NASA TECHNICAL NOTE

NASA TN D-4290

c. 1

NASA TN D-4290

A LOGARITHMIC ENCODER FOR BINARY WORD COMPRESSION

by Joseph A. Sciulli

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MARCH 1968

A LOGARITHMIC ENCODER FOR BINARY WORD COMPRESSION

By Joseph A. Sciulli

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

ABSTRACT

During the past several years efficient information transmission and processing techniques have attracted wide interest and have increased applicability. The present paper describes a logarithmic encoding device which has had particular application in energetic particle detection experiments. The paper provides a generalized encoding error analysis in order to evaluate the performance of the device. Both peak and average error are derived in terms of word size and desired accuracy. The implementation of a flexible logarithmic encoder is also described.

CONTENTS

A LOGARITHMIC ENCODER FOR BINARY WORD COMPRESSION

by
Joseph A. Sciulli
Goddard Space Flight Center

INTRODUCTION

Growing interest in the development of efficient information transmission techniques is stimulating much theoretical and experimental work. Particular emphasis has been placed on the development of video data compression techniques which exploit the predictability (or redundancy) of the data (Reference 1). Many of these techniques are not specifically designed for the video information source and can, in principle, be applied to any information source. For example, recent experiments have applied a simple predictor/encoder compression technique to energetic particle experiment data. The experimental data were obtained from a multi-channel device flown on the 1963-38C-APL satellite in which each channel was designed to detect particles in a specific energy region. Typically, an experiment of this kind requires a rather large word size because of the wide range of possible sample values. (The experiment considered used 16 bits per sample.) The compression simulations showed that average energy savings of approximately 4 db can be achieved with an allowable peak error of approximately 5 percent.

The main difficulty with such an approach, however, is defining an error criterion which is suitable over the entire range of possible sample values. Logarithmic encoding, however, is extremely useful in experiments of this type which involve wide dynamic range. The function of the encoder is to reduce the number of transmitted bits per sample while maintaining a relatively small error in the word reconstructed at the receiver. This paper describes a logarithmic encoding procedure and develops an expression for the error between the original sample value and the sample value reconstructed at the receiver. In addition, several design considerations are outlined and the operation of the device is discussed briefly. This procedure results in a fixed compression, independent of the *predictability* of the data. Since the bit rate at the output of the encoder is constant, there is no need to consider the yet-unsolved buffer queuing problem.

ENCODING PROCEDURE

Suppose each sample is originally represented by an n-bit binary word S with components $(\sigma_{n-1}, \sigma_{n-2}, \cdots \sigma_{n-j}, \sigma_{n-j-1} \cdots \sigma_0)$. The most significant bit (MSB) is σ_{n-1}; σ_0 is the least significant bit (LSB); and σ_{n-j} is the first component (from the left) which is a 1. We now describe an encoder which converts S into two subwords, the characteristic (C) and the mantissa (M). The mantissa is a k-bit word specifying the k MSBs following σ_{n-j}. The position of $\sigma_{n-j-k+1}$ is specified by C, an r-bit word. The encoder output, therefore, consists of $(r + k)$ bits where $(r + k) \leq n$. The operation is best illustrated by an example:

let $n = 19$, $k = 4$, and $r = 4$ and suppose the input word is $S = 1011001110110010110$.

Here the first 1 appears at the MSB, hence $j = 1$. The 4 bits of M must then consist of the second, third, fourth, and fifth MSBs of S (i.e., $M = 0110$). The position of the LSB of M must then be specified by C. Thus, $C = n - k - j + 1 = 15$ or $C = 1111$. The encoder output is then,

$$\underbrace{1111}_{C} \qquad \underbrace{0110}_{M} .$$

Some additional examples are

Input	Output
1111111111111111111	11111111
1000000000000000000	11110000
0000000000110000000	01011000
0000000000000010000	00010000

This example was given by Schaefer (Reference 2).

The general design problem consists of choosing k and r for some specified n. Since k determines how closely the encoded word approximates the input word (n-bits), k should be chosen to satisfy the error specification. After k is fixed, r can be chosen to minimize the number of bits in the output word. The bit compression ratio $n/(r + k)$ then provides a useful system performance measure. In the next section error expressions assuming transmission in a noiseless channel are developed.

ENCODING ERROR

In practice, an experimenter might specify an upper limit on either the average error, the maximum error, or both. Here both the average error and the maximum error for the special case of equally likely samples are calculated.

The input word S for some value of j is, by definition,

$$S = \sum_{i=0}^{n-j} \sigma_i \; 2^i \; . \tag{1}$$

According to the encoding rule, the word reconstructed at the receiver must be

$$\tilde{S} = \sum_{i=n-j-k}^{n-j} \sigma_i \; 2^i \; , \tag{2}$$

and the error for a given j

$$E_j = S - \tilde{S} = \sum_{i=0}^{n-j-k-1} \sigma_i \; 2^i \; . \tag{3}$$

We may now calculate the expected error for a given j according to

$$\overline{E}_j = E\{E_j\} = \sum_{i=0}^{n-j-k-1} E(\sigma_i) \; 2^i \; . \tag{4}$$

If all input words are equally likely, $E(\sigma_i) = 1/2$. Therefore,

$$\overline{E}_j = \frac{1}{2} \sum_{i=0}^{n-j-k-1} 2^i = \frac{1}{2}\left[2^{n-j-k} - 1\right] \; . \tag{5}$$

To determine the average error we must average $\overline{E}_j$ over all values of j for which an error can occur. That is,

$$\overline{E} = \sum_j \overline{E}_j \; p(j) \; ; \qquad p(j) = \begin{cases} \dfrac{1}{2^j} & \text{for } j = 1, 2, \cdots \; (n-k-1) \\[2ex] \dfrac{1}{2^{n-k-1}} & \text{for } j = (n-k) \end{cases}$$

where $\overline{E}_j = 0$ for $j = (n-k)$. Therefore,

$$\overline{E} = \sum_{j=1}^{n-k-1} \frac{1}{2}\left[2^{n-j-k} - 1\right] \frac{1}{2^j}$$

$$= 2^{n-k-1}\left\{\sum_{j=1}^{n-k-1} 2^{-2j}\right\} - \frac{1}{2}\left\{\sum_{j=1}^{n-k-1} 2^{-j}\right\} . \tag{6}$$

Since each term in braces is a geometric series,

$$\overline{E} = 2^{n-k-1}\left\{\frac{1}{3}\left(1 - 2^{-2(n-k-1)}\right)\right\} - \frac{1}{2}\left\{1 - 2^{-(n-k-1)}\right\} , \tag{7}$$

and, collecting terms,

$$\overline{E} = \frac{1}{3}\left[2^{n-k-1} + 2^{-(n-k)} - \frac{3}{2}\right] . \tag{8}$$

We can also calculate the maximum error. Returning to Equation 3, the error for a given j is

$$E_j = \sum_{i=0}^{n-j-k-1} \sigma_i \, 2^i .$$

Now the maximum error must obviously occur for $j = 1$ and $\sigma_i = 1$ for all i. Therefore,

$$E_{max} = \max_j E_j = \sum_{i=0}^{n-k-2} 2^i = \left[2^{n-k-1} - 1\right] . \tag{9}$$

The average error as a fraction of full scale is

$$\overline{E}_F = \frac{\left[2^{n-k-1} + 2^{-(n-k)} - \frac{3}{2}\right]}{3\left(2^n - 1\right)} , \tag{10}$$

and, for $(n - k) >> 1$,

$$\overline{E}_F \doteq \frac{2^{n-k-1}}{3\left(2^n\right)} = \frac{1}{3} \cdot \frac{1}{2^{k+1}} . \tag{11}$$

Similarly, the maximum error as a fraction of full scale is

$$E_{F[max]} = \frac{2^{n-k-1} - 1}{2^n - 1} , \tag{12}$$

and, again, for $(n - k) >> 1$,

$$E_{F[max]} \doteq \frac{1}{2^{k+1}} . \tag{13}$$

The appendix lists both $\overline{E}_F$ and $E_{F[max]}$ for various combinations of n and k. Thus, for a given n, a k can be chosen, according to Equations 10 and 12, to satisfy given error requirements. After the value of k is established r should be chosen as the smallest integer satisfying

$$(n - k) \leq 2^r - 1 \ ,$$

or

$$r \geq \text{Log}_2 \ (n - k + 1) \ . \tag{14}$$

If equality is obtained in Equation 14, then all possible values of C can occur.

IMPLEMENTATION

The implementation of this device is quite simple. A generalized system is shown in Figure 1. The n-bit input word is presented serially (MSB first) to a k-bit register. The input word is shifted until a 1 appears in the kth bit of the register. The next clock pulse changes the state of the control flip flop, inhibiting the shift register and starting the r-bit counter. At the nth clock pulse the shift register contains the k bits of M and the counter contains the r bits of C. The contents of the register and counter can then be transferred to an output register. The elements of the encoder are then reset, and the device is prepared to receive the next input word.

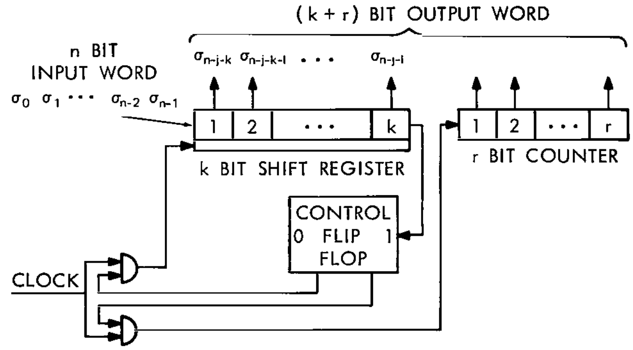

Figure 1–Implementation of the encoder.

CONCLUSIONS

This encoding procedure provides a useful device which achieves fixed, but modest, bit compression. For example, the 16-bit words of the energetic particle experiment could be encoded to 8 bits ($k = 4$, $r = 4$), resulting in a bit compression ratio of 2:1. This gives an average error of approximately 1 percent and a maximum error of about 3 percent. Moreover, this device could be combined with a zero-order predictor/run-length encoder to achieve further compression. The zero order hold compression might even be applied in the bit planes of C since these should be relatively quiescent from sample to sample.

For a slight increase in complexity, the logarithmic encoder could be made more flexible by varying k and r on command from the ground. Thus, the experimenter would have the capability of selecting the allowable error in the data depending on the activity of his experiment at a given time.

Goddard Space Flight Center
National Aeronautics and Space Administration
Greenbelt, Maryland, August 1, 1967
160-44-05-14-51

REFERENCES

1. Kutz, R. L. and Sciulli, J. A., "An Adaptive Image Data Compression System and its Performances in a Noisy Channel," presented at the International Symposium on Information Theory, San Ramo, Italy, September 11-15, 1967, submitted to IEEE Transactions on Information Theory.

2. Schaefer, D. H., "Logarithmic Compression of Binary Numbers," *Proceedings of the IRE* (Correspondence) 49 (7): July 1961.

Appendix

Listing of $\overline{E}_F$ and $E_{F[max]}$ for Combinations of n and k

n	k	$\overline{E}_F$	$E_{F[max]}$
2	1	0.	0.
3	1	0.03571428	0.14285714
3	2	0.	0.
4	1	0.05833333	0.20000000
4	2	0.01666667	0.06666655
4	3	0.	0.
5	1	0.07056452	0.22580644
5	2	0.02822580	0.09677419
5	3	0.00806452	0.03225806
5	4	0.	0.
6	1	0.07688492	0.23809523
6	2	0.03472222	0.11111110
6	3	0.01388889	0.04761904
6	4	0.00396825	0.01587301
6	5	0.	0.
7	1	0.08009350	0.24409448
7	2	0.03813976	0.11811023
7	3	0.01722441	0.05511811
7	4	0.00688976	0.02362204
7	5	0.00196850	0.00787401
7	6	0.	0.
8	1	0.08170956	0.24705882
8	2	0.03988970	0.12156862
8	3	0.01999509	0.05882353
8	4	0.00857843	0.02745098
8	5	0.00343137	0.01176471
8	6	0.00098039	0.00392157
8	7	0.	0.
9	1	0.08252048	0.24853229
9	2	0.04077482	0.12328767
9	3	0.01990582	0.06066536
9	4	0.00947896	0.02935421
9	5	0.00428082	0.01369863
9	6	0.00171232	0.00587084
9	7	0.00048923	0.00195695
9	8	0.	0.
10	1	0.08292667	0.24926686
10	2	0.04121991	0.12414467
10	3	0.02036748	0.06158357
10	4	0.00994318	0.03030302
10	5	0.00473484	0.01466275
10	6	0.00213832	0.00684261
10	7	0.00085533	0.00293255
10	8	0.00024438	0.00097752
10	9	0.	0.
11	1	0.08312994	0.24963360
11	2	0.04144308	0.12457254
11	3	0.02059989	0.06204201
11	4	0.01017877	0.03077674
11	5	0.00496916	0.01514411
11	6	0.00236627	0.00732780
11	7	0.00106864	0.00341964
11	8	0.00042745	0.00146556
11	9	0.00012213	0.00048852
11	10	0.	0.
12	1	0.08323162	0.24981684
12	2	0.04155482	0.12478632
12	3	0.02071647	0.06227106
12	4	0.01029743	0.03101343
12	5	0.00508814	0.01538461
12	6	0.00248397	0.00757021
12	7	0.00118284	0.00366300
12	8	0.00053418	0.00170940
12	9	0.00021368	0.00073260
12	10	0.00006105	0.00024420
12	11	0.	0.
13	1	0.08328247	0.24990843
13	2	0.04161073	0.12489317
13	3	0.02077487	0.06238554
13	4	0.01035697	0.03113173
13	5	0.00514808	0.01550482
13	6	0.00254375	0.00769137
13	7	0.00124183	0.00378464
13	8	0.00059135	0.00183128
13	9	0.00026706	0.00085460
13	10	0.00010682	0.00036626
13	11	0.00003052	0.00012209
13	12	0.	0.
14	1	0.08330790	0.24995422
14	2	0.04163869	0.12494659
14	3	0.02080409	0.06244277
14	4	0.01038680	0.03119086
14	5	0.00517817	0.01556491
14	6	0.00257389	0.00775193
14	7	0.00127180	0.00384545
14	8	0.00062088	0.00189220
14	9	0.00029565	0.00091558
14	10	0.00013352	0.00042727
14	11	0.00005341	0.00018311
14	12	0.00001526	0.00006104
14	13	0.	0.
15	1	0.08332061	0.24997710
15	2	0.04165268	0.12497329
15	3	0.02081871	0.06247138
15	4	0.01040173	0.03122043
15	5	0.00519324	0.01559495
15	6	0.00258900	0.00778221
15	7	0.00128690	0.00387584
15	8	0.00063588	0.00192266
15	9	0.00031043	0.00094607
15	10	0.00014782	0.00045777
15	11	0.00006676	0.00021362
15	12	0.00002670	0.00009155
15	13	0.00000763	0.00003052
15	14	0.	0.

n	k	$\overline{E}_F$	$E_{F[max]}$
16	1	0.08332697	0.24998855
16	2	0.04165967	0.12498664
16	3	0.02082602	0.06248569
16	4	0.01040919	0.03123521
16	5	0.00520078	0.01560997
16	6	0.00259658	0.00779735
16	7	0.00129448	0.00389104
16	8	0.00064344	0.00193789
16	9	0.00031793	0.00096131
16	10	0.00015521	0.00047302
16	11	0.00007391	0.00022888
16	12	0.00003338	0.00010681
16	13	0.00001335	0.00004578
16	14	0.00000381	0.00001526
16	15	0.	0.
17	1	0.08333015	0.24999427
17	2	0.04166316	0.12499332
17	3	0.02082967	0.06249284
17	4	0.01041293	0.03124260
17	5	0.00520456	0.01561748
17	6	0.00260037	0.00780492
17	7	0.00129828	0.00389864
17	8	0.00064723	0.00194550
17	9	0.00032172	0.00096893
17	10	0.00015897	0.00048065
17	11	0.00007761	0.00023651
17	12	0.00003695	0.00011444
17	13	0.00001669	0.00005341
17	14	0.00000668	0.00002289
17	15	0.00000191	0.00000763
17	16	0.	0.
18	1	0.08333174	0.24999713
18	2	0.04166491	0.12499665
18	3	0.02083150	0.06249642
18	4	0.01041479	0.03124630
18	5	0.00520644	0.01562124
18	6	0.00260226	0.00780871
18	7	0.00130018	0.00390244
18	8	0.00064913	0.00194931
18	9	0.00032362	0.00097275
18	10	0.00016086	0.00048447
18	11	0.00007948	0.00024033
18	12	0.00003880	0.00011826
18	13	0.00001848	0.00005722
18	14	0.00000834	0.00002670
18	15	0.00000334	0.00001144
18	16	0.00000095	0.00000381
18	17	0.	0.
19	1	0.08333253	0.24999856
19	2	0.04166579	0.12499832
19	3	0.02083241	0.06249820
19	4	0.01041573	0.03124814
19	5	0.00520738	0.01562312
19	6	0.00260322	0.00781060
19	7	0.00130113	0.00390434
19	8	0.00065009	0.00195122
19	9	0.00032456	0.00097456
19	10	0.00016180	0.00048637
19	11	0.00008043	0.00024223
19	12	0.00003974	0.00012016
19	13	0.00001940	0.00005913
19	14	0.00000924	0.00002361
19	15	0.00000417	0.00001335
19	16	0.00000167	0.00000572
19	17	0.00000048	0.00000191
19	18	0.	0.
20	1	0.08333293	0.24999928
20	2	0.04166622	0.12499916
20	3	0.02083287	0.06249910
20	4	0.01041619	0.03124907
20	5	0.00520786	0.01562405
20	6	0.00260369	0.00781155
20	7	0.00130160	0.00390530
20	8	0.00065056	0.00195217
20	9	0.00032504	0.00097561
20	10	0.00016228	0.00048733
20	11	0.00008090	0.00024319
20	12	0.00004021	0.00012112
20	13	0.00001987	0.00006008
20	14	0.00000970	0.00002955
20	15	0.00000462	0.00001431
20	16	0.00000209	0.00000658
20	17	0.00000083	0.00000286
20	18	0.00000024	0.00000095
20	19	0.	0.

NASA-Langley, 1968 —— 8

National Aeronautics and Space Administration
WASHINGTON, D. C.

OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

04U 001 33 51 3DS 68059 00903
AIR FORCE WEAPONS LABORATORY/AFWL/
KIRTLAND AIR FORCE BASE, NEW MEXICO 87117

ATT MISS MADELINE F. CANOVA, CHIEF TECHNI
LIBRARY /WLIL/

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546