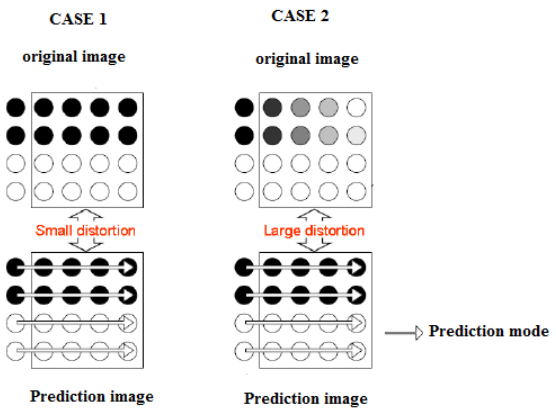

Fig.2 An example of conventional horizontal prediction mode

The All_Intra mode is used for editing purposes and all the frames are intra coded. A picture-wise access to the sequence is therefore enabled. In the low delay_B mode, only the first frame is intra coded while all the following images are bi-predictive inter coded. This is a coding with no structural delay, i.e., the frames coding order is the same as the output order. This configuration is suitable for interactive applications. The random access mode is convenient for entertaining applications such as broadcasting and streaming. One frame is intra coded in a group of 16 pictures for video sequences rate of 20fps (20 Hz), in 32 pictures for 30fps and in 64 pictures for 60fps. The frame coding and output order are different.

The tests were run on a Processor : Intel ® Core ™2 Duo CPU T6670 @ 2.20 GHz with 2 GB RAM and windows 7 32 bits operating system. The C++ compiler used is Microsoft Visual Studio 9.

The deviation from the CTC is justified by the fact that the process is too slow and it was not possible to encode the full video sequences of class A and B as the process runs out of memory space when encoding more than 50 and 76 frames respectively. Video sequences of class E are no longer available and CIF and QCIF sequences were added to enlarge the tests. The class A test sequences are not used for low delay coding tests as they represent application areas like editing or ultra high definition broadcasting [8].

The first set of tests were conducted using the DST-VI as the 4×4 intra predicted luma block transform. Then, it was replaced by the DCT-II and, after a successful build, a second group of simulations were run. The modifications were made in the TLibCommon\Source Files\TComRom.cpp file of the HM_vc9 solution.

A sample of the rate-distortion curves is shown in figures 3, 4 and 5. They indicate that the DCT-II performs almost as well as the DST-VI, particularly in the Low delay and random access modes where the number of intra predicted blocks is highly inferior to that of the all_intra mode.

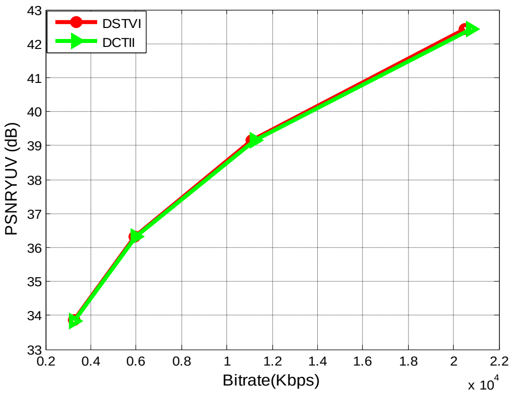

Fig 3. R-D curve for the sequence Basketballdrill in AI mode

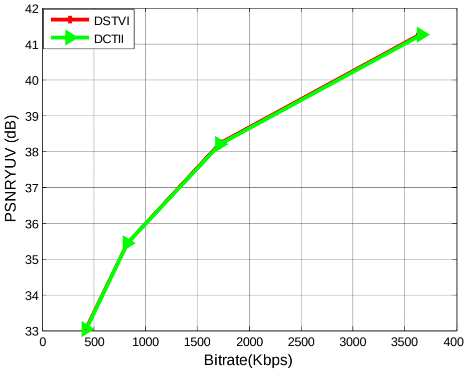

Fig 4. R-D curve for the sequence Basketballdrill in LDB mode

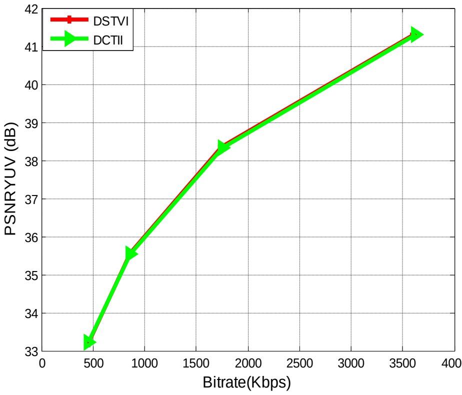

Fig 5. R-D curve for the sequence Basketballdrill in RA mode

To obtain the exact performance differences , the B-D PSNR-YUV and BD-rate were calculated and reported in table 3. The PSNR-YUV is calculated according to the following rule which is applicable in the case of 4: 2: 0 format:

$$PSNR_{YUV} = \frac{4.PSNR\_Y + PSNR\_U + PSNR\_V}{6} \qquad (15)$$

While the next formula

$$PSNR_{YUV} = \frac{6.PSNR\_Y + PSNR\_U + PSNR\_V}{8} \qquad (16)$$

Which has been used in [15] is applicable in the case of 4: 2: 2 format.

Table 2. The video sequences and their characteristics

Video Sequence	*Class*	*Resolution*	*Number of frames*	*Frame rate (Hz)*
Traffic	**A**	**2560x1600**	**300**	**30**
Tennis	**B**	**1920x1080**	**240**	**30**
Basketballdrill	**C**	**832x480**	**500**	**50**
Basketballpass	**D**	**416x240**	**300**	**50**
Foreman	**CIF**	**352x288**	**300**	**24**
Coastguard	**QCIF**	**176x144**	**300**	**30**

Table 3. Simulation results

Video Sequence DSTVI/DCTII	*AI*		*LDB*		*RA*	
	BD-PSNR-YUV (dB)	***BD-rate (%)***	***BD-PSNR-YUV (dB)***	***BD-rate (%)***	***BD-PSNR-YUV (dB)***	***BD-rate (%)***
Traffic	**-0.0152**	**0.5006**			**-0.053**	**1.068**
Tennis	**-0.0139**	**0.4775**	**-0.0039**	**0.1136**	**0.0015**	**-0.0593**
Basketballdrill	**-0.0552**	**1.1896**	**-0.0065**	**0.1758**	**-0.0209**	**0.5416**
Basketballpass	**-0.0542**	**0.9567**	**-7.53.10^{-4}**	**0.0234**	**-0.0249**	**0.5128**
Foreman	**-0.0842**	**1.6589**	**-0.0045**	**0.1114**	**-0.0222**	**0.5677**
Coastguard	**-0.0632**	**0.9216**	**-0.0055**	**0.0875**	**-0.0152**	**0.3778**
Mean value	**-0.0477**	**0.9508**	**-0.0042**	**0.1023**	**-0.0224**	**0.5014**

Table 4. Mean overall BD figures

	BD-PSNR-YUV (dB)	*BD-rate (%)*
Mean Value (all configurations)	**-0.0248**	**0.5182**

A negative value of the BD-PSNR-YUV means that there is a loss in the quality whereas a positive BD-rate reveals an increase in the bitrate.

The results indicated in table 3 came in complete concord with the previsions. Indeed, the highest decrease in the PSNR (-0.0477 dB) and the most important increase in the bitrate (0.9508 %) are found with the all intra mode. It is followed by the random access mode with -0.0224 dB and 0.5014% increase in the bitrate and lastly, the low delay configuration with -0.0042 dB and 0.1023% BD-rate. The low delay coding test uses only one intra predicted frame. The overall figures are shown in table 4.

From a computational complexity point of view, it can easily be stated that the DST-VI is far more complex than the DCT-II. Indeed, the DCT-II transform matrix uses 3 different coefficients (64, 36, 83) and presents a symmetric structure allowing a partial butterfly approach. In contrast, the DST-VI core transform includes 4 different non dyadic coefficients (29, 55, 74, 84) and does not present any symmetry. Moreover, there is no common factor to the 4 coefficients which, if it existed, would contribute to a less complex hardware implementation. Also, any tentative of optimization, such as getting 74 from 55 and 29 (74=55+29) would certainly result in race hazards. The following matrix calculations will permit to determine the number of necessary operations for the DCT-II and the DST-VI.

DCT-II: [16]

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \end{bmatrix} = \begin{bmatrix} 64 & 64 & 64 & 64 \\ 83 & 36 & -36 & -83 \\ 64 & -64 & -64 & 64 \\ 36 & -83 & 83 & -36 \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \end{bmatrix} = \begin{bmatrix} 64a(0) + 64a(1) \\ 83b(0) + 36b(1) \\ 64a(0) - 64a(1) \\ 36b(0) - 83b(1) \end{bmatrix} \quad (17)$$

With $\begin{cases} a(0) = x(0) + x(3) \\ a(1) = x(1) + x(2) \end{cases}$ and $\begin{cases} b(0) = x(0) - x(3) \\ b(1) = x(1) - x(2) \end{cases}$

DST-VI:

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \end{bmatrix} = \begin{bmatrix} 29 & 55 & 74 & 84 \\ 74 & 74 & 0 & -74 \\ 84 & -29 & -74 & 55 \\ 55 & -84 & 74 & -29 \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \end{bmatrix}$$

$$= \begin{bmatrix} 29x(0) + 55x(1) + 74x(2) + 84x(3) \\ 74x(0) + 74x(1) \qquad\qquad - 74x(3) \\ 84x(0) - 29x(1) - 74x(2) + 55x(3) \\ 55x(0) - 84x(1) + 74x(2) - 29x(3) \end{bmatrix} \quad (18)$$

From equations (17) and (18), the number of operations for 1D and 4 outputs is determined and given in table 5:

Table 5. Computational complexity of the DCT-II and the DST-VI

Transform	ADD	MUL	Total
DCT-II	**8**	**6**	**14**
DST-VI	**11**	**13**	**24**

The above results reveal that using the DCT-II decreases the number of operations by 41.67% and the number of multiplications by 53.85%. This is a big saving knowing that the multipliers are great hardware resources and high energy consumers.

IV. Conclusion

In this paper, two sets of tests were carried on 6 video sequences using the HEVC reference model HM.16.3 and the profile main with the configurations All Intra, Low Delay B and Random Access. For the first group of experiments, the DST-VI was used as the 4x4 intra predicted luminance blocks transform. Then, the DST-VI was replaced by the DCT-II and the tests re-conducted under the same conditions. The comparison of the results, using the Bjøntegaard average delta PSNR and the average delta rate have shown that using the DCT-II yields to a slightly lower performance. Indeed, for the main profile, there is an insignificant loss in the quality of 0.0248 dB and a loss in the compression gain represented by an increase of 0.5182 % in the bitrate. Let us recall that the introduction of the DST-VI targets a 1% bitrate reduction. This value is reached with the All Intra configuration used for editing purposes. As far as the quality is concerned, a subjective assessment is of great importance as the DST-VI is not perfectly orthogonal.

On the other hand, the DST-VI possesses a much higher computational complexity than the DCT-II. Its hardware implementation represents a serious challenge if high throughput and low energy are targeted. These two metrics are of great importance in applications such as wireless sensors networks and portable computing applications.

References

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand ,"Overview of the High Efficiency Video Coding (HEVC) Standard" . *IEEE Trans. on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649-1668, Dec.2012.

[2] ITU TSB (2010-05-21). "Joint Collaborative Team on Video Coding". ITU-T..

[3] ISO/IEC 23008-2:2013 Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding (standard has been revised by: ISO/IEC 23008-2:2015)

[4] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding" . *IEEE Journal on Selected Topics in Signal Processing* , vol. 7, no. 6, pp. 1001 – 1016, Dec. 2013.

[5] Gerhard Tech; Krzysztof Wegner; Ying Chen; Sehoon Yea (2015-02-18). "3D-HEVC Draft Text 7". JCT-3V. Retrieved 2015-02-26.

[6] J. Han, A. Saxena, and K. Rose, " Toward Jointly Optimal Spatial Prediction and Adaptive Transform in Video/Image Coding," *2010 IEEE Inter. Conf. on Acoustic Speech and Signal Processing (ICASSP),* Sep. 2010, pp. 726-729.

[7] C. Yao, Y. H. Tan, Z. Li, and S. Rahardja,"Mode-Dependent Fast Separable KLT for Block-Based Intra Coding," *2011 IEEE International Symposium on Circuit and Systems (ISCAS),* May 2011, pp. 621-624.

[8] Mathias Wien , "High Efficiency Video Coding. Coding Tools and Specification", pp.206-212, Springer-Verlag Berlin Heidelberg 2015.

[9] A. Saxena, F. C. Fernandes and Y. A. Reznik, "Fast transforms for Intra-prediction-based image and video coding", *Data Compression Conference (DCC), IEEE Computer Society,* March 2013, pp.13-22.

[10] K. Ugur and A. Saxena, "Summary report of Core Experiment on intra transform mode dependency simplifications," *ITU-T & ISO/IEC JCTVC J0021*, July 2012.

[11] V. Britanak, K. R. Rao, and P. Yip, "Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations", Oxford, UK: Academic Press - Elsevier, 2007.

[12] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD curves", in VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA, Apr. 2001.

[13] http://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

[14] F. Bossen, "Common HM test conditions and software reference configurations," document JCTVC-L1100 of JCT-VC, Geneva, CH, Jan. 2013.

[15] J.-R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and 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,* vol.22 , no.12, Dec. 2012.

[16] P. K.Meher, S. Y. Park, B. K. Mohanty, K. S. Lim, and C. Yeo," Efficient Integer DCT Architectures for HEVC", *IEEE Transactions on Circuits and Systems for Video Technology*, pp.168-178, Jan. 2014