

A Framework for FPGA based Discrete Biorthogonal Wavelet Transforms Implementation

Isa Servan Uzun and Abbes Amira
School of Computer Science
The Queen's University of Belfast
Belfast, BT7 1EN, United Kingdom
Email: isu@ieee.org

Abstract—During the last decade, the wavelet transform has proven to be a valuable tool in many application fields including telecommunication, numerical analysis and most notably image/video compression. This paper describes a high-level framework dedicated to the implementation of 1-D and 2-D Discrete Biorthogonal Wavelet Transforms (DBWTs) on FPGAs. The system hides the low level hardware details of the FPGA structure and thus allows the user to concentrate more on the experimentation rather than on the low-level architecture. The DBWT architectures proposed within the framework are scalable, modular and have less area and time complexity when compared with existing structures. FPGA implementation results based on a Xilinx Virtex-2000E device have shown that proposed system provides an efficient solution for the processing of DBWTs in real-time.

I. INTRODUCTION

In the last decade Discrete Wavelet Transforms (DWT) have become powerful tools in a wide range of applications including image/video processing, numerical analysis and telecommunication. The advantage of DWT over existing transforms such as Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) is that the DWT performs a multiresolution analysis of a signal with localization in both time and frequency [1], [2].

In 1992, Cohen, Daubechies and Feauveau established the theory of biorthogonal wavelet systems [3]. Biorthogonal wavelets have been found to offer improved coding gain and an efficient treatment of boundaries in image coding applications [4]. A significant difference between the orthonormal and biorthogonal wavelet transform lies in the Quadrature Mirror Filter (QMF) relationship. The high-pass filter coefficients in the orthonormal filters are the QMF of the low-pass filter. Biorthogonal filters lack this property.

Due to the demand for real time wavelet processing in applications such as video compression, internet communications, compression, object recognition, and numerical analysis, many architectures for DWT have been proposed[5], [6]. Most of the effort towards the design and hardware implementation (in the form of VLSI and FPGA) of wavelet transforms has been concentrated on the orthonormal wavelet family. One of the main reasons for this is that orthonormal wavelets were the first functions to be implemented in the form of filter banks, whereas biorthogonal wavelet functions are relatively new. Secondly, the properties of biorthogonal wavelet functions

are much more diverse, which complicates the development of their generic architecture. A good survey of the different schemes used for the development of DWT architectures can be found in a recent paper by Weeks and Bayoumi[6].

Reconfigurable hardware, in the form of FPGAs, has been touted as a new and better means of performing high performance computing. FPGA structures provide a reconfigurable hardware with flexible interconnections, with field-programmable ability, which are widely used for rapid prototyping of DSP and computer systems. However, users must program FPGAs at a very low level and have a detailed knowledge of the architecture of the device being used. They do not therefore facilitate easy development of, or experimentation with, signal/image processing algorithms [7].

The main objectives of the research work presented in this paper can be described as follows:

- Developing efficient and scalable VLSI architectures for 1-D and 2-D biorthogonal wavelet transforms, where both area and speed can be estimated with specific design parameters;
- Developing a library of biorthogonal wavelet transforms targeting FPGAs, which can be extended for other types of wavelets; and
- Developing a high-level framework to try to reconcile the dual requirements of high performance and ease of development by enabling the system designers to experiment conveniently with different wavelet filters to investigate the best area/speed trade-offs, rather than concentrating on the low-level and complex structure of FPGAs.

The composition of the rest of the paper is as follows. The proposed system for the implementation biorthogonal wavelets is presented in section 2. The architectures for 1-D and 2-D DBWTs are described in section 3. The results and analysis for the FPGA implementations of the architectures are presented in section 4. Concluding remarks are given in section 5.

II. PROPOSED SYSTEM

The proposed system for mapping the DBWTs on the FPGA as shown in figure 1 consists of:

- Graphical User Interface (GUI): The GUI supports experimentation with different parameters to enable the user to explore system performance e.g. speed, area. The input

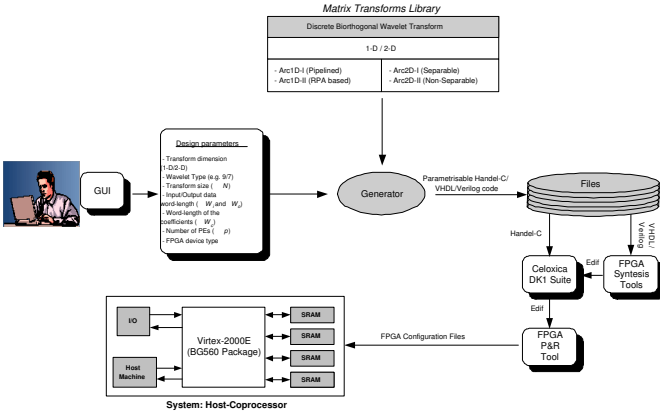


Fig. 1. Rapid prototyping environment for discrete biorthogonal wavelet transforms.

parameters required for the generation of design files include:

- the DBWT dimension (1-D or 2-D);
- the DBWT architecture type;
- the DBWT filter length (e.g. 1-D 9-tap, 2-D 9/7 pair);
- the transform length (N);
- the input and output data wordlength (W_i and W_o) and
- the coefficient wavelength (W_c).
- **DBWT Library** : The library includes the architecture for 1-D and 2-D DBWTs. The application has the ability to choose and download existing files and to generate new files and save them.
- **Generator**: The generator automatically download the necessary modules and then generates the top-level design files given the user selected parameters and settings;
- **FPGA Coprocessor** : The celoxica's RC1000 FPGA-based development board is based on the Xilinx XCV2000E of the Virtex-E family [8].

It is worth mentioning here that although the target hardware in this work is RC1000 board with Xilinx XCV2000E Virtex FPGA, the architecture designs are completely portable and can be implemented on any type of FPGA chip with the use of the proposed system.

III. DBWT ARCHITECTURES

A. Arc1D-I: A Balanced Pipelined Architecture

The 1-D DBWT can be pipelined into J Processing Elements PE_j ($1 \leq j \leq J$) where each PE_j is responsible for the computation of the decomposition level j . The complexity of the decomposition level j is linear with the number of input samples N_j . Because of the decimation by two, the complexity in each decomposition level can be expressed as:

$$C_j = C_{j+1} \quad (1)$$

Therefore, in order to constitute a balance pipelined 1-D DBWT architecture, each PE_j should consist of M_j number

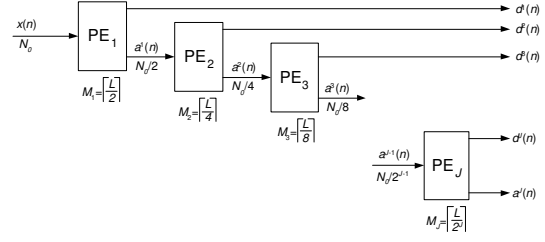


Fig. 2. Arc1D-I: Top-level architecture for 1-D DBWT.

of multipliers where $M_j = 2 \cdot M_{j+1}$.

Since, PE_1 uses $M_1 = \lfloor \frac{L}{2} \rfloor$ which leads to design of a PE_1 having a period of $N_0/2$ ccs with 100% efficiency, each PE_j should have:

$$M_k = \lfloor \frac{L}{2^j} \rfloor \quad j = 1, 2, \dots, J \quad (2)$$

The top-level architecture [9] of pipelined 1-D DBWT is shown in figure 2.

B. Arc1D-II: A Hybrid-Pipeline Architecture

As described in the previous section, 1-D DBWT can be pipelined into J PE_j ($1 \leq j \leq J$), each PE_i being devoted to compute the decomposition level j . Nevertheless, the downsampling occurring in each decomposition level makes the fully-pipelined architectures heavily underutilised, since the stage implementing the decomposition level j is usually clocked by a frequency 2^{j-1} times lower than the clock frequency used in the first level.

In order to avoid this underutilisation, we propose a hybrid-pipeline architecture for 1-D DBWT consisting of two PEs [10]. PE_1 is devoted to perform the first level of decomposition ($j=1$), while the second PE_2 is responsible for the higher level of decompositions ($2 \leq j \leq J$) based on Modified-Recursive Pyramid Algorithm (MRPA) approach. A top-level scheme of the architecture is given in figure 3.

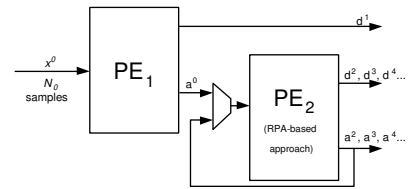


Fig. 3. Arc1D-II: Top-level architecture for 1-D DBWT.

C. Arc2D-I: Separable MRPA-based Architecture

The proposed separable 2-D DBWT architecture is a modified version of the direct architecture described in [11] and it is shown in figure 4. The separable 2-D DBWT architecture consists of a delay line, a filter bank and a memory unit of J register blocks (R_j) in order to store intermediate outputs. The systolic filters, PE_1 , are based on the work presented

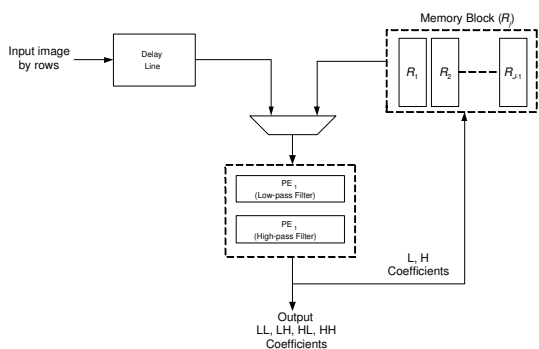


Fig. 4. Arc2D-I: Top-level architecture for separable 2-D DBWT.

in [4]. It exploits the decimation by 2 in wavelet transform and anti/symmetrical property of biorthogonal wavelet filters. Therefore, it achieves a reduction in the number of multipliers by a factor of 4. The memory unit consists of J register blocks, each storing $N_j x (L - 1)$ words, where L is the filter length and N_j is the input data size at decomposition level j . By organising the memory into blocks, the coefficients are automatically transposed into column major format. The inputs to the filter bank are row-based and multiplexed between the output of the delay line and the output of the memory unit. By doing so, a simple control and routing can be achieved without the need for a N^2 memory units.

The computation of different levels of decomposition is scheduled according to row-based RPA scheduling [12]. The entire row of input image (or intermediate LL, LH, HL and HH results) is fed into filter bank at a time. This scheduling uses a buffer to store a single row of LL^j coefficients for each decomposition level required.

D. Arc2D-II: Non-Separable 2-D DBWT Architecture

In the non-separable (or direct) approach, the decomposition is computed by four 2-D convolutions followed by a decimation by 2 in both horizontal and vertical directions which can be mapped into the proposed architecture shown in figure 5 [13]. The design of the non-separable 2-D discrete biorthogonal wavelet filter architecture has been derived from MRPA-like based architecture in [14]. MRPA-like based architecture exploits the down sampling of output sub-bands and perform the first decomposition level interspersed with all other levels by means of only one processing unit. The top-level architecture for the (9x7) non-separable 2-D biorthogonal wavelet is shown in figure 5.

The architecture is composed of a set of $\lceil L/2 \rceil$ (7-tap) 1-D filter processors (P_i) and J sets of row delay circuits, R_j being used for the j^{th} level of decomposition where $j=0,1,\dots,J$. Each row-delay circuit R_j is composed of a pipe of $(L-1)$ row-delay elements with $N/2^j$ memory cells. R_0 stores the rows of the input image while $R_{j>0}$ are used to store the rows of LL^j subband, which are used as input for computing the decomposition level $j+1$.

The even-numbered and odd-numbered rows of the input image are fed simultaneously into processors P_{2i} and P_{2i+1}

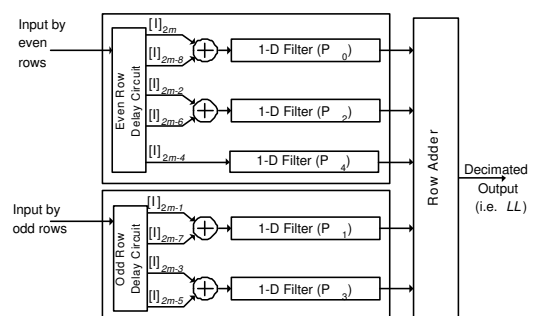


Fig. 5. Arc2D-II: Non-separable (9x7)-tap 2-D DBWT top-level architecture. The even and odd numbered rows of the input 2-D frames are fed simultaneously with the use of splitted data pipes.

in a word-serial fashion by using two distinct row-delay pipes so that the decimated output can be directly computed. Computation of the different levels of decomposition is scheduled according to an algorithm that differs from the MRPA since the processors in this architecture require the parallel input of odd and even rows [14]. The k^{th} row of each subband at the decomposition level j , $[LL^{(j)}]_{2k}$ can only be computed when two adjacent rows from the previous level ($[LL^{(j-1)}]_{2k}$ and $[LL^{(j-1)}]_{2k+1}$) are produced and stored in $[R_{j-1}]_0$ and $[R_{j-1}]_1$. The rows LH^j , HL^j and HH^j are immediately output, while the row LL^j is fed back and stored either in $[R_j]_0$ if k is even or in $[R_j]_1$ if j is odd.

IV. IMPLEMENTATION RESULTS

In order to verify the performance of the DBWT architectures, the biorthogonal wavelet filter designs have been ported to a Virtex-2000E FPGA chip (package: bg560, speed grade 6) using Handel-C [15]. The designs of the architectures have been parameterised in terms of:

- Filter length (L , LxM)
- Input image size (N)
- Number of decomposition levels (J)
- Input and output data wordlength (W_i and W_o)
- Filter coefficients wordlength (W_c)

A 2's complement data format has been used in the implementations. Basically, each processor in the architecture comprises multiply-accumulate operations. The multiplication of two numbers comprising of W_i and W_c bits, results in the output having a wordlength of W_i+W_c bits. To prevent overflows, the wordlength in the "addition path" is defined as $(W_i+W_c+W_a)$, where W_a is the number of adders. However, the Handel-C model simulations showed that for a 9-bits input data wordlength, the wavelet coefficients from all levels are bounded by 2^{16} . Therefore, output data wordlength of 16 bits is enough to represent wavelet coefficients without causing any overflow.

FPGA implementation performances for 1-D DBWTs and comparison with the existing FPGA implementations is shown in table I in terms of area and maximum clock frequency (f_{max}). It can be seen that the proposed 1-D DBWT architectures compares favorably in terms of area and speed

TABLE I

COMPARATIVE EVALUATION OF 1-D DBWT ARCHITECTURES FOR AN L -TAP BIORTHOGONAL FILTER: AREA AND COMPUTATION TIME.

Design	Latency	Multipliers	FPGA Implementation	
			Area (Slices)	f_{max}
Pipelined [16]	N_0	$J \cdot \lceil L/4 \rceil$	785	85.49
PA [17]	JN_0	$\lceil L/4 \rceil$	n/a	n/a
Bit-level [18]	JN_0W_i	$\lceil L/2 \rceil$	69	70.2
Systolic [19]	$N_0/2$	$\lceil L/2 \rceil$	n/a	50
Ach1D-I	$N_0/2$	$\sum_{j=1}^J \lceil L/2^j \rceil$	453	159.058
Ach1D-II	$N_0/2$	$\lceil L/2 \rceil + \lceil L/4 \rceil$	453	159.058

in comparison to implementations proposed in [16] and [17]. Although the bit-level implementations in [18] present a better area/speed ratio, high computation time and low-throughput rate features make them unsuitable for high-speed/high-throughput applications.

Since this study presents the first hardware implementation of non-separable 2-D wavelet transforms on FPGAs in the literature, we can provide performance comparison only with existing separable 2-D DBWT based FPGA implementations. The performance comparisons are shown in Table II based on the criteria of wavelet type, FPGA area, memory, and latency. It can be seen from the table that f_{max} for proposed implementation is nearly as twice as the other implementations. This result has been achieved due to the simple routing and control circuit of the proposed architecture. Also, the proposed architecture has a latency of $2/3N^2$ ccs which is at least %33 better than the others. Although it requires at least %15 more FPGA slices compared to other implementations, the proposed architecture can be used at twice lower frequency ($f_{max}/2$) but still achieving the same performance (in terms of computation time) with the other implementations.

V. CONCLUSION

Reconfigurable hardware, usually in the form of FPGAs, offers an attractive combination of low cost and high performance computing combined with an apparent flexibility. Although, users must program FPGAs at a very low level and have a detailed knowledge of the architecture of the device being used, they remain very good target devices for rapid prototyping. This paper describes the development of a general framework for FPGA based biorthogonal wavelet transforms implementation. With the use of proposed system, efficient

TABLE II

2-D DBWT PERFORMANCE COMPARISON WITH EXISTING FPGA-BASED SEPARABLE DBWT IMPLEMENTATIONS.

Design	Wavelet	FPGA Area			Latency(ccs)
		Slices	BRAMs	f_{max}	
Amphion[20]	9/7 Lifted	3784	24	55	N^2
Cast[21]	9/7 Conv	2293	14	50	$5.7 \cdot N^2$
Cast[22]	5/3 Lifted	971	10	51	$3 \cdot N^2$
McCanny[12]	9/7 Conv	2559	17	44.1	$1.5 \cdot N^2$
Arc2D-I	9/7 Sep	2221	24	78	$2 \cdot N^2$
Arc2D-II	9/7 Non-Sep	4348	24	105	$2/3 \cdot N^2$

scalable and modular 1-D and 2-D DBWT architectures can be automatically generated and mapped on to FPGAs targeting real-time signal and image processing applications.

REFERENCES

- [1] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, 1990.
- [2] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.
- [3] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 45, no. 5, pp. 485–560, 1992.
- [4] S. Masud and J. McCanny, "Finding a suitable wavelet for image compression applications," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '98)*, vol. 5, 1998, pp. 2581–2584.
- [5] C. Chakrabarti, M. Vishwanath, and R. Owens, "A survey of architectures for the discrete and continuous wavelet transforms," *Journal of VLSI Signal Processing Systems*, vol. 3, no. 43, pp. 171–192, 1996.
- [6] M. Weeks and M. Bayoumi, "Discrete wavelet transform: Architectures, design and performance issues," *J. VLSI Signal Process. Syst.*, vol. 35, no. 2, pp. 155–178, 2003.
- [7] A. Amira, "A custom coprocessor for matrix algorithms," Ph.D. dissertation, The Queen's University of Belfast, United Kingdom, 2000. [Online]. Available: <http://www.cs.qub.ac.uk/a.amira>
- [8] *RC1000 Development Platform Product Datasheet*, Celoxica. [Online]. Available: www.celoxica.com
- [9] I.S.Uzun, A. Amira, and A. Bouridane, "A high-speed/low-power architecture for 1-d discrete biorthogonal wavelet transform," in *The 46th The IEEE International Midwest Symposium on Circuits and Systems (MWSCAS 2003)*, 2003.
- [10] —, "An efficient architecture for 1-D discrete biorthogonal wavelet transform," in *IEEE International Symposium on Circuits and Systems (ISCAS 2004)*, vol. 2, 2004, pp. 697–700.
- [11] M. Vishwanath, R. Owens, and M. Irwin, "VLSI architectures for the discrete wavelet transform," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 42, no. 5, pp. 305–316, 1995.
- [12] P. McCanny, S. Masud, and J. McCanny, "Design and implementation of the symmetrically extended 2-d wavelet transform," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 3, 2002, pp. 3108–3111.
- [13] I.S.Uzun and A. Amira, "Design and FPGA implementation of non-separable 2-D biorthogonal wavelet transforms for image/video coding," Accepted for presentation at 2004 IEEE International Conference on Image Processing (ICIP 2004), 2004.
- [14] F. Marino, "Two fast architectures for the direct 2-D discrete wavelet transform," *IEEE Trans. on Signal Processing*, vol. 49, no. 6, pp. 1248–1258, 2001.
- [15] *Handel-C Language Reference Manual*, Celoxica. [Online]. Available: www.celoxica.com
- [16] S. Masud and J. McCanny, "Reusable silicon IP cores for discrete wavelet transform applications," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 51, no. 6, pp. 1114–1124, 2004.
- [17] S. Masud, "VLSI systems for discrete wavelet transforms," Ph.D. dissertation, The Queen's University of Belfast, United Kingdom, 1999.
- [18] A. B. M. Nibouche and O. Nibouche, "Rapid prototyping of biorthogonal discrete wavelet transforms on fpgas," in *Proceedings IEEE International Symposium on Circuits and Systems (ISCAS'01)*, vol. 3, 2001, pp. 1399–1402.
- [19] J. Jou, Y. Shiau, and C.-C. Liu, "Efficient VLSI architectures for the biorthogonal wavelet transform by filter bank and lifting scheme," in *Proceedings IEEE International Symposium on Circuits and Systems (ISCAS'01)*, vol. 2, 2001, pp. 529–532.
- [20] *CS6210 Discrete Wavelet Transform Core Datasheet*, Amphion. [Online]. Available: <http://www.amphion.co.uk>
- [21] *LB.DFDWT Line-Based Programmable Forward DWT Core Datasheet*, Cast Inc. [Online]. Available: <http://www.cast-inc.com>
- [22] *BB.DFDWT Block-Based Forward Discrete Wavelet Transform Core Datasheet*, Cast Inc. [Online]. Available: <http://www.cast-inc.com/>