



High Performance VLSI Architecture for 3-D Discrete Wavelet Transform

¹CHEEKATYALA ANJAN KUMAR, M.Tech Assistant Professor, anjankumarcheekatla@gmail.com

²THOTA SRAVANTI M.Tech Associate Professor, sravanti815@gmail.com

Department-ECE

Pallavi Engineering College Hyderabad, Telangana 501505.

Abstract—This paper presents a high-speed memory efficient VLSI architecture for three dimensional (3-D) discrete wavelet transform. A major strength of the proposed architecture lies in reducing the number and period of clock cycles for the computation of wavelet transform. This five stage pipelined architecture shares the partial load of the next stage with the present stage to reduce computational load at the next stage and critical path delay (CPD). The proposed architecture has replaced the multiplications by optimized shift and add operations to reduce the CPD. Implementation results show that the proposed architecture benefits from the features of reduced memory, low power consumption, low latency, and high throughput over several existing designs. **Index Terms**—3-D DWT, lifting based DWT, VLSI Architecture, flipping structure.

1. INTRODUCTION

Video compression is a major requirement in many applications including medical imaging, studio applications and broadcasting applications. Compression ratio of a video encoder depends on the underlying compression algorithms. The goal of any compression technique is to reduce the immense amount of visual information to a manageable size, so that it can be efficiently stored, transmitted, and displayed. It may be noted that 3-D DWT enables compression in spatial as well as temporal direction, which is very desirable for video compression. Moreover, wavelet based compression provides the scalability with the levels of decomposition. From the last two decades, several hardware designs have been noted for implementation of 2-D DWT and 3-D DWT for different applications. Majority of the designs are developed based on convolution and lifting procedures. Most of the existing architectures involve large memory requirement, low throughput, and complex control circuit. Convolution based implementations [1]-[3] provide the outputs within less time but would require high amount of arithmetic resources. Moreover, it is memory intensive and would occupy large area for implementation. Lifting based [4] implementations require less memory, less computational complexity and possibility to implement in parallel. However, it will require long critical path, and recently more number of contributions are reported to reduce the critical path in lifting based implementations. Several lifting based 3-D DWT architectures are noted in the literature [5]-[11] to reduce the critical path of the 1-D DWT architecture and to decrease the memory requirement of the 3-D architecture. Among the best existing designs of 3-D DWT, Darji et al. [11] produced best results by reducing the memory requirements and gives the throughput of 4 results/cycle. Still it requires the large on-chip memory ($4N^2 + 10N$). In this paper, we propose a high-speed and memory efficient lifting based 3-D DWT architecture, which requires only $2 * (3N + 30P) + 24$ words of on-chip memory and produces 8 results/cycle and reduced the critical path delay to $2T_a$. The proposed 3-D DWT architecture is built with two spatial 2-D DWT (CDF 9/7) processors and four temporal 1-D DWT (Haar) processors. To eliminate the temporal memory and to reduce the latency, Haar wavelet is incorporated in the temporal processor. The resultant architecture has succeeded in reducing the latency, on chip memory and enhance the speed of operation compared to existing 3-D DWT designs. Organization of rest of the paper is as follows. Detailed description of the proposed architecture for 3-D DWT is provided in section II. Results and performance comparisons are given in Section III. Finally, concluding remarks are given in Section IV.

II. PROPOSED ARCHITECTURE FOR 3-D DWT

The proposed architecture for 3-D DWT consisting of two parallel spatial processors (2-D DWT) and four temporal processors (1-D DWT), is depicted in Fig. 1. After applying 2-D DWT on two consecutive frames, each spatial processor (SP) produces 4 sub-bands, viz. LL, HL, LH and HH which are fed to the inputs of four temporal processors (TPs) to perform the temporal transform. Output of these TPs consist of a low frequency frame (L-frame) and a high frequency frame (Hframe). A. Architecture for Spatial Processor In this section, we propose a new high-speed memory efficient lifting based 2-D DWT architecture denoted by spatial processor (SP). It consists of row and

column processors. By introducing pipelining and sharing the computational load within the pipeline stages, the flipping based DWT equations proposed by Huang et al. in [12] are modified to eqns.(1)- (5). The proposed row processor (1-D DWT) and column processor (1-D DWT) have utilized the same equations ((1)- (5)) for implementation.

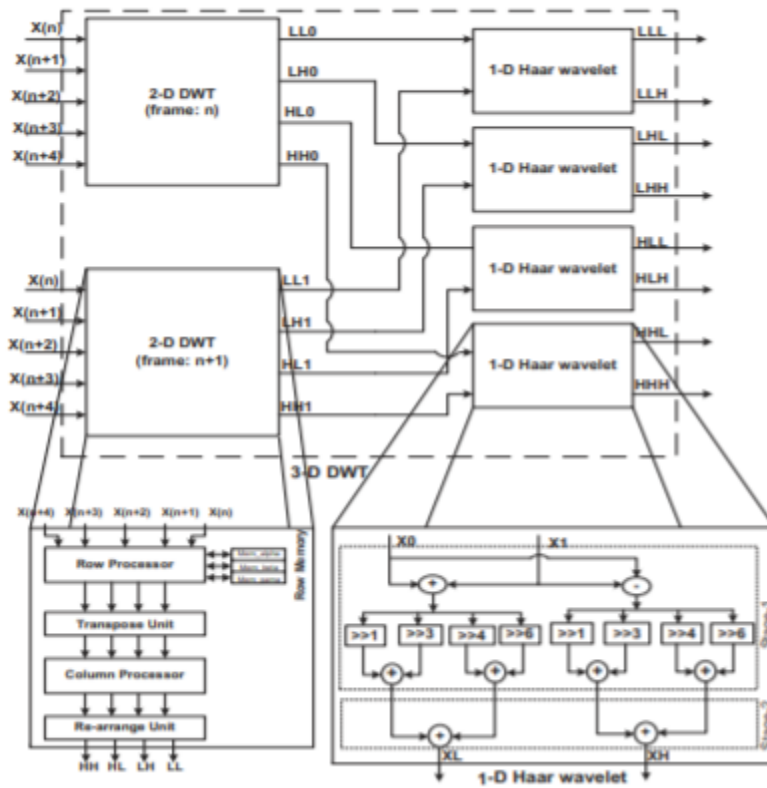


Figure 1. Block diagram for 3-D discrete wavelet transform

$$\left. \begin{aligned} X'[2n-1] &= 0.0078125 * X[2n-1] \\ X''[2n-1] &= 0.625 * X[2n-1] \end{aligned} \right\} 1^{st} \text{ stage} \quad (1)$$

$$\left. \begin{aligned} a' * X[2n-1] &= X'[2n-1] + X''[2n-1] \\ H_1[n] &= a' * X[2n-1] + X[2n] + X[2n-2] \\ A = b' * X[2n] &= 12 * X[2n] \end{aligned} \right\} 2^{nd} \text{ stage} \quad (2)$$

$$\left. \begin{aligned} B = H_1'[n] &= 21 * H_1[n] \\ C = H_1''[n] &= 0.375 * H_1[n] \\ L_1[n] &= A + H_1[n] + H_1[n-1] \end{aligned} \right\} 3^{rd} \text{ stage} \quad (3)$$

$$\left. \begin{aligned} D = d' * L_1[n] = L'_1[n] &= 2.565 * L_1[n] \\ H_2[n] &= L_1[n] + L_1[n-1] + B + C \\ H[n] &= 0.0625 * H_2[n] \end{aligned} \right\} 4^{th} \text{ stage} \quad (4)$$

$$\left. \begin{aligned} L_2[n] &= D + H_2[n] + H_2[n-1] \\ L[n] &= 0.03125 * L_2[n] \end{aligned} \right\} 5^{th} \text{ stage} \quad (5)$$

The proposed architecture utilizes the strip based scanning [14] to enable the trade-off between external memory and internal memory. Pipelined lifting based (9/7) 1-D DWT process has been performed by the processing unit (PU) in the proposed architecture. The proposed PU architecture reduces the CPD to 2Ta (two adder delay). Fig. 2(a) shows the data flow graph (DFG) of the proposed PU architecture and Fig. 2(b) depicts the internal architecture of the proposed PU.

The number of inputs to the spatial processor is equal to $2P+1$, which is also equal to the width of the strip, where P is the number of parallel processing units (PUs) in the row processor as well as column processor. The proposed architecture has been designed with two parallel processing units ($P = 2$). The same structure can be extended to $P = 4, 8, 16$ or 32 depending on external bandwidth. Whenever the row processor

produces the intermediate results, the column processor starts to work on those intermediate results. Row processor takes 5 clocks to produce the temporary results, upon which the column processor takes 5 more clocks to give the 2-D DWT output. Finally, temporal processor takes 2 more clock after 2-D DWT results are available to produce 3-D DWT output. Overall, the proposed 2-D DWT and 3-D DWT architectures have constant latency of 10 and 12 clock cycles respectively, regardless of the image size N and the number of parallel PUs (P). A detailed description of the row processor and the column processor is given in the following sub-sections. 1) Row Processor (RP): Let X be the image of size $N \times N$, which is extended by one column using symmetric extension. Now the image size becomes $N \times (N+1)$. One may refer [14] for the structure of strip based scanning method. The proposed architecture initiates the DWT process row wise through the row processor (RP) and then process the column DWT by the column processor (CP). Fig. 2(c) shows the generalized structure for a row processor with two ($P=2$) PUs. Each PU consists of five pipeline stages and each pipeline stage is processed by one processing element (PE) as depicted in Fig. 2(b). The maximum CPD provided by these PEs is 2Ta. The outputs $H_1[n+1]$, $L_1[n+1]$, and $H_2[n+1]$ corresponding to PE alpha and PE beta of the last PU and PE gama of last PU is saved in the memories Memory alpha, Memory beta and Memory gama respectively. Those stored outputs are inputted for next subsequent columns of the same row. For an $N \times N$ image, the size of each memory is $N \times 1$ words and total row memory to store these outputs is $3N$. Output of each PU is fed to the transposing unit, which has P number of transpose registers (one for each PU). Fig. 3(a) shows the structure of the transpose register, which gives the two H and two L data alternatively to the column processor. 2) Column Processor (CP): The structure of the Column Processor (CP) is shown in Fig. 2(d). To match the RP throughput, CP is also designed with two PUs in our architecture. Each transpose register produces a pair of H and L in an alternative order, which are fed to the inputs of one PU of the CP. The partial results produced are consumed by the next PE after two clock cycles. As such, the shift

registers of length two are needed within the CP between each pair of pipeline stages for caching the partial results (except between 1st and 2nd pipeline stages). At the output of the CP, four subbands are generated in an interleaved pattern, i.e., (HL,HH), (LL,LH), (HL,HH), (LL,LH), and so on. Outputs of the CP are fed to the re-arrange unit. Fig. 3(b) shows the architecture for the re-arrange unit, and it provides the outputs in sub-band order i.e LL, LH, HL and HH simultaneously. B. Architecture for Temporal Processor (TP) According to Sweldens et al. [15], lifting based Haar wavelet transform depends on intensity values of two adjacent pixels. As soon as the spatial processors provide the 2-D DWT results, the temporal processors start processing on the spatial processor outputs (2-D DWT results) and produce the 3-D DWT results. Fig. 1 shows that there is no requirement of temporal buffer, as the sub-band coefficients of two spatial processors are directly connected to the four temporal processors. Temporal processors apply 1-D Haar wavelet on sub-band coefficients, and provide the low frequency sub-band and high frequency sub-band as output. III. RESULTS AND PERFORMANCE COMPARISON The proposed 3-D DWT architecture has been described in Verilog HDL. A uniform word length of 15 bits has been maintained throughout the design. Simulation results have been verified by using Xilinx ISE simulator. We have simulated the Matlab model which is similar to the proposed 3-D DWT hardware architecture and verified the 3-D DWT coefficients. RTL simulation results have been found to exactly

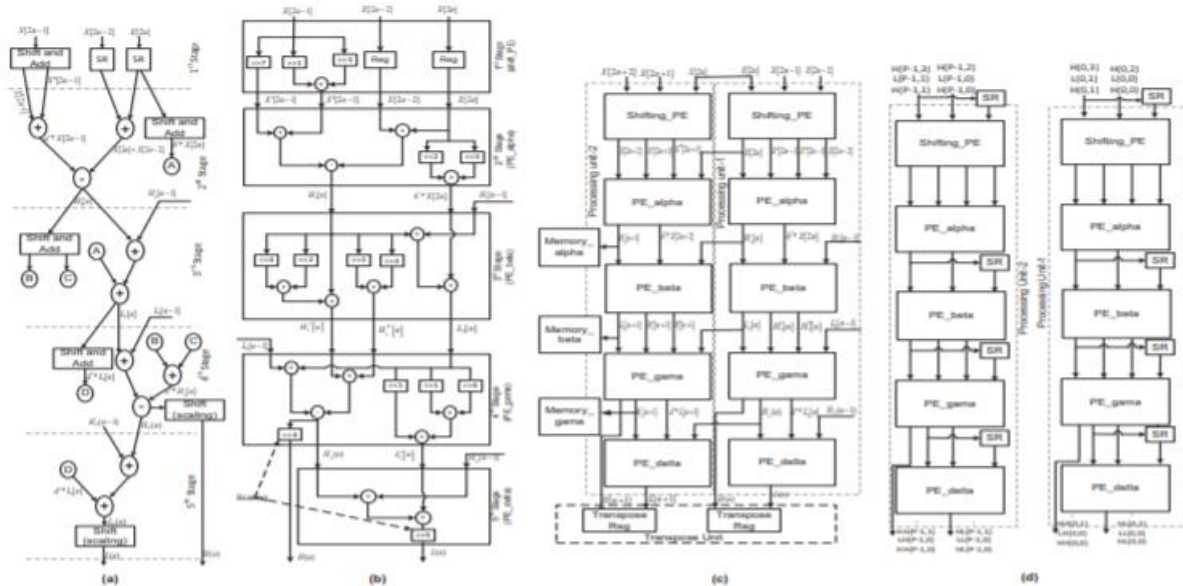


Figure 2. (a) Data Flow Graph of modified 1-D DWT architecture (b) Structure of Processing Unit (c) Row Processor (d) Column Processor

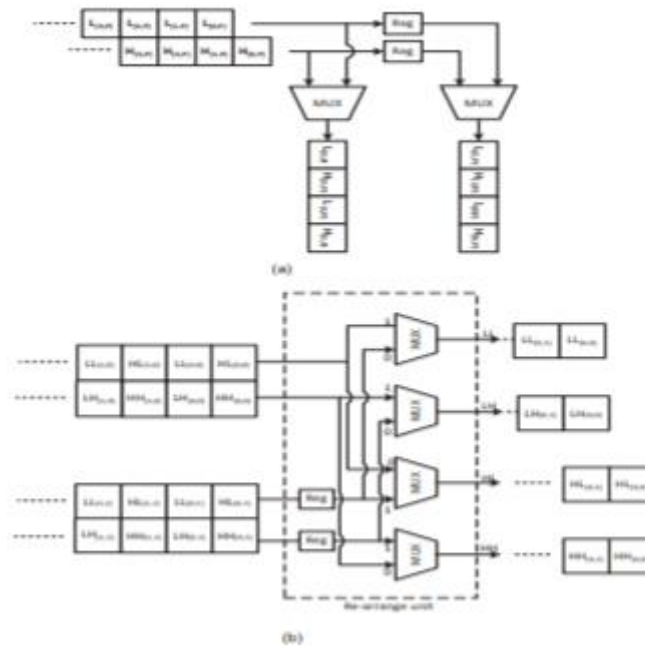


Figure 3. (a) Transpose Register (Ref:[14]) (b) Re-arrange Unit

table 1 device utilisation summary of the proposed architecture

Logic utilized	Used	Available	Utilization
Slice Registers	1958	106400	1%
Number of Slice LUTs	2852	53200	5%
Number of fully used LUT-FF pairs	1137	3673	30%
Number of Block RAM	3	140	2%

table ii comparison of proposed 3-d dwt architecture with existing architectures (for 1-level)



Parameters	Weeks [6]	Taghavi [7]	A.Das [9]	Darji [11]	Proposed
Memory requirement	$6N^2+6l$	$5N^2$	$5N^2 + 5N$	$4N^2 + 10N$	$2*(3N+30P)+24$
Throughput/cycle	-	1 result	2 results	4 results	8 results
Computing time For 2 Frames	$2N^2 + 3l/2$	$6N^2$	$3N^2$	$3N^2$	$N^2/2P$
Latency	$2.5N^2 + 0.5l$	$4N^2$ cycles	$2N^2$ cycles	$3N^2/2$ cycles	12 cycles
Area	-	-	1825 slices	2490 slices	2852 slice LUTs
Operating Frequency	200 MHz (ASIC)	-	321 MHz (FPGA)	91.87 MHz (FPGA)	265 MHz (FPGA)
Multipliers	-	-	Nil	30	Nil
Adders	$6l$ MACs	-	78	48	168
Filter bank	l -length	D-9/7	D-9/7	D-9/7	D-9/7 (2-D) + Haar (1-D)

A.Comparison Table II shows the comparison of the proposed 3-D DWT architecture with several existing 3-D DWT architectures. It is found that the proposed design possesses the features of less memory requirement, high throughput, less computation time and minimal latency compared to [6], [7], [9], and [11]. Though the proposed 3-D DWT architecture has small disadvantage in area and frequency, when compared to [9], the proposed one has a great advantage in remaining all aspects. Table III gives the comparison of synthesis results between the proposed 3-D DWT architecture and [11]. Although the proposed architecture occupies more cell area, but it includes total on chip memory also, whereas in [11], on chip memory is not included. Power consumption of the proposed 3-D architecture is much reduced compared to [11].

table iii synthesis results (design vision) comparison of proposed 3-d dwt architecture with existing

Parameters	Darji[11]	Proposed
Comb. Area	$61351 \mu m^2$	$526419 \mu m^2$
Non Comb. Area	$807223 \mu m^2$	$553078 \mu m^2$
Total Cell Area	$868574 \mu m^2$	$1079498 \mu m^2$
Operating Voltage	1.98 V	1.2 V
Total Dynamic Power	179.75 mW	38.56 mW
Cell Leakage Power	$46.87 \mu W$	4.86 mW

match the Matlab simulation results. The Verilog RTL code is synthesized using Xilinx ISE 14.2 tool and mapped to a Xilinx programmable device (FPGA) 7z020clg484 (zynq board) with speed grade of -3. Table I shows the device utilization summary of the proposed architecture and it operates with a maximum frequency of 265 MHz. The proposed architecture has also been synthesized using synopsys design compiler with 90-nm technology CMOS standard cell library. It consumes 43.42 mW power and occupies an area equivalent to 231.45 K equivalent gate at frequency of 200 MHz.

IV. CONCLUSIONS:

In this paper, we have proposed high-speed and memory efficient architecture for lifting based 3-D DWT. The proposed architecture has not only been implemented on 7z020clg484 FPGA target of zynq family, but has also been synthesized on Synopsys' design vision for ASIC implementation. An efficient design of 2-D spatial processor and 1-D temporal processor reduces the internal memory, latency, critical path delay and complexity of a control unit, and increases the throughput. When compared with several existing architectures, the proposed scheme shows higher performance at the cost of a slight increase in area. The proposed 3-D DWT architecture is capable of computing 60 UHD (3840×2160) frames in a second.

REFERENCES



- [1] Q. Dai, X. Chen, and C. Lin, "A Novel VLSI Architecture for Multidimensional Discrete Wavelet Transform," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 8, pp. 1105-1110, Aug. 2004.
- [2] C. Cheng and K. K. Parhi, "High-speed VLSI implementation of 2-D discrete wavelet transform," IEEE Trans. Signal Process., vol. 56, no. 1, pp. 393-403, Jan. 2008.
- [3] B. K. Mohanty and P. K. Meher, "Memory-Efficient High-Speed Convolution-based Generic Structure for Multilevel 2-D DWT." IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 2, pp. 353-363, Feb. 2013.
- [4] I. Daubechies and W. Sweledens, "Factoring wavelet transforms into lifting schemes," J. Fourier Anal. Appl., vol. 4, no. 3, pp. 247-269, 1998.
- [5] J. Xu, Z.Xiong, S. Li, and Ya-Qin Zhang, "Memory-Constrained 3-D Wavelet Transform for Video Coding Without Boundary Effects," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 9 ,pp. 812-818, Sep. 2002.
- [6] M. Weeks and M. A. Bayoumi, "Three-Dimensional Discrete Wavelet Transform Architectures," IEEE Transactions on Signal Processing, vol. 50, no. 8, pp.2050-2063, Aug. 2002.
- [7] Z. Taghavi and S. Kasaei, "A memory efficient algorithm for multidimensional wavelet transform based on lifting," in Proc. IEEE Int. Conf. Acoust Speech Signal Process. (ICASSP) , vol. 6, pp. 401-404, 2003.
- [8] Q. Dai, X. Chen, and C. Lin, "Novel VLSI architecture for multidimensional discrete wavelet transform," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 8, pp. 1105-1110, Aug. 2004.
- [9] A. Das, A. Hazra, and S. Banerjee, "An Efficient Architecture for 3- D Discrete Wavelet Transform," IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 2, pp. 286-296, Feb. 2010.
- [10] B. K. Mohanty and P. K. Meher, "Memory-Efficient Architecture for 3-D DWT Using Overlapped Grouping of Frames," IEEE Transactions on Signal Processing, vol. 59, no. 11, pp.5605-5616, Nov. 2011.
- [11] A. Darji, S. Shukla, S. N. Merchant and A. N. Chandorkar, "Hardware Efficient VLSI Architecture for 3-D Discrete Wavelet Transform," Proc. of 27th Int. Conf. on VLSI Design and 13th Int. Conf. on Embedded Systems pp. 348-352, 5-9 Jan. 2014.
- [12] C.T. Huang, P.C. Tseng, and L.G. Chen, "Flipping structure: an efficient VLSI architecture for lifting-based discrete wavelet transform," IEEE Trans. Signal Process., vol. 52, no. 4, pp. 1080-1089, Apr. 2004.
- [13] C. Y. Xiong, J. W. Tian, and J. Liu, "A Note on Flipping Structure: An Efficient VLSI Architecture for Lifting-Based Discrete Wavelet Transform," IEEE Transactions on Signal Processing, vol. 54, no. 5, pp. 1910-1916, May 2006
- [14] Y. Hu and C. C. Jong, "A Memory-Efficient Scalable Architecture for Lifting-Based Discrete Wavelet Transform," IEEE Transactions on Circuits and Systems-II: Express Briefs, vol. 60, no. 8, pp. 502-506, Aug. 2013.
- [15] W. Sweldens, "The Lifting Scheme: a Construction of Second Generation of Wavelets," SIAM Journal on Mathematical Analysis, vol.29 no.2, pp. 511-546, 1998.
- [16] C.T. Huang, P.C. Tseng, and L.-G. Chen, "Analysis and VLSI architecture for 1-D and 2-D discrete wavelet transform," IEEE Trans. Signal Process., vol. 53, no. 4, pp. 1575-1586, Apr. 2005.