

Approximate 4x4 1D-DCT Hardware Architecture Using Imprecise LOA Adder

Mateus Leme, Luciano Braatz, Luciano Agostini, Marcelo Porto

Video Technology Research Group (ViTech)

Federal University of Pelotas (Ufpel)

{mtleme, la.braatz, agostini, porto}@inf.ufpel.edu.br

Abstract—Mobile devices that process multimedia applications are becoming more and more present, however, these devices are energy restricted. Multimedia applications can benefit from approximate computing to save energy. This work proposes an approximate 4x4 Two-Dimensional Discrete Cosine Transform hardware using the imprecise Lower Part-OR Adder. This approximate hardware architecture is developed using different imprecision levels. An analysis performed using the High-Efficiency Video Coding standard (HEVC) reference software shows this solution having a BD-Rate impact on the coding efficiency between -0.1% and 3.1%. On the other hand, the synthesis results using a 45nm standard cells technology brought a significant power saving, between 12% and 42%.

Keywords—*approximate computing, imprecise adders, transform, DCT, low power design, video coding.*

I. INTRODUCTION

Nowadays there is an increase in devices and streaming services that process digital videos. Video related data reached 64% of all Internet traffic in 2014, and it can reach 82% by 2020 [1]. Video streaming without compression is prohibitive due to the huge amount of data required to represent uncompressed videos. Therefore, techniques to encode the video are applied with the goal to reduce the data that will be transmitted with a reduced loss in quality. The High-Efficiency Video Coding (HEVC) is one of the most recent video coding standards. It can reduce up to 50% of the encoded video size while maintaining the same visual quality when compared to the previous standard H.264/AVC [2].

One of the important tools of the HEVC encoder is the transform step, responsible for data conditioning which improves the encoder compression efficiency. The HEVC standard uses two transform types in the transform step. The Two-dimensional Discrete Cosine Transform (2D-DCT) supports 4x4, 8x8, 16x16 and 32x32 block sizes. Alternatively, the Discrete Sine Transform (DST) can be used to process 4x4 blocks [3].

Many works focus on the improvement of the direct DCT and the Inverse DCT (IDCT) of the HEVC standard. A high-abstraction architectural solution for four-size DCT is proposed in [4], using an exploitation of smaller operation blocks. Therefore, it reduces area and power dissipation. The hardware architecture in [5] improves the area by reusing the hardware to perform the DCT and the inverse DCT (IDCT). A modular architecture to DCT able to process continuously 32 samples

by cycle is presented in [6]. It can process every HEVC transform size blocks.

Due to the energy constraints of mobile devices, energy efficiency becomes a significant concern. However, multimedia applications demand high energy consumption. Thus, the usage of approximate computing is a promising tool to achieve low-power in digital systems. This paradigm is based on the concept of error-tolerant applications, i.e., the applications must be tolerant to the usage of numerically imprecise results. [7]. Video coding, one of the most important multimedia applications, can greatly benefit energy-wise by using approximate computing techniques. This work develops an approximate hardware architecture for the HEVC 4x4 One-dimensional Discrete Transforms (1D-DCT), partially composed by imprecise Lower-Part-OR Adder (LOA) operators.

II. BACKGROUND OF THE HEVC TRANSFORM

The transform in the HEVC standard is applied to the residues generated in the prediction step. The residuals contain the difference between the current block being coded, and the predicted block. The predicted block is composed based on previously encoded video samples in the same frame (intra-prediction) or in already processed frames (inter-prediction).

A. Block Partition and Residual Coding in HEVC Standard

Digital videos are composed of pixels. Pixels are divided into color channel samples, with different formats. The YUV format is composed of three channels, luminance (Y) that carries the brightness information, and two-color channels (U and V) that carry the colors information. The Y channel is the most important channel, having more impact on the image details, while the impact of colors channels on the image is lower since the human visual system has lower sensitivity to color information [2].

The HEVC standard process video data by subdividing it into blocks of 64x64, 32x32 or 16x16 called Coding Tree Units (CTU). The CTUs can be split into Coding Units (CU). The CU can be divided into Prediction Unit (PU) and Transform Unit (TU). Data processed by the transform step are stored in TU blocks, which can assume 32x32, 16x16, 8x8, or 4x4 block size. The transform step converts data from the spatial domain to the frequency domain, which concentrate the data in low-frequency coefficients, thus making subsequent encoding steps more efficient [3].

B. Discrete Cosine Transform

Commonly used in digital image and video processing, DCT is also used by the HEVC standard. However, due to the block format input, the transform used in the HEVC standard is the Two-dimensional Discrete Cosine Transform (2D-DCT) [3]. On the other hand, the 2D-DCT design uses a large number of arithmetic operators [6]. Since the 2D-DCT is separable, it can be implemented through two 1D-DCT, which has a smaller number of arithmetic operators [7].

The 1D-DCT processes data extracted from the lines of the blocks. One way of processing the 1D-DCT is using a butterfly-based structure. This structure explores similarity among the different size DCT coefficients. Thus, each input is decomposed in even and odd parts, that together build up a Partial Butterfly (PB). The odd part is processed by actual size block specific equations and the even part can be processed using the immediately lower size block equations. Therefore, it is possible to reuse parts of the hardware to process different inputs.

The transformed residues tend to concentrate information on the low-frequency coefficients. It is important to highlight that this operation does not insert data loss in the encoding. The original values can be restored by the usage of the 2D-IDCT (Two Dimension Inverse Discrete Cosine Transform).

C. 4x4 DCT Equations

The 4x4 1D-DCT can be processed by equations in (1), where E_0 , E_1 , O_0 , and O_1 were the butterfly outputs, given by (2) and X_n are the inputs of the 1D-DCT and Y_n the 1D-DCT outputs.

$$\begin{cases} Y_0 = 64E_0 + 64E_1 \\ Y_1 = 36O_0 + 83O_1 \\ Y_2 = 64E_0 - 64E_1 \\ Y_3 = 83O_0 - 36O_1 \end{cases} \quad (1)$$

$$\begin{cases} E_0 = X_0 + X_3 \\ E_1 = X_1 + X_2 \\ O_0 = X_0 - X_3 \\ O_1 = X_1 - X_2 \end{cases} \quad (2)$$

III. APPROXIMATE DCT ARCHITECTURE DESIGN

Error-tolerant digital circuits, such as video encoding applications, benefit from the throughput increase, circuit area and power dissipation savings of approximate computing. Approximate computing can be used in the HEVC transform by using imprecise arithmetic operators since the arithmetic operators are the main responsible for the power dissipation in digital circuits [8]. There are many imprecise operators in the literature, including the Error-Tolerant Adder (ETA) [8], Generic Accuracy Configurable Adder (GA-CA) [9], Lower-Part-OR Adder (LOA) [10], and more. The principal contribution of this work is the use of LOA operator in DCT 4x4 architecture, aiming area-, and power-efficiency.

A. Lower-Part-OR-Adder

The LOA operator presented in Fig.1 divides the sum operation into two parts, precise part and imprecise part. The precise part is composed of an accurate adder and processes the

most significant bits. The imprecise part performs a bitwise OR operation, at the least significant bits. The carry-in of the precise part is created by an AND operation on the most significant bit of the imprecise part [10].

B. DCT 4x4 Operation

The diagram shown in Fig. 2 (a) illustrates the 4x4 1D-DCT diagram block. The 1D-DCT is processed by calculating the butterfly, multiplying the result by the transform constants (represented by the *Multiplier Step* block in Fig. 2(a) and Fig. 2.(b)), combining the multiplication products and adjusting the bit length by adding an offset (represented by the *Multiplier Step* block in Fig. 2(a)) and a final right shift. The data bit width before each operation is 9-, 10-, 17-, and 18-bits for the first 1D-IDCT, and 16-, 17-, 24-, and 25-bits for the second 1D-IDCT, respectively.

C. Approximate Architecture

The approximate 4x4 1D-DCT architecture proposed in this work consists of the removal of the offset adjustment step and the replacement of precise adders by the LOA operator. The developed approximate 4x4 1D-DCT is shown in the design in Fig. 2 (b).

IV. HARDWARE AND CODING EFFICIENCY RESULTS

To evaluate the impact of coding efficiency generated by the application of the proposed approximate architecture, the LOA operator has been described in C++ and applied in HEVC standard reference software, the HM in version 16.17 [11]. Table I shows the five different levels of imprecision were considered for evaluation, along with the number of imprecise bits in the LOAs of the 1st 1D IDCT and 2nd 1D IDCT.

The results of the coding efficiency are obtained by the encoding of six video sequences recommended in Common Test Conditions (CTC) [12]. The test was divided into two classes of resolution, 1080p (1920x1080 pixels) and 4k (3840x2160 pixels). Each of the classes contains three videos, *BasketBallDrive*, *ParkScene*, and *Kimono* in 1080p class and *Suzie*, *Foreman*, and *Jockey* in 4k class. The configuration used

TABLE I- IMPRECISION LEVELS

Level	IMP1	IMP2	IMP3	IMP4	IMP5
1st 1D-DCT	1	3	5	7	9
2nd 1D-DCT	3	5	7	9	11

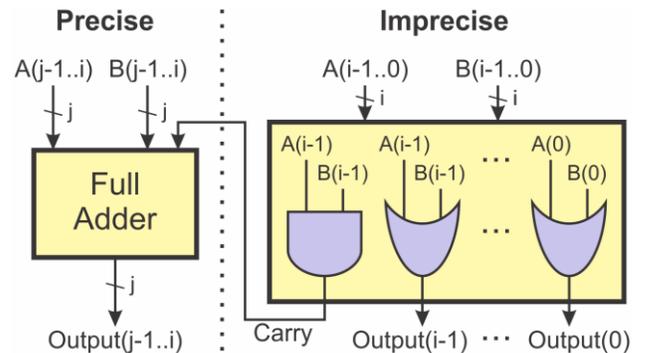


Fig. 1. Lower-Part-OR Adder Structure

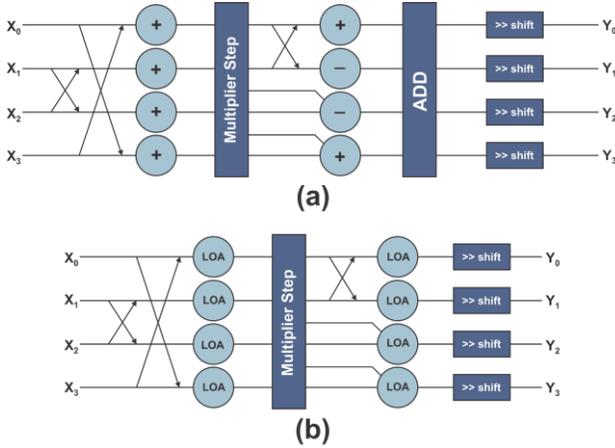


Fig. 2.(a) 1D-DCT 4x4 Operation and (b) Approximate Design of 1D-DCT 4x4.

was the Random Access Main for 1080p and the Random Access Main10 for 4k, using four QP values (22, 27, 32, and 37), as recommended by the CTC. The metric to evaluate the coding efficiency is the Bjøntegaard Distance on bitrate (BD-Rate). The BD-Rate metric shows the bitrate variation required on the test encoding to maintain the same video quality as the reference encoding. A negative BD-Rate represents a reduction on the size of the encoded video, while, maintaining the same video quality.

A. Coding Efficiency Results

The BD-Rate results were analyzed considering YUV results, which consists of a weighted average between the Y, U, and V channels. The YUV is used to cover the heterogeneous coding efficiency degradation among these channels.

The results presented in Table II shows an expected increase of the impact on coding efficiency while the imprecision level rises. According to the video characteristics, the use of imprecision leads to different impacts on coding efficiency. In 1080p videos, the *Kimono* video presents low BD-Rate impact in all of the imprecision levels (up to 0.193%), while the BD-Rate impact of *ParkScene* video is much higher (up to 3.105%). Although both videos present camera movement, there are fine details on the *ParkScene* video, while the moving background on the *Kimono* video lacks fine details. Since 4x4 blocks are less used on 4k videos, the BD-Rate

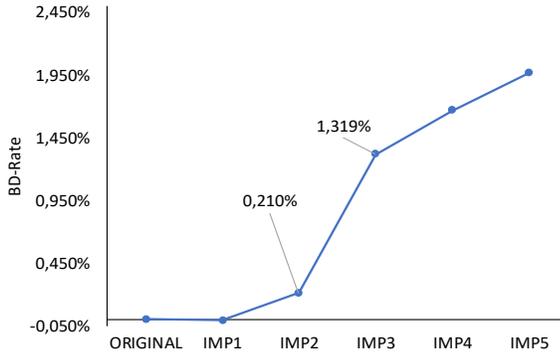


Fig. 4. Average BD-Rate 1080p class

TABLE II- COMPRESSION EFFICIENCY RESULTS

Resolution	BD-rate Increase (%)				
	IMP1	IMP2	IMP3	IMP4	IMP5
Kimono	-0.084	0.002	0.130	0.127	0.193
ParkScene	0.065	0.362	2.163	2.642	3.105
BasketBall	-0.003	0.267	1.665	2.247	2.611
1080p Average	-0.007	0.21	1.319	1.672	1.970
Suzie	0.136	0.044	0.007	0.066	0.369
Jockey	0.104	0.137	-0.125	0.021	0.441
Foreman	-0.105	0.044	-0.036	0.211	0.408
4k Average	0.045	0.075	-0.054	0.099	0.406
Total Average	0.019	0.143	0.633	0.886	1.188

impacts on such videos are smaller and more uniform. For instance, the IMP5 imprecision level produced a maximum BD-Rate impact of 0.441%. There are some cases of negative occurrences of BD-Rate which indicates an improvement in coding efficiency. In such cases the encoder opted for different size blocks, resulting in increase of coding efficiency.

The 1080p average results show an increase of BD-Rate proportional to the imprecision level, ranging from -0.007% to 1.97%. Fig. 4 shows a BD-Rate increase of 630% between two imprecision levels, 0.21% to 1.319% in the IMP2 to IMP3. This proves that increasing the imprecise level by one can result in a big loss in coding efficiency. Thus a thorough evaluation should be done to achieve the best imprecision point where coding efficiency is not critically worsened.

Therefore, the coding efficiency reduction is proportional to the imprecision increase inserted by the LOA operator, although such reduction is shown within an acceptable range.

B. Synthesis and Power Analysis Results

The approximate architecture of 4x4 1D-DCT presented in the previous section was described in VHDL in six different versions, the original version (ORG) and the five versions using the LOA operator with different levels of imprecision. The architectures were synthesized for ASIC the 45nm @ 1.1V/25°C Nangate standard cell library. Cadence Encounter RTL Compiler tool was used for the syntheses, configured to secure the fidelity of the described hardware.

The syntheses results are shown in Table III. The power dissipation and area usage decreased as the imprecision increased. It is expected due to the removal of the offset step and the exchange of accurate operators for LOA operators. Regarding the impact of the area, is possible to note a decrease of 28% in the higher imprecision level in IMP5. A decrease in power dissipation with the increase imprecision level is presented in synthesis results. The power dissipation in the higher imprecision level (IMP5) achieve a decrease of 42% compared to the original architecture. The directly

TABLE III- ARCHITECTURE RESULTS

Parameter	ORG	Variation (%)				
		IMP1	IMP2	IMP3	IMP4	IMP5
Area (μm^2)	2626	-14	-18	-18	-21	-28
Frequency (MHz)	467.5	+7	+16	+24	+25	+20
Power (mW)	1.873	-12	-19	-24	-32	-42

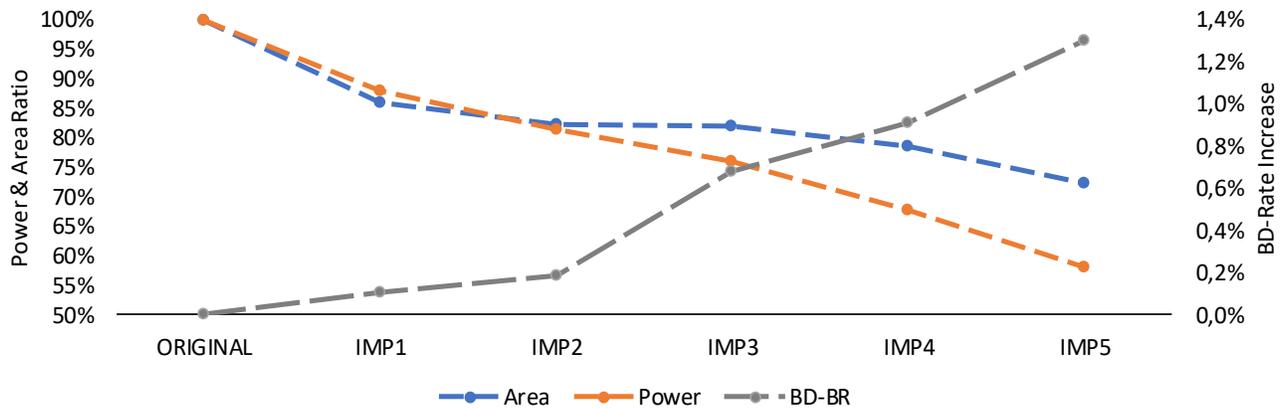


Fig.5. Coding Efficiency and Hardware Results Compared to the Original

proportional relation is noted about the operation frequency behavior and the imprecision level increases, i.e., as the increase of imprecision level the operation frequency also increases. The increase of operation frequency reached 25% in IMP4 level. The minor imprecision level (IMP1) reach 14% of reduction in power dissipation and 12% of area reduction, when compared with original (ORG) DCT architecture. With the hardware performance gains achieved with the proposed approximate solution, it may be concluded that it is a promising strategy for reducing energy consumption.

C. BD-Rate and Hardware Impact Relation

Fig. 5 presents a relation between hardware improvements and the coding efficiency degradation. There is an upward trend of coding efficiency degradation with the increase of imprecision level. On the other hand, a downward trend is present in area values and power dissipation with increasing imprecision level. Assuming 100% as the value of the hardware resources used from original architecture, in the fourth imprecision level (IMP4) the usage of area reduced to 79% and the power dissipation reduced to 68%, although the BD-Rate increased in 0.91%.

V. CONCLUSIONS

This work presented five versions of an approximate 4x4D-DCT using imprecise LOA operators. It proposed the usage of LOA operator in order to reduce the power dissipation. A coding efficiency evaluation was performed using the proposed approximate DCT 4x4 hardware. It showed that this proposal resulted in a negligible impact when compared to the synthesis results which have brought considerable improvements.

The 4x4 DCT proved to be tolerant to imprecision according to the video coding efficiency results. In addition, the use of imprecision has a positive impact on hardware architecture by decreasing area, power, and increasing operating frequency. As the results of this work were very positive, in the future works will be expanded the analyses for all DCT block sizes.

ACKNOWLEDGMENTS

The authors would like to thanks Cnpq and Fapergs for the financial support, which allows the development of this work.

REFERENCES

- [1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology 2016-2021"2018.[Online].Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html>.
- [2] G. J.Sullivan, J. R. Ohm, W. J. Hanand T. Wiegand. 2012. Overview of the High efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)* 22, 12 (2012), 1649 – 1668.
- [3] High Efficiency Video Coding document ITU-T H.265/ISO/IEC 23008-2 HEVC, 2013.
- [4] T. T. T. Do, Y. H. Tan and C. Yeo. 2014. High-throughput and low-cost hardware oriented integer transforms for HEVC. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. IEEE, Paris, France, 2105-2109.
- [5] M. Budagavi and V. Sze. 2012. Unified forward+inverse transform architecture for HEVC. In *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP)*. IEEE, Orlando, FL, USA, 209-212.
- [6] J. Goebel, G. Paim, L. Agostini, B. Zatt and M. Porto. 2016. An HEVC multisize DCT hardware with constant throughput and supporting heterogeneous CUs. In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, Montreal, Canada, 2202-2205.
- [7] J. Han, M. Orshansky. "Approximate computing: an emerging paradigm for energy-efficient design", in *IEEE European Test Symposium*, pp. 1-6, 2013.
- [8] S. Dutt, S. Nandi, G. Trivedi. "A comparative survey of approximate adders", in *IEEE International Conference Radioelektronika*, pp. 61-65, 2016.
- [9] N. Zhu, W. Goh, G. Wang, K. Yeo. "Enhanced low-power high-speed adder for error-tolerant application", in *IEEE International SOC Design Conference*, pp. 323-327, 2010.
- [10] H. Mahdiani, A. Ahmadi, M. Fakhraie, C. Lucas. "Bio-inspired imprecise computational blocks for efficient VLSI implementation of soft-computing applications", in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 4, pp. 850-862, 2010.
- [11] HEVC Test Model, HM 16 [Online] Available <https://hevc.hhi.fraunhofer.de/svn/svnHEVCSoftware/tags/HM-16.7>
- [12] F. Bossen. "Common test conditions and software reference configurations", document JCTVC-L1100 of JCT-VC, 2013.