ECAL Data Compression Study

Ph. Busson, **A. Karar**, **G.B. Kim**

(LPNHE, Palaiseau, France)

Schedule of our Study in 1998

Our study \ Month		1	2	3	4	5	6	7	. . .	12
Time domain	Signal Modeling	→								
	Data compression		→	→	→	→	→	→	→	→
	DSP card				→	→	→	→	→	→
	Lossless data Compresion ICS							→	→	→
Space domain	Simulation the CMSIM							→	→	→

* We started studying data compression in space domain by using the CMSIM with **D.W. Kim, S.C. Lee .**

(Kangnung National University, South Korea)

PLAN

Different methods of data compression	Review	PDPCM (Predictive Differential Pulse Code Modulation) : interesting of USES IC
	New idea	Run-Length coding
		Mixed coding with Run-length and 8bit coding
Losseless compression ICs	Review	ALDC1_40S (IBM)
		AHA3231 (AHA) : data processing mode
	New IC	1. What is USES 2. How does it work ? 3. Result of software 4. Improvement of USES
Conclusion		Future test of lossless compression IC

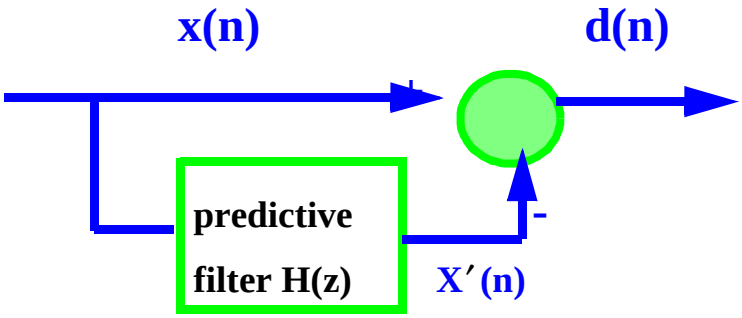
1. Different methods of data compression

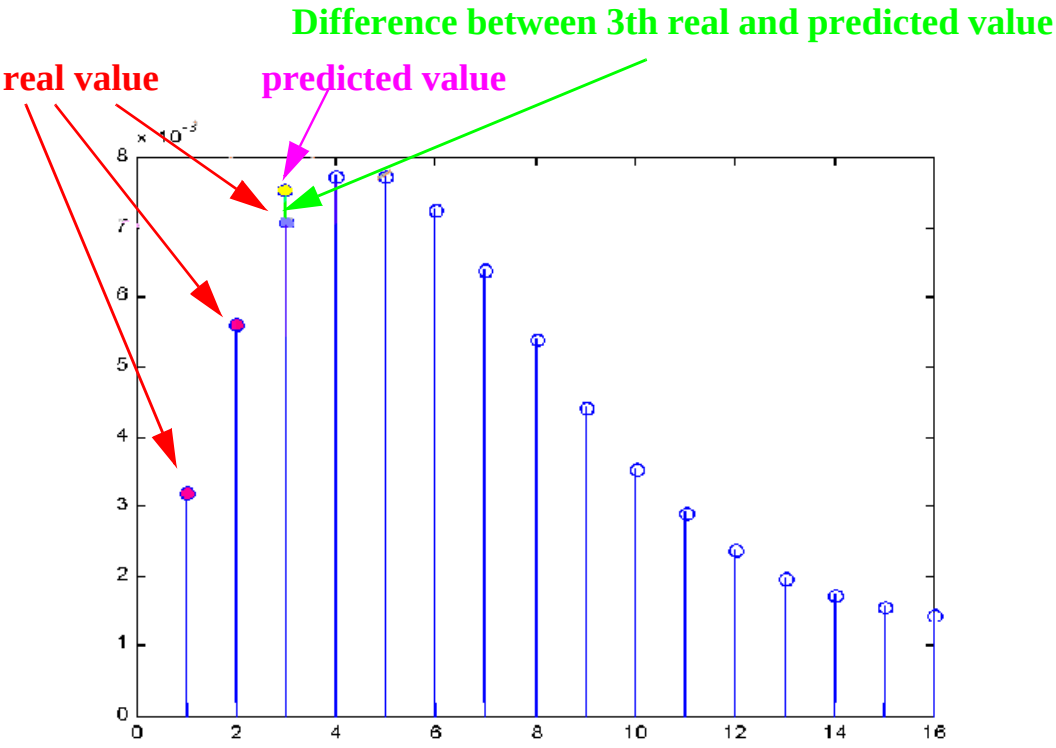
1.1 PDPCM (Predictive Differential Pulse Code Modulation)

This method is interesting in the application of USES chips.

The main idea of PDPCM is

to use a finite set of present and past samples to predict a sample in the future.





This method allows to reduce the code length.

1.2 New idea

If we will use FPU+ADC system, the Mixed coding with Run-length and 8 bits coding could be useful.

1.2.1 Run-length coding

The main idea : replace redundant data with tokens.

For example in the case of the following string:

A B C C C C C C A B C A B C

→A B !6C A B C A B C (after run-length coding)

using run-length coding, compression of the long string of C's can be done using some meta character.

1.2.2 Mixed coding with Run-length coding and 8 bits coding

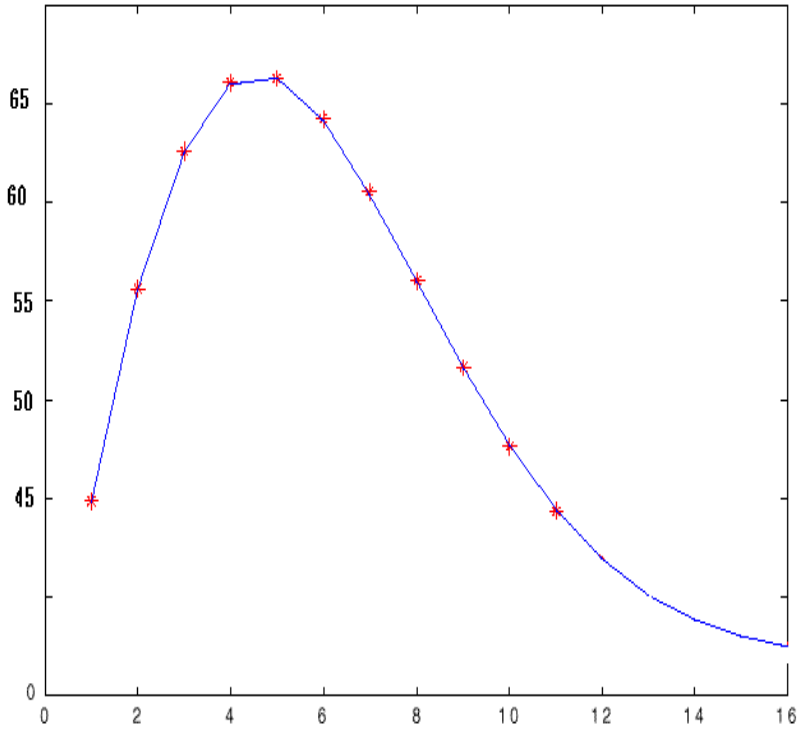
Main idea : for the values bigger than 8bit

Run-length coding could be useful

because there is a large number of redundant data with 0 value.

For example we consider 10 samples with following values:

The value
45
56
62
66
67
64
60
56
52
47



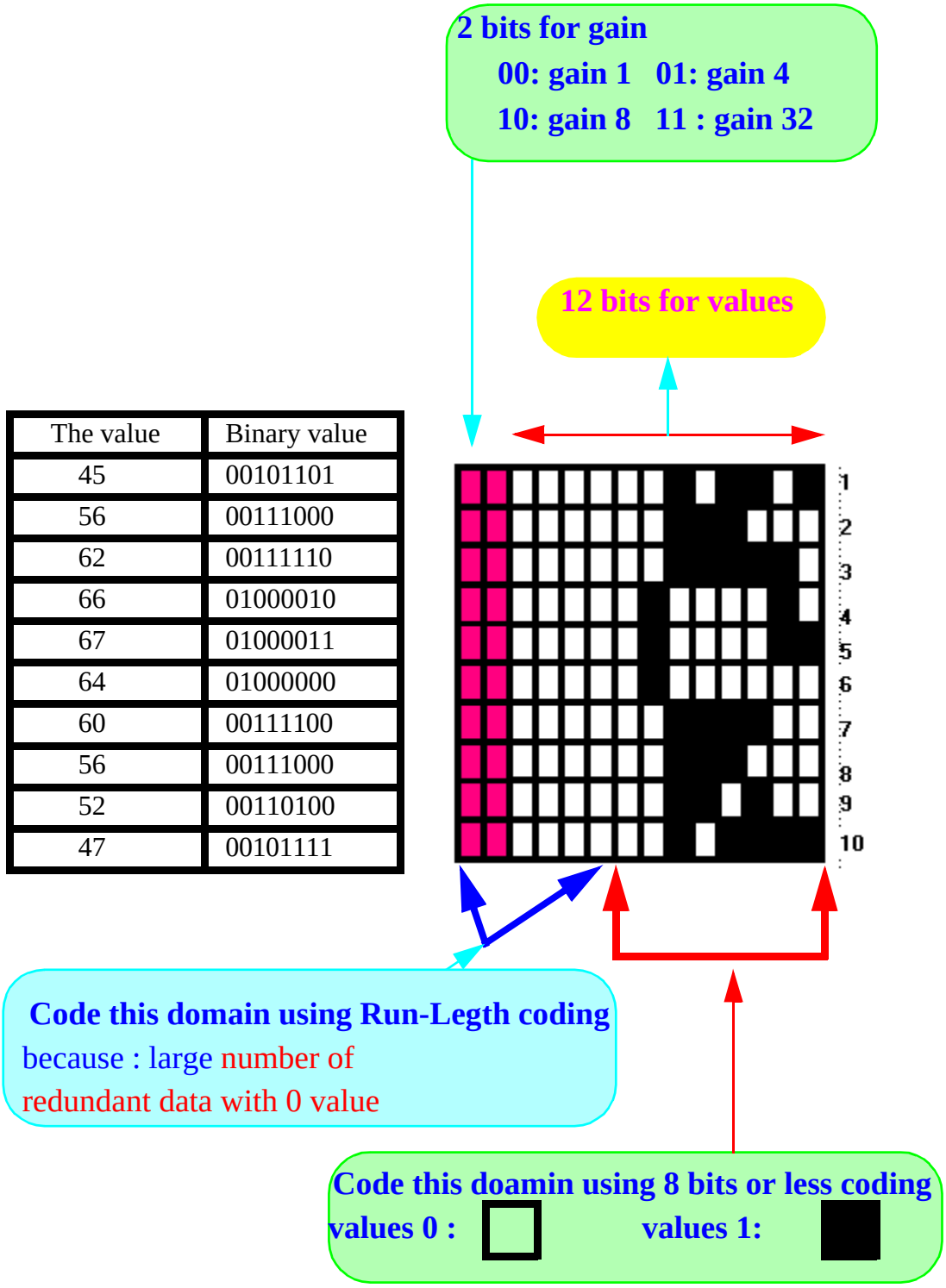
How we can code these values?

1. Transform these values to binary values.

In CMS, we will have 14 bits : 2 bits for gain, 12 bits for values

But in this typical example, the values can be coded with 8 bits.

The value	Binary value
45	00101101
56	00111000
62	00111110
66	01000010
67	01000011
64	01000000
60	00111100
56	00111000
52	00110100
47	00101111



2. Lossless data compression ICs

2.1 Review

We have already introduced two lossless compression ICs :

2.1.1 ALDC1_40S (Adaptive Lossless Data Compression)

ALDC1-40S is a member of a IBM's growing family of high performance general purpose "lossless" data compression products.

Performance:

* Lempel-Ziv algorithm (The dictionary method)
* Compression speeds up to 40 Mbytes/sec
* Decompression output data rate up to 40 Mbytes/sec
* Clock speeds up to 40 MHz
* Evaluation Software Available

2.1.2 AHA3231

The AHA3231 is produced by Advanced Hardware Architectures Inc.

Performance:

* Standard DCLZ (Data Compression Lempel-Ziv) algorithm
* Compression speeds up to 20 Mbytes/sec
* Decompression output data rate up to 20 Mbytes/sec
* Clock speeds up to 40 MHz
* Evaluation Software Available

Our conclusion using the files from CMSIM coded with 16 bits was :

DCLZ algorithm is better than the ALDC algorithm in case of our data.

2.2 Data processing Mode of AHA3212

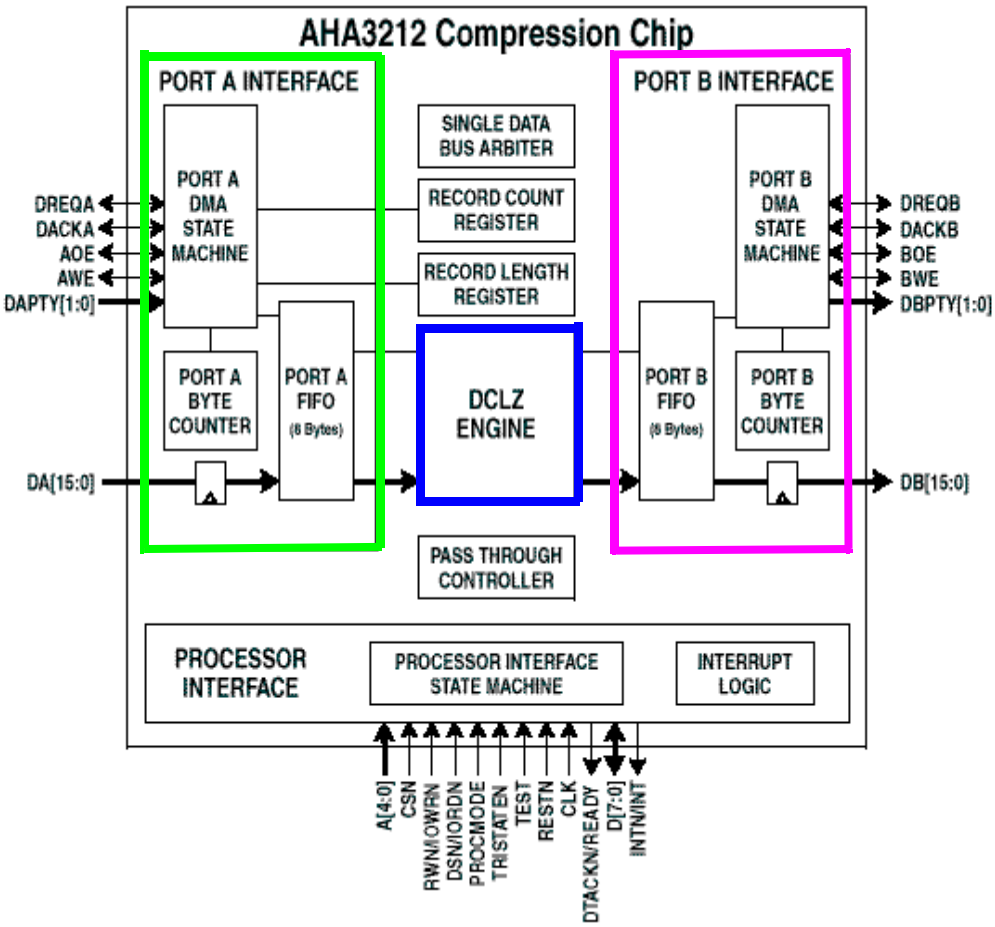
This chip has two independent DMA interface : Port A and Port B

For compression and decompression

Port A transfers uncompressed data

Port B transfers compressed codes

The configuration of the DMA interface is programmable



During compression mode

1. **uncompressed data flows into Port A**
2. **It is compressed by the DCLZ engine**
3. **The resulting compressed data is transferred out of Port B**

2.2 Introduction of new lossless compression IC

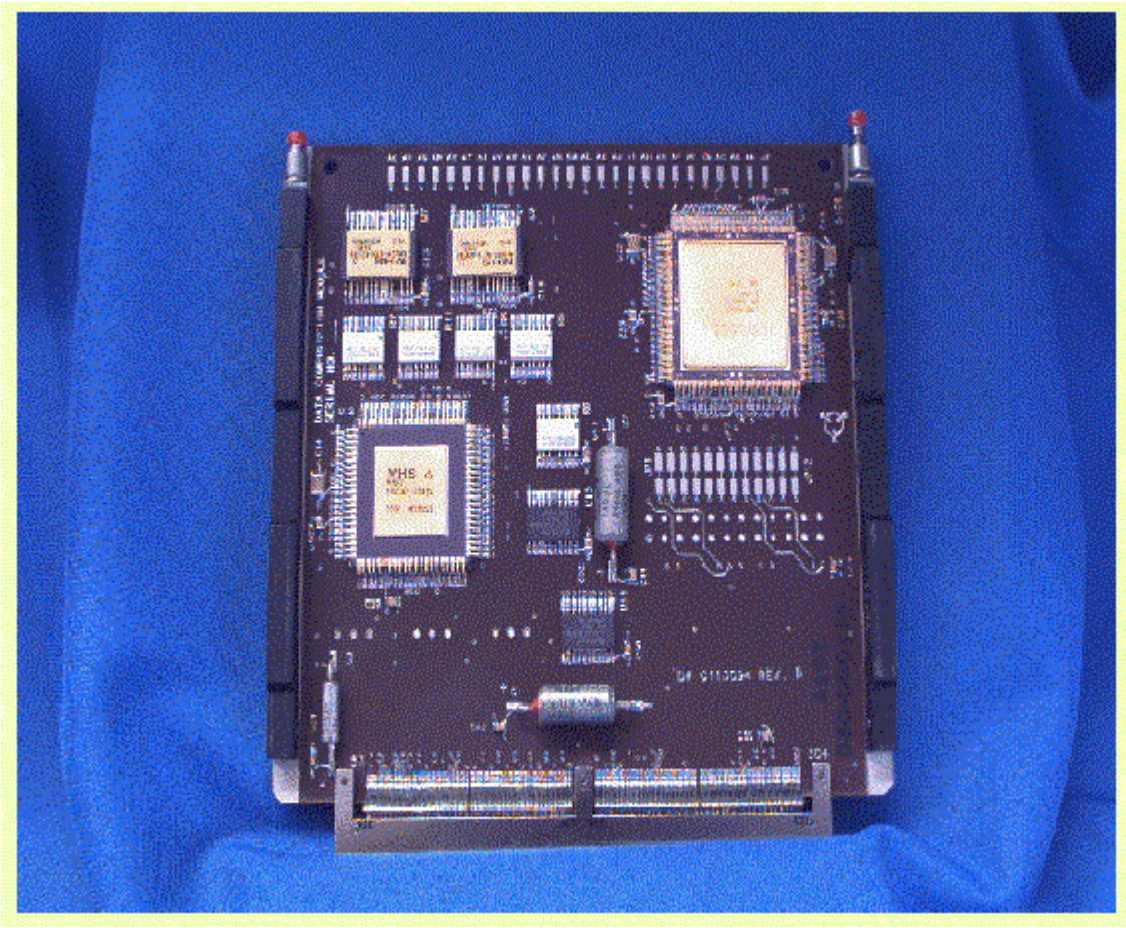
- **USES(Universal Source Encoding for Science Data) chip**

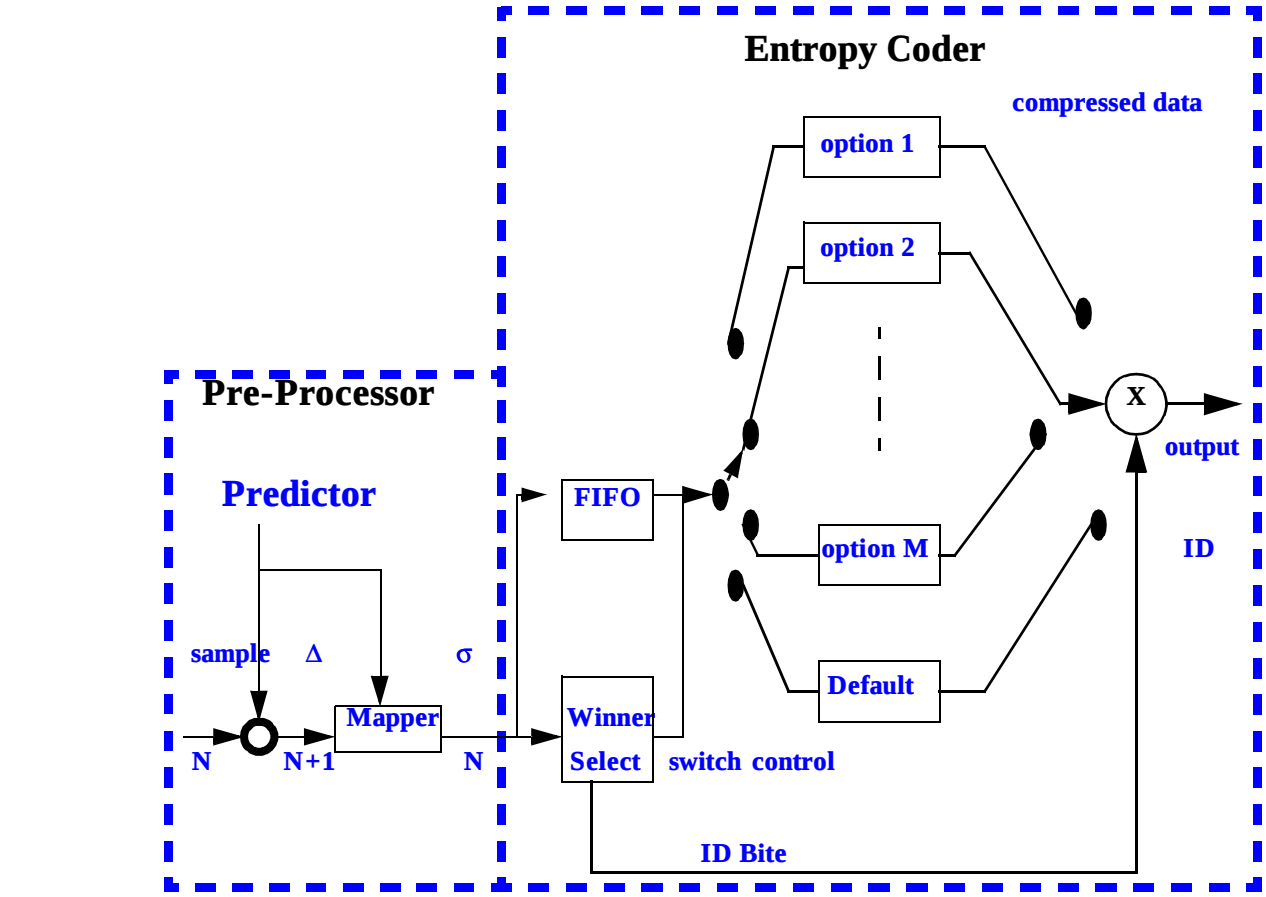
  This chip is produced by Microelectronics Research Center (MRC), was established by NASA.

2.2.1 What is USES ?

Performance:

* uses the mixed algorithm (Predictive coding and Entropy coding)
* Lossless compression up to 20 Msamples/sec
* Works on data of any quantization levels: 8-bit, 10-bit, 16-bit or more
* Speed relatively independent of quatization levels
* Evaluation Software Available (named szip.exe used in DOS)





2.2.2 How does it work?

1. **The architecture generally consists of two parts:**
   a preprocessor and an entropy coder

2. **Preprocessor: used to decorrelate data, can be user-defined**

3. **Entropy coder: has a set of options**

4. **Each entropy option is a unique Huffman coder WITHOUT a code table**

2.2.3 The result of evaluation software

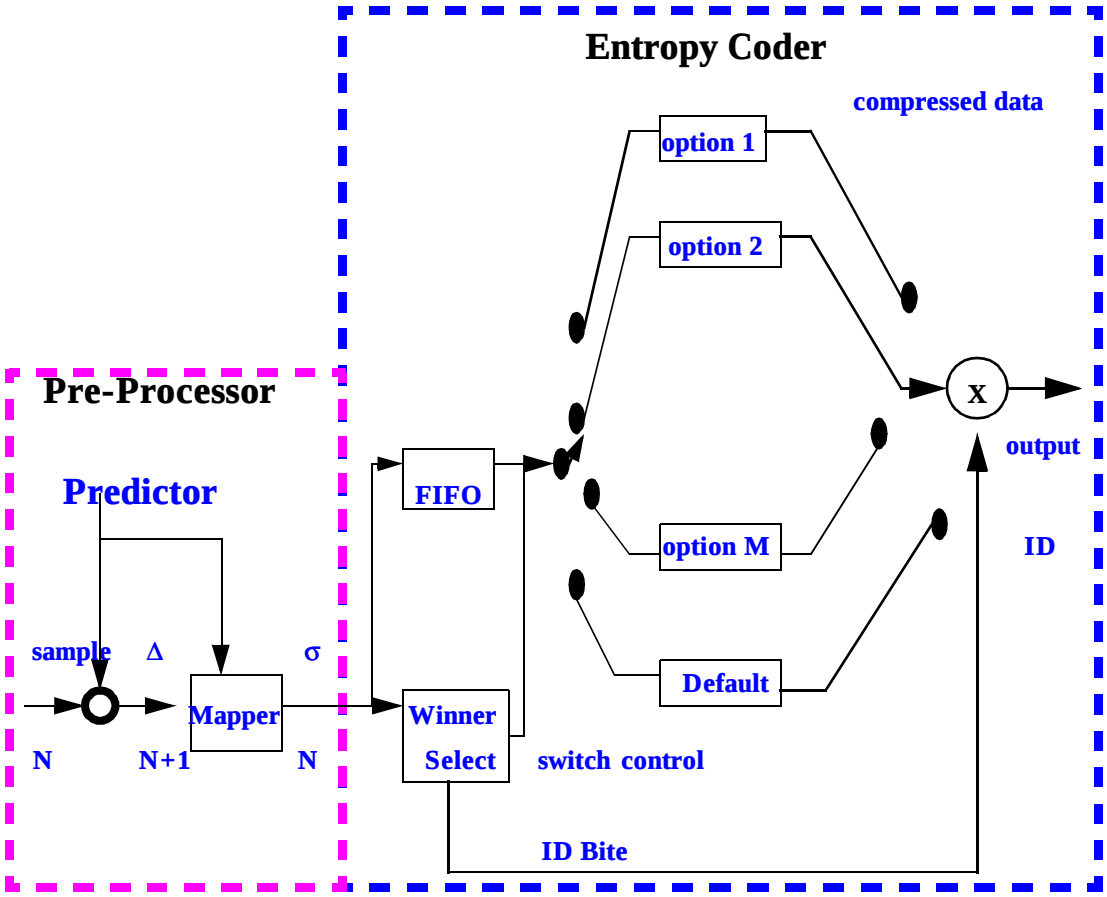
* Using the CMSIM files coded with 16 bits

File name	File size	DCLZ file size	DCLZ comp. factor	USES file size	USES comp. factor
S_0001.dat	99 148	29 025	3.41	83 576	1.19
S_0002.dat	126 568	39 046	3.24	106 638	1.19
S_0003.dat	86 596	25 689	3.37	72 748	1.19
S_0004.dat	58 648	16 757	3.50	49 386	1.19
S_0005.dat	87 282	24 592	3.54	73 282	1.19
S_0006.dat	115 590	28 823	4.01	97 288	1.19
S_0007.dat	67 204	18 548	3.62	56 557	1.19
S_0008.dat	87 492	22 695	3.85	73 621	1.19
S_0009.dat	108 666	33 720	3.22	91 692	1.19
S_0010.dat	113 644	32 864	3.45	98 809	1.15
S_0011.dat	114 440	34 707	3.29	96 215	1.19
S_0012.dat	130 870	37 376	3.50	110 161	1.19
S_0013.dat	142 806	36 463	3.91	120 169	1.19
S_0014.dat	159 260	48 337	3.29	134 029	1.19
S_0015.dat	213 552	63 747	3.35	179 179	1.19
S_0016.dat	161 104	49 181	3.27	135 723	1.19
S_0017.dat	103 080	31 332	3.29	86 883	1.19
S_0018.dat	162 602	45 416	3.58	136 688	1.19
S_0019.dat	86 622	24 236	3.57	72 713	1.19
S_0020.dat	106 814	31 662	3.37	89 567	1.19
	116 599	33 710	3.48	98 246	1.188

in case of our data, DCLZ algorithm is better than the USES algorithm .

But we think that the compression rate of USES algorithm can be increased by changing the existing predictive coding with our predictive coding algorithm.

2.2.4 Improvement of USES algorithm



We can change the predictor in the pre-processor with our predictive algorithm to increase the compression rate

Our Pre-Processor

our predictive algorithm

PDPCM(Predictive Differential Pulse Modulation)

Conclusion

We are interested in the lossless data compression ICs offered by :

1. IBM Microelectronics

(http://www.chips.ibm.com)

- **ALDC (Adaptive Lossless data compression) algorithm**
  - based on the Dictionary method

2. Advanced Hardware Achitectures

(http://www.aha.com)

- **DCLZ (Data compression Lemple Ziv) algorithm**
  - based on the Dictionary method

3. Microelectronics Research Center (MRC)

(http://www.mrc.unm.edu)

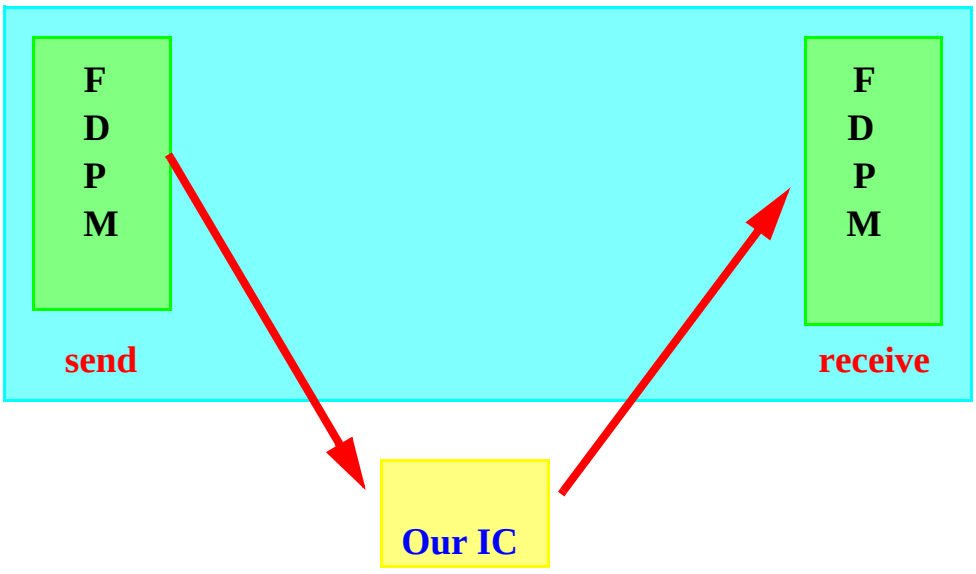
- **USES (Universal Source Encoding for Science Data) algorithm**
  - based on the predictive and entropy coding

- **Purchase the evaluation or design card**

Future test of the lossless compression IC

VME

Clock speeds up to 100 MHz for 32bits
: 20 MHz for 16 bits



To programme : use Labview in MAC and PC