An Area Efficient Approach on PACC RLE Encoder

Ancy Mathew[1], Rafeekha M J[2]

[1]P G Scholar, VLSI and Embedded Systems, Department of ECE, T K M Institute of Technology, Kollam, India

[2]Assistant Professor, Department of ECE, T K M Institute of Technology, Kollam, India

Abstract: ***Non volatile processor can retain its state when the power is off. Design challenge of non volatile processor is the excess area consumed by non volatile registers in it. Compared to regular CMOS flip flop, a ferro electric non volatile flip flop takes a large area due to its hybrid structure. Later Non volatile processor based on floating gate transistor was designed. But it also suffer from 40 percentage memory area overhead. Parallel Compare and Compress architecture is used to reduce the area of non volatile registers. The main block in PaCC architecture are PaCC encoder, PaCC decoder, Volatile registers, non volatile registers and non volatile flip flop controller. PaCC encoder is the work that is mainly focused. PaCC architecture uses Run length encoder(RLE), that compresses the data efficiently. Run length encoder is a simple form of data compression in which runs of data are stored as a single data value and count. This framework can be extended by replacing RLE encoder by Borrow wheeler transform and a comparative study on the area of both these encoders have to be done. The architecture was modeled using VHDL Coding and simulated using Modelsim.***

Keywords: Run length encoder, parallel compare and compress encoder, nonvolatile flipflops, Xilinx ISE13.1, modelsim 6.3f

1. Introduction

With the emerging memory technologies, nonvolatile processors are receiving more and more attention. Compared with the volatile ones, the nonvolatile processors are manufactured with nonvolatile registers and have the following advantages: I) zero-standby power: the processor can retain its state when not powered, while the traditional ones suffer from the increasing leakage power to keep data; II) instant on and off: the processor can resume its work within several cycles from the stalled point, while the traditional one needs millions of initializing cycles; III) high resilience to power failures: the processor can work reliably under the environments with frequency power interrupts, such as energy harvesting and wireless powered applications; IV) fine-grained power management: the processor can be shut down whenever possible due to the ultra-low energy and fast recovering characteristics.

Ferroelectric random access memory (FeRAM) is a low-cost nonvolatile memory technology fabricated with two additional masks to a standard CMOS technology , which has been widely used in low-cost and low power applications. These flip-flops are used in a distributed fashion and are able to maintain system states without any power supply indefinitely. An efficient controller is employed to achieve parallel reads and writes to the flip-flops. A reconfigurable voltage detection system is designed for the automatic system backup during power failures.The data transfer between a volatile processor and secondary NV storage cause redundancy over time .The PaCC architecture to reduce the number of the bits to be stored in the NVFFs, and hence the number and area of NV registers. The architecture adopts a compare and compress strategy to improve the compression ratio. A compression codec based on a parallel run length encoding is used . An area efficient two-stage shifting network is designed to minimize the codecs area, which reduces the area of the original barrel shifting network. A configurable state table structure in PaCC to store the reference vectors used for comparison. For specific application to formulate the reference vector selection problem as an optimization problem and develop heuristic algorithms to solve it.

2. Theory

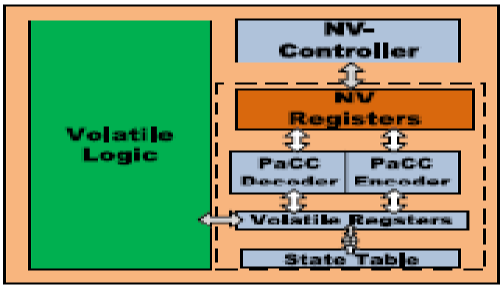
Compression is a way to reduce the number of bits in a frame but retaining its meaning .It decreases space, time to transmit, and cost. It is a technique to identify redundancy and to eliminate it. Data compression is the art of reducing the number of bits needed to store or transmit data. When data compression is used in a data transmission application, speed is the primary goal. Speed of transmission depends upon the number of bits sent, the time required for the encoder to generate the coded message and the time required for the decoder to recover the original ensemble. In a data storage application, the degree of compression is the primary concern .Various lossless data compression algorithms have been proposed and used. Some of the main techniques in use are the Huffman Coding, Run Length Encoding, Arithmetic Encoding and Dictionary Based Encoding.

Lossless data compression is nothing but the orginal data can be reconstructed exactly from compressed data. Lossless compression is generally used for applications that cannot tolerate any difference between the original and reconstructed data. Text compression is an important area for lossless compression. In lossless methods, original data and the data after compression and decompression are exactly the same. Redundant data is removed in compression and added during decompression.

Lossy data compression in which data after compression and decompression are exactly the same. Lossy data compression in which data after compression and then decompression retrieves a file that is not exactly as the orginal data as there will be loss of data. These methods are cheaper, less time and space.

3. Methodology

3.1 Pacc Architecture



The PaCC architecture consists of volatile registers which stores the current system state V, a state table, a compression codec (divided into a PaCC encoder and a PaCC decoder), and a small set of NV registers. The state table is used to store $_{\text{Vref}}$; the compression codec is used to make conversion between V_{diff} and V^{c}_{diff} ; and the comparison is done by the bitwise XORs. Though the volatile registers, the state table, and compression codec may increase the area, the significant reduction in the number of NV registers leads to a much smaller overall chip area. Pacc encoder in the Pacc architecture is the work that is mainly focused.

3.2 Block Diagram

The basic block diagram for the Parallel compare and compress encoder is as shown in figure 2. Block diagram consist of input end shifting network, output end shifting network, Run Length Encoder, All 0/1 detector block and length control. Two stage shifting network consist of Fixed N bit shifting network and 2N bit barrel shifter. Fixed N bit shifting network is used to simply shift the input by N .This shifter is extremely area efficient and no multiplexor is needed. The fixed shifting network is used to update the remaining part of the volatile registers. A Barrel shifter is a digital circuit that can shift a data word by a specified number of bits in one clock cycle .It can be implemented as a sequence of multiplexers and in such an implementation the output of one mux is connected to the input of the next multiplexer .The data to be compressed is shifted by the barrel shifter.

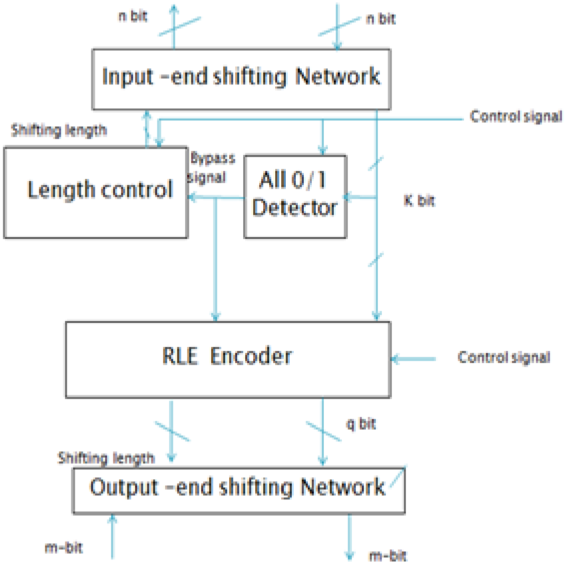


Figure.2: PaCC Encoder

1) Input-end shifting network
It shifts n-bit differential vector, V_{diff} from the volatile registers to the RLE encoding module. The n-bit output is the shifted value for updating the volatile registers.

2) Output-End Shifting Network
The output-end shifting network shifts the m-bit compression results to the NV registers.

3) All 0/1 detector block
The all 0/1 detector block helps to execute k-bit parallel observation and generate a bypass signal to the RLE encoder and length controller

4) Length controller
The length controller provides the shifting length to the input end shifting network according to the bypass signal as well as Observation window width k.

5) Run Length Encoder
The RLE encoder compresses the k-bit input serially when the bypass signal is disabled, otherwise bypasses the k-bit input. The format of the compression result is based on the threshold, L_{th} . Once the RLE encoder accomplishes the compression, it sends the q-bit compressed segment. Run length encoding is an easy, well known data compression method used in numerous application such as data transfer or efficient image storing. The output of a run length encoder is a sequence of runs. Run-length encoding (RLE) is a very simple form of data compression in which runs of data are stored as a single data value and count, rather than as the original run.

3.3 Two stage shifting Network

Two stage shifting network consist of Fixed N bit shifting network and 2N bit barrel shifter.

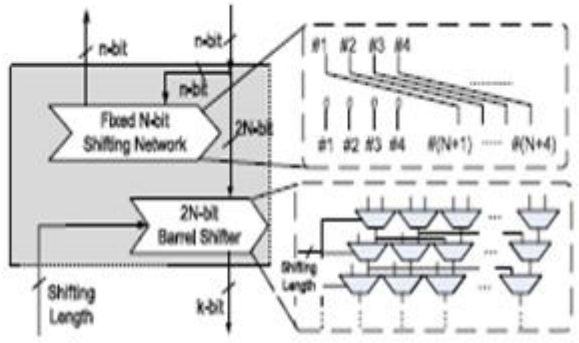


Figure 3: Two stage shifting network

Fixed N bit shifting network

Fixed N bit shifting network is used to simply shift the input by N. This shifter is extremely area efficient and no multiplexor is needed. The fixed shifting network is used to update the remaining part of the volatile registers.

2N bit barrel shifter

A Barrel shifter is a digital circuit that can shift a data word by a specified number of bits in one clock cycle. It can be implemented as a sequence of multiplexers and in such an implementation the output of one mux is connected to the input of the next mux. The data to be compressed is shifted by the barrel shifter

3.4 Data compression and decompression using Run length encoding(RLE)

Run-Length Encoding (RLE) is created especially for data with strings of repeated symbols. For example ,if we were to compress the string AAAABBBAACCCBBBB we would get the compressed string4A3B2A3C4B. Here one can see that the string of 15 bytes can be expressed as a string of 10 bytes. The main advantage of RLE is that it performs lossless data compression in which the orginal data can be perfectly reconstructed from the compressed data.

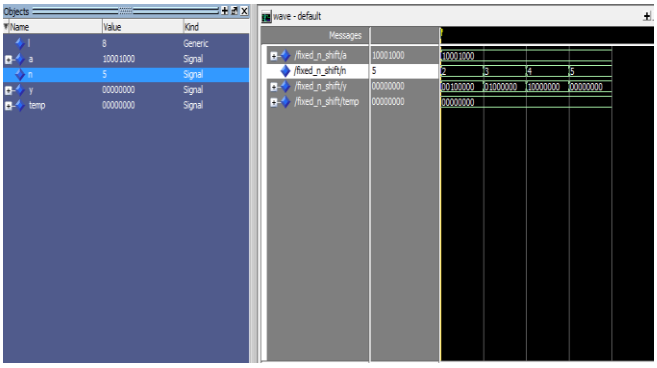
Run length encoding has a weakness of its reliability on the nature of the input data. In some cases, it can even give a larger compressed string than the orginal. For example if the input string is ABDBAC, the compressed string would be 1A1B1D1B1A1C,which is twice as large as the orginal. RLE is in the same way very efficient in an opposite scenario. If the input data is very similar then the compression rate is better than other algorithms.As an example if the input string isAAAAAAAAAA,the compressed data would be 10A,which is one fifth the size of the orginal.gate.

4. Results and Discussion

The modules are modelled using VHDL in Xilinx ISE Design Suite 12.1 and the simulation of the design is performed using Modelsim SE 6.3f to verify the functionality of the design. Here a structural model is used for the coding purpose.The parallel compare and compress encoder consist of shifting Networks,Run Length Encoder ,All 0/1 detector and length control. Coding for input end shifting Network, output end shifting network and Run Length Encoder is done. PACC Encoder is based on Threshold based parallel RLE algorithm. Then the whole block of PACC encoder was simulated.

4.1 Simulation Result of Fixed N bit shifting network

Fixed N bit shifting network is used to simply shift the input by N.This shifter is extremely area efficient and no multiplexor is needed. Let the 8 bit input data is a=10001000 shifting length,n=3 then output y=01000000

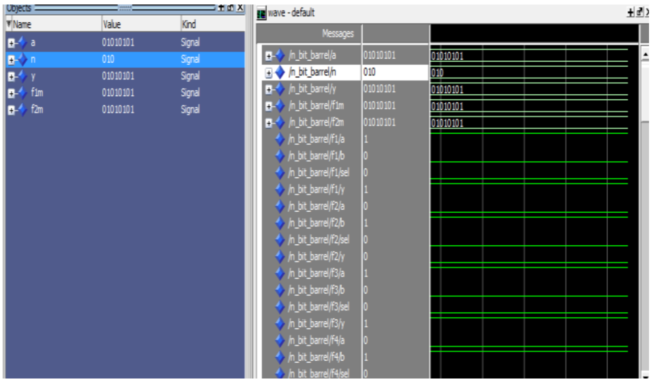


Name	Value	Kind
l	8	Generic
a	10001000	Signal
n	5	Signal
y	00000000	Signal
temp	00000000	Signal

Messages					
/fixed_n_shift/a	10001000	10001000			
/fixed_n_shift/n	5	2	3	4	5
/fixed_n_shift/y	00000000	00100000	01000000	10000000	00000000
/fixed_n_shift/temp	00000000	00000000			

Figure 4: Fixed N bit shifting network

4.2 Simulation Result of N bit barrel shifting network



Name	Value	Kind
a	01010101	Signal
n	010	Signal
y	01010101	Signal
f1m	01010101	Signal
f2m	01010101	Signal

Messages		
/n_bit_barrel/a	01010101	01010101
/n_bit_barrel/n	010	010
/n_bit_barrel/y	01010101	01010101
/n_bit_barrel/f1m	01010101	01010101
/n_bit_barrel/f2m	01010101	01010101
/n_bit_barrel/f1/a	1	
/n_bit_barrel/f1/b	0	
/n_bit_barrel/f1/sel	0	
/n_bit_barrel/f1/y	1	
/n_bit_barrel/f2/a	0	
/n_bit_barrel/f2/b	1	
/n_bit_barrel/f2/sel	0	
/n_bit_barrel/f2/y	0	
/n_bit_barrel/f3/a	1	
/n_bit_barrel/f3/b	0	
/n_bit_barrel/f3/sel	0	
/n_bit_barrel/f3/y	1	
/n_bit_barrel/f4/a	0	
/n_bit_barrel/f4/b	1	
/n_bit_barrel/f4/sel	0	

Figure 5: N bit barrel shifting network

A Barrel shifter is a digital circuit that can shift a data word by a specified number of bits in one clock cycle. It can be implemented as a sequence of multiplexers and in such an implementation the output of one multiplexer is connected to the input of the next multiplexer Let the 8 bit input data is a=01010101 shifting length,n=2(010) then output y=0101010

4.3 Simulation Result of Run Length Encoder

Run Length Encoding is a simple form of data compression. Same data value occurs in many consecutive data elements are stored as a single data value and count. Let the 32 bit input data is given as input a=10101011111100001000111111110000, Find how many zero count and one count are there in the sequence
zero count = 4 3 4 1 1 1 1 0 0 0 0
one count = 8 1 6 1 1 0 0 0 0 0 0

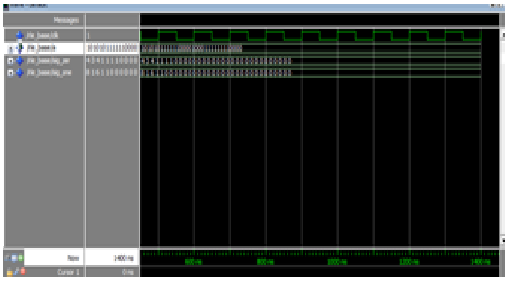
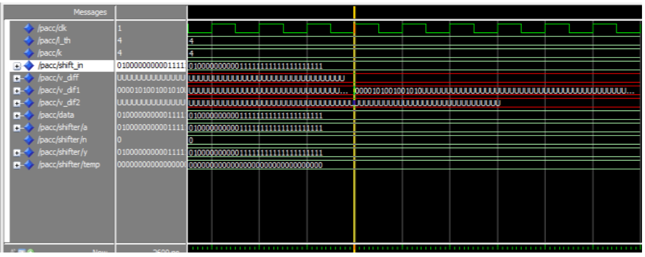


Figure 6: Run Length Encoder

4.4 Simulation Result of PaCC Encoder

Non volatile processors are manufactured with nonvolatile registers .Parallel compare and compress architecture is used to reduce the area of nonvolatile registers. PaCC encoder is based on threshold based parallel RLE algorithm. PaCC Encoder consist of input end shifting network, output end shifting network ,run length encoder,length control and all 0/1 detector. Let the 32 bit data is given as input. shift in = 01000000000011111111111111111111



```
Messages
/pacc/clk            1                  
/pacc/l_th           4                  4
/pacc/k              4                  4
/pacc/shift_in       0100000000001111   01000000000011111111111111111111
/pacc/v_diff         UUUUUUUUUUUUUU     UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
/pacc/v_dif1         0000101001001010   UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU...  0000101001001010UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU...
/pacc/v_dif2         UUUUUUUUUUUUUU     UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU
/pacc/data           0100000000001111   01000000000011111111111111111111
/pacc/shifter/a      0100000000001111   01000000000011111111111111111111
/pacc/shifter/n      0                  0
/pacc/shifter/y      0100000000001111   01000000000011111111111111111111
/pacc/shifter/temp   0000000000000000   00000000000000000000000000000000
```

Figure 7: PaCC Encoder

5. Conclusion

Parallel compare and compress architecture is used to reduce excess area of non volatile processors .Non volatile processors are manufactured with nonvolatile registers. A novel pacc codec and a reconfigurable state table were introduced and the corresponding heuristics were developed to optimize reference vector selection. PaCC architecture is used to reduce the number of the bit to be stored in the nonvolatile flipflops. Compression codec based on a parallel run length encoding scheme is used. An area-efficient two-stage shifting network is designed to minimize the codec area. Hybrid nonvolatile flipflops lead to significant area increase because they contain nonvolatile storage besides standard flipflops. The nonvolatile storage such as ferro electric capacitors ,magnetic tunnel junctions or floating gate transistors usually occupies a very large area. While self powered embedded system use on chip ferro electric random access memory to prevent data loss during power failures. By using pacc architecture in nonvolatile processors, it may reduce area overhead and increase the speed.

References

[1] Y Wang, Y. Liu, Y. Liu, D. Zhang, S. Li, B. Sai, M.-F. Chiang, and H. Yang,"Pacc: A Parallel compare and compress codec for area reduction in Nonvolatile processors", IEEE Transactions on Very Large Scale Integration Systems, vol. 22,no.7,July 2014.

[2] M. Poremba and Y. Xie NVMain: An architectural-level main memory simulator for emerging non-volatile memories,in Proc. ISVLSI, 2012.

[3] Y. Wang, Y. Liu, Y. Liu, D. Zhang, S. Li, B. Sai, M.-F. Chiang, and H. Yang, " A compression-based area efficient recovery architecture for nonvolatile processors",in Proc. DATE, Mar. 2012.

[4] Y. Wang, Y. Liu, S. Li, D. Zhang, B. Zhao, B. Sai, M.-F. Chiang,Y. Yan, and H. Yang, A 3us wake-up time nonvolatile processor basedon ferroelectric flip-flops, in Proc. ESSCIRC, Sep. 2012.

[5] W Yu, S. Rajwade, S. Wang, B. Lian, G. Suh, and E. Kan," non-volatile microcontroller with integrated floating-gate transistors, Proc. 5th Workshop Dependable Secure Nanocomput, 2011.

[6] M. Zwerg, A. Baumann, R. Kuhn, M. Arnold, R. Nerlich, M. Herzog,R. Ledwa, C. Sichert, V. Rzehak, P. Thanigai, and B. Eversmann,An 82 A/MHz microcontroller with embedded FeRAM for energyharvesting applications, in Proc. ISSCC, Feb. 2011.

[7] J Wang, Y. Liu, H. Yang, and H. Wang,"A compare-and-write ferroelectric non-volatile flipflop for energy-harvesting applications",in Proc. ICGCS, 2010

[8] X. Guo, E. Ipek, and T. Soyata, Resistive computation: Avoiding the power wall with lowleakage, STT-MRAM based computing, in Proc.37th AnnuISCA, 2010

[9] N. Sakimura, T. Sugibayashi, R. Nebashi, and N. Kasai, Nonvolatile magnetic flip-flop for standby-power-free SoCs, IEEE J. Solid-State Circuits, vol. 44, no. 8, Aug. 2009. 32

[10] J. Trein, A. Schwarzbacher, B. Hoppe, and K. Noff A hardware implementation of a run length encoding compression algorithm with parallel inputs, in Proc. ISSC, Jun. 2008.

[11] W. Zhao, E. Belhaire, V. Javerliac, C. Chappert, and B. Dieny ,A nonvolatile flip-flop in magnetic FPGA chip, in Proc. DTIS, Sep. 2006.