Design of 2-D DWT VLSI Architecture for Image Compression

ThirumaraiSelvi Chandraraju, and Sudhakar RadhaKrishnan

***Abstract*—**In this paper, a new data access scheme for the computation of lifting 2-D DWT (Discrete Wavelet Transform) using systolic arrays with block processing is suggested. From DG (dependence graph) linear systolic array is directly derived. For parallel and pipeline implementation of 1-D DWT from suitably segmented DG is used for deriving 2-D systolic arrays. Above two systolic arrays are used as building blocks to derive the lifting 2-D DWT. The proposed architecture requires a small on-chip memory of (4N + 8P) where N is the image width, process a block of P samples in every cycle. Compared to existing structures it has high throughput, low latency and less computational complexity. The synthesis is performed in Xilinx 8.1i, Spartan 2E hardware with XC2S50E device and FT256 package and simulation results are obtained using Mat lab 7.10 and Modelsim 6.3f. The image size is 512 X 512 and block size is 4 with area is 987500.22 u.sqm, power consumed is 8.34027 mw and delay count is 16.11 ns

***Keywords*—**Block processing, Discrete Wavelet Transform (DWT), Lifting, Systolic VLSI, 2-D DWT.

I. Introduction

TWO dimensional discrete wavelet transform (2-D DWT) has evolved as an effective and powerful tool in many applications especially in image processing and compression. This is mainly due to its better computational efficiency achieved by factoring wavelet transforms. Mainly two types of DWT structures are classified in to (i) convolution and (ii) lifting. Lifting scheme facilitates high speed and efficient implementation of wavelet transform and it is attractive for both high throughput and low power applications. Lifting requires less arithmetic and memory resources. Compared to convolution [1], [2] hardware components of 2-D DWT are broadly classified into arithmetic components and memory components. Arithmetic components have multipliers and memory components have transposition memory and temporal memory. Transposition memory used to store input/intermediate coefficients whereas temporal memory stores partial results of filter output.

ThirumaraiSelvi Chandraraju is with the Electronics and Communication Engineering Department, Sri Krishna College of Engineering and Technology, Coimbatore-641008,Tamilnadu, INDIA (Phone: 9944192456; e-mail: selvichand@gmail.com).

Sudhakar RadhaKrishnan, was with Electronics and Communication Engineering Department, Dr.Mahalingam College of Engineering and Technology,Pollachi,Tamilnadu,INDIA

(e-mail:sudha_radha2000@yahoo.co.in).

Parallel data access scheme Cheng et al [5] in which the size of transposition memory is reduced and temporal memory remains independent of data access scheme and input block size. Therefore, in 2-D DWT structures the on-chip memory is based on parallel data access scheme is dominated by temporal memory. The line based structure in [4] requires temporal memory of size 3N to process the 4 samples per cycle and parallel scanning lifting scheme [6] involves same size of temporal memory as in line based. The proposed systolic arrays block processing system is used to utilize temporal memory to reduce area-time complexity of 2-D DWT structure. The block based methods of parallel and pipelined architecture are used in the implementation of 2-D DWT [7], [8]. Both these structures have same throughput rate and same arithmetic resources but different sizes of transposition memory is varied according to the size of input data matrix. Mohanty et al [7] obtained data blocks by folding rows, size of temporal memory is 3N and transposition memory is 2.5N for 1 level 2-D DWT. Tian [8] derived the data blocks from P-rows parallel data access, transposition memory size is [N (P + 2)/2] and temporal memory size 3N, P is the block size. Structure [8] requires transposition memory to buffer the intermediate blocks and the processing of blocks is different order than the input data matrix. Transposition memory size depends on block size as well as the on the image size. On chip memory in [8] depends on block size and for block size>=4, on chip memory is independent of block size in [7] and has less block size compared to [8]. This paper suggest a data access scheme of suitably partitioning and mapping of appropriate computation of hardware architecture to derive the memory and area-power efficient block based 2-D DWT structure.

II. Existing Work

A modular and pipeline architecture of lifting based multilevel 2-D DWT [7] structure provides appropriate partitioning and scheduling is performed at each decomposition levels. The different levels at which the processing is performed using cascaded pipeline architecture. The proposed structure uses pyramid algorithm and one recursive pyramid algorithm. Then the entire processing is based on unit input block size. It has local register and RAM for storage of data instead of buffers which processes image of size 512 X 512.The main drawback of this method is large size on-chip memory which requires more area and power for

processing. To overcome the drawback we are moving onto lifting based systolic arrays of block processing.

III. Proposed Work

The proposed structure consists of one row processor and one column processor. The row processor is composed of M 1-D systolic arrays.

A. One-Dimensional Systolic Array

Data dependence graph (DG) of N-point lifting DWT is shown in Fig. 1(a). It consists of N/2 identical sections, and each section has four identical nodes.

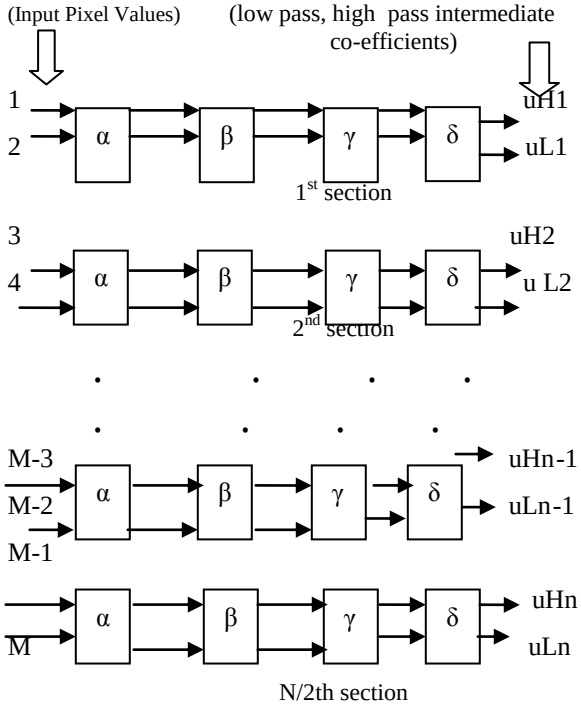

Fig. 1(a) DG of *N*-point lifting DWT

Projecting the nodes of the DG and using a systolic schedule, we obtain a 1-D systolic array [shown in Fig. 1(b)].

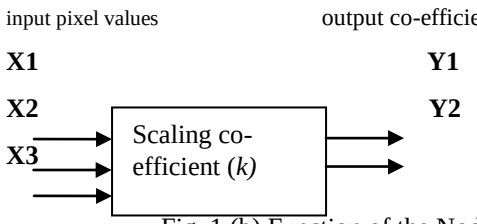

Fig. 1 (b) Function of the Node

The function of the processing element (PE) is shown in Fig. 1(c).

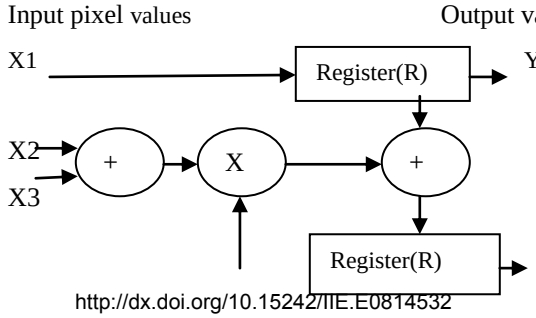

Fig. 1 (c) Function of the PE

The 1-D systolic array consists of four PEs and processes two samples in every cycle, where one cycle period is T = TM + 2TA, with TM and TA being the time required to perform one multiplication and the time required to perform addition in a PE, respectively.

Lifting computation of 1-D DWT is illustrated in Fig.1 (d) and expressed as

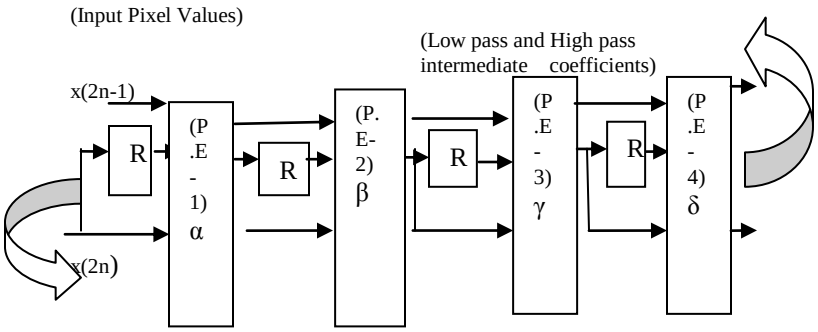

Fig. 1 (d) One-dimensional systolic array

$$s1(n) = x(2n-1) + \alpha((x(2n) + x(2n-2)) \quad (1)$$

$$s2(n) = x(2n-2) + \beta((s1(n) + s1(n-1)) \quad (2)$$

$$uH(n) = s1(n-1) + \gamma((s2(n) + s2(n-1)) \quad (3)$$

$$uL(n) = s2(n-1) + \delta((uH(n) + uH(n-1)) \quad (4)$$

where

α= -1.586134342, β = -0.05298011854, γ=0.8829110762, and δ =0.4435068522 are lifting constants x(n) represents the low sub band components of the (j – 1)th level. The notation uL(n)and uH(n) represent the low-pass and high-pass intermediate outputs, respectively, corresponding to the input x(n). The coefficients uL(n) and uH (n) are scaled by constants K and K−1, respectively, for normalizing their magnitude. The column processor consists of one high-pass block and one low-pass block. The high-pass block processes the high pass intermediate matrix and generates a pair of subband matrices vHL and vHH, while the low-pass block processes the low-pass intermediate matrix and generates the subband matrices vLH and vLL. The high-pass block and low-pass block are identical in structure and derived from the 2-D systolic array shown in Fig. 2.

http://dx.doi.org/10.15242/IIE.E0814532

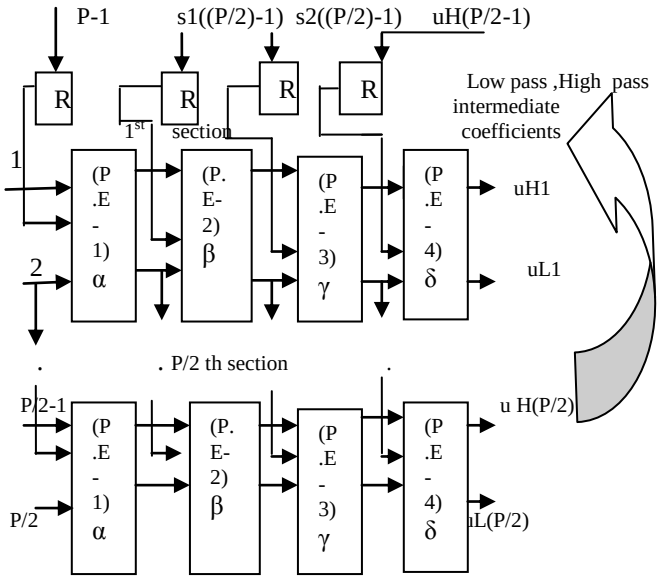

Fig. 2 Two-Dimensional Systolic Array

B. Two-Dimensional Systolic Array

The DG is partitioned into Q segments of (P/2) sections each, where N = PQ. As shown in Fig. 3,

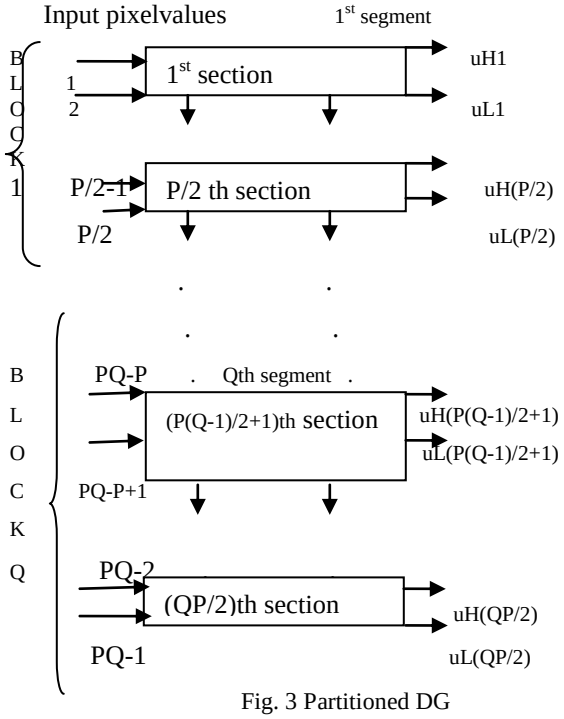

Fig. 3 Partitioned DG

The four registers of the 2-D systolic array of block size (P/2) are replaced by four shift registers (SRs) of (N/2) words each to have a low-pass/high-pass block. SRs are used by the low-pass/high-pass block to provide the necessary row delay to the intermediate coefficients and partial results. In both the low-pass and high-pass blocks, one SR is used for storing the intermediate coefficients, and other three SRs are used for storing the partial results. The sizes of transposition memory and temporal memory of the proposed structure are N and 3N, respectively. The low-pass and high-pass outputs of row processor and column processor need to be scaled according to the lifting scheme for the 9/7 filters of proposed structure is represented in Fig.4. The high-pass block computes P/4 rows of other two subband matrices (vHL and vHH). The column processor generates four subband matrices (vLL, vLH, vHL, and vHH) each of size (M/2 × N/2) in NQ/2 cycles.

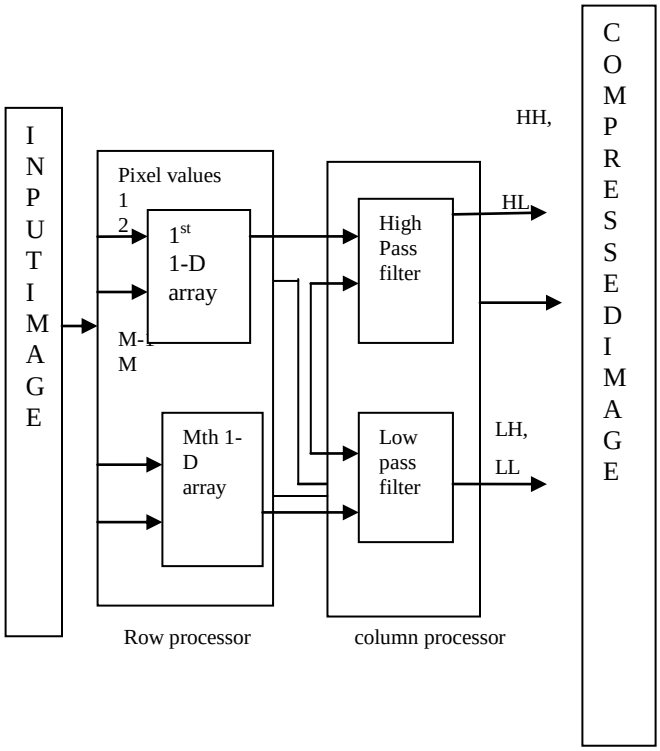

Fig. 4 Proposed structure for computation of lifting 2-D DWT

Both the row processor and the column processor work concurrently. The proposed structure, therefore, computes the 1-level lifting 2-D DWT of an image of size (M × N) in NQ/2 cycles with a latency of nine cycles, where Q = 2M/P.

IV. Results And Comparison

The proposed structure involves the same arithmetic resource (multiplier and adder) and offers the same throughput.

However, the proposed structure involves nearly 1.5N less on-chip memory words than those in and does not involve MUXes like other structures. Due to less on-chip memory, substantial amount of area and power could be saved using the proposed scheme than the other existing schemes.

The input image and the compressed image in 1-level is represented below"

Fig. 5 Input Image

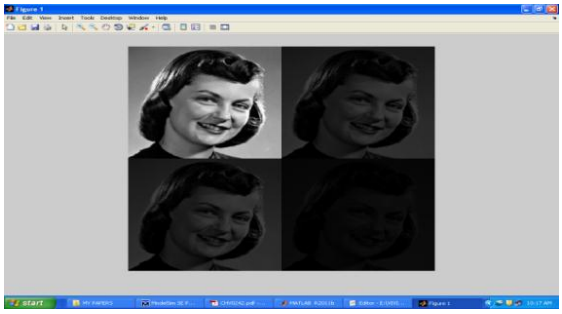

Fig. 5.1 Compressed 1-Level Image of 512X512

in 2-D DWT

In the above figure5.1compressed image along with the approximated co-efficient are represented in various columns and Fig.6 shows the simulation result of proposed work.

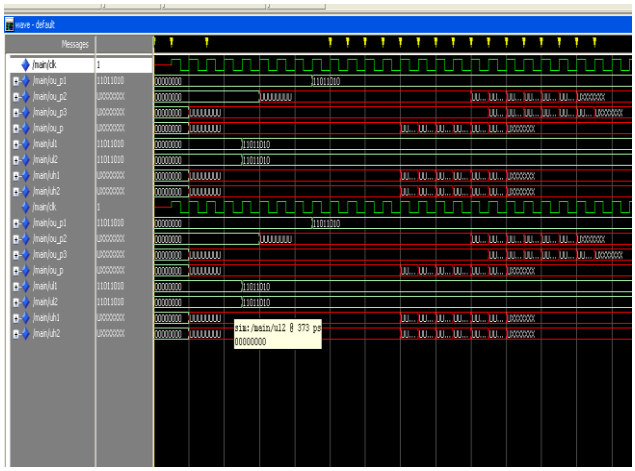

Fig. 6 Simulation Result of Proposed Work

The proposed structure in Table-I, therefore, involves (9P/2) multipliers, 8P adders, 8P pipeline/data registers, N transposition-memory words, and N temporal-memory words. It processes a block of P pixels in every cycle and requires (MN/P) cycles to compute 1-level 2-D DWT of an image of size M × N, where M= PQ/2. It has an initial delay of nine cycles, where one cycle is defined as T = TM + 2TA.

TABLE I

COMPARISON OF HARDWARE AND TIME COMPLEXITIES OF THE PROPOSED AND THE EXISTING STRUCTURES FOR 1-LEVEL LIFTING 2-D DWT USING 9/7 FILTERS. [(M) Image height and (N) Image width,(P) Block Size]

Structures	Mult	Add	Reg	Transpositi on memory	On-chip memory	Multiplexer
Cheng	12	16	24	N	4N+ 24	2
Lai	10	16	44	N	4N+44	2
Tian	6P	8P	10P	N(P+2)/2	8P+N(P+8)/2	P(6+(P-2)/2)
Mohanty	4.5P	8P	7.5P	2.5P	7.5P+5.5N	3P
Proposed	4.5P	8P	8P	N	4N+8P	0

LEGEND : Mult : Multiplier, Add : Adder, Reg: Register

TABLE II

COMPARISON OF SYNTHESIS RESULTS OF THE PROPOSED STRUCTURE AND THE EXISTING STRUCTURES FOR IMAGE SIZE (512 × 512) AND POWER ESTIMATED AT 20-MHZ CLOCK.

Designs	**Block size**	**Area (u.sqm)**	**Power (mw)**	**Delay (ns)**	**I/O pins**
Lai	2	987629.82	8.973	11.06	66
Tian	4	1593435.98	13.416	15.52	84
Mohanty	4	1385941.29	11.6415	16.95	83
Proposed	4	987500.22	8.34027	16.11	79

From the above Table-II, for the block sizes of various existing structures, the proposed structure involves less area, power, delay and less number of I/O pins at block size 4.

TABLE III

COMPARISON OF SYNTHESIS RESULTS OF THE PROPOSED STRUCTURE FOR DIFFERENT IMAGE SIZES AND POWER ESTIMATED AT 20-MHZ CLOCK

Input images	**Block size**	**Area (u.sqm)**	**Power (mW)**	**Delay (ns)**	**I/o pins**
IMAGE SIZE:8X8					
girl.bmp	4	16,592.2	5.008	5.067	23
rice.png	4	16,972.3	5.09	5.15	21
Sunset.jpg	4	16,157.0	5.95	6.16	25
Mri.tif	4	15,786.4	5.32	5.09	20
IMAGE SIZE:16X16					
girl.bmp	4	47,756.2	6.07	6.12	36
rice.png	4	49,456.0	6.08	6.1	39
Sunset.jpg	4	51,999.0	6.3	6.45	41
Mri.tif	4	48,876.6	6.09	6.1	37

IMAGE SIZE:256 X 256					
girl.bmp	4	3,56,456	8.01	8.1	65
rice.png	4	3,61,456	8.00	8.09	61
Sunset.jpg	4	3,62,678	8.11	8.13	64
Mri.tif	4	3,59,006	8.005	8.123	63

IMAGE SIZE:64X64					
girl.bmp	4	87,234.1	7.94	11.1	48
rice.png	4	85,234.0	7.85	10.95	53
Sunset.jpg	4	86,546.7	7.89	11.04	50
Mri.tif	4	85,500	7.92	11.29	47

IMAGE SIZE:512X512					
girl.bmp	4	9,87,500.22	8.34027	16.11	79
rice.png	4	9,86,008	8.314	15.98	77
Sunset.jpg	4	9,87,070.7	8.3390	16.10	78
Mri.tif	4	9,85,446.09	8.312	16.09	76

Synthesis result various images of different sizes in Table-III shows that the proposed structure for block size 4 is regular and is modular.

V. Conclusion

A new data access scheme for the computation of lifting 2-D DWT (discrete wavelet transform) using systolic arrays with block processing is suggested. From DG (dependence graph) linear systolic array is directly derived. for the parallel and pipeline implementation of 1-D DWT from suitably segmented DG is used for deriving 2-D systolic arrays. Above two systolic arrays are used as building blocks to derive the lifting 2-D DWT. The proposed structure involves only a small on-chip memory of size (4N + 8P) and processes a block of P samples in every cycle, where N is the image width. The proposed structure involves the same number of multipliers and adders and 1.5N less on-chip memory and synthesis result shows that the proposed structure of image size 512 x 512 having a block size of 4 with area is 987500.22 u.sqm, power consumed is 8.34027 mw and delay count is 16.11 ns is better than the best of the existing structures. The proposed structure is regular, modular, and can be easily configured for different image sizes.

References

[1] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal.*, vol. 3, no. 2 , Apr. 1996, pp. 186–200. http://dx.doi.org/10.1006/acha.1996.0015

[2] P. C. Wu and L.G. Chen, "An efficient architecture for two-dimensional discrete wavelet transform," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, Apr. 2001,pp. 536–545. http://dx.doi.org/10.1109/76.915359

[3] 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, Apr. 2005, pp. 1575–1586. http://dx.doi.org/10.1109/TSP.2005.843704

[4] C.Y. Xiong, J.W. Tian, and J. Liu, "Efficient architecture for twodimensional discrete wavelet transform using lifting scheme," *IEEE Trans. Image Process.*, vol. 16, no. 3, , Mar. 2007, pp. 607–614. http://dx.doi.org/10.1109/TIP.2007.891069

[5] C.C. Cheng, C.-T. Huang, C.Y. Cheng, C.J. Lian, and L.G. Chen, "Onchip memory optimization scheme for VLSI implementation of line-based two dimensional discrete wavelet transform," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 17, no. 7,, Jul. 2007, pp. 814–822. http://dx.doi.org/10.1109/TCSVT.2007.897106

[6] Y.K. Lai, L.F. Chen, and Y.C. Shih, "A high-performance and memoryefficient VLSI architecture with parallel scanning method for 2-D lifting based discrete wavelet transform," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, May 2009, pp. 400–407. http://dx.doi.org/10.1109/TCE.2009.5174400

[7] B. K. Mohanty and P.K.Meher, "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, , May 2011,pp. 2072– 2084. http://dx.doi.org/10.1109/TSP.2011.2109953

[8] X.Tian, L.Wu, Y.H. Tan, and J.-W. Tian, "Efficient multi-input/multioutput VLSI architecture for two-dimensional lifting-based discrete wavelet transform," *IEEE Trans. Comput.*, vol. 60, no. 8, Aug. 2011 ,pp. 1207–1211. http://dx.doi.org/10.1109/TC.2010.178