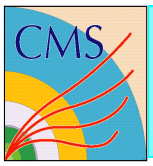


ECAL Data Compression Study

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)

1. Three Companies offer lossless data compression ICS

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. Stac Electronics

(http://www.stac.com)

2. What is the dictionary method ?

(Electrical Engineering and Computing in Croatia: http:/rasip.fer.hr)

2.1 Main Idea : These compression methods use the property of many data types to contain repeating code sequences.

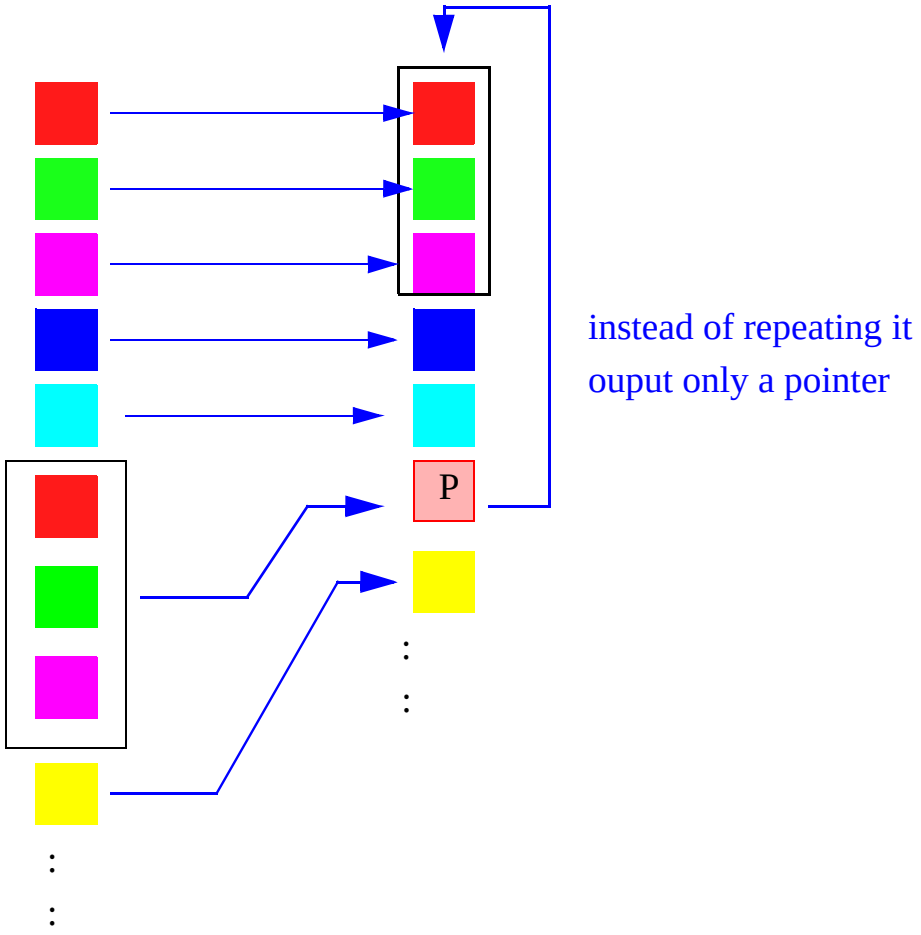
2.2 Two main groups

All the dictionary methods can be subdivided into two main groups.

* The first group : try to find if the character sequence currently being compressed has already occurred earlier in the input data instead of repeating it output only a pointer.

For example

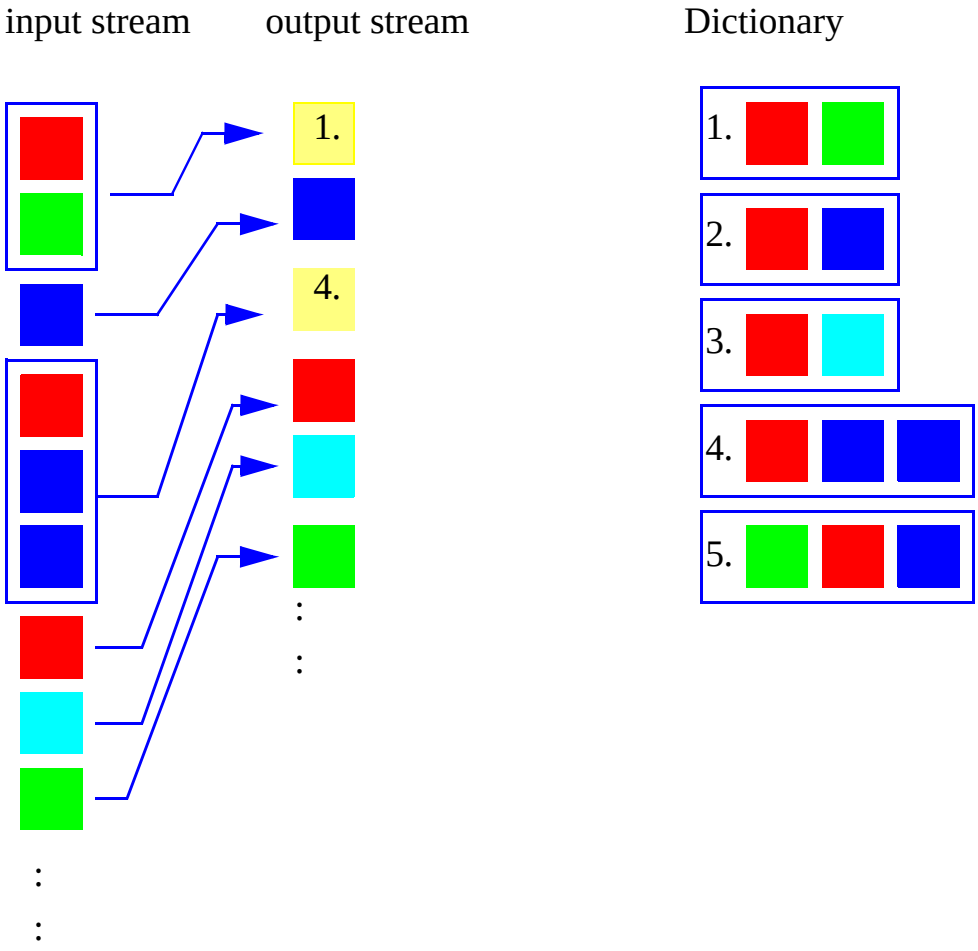
input stream output stream



* All the methods of this group are based on the algorithm developed and published in 1977 by Abraham Lempel and Jakob Ziv

* The second group : create a dictionary of the phrases that occurr in the input data.
When they encounter a phrase already present in the dictionary, they just ouput the index number of the phrase in the dictionary.

For example:



* All the methods of this group are based on the algorithm developed and published in 1978 by Abraham Lempel and Jakob Ziv

3. ALDC1_40S and AHA3231 IC

3.1 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

3.2 AHA3231

The AHA3231 is produced by Advanced Hardware Architectures Inc.

Performance:

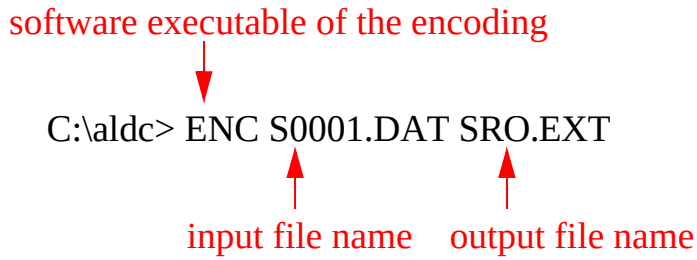
* Standard DCLZ (Data Compression Lempel-Ziv) algorithm
* Compression speeds up to 20 Mbytes/sec
* Decompression output data rate up to 20 Mbytes/sec
* Clock speeds up to 40 MHz
* Evaluation Software Available

4. Evaluation software available

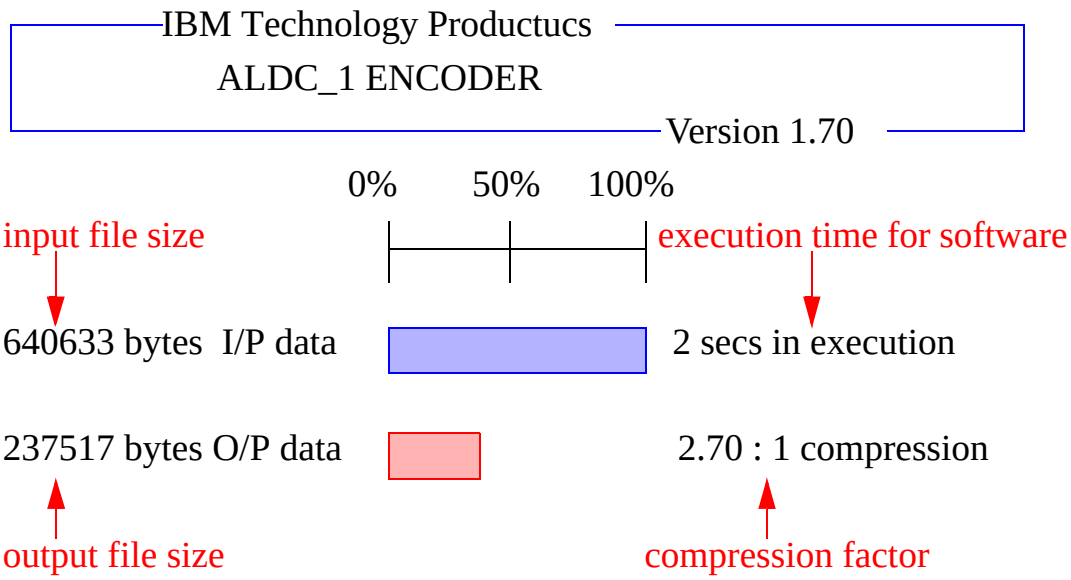
4.1 Evaluation software for ALDC produced by IBM

The evaluation software for ALDC algorithm can be used in DOS Software executable of the encoder and decoder is named ENC.EXE and DEC.EXE

For example : In DOS system

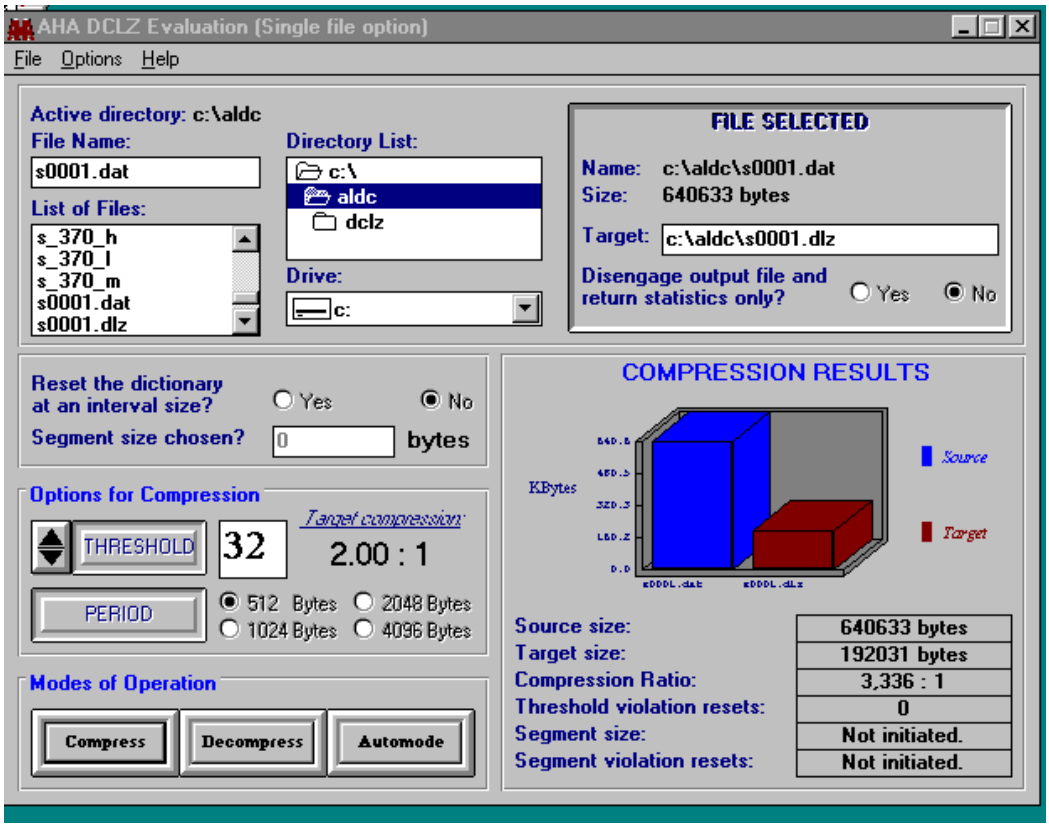


The result



4.2 Evaluation software for DCLZ produced by Advanced Hardware Achitectures

The evaluation software for DCLZ algorithm can be used in Window 95



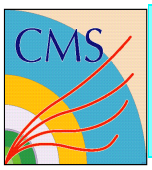
5. The result of software for DCLZ and ALDC algorithm

5.1 Using the files saved by 16 bits from CMSIM

File Name	File Size (byte)	DCLZ File Size (byte)	DCLZ Compression factor	ALDC File size (byte)	ALDC Compression Factor
SRO_16__0001.DAT	640 633	192 031	3.33	237 517	2.70
SRO_16__0002.DAT	747 422	220 521	3.38	276 638	2.70
SRO_16__0003.DAT	493 827	148 491	3.32	189 895	2.60
SRO_16__0004.DAT	543 177	158 844	3.42	202 069	2.69
SRO_16__0005.DAT	454 954	120 397	3.77	168 566	2.70
SRO_16__0006.DAT	648 307	185 035	3.54	237 767	2.73
SRO_16__0007.DAT	616 173	177 492	3.47	224 857	2.74
SRO_16__0008.DAT	428 233	124 539	3.43	161 130	2.66
SRO_16__0009.DAT	676 336	182 534	3.70	251 782	2.69
SRO_16__0010.DAT	719 341	207 501	3.46	264 141	2.72
SRO_16__0011.DAT	685 241	179 504	3.81	252 461	2.71
SRO_16__0012.DAT	683 318	202 889	3.36	254 999	2.68
SRO_16__0013.DAT	672 295	195 863	3.43	249 678	2.69
SRO_16__0014.DAT	797 847	233 569	3.41	296 501	2.69
SRO_16__0015.DAT	898 654	265 294	3.38	331 648	2.71
SRO_16__0016.DAT	787 450	236 471	3.33	293 601	2.68
SRO_16__0017.DAT	617 017	183 806	3.35	230 819	2.67
SRO_16__0018.DAT	793 231	184 417	3.81	260 079	2.70
SRO_16__0019.DAT	638 857	186 338	3.42	236 899	2.70
SRO_16__0020.DAT	710 997	209 233	3.39	264 428	2.69
SRO_16__0021.DAT	657 427	188 212	3.49	241 478	2.72
SRO_16__0022.DAT	748 527	224 601	3.33	279 955	2.67
SRO_16__0023.DAT	570 782	164 514	3.47	210 708	2.71
SRO_16__0024.DAT	606 985	173 643	3.49	230 735	2.63
SRO_16__0025.DAT	720 501	212 347	3.39	268 047	2.69
SRO_16__0026.DAT	492 318	144 238	3.41	185 458	2.65
SRO_16__0027.DAT	548 048	171 801	3.19	209 917	2.61
SRO_16__0028.DAT	660 148	197 330	3.34	249 970	2.64
SRO_16__0029.DAT	633 381	183 436	3.45	233 781	2.71
SRO_16__0030.DAT	615 621	179 866	3.42	218 166	2.70
SRO_16__0031.DAT	750 944	219 878	3.41	274 519	2.74
SRO_16__0032.DAT	690 565	203 580	3.39	256 839	2.69
SRO_16__0033.DAT	614 202	181 332	3.38	229 815	2.67
SRO_16__0034.DAT	957 869	295 571	3.24	353 182	2.71
SRO_16__0035.DAT	641 046	187 603	3.41	236 162	2.71
SRO_16__0036.DAT	798 204	243 154	3.28	296 869	2.69
SRO_16__0037.DAT	751 941	215 846	3.48	281 114	2.67
SRO_16__0038.DAT	560 231	161 994	3.45	206 746	2.71
SRO_16__0039.DAT	637 706	187 003	3.41	237 100	2.69
SRO_16__0040.DAT	655 223	195 041	3.35	241 660	2.71
Mean value	661 870	193 140	3.43	245 690	2.69

5.2 Using the files saved by dynamic coding from CMSIM

File Name	File (byte)	DCLZ File Size (byte)	DCLZ Compression factor	ALDC File size (byte)	ALDC Compression factor
SRO_DYN_0001.DAT	321 847	150 731	2.13	183 280	1.76
SRO_DYN_0002.DAT	375 487	174 927	2.14	213 450	1.76
SRO_DYN_0003.DAT	248 274	123 161	2.01	146 335	1.70
SRO_DYN_0004.DAT	272 956	127 992	2.13	156 160	1.75
SRO_DYN_0005.DAT	228 601	106 277	2.15	130 090	1.76
SRO_DYN_0006.DAT	325 669	149 469	2.17	183 689	1.77
SRO_DYN_0007.DAT	309 598	141 774	2.18	173 777	1.78
SRO_DYN_0008.DAT	215 159	102 462	2.10	124 260	1.73
SRO_DYN_0009.DAT	339 806	159 767	2.12	194 277	1.75
SRO_DYN_0010.DAT	361 418	168 512	2.14	204 251	1.77
SRO_DYN_0011.DAT	344 269	160 061	2.15	194 951	1.77
SRO_DYN_0012.DAT	343 363	162 278	2.11	196 707	1.75
SRO_DYN_0013.DAT	337 772	158 622	2.12	192 736	1.75
SRO_DYN_0014.DAT	400 811	188 738	2.12	228 615	1.75
SRO_DYN_0015.DAT	451 448	209 867	2.15	255 812	1.76
SRO_DYN_0016.DAT	395 590	185 569	2.13	226 595	1.75
SRO_DYN_0017.DAT	310 036	146 153	2.12	178 069	1.74
SRO_DYN_0018.DAT	353 305	164 041	2.15	200 854	1.76
SRO_DYN_0019.DAT	320 996	151 141	2.12	182 866	1.76
SRO_DYN_0020.DAT	357 184	167 499	2.13	204 032	1.75
SRO_DYN_0021.DAT	330 265	153 079	2.16	186 607	1.77
SRO_DYN_0022.DAT	376 058	178 796	2.10	215 769	1.74
SRO_DYN_0023.DAT	286 801	134 861	2.12	162 794	1.76
SRO_DYN_0024.DAT	305 064	148 983	2.04	177 725	1.72
SRO_DYN_0025.DAT	362 032	171 206	2.15	206 858	1.75
SRO_DYN_0026.DAT	247 392	119 205	2.04	143 002	1.73
SRO_DYN_0027.DAT	275 481	134 929	2.04	161 638	1.70
SRO_DYN_0028.DAT	331 707	159 999	2.07	192 670	1.72
SRO_DYN_0029.DAT	318 200	147 660	2.15	180 509	1.76
SRO_DYN_0030.DAT	309 316	143 746	2.15	176 174	1.76
SRO_DYN_0031.DAT	377 287	173 782	2.17	211 872	1.78
SRO_DYN_0032.DAT	346 935	162 832	2.13	198 261	1.75
SRO_DYN_0033.DAT	308 598	145 817	2.11	177 102	1.74
SRO_DYN_0034.DAT	481 208	221 820	2.16	272 600	1.77
SRO_DYN_0035.DAT	322 095	149 759	2.15	182 285	1.75
SRO_DYN_0036.DAT	401 062	188 545	2.12	228 949	1.74
SRO_DYN_0037.DAT	377 873	178 779	2.11	216 748	1.76
SRO_DYN_0038.DAT	281 443	131 205	2.14	159 572	1.74
SRO_DYN_0039.DAT	320 464	150 372	2.13	182 865	1.75
SRO_DYN_0040.DAT	329 200	152 374	2.16	186 714	1.76
Mean value	332 550	156 170	2.12	189 788	1.75



Conclusion

In case of our data,

DCLZ algorithm is better than the ALDCalgorithm.

3.43 compression factors in case without dynamic coding

2.12 compression factors in case with dynamic coding

If we consider both together DCLZ algorithm and dynamic coding

we can have about 4 compression factors :

2 compression factors from dynamic coding

\+ 2 compression factors from DCLZ algorithm

4 compression factors

Future of our study

- **find the best algorithm for our data**
- **buy or design an evaluation card**