International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958, Volume-1, Issue-6, August 2012

Design and FPGA Implementation of Systolic Array Architecture for Matrix Multiplication

Rakesh Birle, Lalit Bandil

Abstract: Matrix multiplication is the kernel operation used in many image and signal processing applications. This paper demonstrates an effective design for the Matrix Multiplication using Systolic Architecture. This architecture increases the computing speed by using the concept of parallel processing and pipelining into a single concept. The selected platform is a FPGA (Field Programmable Gate Array) device since, in systolic computing, FPGAs can be used as dedicated computers in order to perform certain computations at very high frequencies. The description language used as an entry tool to model the hardware architecture is VERILOG HDL.

Key words: FPGA implementation, Matrix multiplication, Systolic Arrays, VERILOG HDL.

I. INTRODUCTION

Matrix multiplication which is use in many image and signal processing is a computationally intensive problem, especially the design and efficient implementation on an FPGA where resources are very limited, has been more demanding. In recent years, FPGAs have been improved considerably in speed, density, and functionality, which makes them ideal for system-on-a-programmable-chip (SOPC) designs for a wide range of applications. Designers can benefit from high density and high performance FPGAs instead of costly multicore Digital Signal Processing (DSP) systems.[1-3]

In this work we demonstrate an effective design and efficient implementation of the matrix Multiplication using systolic Architecture and Ancient mathematics. For efficient implementation and maximum speed-up, integer arithmetic was used. Three main steps of the work, i.e. design, simulation and implementation, were accomplished.

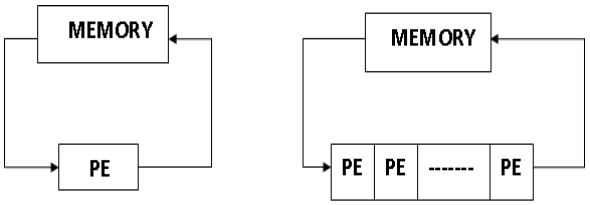
This paper is organized as follows. Section 2 present the matrix multiplication using proposed Systolic Architecture. Section 3 present Implementation Scheme. Section 4 present results and discussion and finally conclusion are presented in section 5.

II. PROPOSED ARCHITECTURE

Systolic array are regular arrays of simple finite state machine, where each finite state machine in the array is identical. A systolic algorithm relies on data from different directions arriving at cells in the array at regular intervals and being combined. It is a class of parallel pipelined architecture. by this pipelining processing it may proceed concurrently with input and output ,and consequently over all execution time is minimized.

Systolic architecture is characterized by processing data input in pipeline and comprised of regularly arrayed PE. Where neighbor PEs are connected with each other by shortest line and therefore mass data has no need to be stored before processing. Decrease of distance between the PEs in an array greatly reduces the internal communication delay and improves the utility of processing units. It also removes time consumption for controlling the establishment of data stream [6-9].

The basic principle of systolic array is doing multiple operations in one second [5, 7].



A
100ns
5MOPS

B
100ns
30MOPS

A basic principle of conventional design
B basic principle of Systolic array design

Figure 1. Basic principle

III. IMPLEMENTATION SCHEME

A general block diagram for systolic array is shown in figure. Here UART (Universal Asynchronous Receiver/Transmitter) receive a serially data coming from PC and transmit them parallel in to the memory. The data will be store either in memory A or memory B which is control by memory controller. After storing the data in to the memory it will proceed for matrix multiplication. The output of matrix multiplication is store in to the memory C and the output of memory C is further transmitting in PC through UART.

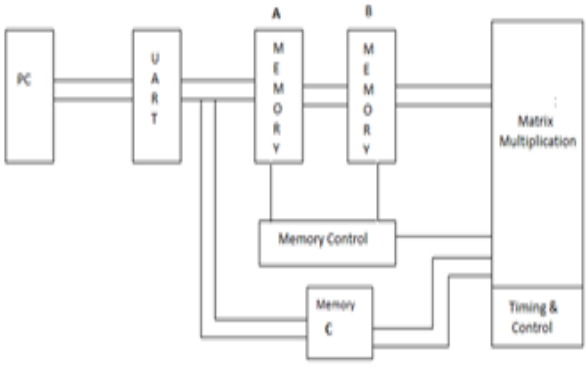


Figure 2. Block diagram of systolic array

In this paper, our aim to compute the equation (1) with a two dimensional systolic array.

$$C_{mxn} = A_{mxk} \times B_{kxn} \qquad \text{------- (1)}$$

Where A, B and C are the matrices with order m x k, and k x n respectively. Each PE of systolic array computes the

Manuscript received on August, 2012.

Rakesh Birle, [Mtech IVth Sem] Specialization in embedded system & vlsi, AITS Academy, Indore (M.P.) India.

Lalit Bandil, Associate Professor (Dept. of Electronics and Communication Engg.) AITS Academy, Indore (M.P.) India.

multiplication of elements and accumulates to the corresponding element and then elements will be passed to neighbor PE in the systolic array. First elements $a_{i,j}$ in row i of matrix A are injected first into PE as pipeline with the sequence of $a_{i,K}$ and the input time to the element of $a_{i+1,J}$ is one time unit later than $a_{i,j}$ .Similarly, elements $b_{i,j}$ in column j of matrix B are injected first into PE as pipeline with the sequence of $b_{K,j}$ and the input time to the element of the sequence of $b_{k+1,j}$ is one time unit later than $b_{k,j}$, The architecture of PE in this approach is shown in figure which perform multiplication and accumulation on data.

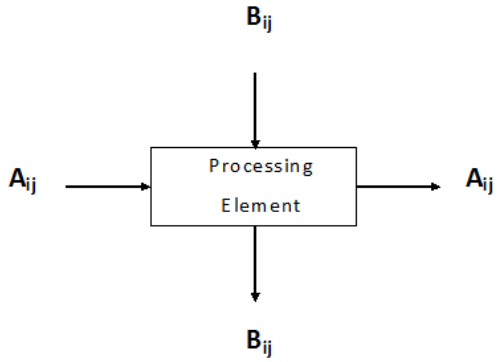


Figure 3. PE of Systolic Architecture

3.1. Systolic array architecture:

A systolic architecture is an array of Processing Elements, each called as a cell. Each cell is connected to a small number of nearest neighbors in a mesh like topology. Each cell performs a sequence of operations on data that flows between them. PE at each step takes input data from one or more neighbors (e.g. Left and Top), processes it and, in the next step, outputs results in the opposite direction (Right and Bottom) [9]. The Proposed two dimensional systolic Architecture for 3 by 3 matrixes is given in Fig 4.

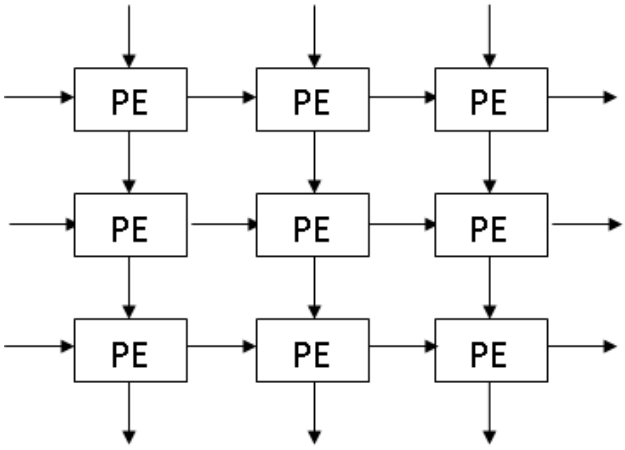


Figure 4. Two-dimensional Systolic Array

In this architecture each processing elements are performing following operation.

(1) FETCH

(2) MULTIPLICATION

(3) SHIFT

(4) ADDITION

Data is fetch from respective memory that is Memory A and Memory B, which is then passed to multiplication of a cell and also shifted to other respective cell. The result of

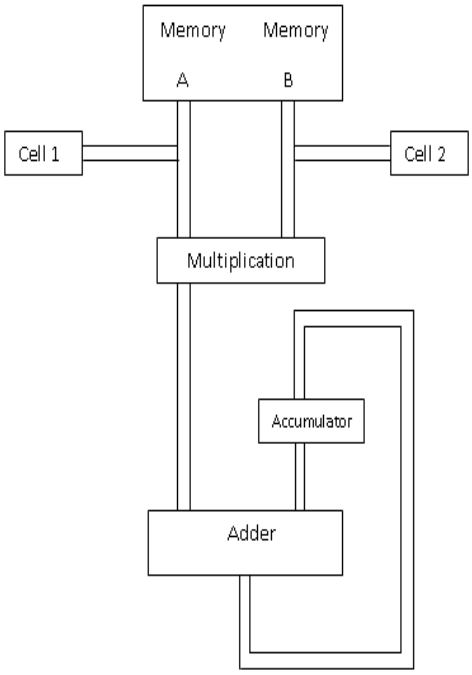


Figure 5. Functional block diagram of cell

IV. RESULT AND DISCUSSION

The implementation of Matrix Multiplication is done in both methods i.e. Conventional and Systolic Architecture, as described above, on FPGA but The simulation results have given that, the Systolic architecture implementation requires less number of clock cycles then Conventional method. The simulation result of systolic architecture is shown in figure 6. This result exposes the parallel processing and pipelining by the systolic array architecture and also the input and output matrices A_{3X3}, B_{3X3} and C_{3X3} respectively. Where the matrix elements are of 8 bit each.

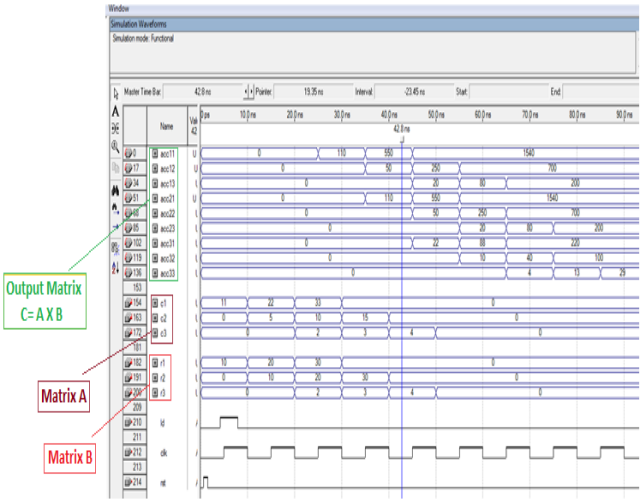


Figure 6. Simulation waveform of systolic array architecture for matrix multiplication

REFERENCE

[1] S. Ogrenci, A. K. Katsaggelos, and M. Sarrafzadeh, "Analysis and FPGA Implementation of Image restoration under resource constraint," IEEE Trans. on Computers, Vol. 52, No. 3, pp. 390-399, 2003.

[2] C. Ebeling, C. Fisher, G. Xing, M. Shen, and H. Liu, "Implementing an OFDM Receiver on the Rapid Reconfigurable Architecture," IEEE Trans. on Computers, Vol. 53, No. 11, pp. 1436-1448, 2004.

[3] G. R. Goslin, "A Guide to Using Field Programmable Gate Arrays for Application-Specific Digital Signal Processing Performance," Microelectronics Journal, Vol. 28, Issue 4, pp. 24-35, 1997.
[4] T. Komarek, P. Pirsch, "Array architectures for block matching algorithms," IEEE Transactions on Circuits and Systems, Vol. 36, No. 10, October 1989.
[5] Kuani Lee, Algorithm and VLSI architecture design for H.264/AVC Inter Frame Coding, A PhD Thesis at National Cheng Kung University, Tainan, Taiwan, in 2007.
[6] Ganapathi Hegde, Cyril Prasanna Raj P and 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] Feifei Dong, Sihan Zhang and Cheng Chen, Improved Design and Analyse of Parallel Matrix Multiplication on Systolic Array Matrix, IEEE, 2009.
[8] Ziad Al-Qadi and and Musbah Aqel, performances Analysis of Parallel Matrix Multiplication Algorithms Used in Image Processing, World Applied Sciences Journal 6 (1): 45-52, 2009.
[9] 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.