Design of Systolic Array Architecture for Matrix Multiplication Inference on FPGA

Mahendra Pratap Singh[1] Nikhil Kumar[2] Deepam Dubey[3]
[1,2]M.Tech Student [3]Scientist B
[1,2,3]Department of Electronics Design & Technology
[1,2,3]NIELIT, Gorakhpur, India

Abstract— For data transmission, an ideal communication relies on Low Power Design. Systolic array multiplier with logic is widely known among those techniques for synchronizing signals in DSP processor applications. Low power circuit design yields many favorable conditions like increased performance, system capacity, minimized cost etc. Among basic arithmetic operations, Multiplication demands more processing time and seek complex hardware. As Conventional Systolic Array Multiplier is designed using logic gates, there is undesired power dissipation. . FPGA is used to obtain simulation results. Power for the proposed design when the power of individual components. In recent years, the FPGA implementation has attracted much attention because of its high performance and energy efficiency. However, existing implementations have difficulty to fully leverage the computation power of the latest FPGAs So, to improvise this downside, this paper illuminates the design of low power Systolic Array Multiplier using logic gates which performs data processing in concurrent manner. Power dissipation, time delay, outputs and cost is calculated mathematically. In this paper we implement on an FPGA using a systolic array architecture, which can achieve high clock frequency under high resource utilization. We provide an analytical model for performance and resource utilization and develop an automatic design space exploration.

Key words: Systolic Array Architecture, Matrix Multiplication Inference, FPGA

I. INTRODUCTION

Parallel array multipliers are used to achieve low power and high speed. Multipliers are devolved to meet the needs in various DSP systems. The basic principle behind multiplier is multiplication in algorithmic and structural levels[1]. DSP applications are mainly designed with low power dissipation Multipliers. Systolic algorithm is linear sequence of channels, sometimes in multi-dimension. The Systolic Array is the linear arrangement of the blocks or units like the pipelining processing. It is a homogeneous network. Each block in the arrangement computes a partial result from the inputs given to it from top of the array and the result is stored within itself. Later, the stored result is passed to the adjacent blocks down the array in the arrangement. The blocks in the design are usually similar and fixed. The operations at each stage in the arrangement are done simultaneously, which increases the speed and reduces the processing time without compromising on the results. They are used to perform correlation, matrix multiplication, data sorting tasks and convolution. In this paper a 3-bit systolic array multiplier using gate logic is designed and implemented. In systolic array multiplication [2], the multiplicand and the multiplier are linearly arranged like in an array and each bit of both multiplier is multiplied with the multiplicand to obtain partial products. The carry is generated from the column. The final output is obtained by adding the partial products and carry. While transmitting data certain amount of energy is dissipated for each bit lost. The formula for calculating the energy dissipated for loss of each bit is

KT*log2

Where T is absolute temperature and K is Boltzmann's constant. Heat dissipation in a digital circuit can be minimized or eliminated by performing all the computations based on logic. It improves the overall performance of the circuitry by allowing higher densities and higher speeds by reducing power dissipation. In this paper the design flow of Systolic Array Multiplier circuit is explained and it is divided into three sections. Existing work is discussed in Section II. The Systolic array is given in Section III. The mathematical calculations for Systolic Array Multiplier is given in Section IV and conclusion is given in section V.

II. EXISTING WORK

In the existing paper, Systolic Array Multiplier circuit is implemented using conventional logic gates. To perform multiplication of bits by consuming the low power, the circuit is developed with logic gates. This multiplication process is enhanced by Pipelining process and involves data transmission in high Parallelism. The multiplier involves minimum 4 stages which are required to implement 3*3 systolic array multiplier. Circuits store the bits rather than throwing them away. The bits which are not computed are used to trace back the inputs from outputs. For the proposed design in this paper, the parameters like power and delay, cost, number of gates, outputs based on the definitions given.

III. SYSTOLIC ARRAY

Systolic array in parallel computer architectures [3], a systolic array is a homogeneous network of tightly coupled data processing units (DPUs) called cells or nodes. Each node or DPU independently computes a partial result as a function of the data received from its upstream neighbors, stores the result within itself and passes it downstream. The parallel input data flows through a network of hard-wired processor nodes, which combine, process, merge or sort the input data into a derived result. Because the wave-like propagation of data through a systolic array resembles the pulse of the human circulatory system, the name systolic was coined from medical terminology. The name is derived from systole as an analogy to the regular pumping of blood by the heart.

Architecture Replace single processor with an array of regular processing elements Orchestrate data flow for high throughput with less memory Access Different from pipelining Nonlinear array structure, multi direction data flow

[4], each PE may have (small) local instruction and data memory as shown in figure 1.

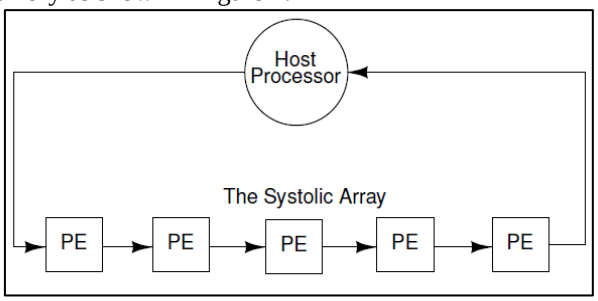

Fig. 1: Systolic Array Architecture

- Different from SIMD: each PE may do something different.
- Initial motivation: VLSI enables inexpensive special-purpose chips.
- Represent algorithms directly by chips connected in regular pattern.
- 3x3 Systolic Array Matrix Multiplication Processors arranged in a 2-D grid.
- Each processor accumulates one element of the product.

IV. PROPOSED THE MULTIPLIER CELL BLOCK

To achieve our proposed circuit, the cascading takes place in between gates and Full adder cells. It has two gates and 1 Full Adder gate Figure 2 explains detail schematic of Multiplier cell block.

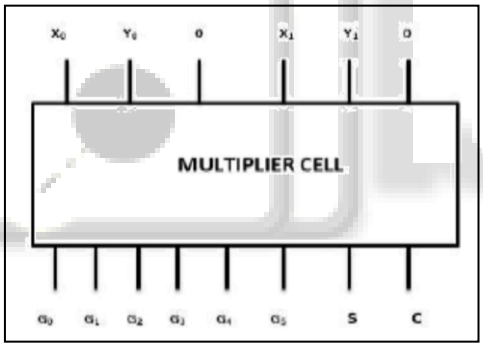

Fig. 2: Multiplier Cell Block

To design this Multiplier, it requires three multiplier cells, seven Gates, nine Full Adder are used, which succeeds in optimizing power by keeping the cost as low as possible and eliminating outputs. It has 6 inputs and 6 outputs. The 6 inputs are data bits. The output is 6 bits of Sum and 1 bit of Carry. The architecture of Systolic array multiplier with gate logic is shown in figure 3.

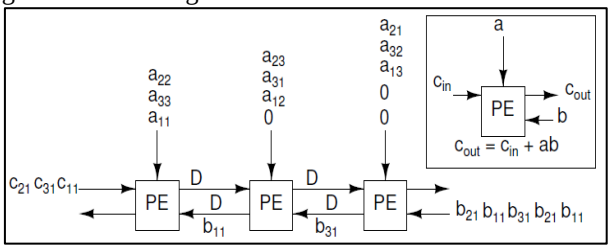

Fig. 3: Systolic Array Multiplier with Gate Logic

V. MATHEMATICAL ANALYSIS FOR SYSTOLIC ARRAY MULTIPLIER

Number of Gates (NOG) - Number of gates is calculated by counting all the gates used in individual cells. In designing Full Adder gate, two cascaded gates are used. So, to calculate total number of gates, the expression is given as follows

NOGSYS = NOGMC + NOGPFA + NOGTG
= 4[2(TG) + 1(PFA)] +12(PFA) + 9(TG)
= 33 (1)

A. Output Calculation

The total Output for the proposed Systolic Array Multiplier with gate Logic circuit is given by the following expression

GOSYS = GOMC + GOPFA+GOTG
= 4□6+ 12□2 + 9□2
= 66 (2)

B. Calculating Cost

Cost for the proposed Systolic Array Multiplier with Gate Logic is

QCSYS = QCMC + QCPFA + QCTG
= 4[2(QCTG) + 1(QCPFA)] + 12(QCPFA) +9(QCTG)
= 4[2(5) + 1(8)] + 12(8) + 9(5)
= 213 (3)

C. Power Calculation

The total power for proposed systolic array multiplier circuit using gat logic is given by the following expression

PSYS = PMC + PPFA + PTG

The power of individual cells is given below.

PMC = 1.147 W
PPFA = 1.169 W
PTG = 500 nW
PSYS = 1.147 W + 1.169W +500 nW
= 2.816 W

VI. CONCLUSION

This paper provides unprecedented approach to reduce power consumption in Systolic Array Multiplier using logic gates. Proposed Systolic Array Multiplier circuit has 33 logic gates and overall power consumption is 14.7 W. Output is 66 and cost is 213. Simulations of the proposed designs are implemented with cadence virtuoso 90nm CMOS technology.

REFERENCES

[1] K. S. Ganesh Kumar1, J. Deva Prasannam2, M. Anitha Christy "Analysis of Low Power, Area and High Speed Multipliers For DSP Applications", International Journal of emerging technology Advanced Engineering, March 2014, Volume: 4, Issue: 3.
[2] H. T. Kung, "Why systolic architectures?" IEEE Computers Magazine, vol. 15, pp. 37–45, Jan. 1982.
[3] E. I. Milovanovic et. al, "Matrix Multiplication on Linear Bidirectional Systolic Arrays," Ser. A: Appl. Math. Inform. And Mech, vol. 2, no. 1, pp. 11–20, 2010.
[4] www.wikipedia.org

