International Journal of Electronics and Electical Engineering

Volume 1 | Issue 1 Article 14

July 2012

Bit-Level Systolic Architecture for a Matrix-Matrix Multiplier

M.N. Murty
Department of Physics, NIST, Berhampur-761008, Orissa, India, mnarayanamurty@rediffmail.com

S.S. Nayak
Department of Physics, JITM, Paralakhemundi, Orissa, India, sudhansusekharnayak@yahoo.com

Binayak Padhy
Department of Physics, Khallikote (Auto) College, Berhampur-760001, Orissa, India, binayak_padhy2@rediffmail.com

S.N. Panda
Department of Physics, Gunupur College, Gunupur, Orissa, India, snpanda@gmail.com

Bit-Level Systolic Architecture for a Matrix-Matrix Multiplier

M.N. Murty
Department of Physics
NIST, Berhampur-761008, Orissa, India
mnarayanamurty@rediffmail.com

S.S. Nayak
Department of Physics
JITM, Paralakhemundi, Orissa, India
sudhansusekharnayak@yahoo.com

B. Padhy
Department of Physics
Khallikote (Auto) College,
Berhampur-760001, Orissa, India
binayak_padhy2@rediffmail.com

S.N. Panda
Department of Physics,
Gunupur College, Gunupur,
Orissa, India

Abstract - Highly efficient arithmetic operations are necessary to achieve the desired performance in many real-time systems and digital image processing applications. In all these applications, one of the important arithmetic operations frequently performed is to multiply and accumulate with small computational time. In this paper, a 4-bit serial - parallel multiplier, which can perform both positive and negative multiplications, is presented. Baugh-Wooley algorithm necessitates complementation of last bit of each partial product except the last partial product in which all but the last bit are complemented. In the proposed algorithm all bits of the last partial product are complemented. This modification results in considerable reduction in hardware compared to Baugh-Wooley multiplier. This multiplier can be used for implementation of discrete orthogonal transforms, which are used in many applications, including image and signal processing. This paper presents a 2D bit-level systolic architecture for a matrix-matrix multiplier. A comparison with similar structures has shown that the proposed structure performs better.

Keywords - signal processing; serial-parallel multiplier; discrete orthogonal transforms; systolic architecture

1. INTRODUCTION

In digital signal processing (DSP) numerous arithmetic operations require the extensive use of multiplications. The multiplier is therefore an essential part of any DSP circuit. In the past much attention has been paid to the development of multiplication strategies, among others the algorithms of Booth [1], Chen [2], Wallace[3], Braun[4] and Baugh-Wooley[5] are most popular. A number of DSP applications demand high-speed multiplication while others require less hardware and moderate speed. Performance is greatly affected by the multiplication strategy and the type of multiplier used. A serial multiplier is employed to minimise the chip area and hence the hardware cost. On the other hand, a parallel multiplier presents a high speed but is expensive in terms of silicon area. A serial-parallel multiplier is a good trade-off between a time consuming serial multiplier and an area consuming parallel multiplier. In this device one factor is entered serially and the other is stored in parallel with a fixed number of bits. The multiplier is therefore designed for a given word length.

This paper proposes a serial-parallel multiplier designed to achieve precision with 4 bits. This multiplier is configured to accept and produce unsigned or signed two's complement numbers. The hardware used in this multiplier is less compared with Baugh-Wooley multiplier [5]. This multiplier can be used for implementation of discrete orthogonal transforms (DOTs).

DOTs are used in speech and image processing applications. The discrete Fourier transform (DFT) and the discrete Hartley transform (DHT) are popular orthogonal transforms. Of all the sinusoidal transforms, the discrete cosine transform (DCT) is most efficient transform for the compression of speech and image data. All these transforms are based on matrix-vector multiplication.

The systolic arrays are used in the design and implementation of high performance digital signal processing equipment. The systolic array concept can also be exploited at bit level in the design of individual VLSI chips[6]. Bit-level systolic arrays exhibit a number of attractive features such as :

i) The basic processing element (PE) is small.

ii) The computation time of the PE is small.

iii) The highly regular structure of the circuits renders them comparatively easy to design and test.

This paper presents a bit-level systolic architecture for a matrix-matrix multiplier based on the proposed algorithm for both positive and negative multiplications. This architecture exhibits the following features :

- Simplicity, regularity and modularity of the structure.
- Massive bit-level processing concurrency to achieve ultra-high throughput rate.
- Flexible implementation for any desired transform length.

The rest of the paper is organised as follows. The multiplication algorithm is presented in Section-2. The proposed serial-parallel multiplier is presented in Section-3. Section-4 deals with the systolic architecture for DOTs based matrix-matrix product. Finally we conclude our paper in Section-5.

2. Algorithm for Matrix - Matrix Multiplication

Consider two N x N matrices $\boldsymbol{A} = [A_{ij}]$ and $\boldsymbol{B} = [B_{ij}]$ where $\boldsymbol{A}$ is the kernel matrix of an orthogonal transform. The product $\boldsymbol{C} = [C_{ij}]$ of $\boldsymbol{A}$ and $\boldsymbol{B}$ is given by.

$$\boldsymbol{C = AB} \qquad (1)$$

Such that $C_{ij} = \sum_{k=0}^{N-1} A_{ik} B_{kj}$ (2)

If the elements of the two matrices are represented using 2's complement number system representation, then

$$A_{ik} = -a_{ik}^{n-1} 2^{n-1} + \sum_{\ell=0}^{n-2} a_{ik}^{\ell} 2^{\ell} \qquad (3)$$

$$B_{ik} = -b_{kj}^{n-1} 2^{n-1} + \sum_{m=0}^{n-2} b_{kj}^{m} 2^{m} \qquad (4)$$

Where a_{ik}^{ℓ} and b_{kj}^{m} are the ℓth bit of A_{ik} and mth bit of B_{kj} respectively (which are either 0 or 1). a_{ik}^{n-1} and b_{kj}^{n-1} are the sign bits, where n is the word length.

By substituting (3) and (4) in (2), the element of the matrix product C_{ij} can be computed as follows:

$$C_{ij} = \sum_{k=0}^{N-1}\left[-a_{ik}^{n-1}2^{n-1} + \sum_{\ell=0}^{n-2} a_{ik}^{\ell}2^{\ell}\right] \text{ x } \left[-b_{kj}^{n-1}2^{n-1} + \sum_{m=0}^{n-2} b_{kj}^{m}2^{m}\right]$$

$$= \sum_{k=0}^{N-1}\left[a_{ik}^{n-1}b_{kj}^{n-1}2^{2n-2} + 2^{n-1}\sum_{\ell=0}^{n-2}2^{\ell}\left(-a_{ik}^{\ell}b_{kj}^{n-1}\right) + 2^{n-1}\sum_{m=0}^{n-2}2^{m}\left(-a_{ik}^{n-1}b_{kj}^{m}\right) + \sum_{\ell=0}^{n-2}\sum_{m=0}^{n-2}2^{\ell+m}a_{ik}^{\ell}b_{kj}^{m}\right]$$

Using the relation $A + \bar{A} = 1$ from Boolean algebra,

$$C_{ij} = \sum_{k=0}^{N-1}\left[\left(-\overline{a_{ik}^{n-1}b_{kj}^{n-1}}+1\right)2^{2n-2} + 2^{n-1}\sum_{\ell=0}^{n-2}2^{\ell}\left(\overline{a_{ik}^{\ell}b_{kj}^{n-1}}-1\right) + 2^{n-1}\sum_{m=0}^{n-2}2^{m}\left(\overline{a_{ik}^{n-1}b_{kj}^{m}}-1\right) + \sum_{\ell=0}^{n-2}\sum_{m=0}^{n-2}2^{\ell+m}a_{ik}^{\ell}b_{kj}^{m}\right]$$

$$= \sum_{k=0}^{N-1}\left[\sum_{\ell=0}^{n-2}\sum_{m=0}^{n-2}2^{\ell+m}a_{ik}^{\ell}b_{kj}^{m} + 2^{n-1}\sum_{\ell=0}^{n-2}2^{\ell}\overline{a_{ik}^{\ell}b_{kj}^{n-1}} + 2^{n-1}\sum_{m=0}^{n-2}2^{m}\overline{a_{ik}^{n-1}b_{kj}^{m}} - 2^{2n-2}\overline{a_{ik}^{n-1}b_{kj}^{n-1}} + 2^{n} - 2^{2n-2}\right] \qquad (5)$$

Equation (5) can be mapped into a systolic architecture.

3. Proposed Serial - Parallel Multiplier

For n=4 and N=4, (5) can be written as

$$C_{ij} = \sum_{k=0}^{3}\left[2^{0}a_{ik}^{0}b_{kj}^{0} + 2^{1}\left(a_{ik}^{1}b_{kj}^{0} + a_{ik}^{0}b_{kj}^{1}\right) + 2^{2}\left(a_{ik}^{2}b_{kj}^{0} + a_{ik}^{1}b_{kj}^{1} + a_{ik}^{0}b_{kj}^{2}\right) + 2^{3}\left(\overline{a_{ik}^{3}b_{kj}^{0}} + a_{ik}^{2}b_{kj}^{1} + a_{ik}^{1}b_{kj}^{2} + \overline{a_{ik}^{0}b_{kj}^{3}}\right) + 2^{4}\left(\overline{a_{ik}^{3}b_{kj}^{1}} + a_{ik}^{2}b_{kj}^{2} + \overline{a_{ik}^{1}b_{kj}^{3}} + 1\right) + 2^{5}\left(\overline{a_{ik}^{3}b_{kj}^{2}} + \overline{a_{ik}^{2}b_{kj}^{3}}\right) - 2^{6}\left(\overline{a_{ik}^{3}b_{kj}^{3}} + 1\right)\right] \qquad (6)$$

The multiplication algorithm (6) resulting from (5) for a word length 4 is illustrated by the multiplication table shown in Table 1. The partial product terms are formed by ANDing each multiplicand bit with each multiplier bit. For a two's complement multiplication, one NAND and three AND operations are performed in each of the first

three rows of the bit product, and four NAND operations in the fourth row. The final product is computed by adding '1' to the fifth and the seventh columns along with all partial product terms.

Figure 1 shows the steps of multiplication procedures for four bit tow's complement numbers.

The 2's complement serial - parallel multiplier comprising of a logic unit and an adder unit is shown in Fig. 2. The logic unit consist of three AND gates, one NAND gate, and three XOR gates. In this multiplier the matrix element A_{ik} is stored in parallel and the matrix element B_{kj} is entered serially from the least-significant bit to the most-significant bit. A flip-flop (FF) saves the carry bit and a full adder adds the result of the partial product. The output of the multiplier is obtained from the adder unit using a carry-save and add-shift technique (the CSAS technique) together with one control signal Q. In the first three and last two clock cycles Q=0, but in the fourth, fifth and sixth clock cycles Q=1. The extra '1' for the fifth and seventh columns are automatically provided by the control signal Q. Four zeros are appended to the left of the MSB of B_{kj} for input / output synchronisation. The excessive carry-out bit out of (2n-1) bit product is ignored.

TABLE 1 : 4 X 4 BIT TWO'S COMPLEMENT MULTIPLICATION

			b_{kj}^3	b_{kj}^2	b_{kj}^1	b_{kj}^0
			a_{ik}^3	a_{ik}^2	a_{ik}^1	a_{ik}^0
			$\overline{a_{ik}^3 b_{kj}^0}$	$a_{ik}^2 b_{kj}^0$	$a_{ik}^1 b_{kj}^0$	$a_{ik}^0 b_{kj}^0$
		$\overline{a_{ik}^3 b_{kj}^1}$	$a_{ik}^2 b_{kj}^1$	$a_{ik}^1 b_{kj}^1$	$a_{ik}^0 b_{kj}^1$	
	$\overline{a_{ik}^3 b_{kj}^2}$	$a_{ik}^2 b_{kj}^2$	$a_{ik}^1 b_{kj}^2$	$a_{ik}^0 b_{kj}^2$		
$\overline{a_{ik}^3 b_{kj}^3}$	$\overline{a_{ik}^2 b_{kj}^3}$	$\overline{a_{ik}^1 b_{kj}^3}$	$\overline{a_{ik}^0 b_{kj}^3}$			
1		1				
P_6	P_5	P_4	P_3	P_2	P_1	P_0

			1	1	0	1	(-3)
			0	1	0	1	(5)
			1	1	0	1	
		1	0	0	0		
	1	1	0	1			
1	0	1	0				
1		1					
1	1	1	0	0	0	1	(-15)

			1	0	1	1	(-5)
			1	1	0	0	(-4)
			0	1	0	0	
		0	1	0	0		
	1	0	0	0			
0	0	1	1				
1		1					
0	0	1	0	1	0	0	(20)

Figure 1 : Steps in computation of -3 x 5 and -5 x -4

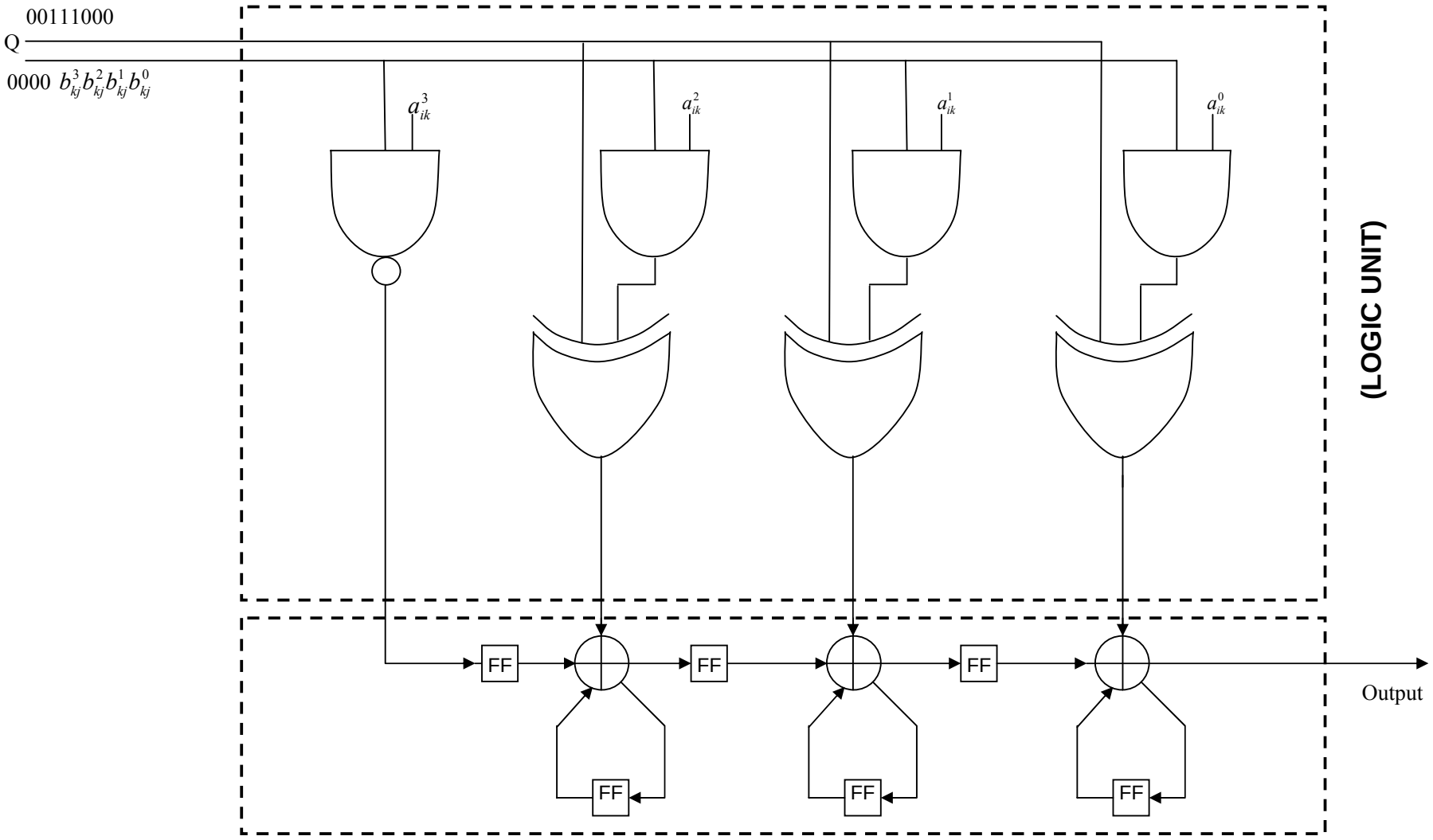

Figure 2. Serial - Parallel 4-bit multiplier

4. Proposed Systolic Architecture for Matrix-Matrix Product

From (1)and (2), we have for N=3

$$\begin{bmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \\ C_{20} & C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} A_{00} & A_{01} & A_{02} \\ A_{10} & A_{11} & A_{12} \\ A_{20} & A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{00} & B_{01} & B_{02} \\ B_{10} & B_{11} & B_{12} \\ B_{20} & B_{21} & B_{22} \end{bmatrix} \quad (7)$$

Equation (5) can be mapped into the proposed systolic architecture as shown in Fig. 3 for the case of N=3 and the matrix - matrix product (7) can be implemented by this structure. It consists of three identical blocks. Each block has three identical processing elements (PEs). Function of each PE is shown in Fig. 4. Each PE comprises a serial - parallel multiplier M as shown in Fig. 2, a flip-flop (FF) for saving the carry bit and a full adder that adds the result of the partial product and the result generated from the previous PE.

The matrix elements B_{kj} are fed in a parallel / serial fashion bit by bit with LSB first through 3-stage parallel - serial S shift registers of each block, while the matrix elements A_{ik} are fed in parallel fashion and remain fixed in their corresponding PE cell during the entire computation of the operation. The data in the S shift registers are bit-serially shifted out with least significant bit first, staggered by one addition delay 'T_A' with respect to the preceding location of the shift register. Q signal is provided to each PE for regulating the multiplications. Each bit of the final product of the PE is fed to the full adder of the preceding PE so that the corresponding output of each PE are added to compute the desired output.

The proposed systolic architecture consists of N number of S shift registers and N^2 number of PEs. The area complexity of the structure for N-point transform is $O(N^2)$ as the area of the shift registers can be treated small compared with the area of the PEs.

During the first seven clock cycles three inner products $[C_{io}]$ (i = 0,1,2) are computed. Then during the second seven cycles the three inner products $[C_{i1}]$ are computed. Finally, the three inner products $[C_{i2}]$ are computed during the third seven cycles. This architecture produces three elements of the matrix $\boldsymbol{C}$ in every seven clock cycles based on the multiple accumulate technique and therefore the entire computation can be carried out in (2n-1)N clock cycles with a structure requiring N^2 PEs. The setup time to load the matrix elements B_{kj} into the parallel-serial shift registers is ignored.

Table 2 illustrates the performance comparison with similar structures in terms of computation time and area complexity.

TABLE 2 : COMPARISON OF PROPOSED MATRIX - MATRIX PRODUCT STRUCTURE WITH SOME EXISTING STRUCTURES

Features	Proposed Structure	Structure of [7]a	Structure of [7]b	Structure of [8]
Computation time	$[(2n\text{-}1)N]\,T$	$(N^2 \log_2 n)T$	$(N \log_2 n)T$	$(2nN)T$
Area complexity	$O(N^2)$	$O(nN \log_2 n)$	$O(N^2 n \log_2 n)$	$O(N^2)$

It can be seen that the proposed structure requires less area when compared with [7]a and [7]b. The computation time of the proposed structure is less than the structure of [8].

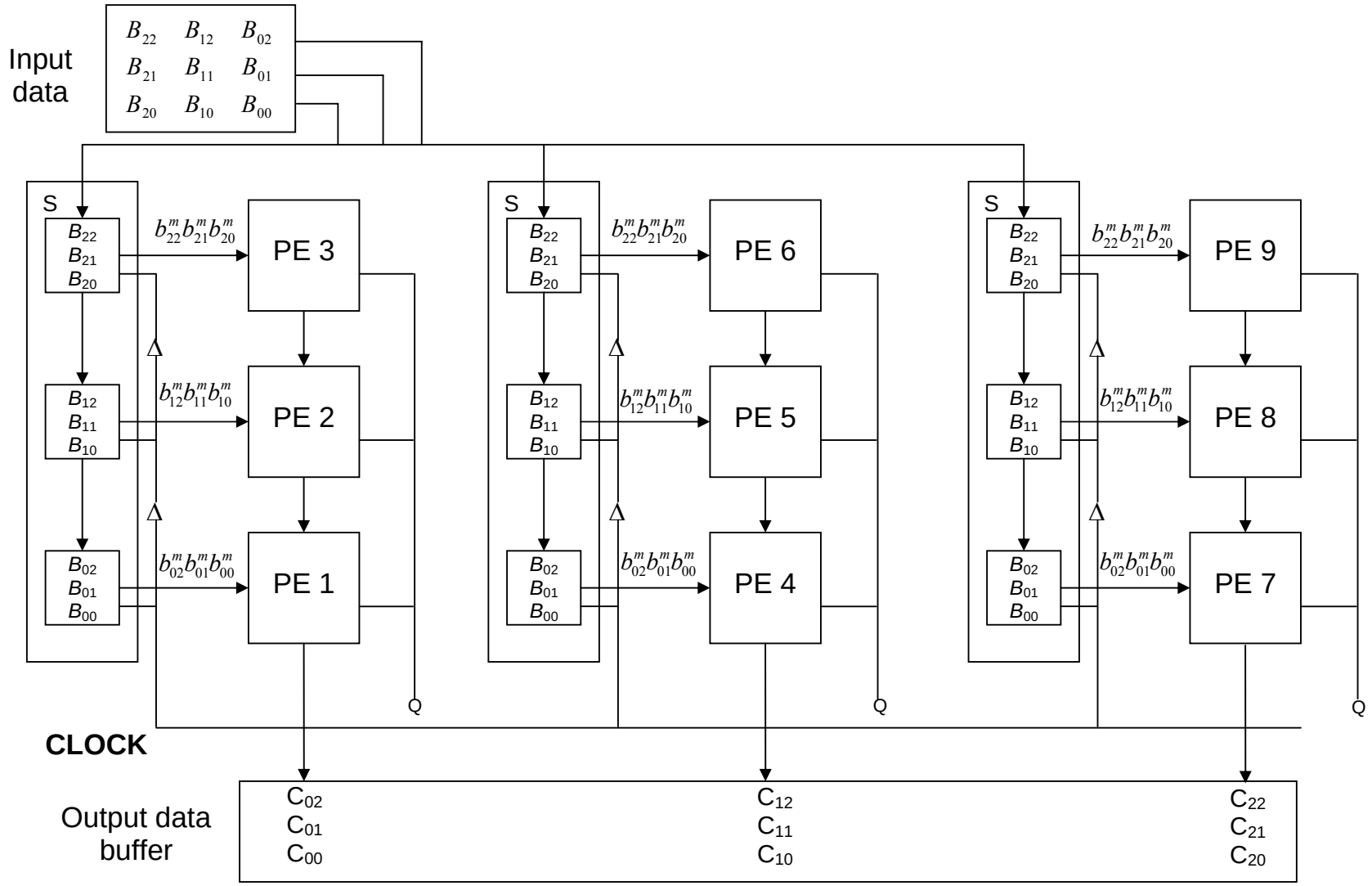

Figure 3. Systolic architecture for implementation of DOT based matrix-matrix product for N=3. Δ implies delay of duration T_A.

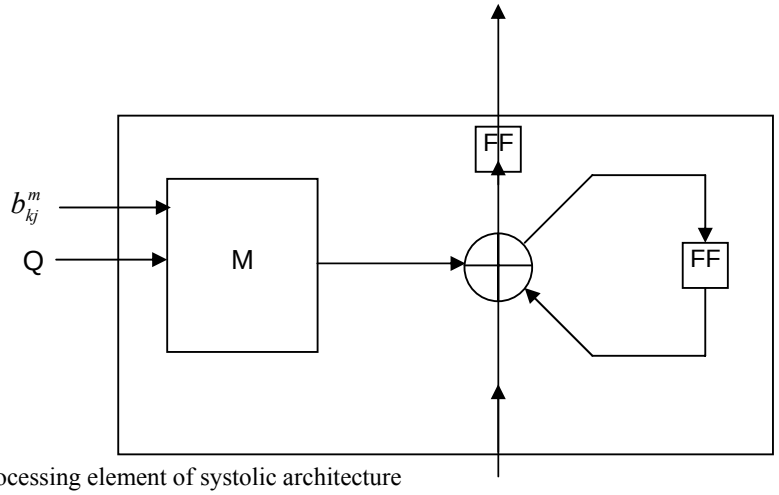

Figure 4. Processing element of systolic architecture

5. CONCLUSION

A serial - parallel multiplier, which can perform either signed or unsigned multiplications, is proposed. The hardware used in this multiplier is less compared with Baugh - Wooley multiplier. Also we have presented a bit - level systolic architecture for the computation of DOTs based matrix-matrix product. It has been found that, this structure requires less area and time complexity when compared with some existing structures. Due to simplicity, regularity and modularity, the proposed systolic architecture is suitable for VLSI signal processing applications.

REFERENCES

[1] A.D. Booth, "A signed binary multiplication technique", Q.J. Mech. Appl. Math., vol. 4, Part-2, pp. 236-240, May 1951.

[2] T.C. Chen, "A binary multiplication scheme based on squaring", IEEE Tans., vol. C-20, pp. 678-680, 1971.

[3] C.S. Wallace, "A suggestion for a fast multiplier", IEEE Trans., vol. EC-13, pp. 14-17, February 1964.

[4] E.L. Braun, "Digital computer design", Academic Press, New York, 1963.

[5] C.R. Baugh and B.A. Wooley, "A two's complement parallel array multiplication algorithm", IEEE Trans., vol. C-22, no. 1-2, pp. 1045-1047, December 1973.

[6] J. McCanny, J. McWhirter, and S. Kung, "The use of data dependence graphs in the design of bit-level systolic arrays", IEEE Trans. Acoust., Speech., Signal Process., vol. 18, pp. 787-793, 1990.

[7] R.S. Grover, W. Shang, and Q.Li, "A comparison of FPGA implementation of two's complement bit-level and word-level matrix multiplier", Proceedings of the 9th ACM/SIGDA International Symposium on Field Programmable Gate Arrays, (FPGA'01), Monterey, California, USA, p. 223, 11-13 February 2001.

[8] A. Amira, A. Bouridane, P. Milligan, and A. Belatreche, "Design of efficient architectures for discrete orthogonal transforms using bit level systolic structures", IEE Proc. - Comput. Digit. Tech., vol. 149, no. 1, pp. 17-24, January 2002.

[9] S.S. Nayak and P.K. Meher, "High throughput VLSI implementation of discrete orthogonal transforms using bit-level vector - matrix multiplier", IEEE Trans. Circuits Syst. II, Analog and Digital Signal Processing, vol. 46, no. 5, pp. 655-658, May 1999.

[10] S.S. Nayak, M.N.Murty, and H. Panda, "Bit - level systolic implementation of discrete orthogonal transforms", Signal Processing, vol. 81, pp. 2437-2443, 2001.