LOW COMPLEXITY TRANSFORM COMPETITION FOR HEVC

Adrià Arrufat[⋆†], *Pierrick Philippe*[⋆∗], *Kevin Reuzé*[⋆†]

[⋆]Orange Labs
4, Rue du Clos Courtel
35512 Cesson-Sévigné — FRANCE

Olivier Déforges[†]

[†]IETR/INSA
UMR CNRS 6164
35043 Rennes — FRANCE

ABSTRACT

The use of multiple transforms in video coding can lead to substantial bit-rate savings. However, these savings come at the expense of increased coding complexity and storage requirements, which challenge the usability of this approach. In this paper, a systematic procedure is proposed to design low complexity systems making use of transform competition. Multiple trade-offs accommodating the complexity are unveiled and it is demonstrated that they can keep a certain level of performance. Compared to the HEVC standard, some of them provide bit-rate savings around 2% with a 50% increase in the encoding time, using less than 4 kB of extra ROM and no added decoding complexity.

***Index Terms*—** DTT, transform, video, coding, HEVC

1. INTRODUCTION

The use of multiple transforms can improve state-of-the-art video coders, such as the High Efficiency Video Coding (HEVC) standard [1]. Work introduced in [2] evidenced the interest to extend HEVC coding possibilities with multiple transforms in addition to the existing block sizes and prediction choices. Transforms used to build the system were based on the rate-distortion optimised transform (RDOT) design method presented in [3]. The studies carried out in [4] extended the mode dependent directional transform (MDDT) [5] by providing a set of transforms in each intra prediction mode (IPM). This technique, named mode dependent transform competition (MDTC), leads to BD-rate savings of over 7% for non-separable transforms and 4% for separable transforms, relative to HEVC.

However, the encoder complexity was affected by a factor of 10 and the decoding time increased by at least 5%. The amount of storage requirements for the transforms was also reported as significant: more than 300 kB were required. Consequently, the main motivation of this article is to provide a low complexity alternative to the MDTC system while keeping a reasonable level of performance.

[⋆]This author performed the work while at B-COM, 1219 avenue Champs Blancs, 35510 Cesson-Sévigné — FRANCE

The objective is to conceive a system using no more than 16 kB to store the transforms and a decoding time equivalent to HEVC, while accommodating the encoding complexity.

2. SYSTEM SIMPLIFICATION

In order to conceive a low complexity system that makes use of transform competition, some concessions are required.

This section proposes simplifications to the areas leading to significant complexity increases in transform competition systems, especially: the storage requirements, the decoding time and the encoding time.

The first approach for simplification is to avoid non-separable transforms, even though they provide notably better performances than their separable counterparts [4, 6]. These transforms can definitively not accommodate a system with a storage limitation in a range below 16 kB.

Two propositions are presented below to further reduce the storage requirements of transforms and the coding complexity.

2.1. ROM reduction exploiting prediction residual symmetries

A disadvantage of the MDTC system proposed in [4] was the large number of different transforms that needed to be stored, since each IPM used a different set.

In order to limit the ROM impact of the transforms, the proposed system takes advantage of geometrical symmetries existing amongst HEVC IPMs: within the 35 IPMs in HEVC, symmetries in prediction residuals can be observed for directional modes (2–34).

It can be stated that residuals issued from the first half (2–18) are closely related to the transposed version of residuals issued from the second half (18–34). Symmetries can be taken one step beyond and be applied inside the first half. These intra prediction residuals are symmetrical with respect to IPM 10: residuals from IPMs above 10 can be related to those below 10 by applying a horizontal mirroring (top-bottom) or reflection. Since the second half is related to the first half transposed, this property applies around IPM 26 in a similar fashion, through a vertical mirroring (left-right).

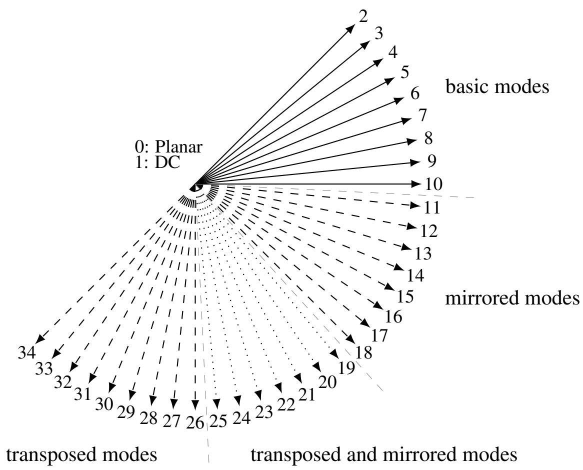

Fig. 1: The 35 IPMs in HEVC and their symmetries

IPM	$(\cdot)^T$	**Mirror**	$(\cdot)^T$ **+ mirror**
0	—	—	—
1	—	—	—
2	34	18	—
3	33	17	19
4	32	16	20
5	31	15	21
6	30	14	22
7	29	13	23
8	28	12	24
9	27	11	25
10	26	—	—

Table 1: Symmetrical relations among IPM residuals

Figure 1 contains a simplified version of the HEVC IPMs, which illustrates the proposed symmetrical relations.

When both symmetries are exploited, the number of transforms is no longer affected by a factor of 35 (the number of modes), but by a factor of 11 (the number of basic modes), as presented in table 1. Using these symmetries between IPMs implies manipulating the residual before the transform stage. Before transforming a residual, the encoder will consider the following cases, depending on the IPM:

$$\mathbf{X} = \begin{cases} \mathbf{A}\,\mathbf{x} & 0 \leq \text{IPM} \leq 10 \\ \mathbf{A}\ \zeta\,\mathbf{x} & 11 \leq \text{IPM} \leq 18 \\ \mathbf{A}\ \zeta\,\mathbf{x}^T & 19 \leq \text{IPM} \leq 25 \\ \mathbf{A}\,\mathbf{x}^T & 26 \leq \text{IPM} \leq 34 \end{cases}$$

Where $\mathbf{A}$ is a transform designed for the basic IPM set, $\mathbf{x}$ is the current residual and ζ represents the horizontal mirroring operator (top-bottom). Mirroring and transposing operations are used to make residuals compatible with the transforms learnt for the basic IPM set. These operations are only different ways of re-arranging the residual pixels in a consistent way, which come at no computational cost.

2.2. Simplification using fast trigonometric transforms

Work in [6] highlighted the appropriateness of the RDOT design method over the Karhunen-Loève transform (KLT) for video coding. However, for the sake of simplicity, an additional family of transforms is considered in this work: the discrete trigonometric transforms (DTTs).

DTTs are orthogonal transforms based on trigonometric functions. This family of transforms consists of 8 types (I to VIII) of discrete cosine transforms (DCTs) and discrete sine transforms (DSTs) [7].

Historically, the DCT-II has been the *de facto* standard transform for image and video coding applications. Recently, other transforms from the DTT family are starting to arise the interest in video coding applications: The DST-VII is used in HEVC for 4×4 intra prediction luma residuals and the DST-III has been proposed for inter-layer prediction residuals in scalable video coding [8].

The interest of DTTs is motivated by the existing fast algorithms for transform implementation, which are notably less complex than a full matrix multiplication required by generic block transforms. The algorithmic complexity for a DTT is in the order of $N \log_2(N)$, where N stands for the transform size [9]. Nevertheless, these transforms are a restrained subset of the orthogonal transform class, as such their performance is expected to be lower than that of RDOTs.

Since DTTs coefficients can be computed using an analytical formula their storage requirement is negligible. However, a scanning matrix is needed per transform in each IPM, as in the separable RDOTs [4]. This scanning matrix is required to sort the transformed coefficients in a globally monotonic order to ease the lossless coding stage. A result, the storage amount necessary per transform is N^2 bytes.

For the RDOTs, the memory required is $3N^2$ bytes per transform, as the vertical and horizontal transforms are stored along with a dedicated scanning pattern.

3. MDTC DESIGN METHOD

The RDOT equation, described in (1) has proved to be a good way of designing transforms that offer a balance between distortion introduced by quantisation and the sparsity of transformed coefficients. This equation was first used to design RDOTs in [3] and later in the MDTC system from [4].

$$\overset{\mathbf{A}_v,\mathbf{A}_h}{opt} = \arg\min_{\mathbf{A}_v,\mathbf{A}_h} \sum_{\forall i} \min_{\mathbf{c}_i} \left(\|\mathbf{x}_i - \mathbf{A}_v^T \mathbf{c}_i \mathbf{A}_h\|_2^2 + \lambda \|\mathbf{c}_i\|_0 \right) \quad (1)$$

Where $\mathbf{x}_i$ is a residual block from a training set, $\mathbf{c}_i$ are the quantized transformed coefficients using $\mathbf{A}_h$ and $\mathbf{A}_v$, the horizontal and vertical transforms, respectively. Transforms are chosen orthogonal, to guarantee the energy preserving property and to be invertible. The constraint in the cost function is the ℓ_0 norm of the coefficients, i.e. the number of non-zero coefficients. The Lagrange multiplier λ depends on the quantisation accuracy applied to the transformed coefficients [3].

Nonetheless, this equation cannot only be used for transform design, but also to measure the appropriateness of a given transform to compactly represent a residual in the rate-distortion plane. The measure $\delta_{i,k}$ for a residual $\mathbf{x}_i$ and a pair of horizontal and vertical transforms $\mathbf{A}_{hk}, \mathbf{A}_{vk}$ is given by:

$$\delta_{i,k} = \|\mathbf{x}_i - \mathbf{A}_{vk}^T \mathbf{c}_i \mathbf{A}_{hk}\|^2 + \lambda \|\mathbf{c}_i\|_0 \quad (2)$$

By using the metric described in (2), one can evaluate the performance of a transform on a residual and define a set of residuals for the transform which give the best rate-distortion.

input : Residuals $\mathbf{x}$ from a given intra prediction mode
output: Set of N separable pairs of transforms $\mathbf{A}_{hk}, \mathbf{A}_{vk}$

Initial random classification into $1+M$ classes

while *!convergence* **do**
  for $m = 1$ **to** M **do**
    Use DTTs or, based on (1), learn a RDOT on Class_m
  end
  foreach *block* $\mathbf{x}$ **do**
    for $m = 0$ **to** M **do**
      $\delta_m = \|\mathbf{x} - \mathbf{A}_{vm}^T \mathbf{c} \mathbf{A}_{hm}\|^2 + \lambda \|\mathbf{c}\|_0$
    end
    $m^* = \arg\min_m(\delta_m)$
    Class_{m^*}.append ($\mathbf{x}$)
  end
end

Algorithm 1: Multiple transform design and classification

Two systems are targeted in this publication leading to different approaches for the transform selection process:

- RDOT-based system, where the transforms are learnt considering the residuals belonging to the class. This process implies an iterative algorithm based on a Singular Value Decomposition as described in [3].
- DTT-based system, where given a set of residuals the appropriate pair of horizontal and vertical among the DTT set is selected based on the metric of (2).

Algorithm 1 details the learning process in which M transforms are designed in addition to the default HEVC transform. For DTTs, since 8 types of DCT and DST are considered, 256 combinations are possible when combining vertical and horizontal transforms. As in vector quantization design methods, initial conditions, here the initial clustering, need to be carefully set up. Multiple attempts involving clustering based on geometrical properties of the residual plus random distribution were used for the learnings in this paper. This topic is still under consideration as a possible improvement to the performances presented here.

4. EXPERIMENTAL RESULTS

The previous section has presented a metric which has served to design multiple RDOTs adapted to a set of residuals. In addition, a selection procedure of the best combination of horizontal and vertical DTT has also been described.

In this section, a low complexity system has been chosen amongst several configurations. The maximum number of transforms used in each IPM has been set to 8 to limit the system complexity and the ROM footprint. The amount of ROM, expressed in bytes, required to store the transforms and the scanning matrices computes as follows, where M is the number of transforms per IPM and N the size of the transform:

- for a RDOT system, ROM = $11 \times 3 \times M \times N^2$ bytes are requested, as 11 independent transforms sets are considered for which two core transforms (horizontal and vertical) are considered along with a scanning pattern.
- for a DTT system, ROM = $11 \times M \times N^2$ bytes are requested, as 11 independent transforms sets are considered for which a scanning pattern needs to be stored.

All proposed systems have been designed using a learning set formed by residuals derived from an HEVC encoding of the *Tears of Steel* sequence [10], providing over 140 million 4×4 residuals and over 340 million 8×8 residuals. Transform competition has only been enabled for these transform unit (TU) sizes.

As reported in [4], for a given block size and IPM, the encoder selects the best transform among available ones using a rate-distortion selection method. This notably increases the encoding time as reported below. Once the best transform for a block is selected, its index is signalled to the decoder using a flag plus fixed-length code approach per TU. The flag indicates whether the legacy HEVC transform (DST-VII and DCT-II respectively for 4×4 and 8×8 blocks) is used or not, in which case, the index of the transform is explicitly signalled.

Results reported here use the coding configurations established by the Joint Collaborative Team on Video Coding (JCT-VC) standardisation group [11]. They consist in encoding the sequences at four quantisation parameter (QP) points and computing the average bit-rate reduction [12].

Table 2 compares the performances of the DTT and RDOT systems. For each transform combination, the ROM requirement is provided, with the encoding complexity and the average encoding time on the HEVC test set in intra coding.

The decoding time is nearly independent from the number of transforms. It only depends on the type of transform: RDOTs cause an increase in the decoding time of around 3%, whereas DTTs lead to a marginal impact, due to their fast algorithms.

To ease the decision of the system that offers a suitable trade-off, figure 2 maps the bit-rate savings of each system

Tr. size Num. 4×4	Num. 8×8	DTT ROM (kB)	DTT Compl. (%)	DTT BD-r (%)	RDOT ROM (kB)	RDOT Compl. (%)	RDOT BD-r (%)
1	0	0.17	109.54	-0.50	0.52	125.75	-0.82
2	0	0.34	114.32	-0.69	1.03	138.67	-1.08
4	0	0.69	123.78	-0.80	2.06	164.21	-1.36
8	0	1.38	142.52	-0.84	4.13	214.79	-1.48
0	1	0.69	109.18	-0.69	2.06	124.79	-0.72
1	1	0.86	119.03	-1.14	2.58	151.38	-1.46
2	1	1.03	124.01	-1.33	3.09	164.83	-1.72
4	1	1.38	133.44	-1.44	4.13	190.29	-1.99
8	1	2.06	152.08	-1.47	6.19	240.63	-2.09
0	2	1.38	114.68	-1.01	4.13	139.63	-1.12
1	2	1.55	124.60	-1.41	4.64	166.42	-1.82
2	2	1.72	129.57	-1.59	5.16	179.83	-2.07
4	2	2.06	138.87	-1.69	6.19	204.96	-2.31
8	2	2.75	157.67	-1.75	8.25	255.71	-2.45
0	4	2.75	125.59	-1.20	8.25	169.08	-1.46
1	4	2.92	135.35	-1.58	8.77	195.46	-2.12
2	4	3.09	140.28	-1.78	9.28	208.75	-2.37
4	**4**	**3.44**	**149.65**	**-1.92**	10.31	234.04	-2.59
8	4	4.13	168.64	-1.94	12.38	285.33	-2.74
0	8	5.50	147.08	-1.30	16.50	227.13	-1.82
1	8	5.67	156.77	-1.67	17.02	253.29	-2.43
2	8	5.84	161.71	-1.86	17.53	266.63	-2.67
4	8	6.19	171.17	-1.99	18.56	292.17	-2.90
8	**8**	6.88	190.02	-2.03	**20.63**	**343.04**	**-3.06**

Table 2: DTT and RDOT systems relative to HEVC

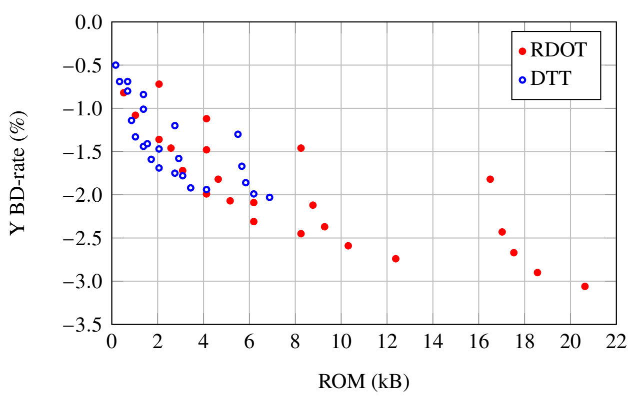

Fig. 2: Different ROM – BD-rate trade-offs

with its storage requirements. Figure 3 displays the relation between bit-rate savings and encoding complexity.

Limiting the ROM to 16 kB and the encoding complexity to 150% of that of HEVC, the system that offers the best trade-off is the one composed by 4 DTTs for both 4×4 and 8×8 blocks. For this system, apart from the legacy HEVC transforms, the most frequently selected DTT is the DCT-IV which represents more than 50% of the transforms for 4×4 TUs. For the 8×8 TUs, the combination of vertical and horizontal transforms based on the DCT-IV and DST-VII represent 61% of the usage. With this system, bit-rate savings of 1.92% can be obtained, at no impact on the decoding time and 3.44 kB of ROM. At almost 3.5 times the complexity of HEVC, an improvement of 3.06% can be obtained with a RDOT-based system using 8 transforms for both 4×4 and 8×8 TUs, with storage requirements of 20.63 kB.

Table 3 contains the detailed performances for these two systems on the HEVC test set for all intra (AI) and random access (RA) coding configurations.

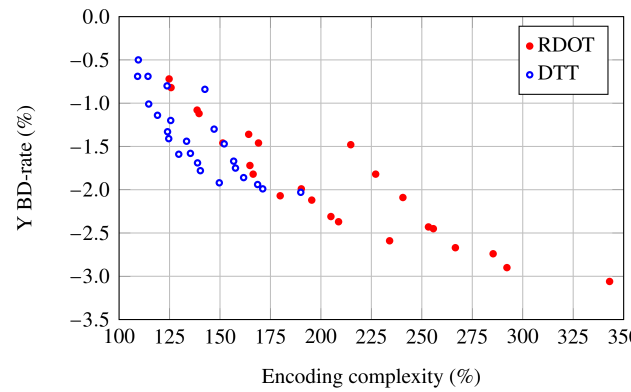

Fig. 3: Different Complexity – BD-rate trade-offs

	Sequence	DTT: 4–4 Y BD-rate (%) AI	RA	RDOT: 8–8 Y BD-rate (%) AI	RA
Class A (2560×1600)	NebutaFestival	-0.78	-0.07	-1.12	-0.11
	PeopleOnStreet	-2.56	-0.91	-4.17	-1.57
	SteamLocTrain	-0.58	0.27	-0.67	0.02
	Traffic	-2.44	-1.84	-4.07	-3.08
Class B (1920×1080)	BasketballDrive	-1.16	-0.17	-1.98	-0.50
	BQTerrace	-1.89	-1.11	-2.46	-1.66
	Cactus	-2.41	-1.40	-3.21	-2.05
	Kimono1	-0.68	-0.40	-1.09	-0.64
	ParkScene	-2.75	-1.75	-3.52	-2.49
Class C (832×480)	BasketballDrill	-2.32	-1.70	-2.83	-2.40
	BQMall	-1.95	-1.12	-3.51	-2.12
	PartyScene	-2.24	-1.62	-3.59	-2.59
	RaceHorses	-2.41	-0.96	-2.82	-1.18
Class D (416×240)	BasketballPass	-1.72	-0.69	-3.05	-1.38
	BlowingBubbles	-2.12	-1.39	-3.26	-2.15
	BQSquare	-2.10	-1.38	-3.78	-2.32
	RaceHorses	-2.35	-0.87	-2.89	-1.13
Class E (1280×720)	FourPeople	-2.47	-2.54	-3.56	-4.08
	Johnny	-1.46	-1.97	-2.30	-3.20
	KristenAndSara	-1.71	-2.18	-2.67	-3.54
Class F (various resolutions)	BasketDrillText	-2.50	-1.78	-3.64	-2.61
	ChinaSpeed	-1.55	-1.15	-3.81	-2.38
	SlideEditing	-1.81	-1.89	-4.40	-4.51
	SlideShow	-2.18	-2.23	-4.90	-4.94
All sequences	**Overall**	**-1.92**	**-1.29**	**-3.06**	**-2.19**

Table 3: Compression gains for two proposed configurations

5. CONCLUSION

MDTC exhibits promising bit-rate savings over HEVC. In this work, a systematic simplification of its main flaws (storage requirements and coding complexity) has been proposed. Depending on the working point, several solutions exist at different performance levels: about 2% can be obtained with an increase of 50% in the encoding complexity, and more than 3% at 3 times the complexity of HEVC with a 3% increase in the decoding time. Further work will necessarily focus on the encoder side to design fast decision mechanisms and reduce the complexity in the selection of the best transform. Also, the transform learning algorithm is subject to improvement since local minimums during the learning phase are often encountered.

6. REFERENCES

[1] G. Sullivan, J. Ohm, W. J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] A. Arrufat, P. Philippe, and O. Déforges, “Rate-distortion optimised transform competition for intra coding in HEVC,” in *Visual Communications and Image Processing Conference, 2014 IEEE*, Dec 2014, pp. 73–76.

[3] O. Sezer, O. Harmanci, and O. Guleryuz, “Sparse orthonormal transforms for image compression,” in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008, pp. 149–152.

[4] A. Arrufat, P. Philippe, and O. Déforges, “Mode-dependent transform competition in HEVC,” in *Image Processing, 2015. ICIP 2015. 22th IEEE International Conference on*, 2015.

[5] C. Yeo, Y. H. Tan, Z. Li, and S. Rahardja, “Mode-dependent transforms for coding directional intra prediction residuals,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 4, pp. 545–554, April 2012.

[6] A. Arrufat, P. Philippe, and O. Déforges, “Non-separable mode dependent transforms for intra coding in HEVC,” in *Visual Communications and Image Processing Conference, 2014 IEEE*, Dec 2014, pp. 61–64.

[7] K. R. Rao and P. Yip, *The Transform and Data Compression Handbook*. Boca Raton, 2001.

[8] L. Guo, M. Karczewicz, and J. Chen, “Transform for inter-layer prediction residues in scalable video coding,” in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, July 2013, pp. 1–6.

[9] M. Puschel and J. Moura, “Algebraic Signal Processing Theory: Cooley–Tukey Type Algorithms for DCTs and DSTs,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 4, pp. 1502–1521, April 2008.

[10] (2014) Tears of Steel. [Online]. Available: http://www.tearsofsteel.org

[11] F. Bossen, “Common test conditions and software reference configurations,” ITU-T, Geneva, Switzerland, Tech. Rep. JCT-VC I1100, May 2012.

[12] G. Bjøntegaard, “Calculation of average PSNR differences between RD-curves,” ITU-T SG16/Q6 VCEG, Austin, Texas, Tech. Rep. VCEG-M33, April 2001.