

VLSI Architecture for High-Throughput Implementation of Lifting 3D-DWT

N.Vairamani, A.Taksala Devapriya, V.Muthukumar

Abstract—In this paper, we present a throughput-scalable parallel and pipeline architecture for high-throughput computation of multilevel three-dimensional discrete wavelet transform (3-D DWT). The computation of 3-D DWT for each level of decomposition is split into three distinct stages, and all the three stages are implemented in parallel by a processing unit consisting of an array of processing modules. The processing unit for the first level decomposition of a video stream of frame-size ($M \times N$) consists of $Q/2$ processing modules, where Q is the number of input samples available to the structure in each clock cycle. The processing unit for a higher level of decomposition requires $1/8$ times the number of processing modules required by the processing unit for its preceding level. For J level 3-D DWT of a video stream, each of the proposed structures involves J processing units in a cascaded pipeline. The proposed structures have small output latency, and can perform multilevel 3-D DWT Computation with 100% hardware utilization efficiency. The throughput rate of proposed structures is $Q/7$ time higher than the best of the corresponding existing structures. Interestingly, the proposed structures involve a frame-buffer of $O(MN)$ while the frame-buffer size of the existing structures is $O(MNR)$. Besides, the on-chip storage and the frame-buffer size of the proposed structure is independent of the input-block size and this favours to derive highly concurrent parallel architecture for high-throughput implementation. The overall area-delay products of proposed structure are significantly lower than the existing structures. The throughput rate of the proposed structure can easily be scaled without increasing the on-chip storage and frame-memory by using more number of processing modules; and it provides greater advantage over the existing designs for higher-frame rates and higher input block-size. The full-parallel implementation of proposed scalable structure provides the best of its performance.

Index Terms — 3-D DWT, Discrete wavelet transform, Lifting, Parallel and pipeline architecture

1 INTRODUCTION

The discrete wavelet transform (DWT) is widely used due to its remarkable advantage over the unitary transforms like discrete Fourier transform (DFT), discrete cosine transform (DCT) and discrete sine transform (DST) for various applications due to its multiple time-frequency resolution. DWT of different dimensions has emerged as a powerful tool for speech and image coding in recent years. The 3-dimensional (3-D) DWT is found to provide superior performance in video compression by eliminating the temporal redundancies within the video sequences for motion compensation. Apart from that, 3-D DWT has been used popularly for the compression of 3-D and 4-D medical images, volumetric image compression, and video watermarking. The multidimensional DWTs are particularly more computation intensive and, therefore, require to be implemented in VLSI systems for real-time applications.

2 DISCRETE WAVELET TRANSFORM

The wavelet transform is computed separately for different segments of the time-domain signal at different frequencies. Multi-resolution analysis: analyzes the signal at different frequencies giving different resolutions. Multi-resolution analysis is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. Good for signal having high frequency components for short durations and low frequency components for long duration, e.g. Images and video frames.

A 'wavelet' is a small wave which has its energy concentrated in time. It has an oscillating wave like characteristic but also has the ability to allow simultaneous time and frequency anal-

ysis and it is a suitable tool for transient, non-stationary or time-varying phenomena

3 PROPOSED ALGORITHM

The lifting scheme was first proposed by Daubechies and Sweldens in 1996. It shows that every finite-impulse response wavelet or filter bank can be factored into a cascade of lifting steps. That means the polyphase matrices for the wavelet filters can be decomposed into a sequence of alternating upper and lower triangular matrices multiplied by a diagonal normalization matrix. The proposed algorithm combines the predictor with the updater. The high-pass signal and the low-pass signal can be calculated in parallel through the two-input/two-output architecture. At the same time, the coefficients of even items are changed by inversion of the factors.

The on chip storage and the frame-buffer contribute more than 90% to the total area of the existing structures. Significant amount of memory-bandwidth and computation-time are also wasted for accessing the external frame-buffer. It is also observed that, the on-chip storage and frame-buffer size is remain independent of the throughput rate. This motivates us to apply concurrent design method and this has two fold advantages:

Using concurrent design method the frame-buffer size could be reduced and the on-chip memory of the 3-D structure can be used more efficiently to calculate multiple outputs per cycle to improve the overall performance of the chip. Hence, it may be considered as an appropriate strategy to design parallel architectures where area can be traded either for time or for

power if faster computation is not required by the application. If high throughput is not required for a given application, then the clock frequency could be reduced and lower operating voltage could be used for reducing the power consumption. Keeping this in mind, we have proposed a parallel architecture for multilevel 3-D DWT. The key ideas used in our proposed approach are:

- 1) To process each decomposition level of 3-D DWT in separate computing blocks in cascaded pipeline structure for concurrent computation of multilevel DWT computation in order to reduce the size of the frame-buffer used for buffering of the sub band components and to maximize hardware utilization efficiency
- 2) The input rows for each level are appropriately folded to meet the desired throughput rate and to achieve 100% HUE of the processing unit.

Using the above approach, we have reduced the frame buffer size, and have obtained Q/7 times higher throughput compared with the best of the existing structures, using the on chip memory of the same order. It is shown that the proposed structure can calculate DWT coefficient of an input video signal of size (M×N×R) in MNR/Q cycles.

The proposed parallel implementation of 3-D DWT structures is of additional advantage, since the size of the frame-buffer could be reduced, and it does not demand for higher on-chip memory and frame buffer for higher input block sizes, which contribute the most of the hardware in the existing designs.

4 MATHEMATICAL FORMULATIONS

The 3D-DWT coefficients of any decomposition level can be obtained from the scaling coefficients of its previous level according to the pyramid algorithm can be given by

$$lll^j(n_1, n_2, n_3) = \sum_{i_1=0}^{k_{h-1}} \sum_{i_2=0}^{k_{h-1}} \sum_{i_3=0}^{k_{h-1}} h_1(i_1)h_2(i_2)h_3(i_3).lll^{j-1}(2n_1-i_1, 2n_2-i_2, 2n_3-i_3) \quad (1)$$

$$llh^j(n_1, n_2, n_3) = \sum_{i_1=0}^{k_{h-1}} \sum_{i_2=0}^{k_{h-1}} \sum_{i_3=0}^{k_{g-1}} h_1(i_1)h_2(i_2)g_3(i_3).lll^{j-1}(2n_1-i_1, 2n_2-i_2, 2n_3-i_3) \quad (2)$$

$$hhl^j(n_1, n_2, n_3) = \sum_{i_1=0}^{k_{g-1}} \sum_{i_2=0}^{k_{g-1}} \sum_{i_3=0}^{k_{g-1}} g_1(i_1)g_2(i_2)g_3(i_3).lll^{j-1}(2n_1-i_1, 2n_2-i_2, 2n_3-i_3) \quad (3)$$

$$n_1 = 0, 1, \dots, \left(\frac{R}{2}\right) - 1, n_2 = 0, 1, \dots, \left(\frac{M}{2}\right) - 1, \text{ and}$$

$$n_3 = 0, 1, \dots, \left(\frac{N}{2}\right) - 1, \text{ where } k_h \text{ and } k_g \text{ are, respectively,}$$

The lengths of the low pass filter and high pass filters M and N, respectively, the height and width of the image and R is the frame-rate of the video stream. Assuming $K = k_h = k_g$ (1)-(3) can be represented in a generalized form

$$z(n_1, n_2, n_3) = \sum_{i_1=0}^{K-1} \sum_{i_2=0}^{K-1} \sum_{i_3=0}^{K-1} w_1(i_1)w_2(i_2)w_3(i_3).x(2n_1-i_1, 2n_2-i_2, 2n_3-i_3) \quad (4)$$

The computations of (2) can be composed into three distinct

stages as

$$z(n_1, n_2, n_3) = \sum_{i=0}^{K-1} w_3(i).v(2n_1-i, n_2, n_3) \quad (5)$$

$$v(n_1, n_2, n_3) = \sum_{i=0}^{K-1} w_2(i).u(n_1, 2n_2-i, n_3) \quad (6)$$

$$u(n_1, n_2, n_3) = \sum_{i=0}^{K-1} w_1(i).x(n_1, n_2, 2n_3-i) \quad (7)$$

$$u(n_1, n_2, n_3) = \sum_{i=0}^{K-1} w_1(i).x(n_1, n_2, 2n_3-i) \quad (7)$$

$[u(n_1, n_2, n_3)]$ represents the low pass and high pass output matrices.

Using the decomposition scheme of (5)-(7), the computation of 3D-DWT can be performed in three distinct stages as follows.

- 1) In stage-1 low pass and high pass filtering is performed row-wise on each input frame to produce Intermediate matrix $[U_l]$ and $[U_h]$ according to (7)
- 2) In stage-2 low and high pass filtering is performed in column-wise on each of the intermediate matrix $[U_l]$ and $[U_h]$ to generate four sub band matrices $[V_{ll}], [V_{lh}], [V_{hl}]$ and $[V_{hh}]$ according to (6)
- 3) Finally, in stage-3 low pass and high pass filtering is performed on inter-frame sub-bands to obtain eight oriented selective sub-band matrices $[Z_{lll}], [Z_{llh}], [Z_{lhl}], [Z_{lhh}], [Z_{hll}], [Z_{hlh}], [Z_{hhl}]$ and $[Z_{hhh}]$ of size $[M/2 \times N/2 \times R/2]$.

5 VLSI IMPLEMENTATION OF LIFTING 3D-DWT

The input image file is converted into text file by using Matlab software. The text file consisting of pixels information corresponding to the image file



Fig.1 Input image

The figure 1 shows the corresponding input image. The initial block of the design is that the Discrete Wavelet Transform (DWT) block which is mainly used for the transformation of the image.

In this process, the image will be transformed and hence the high pass coefficients and the low pass coefficients were generated. The DWT consists of registers and adders. When ever the input is send, the data divided into even data and odd data. The even data and odd data is stored in the temporary registers. When the reset is high the temporary register value

consists of zero whenever the reset is low the input data split into the even data and odd data. The input data read up to sixteen clock cycles after that the data read according to the lifting scheme. The output data consists of low pass and high pass elements. This is the 1-D discrete wavelet transform.

The 2-D discrete wavelet transform is that the low pass and the high pass again divided into LL, LH and HH, HL. The output is verified in the Modelsim software. For this DWT block, the clock and reset were the primary inputs. The pixel values of the image, that is, the input data will be given to this block and hence these values will be split in to even and odd pixel values. In the design, this even and odd were taken as a array which will store its pixel values in it and once all the input pixel values over, then load will be made high which represents that the system is ready for the further process. Once the load signal is set to high, then the each value from the even and odd array will be taken used for the Low Pass Coefficients generation process. Hence each value will be given to the adder and in turn given to the multiplication process with the filter coefficients. Finally the Low Pass Coefficients will be achieved from the addition process of multiplied output and the odd pixel value.

Again this Low Pass Coefficient will be taken and it will be multiplied with the filter coefficients. The resultant will be added with the even pixel value which gives the High Pass Coefficient. Hence all the values from even and odd array will be taken and then above said process will be carried out in order to achieve the High and Low Pass Coefficients of the image. Now these low pass coefficients and the high pass coefficients were taken as the input for the further process. Hence for the DWT-2 process, low pass coefficients will be taken as the inputs and will do the process in order to calculate the low pass and high pass coefficients from the transformed coefficients of DWT-1. Similarly the same process is carried out for both DWT-2 and DWT-3.

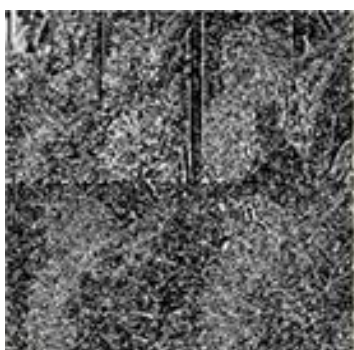


Fig.2. Compressed image

The figure 2 shows the compressed image using 3D-DWT.

The corresponding 3D-DWT is simulated using Modelsim simulator. Both Discrete wavelet transform and Inverse discrete wavelet transform was simulated by using VerilogHDL coding.

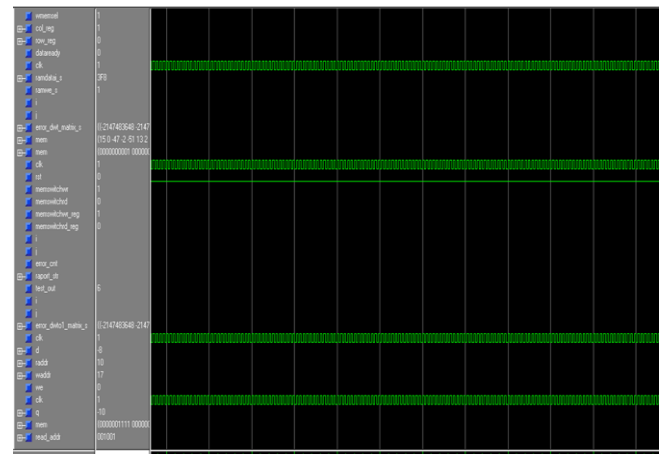


Fig.3 Simulation result of DWT block.

The figure shows the simulation result of DWT block.

From the compressed image, original image was reconstructed by using Inverse Discrete wavelet transform.

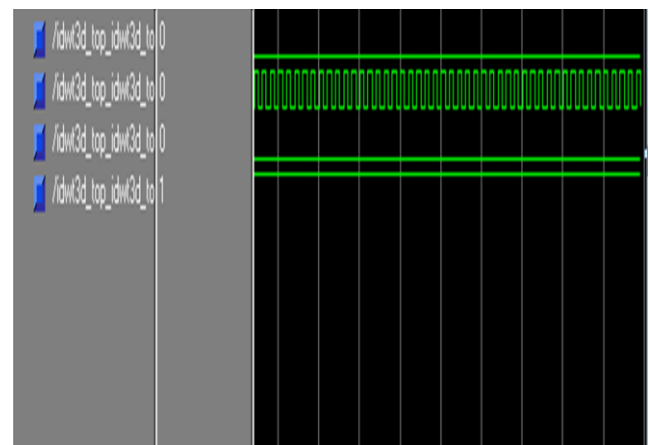


Fig.4 Simulation result of IDWT block.

The figure shows the simulation result of IDWT block.

6 SYNTHESIS RESULT

The simulation result was synthesized using Xilinx Synthesis technology. The developed DWT is synthesized using the Xilinx ISE tool. Here in this Spartan 3E family, many different devices were available in the Xilinx ISE tool. In order to synthesize this DWT and IDWT design the device named as "XC3S500E" has been chosen and the package as "FG320" with the device speed such as "-3".

The design of DWT is synthesized and its results were analyzed as follows.

Timing Summary:

Speed Grade: -3

Minimum period: 4.44ns

Maximum Frequency: 225.195 MHz

Minimum input arrival time before clock: 4.147 ns

Maximum output required time after clock: 14.682 ns
Maximum combinational path delay: No path found.

The maximum operating frequency of this synthesized design is given as 225.195 MHz and the minimum period as 4.44 ns. Here, OFFSET IN is the minimum input arrival time before clock and OFFSET OUT is maximum output required time after clock.

Device utilization summary:

tst1 Project Status (04/21/2010 - 17:41:36)				
Project File:	tst1.isc	Current State:	Placed and Routed	
Module Name:	top_dwt	• Errors:	No Errors	
Target Device:	xc3s500e-4lg320	• Warnings:	5 Warnings	
Product Version:	ISE 10.1 - Foundation Simulator	• Routing Results:	All Signals Completely Routed	
Design Goal:	Balanced	• Timing Constraints:	All Constraints Met	
Design Strategy:	Xilinx Default (unlocked)	• Final Timing Score:	0 (Timing Report)	

tst1 Partition Summary		
No partition information was found.		

Device Utilization Summary				
Logic Utilization	Used	Available	Utilization	Note(s)
Number of Slice Flip Flops	386	9,312	4%	
Number of 4 input LUTs	532	9,312	5%	
Logic Distribution				
Number of occupied Slices	446	4,656	9%	
Number of Slices containing only related logic	446	446	100%	
Number of Slices containing unrelated logic	0	446	0%	
Total Number of 4 input LUTs	532	9,312	5%	
Number used as logic	292			
Number used for Dual Port RAMs	240			
Number of bonded I/Os	104	232	44%	
Number of BUFMUXs	1	24	4%	

The device utilization summary is shown above in which it gives the details of number of devices used from the available devices and also represented in %. Hence as the result of the synthesis process, the device utilization in the used device and package is shown above

TABLE 1

Comparison of DWT output

Existing	Area usage in u.sqm	Power in mW	Delay in ns
RWTU	146316	29.95	9.56
DWT CONTROL	146386	61.12	3.85
DWT 3D CONTROL	133704	44.09	3.85
Proposed	Area usage in u.sqm	Power in mW	Delay in ns
RWTU	131020	4.6	9.56
DWT CONTROL	132684	23.61	3.86
DWT 3D CONTROL	132684	19.07	3.86

TABLE 2

Comparison of IDWT output

Existing	Area usage in u.sqm	Power in mW	Delay in ns
IRWTU	175292	30.01	13.904
IDWT CONTROL	132684	61.18	3.85
IDWT 3D CONTROL	147340	44.03	3.85
Proposed	Area usage in u.sqm	Power in Mw	Delay in ns
IRWTU	131020	4.6	13.40
IDWT CONTROL	131660	23.61	3.85
IDWT 3D CONTROL	132684	19.00	3.85

TABLE 3

Metrics of proposed architecture

Proposed	Throughput	Latency	Power saved(mw)
3D-DWT	3.2N	0.06	55.97
3D-IDWT	3.59N	0.05	55.97

Where N in Input Frame Size

Summary:

- The developed DWT is modeled and simulated using the Modelsim tool.
- The model is synthesized using the Xilinx tool in Spartan 4E and their synthesis results were discussed.

7 CONCLUSION

The parallel and pipeline architecture was proposed for high-throughput computation 3-D DWT. Each level of 3-D DWT computation is split into three distinct stages and the computations of all the three stages are implemented concurrently in a parallel array of pipelined processing modules. We have proposed a cascaded structure where each level of decomposition is performed by a processing unit in a separate pipeline stage. An interesting feature of the proposed structure is that it involves relatively small frame-buffer than the existing structures; and size of the frame-buffer is independent of frame-rate.

Besides, the size of the on-chip storage and frame-buffer is independent of the input block size. The latency of the proposed structures for 3-D DWT is $O(MN/Q)$ while that of the existing structures is $O(MNR)$, where $(M \times N)$ is the frame-size. Compared with best of the existing structures. The throughput rate of the proposed structures can easily be scaled without increasing the On-chip storage and frame-

memory by using more number of processing modules; and for higher-frame rates and higher input block-size it provides greater advantage over the existing designs. It is also found that the full-parallel implementation of proposed scalable structure provides the best of its performance.

REFERENCES

- [1] AnirbanDas, AnindyaHazara ,Swapna Banarjee, (2010), "An efficient architecture for 3-DDWT Wavelet transform", In IEEE Transactions on circuits and systems for video technology, Vol 20,No2
- [2] BasantK.Mohanty, AnuragMahajan,and PramodK.Meher, (2012), "Area and power efficient architecture for High- Throughput implementation of Lifting 2-D DWT.9", ,In IEEE Transactions on circuits and systems-2,Vol 59,No 7
- [3] BasantK.Mohanty,Pramod K.Meher ,(2010) ,"Parallel and pipeline architecture for high throughput computation of Multilevel 3-DDWT", In IEEE Transactions on circuits and systems for video technology, Vol 20,N0 9
- [4] Dhaha Dia, Zeghid, Taoufik saidani, Mohamed Atri, Mohsen Machout,(2009),"Multi level Discrete Wavelet Transform Architecture design", Proceedings of the world congress on engineering, Vol 1, 978-988
- [5] Janardan.M, AshokBabu.K, (2011), "An efficient architecture for 3-DDWT Lifting based discrete wavelet transform",Comp tech Appl, Vol 2(5), 1439-1458.
- [6] Mansouri.A, Ahaitouf.A and Abdi.F (2009), "An efficient VLSI architecture and FPGA implementation of High-Speed and Low Power 2-D DWT for(9,7)
- [7] wavelet filter", IJCSNS International Journal of computer science and network security, Vol 9,No.3
- [8] Mohanty.B.K and Meher.P.K ,(2011), "Memory- efficient modular VLSI architecture for high-throughput and low latency implementation of multilevel lifting 2-D DWT," IEEE Trans. Signal Process., vol. 59, no. 5, pp. 2072– 2084

-
- N.Vairamani is currently pursuing masters degree program in commnnication system engineering in Mount Zion Engineering College affiliated to Anna University, Chennai,Tamilnadu,India. E-mail: anoushkasrinidhi@gmail.com
 - A.Taksala Devapriya is currently working as Assistant Professor in Department of Electronics and Communication Engineering ,Mount Zion Engineering College affiliated to Anna University, Chennai,Tamilnadu,India. E-mail : taksala@gmail.com
 - V.Muthukumar is currently working as a Senior Lecture in Faculty of Engineering and Computer Technology,AIMST University,Malaysia. His area of research interest is VLSI based image processing and Bio-MEMS.He got 12 years of vast teaching experience. E-mail: gvmkumaran@gmail.com.