1998 International Computer Symposium
Workshop on Computer Architecture
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

A Study for Carry Propagation Delay of Digital Ratioed Compressors

Chua-Chin Wang, Chenn-Jung Huang, and Po-Ming Lee

Department of Electrical Engineering,
National Sun Yat-Sen University, Kaohsiung, Taiwan R.O.C
E-Mail: ccwang@ee.nsysu.edu.tw

Abstract

Inner product calculations are often required in digital neural computing. The critical path of the inner product of two binary vectors is the carry propagation delay generated from individual product terms. In this work, a hierarchical structure of ratioed compressor building blocks is proposed, and the carry propagation delay estimation of these compressor building blocks is derived. A general delay formula is obtained, which leads a minimal delay of calculation. The theoretical derivation and Verilog simulation both indicate that the 3-2 compressor might be an ideal candidate for the basic building blocks used in digital hardware realization of the inner product computation.

1. Introduction

Many efforts have been thrown on the realization of neural networks mainly owing to their attractive pattern recognition features, [1], [2]. In the computation of neural networks, the inner product of two vectors might be one of the most frequently used mathematical operations. Unavoidably the carry propagation will occur if the neural networks are dedicated for either discrete or digital signals. For instance, the recall of pattern pairs stored in discrete bidirectional associative memory (BAM) needs to compute a summation in the form as $Y = th\left(\sum_{i=1}^{n} Y_i \cdot (X_i \cdot X)\right)$, where X is the input pattern, Y is the output pattern, X_i's and Y_i's are stored pattern pairs, and *th()* is a threshold function. Notably, the components of every vector are either bipolar or binary. If n is large in the above calculation, then the carry propagation of the inner product of the vectors will likely become the critical delay of the entire neural computing.

Since neural computing is composed of mass amount of inner product calculations, the demand of shortening the delay therewith becomes urgent. Many high-speed logic design styles have been announced. However, these logics suffer from different difficulties. For example, domino logic [3] can not be non-inverting; NORA [4] has the charge sharing problem; all-N-logic [5] and robust single phase clocking [6] cannot operate correctly under clocks with short rise time or fall time, which can not be easily integrated with other part of logic design; single-phase logic [7] and Zipper CMOS [8] contain slow P-logic blocks. Though Zhang *et al.* [9] proposed a design of compressor to resolve such a problem by employing a so-call C^2PL (complex CPL), several physical design factors are not fully considered or implemented. First, the sizes of the NMOS transistors are impossible to be minimized. Second, the driving inverters' sizes have to be properly tuned. Third, the original design of [9] not only gives a poor fan-in and fan-out capability, but also produces very asymmetrical rise delay and fall delay which will very much likely cause glitch hazards and unwanted power consumption. Fourth, no investigation on topics related to compressor building blocks with minimal carry propagation delay is made. In this paper, a robust hierarchical architecture of the building blocks based on the 3-2 compressors is proposed; an analytical form of

carry propagation delay estimation for this architecture is derived; and the HSPICE and Verilog simulation results are also presented to verify our theoretical analysis.

2. Framework of Ratioed Compressor Building Blocks

2.1 Basic compressor building block design

A *3-2* compressor is basically a full adder. The equations of a full adder are shown as follows:

$$S = (a \oplus c)\, b' + (a \oplus c)'\, b = F\, b' + F'\, b$$

$$C = (a \oplus c)\, b + (a \oplus c)'\, c = F\, b + F'\, c \qquad (1)$$

where F denotes $(a \oplus c)$. Then, a typical *3-2* compressor is shown in Fig. 1. The feature of such a compressor is that the output represents the number of 1's given in inputs.

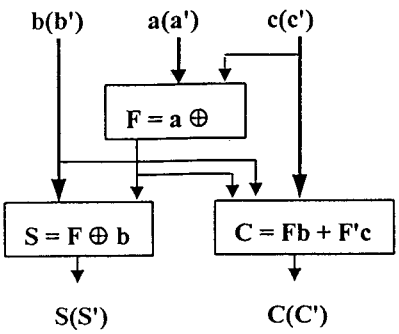

Fig. 1. 3-2 compressor building block

A *7-3* compressor building block is constructed by cascading four *3-2* compressors as shown in Fig. 2. A *15-4* compressor building block can also be formed similarly with two *7-3* compressors and two *3-2* compressors, as shown in Fig. 3. Basing on this design methodology, a general form for a $(2^n\text{-}1)\text{-}n$ compressor building block is composed of two $(2^{n-1}\text{-}1)\text{-}(n\text{-}1)$ compressors and n - *1 3-2* compressors, as shown in Fig. 4. Note that the counts of logic layers can be reduced from two to one when the serialized single bit input passes through every *3-2* compressor shown on the bottom of the figures.

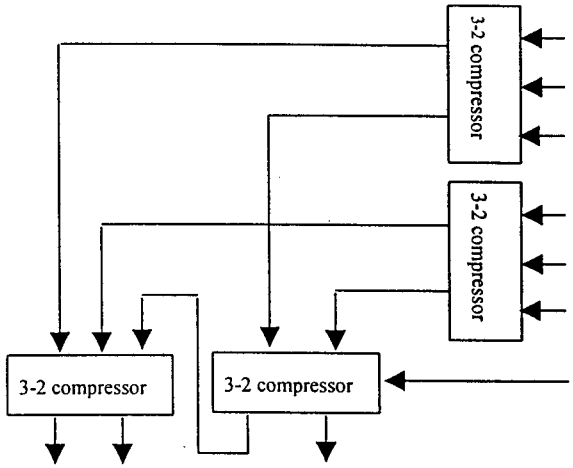

Fig. 2. 7-3 compressor building block

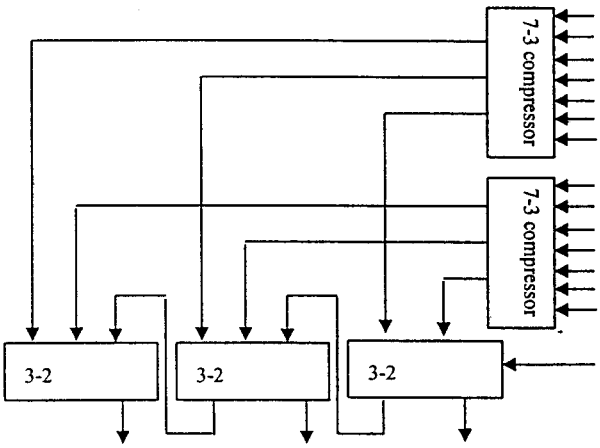

Fig. 3. 15-4 compressor building

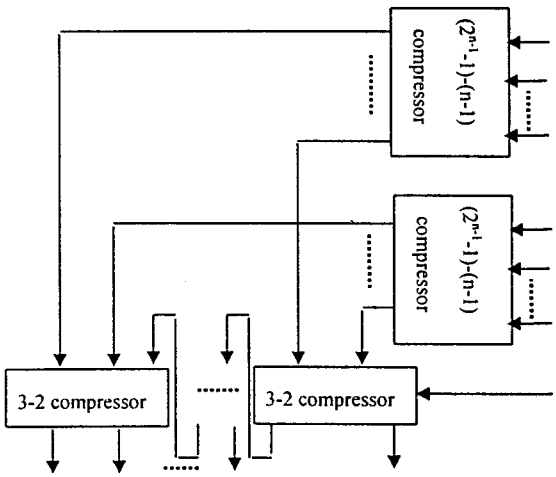

Fig. 4. $(2^n\text{-}1)$-n compressor building block

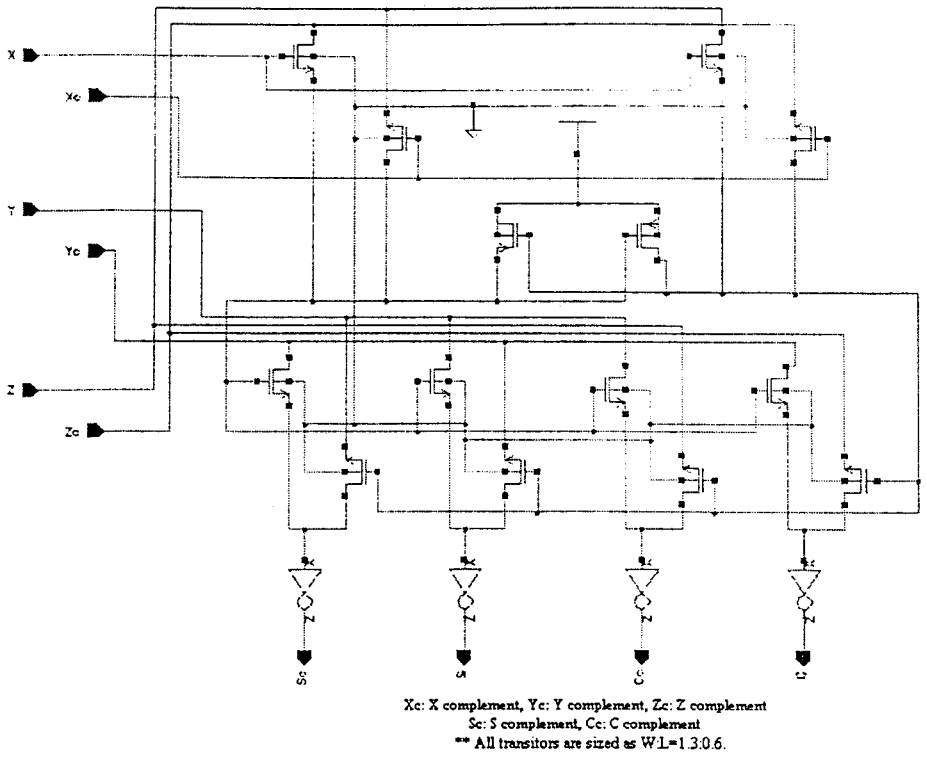

Fig. 5. Circuit of $C^2PL(1)$ 3-2 compressor in original design

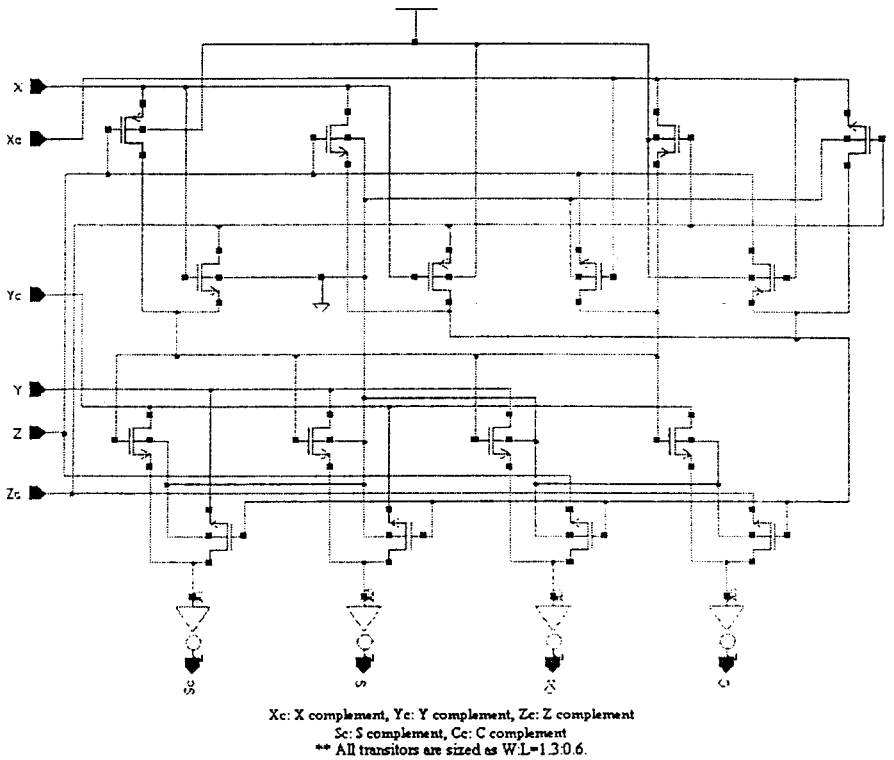

Fig. 6. Circuit of $C^2PL(2)$ 3-2 compressor in original design

2.2 Ratioed 3-2 compressor design

Though a *3-2* compressor can be realized by a full adder, and Zhang *et al.* [9] proposed a C²PL design for *3-2* and *7-3* compressors, several design issues as addressed in Section 1 are still ignored in their work. Fig. 5, and 6 show the circuit design of two types of *3-2* compressors based on complex complementary pass-transistor logic (C²PL) proposed in [9]. We use TSMC 0.6 μm 1P3M technology to re-design the *3-2* compressors, and the corresponding circuit design of the ratioed *3-2* compressors are shown in Fig. 7, and 8. In Section 3 of this study, we will demonstrate that the re-designed *3-2* compressor circuits will overcome all of the problems mentioned in Section 1.

2.3 Carry propagation delay equations

Since the total delay of such design is approximately proportional to the count of *3-2* compressors that the critical path resides, we assume D_n denotes the count of *3-2* compressors when *2ⁿ-1* bits are applied on the *(2ⁿ-1)-n* compressor block. By observing the structure of the compressor blocks, we can deduce D_2, D_3, , D_4, and D_n as follows:

$$D_2 = 1$$

$$D_3 = 3 = 2 + 1 = 2 + D_2$$

$$D_4 = 6 = 1 + 1 + 1 + 3 = 3 + D_3$$

$$D_n = n - 1 + D_{n-1}, \qquad n \geq 3. \tag{2}$$

By solving the above recurrence relation, we obtain

$$D_n = \frac{n(n-1)}{2} \tag{3}$$

Apart from the delay for the single building block, we have to count in the processing stages needed for summing individual inner product terms. The numbers of processing stages is roughly estimated as $\frac{\ln \frac{n}{M}}{\ln \frac{n}{2^n - 1}}$, where n denotes the total bits of the basic building block output, and M represents the bit count of data inputs.

Therefore, the count of *3-2* compressors when M bits are applied on the *(2ⁿ-1)-n* compressor building blocks can be shown as follows:

$$D_{M,n} = \frac{\ln \frac{n}{M}}{\ln \frac{n}{2^n - 1}} \cdot \frac{n(n-1)}{2} \tag{4}$$

2.4 Performance analysis

The comparison of delay for different size of building blocks can be made by applying the same amount of inner product calculations to each combination of compressor blocks. Basing on Eqn. (4), a 3-D mesh of delay computation for n from *2* to *10* and M from *500* to *10000* can be built up as shown in Fig. 9. As we can see, the *3-2* compressor is apparently the one possessing the minimal delay. The reason why other building blocks cannot introduce a shorter delay than that of the basic *3-2* compressor is that the construction of the serialized *3-2* compressor connections in the bottom part of *(2ⁿ-1)-n* compressor blocks generates the sluggish carry propagation along the critical path of summation.

3. Simulation and Analysis

In order to verify the correctness of our theoretical analysis, we tend to use HSPICE and Verilog to conduct a series of simulations. The improvement of asymmetrical rise delay and fall delay in the original design can be illustrated through HSPICE simulations. The simulation results are tabulated as Table 1 shows.

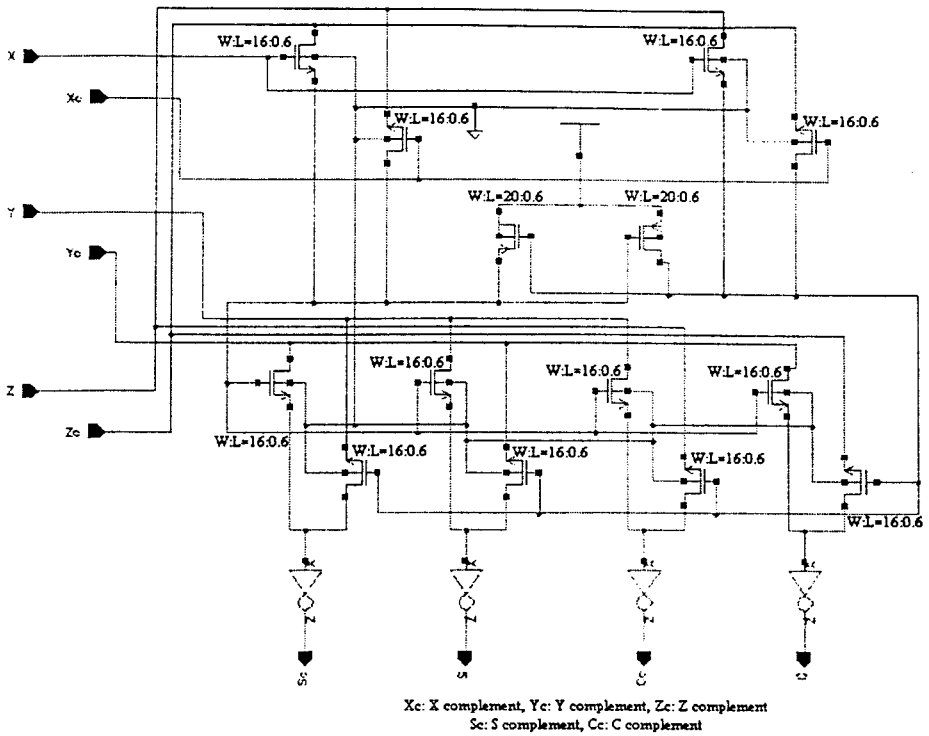

Fig. 7. Circuit of re-designed C²PL(1) 3-2 compressor

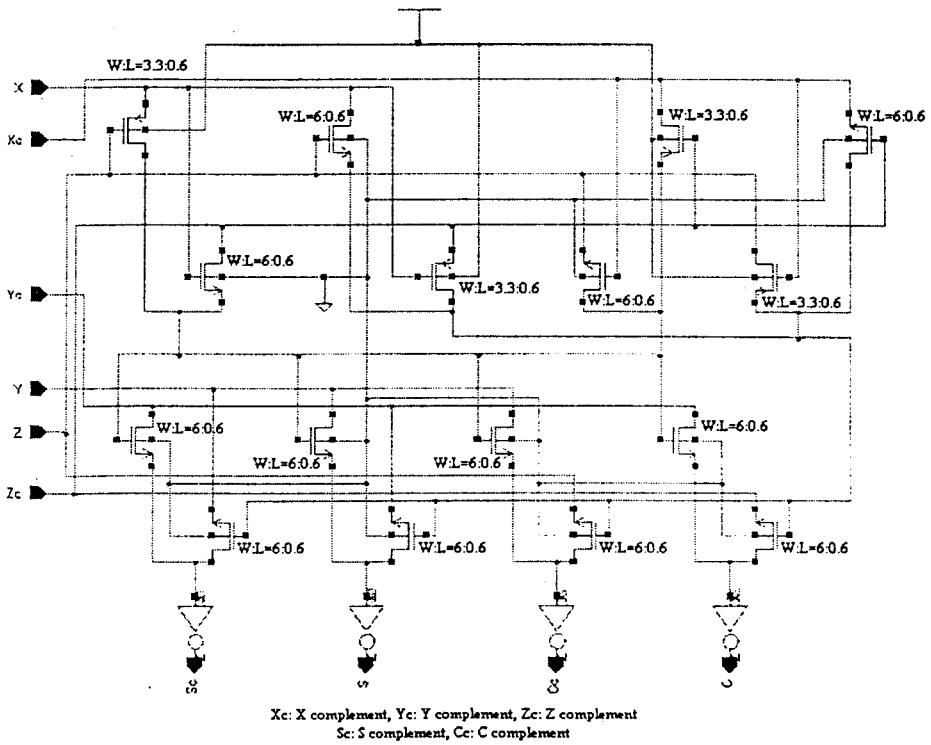

Fig. 8. Circuit of re-designed C²PL(2) 3-2 compressor

circuits / delay	The original 3-2 compressor				The re-designed 3-2 compressor			
	$C^2PL(1)$		$C^2PL(2)$		$C^2PL(1)$		$C^2PL(2)$	
	carry	sum	carry	sum	carry	sum	carry	sum
rise delay	0.26ns	0.31ns	0.42ns	0.35ns	0.32ns	0.36ns	0.41ns	0.34ns
Fall delay	0.87ns	0.83ns	0.87ns	0.87ns	0.24ns	0.43ns	0.39ns	0.42ns

Table 1 The comparison of rise delay and fall delay in the original design and the re-designed 3-2 compressor

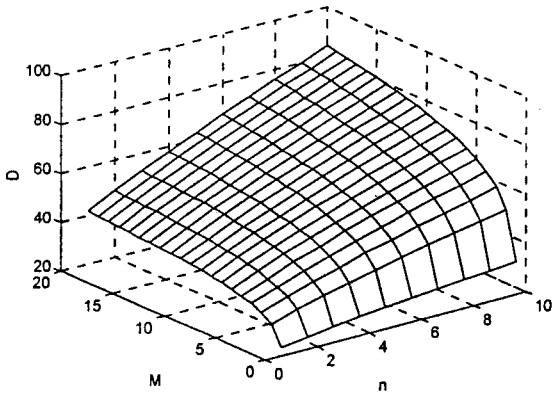

Fig. 9. Delay of M data inputs using (2^n-1)-n compressors

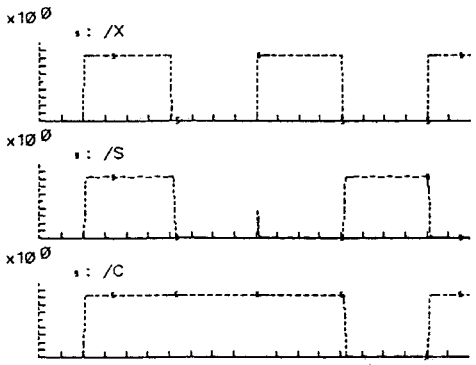

Fig. 10. Waveform diagram of $C^2PL(1)$ 3-2 compressor in original design

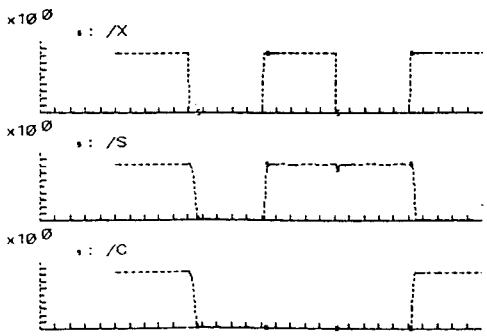

Fig. 11. Waveform diagram of $C^2PL(2)$ 3-2 compressor in original design

Fig. 10, and 11 show the waveform diagrams of the original work in [9], while Fig. 12, and 13 demonstrate the correctness of the undesirable property mentioned above in our design. The Verilog simulations are performed *20000* iterations for *3-2*, *7-3*, *15-4*, and *31-5* compressor building blocks, respectively. Fig. 14 illustrates the total delay time in nanosecond for different size of compressor building blocks used in *128* data inputs summation. The results demonstrate that both C^2PL *3-2* compressors indeed lead the minimal carry propagation delay.

4. Conclusion

In this paper we have proposed a novel organization of basic compressor building blocks which can be adopted in the implementation of digital neural networks. The re-designed *3-2* compressor is presented to correct several problems appearing in Zhang's work in [9]. The equation for counting the number of *3-2* compressors is derived and used for exploration of the basic building block with minimal carry propagation delay. Our simulation results show that *3-2* compressor building block is the winner due to the influential overhead caused by the serialized connection of 3-2 compressors portion in other kind of ratioed compressor blocks. However, this study still gives us a clue for further improvement on the construction of basic compressor building blocks design for inner product calculations in the future work.

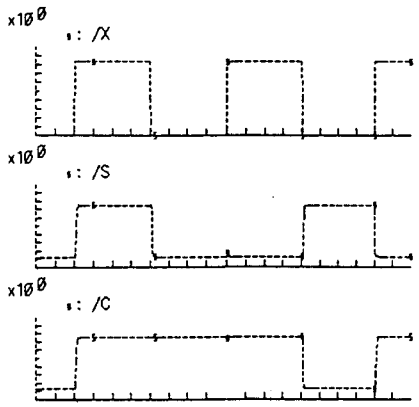

Fig. 12. Waveform diagram of re-designed $C^2PL(1)$ 3-2 compressor

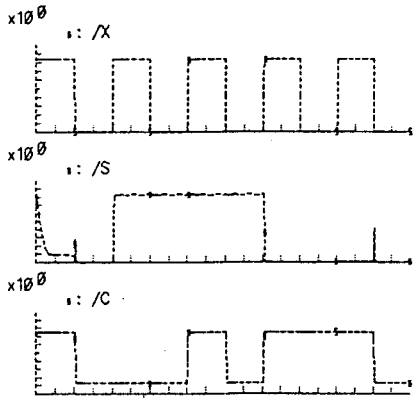

Fig. 13. Waveform diagram of re-designed $C^2PL(2)$ 3-2 compressor

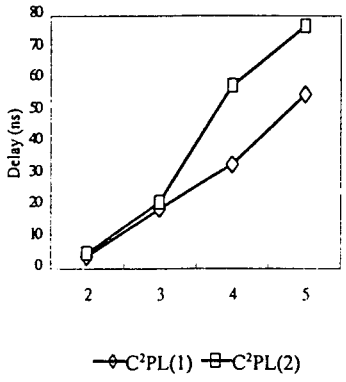

Fig. 14. Delay of 128 data inputs using C^2PL (2^n-1)-n compressors

5. References

[1] B. Kosko, "Bidirectional associative memory," IEEE Trans. System Man Cybernet, vol. 18, no. 1, pp.49-60, Jan./Feb. 1988.

[2] C.-C. Wang, and H.-S. Don, "An analysis of high-capacity discrete exponential BAM," IEEE Trans. on Neural Networks, vol. 6, no. 2, pp. 492-496, Mar. 1995.

[3] R. H. Krambeck, C. M. Lee, and H.-S. Law, "High-speed compact circuits with CMOS," IEEE J. Solid-State Circuits, vol. 17, pp. 614-619, June 1982.

[4] N. F. Goncalves, and H. J. De Man, "NORA: A race-free dynamic CMOS technology for pipelined logic structures," IEEE J. on Solid-State Circuits, vol. 18, pp. 261-266, June 1983.

[5] R. X. Gu, and M. I. Elmasry, "All-N-logic high-speed true-single-phase dynamic CMOS logic," IEEE J. on Solid-State Circuits, vol. 31, no. 2, pp. 221-229, Feb. 1996.

[6] M. Afghahi, "A robust single phase clocking for low power high-speed VLSI application," IEEE J. of Solid-State Circuits, vol. 31, no. 2, pp. 247-253, Feb. 1996.

[7] J. Yuan, and C. Svensson, "High-speed CMOS circuit technique," IEEE J. on Solid-State Circuits, vol. 24, pp. 62-70, Feb. 1989.

[8] C. M. Lee, and E. W. Szeto, "Zipper CMOS," IEEE Circuits Devices Mag., pp. 10-16, May 1986.

[9] D. Zhang, and M. I. Elmasry, "VLSI compressor design with applications to digital neural networks," IEEE Trans. on VLSI Systems, vol. 5, no. 2, pp. 230-233, June 1997.