

# VLSI Architecture of Low Memory and High Speed 2-D Lifting-Based Discrete Wavelet Transform for JPEG2000 Applications

Jen-Shiun Chiang, Chih-Hsien Hsia, Hsin-Jung Chen, and Te-Jung Lo  
Department of Electrical Engineering, Multimedia IC Design Lab.  
Tamkang University  
Tamsui, Taipei, Taiwan  
E-mail: {chiang, chhsia}@ee.tku.edu.tw

**Abstract**—This work presents a low memory and high speed VLSI architecture for 2-D lifting-based lossless 5/3 filter discrete wavelet transform (DWT). The architecture is based on the proposed interlaced read scan algorithm (IRSA) and parallel scheme processing to achieve low memory size and high speed operation. The proposed lifting-based DWT architecture has the advantages of lower computational complexity, transforming signal with extension, and regular data flow, and is suitable for VLSI implementation. It can be applied to real time image/video operating of JPEG2000 and MPEG4 applications. Basing on the proposed architecture, we designed and simulated a 2-D DWT VLSI chip by 0.35 $\mu$ m 1P4M CMOS technology. The memory requirement of the  $N \times N$  2-D DWT is  $N$  and it can operate at 100MHz clock frequency.

## I. INTRODUCTION

In the last few years, discrete wavelet transform (DWT) [1] has been used for a wide range of applications including image coding and compression, speech analysis, pattern recognition, and computer vision. DWT can be viewed as a multi-resolution decomposition of a signal. This means that it decomposes a signal into several components in different frequency bands. It always needs a large amount of computations and memory to perform the DWT. In order to achieve the real-time processing, reducing of memory and computational complexity and increasing the efficient hardware utilization are necessary.

In this paper, we propose an efficient 2-D lifting-based DWT VLSI architecture which uses lossless 5/3 filter processing. Memory requirement and hardware utilization are the major concerns of the DWT implementation. For an  $N \times N$  2-D lifting-based DWT, people like to use row-wise data flow operation. Under this approach the memory requirement is ranged from  $3.5N$  to  $N^2$  [2, 8, 11, 12, 13, 14]. Lee et al. [12] proposed a new data flow operation approach for the DWT implementation, and they used only memory

size of  $N$  for an  $N \times N$  2-D DWT. However, according to the characteristics of their data flow operation, the operating clock rate is only 85 MHz. In this work we revise the data flow from row-wise to mixed row-wise and column-wise, and further propose a new approach, interlaced read scan algorithm (IRSA), to reduce the memory requirement for a 2-D DWT. In the IRSA approach, for an  $N \times N$  DWT, it needs only memory size of  $N$ . Parallel schemes are adopted in our 2-D DWT architecture to increase the operation speed. We also use shifters and adders to replace multipliers in the computation to accomplish high hardware utilization. The advantages of the proposed DWT have the characteristics of higher hardware utilization, less memory requirement, and regular data flow.

The rest of this paper is organized as follows. In Section II, we briefly describe the DWT and lifting scheme. Then, the proposed IRSA approach and the efficient VLSI architecture for the 2-D DWT are presented in Section III. Section IV compares the simulation result with other existed 2-D DWT architectures. Finally, brief summaries are given in Section V to conclude this paper.

## II. DISCRETE WAVELET TRANSFORM

In a classical DWT, filtering and convolution are applied to achieve the signal decomposition. Meyer and Mallat [4] found in 1986 that the orthonormal wavelet decomposition and reconstruction could be implemented in the multi-resolution signal analysis framework. The multi-resolution analysis is now a standard method to construct the orthonormal wavelet bases. In the analysis process, Figure 1 shows the classical 2-D DWT. For the decomposition process, the low-pass filter  $H$  and high-pass filter  $G$  are constructed from the scaling functions and the corresponding wavelets.

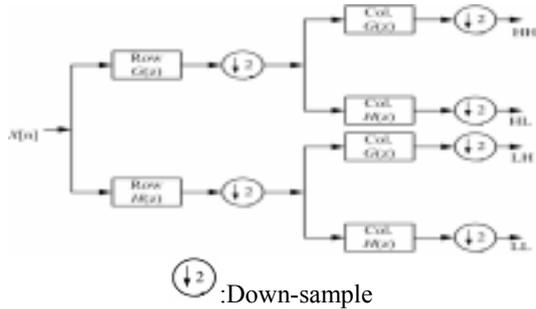


Figure 1. 2-D DWT decomposition

Suppose that the filter length  $L$  is 4, and the transfer functions of filters  $H$  and  $G$  can be represented as follows:

$$H(z) = h_0 + h_1z^{-1} + h_2z^{-2} + h_3z^{-3} \quad (1)$$

$$G(z) = g_0 + g_1z^{-1} + g_2z^{-2} + g_3z^{-3} \quad (2)$$

The down-sampling operation is used to decimate half of the filtered results. Each level of the wavelet transform decomposes the input signal into Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH) signals. Let the input image ( $X$ ) be of size  $N \times N$ , and the numbers of rows and columns at each stage location are shown in Figure 2.

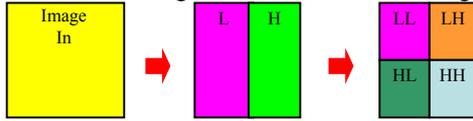


Figure 2. 2-D 1-Level DWT image decomposition

### A. Lifting representation

The lifting scheme proposed by Daubechies and Sweldens [5-7] requires fewer computations compared to the conventional convolution-based approach. We can use the Euclidean algorithm to factorize the poly-phase matrix of a DWT filter into a sequence of alternating upper and lower triangular matrices and a diagonal matrix. In Equation (3),  $h(z)$  and  $g(z)$  are the low-pass and high-pass synthesis filters, and these synthesis filters can be divided into even and odd parts to form a poly-phase matrix  $P(z)$  as defined in Equation (4).

$$\begin{aligned} g(z) &= g_e(z^2) + z^{-1} g_o(z^2) \\ h(z) &= h_e(z^2) + z^{-1} h_o(z^2) \end{aligned} \quad (3)$$

$$P(z) = \begin{bmatrix} h_e(z) & g_e(z) \\ h_o(z) & g_o(z) \end{bmatrix} \quad (4)$$

Since  $h(z)$  and  $g(z)$  form a complementary filter pair,  $P(z)$  can be factorized into Equation (5).

$$P(z) = \prod_{i=1}^m \begin{bmatrix} 1 & Si(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ ti(z) & 1 \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & 1/k \end{bmatrix} \quad (5)$$

Therefore, we can factorize the filter bank into three lifting steps. As illustrated in Figure 3, a lifting scheme is composed of the following four stages:

1. Split: Divide the original signal into two disjoint subsets. Note that  $X_e$  stands for the set of even samples and  $X_o$  stands for the set of odd samples.

2. Predict: The predicting operator  $P$  is applied on the subset  $X_o$  to obtain the wavelet coefficients  $d[n]$ , as shown in Equation (6).

$$d[n] = X_o[n] + P \times (X_e[n]) \quad (6)$$

3. Update: We can combine  $X_e[n]$  and  $d[n]$  to obtain the scaling coefficients  $s[n]$  after an update operator  $U$ , as shown in Equation (7).

$$s[n] = X_e[n] + U \times (d[n]) \quad (7)$$

4. Scaling: In the last step, we apply the normalization factor on  $s[n]$  and  $d[n]$  to obtain the wavelet coefficients. Equations (8) and (9) describe the implementation of the 5/3 lifting analysis DWT. Equation (8) is used for the computation of the odd coefficients (high-pass coefficients), whereas Equation (9) is used for the computation of the even coefficients (low-pass coefficients).

$$d^*[n] = X(2n+1) - \lfloor X(2n) + X(2n+2)/2 \rfloor \quad (8)$$

$$s^*[n] = X(2n) + \lfloor d(2n-1) + d(2n+1) + 2/4 \rfloor \quad (9)$$

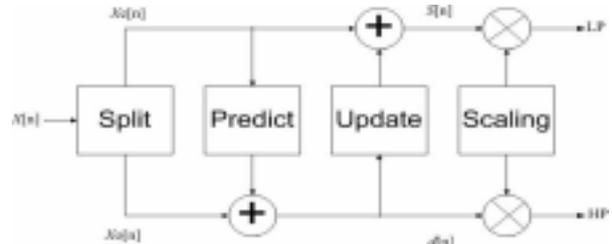


Figure 3. Block diagram of the lifting-based DWT

### B. Boundary treatment

In order to keep the number of the wavelet coefficients the same as that of the original signal, an appropriate signal extension is necessary. The embedded data extension algorithm [10] can be used to compute the boundary of the image. The data extension is needed only at the beginning and ending of the process. For JPEG2000 standard data extension should be symmetrical. Without loss of generality, the lossless 5/3 lifting-based DWT of JPEG2000 is discussed here. The dependence graph (DG) is used to describe the corresponding lifting factored wavelet filter, and is shown in Figure 4. Figure 4 depicts a basic DG to describe a lifting step constructed by the specific lifting factorization, and the DG operation is  $H1 = B - (1/2)(A+C)$ ,  $L0 = A + (1/4)(H1+H1+2)$ , as shown in Figure 4.

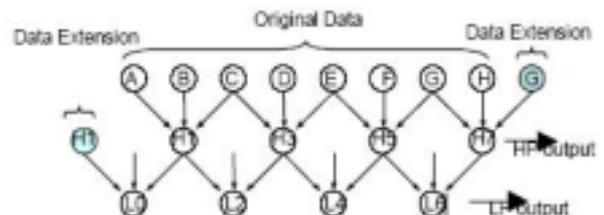


Figure 4. Dependence graph of the 5/3 Lifting-Based DWT

### III. PROPOSED VLSI ARCHITECTURE FOR THE 2-D DWT

During the last decade, although many 2-D DWT architectures have been proposed, most of them need a transpose memory for the decomposition, and the memory size is proportional to the size of the image. Recently people like to use the lifting-based approach to implement the 2-D DWT [2, 8-14]. In the lifting DWT scheme, the row data are processed first and then these data must be stored in an external memory block for matrix transpose. The high-frequency and low-frequency data are reserved for the column data transformation, and then the previous row data stored in the external memory block are used together for the column data transformation. If the pixels of the input image are  $N \times N$ , a  $N^2$  memory block is needed. Because the large memory size, the processing time is long and it may cause high power consumption.

Since the major problem of a DWT is the requirement of a large memory block (an  $N \times N$  2-D DWT needs a memory block of  $N^2$ ), to reduce the memory requirement becomes a key issue. In order to improve the efficiency of a DWT we propose a new approach called interlaced read scan algorithm (IRSA). In a conventional lifting-based DWT, the data operation is row-wised, and therefore it needs an  $N^2$  memory for matrix transpose. The IRSA reads data row-by-row but outputs data column-by-column, and therefore it needs no transpose memory. For a 1-D DWT calculation, DWT takes 3 pixels to find a DWT coefficient, and we can use this characteristic to build the IRSA. The IRSA sequentially reads three pixels to calculate the new DWT coefficient, and then it jumps to the next row to read the next three pixels to find the DWT coefficient, and then it jumps to the following row and does the same operation. When it finishes the last row's operation, it goes back to the first row to read the following three pixels to find the corresponding DWT coefficient, and then it jumps to the next row to do it all over again till it finishes the calculation of the whole image. Let us take an example. If we have a  $5 \times 3$  image, IRSA first reads pixels 1, 2, and 3 in the first row to calculate the DWT coefficient, and then it jumps to row two to read pixels 1, 2, and 3 to calculate the second DWT coefficient, and then it jumps to row three to find the DWT coefficient. After it finishes the operation of row three, it goes back to row one to read pixels 3, 4, and 5 to find the DWT coefficient, and then it jumps to row two to read pixels 3, 4, and 5 to calculate the corresponding DWT coefficient, and then it jumps to row 3 to finish the calculation of the DWT coefficient. In order to increase the operation speed, we adopt the parallel scheme in our IRSA architecture. Here two processors are used and therefore it can operate two rows of data a time. An example of a  $5 \times 5$  image IRSA 1-D DWT calculation is shown in Figure 5. By the IRSA approach, the data flow is changed to row-by-row

input and column-by-column output and thus it does not need transpose memory.

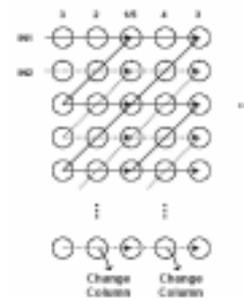


Figure 5. Proposed interlaced read scan algorithm (IRSA) of the 2-D DWT

From the 1-D DWT operation, the high and low frequency coefficients can be calculated, and these data can be fed to some FIFO's (first-in-first-out queue) for the 2-D operation. Based on the IRSA approach, a 2-D DWT processor is proposed and is shown in Figure 6. There are input signal unit, multiplication and accumulation unit (MAC), control unit (counter and multiplexer), and queue unit in the proposed 2-D DWT processor. The input signal unit is used to assign the input pixel. The MAC unit is used to perform specific mathematical function calculations required for the wavelet transform, and the block diagram is shown in Figure 7. In order to improve the hardware utilization, the proposed 2-D DWT architecture consists of overlap schemes and thus the multiplier is replaced by shifters and adders. All these arrangements cannot only reduce the hardware cost but also increase the operation speed. The queue unit is used for temporary data storage.

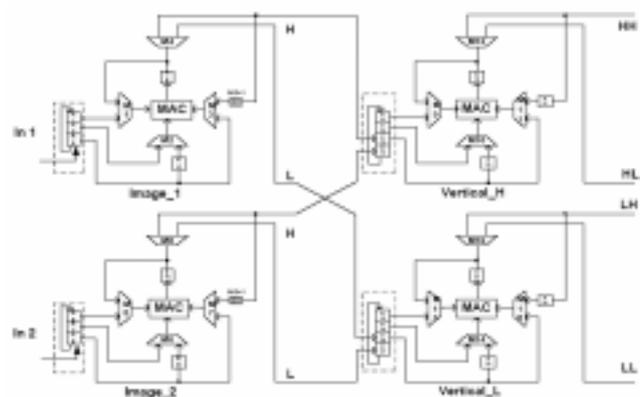


Figure 6. The proposed the architecture of 2-D DWT

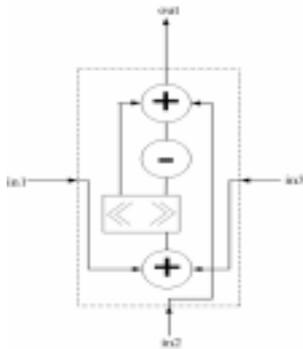


Figure 7. Proposed multiply-and-accumulate (MAC) unit

#### IV. RESULTS AND COMPARISONS

Based on IRSA and the proposed 2-D DWT architecture, a  $128 \times 128$  lifting-based lossless  $5/3$  filter DWT is designed. We designed and simulated the DWT by the  $0.35\mu\text{m}$  1P4M standard CMOS process. The operating frequency of this designed DWT can reach 100MHz. For an  $N \times N$  2-D DWT operation, our proposed architecture takes  $(3/4)N^2 + 6$  cycles. For an  $8 \times 8$  2-D DWT it takes 54 clock cycles to finish the calculation. The comparisons of the memory size and maximum clock rate of the proposed DWT and some typical DWT's are listed in Table I. Our proposed DWT has the advantages of low-memory and high-speed.

Table I  
Comparisons

| Proposed    | Wave stage | 2-D DWT | Memory size | Max. clock rate | years       |
|-------------|------------|---------|-------------|-----------------|-------------|
| <b>Ours</b> | 5/3        | Lifting | N           | <b>100MHz</b>   | <b>2004</b> |
| [2]         | 5/3        | Lifting | $N^2$       | N/A             | 2000        |
| [8]         | 5/3        | Lifting | $3.5N$      | N/A             | 2000        |
| [11]        | 5/3        | Lifting | $3.5N$      | N/A             | 2002        |
| [12]        | 5/3        | Lifting | N           | 83MHz           | 2003        |
| [13]        | 5/3        | Lifting | $3.5N$      | N/A             | 2001        |
| [14]        | 5/3        | Lifting | $2.5N$      | N/A             | 2002        |

#### V. CONCLUSION

We propose an efficient VLSI architecture for 2-D lifting-based lossless  $5/3$  filter discrete wavelet transform (DWT) in this paper. The architecture is based on the proposed interlaced read scan algorithm (IRSA) and a parallel scheme is adopted. It has the characteristics of low-memory requirement, regular memory access order, and high-speed operation. Based on the architecture, a  $128 \times 128$  2-D lifting-based lossless  $5/3$  filter DWT is designed by the  $0.35\mu\text{m}$  1P4M standard CMOS technology. For an  $N \times N$  DWT, our proposed architecture reduces memory significantly to only memory size of  $N$  and can operate at 100 MHz clock frequency. The experimental results show that the architecture is very suitable for various real-time video/image applications such as JPEG2000 and MPEG4.

#### REFERENCES

- [1] S. Mallat, "A theory for multi-resolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 11, no. 7, pp. 674-693, July 1989.
- [2] JPEG 2000 Part 1 Final Committee Draft Version 1.0, ISO/IEC JTC1/SC29/WG1N1646R.
- [3] ISO/IEC JTC1/SC29/WG1 WgIn 1684, "JPEG 2000 Verification Model 7.0," 2000.
- [4] S. G. Mallat, "Multi-frequency channel decompositions of images and wavelet models," *IEEE Trans. On Acous. Speech and Proc.* ASSP-37, 12, pp. 2091- 2110, 1989.
- [5] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting scheme," *The J. of Fourier Analysis and Applications*, vol. 4, pp. 247- 269, 1998.
- [6] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Applied and Computation Harmonic Analysis*, vol. 3, pp.186- 200, 1996.
- [7] W. Sweldens, "The lifting scheme: A construction of second-generation wavelets," *Siam Journal of Mathematical Analysis*, vol. 29, pp. 511- 200, 1997.
- [8] K. Andra, C. Chakrabarti and T. Acharya, "A VLSI Architecture for Lifting-Based Wavelet Transform," *IEEE Workshop on Signal Processing Systems*, pp. 70-79, 2000.
- [9] C.-J. Lian, K.-F. Chen, H.-H. Chen, and L.-G. Chen, "Lifting based discrete wavelet transform architecture for JPEG2000," *IEEE International Symposium on Circuits and Systems*, pp.445- 448, May 2001.
- [10] K.C.B. Tan and T. Arslan "An embedded extension algorithm for the lifting based discrete wavelet transform in JPEG2000," *ICASSP*, vol.4, pp.3513-3516, 2002.
- [11] K. Andra, C. Chakrabarti and T. Acharya, "A VLSI architecture for lifting-based forward and inverse wavelet transform," *IEEE Trans. on Signal Processing*, vol. 50, pp. 966- 977, 2002.
- [12] We-Tan Lee, Hug-Ying Chen, Pei-Yung Hsiao and Chia-Chun Tsai, "An Efficient Lifting Based Architecture for 2-D DWT Used in JPEG2000," *Proceeding of the 2003 VLSI Design/ CAD Symposium*, pp. 577- 580, August 2003.
- [13] C. Diou, L. Torres and M. Robert, "An embeded core for the 2-D Wavelet Transform," *IEEE on Emerging Technologies and Factory Automation*, vol. 2, pp. 179-186, 2001.
- [14] Shung-Chih Chen and Chung-Cheng Wu, "An Architecture of 2-D 3-Level Lifting-based Discrete Wavelet Transform," *Proceeding of the 2002 VLSI Design/ CAD Symposium*, pp. 351- 354, August 2002.