ISSN: 2278 – 909X
International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 4, Issue 5, May 2015

A Survey on Systolic Array Multiplier and its Implementation on FPGA

Pritam H. Langade, S. B. Patil

Abstract*— The evolution of computer and Internet has brought demand for powerful and high speed data processing, but in such complex environment some methods can provide pristine solution. To handle above addressed situation, parallel computing is proposed as a solution to the contradiction. This paper review the implementation issues in systolic array multiplier for high speed data processing. This also introduces the concept of column compression and pipelining which improves the speed of execution. Multiplication is most commonly used operation in data processing. Systolic algorithms are the efficient algorithms to perform the binary multiplication. Systolic array is an arrangement of processors in an array where data flows synchronously across the array between the processors This data usually flows in different directions. Each processor at each step receives in data from one or more neighbours (e.g. North and West), processes it and, in the next step, we get outputs in the opposite direction (South and East).

Index Terms*—- Parallel Computing, Pipelining, Systolic Array, Multiplication, Algorithm, Column Compression.

I. INTRODUCTION

High-performance, special-purpose computer systems are typically used to meet specific application requirements or to off-load computations that are especially taxing to general-purpose computers. Systolic arrays are used in image and signal processing graph algorithms, solutions of differential equations and in other computationally intensive tasks. As hardware cost and size continue to drop and processing requirements become well-understood, more special-purpose systems are being constructed. In the recent technology, there is a need of high speed and powerful data processing. This complex problem is overcome by using parallel computing technology which uses the concept of pipelining for this application. Multiplication is most commonly used operation in mathematics [9]. Binary multiplication is the basic multiplication used for the integer multiplication. Systolic algorithms are the efficient algorithms to perform the binary multiplication. Systolic algorithm is the form of pipelining, which is in more than one dimension. In this algorithm data flows from a memory in a rhythmic way and it passes through many processing elements before it returns to memory. Systolic array is used in every Sequential algorithm that can be transformed to a parallel version which are very easy to run on array processors that execute operations in the systolic way and systolic array is one of solutions to the requirement for a highly parallel computational power. Systolic array is such an architecture which is highly scalable. A systolic architecture is an array composed of matrix like rows of cells. Here, the Processing Elements is equivalent to central processing units (CPUs) (except for the usual lack of a program counter, instruction register, control unit etc. Each cell shares the information with its neighbors immediately after processing. Systolic arrays have balanced, uniform, grid like architectures in which each line indicates a communication path and each intersection represents a cell or systolic elements. In a systolic array, all systolic cells perform computations concurrently while data like initial inputs, partial results, and final outputs, is being passed from cell to cell. When partial results are moved across the cells, they are calculated over these cells in a pipeline manner.

Recently, Field Programmable Gate Arrays (FPGAs) became a platform of alternative for hardware realization of computation-intensive applications. Especially, when the planning at hand needs very high performance, designers will like high density and high performance. FPGAs modify a high degree of parallelism and can meet orders of magnitude speedup over GPPs. This is as a result of the increasing embedded resources on FPGA. FPGA have the advantages of the hardware speed and the software flexibility, jointly they have a price/performance ratio much more favorable than Application Specific Integrated Circuits (ASICs). Since the key resources for implementing computation-intensive algorithms are embedded on FPGA, latency related to device communication has been eliminated. However, these embedded resources are limited hence it is important to use these resources optimally [2]. Specification of FPGA configuration is done by using a hardware description language. Verilog is the hardware description language used for designing as well as simulation purposes. FPGAs contains logic blocks (flip flops, gates, memory elements)that are used to implement any logic functionality.

II. LITERATURE SURVEY

In order to achieve the demand of high speed and low power in DSP applications, parallel array multipliers are widely used. These multipliers possess common properties like regularity, locality and reclusiveness. Previous work was targeted on the literature survey conducted on various

Manuscript received
Pritam H. Langade, M.E. Student, Department of Electronics and Tele-communication, S.S.G.M.C.E, Shegaon, S.G.B. Amravati University, Amravati(Maharashtra State),India.
S. B. Patil, Associate Professor, Department of Electronics And Tele-communication, S.S.G.M.C.E, Shegaon, S.G.B. Amravati University, Amravati(Maharashtra State),India

multipliers in VLSI. It included the study of the basic technical aspects behind the design approaches of proposed systolic array multiplier.

Systolic array is defined as a connected set of processors with rhythmical data computation and propagation along the system. They result in cost effective high performance special purpose system for a wide range of problems. By using this systolic array, concurrency and communication is achieved and special purpose design cost can be reduced by the use of special architecture. But one must provide a convenient means for incorporating high performance systolic processors into a complete system [1].

The systolic array designs are optimized for speed. After studying these designs, it is concluded that for multiplication of large matrices memory architectures is quiet efficient than systolic array. It requires several clock cycles [2].

There are various systolic architectures. These gives solution for the addressed issues of demand for high speed data processing. Also the number of components required for the matrix multiplication in conventional method is more than that of required in systolic method. For example critical path delay in conventional method is 9.831ns and in systolic array, it is 4.757ns.This system enhance the speed and reduce the complexity. It requires less number of clock cycles. It requires more number of accumulator than conventional method [3].

In sequential algorithm, the complexity depends upon the required computation and storage capacity. The effectiveness of the algorithm may be improved by using special updating techniques. Several parallel computer architecture may vary according to the applied processing elements, reconfigurability, data interchange connections. Digital image processing is a data-oriented computing problem, so architectures with global data interchange are to be omitted. In this paper, triangular array of QR decomposition is used for speed up performance. Less number of processors are required and they are n(n-1)/2. The basic approach to mapping techniques and some possible applications are presented in this paper. It is efficient in sequential algorithm but less efficient in parallel execution [4].

The approach of an optimized systolic array architecture for Full Search Block Matching Algorithm is shown. This Array Architecture is implemented by RTL-level VHDL for using as a motion estimation unit in low bit rate and real-time applications such as video telephony. AS1 architecture for Full Search Block Matching Algorithm and a modification of AS1 is implemented on FPGA. The results show that the FPGA is suitable for real-time motion estimation implementation. The results also show that the area occupied on a FPGA for motion estimation is about %11 of the chip area for Vertex family and %51 for Spartan family (N = P = 8). Optimization results show that this modification decreases hardware about %20 without any effect in computation time and operating frequency. With this approach, it is possible to develop systems with high efficiency and low price in a short period of time on a single FPGA chip [5].

The two dimensional systolic array architecture with serial input is used to perform Full Search Block Matching Algorithm. The system has single clock and reset. Parallel matrix multiplication on systolic array is characterized by processing data input in the pipeline and it is made up of regular array of PE which are connected with each other by shortest line and therefore mass data has no need to be stored before it is processing. It takes inputs serially to reduce IO pin counts and processes data in parallel. The best match selection unit outputs the address value which is the top left corner of the best match position for the present reference block [6].

Binary multiplier has intrinsic regularity and simplicity and may be extended for any number of bits. It is modification of the array of full adder scheme which allows the operation with two's complement numbers in systolic mode.The relatively poor performance obtained comes from use of automatic placement and routing srtatagy and standard cells since it breaks the inherent regularity of the systolic array.To reduce the delay and area, one should use registers with capacitive effects.Although the use of standrd cell is useful to validate the design, only full custom implementation would exploit the potential characteristics[7].

Multiplication of two n bit numbers plays a very important role in various applications and with the advancement in VLSI technology, parallel algorithms for multiplication is becoming very important.So, the algorithms using the technique of column compression is designed.This technique leads to less execution time than iterative array algorithms.The desined algorithms has almost regular interconnections between two types of cell and hence they are suitable for single chip VLSI implementation[8].

The problem of implementing fixed point matrix multiplication algorithm as deeply nested loops and then mapping the algorithm to parallel architecture on FPGA is considered in this paper. The design methodology is based on a parallel array design that maps a nested loop algorithm onto the parallel architecture. Parallel array design effectively exploits the inherent parallelism of the matrix multiplication. FPGAs can be used efficiently to implement these fine grain arrays since they inherently possess the same regular structure. The proposed architecture provides better speed as compared to previous implementation of matrix multiplication and it requires less area. A high throughput architecture is developed using advanced design techniques[12].

To use in a linear, purely systolic array forming a digit-serial multiplier for unsigned or 2'complement operand, a very simple multiplier is used. Two digit product term is produced by each cell and accumulated these into a previous sum of the same weight, developing the product least significant digit first. It is also shown how the multiplier, with some simple back-end connections can compute modular inverses and perform modular division for a power of two as modulus [13].

The design of 4 bit Systolic Array Multiplier working with frequency of 211.035 MHz and the proposed design is optimized using structural style compared with behavioural style. If this prototype is implemented in real time then there will n number of advantages benefited to the mankind [14].

Faster column compression multiplication is achieved by using technique of partition of the partial products into two parts for independent parallel column compression and acceleration of the final addition using a hybrid adder. The performance of the proposed multiplier is analyzed by evaluating the delay, area and power, with 180 nm process technologies on interconnect and layout using industry standard design and layout tools. Also the power-delay

product of the proposed design is significantly lower than that of the regular Dadda multiplier[15].

The effective design for binary multiplication using modified booth's algorithm and systolic multiplier is discussed in this paper. The design is simulated using modelsim by mentor graphics tool and Xilinx is used for the implementation of the code in FPGA [16].

III. PROPOSED WORK

In computer architecture, a systolic architecture is a pipelined network arrangement of Processing Elements (PEs) called cells. It is a specialized form of parallel computing, where cells compute the data which is coming as input and store them independently. In systolic multiplication, to carry out the multiplication and get the final product, the multiplicand and multiplier are arranged in the form of array.

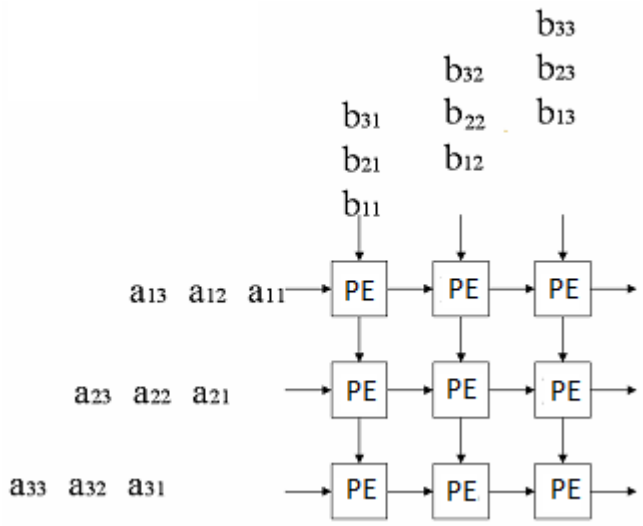

Fig.1: Systolic Array Multiplier.

The systolic array shown in the above figure, takes inputs in parallel, perform parallel processing and outputs the result. Each unit is an independent processing performing the required operations on the data. Thus computation takes place simultaneously within the rows and columns. All the inputs are applied simultaneously, therefore registers are not required. Availability of inputs at the start of the computation and the systolic array structure helps the multiplier to perform fast compared to other multipliers.

For example, when we multiply two 3*3 matrix, we need 27 operations according to the given formula,

```
For I = 1 to N
    For J = 1to N
        For K= 1 to N
          C[ I,J ] = C[ I,J ] + A [ J,K ]*B [ K,J ] ;
        End
    End
End
```

But by using systolic array, it can be calculated by only 9 clock pulses.

High speed multiplication is a basic requirement of digital systems giving high performance. Parallel multiplication schemes are developed for the fulfillment of this need . There are two classes of parallel multipliers, which are array multipliers and tree multipliers. Tree multipliers are also known as column compression multipliers[8], well known for their higher speeds making them very useful in high speed computations. Their propagation delay is proportional to the logarithm of the operand word length. In array multipliers delay is directly proportional to operand word length. Column compression multipliers are faster than array multipliers. They have an irregular structure and hence their design is difficult. With the improvement in VLSI design techniques and process technology, designs which were previously infeasible or too difficult to be implemented by manual layout can now be implemented through automated synthesis. Two of the most well-known column compression multipliers have been presented by Wallace and Dadda . Both architectures are similar with the difference occurring in the procedure of reduction of the partial products and the size of the final adder[11]. In Wallace scheme, the partial products are reduced as soon as possible[10]. On the other hand, Dadda's method does minimum reduction necessary at each level. Since the Dadda multiplier has a faster performance, we implement the proposed techniques using this multiplier.

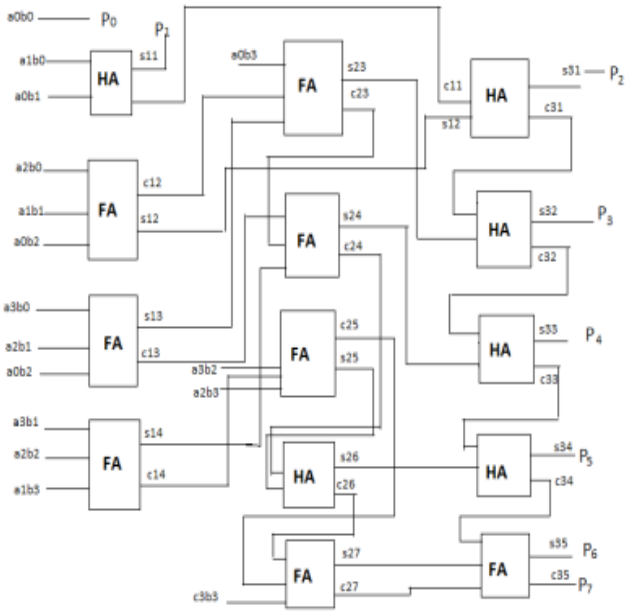

Fig.2 : 4 bit Dadda multiplier.

Here, we are multipliying two 3*3 matrix whose elements are of 4 bit.Each bit of multiplicand is multiplied with each bit of multiplier to obtain the partial products. The partial products of the same column are added along with carry generated. So the ultimate output by adding partial products and the carry, is the final product of the two binary numbers. This multiplication is carried out by Dadda multiplier which is the processing element of proposed architecture. Partial products are compressed in such a structural arrangement that they will provide high speed.

IV CONCLUSION

In current technology, there is a great influence on high speed and powerful data processing to meet the requirement of minimum delay. This complex problem is overcome by using parallel computing technology which uses the concept

of pipelining.Systolic array is the form of pipelining which is highly scalable and fast architecture. So they are widely used in various applications like matrix multiplication, filter designing, convolution, image processing. The proposed architecture is designed for high speed by using the concept of column compression. It is simulated and is targeted to the Field Programmable Gate Array device. The parallel processing and pipelining is introduced into the proposed systolic architecture to enhance the speed and reduce the complexity of the multiplier.

REFERENCES

[1] H. T. Kung, "Why Systolic Architecture?" IEEE Computer,15(1),(1982)37-46.
[2] Syed M. Quasim,Ahmed A. Telba and Abdulhameed Y. AIMazroo," FPGA implementation of matrix multiplier architectures for use in image and signal processing application", IJCSNS International Journal of Computer Science And Network Security, VOL.10 No.2,February 2010.
[3] Mahendra Vucha, Arvinda Rajawat ,"Design and implementation of systolic array architecture for matrix multiplication", IJCA International Journal of Computer Applications (0975-8887),Volume 26-No.3,July 2011.
[4] Ziad Al-Qadi and Musbah Aqel., "Performance analysis of Parallel Matrix Multiplication Algorithms Used in Image Processing",World Applied Sciences Journal 6 (1) : 45-52, 2009.
[5] Mohammad Mahdi Azadfar , "Implementation of A Optimized Systolic Array Architecture for FSBMA using FPGA for Real-time Applications", IJCSNS International Journal of Computer Science and Network Security, VOL.8 NO.3,March 2008.
[6] Ganapathi Hegde, Cyril Prasanna Raj P , P.R. Vaya: "Implementation of Systolic Array Architecture for Full Search Block Matching Algorithm on FPGA", European Journal of Scientific Research,Vol.33 No.4(2009),pp.606-616.
[7] Arechabala,E.I.Boemo, J.Meneses, F.Moreno,C. Lopez Barrio: "Full systolic binary multiplier",8 Oct 1991.
[8] Bhabani P. Sinha, Pradip K. Srimani, "Fast Parallel Algorithms for Binary Multiplication and Their Implementation on Systolic Architectures", IEEE Transactions on computers, Vol.38. No.3. March 1989.
[9] Kurtis T. Johnson and A.R.Hurson, Pennsylvania State University, "General Purpose Systolic Arrays" IEEE, Nov.1993.
[10] K.Gopi Krishna, B.Santhoshand V.S ridhar, "Design of Wallace Tree Multiplier using Compressors", International journal of engineering sciences & research Technology,September 2013.
[11] Jasbir Kaur and Kavita, " Structural VHDL Implementation of Wallace Multiplier" International Journal of Scientific & Engineering Research, Volume 4, Issue 4, April-2013 1829 ISSN 2229-5518.
[12] Syed Manzoor Qasim, Shuja Ahmad Abbasi and Bandar Almashary, " A Proposed FPGA-based Parallel Architecture for Matrix Multiplication" 978-1-4244-2342-2/08/$25.00 ©2008 IEEE.
[13] Peter Komerup, Member, IEEE, "A Systolic, Linear-Array Multiplier for a Class of Right-Shift Algorithms", IEEE TRANSACTIONS ON COMPUTERS, VOL. 43, NO. 8, AUGUST 1994.
[14] Bairu K. Saptalakar, Deepak kale, Mahesh Rachannavar , Pavankumar M. K, "Design and Implementation of VLSI Systolic Array Multiplier for DSP Applications", International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) Volume 2 Issue 3, PP : 156-159 1 April 2013.
[15] B. Ramkumar, V. Sreedeep and Harish M Kittur, Member, IEEE, " A Design Technique for Faster Dadda Multiplier ".
[16] Himani Harmanbir Singh Sidhu, "Design and Implementation Modified Booth algorithm and systolic multiplier using FPGA" International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 11, November – 2013.