Attention: Mr. William Alford

Subject: Preliminary Results of the Computer Simulation of Data Compression Using ERTS-A MSS Digital Tapes

Date: 3 January 1973

Contract: ERTS Image Data Compression Technique Evaluation

Principal Investigator: Dr. Donald J. Spencer, GSFC ID PR512

Prepared by: Dr. Curtis L. May
TRW Systems
One Space Park
Redondo Beach, California

1. INTRODUCTION

This report presents some preliminary results on the data compression simulation program prepared by TRW for the NASA contract "ERTS Image Data Compression Technique Evaluation." This report is intended to illustrate the typical computer output for each scene processed.

2. BACKGROUND

As specified in the TRW proposal, the computer program should be capable of generating various statistical characterizations of the MSS data and of the compression algorithms. These measures are:

- Data mean and variance in each spectral band and over all bands.
- First difference probability density functions (pdf) for each spectral band using the SSDI, SSDIA, and SSDIAM transforms.
- Joint spectral-spatial correlation along the scan lines.
- First difference joint probability ellipsoids.

(E73-10011) PRELIMINARY RESULTS OF THE COMPUTER SIMULATION OF DATA COMPRESSION USING ERTS-A MSS DIGITAL TAPES (TRW Systems Group) 43 p HC $4.25 CSCL 05B

N73-15347

Unclas
G3/13 00011

- Overall pdf for SSDI, SSDIA, SSDIAM and Shell symbols.
- Huffman codes for SSDI, SSDIA, SSDIAM, and Shell symbols.
- Scene entropy and average code lengths for SSDI, SSDIA, SSDIAM, and Shell transforms.
- Time-varying data compression for the scene.

These statistical measures will be computed for several 5 nmi x 5 nmi object classes and 25 nmi x 25 nmi scenes to be extracted from MSS tapes. In addition, compressed tapes will be generated and reconstructed tapes will be made for a few selected scenes.

3. RELEVANCE TO FUTURE WORK

The MSS tapes processed during the data analysis phase of the contract will produce similar computer output unless changes are made as a result of the Data Analysis Plan due 26 January 1973. As a result of the preliminary analysis several additional statistical measures such as line-to-line and column-to-column correlation as well as covariance matrices would be desirable, at least for selected scenes.

Using ERTS-A tape 1025-1S103, several subscenes have been processed with the TRW computer program. These scenes corresponded to segments of the image having varying degrees of data activity. The compressions achieved on these 5 nmi x 5 nmi subscenes produce average output bit rates varying from 1.8 bits per sample to 4.4 bits/per sample. The example case given in section 4 is an intermediate case requiring an output rate on the order of 3 bits per sample.

4. COMPUTER SIMULATION RESULTS

The master computer programs have been used to simulate the various compression algorithms and to compute the desired statistics on several segments of ERTS-A multispectral data. These scenes have been taken from the ERTS digital bulk MSS tape number 1025-15103 which covers the Lake St. John area in Quebec, Canada and includes the cities of Alma and Chicoutioni and the Saguenay River. The area is shown in Figure 4.1.

The various printout generated by the CDC-6500 computer is included in this section. The results will first be given and then the interpretation of these results will be presented and compared to results obtained on other sections of the scene. The output shown is from a 5 nmi x 5 nmi high detailed section of the scene centered at 49.2° and 71.1°longitude.

Figure 4.2 shows the cross spectral-spatial correlation σ_k of the input data as a function of distance along a scan line. This data is plotted in Figure 4.3. The correlation is formed by obtaining the normalized dot products of the intensity vectors $\underline{I_i}$ and $\underline{I_{i+k}}$, corresponding to pairs of intensity vectors separated along the scan line by k-1 intervening pixels. Normalization removes the effects of scene illumination and the closer the dot product is to unity (100%), the higher the correlation between intensity vectors. If the data is very active spectrally, the pair of vectors can be widely separated. The curves correspond to the percentages of vectors a distance k apart having normed dot products greater than σ_k, as averaged over the entire scene.

	CORRELATION				
STEPSIZE	1 PROB	2 PROB	4 PROB	6 PROB	8 PROB
ERCENT					
00.000	0.000	0.000	0.000	0.000	0.000
99.000	.666	.398	.267	.222	.191
98.000	.838	.553	.394	.340	.305
97.000	.915	.655	.479	.424	.384
96.000	.953	.731	.546	.489	.448
95.000	.973	.791	.601	.544	.507
94.000	.984	.837	.649	.592	.561
93.000	.991	.871	.691	.637	.611
92.000	.994	.898	.729	.676	.655
91.000	.996	.918	.760	.709	.692
90.000	.998	.935	.789	.740	.725
89.000	.999	.949	.816	.767	.754
88.000	.999	.960	.839	.792	.782
87.000	.999	.968	.861	.814	.806
86.000	1.000	.974	.878	.835	.827
85.000	1.000	.979	.893	.854	.847
84.000	1.000	.983	.907	.871	.865
83.000	1.000	.986	.918	.886	.880
2.000	1.000	.989	.928	.899	.894
1.000	1.000	.991	.938	.911	.906
80.000	1.000	.993	.946	.921	.918
79.000	1.000	.994	.952	.931	.927
78.000	1.000	.995	.958	.939	.935
77.000	1.000	.996	.963	.946	.942
76.000	1.000	.997	.968	.952	.949
75.000	1.000	.998	.972	.958	.955
74.000	1.000	.998	.976	.963	.961
73.000	1.000	.999	.979	.967	.966
72.000	1.000	.999	.982	.971	.970
71.000	1.000	.999	.985	.975	.974

Figure 4.2 Cross Spectral-Spatial Correlation of the Data

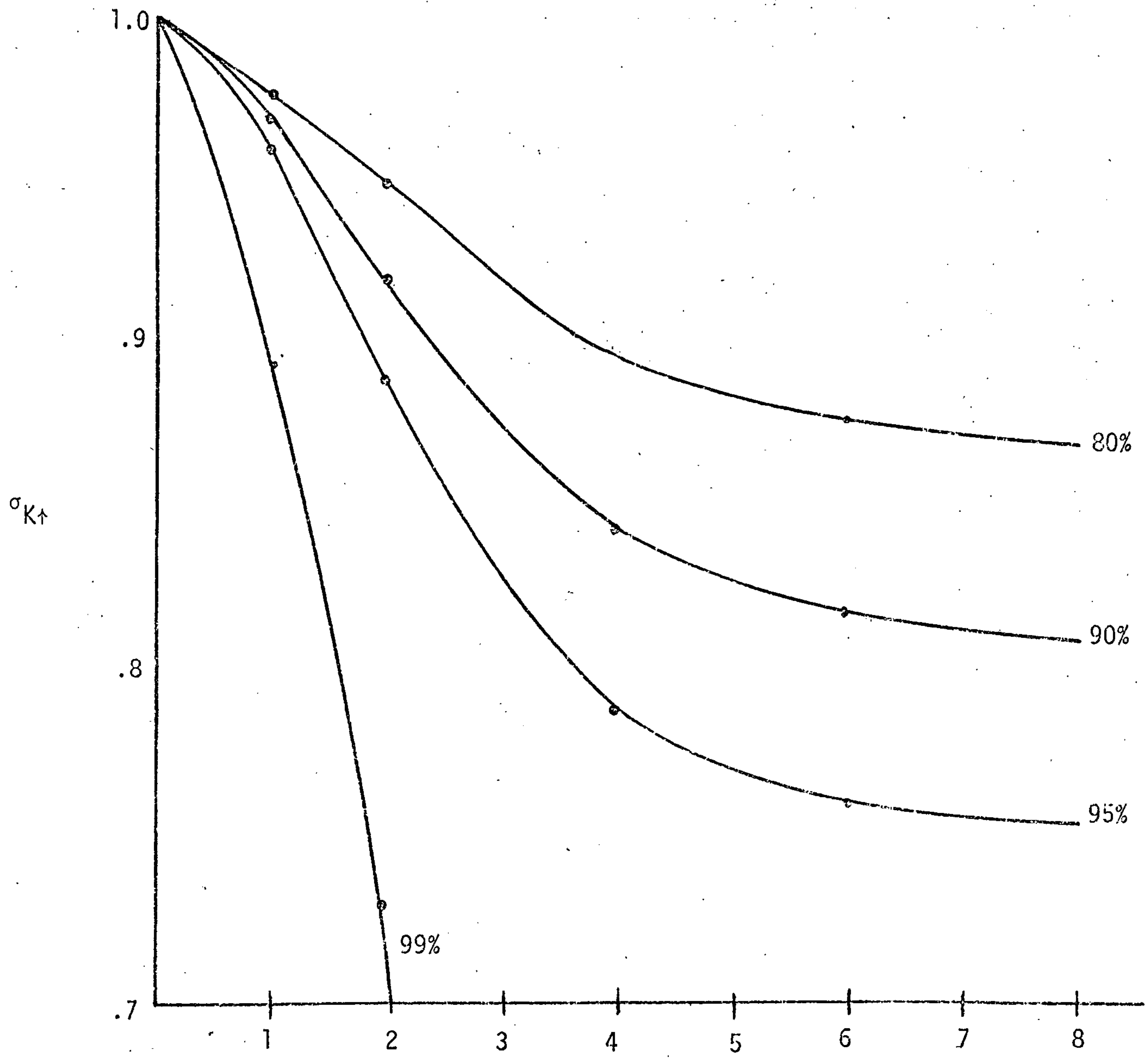


Figure 4.3 Plot of Correlation σ_K

The joint probability distribution function (pdf) of the first order differences (as obtained by the SSDI algorithm) over the scene is given in Figures 4.4 through 4.9. The set of output products spans all six possible pairs of bands. In each plot the joint occurrence of (0, 0) is normalized to 100 and this normalizing factor is used to multiply all other joint output occurrences. If a pair occurs less than one percent of the occurrence of (0, 0) it is not displayed in order to simplify the figures.

Figure 4.10 shows the mean and variance of each band as well as the overall mean and variance of the scene. Figure 4.11 gives the pdf of the first differences as obtained by the SSDI algorithm. Only difference levels in the range [-18, 18] are given since levels beyond these normally occur far less than one per cent of the time. Figures 4.12 and 4.13 give the first difference pdf as obtained by the SSDIA and SSDIAM algorithms. Note that the SSDIA first differences have a much smaller variance than those obtained by the SSDI and the one percent occurence cuts off at a lower level. The SSDIAM further decreases the variance but increases the probability of +1 and -1 due to the one bit mapping of this algorithm. Again, the SSDIAM produces a one percent cutoff at a lower difference level than the SSDIA. In general, compression increases as the variance decreases.

The pdf of the SSDI, SSDIA, and SSDIAM symbols are given in Figure 4.14. As in Figures 4.11 through 4.13, an improvement can be observed in the

Figure 4.4 JOINT PROBABILITY ELLIPSE - BAND 1 .VS. BAND 2

```
    -16-15-14-13-12-11-10 -9 -8 -7 -6 -5 -4 -3 -2 -1  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
-16
-15
-14
-13
-12
-11
-10
 -9
 -8
 -7
 -6                                                   1
 -5                                                   2
 -4                                                1  3
 -3                                             1  2  5  2  1
 -2                                             2  3  9  3  2
 -1                                             2  7 19  7  3  2
  0                     1        1  1     2  3  9 23100 25 11  4  3  2  1  1     1  1
  1                                          2  3  7 17  6  3  1
  2                                          1  2  3  7  3  2
  3                                             1  1  4  1
  4                                                   2
  5                                                   1
  6
  7
  8
  9
 10
 11
 12
 13
 14
 15
 16
```

Figure 4.5 JOINT PROBABILITY ELLIPSE - BAND 1 .VS. BAND 3

	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-16																																	
-15																																	
-14																	1																
-13																	2																
-12																	1																
-11																	2																
-10																	2																
-9																	1																
-8																	1																
-7																	2																
-6																	2																
-5																	2																
-4													1	1	1	2	4		2														
-3													1	2	2	2	5	2	2	1		1											
-2											1	1	2	2	4	3	9	3	4	2	2	2	1										
-1								1		1	2	2	3	2	8	6	16	6	7	4	3	2	2	1	1	1	1						
0	2	2	2	2	3	4	3	5	5	5	10	8	11	14	25	17	100	17	26	13	12	9	9	5	6	4	3	4	3	3	2	2	2
1									2	1	2	2	4	3	8	5	16	3	8	3	4	2	2	2	1	1	1						
2										1	1	1	2	2	4	2	8	2	3	2	2	1	1										
3											1		1	1	3		5		2	1	1												
4															1		4		1														
5																	2																
6																	2																
7																	3																
8																	2																
9																	2																
10																	2																
11																	2																
12																	2																
13																	2																
14																	2																
15																	1																
16																	1																

Figure 4.6

JOINT PROBABILITY ELLIPSE - BAND 1 .VS. BAND 4

	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-16																	1																
-15																	1																
-14																2	2																
-13																1	1																
-12																	1																
-11																1	2																
-10															1	1	2	2															
-9																1	2																
-8																	2	1															
-7															1	2	4	1															
-6															2	1	3	1															
-5													1			2	3	1	1	1													
-4												1	1	1	3	4	5	4	2	1	1	1											
-3											1	1	1	2	3	6	7	5	3	2	2	1		1									
-2										1		1	3	3	5	8	12	8	4	3	2	2	1										
-1											2	2	4	5	9	15	20	15	8	6	4	3	2	2	1	2							
0	1	2	2	2	3	4	5	6	7	9	11	14	19	24	34	57	100	54	34	25	18	14	11	9	7	5	4	3	3	2	2	1	1
1									1	1	2	2	3	6	8	15	19	14	8	6	4	4	3	2	1	1							
2										1		2	2	3	4	7	10	6	4	4	3	2	1	2									
3											1	1	2	2	3	5	7	4	3	2	2	1	1										
4													1	2	2	2	5	3	1	1		1											
5															1	2	3	2															
6															1	2	3	1															
7														1		1	3																
8															1		2																
9																	2																
10																	2																
11																	2																
12																	2																
13																	2																
14																	2																
15																	1																
16																	1																

Figure 4.7 JOINT PROBABILITY ELLIPSE - BAND 2 .VS. BAND 3

	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-16																																	
-15																																	
-14																																	
-13																																	
-12																																	
-11																																	
-10																	1																
-9																																	
-8																	1																
-7																	1																
-6																	1																
-5															1		2																
-4															2		3		1														
-3											1				1	2	3	1	2														
-2											1	1	1	2	4	3	7	2	3	2	1	1	1										
-1									1	1	2	2	4	3	8	7	18	5	8	3	4	3	2	1	1	1		1					
0	1	2	1	1	3	2	2	3	3	3	7	5	8	11	20	15	100	16	20	10	10	6	7	4	4	3	2	2	2	2	1	1	
1										1	2	2	4	3	8	4	16	5	8	3	4	3	2	1	2	1							
2												1	2	1	4	2	7	1	4	2	1	1	1										
3															1		3		2	1													
4															1		2																
5																	2																
6																	2																
7																	1																
8																																	
9																																	
10																	1																
11																																	
12																																	
13																																	
14																																	
15																																	
16																																	

Figure 4.8 JOINT PROBABILITY ELLIPSE - BAND 2 .VS. BAND 4

	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-16																																	
-15																	1																
-14																																	
-13																	1																
-12															1	1	1	1															
-11																	2																
-10																1	2	1															
-9																	1																
-8																2	2	1															
-7																2	2	2															
-6															1	2	2	2	1														
-5														1	2	2	3	2	1														
-4														1	2	3	4	3	1	1													
-3											1		1	1	3	5	6	4	2	2	1												
-2											1	1	1	3	5	8	11	8	5	3	2	2	1										
-1										1	2	3	4	6	9	18	25	17	10	7	6	4	3	2	2	1							
0	2	2	2	3	4	4	5	5	7	8	10	12	16	19	31	50	100	47	28	20	14	11	9	7	5	5	3	2	1	2	1	1	
1										1	2	3	4	6	10	15	21	16	9	6	5	4	3	2	1	2	1						
2										1	1	2	2	4	4	7	10	8	4	4	2	2	1	1	1								
3												1	2	2	2	4	5	2	2	2	1												
4													1	1	2	2	4	2	1														
5																2	2	1	1	1													
6															1	2	2																
7																	2																
8																	2																
9															1		1																
10																	2	1															
11																																	
12																																	
13																	1																
14																	1																
15																																	
16																	1																

Figure 4.9 JOINT PROBABILITY ELLIPSE - BAND 3 .VS. BAND 4

	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-16																																	
-15																1	1																
-14															1	1	1																
-13															1	1	2																
-12															1	2	1	1															
-11														1		2	1	2		1													
-10															1	2	2	1															
-9													1	2	2	2	3	2	1			1											
-8															2	3	4	3	2	1													
-7														1	2	3	3	2	2	1		1											
-6												1	2	2	3	5	5	4	3	2	1	2											
-5												1	2	2	3	5	5	5	2	2	2	1											
-4													2	2	3	6	8	6	4	4	2	2											
-3											1	2	2	3	4	8	8	7	4	3	2	2	2										
-2										1	1	2	3	5	6	13	19	15	9	6	4	3	2	2	1								
-1														1	3	10	18	12	4	3	2	1	1	1									
0	2	3	2	2	4	4	4	5	5	6	9	8	10	12	19	31	100	31	17	12	9	6	4	5	3	2	1						
1														2	3	9	15	8	4	2	2	2	1	1									
2											2	3	3	5	9	13	15	10	7	5	4	2	2	2		1							
3											1	1	2	3	4	7	7	5	3	3	2	2	1	1									
4											1	2	2	3	4	6	6	4	3	2	1	1	1										
5												1	1	3	4	3	4	3	2	2	2												
6										1	1	1	3	2	3	3	5	3	3	2	2	1											
7											1	1	1	2	2	2	3	2	2	1													
8													1	2	2	2	2	2	2	1	1												
9											1			1	1	2	3	2	1	1													
10																1	1	2	1														
11												1		1	1	1	1	1															
12														1		1	2	1															
13																1	1																
14																																	
15																	1																
16																																	

```
         BAND1    BAND2    BAND3    BAND4    TOTAL

 MU      49.439   41.653   51.162   24.826   41.770

SIGMA  1438.246 1731.235 1224.019  277.459 1276.277
```

Figure 4.10 Mean and Variance of the Data

Figure 4.11 FIRST DIFFERENCE PROBABILITY - SSDI

LEVEL	PROB. BAND1	PROB. BAND2	PROB. BAND3	PROB. BAND4
-18	.00	.00	.00	.00
-17	.00	.00	.00	.00
-16	.00	.00	.00	.00
-15	.01	.01	.01	.00
-14	.01	.01	.01	.00
-13	.01	.01	.01	.00
-12	.01	.01	.01	.00
-11	.01	.01	.01	.01
-10	.01	.01	.01	.01
-9	.01	.01	.01	.01
-8	.01	.01	.01	.01
-7	.01	.01	.02	.01
-6	.01	.01	.02	.02
-5	.01	.01	.02	.03
-4	.02	.02	.03	.03
-3	.02	.02	.04	.04
-2	.04	.03	.05	.06
-1	.07	.06	.07	.09
0	.43	.43	.25	.34
1	.07	.06	.07	.09
2	.03	.03	.05	.06
3	.02	.02	.04	.04
4	.02	.02	.03	.03
5	.01	.01	.02	.03
6	.01	.01	.02	.02
7	.01	.01	.02	.01
8	.01	.01	.01	.01
9	.01	.01	.01	.01
10	.01	.01	.01	.01
11	.01	.01	.01	.01
12	.01	.01	.01	.00
13	.01	.01	.01	.00
14	.01	.01	.01	.00
15	.00	.00	.01	.00
16	.00	.00	.00	.00
17	.00	.00	.00	.00
18	.00	.00	.00	.00

Figure 4-12 FIRST DIFFERENCE PROBABILITY - SSDIA

LEVEL	PROB. BAND1	PROB. BAND2	PROB. BAND3	PROB. BAND4
-18	.00	.00	.00	0.00
-17	.00	.00	.00	0.00
-16	.00	.00	.00	0.00
-15	.00	.00	.00	0.00
-14	.00	.00	.00	0.00
-13	.00	.00	.00	0.00
-12	.00	.00	.00	0.00
-11	.00	.00	.00	0.00
-10	.00	.00	.00	.00
-9	.01	.01	.01	.00
-8	.01	.01	.01	.00
-7	.01	.01	.01	.00
-6	.01	.01	.01	.00
-5	.01	.01	.02	.01
-4	.02	.02	.02	.01
-3	.02	.02	.04	.03
-2	.03	.03	.06	.05
-1	.07	.06	.10	.11
0	.59	.58	.42	.56
1	.06	.06	.10	.11
2	.03	.03	.06	.06
3	.02	.02	.04	.03
4	.02	.02	.03	.01
5	.02	.02	.02	.01
6	.01	.01	.01	.00
7	.01	.01	.01	.00
8	.01	.01	.01	.00
9	.01	.01	.01	.00
10	.00	.01	.00	.00
11	.00	.00	.00	0.00
12	.00	.00	.00	0.00
13	.00	.00	.00	0.00
14	.00	.00	.00	0.00
15	.00	.00	.00	0.00
16	.00	.00	.00	0.00
17	.00	.00	.00	0.00
18	.00	.00	.00	0.00

Figure 4.13 FIRST DIFFERENCE PROBABILITY - SSDIAM

LEVEL	PROB. BAND1	PROB. BAND2	PROB. BAND3	PROB. BAND4
-18	0.00	0.00	0.00	0.00
-17	0.00	.00	0.00	0.00
-16	0.00	.00	0.00	0.00
-15	0.00	0.00	0.00	0.00
-14	0.00	.00	0.00	0.00
-13	0.00	.00	0.00	0.00
-12	.00	.00	.00	0.00
-11	.00	.00	.00	0.00
-10	.00	.00	.00	0.00
-9	.00	.00	.00	0.00
-8	.00	.00	.00	0.00
-7	.00	.01	.00	0.00
-6	.01	.01	.01	.00
-5	.01	.01	.01	.00
-4	.02	.02	.02	.00
-3	.02	.03	.03	.01
-2	.04	.04	.06	.04
-1	.12	.12	.17	.18
0	.54	.53	.40	.52
1	.11	.11	.17	.19
2	.04	.04	.06	.05
3	.03	.03	.03	.01
4	.02	.02	.02	.00
5	.01	.01	.01	.00
6	.01	.01	.01	.00
7	.00	.01	.00	0.00
8	.00	.00	.00	0.00
9	.00	.00	.00	0.00
10	.00	.00	.00	0.00
11	.00	.00	.00	0.00
12	.00	.00	.00	0.00
13	0.00	.00	0.00	0.00
14	0.00	.00	0.00	0.00
15	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00
18	0.00	0.00	0.00	0.00

	OVERALL PROBABILITY		
LEVEL	SSDI	SSDIA	SSDIAM
-18	.00	.00	0.00
-17	.00	.00	.00
-16	.00	.00	.00
-15	.00	.00	0.00
-14	.01	.00	.00
-13	.01	.00	.00
-12	.01	.00	.00
-11	.01	.00	.00
-10	.01	.00	.00
-9	.01	.00	.00
-8	.01	.01	.00
-7	.01	.01	.00
-6	.02	.01	.00
-5	.02	.01	.01
-4	.02	.02	.01
-3	.03	.03	.02
-2	.05	.04	.05
-1	.07	.09	.15
0	.36	.54	.49
1	.07	.08	.15
2	.04	.04	.05
3	.03	.03	.03
4	.02	.02	.01
5	.02	.01	.01
6	.01	.01	.00
7	.01	.01	.00
8	.01	.01	.00
9	.01	.00	.00
10	.01	.00	.00
11	.01	.00	.00
12	.01	.00	.00
13	.01	.00	.00
14	.00	.00	.00
15	.00	.00	0.00
16	.00	.00	0.00
17	.00	.00	0.00
18	.00	.00	0.00

Figure 4.14 Probability Distributions of the SSDI, SSDIA, and SSDIAM Symbols

SHELL LEVEL	PROBABILITY
1	.057
2	.094
3	.088
4	.075
5	.063
6	.056
7	.046
8	.038
9	.035
10	.030
11	.029
12	.027
13	.025
14	.025
15	.021
16	.021
17	.020
18	.018
19	.017
20	.017
21	.016
22	.016
23	.014
24	.014
25	.012

Figure 4.15 Probability Distribution of Shell Locations

```
UFFMAN CODES FOR SHELL

AVERAGE CODE LENGTH  3.652

LEVEL  PROB.     LENGTH

  1    .057        4     0110
  2    .094        3     001
  3    .088        3     000
  4    .075        4     0101
  5    .063        4     0100
  6    .056        4     1000
  7    .046        4     0111
  8    .038        5     11001
  9    .035        5     11000
 10    .030        5     10111
 11    .029        5     10110
 12    .027        5     10101
 13    .025        5     10100
 14    .025        5     10011
 15    .021        5     10010
 16    .021        5     11011
 17    .020        5     11010
 18    .018        6     111111
 19    .017        6     111110
 20    .017        6     111101
 21    .016        6     111100
 22    .016        6     111011
 23    .014        6     111010
 24    .014        6     111001
 25    .012        6     111000
```

Figure 4.16 Shell Huffman Code

HUFFMAN CODES FOR SSDI

AVERAGE CODE LENGTH 4.193 ENTROPY 4.158

LEVEL	PROB.	LENGTH	
-20	.003	8	11111111
-19	.003	8	11111110
-18	.003	8	11111101
-17	.004	8	11111100
-16	.004	8	11111011
-15	.004	7	1110010
-14	.005	7	1110001
-13	.006	7	1110000
-12	.006	7	1111011
-11	.007	7	1111010
-10	.008	7	1111001
-9	.010	6	110000
-8	.011	6	110111
-7	.012	6	110110
-6	.016	6	110101
-5	.020	5	10010
-4	.025	5	10111
-3	.032	5	10110
-2	.046	4	1000
-1	.075	4	0111
0	.364	2	00
1	.073	4	0110
2	.044	4	0101
3	.031	5	10101
4	.023	5	10100
5	.019	5	10011
6	.014	6	110100
7	.012	6	110011
8	.011	6	110010
9	.009	6	110001
10	.008	7	1111000
11	.007	7	1110111
12	.006	7	1110110
13	.005	7	1110101
14	.005	7	1110100
15	.004	7	1110011
16	.004	8	11111010
17	.003	8	11111001
18	.003	8	11111000
19	.003	4*	0100
20	.003	4*	0100

*GROUPED PROBABILITY = .055 CODE LENGTH = 4

Figure 4.17 SSDI Huffman Code

HUFFMAN
[redacted] CODES FOR SSDIA

AVERAGE CODE LENGTH 2.886   ENTROPY 2.832

LEVEL	PROB.	LENGTH	
-20	.000	6*	110110
-19	.000	6*	110110
-18	.000	6*	110110
-17	.001	6*	110110
-16	.001	6*	110110
-15	.001	6*	110110
-14	.002	6*	110110
-13	.002	6*	110110
-12	.002	6*	110110
-11	.003	8	11101011
-10	.003	8	11101010
-9	.004	8	11101001
-8	.006	7	1110000
-7	.008	7	1101111
-6	.009	7	1101110
-5	.014	6	111011
-4	.018	6	111010
-3	.027	5	11001
-2	.044	5	11000
-1	.085	3	100
0	.537	1	0
1	.081	4	1011
2	.045	4	1010
3	.029	5	11010
4	.019	6	111001
5	.014	6	111000
6	.010	6	110111
7	.008	7	1110010
8	.006	7	1110001
9	.005	8	11101000
10	.004	8	11100111
11	.003	8	11100110
12	.002	6*	110110
13	.002	6*	110110
14	.001	6*	110110
15	.001	6*	110110
16	.001	6*	110110
17	.001	6*	110110
18	.000	6*	110110
19	.000	6*	110110
20	.000	6*	110110

*GROUPED PROBABILITY = .018 CODE LENGTH = 6

Figure 4.18 SSDIA Huffman Code

HUFFMAN CODES FOR SSDIAM

AVERAGE CODE LENGTH 2.542 ENTROPY 2.484

LEVEL	PROB.	LENGTH	
-20	0.000	6*	111100
-19	0.000	6*	111100
-18	0.000	6*	111100
-17	.000	6*	111100
-16	.000	6*	111100
-15	0.000	6*	111100
-14	.000	6*	111100
-13	.000	6*	111100
-12	.000	6*	111100
-11	.000	6*	111100
-10	.001	6*	111100
-9	.001	6*	111100
-8	.002	6*	111100
-7	.003	6*	111100
-6	.005	8	11110111
-5	.008	8	11110110
-4	.013	6	111110
-3	.023	5	11100
-2	.047	4	1101
-1	.148	3	100
0	.494	1	0
1	.145	3	101
2	.049	4	1100
3	.025	5	11101
4	.014	6	111101
5	.009	7	1111010
6	.005	6*	111100
7	.003	6*	111100
8	.002	6*	111100
9	.001	6*	111100
10	.000	6*	111100
11	.000	6*	111100
12	.000	6*	111100
13	.000	6*	111100
14	.000	6*	111100
15	0.000	6*	111100
16	0.000	6*	111100
17	0.000	6*	111100
18	0.000	6*	111100
19	0.000	6*	111100
20	.000	6*	111100

*GROUPED PROBABILITY = .019 CODE LENGTH = 6

Figure 4.19 SSDIAM Huffman Code

distribution as we progress from SSDI to SSDIAM, implying increasing compression. Figure 4.15 shows the probability of shell locations for the given scene. Level 1 implies that all SSDI symbols are simultaneously zero for a pixel. Level 2 implies that the greatest symbol magnitude is 1 for a pixel. The probability distribution resembles a χ^2 distribution with peak at level 2 and a slow fall off of the tail.

Figures 4.16 through 4.19 give the Huffman codes for the scene using SHELL, SSDI, SSDIA, and SSDIAM transforms. Parameters listed are the symbol or shell level, its probability, the length of the symbol code word in bits, and the actual binary code word assigned to the symbol. To conserve space, only symbol levels between -20 and +20 are given but code words are assigned to all symbols. The least probable symbols can be grouped together under a lumped Huffman prefix code word as described in Appendix B. Such is the case for the SSDI, SSDIA, and the SSDIAM codes. All grouped code words displayed are given an asterisk following the symbol length and the lumped prefix code is given. Following the Huffman code, the total probability of the grouped symbols is given as well as the prefix code length in bits. The entropy of the symbol distribution is also displayed.

Figure 4.20 shows several time varying statistics of the coding techniques. The buffering statistics and the average bits per sample are given for each scan line of 180 pixels in each spectral band. Figure 4.20 only gives the first 119 scan lines of data using SSDIA symbols. If desired, the statistics can be presented for the SSDI or the SSDIAM symbols.

Figure 4.20 permits a comparison of the line by line average bit rate for the global Huffman, the adaptive Huffman, and the Rice encoding algorithms. For the data used, the average number of bits per sample varies from 2.214 bits to 3.905 bits for global Huffman coding, from 2.105 bits to 3.912 bits for adaptive Huffman coding, and from 2.249 bits to 4.321 bits for Rice encoding. In general, the average bits/sample varies rather slowly from line to line, following trends in the source data activity. Figure 4.21 presents the average data compression achieved over the scene by the adaptive coding techniques. The adaptive Huffman coding achieves a lower average bit rate than the Rice coding because of the necessary overhead which must be transmitted with Rice encoding. In addition, statistics are given concerning the occurrence of the various Rice modes. For the given data, the fundamental sequence (FS) was transmitted 31.7 percent of the time, the coded fundamental sequence (FSC) was transmitted 57.5 percent of the time, and the complemented fundamental sequence (FSCB) was used 10.8 percent of the time. For this data the split-pixel modes (6, 1), (4, 3) and (3, 4) did not occur.

SSDIA COMPRESSION STATISTICS
CASE ID: NW 180X180

OUTPUT BITS/PIXEL = 7

SCAN LINE	BITS IN BUF.	HUF. (BITS/PIX)	ADAPT. HUF. (BITS/PIX)	RICE (BITS/PIX)
1	0	2.557	2.503	2.571
2	0	2.426	2.313	2.581
3	0	2.541	2.426	2.655
4	0	2.558	2.491	2.634
5	0	2.274	2.213	2.234
6	0	2.379	2.230	2.443
7	0	2.844	2.638	2.952
8	0	2.719	2.631	3.051
9	0	2.675	2.690	3.017
10	0	2.842	2.797	3.143
11	0	2.987	2.959	3.352
12	0	2.982	2.935	3.365
13	0	2.801	2.804	2.963
14	0	2.808	2.763	3.011
15	0	2.750	2.702	2.837
16	0	2.714	2.612	2.641
17	0	2.761	2.668	3.031
18	0	3.028	2.925	3.322
19	0	2.946	2.889	3.259
20	0	2.804	2.760	3.075
21	0	2.827	2.773	2.949
22	0	3.518	3.450	3.648
23	0	3.428	3.385	3.534
24	0	3.601	3.547	3.753
25	0	3.621	3.536	3.832
26	0	3.193	3.128	3.430
27	0	3.107	3.072	3.419
28	0	3.014	2.972	3.297
29	0	3.359	3.375	3.476
30	0	3.534	3.443	3.726
31	0	3.808	3.727	3.889
32	0	3.550	3.564	3.707
33	0	3.692	3.678	4.033
34	0	3.905	3.912	4.321
35	0	3.636	3.740	3.928

Figure 4.20 Time-Varying Statistics of the Global Huffman, Adaptive Huffman and Rice Codes (SSDIA Transform)

Figure 20 (Continued)

36	0	3.416	3.534	3.547
37	0	2.930	2.925	2.956
38	0	3.085	2.952	3.291
39	0	3.268	3.266	3.474
40	0	3.196	3.151	3.253
41	0	3.188	3.128	3.330
42	0	3.114	3.055	3.332
43	0	2.976	2.969	3.142
44	0	2.727	2.703	3.000
45	0	3.034	3.033	3.233
46	0	3.121	3.057	3.216
47	0	2.990	2.946	3.182
48	0	2.794	2.754	2.834
49	0	2.858	2.780	2.841
50	0	2.847	2.739	2.923
51	0	3.190	3.089	3.270
52	0	3.063	3.026	3.053
53	0	2.801	2.793	2.821
54	0	2.621	2.514	2.634
55	0	2.798	2.656	2.838
56	0	2.844	2.699	2.952
57	0	2.862	2.811	2.945
58	0	2.849	2.780	2.935
59	0	2.658	2.568	2.847
60	0	2.548	2.460	2.763
61	0	2.389	2.349	2.602
62	0	2.489	2.392	2.626
63	0	2.393	2.315	2.533
64	0	2.227	2.112	2.271
65	0	2.399	2.202	2.372
66	0	2.482	2.349	2.453
67	0	2.473	2.322	2.499
68	0	2.656	2.510	2.774
69	0	2.773	2.726	2.753
70	0	2.702	2.636	2.744
71	0	2.511	2.466	2.469
72	0	2.597	2.509	2.548
73	0	2.649	2.541	2.641
74	0	2.690	2.625	2.665
75	0	2.920	2.886	2.922
76	0	2.760	2.669	2.848
77	0	3.097	2.953	3.162

78	0	3.223	3.098	3.222
79	0	3.240	3.084	3.337
80	0	3.303	3.229	3.483
81	0	3.074	3.047	3.378
82	0	3.142	3.109	3.388
83	0	3.597	3.565	3.831
84	0	3.351	3.327	3.540
85	0	3.276	3.246	3.568
86	0	3.078	3.082	3.359
87	0	3.274	3.318	3.422
88	0	2.898	2.815	2.964
89	0	2.817	2.705	2.871
90	0	2.773	2.746	2.895
91	0	2.619	2.540	2.624
92	0	2.268	2.129	2.222
93	0	2.318	2.145	2.408
94	0	2.669	2.442	2.717
95	0	2.770	2.629	2.864
96	0	2.784	2.683	2.932
97	0	2.861	2.808	3.028
98	0	2.952	2.891	3.124
99	0	3.401	3.446	3.607
100	0	3.298	3.276	3.455
101	0	3.153	3.136	3.237
102	0	3.134	3.034	3.250
103	0	2.989	2.859	3.102
104	0	2.793	2.680	2.811
105	0	2.888	2.747	2.953
106	0	2.747	2.682	2.824
107	0	2.673	2.611	2.781
108	0	2.551	2.516	2.614
109	0	2.497	2.496	2.385
110	0	2.214	2.105	2.249
111	0	2.359	2.222	2.524
112	0	2.375	2.288	2.470
113	0	2.426	2.278	2.622
114	0	2.459	2.438	2.594
115	0	2.541	2.436	2.680
116	0	2.804	2.753	3.099
117	0	2.831	2.776	3.020
118	0	2.891	2.828	3.105
119	0	2.759	2.739	3.020

Figure 20 (Continued)

```
OVERALL DATA COMPRESSION (BITS/SAMPLE)

  ADAPTIVE HUFFMAN =  2.8  RICE =  3.0

PERCENT OCCORRENCE OF RICE MODES

  FS   FSC  FSCB SP71 SP53 SP44
 31.7 57.5 10.8  0.0  0.0  0.0
```

Figure 4.21 Average Data Compressions for the Scene and Rice Mode Statistics

A sample of the output table from program BLDTAB is given in Figure 4.24 This decoding table is of length 2^{12} and each entry gives the appropriate SSDI symbol and the number of shifts required to reposition the compressed bit stream for the next decoding operation. The beginning segment of the table gives symbols included under the lumped prefix. This lumped prefix has four bits. The following eight bits separates the lumped symbols. As shown the overall code word is of length twelve bits so that twelve shifts would be required in decoding. The other decodable words in this segment of the decoding table are -2 and +2, each of length four bits.

Figures 4.22 and 4.23 show the same segment of data from the input scene. Figure 4.22 gives the source digital data values in the first spectral band and Figure 4.23 gives the reconstructed digital data for the same band. Since the simulation is strictly information preserving, no errors have occurred in the data.

127	127	94	36	23	29	20	16	16	16	16	19	20	19	17	16	16	16	14	14	14	16	16	14
127	127	86	35	18	21	15	12	15	15	16	16	18	24	20	15	15	16	14	15	15	15	16	15
80	57	26	21	20	15	14	12	11	12	12	12	12	15	15	14	16	18	16	16	15	15	16	18
10	14	21	20	12	12	7	1	0	0	0	0	3	3	1	3	4	6	4	4	6	7	6	6
49	51	65	83	65	41	20	14	13	12	12	13	13	15	15	15	15	15	15	15	16	16	18	18
65	69	105	124	102	69	37	19	12	12	12	12	14	14	15	15	16	16	16	18	16	16	18	16
53	72	106	127	127	127	100	49	13	12	12	11	12	13	13	14	14	14	14	17	16	14	16	16
68	103	127	127	127	127	113	62	22	14	11	13	11	11	13	14	15	15	16	15	15	15	16	15
79	116	127	127	127	122	127	112	67	28	13	11	11	11	13	14	15	16	16	15	15	15	15	16
127	127	127	127	127	127	127	127	73	16	0	0	0	0	0	1	3	3	3	1	3	3	5	7
92	120	123	125	120	125	127	104	57	21	14	13	13	14	14	15	16	15	15	15	15	15	15	16
45	69	81	105	99	109	118	84	37	18	15	13	13	13	14	15	15	14	15	15	15	14	15	15
17	39	81	96	94	96	100	87	41	16	16	17	14	12	13	13	13	14	13	13	13	14	14	13
15	37	70	86	93	97	77	35	16	16	18	16	16	15	14	13	14	14	13	13	14	14	16	16
11	20	31	36	28	21	18	15	14	15	20	18	21	18	14	13	14	14	14	14	15	14	15	18
1	7	3	1	3	1	0	5	18	53	69	34	13	7	5	3	3	3	1	3	3	1	3	13
14	13	14	16	16	16	16	23	51	63	74	71	57	58	47	20	15	16	15	15	14	15	15	21
15	14	15	16	16	16	15	21	37	49	77	90	99	94	58	23	18	16	15	16	16	15	15	16
14	14	14	16	14	16	16	14	16	23	49	77	84	77	81	69	65	57	29	17	17	16	14	14
14	14	14	15	14	15	14	14	15	27	59	70	52	50	73	97	110	113	97	62	27	18	16	16
15	15	15	15	14	15	15	14	19	43	74	74	61	77	105	127	127	125	96	65	28	16	15	15
4	4	5	4	4	5	8	4	8	23	59	65	65	76	100	115	127	127	72	27	15	9	5	4
15	14	14	14	15	15	15	15	15	19	28	33	33	31	47	89	107	92	62	33	21	23	18	14
15	14	14	14	15	14	15	15	14	15	15	18	16	14	21	45	68	68	33	19	25	31	19	15
14	13	14	16	14	13	14	14	14	14	14	16	13	11	12	23	36	46	36	22	20	24	30	26
14	15	15	15	14	14	14	15	14	14	14	13	11	10	10	13	15	22	24	21	18	20	27	24
16	14	14	14	14	14	14	15	15	15	14	14	13	13	14	13	13	15	16	15	15	22	25	16
5	5	5	5	4	4	5	4	5	4	4	5	4	8	5	4	4	4	5	4	5	8	9	9
30	28	16	15	15	16	15	16	16	15	15	14	15	18	15	15	15	15	15	16	15	16	15	16
63	35	15	14	14	14	14	16	15	15	15	15	15	15	14	14	14	15	16	15	16	15	14	15
96	44	17	14	12	12	13	14	14	17	17	14	14	16	16	13	13	13	14	14	14	14	14	13
86	44	27	20	13	10	12	10	12	13	13	14	15	14	15	14	13	14	14	14	15	15	15	15
36	28	25	16	12	13	12	10	12	9	10	12	13	15	15	15	15	15	15	15	15	14	15	15
13	16	21	13	4	3	3	1	1	1	1	3	3	4	5	5	7	7	7	7	5	5	9	9
27	44	64	51	23	12	12	12	11	11	12	12	12	12	13	12	13	15	15	16	16	15	14	14
15	15	12	12	10	10	10	10	10	9	9	10	10	10	11	10	10	10	11	12	15	14	11	10
11	11	11	10	10	10	10	11	12	10	10	10	10	11	11	12	12	10	11	11	11	10	10	10
10	10	10	10	10	9	9	10	10	10	10	9	10	9	10	10	10	10	12	13	10	9	10	10
10	9	10	10	10	10	12	10	10	12	10	10	9	10	9	12	13	13	13	14	13	10	9	10
4	1	1	1	1	1	1	1	0	1	3	3	1	0	3	4	4	7	11	9	3	1	1	1
16	14	13	13	11	11	11	11	11	11	12	13	11	12	14	16	18	16	16	16	15	13	15	15
18	18	15	15	14	12	11	12	11	11	12	15	14	14	15	16	21	25	31	51	57	59	71	57

Figure 4.22 A Segment of Input Data from Scene (Band 2)

127	127	94	36	23	29	20	16	16	16	16	19	20	19	17	16	16	16	14	14	14	16	16	14
127	127	86	35	18	21	15	12	15	15	16	16	18	24	20	15	15	16	14	15	15	15	16	15
80	57	26	21	20	15	14	12	11	12	12	12	12	15	15	14	16	18	16	16	15	15	16	18
10	14	21	20	12	12	7	1	0	0	0	0	3	3	1	3	4	6	4	4	6	7	6	6
49	51	65	83	65	41	20	14	13	12	12	13	13	15	15	15	15	15	15	15	16	16	18	18
65	69	105	124	102	69	37	19	12	12	12	12	14	14	15	15	16	16	16	18	16	16	18	16
53	72	106	127	127	127	100	49	13	12	12	11	12	13	13	14	14	14	14	17	16	14	16	16
68	103	127	127	127	127	113	62	22	14	11	13	11	11	13	14	15	15	16	15	15	15	16	15
79	116	127	127	127	127	127	112	67	28	13	11	11	11	13	14	15	16	16	15	15	15	15	16
127	127	127	127	127	127	127	127	73	16	0	0	0	0	0	1	3	3	3	1	3	3	5	7
92	120	123	125	120	125	127	104	57	21	14	13	13	14	14	15	16	15	15	15	15	15	15	16
45	69	81	105	99	109	118	84	37	18	15	13	13	13	14	15	15	14	15	15	15	14	15	15
17	39	81	96	94	96	100	87	41	16	16	17	14	12	13	13	13	14	13	13	13	14	14	13
15	37	70	86	93	97	77	35	16	16	18	16	16	15	14	13	14	14	13	13	14	14	16	16
11	20	31	36	28	21	18	15	14	15	20	18	21	18	14	13	14	14	14	14	15	14	15	18
1	7	3	1	3	1	0	5	18	53	69	34	13	7	5	3	6	3	1	3	3	1	3	13
14	13	14	16	16	16	16	23	51	63	74	71	57	58	47	20	15	16	15	15	14	15	15	21
15	14	15	16	16	16	15	21	37	49	77	90	99	94	58	23	16	16	15	16	16	15	15	16
14	14	14	16	14	16	16	14	16	23	49	77	84	77	81	69	65	57	29	17	17	16	14	14
14	14	14	15	14	15	14	14	15	27	59	70	52	50	73	97	110	113	97	62	27	18	16	16
15	15	15	15	14	15	15	14	19	43	74	74	61	77	105	127	127	125	96	65	28	16	15	15
4	4	5	4	4	5	8	4	8	23	59	65	65	76	100	115	127	127	72	27	16	9	5	4
15	14	14	14	15	15	15	15	15	19	28	33	36	31	47	89	107	92	62	33	21	23	18	14
15	14	14	14	15	14	15	15	14	15	15	18	16	14	21	45	68	68	33	19	25	31	19	15
14	13	14	16	14	13	14	14	14	14	14	16	13	11	12	23	36	46	36	22	20	24	30	26
14	15	15	15	14	14	14	15	14	14	14	13	11	10	10	13	15	22	24	21	18	30	27	24
16	14	14	14	14	14	14	15	15	15	14	14	13	13	14	13	13	15	16	15	15	22	25	16
5	5	5	5	4	4	5	4	8	4	4	5	4	8	5	4	9	4	5	4	5	8	9	9
30	28	16	15	15	16	15	16	16	15	15	14	15	18	15	15	15	15	15	16	15	16	15	16
63	35	15	14	14	14	14	16	15	15	15	15	15	15	14	14	14	15	16	15	16	15	14	15
96	44	17	14	12	12	13	14	14	17	17	14	14	16	16	13	13	13	14	14	14	14	14	13
86	44	27	20	13	10	12	10	12	13	13	14	15	14	15	14	13	14	14	14	15	15	15	15
36	28	25	16	12	13	12	10	12	9	10	12	13	15	15	15	15	15	15	15	15	14	15	15
13	16	21	13	4	3	3	1	1	1	1	3	3	4	6	5	7	7	7	7	5	5	9	9
27	44	64	51	23	12	12	12	11	11	12	12	12	12	13	12	13	15	15	16	16	15	14	14
15	15	12	12	10	10	10	10	10	9	9	10	10	10	11	10	10	10	11	12	15	14	11	10
11	11	11	10	10	10	10	11	12	10	10	10	10	11	11	12	12	10	11	11	11	10	10	10
10	10	10	10	10	9	9	10	10	10	10	9	10	9	10	10	10	10	12	13	10	9	10	10
10	9	10	10	10	10	12	10	10	12	10	10	9	10	9	12	13	13	13	14	13	10	9	10
4	1	1	1	1	1	1	1	0	1	3	3	1	0	3	4	4	7	11	9	3	1	1	1
16	14	13	13	11	11	11	11	11	11	12	13	11	12	14	15	18	16	16	16	15	13	15	15
18	18	15	15	14	12	11	12	11	11	12	15	14	14	15	16	21	25	31	51	57	59	71	57

Figure 4.23 A Segment of Reconstructed Data (Band 2)

```
 17,12    18,12    19,12    20,12    21,12    22,12    23,12    24,12    25,12    26,12
 27,12    28,12    29,12    30,12    31,12    32,12    33,12    34,12    35,12    36,12
 37,12    38,12    39,12    40,12    41,12    42,12    43,12    44,12    45,12    46,12
 47,12    48,12    49,12    50,12    51,12    52,12    53,12    54,12    55,12    56,12
 57,12    58,12    59,12    60,12    61,12    62,12    63,12    64,12    65,12    66,12
 67,12    68,12    69,12    70,12    71,12    72,12    73,12    74,12    75,12    76,12
 77,12    78,12    79,12    80,12    81,12    82,12    83,12    84,12    85,12    86,12
 87,12    88,12    89,12    90,12    91,12    92,12    93,12    94,12    95,12    96,12
 97,12    98,12    99,12   100,12   101,12   102,12   103,12   104,12   105,12   106,12
107,12   108,12   109,12   110,12   111,12   112,12   113,12   114,12   115,12   116,12
117,12   118,12   119,12   120,12   121,12   122,12   123,12   124,12   125,12   126,12
127,12   128,12    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4
 -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4    -2, 4     2, 4     2, 4
  2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4
  2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4
  2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4
  2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4
  2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4     2, 4
```

Figure 4.24 A Segment of Table ITAB

APPENDIX B: HUFFMAN SOURCE CODING

Several algorithms exist for efficiently coding sources whose statistics are known. These techniques have been investigated at TRW and the Huffman code was chosen as being the most desirable algorithm for ground processing. The Huffman code has all the properties required to ensure unique decoding with the minimum number of bits that can be obtained, coding each symbol at a time, and permits use of a "table look-up" decoding algorithm which can be performed rapidly.

A difficulty encountered in practical applications is the cumbersome algorithm required for the classical synthesis of a Huffman code given the statistics of the source symbols S. TRW has developed a more efficient technique for generation of Huffman codes. This algorithm also permits grouping of low probability symbols together for simplified decoding. Following a discussion of the classical Huffman code synthesis, the new algorithm will be described.

The Classical Synthesis of Huffman Codes [3]

Consider the source S with symbols $S_1, S_2, \ldots, S_q$ and symbol probabilities $P_1, P_2, \ldots, P_q$ and $\sum_{i=1}^{q} P_i = 1$. Let the symbols be ordered so that $P_1 \geq P_2 \geq \cdots \geq P_q$. By regarding the last two symbols of S as combined into one symbol, we obtain a new source from S containing only q-1 symbols. This new source is called a reduction of S. The symbols of this reduction of S may be re-ordered again in terms of decreasing probability and again the two least probable symbols of the reduced S are combined to form a second reduction. By continuing this reduction process, a sequence of sources is formed, each containing one less symbol than the previous reduction. The process is finished when a reduced source contains only two symbols.

A compact instantaneous binary code for the final reduction is the trivial code with words 0 and 1. Working backward from this final reduction, the Huffman code is synthesized as follows. Assume that a compact instantaneous code has been found for S_i, one of the sources in a sequence of reduced sources. One of the symbols of S_i, say S_α, is formed from two symbols of the preceding source S_{i-1}. Call these symbols $S_{\alpha 0}$ and $S_{\alpha 1}$. Each of the other symbols of S_i corresponds to one of the remaining symbols of S_{i-1}. The compact instantaneous code for S_{i-1} is formed from the code derived for S_i as follows:

Assign to each symbol of S_{i-1} (except $S_{\alpha 0}$ and $S_{\alpha 1}$) the codeword used by the corresponding symbol of S_i. The codewords used by $S_{\alpha 0}$ and $S_{\alpha 1}$ are formed by adding a 0 and 1, respectively, to the codeword used for S_α. An example of the synthesis procedure for a given source is illustrated in Figure B1. Each symbol S_i of the source S is assigned a codeword of length ℓ_i. The average code length for this source is therefore

$$\overline{L} = \sum_{i=1}^{q} P_i \ \ell_i$$

where $\overline{L}$ satisfies the inequality

$$0 \leq \overline{L} \leq H = - \sum_{i=1}^{q} P_i \log_2 P_i$$

where H is the entropy of the source S.

The difficulty imposed by the classical Huffman synthesis involves the forward flow of the code generation between successive reduced sources. This procedure is very inefficient of time and storage when used as the basis of a computer algorithm for coding a source.

An Improved Huffman Algorithm for Computers

The new algorithm separates the source reductions from the code synthesis. The first part of the algorithm keeps track of the number of times each symbol in the original source is grouped during the sequence of source reductions. This contains all information as to the length of the codeword assigned to that symbol in the resulting Huffman code. The second part of the algorithm uses these lengths, ℓ_i, to generate a Huffman code C for the source S.

Note that the resulting Huffman code may or may not be identical to the code generated by the classical synthesis procedure, but the average code length is identical. Using the classical technique, many different Huffman codes can also be generated, depending on the assignment of 0 and 1 in each reduced source.

An example of the determination of codeword lengths, ℓ_i, is given in Figure B2 for the same source used in Figure B1. The second part of the algorithm is illustrated in Figure B3. This part of the algorithm operates as follows:

1. The lengths of ℓ_i are ranked in the order of increasing length.
2. Symbol S_k of minimum length, ℓ_k, is assigned ℓ_k zeros.
3. Each successive symbol S_m has a code formed as

$$C_m = (C_{m-1}+1) + (\ell_m-\ell_{m-1}) \text{ zeros.}$$

This algorithm is very fast and essentially separates the problem of code generation from that of source reduction. The only information which need be stored from the source reduction portion of the algorithm is the vector of code lengths.

Low Probability Symbol Grouping

Often the total number of symbols S_i in source S is quite large and many of these symbols have probabilities of a small fraction of one percent. To save time in the encoding/decoding process at the expense of a small increase in average code length, these low probability symbols can be lumped into a single symbol. As an example, after ordering symbols with decreasing probability of occurrence, the first J symbols are directly coded, where $\sum_{i=1}^{J} P_i \geq .99$. The remaining symbols, having a total probability P_{J+1} of one percent or less, or grouped into symbol S_{J+1}. If M symbols are lumped into S_{J+1}, R bits must be used to describe these M symbols, where $R = \left\{\log_2 M\right\}^*$. During transmission, codeword C_{J+1} is followed by R bits to describe which of the M symbols occurred. The average code length is lengthened by such a grouping by less than P_i R.

* $\{\ \}$ means next larger integer.

The advantage of grouping symbols which seldom occur is that the maximum length of any code word can be held to some predetermined length N. This simplifies the decoding algorithm and keeps the length of the required look-up table to length 2^N. These advantages in decoding are obtained at the possible expense of a slightly increased average code length.

During the encoding of symbols S, whenever one of the symbols occurs which is in the grouping the compressor transmits the sequence of bits forming code word C_{J+1} followed by R bits to describe which grouped symbol occurred. When the decoder encounters code word C_{J+1}, it uses the next R bits to decode this grouped symbol.

COMPUTER PROGRAM

A computer program has been developed and tested which accepts an array of symbols and generates the Huffman code. The program allows the operator to group symbols if desired and generates the grouped Huffman code and the average bit rate if R bits are used to separate the lumped symbols.

The flowchart describing the program is given in Figure B4. The inputs required are the source symbols S, their associated probabilities P, and the maximum codeword length acceptable N. The program outputs the Huffman coded Table MUF, which contains the coded bit stream C associated with the source symbols S.

Two major subroutines are used in this program. Subroutine ORDER re-orders the symbols and their probabilities in a decreasing order so that the most probable symbols are at the top of an array O. Subroutine GROUP adds the two least probable symbols in the array O to form a source reduction. This subroutine also keeps count of the number of source reductions performed and keeps track of the original source symbols which have been combined to form each reduced symbol. Each symbol is given a bit position in an array V. If symbols S_1, S_3 and S_5 have been combined in a source reduction, that reduced symbol is represented in V as the binary word $(\cdot\ \cdot\ \cdot\ \cdot\ 1\ 0\ 1\ 0\ 1)$. This representation allows a compact designation of groupings at each stage in the reduction.

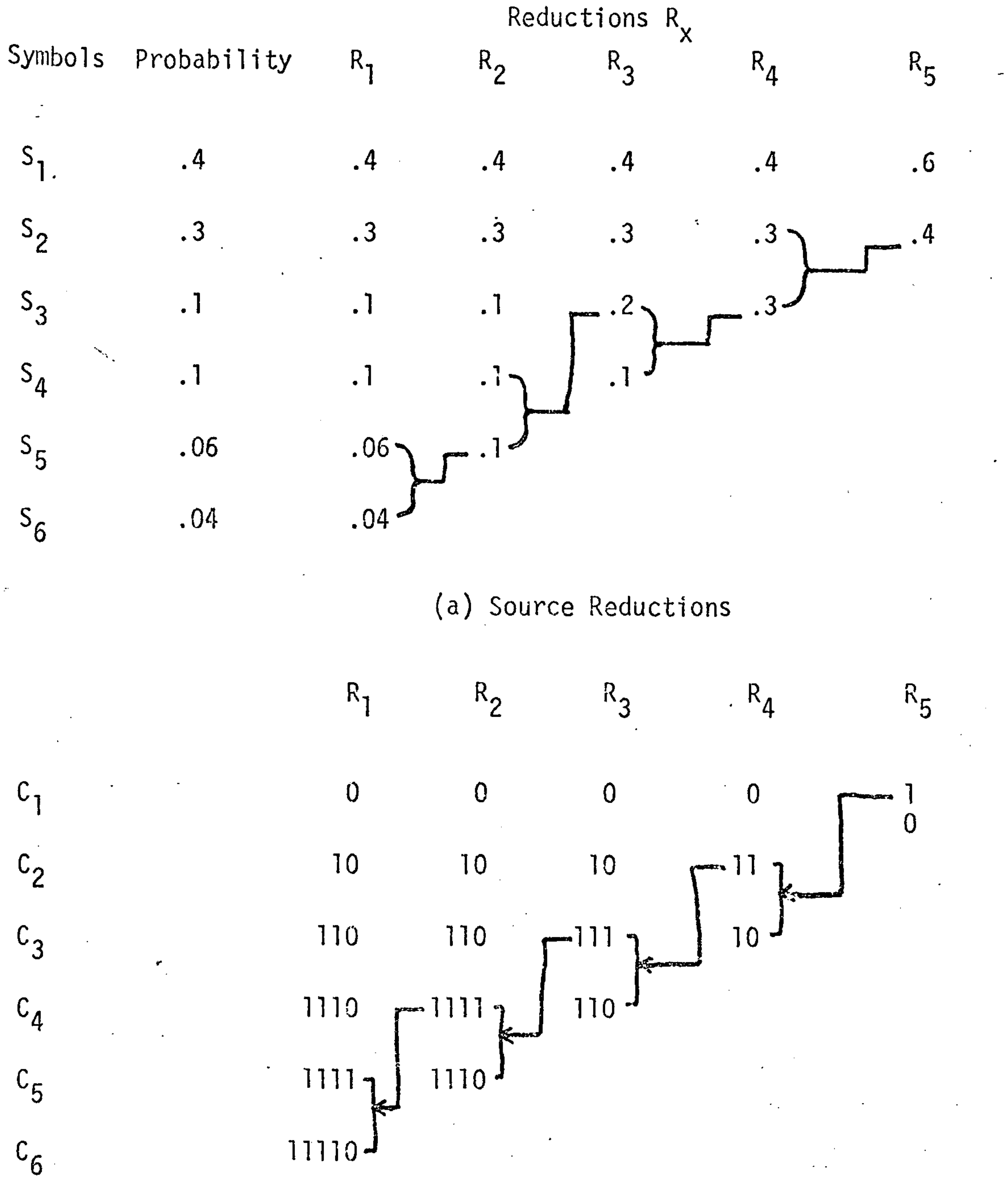
In operation, the program takes the array of input symbols and their probabilities, calls ORDER to rank them, and combines the M least probable symbols to form the grouped symbol S_{J-M+1} of probability $P_{J-M+1} = \sum_{i=M}^{J} P_i$ (assuming $P_1 \geq P_2 \geq \ldots \geq P_{J-1} \geq P_J$). This new set of J-M+1 symbols forms the input to the basic algorithm in which successive calls to subroutines GROUP and ORDER generate successive source reductions until only two reduced symbols remain. At each stage of the reduction, array LENGTH is updated by one for each symbol in S which has been combined to form one of the reduced symbols which have been grouped in that step.

Following the reduction process, the array LENGTH is used to compute the binary codeword associated with all of the J-M+1 non-grouped source symbols. LENGTH is re-ordered so that the most probable symbols which have the shortest code lengths are at the top of the array. A test takes place after LENGTH is re-ordered. If the longest codeword exceeds N bits, more source symbols are grouped and the source reductions performed again until the maximum codeword length is N or less. With 256 source symbols, such an occurrence is guaranteed at some stage of grouping.

The generation of the codes then begins with the minimum length codeword and proceeds from word to word with the successive steps of adding 1 to the previous codeword and adding the required number of zeros to fill the word.

Table HUF is then generated where all entries corresponding to non-grouped symbols contain the computed Huffman codeword. For all grouped symbols, the entry in HUF contains the lumped prefix codeword C_{J-M+1} followed by 8 bits giving the symbol directly.

This program has been written by TRW and tested using SSDI encoding of subscenes from the MSS tape ERTS E-1025-15103.



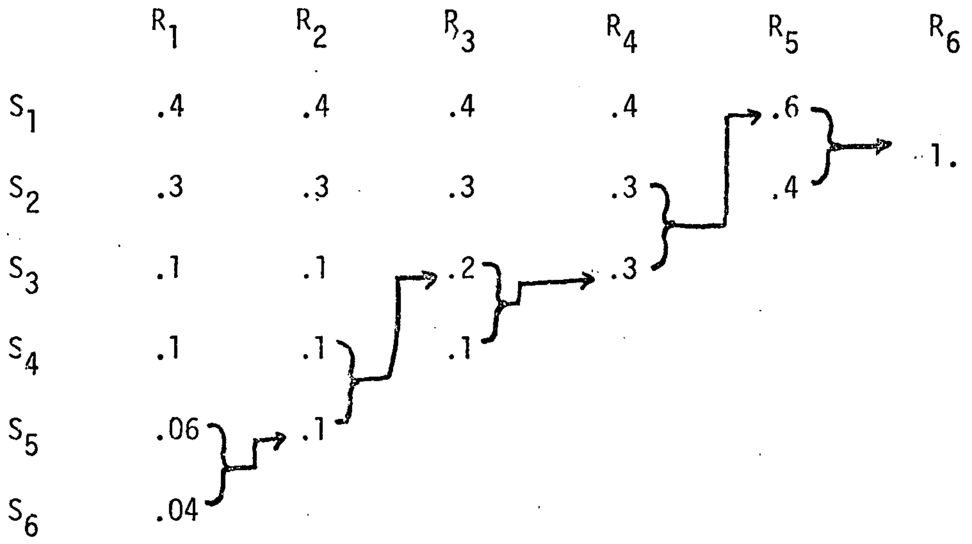
		Reductions R_x				
Symbols	Probability	R_1	R_2	R_3	R_4	R_5
S_1	.4	.4	.4	.4	.4	.6
S_2	.3	.3	.3	.3	.3	.4
S_3	.1	.1	.1	.2	.3	
S_4	.1	.1	.1	.1		
S_5	.06	.06	.1			
S_6	.04	.04				

(a) Source Reductions

	R_1	R_2	R_3	R_4	R_5
C_1	0	0	0	0	1 0
C_2	10	10	10	11	
C_3	110	110	111	10	
C_4	1110	1111	110		
C_5	1111	1110			
C_6	11110				

(b) Code Synthesis

Figure B1: Classical Huffman Code Synthesis



	R_1	R_2	R_3	R_4	R_5	R_6
S_1	.4	.4	.4	.4	.6	1.
S_2	.3	.3	.3	.3	.4	
S_3	.1	.1	.2	.3		
S_4	.1	.1	.1			
S_5	.06	.1				
S_6	.04					

ℓ_1	0	0	0	0	0	1
ℓ_2	0	0	0	0	1	2
ℓ_3	0	0	0	1	2	3
ℓ_4	0	0	1	2	3	4
ℓ_5	0	1	2	3	4	5
ℓ_6	0	1	2	3	4	5

Figure B2: Determination of Code Word Lengths, ℓ

SYMBOL	LENGTH, ℓ_i	OPERATION	CODEWORD
S_1	1	0	0
S_2	2	0 + 1 and 1 shift	10
S_3	3	10 + 1 and 1 shift	110
S_4	4	110 + 1 and 1 shift	1110
S_5	5	1110 + 1 and 1 shift	11110
S_6	5	11110 + 1 and no shift	11111

Figure B3: Huffman Code Synthesis Using Code Word Lengths ℓ_i

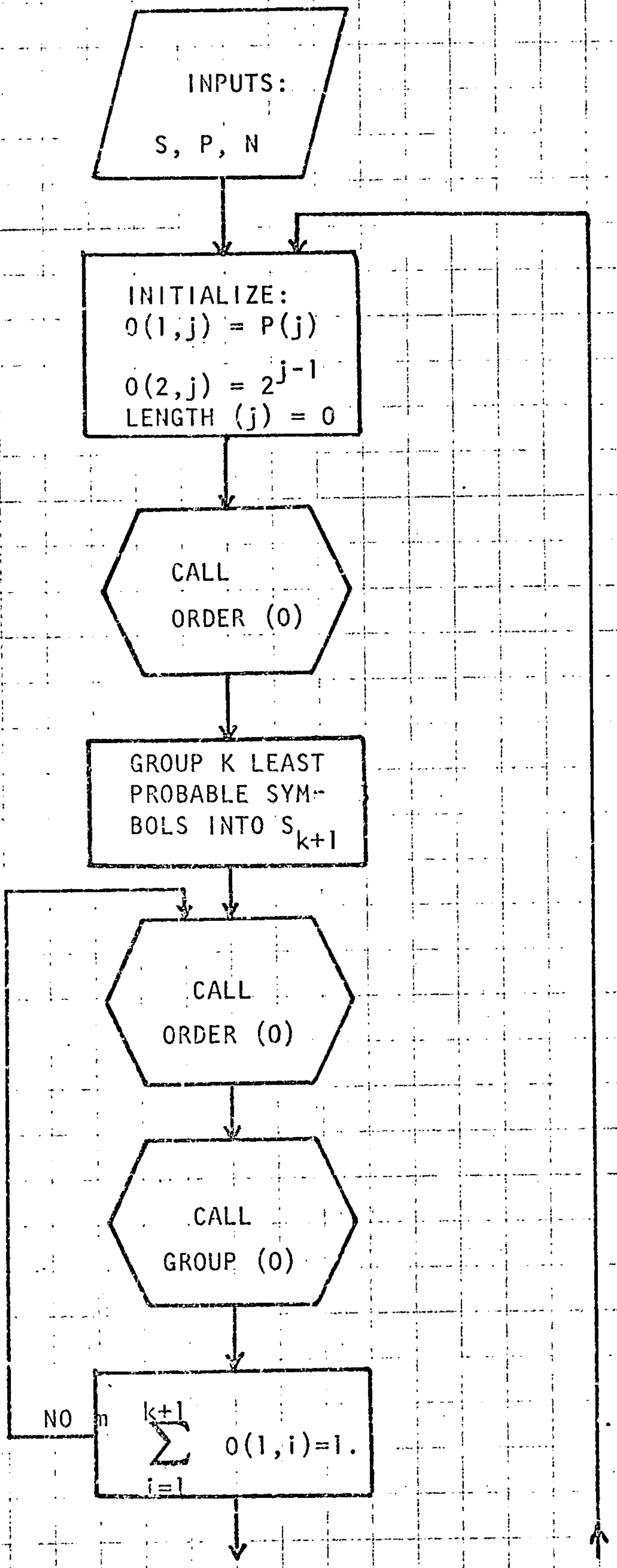


Figure B4: PROGRAM FOR HUFFMAN CODING

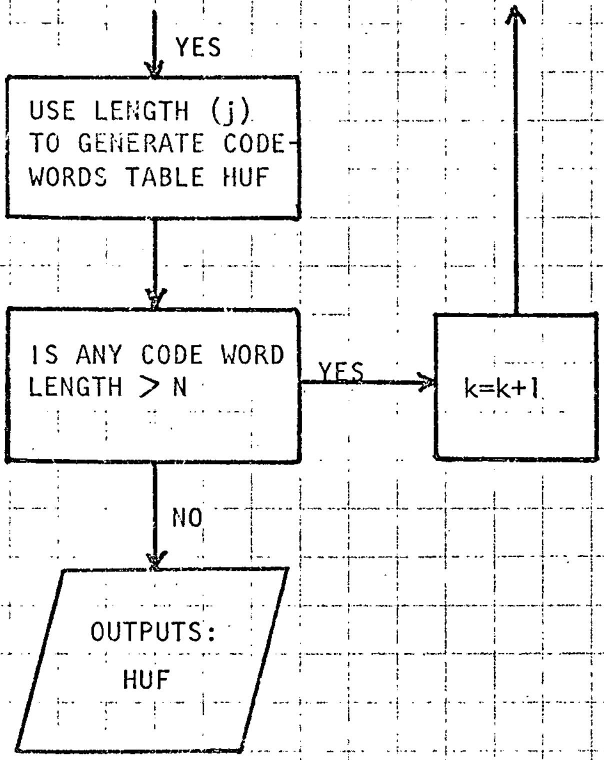

YES
USE LENGTH (j) TO GENERATE CODE-WORDS TABLE HUF
IS ANY CODE WORD LENGTH > N
YES
k=k+1
NO
OUTPUTS:
HUF

Figure B4: PROGRAM FOR HUFFMAN CODING