SYSTOLIC ARITHMETIC ARCHITECTURES

Khaled M. Elleithy

Computer Engineering Department
King Fahd University of Petroleum and Minerals
Box 1967, Dhahran 31261, Saudi Arabia

Abstract: In this paper parallelism on the algorithmic, architectural, and arithmetic levels is exploited in the design of a Residue Number System (RNS) based architecture. The architecture is based on modulo processors. Each modulo processor is implemented by two dimensional systolic array composed of very simple cells. The decoding stage is implemented using a 2-D array, too. The decoding bottleneck is eliminated. The whole architecture is pipelined which lead to high throughput rate.

1. Introduction: In this paper, Residue Number System (RNS) has been used to achieve arithmetic parallelism. Figure 1 shows a general RNS based architecture. The algebraic properties of RNS provide both high speed computation and parallel operations. In our implementation, parallelism is achieved in three directions; each modulo processor is implemented by two dimensional systolic array composed of very simple cells. The decoding stage is implemented using a 2-D array, too. The decoding bottleneck will be eliminated. The whole architecture will be pipelined which lead to high throughput rate.

2. Modulo Adder: The modulo adder performs addition in time complexity O(1). Figure 2 shows the modulo adder architecture. It is composed of fine 1-D arrays independent of the size of the moduli.

3. Modulo Multiplier: The modulo multiplier is using this modulo adder as a computational kernel. The multiplier consists of two stages as shown in Figure 3. In the first stage is an SIMD array of AND gates used to obtain the partial products. The second stage of the adder is a binary tree of modulo adders used to perform the addition of the n partial products.

4. Chinese Remainder Theorem: The proposed decoding stage is based on the Chinese Remainder Theorem(CRT)[1]. Although the CRT provides a direct, fast, and simple conversion formula, the lack of large and fast modulo M adder has held back this approach. Systolizing the CRT will overcome this problem. The proposed systolic CRT architecture can produce both sign-magnitude and 2's complement data representation. The overall structure is shown in Figure 4. The Inputs to the residue decoder are the residues and a control line, C, which determines the output to be in sign magnitude or 2's complement representation. The partial sum generator computes the partial sums. The partial sum addition is computed using the developed modulo adder. The range determinator enforce the correctness of data. The input data to this stage is decomposed into groups of bits to be processed in parallel. The overall time complexity

0-8186-2465-5/91 $1.00 © 1991 IEEE

of the CRT systolic architecture is O(log n), where n is the number of moduli.

Conclusions: An SIMD structure has been presented which exhibits two level parallelism; architectural (systolic array) and arithmetic (RNS). Three optimal algorithms are introduced in this paper; O(1) modulo addition algorithm, O(log n) modulo multiplication and O(log n) CRT decoder.

References

[1] N. S. Szabo and R. I. Tanaka, Residue Arithmetic and its Applications to Computer Technology, McGraw-Hill, 1967.

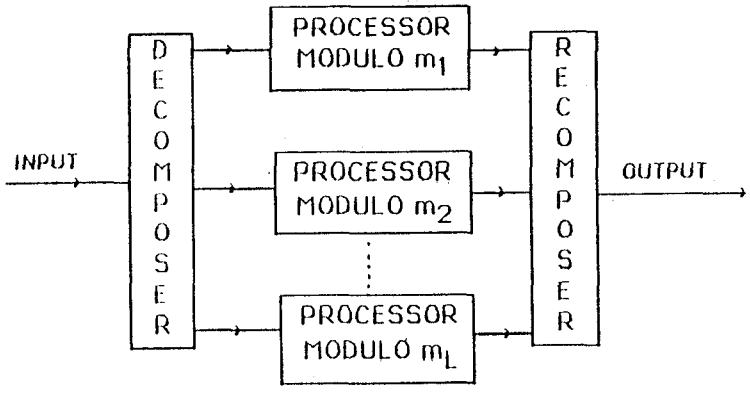


Fig. 1. General RNS Based Architectures.

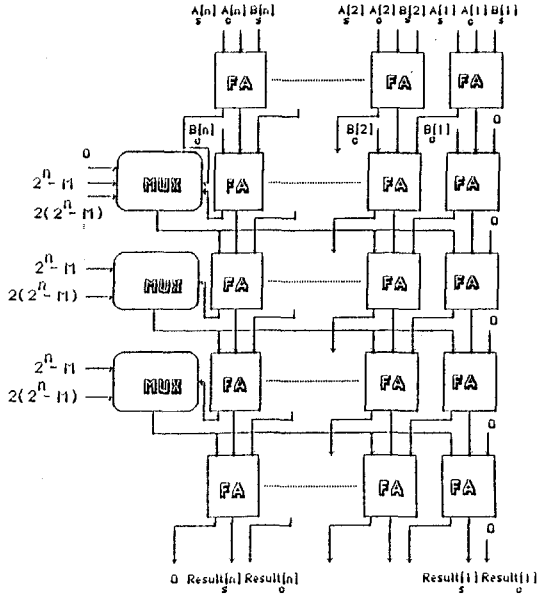
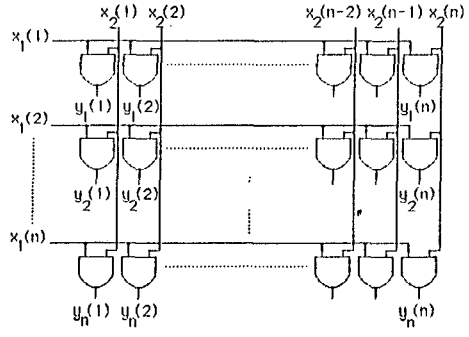
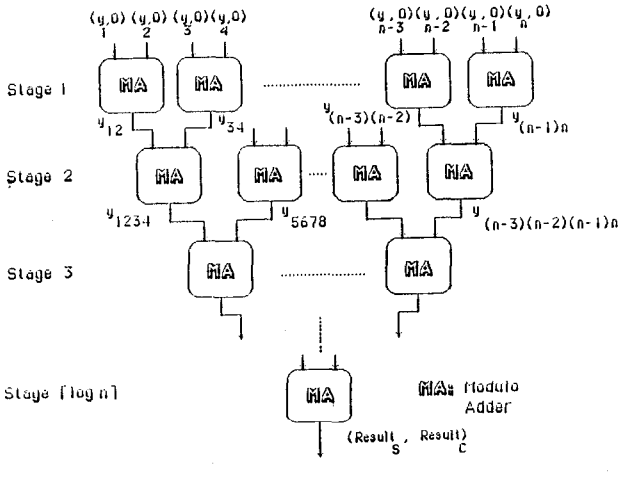


Fig. 2. The Modulo Adder.



3.(a) Partial Product Generator.



3.(b) Addition of Partial Products using Modulo Adders.

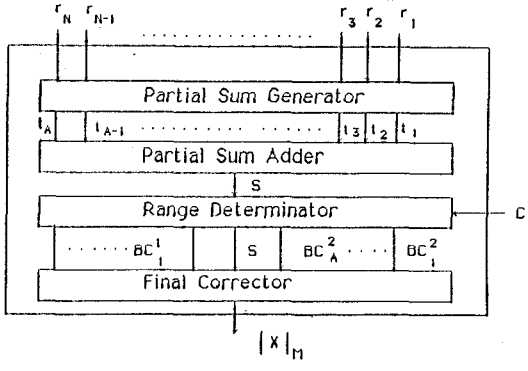


Fig. 4. The Residue Decoder.