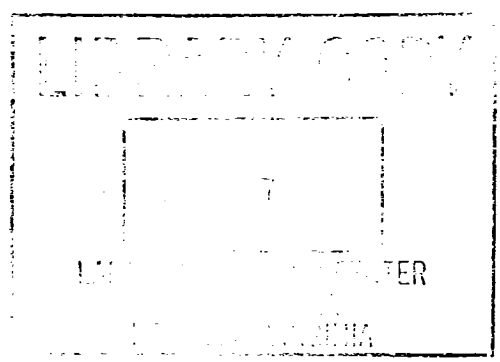
NASA-TP-3441 19940017310

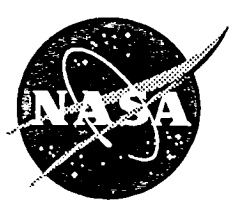
NASA Technical Paper 3441

December 1993

Application Guide for Universal Source Encoding for Space

Pen-Shu Yeh and Warner H. Miller



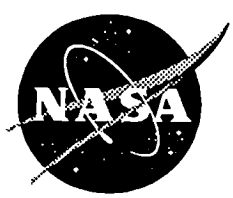


NASA Technical Paper 3441

1993

Application Guide for Universal Source Encoding for Space

Pen-Shu Yeh and
Warner H. Miller
Goddard Space Flight Center
Greenbelt, Maryland



National Aeronautics and Space Administration

Scientific and Technical Information Branch

ABSTRACT

Lossless data compression has been studied for many NASA missions. The Rice algorithm has been demonstrated to provide better performance than other available techniques on most scientific data. A top-level description of the Rice algorithm is first given, along with some new capabilities implemented in both software and hardware forms. The document then addresses systems issues important for onboard implementation including sensor calibration, error propagation and data packetization. The latter part of the guide provides twelve case study examples drawn from a broad spectrum of science instruments.

ACKNOWLEDGEMENT

The authors wish to thank the following colleagues for providing technical information or data on the science instruments used in case study examples. Some also gave valuable comments which greatly improve the content and readability of the document.

John Barker
Gorden Chin
Davila Joseph
Larry Evans (with CSC)
Keith Feggans (with ACC)
Robert Griswold (with GST, Inc.)
David Hancock, 3rd
Jim Irons
Keith Lai (with HSTX)
Tom Lynch (with Hughes)
Ronald Muller
Robert Rice (with JPL)
Keith Strong (with Lockheed Palo Alto Research Lab.)
Leslie Thompson
Paul Westmeyer

TABLE OF CONTENTS

1.0 Introduction 1

2.0 Algorithm 2

2.1 The Preprocessor 3

2.1.1 Predictor 3

2.1.2 References 4

2.1.3 The Mapper 4

2.2 Entropy Coder 4

2.2.1 The Sample-Split Options 4

2.2.2 Option Selection 5

2.2.3 Low Entropy Option 5

2.2.4 Default Option 5

3.0 Systems Issues 5

3.1 Sensor Characteristics and Calibration 5

3.2 Error Propagation 6

3.3 Data Packetization and Compression Performance 7

4.0 Compression Illustrations 8

4.1 Landsat Thematic Mapper 9

4.2 Heat Capacity Mapping Radiometer on Heat Capacity Mapping Mission 9

4.3 Moderate-Resolution Imaging Spectrometer (MODIS) on EOS 12

4.4 Advanced Solid-State Array Spectrometer (ASAS) on C-130B 12

4.5 NS001 Multispectral Scanner on C-130B 15

4.6 Wide-Field Planetary Camera (WFPC) on Hubble Space Telescope 17

4.7 Soft X-ray Telescope (SXT) on Solar-A Mission 20

4.8 Goddard High-Resolution Spectrometer (GHRS) on HST 20

4.9 Acousto-Optical Spectrometer (AOS) on Submillimeter Wave Astronomy Satellite (SWAS) 23

4.10 Gamma-Ray Spectrometer on Mars Observer 24

4.11 Radar Altimeter on TOPEX/POSEIDON 26

4.12 Gas Chromatograph-Mass Spectrometer on Cassini 28

5.0 References 29

Application Guide for Universal Source Encoding for Space

1.0 Introduction

Lossless data compression has been suggested for many NASA applications to either increase the science return or to reduce the requirement on onboard memory, station contact time, and data archival volume. This type of compression technique guarantees a full reconstruction of the original data without incurring any distortion in the process.

Of the many available lossless data compression algorithms, the Rice algorithm has been demonstrated to perform better in most studies on scientific data. The algorithm has also been implemented in a number of NASA's science exploration missions. The earlier applications implemented the algorithm in software form executed by an onboard microprocessor. In 1991, a hardware engineering model was built in an Application Specific Integrated Circuit (ASIC) for proof of concept. This particular chip set was named as Universal Source Encoder/Universal Source Decoder (USE/USD). Later, it was redesigned with several additional capabilities and implemented in Very Large Scale Integration (VLSI) circuits using gate arrays suitable for space missions. The flight circuit is referred to as Universal Source Encoder for Space (USES).

This document aims to provide first a top-level description of the Rice algorithm architecture, along with some new capabilities. It then addresses systems issues important for onboard implementation. It also aims to serve as an application guide. The latter part of the document provides case study examples drawn from a broad spectrum of science instruments. The examples demonstrate how to obtain optimal compression performance for various scientific space applications. Some of these studies have resulted in an onboard implementation; others provided input for future missions.

In the case study examples, whenever feasible, a compression performance comparison with another commercially available technique is included. No attempt was made to exhaust all existing techniques for this purpose. It should be noted that there may exist different means to apply the Rice algorithm to the data, mainly, in preprocessing or reformatting the data before presenting it to the entropy coding module. The users should be aware that the entropy coding scheme in the Rice algorithm, as discussed in reference 2, is a subset of Huffman codes optimal for Laplacian symbol sets. This set of Huffman codes also performs well on Gaussian or Poisson symbol sets. To optimize performance, one would try to preprocess the data set so that it conforms closely to a Laplacian symbol set distribution. The technique for approaching this goal is only limited by the user's imagination.

The types of scientific instruments studied include Charge Coupled Device (CCD) imager, radio wave spectrometer, radar altimeter, gamma ray spectrometer and others. Only mis-

sion and instrument titles will be listed in this document; details on each mission and instrument are not provided.

Users should consult the references for further information on the algorithm and related issues. Reference 1 provides a general overview of the Rice algorithm. Reference 2 details performance analysis. An analysis on the requirement for data flow smoothing buffer is given in reference 3. Reference 4 describes the USE/USD VLSI hardware. Reference 5 provides available hardware specification, while reference 6 provides a source for software code. The data structure standardized by the international space committee Consultative Committee for Space Data Systems (CCSDS) is given in reference 7.

2.0 Algorithm

A block diagram of the architecture of the Rice algorithm is given in Figure 2.1. It consists of a preprocessor to decorrelate data samples and subsequently map them into symbols suitable for the following stage of entropy coding. The entropy coding module is a collection of options operating in parallel over a large entropy range. The option yielding the least number of coding bits will be selected. This selection is performed over a block of J samples to achieve adaptability to scene statistics. An Identification (ID) bit pattern is used to identify the option selected for each block of J input data.

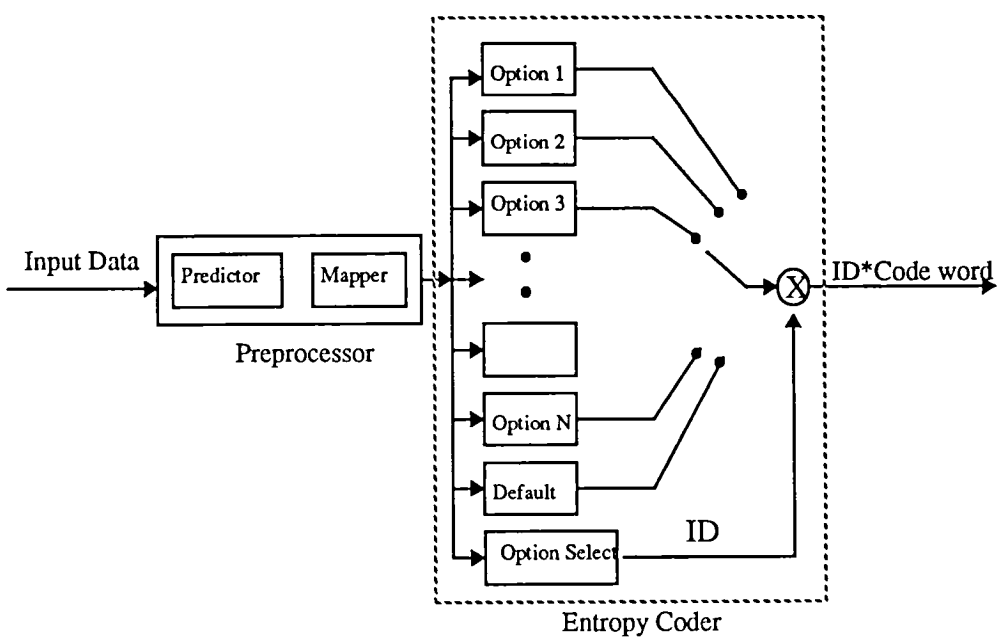


Figure 2.1. The encoder architecture.

2.1 The Preprocessor

2.1.1 Predictor

Lossless source coding for data compression is a method by which only redundancy is removed from the source data in order to achieve the objective that the reconstructed data is identical to the input data, or in other words, the combined pre- and postprocessor procedure has to be completely reversible. There are many types of pre/postprocessors that have this reversible property.

The decorrelation function of the preprocessor can be implemented by a judicious choice of the predictor for an expected set of data samples. The decision should take into account possible variations in the background noise and the gain of the sensors for acquiring the data. The predictor should be chosen to minimize the randomness of the noise resulting from the sensor nonuniformity. An optimal predictor would give small prediction errors with a probability distribution resembling a Laplacian function. There are also some data types that require no decorrelation. These types can be routed directly to the entropy coding module.

A technique widely used as a basic lossless preprocessor function is predictive coding. The simplest predictive coder is a unit-delay predictor shown in Figure 2.2. The output, Δ_i, will be the difference between the input data symbol and the preceding data symbol. The input data signal is assumed to be already linearly quantized. The inherent compression ability of such a predictive coder occurs typically from there being statistically fewer quantization levels used after the differencing operation than in the original input data samples.

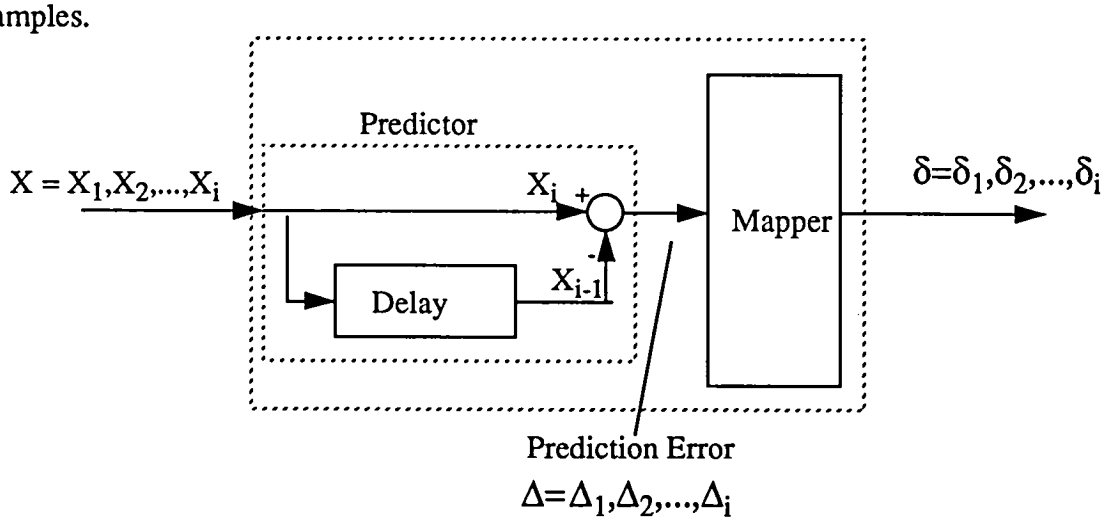


Figure 2.2. A unit-delay predictive preprocessor.

Several other types of prediction modes exist in the USES implementation. These include:

1. a default predictor that takes the previous data sample as the predictor value for the current data sample.
2. an external predictor supplied by the user.

3. a 2D predictor that uses the average of the previous adjacent sample and a user-supplied value as the predictor value.
4. a multispectral (multi-source) predictor which uses data from one channel as input to a higher-order prediction function for the data in a second channel.

These prediction modes are user-selectable during the initialization period for using the hardware. For software implementation, users can easily program any prediction scheme that best decorrelates the input data stream. Users would subsequently select the prediction mode as a run-time parameter.

2.1.2 References

A reference symbol is an unaltered data symbol upon which succeeding symbol differences are based. References are required by the decoder to recover the absolute values from difference values except when the source encoder is operating in the entropy coding mode in which no prediction on data will be performed. The user must determine how often to insert references. When inserted, the reference shall precede the first symbol of a block of J symbols. In packetized formats, the reference shall be applied to the first sample in the packet data fields.

2.1.3 The Mapper

The function of the predictive coder is to decorrelate the incoming data stream by taking the difference between data symbols. The mapper takes these difference values, both positive and negative, and orders them, based on predictive values, sequentially as positive integers. The mapper shall map the positive differences, Δ_i, into even δ_i, and the negative differences into odd δ_i. For N-bit quantization, the mapped symbol set shall have 2^N elements $\delta_0, \delta_1, ..., \delta_n$ ($n = 2^N - 1$). These elements preferably shall have corresponding probabilities approximating the following relationship:

$$P_0 \geq P_1 \geq P_2 \geq ... \geq P_n ,$$

where P_i is the probability of occurrence for symbol δ_i.

2.2 Entropy Coder

2.2.1 The Sample-Split Options

The sample-split option in the Rice algorithm takes a block of J preprocessed data samples, splits off the *k* least significant bits, and codes the remaining higher order bits with a simple comma code before appending the split-off bits to the coded data stream. Each sample-split option in the Rice algorithm is optimal in an entropy range about one bit/sample; only the one yielding the least amount of coding bits will be chosen and identified for a J-sample data block by the option select logic. This assures that the block will be coded with one of the available Huffman codes, whose performance is better than other available

options on the same block of data. The $k = 0$ option is optimal in the entropy range of 1.5 - 2.5 bit/sample; the $k = 1$ option is optimal in the range of 2.5 - 3.5 bit/sample, and so on for other k values. For the developed ASIC hardware that implements the split sample up to $k = 12$, the effective range for this scheme extends from 1.5 bit/sample to 14.5 bit/sample. The users should be made aware that the "entropy" is the information rate for an ideal Laplacian symbol set, and usually can be approximated by the entropy of the prediction errors from an actual data source.

2.2.2 Option Selection

The entropy coder includes a code-word selection, which selects the option that performs best on the current block of symbols. The selection is made on the basis of the number of bits that the selected parameter will use to code the current block of symbols. An identifier specifies which of the optional parameters was used to code the accompanying set of code words.

2.2.3 Low Entropy Option

The low entropy option implemented in the hardware extends the performance below 1.5 bit/sample. When the prediction error comprises only small values, this low entropy option can be very effective. The special case of a block of J errors of zero prediction values is also handled in this option by specifying only an ID code. This mode is especially useful for compressing thresholded imagery.

2.2.4 Default Option

The option, not to apply any parameter, is the default case. If it is the selected option, the preprocessed block of symbols is passed through the encoder process without alteration but with an appended identifier.

3.0 Systems Issues

Several systems issues related to embedding data compression scheme onboard a spacecraft should be addressed. These include the relation between the focal-plane array arrangement and the data sampling/prediction direction, the subsequent data packetizing scheme and how it relates to error propagation in case of bit error incurred in the communication channel, and the amount of smoothing buffer required when the compressed data is passed to a constant-rate communication link without prior buffering.

3.1 Sensor Characteristics and Calibration

Advanced imaging instruments and spectrometers often use arrays of individual detectors arranged in a 1D or 2D configuration; one example is the Charge Coupled Device (CCD). These individual detectors tend to have slight differences in response to the same input photon intensity. For instance, CCDs usually have a different gain and dark current value

for each individual detector element. It is important to have the sensor well calibrated so that the data reflects the actual signals received by the sensor. Simulations have shown that for CCD types of sensors, gain and dark current variations as small as 0.2% of the full dynamic range can render the prediction scheme less effective for data compression. Besides calibration, in order to maximize the compression gain, the prediction scheme used in the preprocessor in Figure 2.1 will be much more effective when data acquired on one detector element is used as the predictor for data acquired on the same detector whose characteristics are stationary, in general, over the data collection period. An example is the push-broom scheme used in many Earth-viewing satellite systems depicted in Figure 3.1. An array of detectors is arranged in the cross-track direction, which also is the detector data readout direction. The scanning of the ground scene is achieved by the motion of the satellite in the along-track direction. In such configuration, predictive coding will be most effective in the along-track direction, in which data from the same detector is used as the predictor. An example of how to optimize compression performance in the presence of detector nonuniformity is given in Section 4.4.

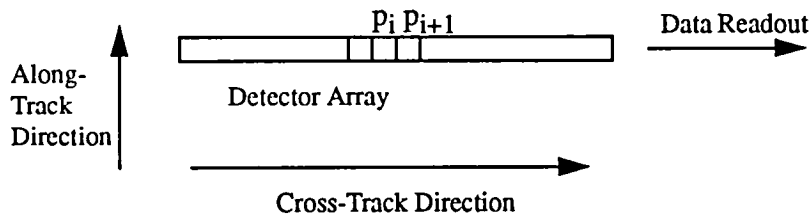


Figure 3.1. Push-broom scanning scheme.

3.2 Error Propagation

User acceptability to distortion resulting from channel bit errors is mission-dependent, and is strongly a function of received Signal-to-Noise-Ratio (SNR), the data format, the error detection and correction technique and the percent of distorted data that is tolerable. A major concern in using data compression is the possibility of error propagation in the event of a single bit error in the compressed data stream. During the decompression process, a single bit error can lead to a reconstruction error over extended runs of data points. There are two general approaches to minimizing such effect. First, it is important to have a clean channel for data communication. Some form of error detection/correction scheme is recommended for compressed data. Second, packetization of compressed data in a proper format in conjunction with the error control coding, will prevent error propagation into the next packet. An example for space application is to packetize the compressed data of one scanline in a CCSDS recommended data packet format (reference 7) and the compressed data of the next scanline into a second packet. In case of a bit error in one packet, the

decompression error will only propagate to the end of the packet, providing that the compression/decompression algorithm does not depend on information from a second packet.

3.3 Data Packetization and Compression Performance

The total number of coding bits resulting from losslessly compressing a fixed number of data samples is usually a variable. The CCSDS data architecture provides a structure to packetize the variable length data stream into a packet, which then can be put into a fixed length frame. Packetization is essential for preventing error propagation as mentioned earlier; it also can impact the compression performance for the data set. The Rice algorithm and the low-entropy option process a block of data of a predetermined number of samples, typically 8, 16, 24, etc., at one time. To facilitate decoding, packetization should be formed on compressed data of multiple blocks of as short as one block of samples or as long as the user desires. However, to prevent error propagation in the event of bit error, a reasonable length of data blocks will be chosen for a packet. The compression performance is optimal for every block of samples and is not dependent on the packet length.

There exist other available compression algorithms whose performance depends on establishing the long-term statistics of the samples in a file. In general such schemes will give good compression performance for a large file and poor performance for a relatively smaller file. And if packetization, as a means of preventing error propagation, is used in conjunction with these algorithms, one would expect poor performance for a shorter packet and a better performance for a larger packet. The drawback is that the loss of data caused by error propagation may be intolerable for the larger packet.

An example of how a data packetization scheme affects compression performance, as a comparison between the current technique and a known technique, is given in Section 4.6.

An alternative to using a variable-length packet is to fix the length of the packet. Fixed-length packets typically will have compressed data followed by fill bits to bring the packet to a fixed length. If the compressed-bit count is greater than the fixed length, truncation occurs with loss of data. The decoder recognizes the truncation condition and signals when a block has been truncated by outputting dummy symbols. The decoder will output as many dummy symbols as would be in a typical packet that was not truncated. If one or more blocks were truncated, corresponding blocks of dummy symbols will be outputted to fill out the required blocks-per-packet.

Depending on the mission requirement, the variable-length packets resulting from packing the compressed bit stream into CCSDS packets can be stored in a large onboard memory or multiplexed with other data packets before being stored in the memory device. In this case, the large-capacity memory device will smooth out the variation in the packet length. Subsequent readout from the memory will be performed at a fixed rate. In other situations, the variable-length packets may have to be temporarily buffered before a direct transmission to the communication link. The temporary buffer serves as a smoothing buffer for the link. Occasionally fill bits are inserted in the data stream to provide a constant readout rate. The amount of buffer needed is a variable of the incoming packet rate, the packet

length statistics, and the readout rate. An analysis on the buffer requirement is given in reference 3.

4.0 Compression Illustrations

This section contains compression study results for several different instruments. The compression performance is expressed as Compression Ratio (CR). It is defined as the ratio of the quantization level in bits to the average code word length, also in bits. When necessary, the performance of the enhanced Rice algorithm is compared with the commonly known and commercially available Lample-Ziv-Welch (LZW) algorithm. The LZW algorithm is particularly efficient in compressing text file that has grammatical structure. It is available on most Unix machines as a system utility. All the studies were performed in software simulator implemented in language C on a Sun Sparc workstation.

Twelve examples are included in this document. Table 4.0 summarizes the result. Notice that the compression ratio is instrument data dependent. For a specific instrument type, the compression ratio can vary from one test data set to the next. The readers should refer to each individual example for more details on each study.

Table 4.0. Summary of Compression Ratio

Instrument	USES Compression Ratio
Thematic Mapper	1.83
Heat Capacity Mapping Radiometer	2.19
Moderate-Resolution Imaging Spectrometer	1.89/band1 1.37/band2
Advanced Solid-State Array Spectrometer	1.74/broad band
NS001 Multispectral Scanner	1.60/averged over 8 bands
Wide-Field Planetary Camera	2.97/no threshold
Soft X-ray Telescope	4.96
Goddard High-Resolution Spectrometer	1.72/spectrum only
Acousto-Optical Spectrometer	2.3
Gamma-ray Spectrometer	26/at 5-sec collection
Radar Altimeter	1.41
Gas Chromatograph-Mass Spectrometer	1.94

4.1 Landsat Thematic Mapper

Mission Purpose: The Landsat program was initiated for the study of Earth's surface and resources. Landsat-1, Landsat-2, and Landsat-3 were launched between 1972 and 1978. Landsat-4 was launched in 1982, and Landsat-5 in 1984.

Landsat Thematic Mapper (TM) on Landsat-4, 5, and 6: The TM data represent typical land observation data acquired by projecting mirror-scanned imagery to solid-state detectors. The sensor for band 1-5 is a staggered 16-element photo-diode array; it is a staggered 4-element array for band 6. An image acquired on Landsat-4 at 30m ground resolution for band 1 in the wavelength region of 0.45 - 0.52 μm is shown in Figure 4.1. This 8-bit 512x512 image was taken over Sierra Nevada in California and has relatively high information over the mountainous area.

Compression Study: Using a 1D predictor in the horizontal direction, setting a block size of 16 samples and inserting one reference per every image line, the lossless compression gives a compression ratio at 1.83 for the 8-bit image. In contrast, a direct application of the LZW algorithm, available on Unix system as *compress*, gives a compression ratio at 1.51.

4.2 Heat Capacity Mapping Radiometer on Heat Capacity Mapping Mission

Mission Purpose: Launched in 1978, the mission supported exploratory scientific investigations to establish the feasibility of utilizing thermal infrared remote sensor-derived temperature measurements of the Earth's surface within a 12-hour interval, and to apply the night/day temperature difference measurements to the determination of thermal inertia of Earth's surface.

Heat Capacity Mapping Radiometer: The sensor is a solid state photo-detector sensitive in either the visible or the infrared region. A typical 8-bit image taken in the visible light region is given in Figure 4.2, at a ground resolution of 500m over the Chesapeake Bay area on the East Coast.

Compression Study:

Prediction on 1D: Choosing 1D default prediction in the horizontal direction, and setting block size J at 16, lossless compression gives a compression ratio of 2.19. A direct application of Unix *compress* gives a compression ratio of 1.95.

Prediction on 2D: Using a 2D predictor that takes the average of the previous sample and the sample on the previous scan line, and keeping other parameters the same, USES compression gives a CR of 2.28, about 5% increase from using a 1D predictor.

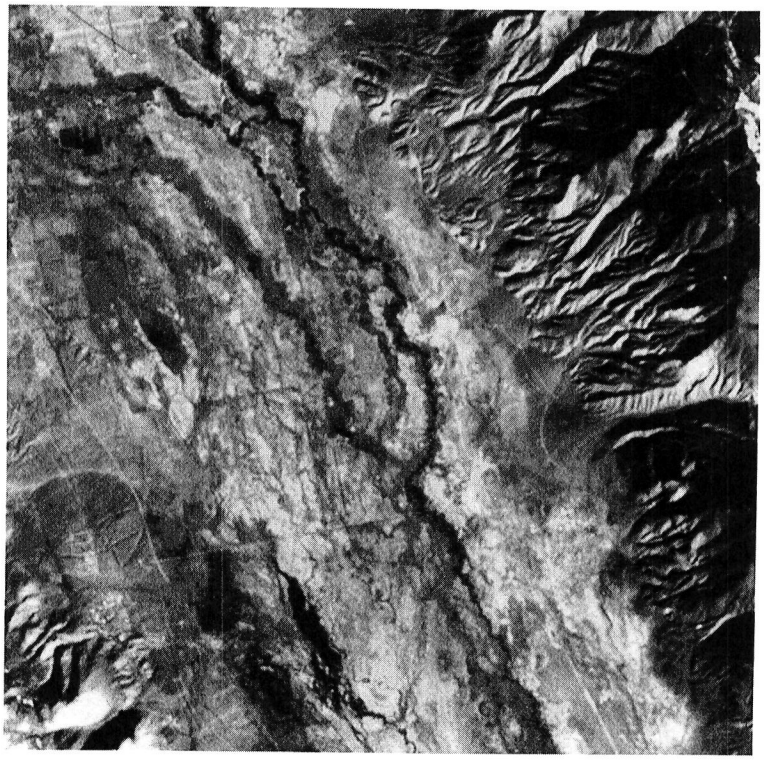


Figure 4.1. Thematic mapper image.

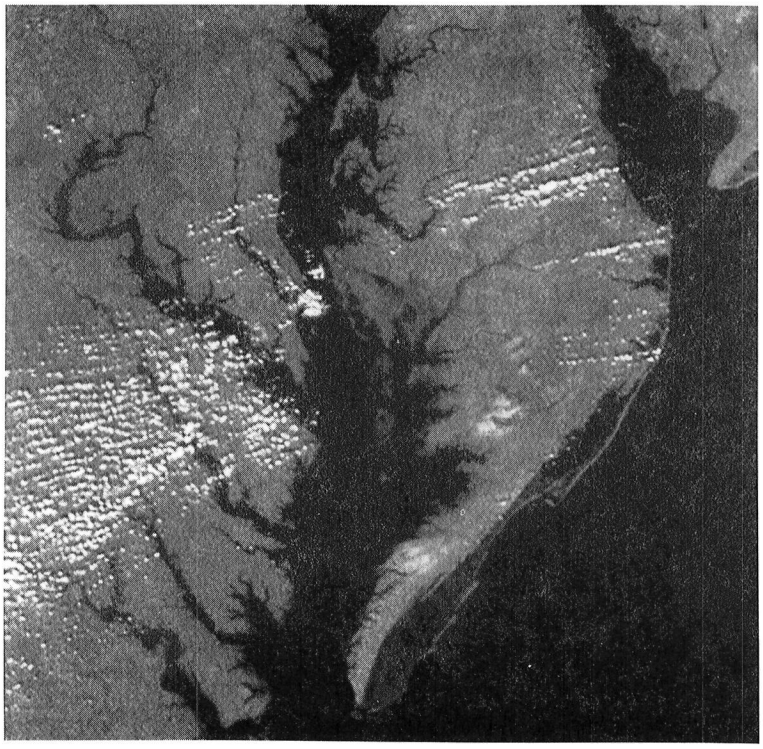


Figure 4.2. HCMR image.

4.3 Moderate-Resolution Imaging Spectrometer (MODIS) on EOS

Mission Purpose: MODIS is one of the instruments planned for the Earth Observing System (EOS). The EOS mission is commissioned to study the global change of the Earth over a prolonged period of time through various sensors for detecting changes in the Earth's environment.

Moderate-Resolution Imaging Spectrometer: The spectrometer has 36 bands over the visible and infrared wavelength region. The sensor uses monolithic focal-plane CCD or photo-diode arrays of sizes 10, 20, or 40 elements designated for different bands. A simulated image at 250m resolution over eastern Florida is given in Figure 4.3 for band 1 centered at 0.659 μm and band 2 centered at 0.865 μm. MODIS will generate data at a 12-bit dynamic range. This test image size is 256x256.

Compression Study:

Prediction on 1D: Using 1D default prediction in the horizontal direction, and at a block size J of 16, USES compression gives a compression ratio of 1.89 for band 1 test image, and a ratio of 1.56 for band 2. The Unix *compress* gives a ratio of 1.71 for band 1 and 1.37 for band 2 test image.

Multispectral prediction: Using band 1 as input to the higher order predictor for band 2, and keeping other parameters the same as in 1D prediction, the compression ratio increases from 1.56 to 1.67, a 7% increase in this case. The efficiency of the multispectral prediction technique will be further explored later.

4.4 Advanced Solid-State Array Spectrometer (ASAS) on C-130B

Mission Purpose: The NASA Earth Resources Aircraft Program at Ames Research Center operates a C-130B aircraft to acquire data for earth science research. It provides a platform for a variety of sensors that collect data in support of scientific projects sponsored by NASA, as well as federal, state, university, and industry investigators.

Advanced Solid-State Array Spectrometer (ASAS): The sensor collects data at 10-bit dynamic range, with 29 or more spatially registered bands. The image in Figure 4.4 was acquired for the geologic remote-sensing field experiment. It's ground resolution is 5.5m across track and the size of the image is 512x360. Images from several bands are combined to give a broader bandwidth of 500~870nm to simulate a possible panchromatic band for the Landsat-7 program. It is evident from the image that, because of either detector nonuniformity or insufficient calibration, streaking noise appears across the image in the vertical direction.

Compression Study:

Prediction in horizontal direction: This prediction scheme does not try to optimize the predictor performance. The compression ratio is 1.42 using 10-bit as base in calculating the ratio. On computer systems, these types of 10-bit/pixel data are often stored as integer*2 data using two bytes for each datum, the algorithm will compress the file at a ratio of 2.27, compared to 1.58 achieved by the Unix *compress* utility.

Prediction in vertical direction: To reduce the prediction error, a more preferable predictor applies the prediction on data from the same detector. The compression ratio now increases from 1.42 to 1.74 and the file size reduction ratio improves to 2.78.

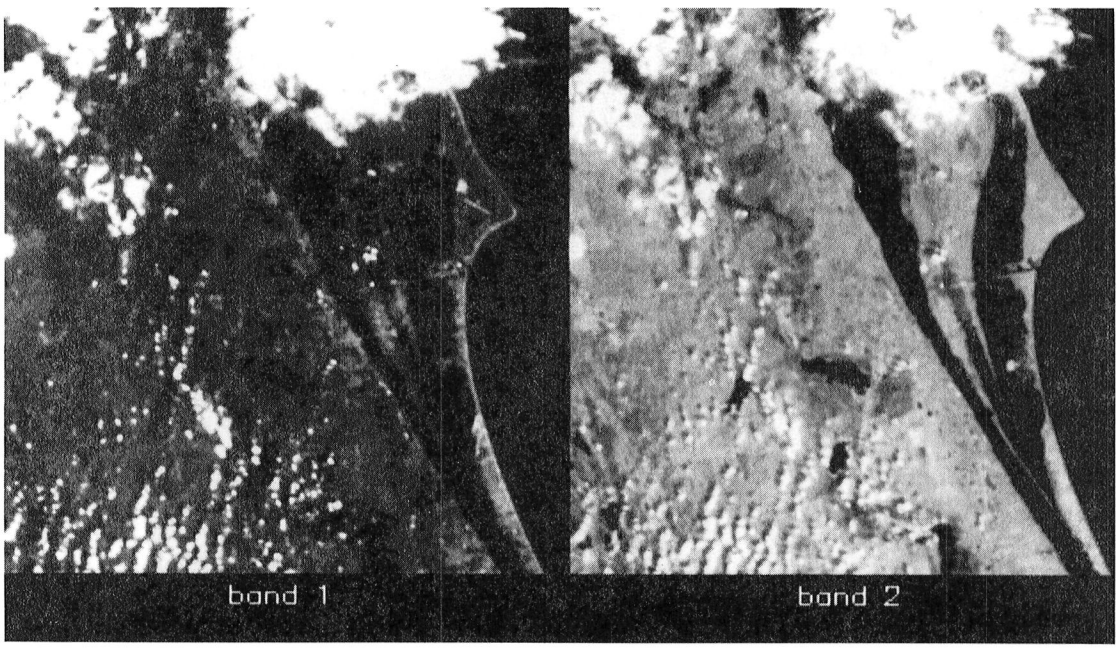


Figure 4.3. MODIS test image.

Figure 4.4. ASAS image.

4.5 NS001 Multispectral Scanner on C-130B

Mission Purpose: Same as Section 4.4.

NS001 Multispectral Scanner: The scanner contains the seven Landsat-4 Thematic Mapper bands plus a band from 1.13 to 1.35 μm. The specific bands are as follows:

Table 4.5.1. NS001 Bands

Band	Spectral Bandwidth, μm
1	0.458 - 0.519
2	0.529 - 0.603
3	0.633 - 0.697
4	0.767 - 0.910
5	1.13 - 1.35
6	1.57 - 1.71
7	2.10 - 2.38
8	10.9 - 12.3

An 8-band image acquired at ground resolution of 2.36m was first low-pass filtered to obtain a ground resolution at approximately 4.72m. This set of images was subsampled 2:1 and displayed in Figure 4.5. The image was taken over Mountain View and Moffet field in California. The compression study, however, was performed on the 4.72m image. The image has an 8-bit dynamic range.

Compression Study:

In-band prediction: Selecting the block size to be 16, the lossless compression within each band of this set of images gives the performance listed in Table 4.5.2 under USES (in-band). The within-band predictor used is the 1D previous pixel within each horizontal line. For comparison purposes, the performance obtained by the Unix *compress* is also given.

Cross-band prediction: When adjacent band is used as a predictor in the multispectral mode of the lossless compression technique, the performance improvement is obvious, as shown in the same table under USES (cross-band). In this study, band 1 is used as the predictor for band 2, band 2 is used as the predictor for band 3, etc. The users should be made aware that this study only suggests which band is more compressible with additional information from the other bands. It is evident that for the particular urban scene in Figure 4.5, band 5 and band 7 can acquire substantial improvement in reducing the data volume, while bands 2, 3, 4, and 6 receive moderate improvement. For the far infrared band 8, the study shows it is more effective to use only in-band compression.

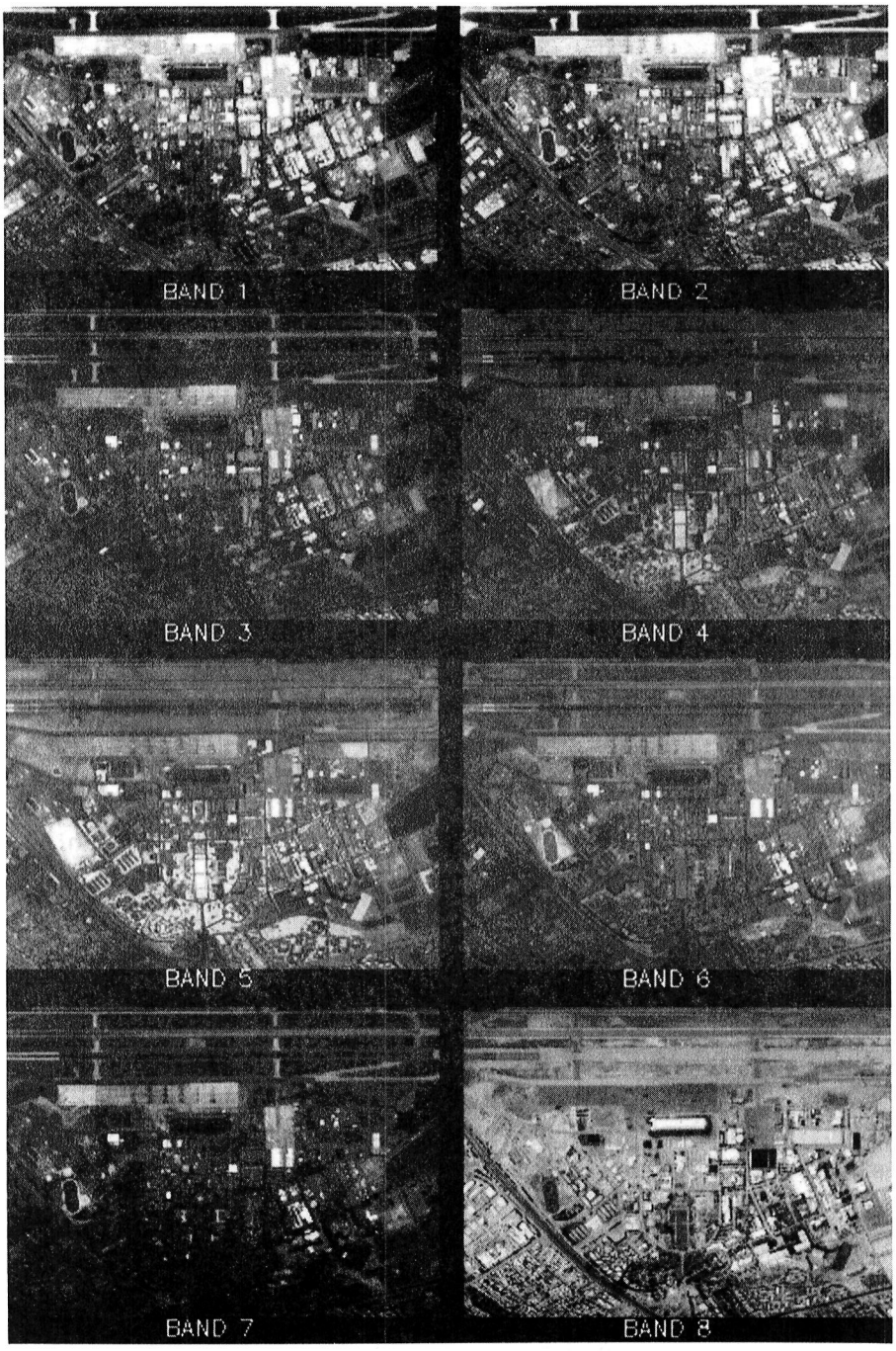


Figure 4.5. NS001 multispectral image.

Table 4.5.2. Compression Ratio on TM Test Data

Band	UNIX *compress*	USES (in-band)	USES (cross-band)	% CR Increase
1	1.04	1.40	/	/
2	1.06	1.36	1.57	15.4%
3	1.10	1.46	1.66	13.7%
4	1.09	1.43	1.51	5.6%
5	1.07	1.42	1.90	33.8%
6	1.11	1.45	1.58	9.0%
7	1.08	1.44	1.80	25.0%
8	1.08	1.43	1.36	-4.9%

4.6 Wide-Field Planetary Camera (WFPC) on Hubble Space Telescope

Mission Purpose: The Hubble Space Telescope (HST), launched in 1990, is aimed at acquiring astronomy data at a resolution never achieved before and observing objects that are 100 times fainter than other observatories have observed.

Wide-Field Planetary Camera (WFPC): The camera uses four area array CCD detectors. A typical star field observed by this camera is shown in Figure 4.6. This particular image data has maximum value of a 12-bit dynamic range, yet the background minimum value is of 9-bit value.

Compression Study: Two studies are performed:

Threshold Study: The effect of thresholding a star field on the compression performance is explored in this study. Different threshold values will be applied to this image, and the values of the data will be clipped at the threshold when they are smaller than the threshold values. The thresholding operation does not change the visual quality of the image. All the bright stars are unaffected. The resultant image will be compressed losslessly and the compression will make use of the low-entropy option frequently over the image. The original image in Figure 4.6 has a minimum quantized value of 423 and an array size of 800x800. The compression was performed using a block size of 16 and a default predictor in the horizontal line direction. Results are summarized in the following table.

Table 4.6.1. Compression Ratio at Different Threshold Values

Threshold Values	USES	LZW
None	2.97	3.30
430	2.99	3.33
440	11.92	17.15
450	41.20	63.53
460	53.52	91.48
470	63.66	110.73

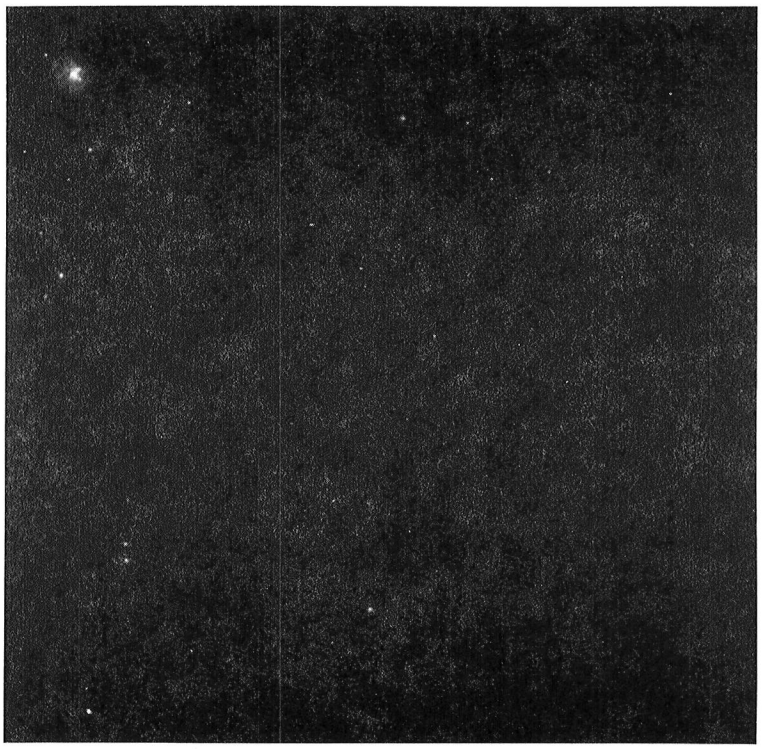


Figure 4.6. WFPC image.

These results are also compared with those obtained using the Unix *compress* command. One notices that the performance of the LZW algorithm over the whole image is better than that of USES. However, after considering the error propagation and packetization issues involved in implementing a lossless algorithm, the results are reversed.

Packetization Study: When a packet data format is used for the compressed data, it is expected that during decompression, the reconstruction error caused by a single bit error in the packet will not propagate to the next packet. It implies that the data packets are independent of each other, that no code tables or statistics of data is passed from one packet to the next. To explore the effect of packet size on the compression performance, several individual lines from Figure 4.6 are extracted and compressed separately. The results in Table 4.6.2 are obtained on lines 101 through 104, which cover partially the brightest star cluster on the upper left corner of Figure 4.6.

Table 4.6.2. Compression Ratio vs. Packet Size

Threshold	Packet Size, No. of Lines	USES	LZW
None	1 (Line 101)	2.89	1.94
None	1 (Line 102)	2.90	1.92
None	1 (Line 103)	2.86	1.90
None	1 (Line 104)	2.88	1.90
None	2 (Line 101-102)	2.90	2.13
None	2 (Line 103-104)	2.87	2.09
None	4 (Line 101-104)	2.89	2.29
470	1 (Line 101)	16.24	6.98
470	1 (Line 102)	16.92	6.98
470	1 (Line 103)	15.42	6.94
470	1 (Line 104)	16.24	6.67
470	2 (Line 101-102)	16.45	8.70
470	2 (Line 103-104)	15.80	8.60
470	4 (Line 101-104)	16.06	10.30

4.7 Soft X-ray Telescope (SXT) on Solar-A Mission

Mission Purpose: The Solar-A mission, renamed as Yohkoh mission after its successful launch in August, 1991, is dedicated to the study of solar flares, especially of high-energy phenomena observed in the X- and gamma-ray ranges.

Soft X-ray Telescope (SXT): The instrument is a grazing-incidence reflecting telescope for the detection of X-ray in the wavelength range of 3-60 Angstrom. It uses a 1024x1024 CCD detector array to cover the whole Solar disk. Data acquired from the CCD is of 12-bit quantization and is processed on board to provide 8-bit telemetry data. The image in Figure 4.7 is an averaged image of size 512x512 with dynamic range up to 15 bits in floating point format as a result of further ground processing.

Compression Study: The test image is first rounded to the nearest integer. Then a 1D default predictor is applied to this seemingly high-contrast image. A compression ratio of 4.69 is achieved, which means that only 3.2 bits are needed per pixel to provide the full precision of this 15-bit image.

4.8 Goddard High-Resolution Spectrometer (GHRS) on HST

Mission Purpose: Same as Section 3.6.

Goddard High-Resolution Spectrometer (GHRS): The primary scientific objective of this instrument is to investigate the interstellar medium, stellar winds, evolution, and extra-galactic sources. Its sensors are two photo-diode arrays optimal in different wavelength regions, both in the UV range. Figure 4.8 gives examples of a typical spectrum and a background trace. These traces have 512 data values; each is capable of storing digital counts in a dynamic range of 10^7. Spectral shapes usually do not vary much. However, subtle variations of the spectrum in such large dynamic range presents a challenge for any lossless data compression algorithm. Two different prediction schemes are applied: the first uses the previous sample within the same trace, and the second uses the previous trace in the same category (spectrum or background) as a predictor. The results are summarized below in Table 4.8. The test data set has data maximum value smaller than 16-bit dynamic range, the compression ratio was thus calculated relative to 16 bits.

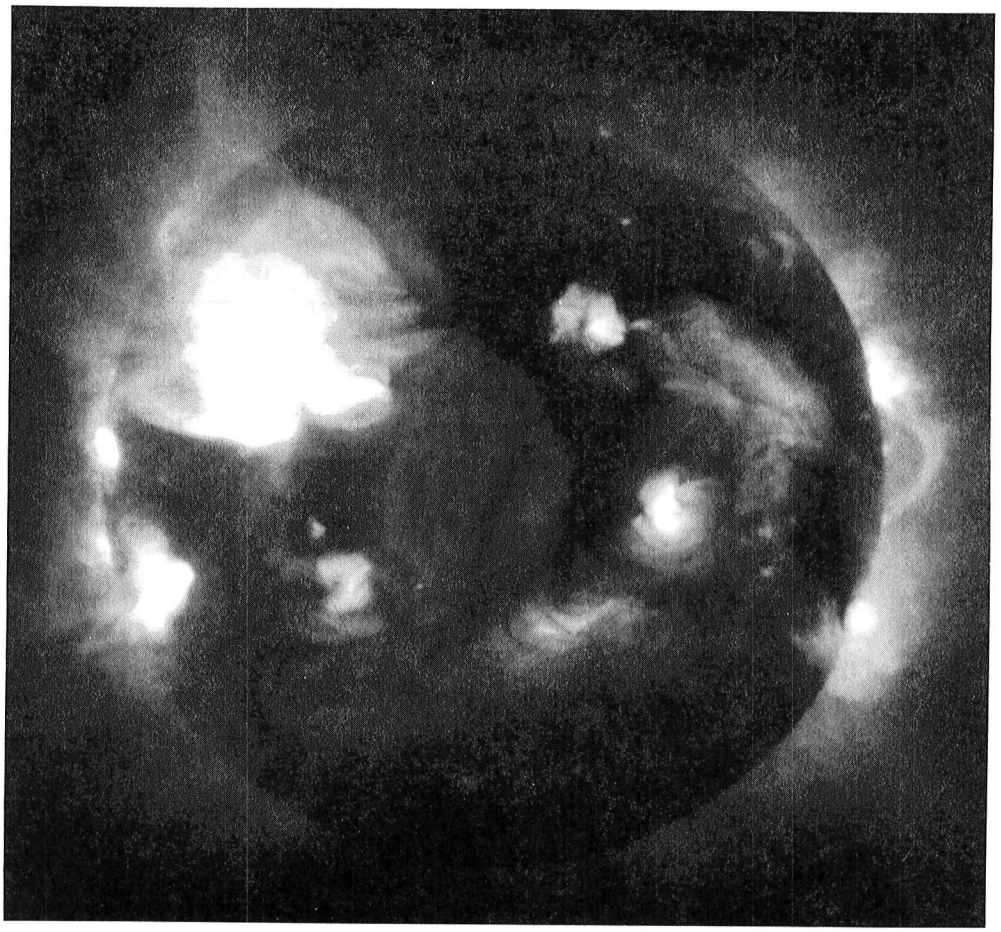


Figure 4.7. Solar-A X-ray image.

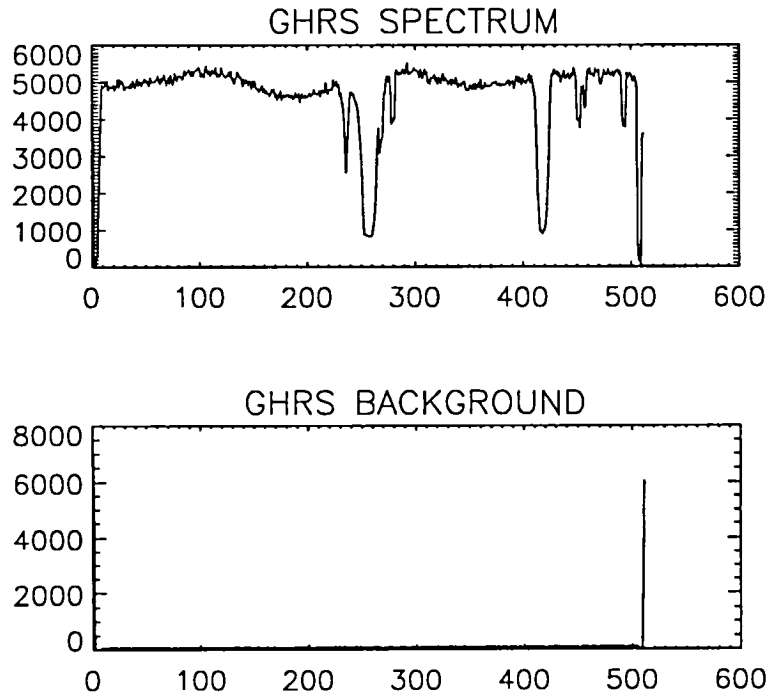


Figure 4.8. GHRS data samples.

Table 4.8. Compression Ratio for GHRS

Trace	Predictor	USES	LZW
spectrum 1	previous sample	1.64	<1.0
spectrum 2	previous sample	1.63	<1.0
spectrum 2	spectrum 1	1.72	NA
background 1	previous sample	2.51	1.63
background 2	previous sample	2.51	1.68
background 2	background 1	1.53	NA

4.9 Acousto-Optical Spectrometer (AOS) on Submillimeter Wave Astronomy Satellite (SWAS)

Mission Purpose: The Submillimeter Wave Astronomy Satellite (SWAS) is a Small Explorer (SMEX) mission, scheduled for launch in the summer of 1995 aboard a Pegasus launcher. The objective of the SWAS is to study the energy balance and physical conditions of the molecular clouds in the Galaxy by observing the radio-wave spectrum specific to certain molecules. It will relate these results to theories of star formation and planetary systems. The SMEX platform provides relatively inexpensive and shorter development time for the science mission. SWAS is a pioneering mission that packages a complete radio astronomy observatory into a small payload.

Acousto-Optical Spectrometer: The AOS utilizes a Bragg cell to convert the radio frequency energy from the SWAS submillimeter receiver into an acoustic wave, which then diffracts a laser beam onto a CCD array. The sensor has 1450 elements with 16-bit readout. A typical spectrum is shown in Figure 4.9(a). An expanded view of a portion of two spectral traces is given in Figure 4.9(b). Because of the detector nonuniformity, the difference in the Analog-to-Digital Converter (ADC) gain between even-odd channels, and effects caused by temperature variations, the spectra have nonuniform offset values between traces, in addition to the saw-tooth-shaped variation between samples within a trace. Because of limited available onboard memory, a compression ratio of over 2:1 is required for this mission.

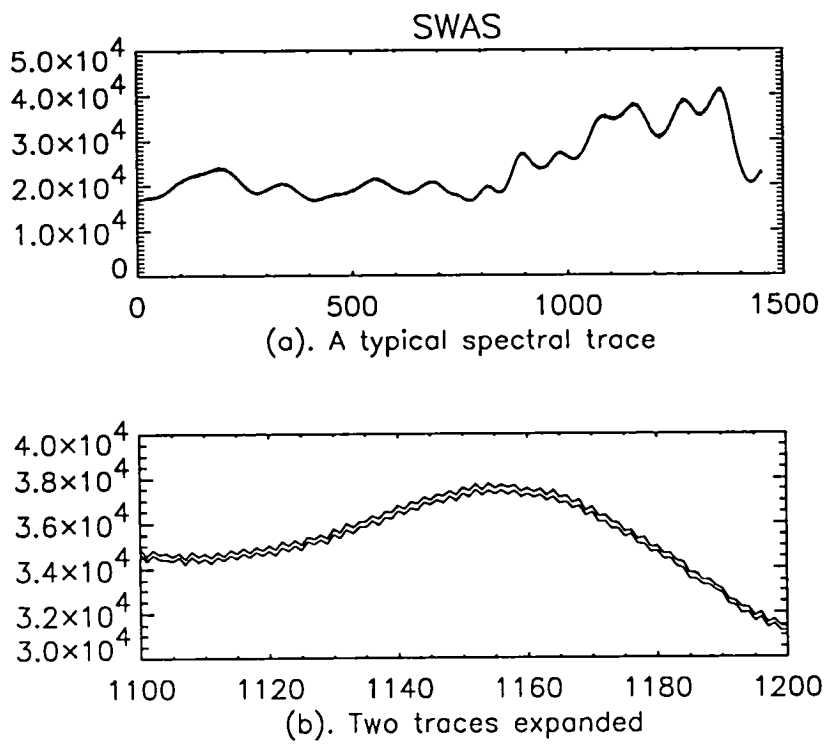


Figure 4.9. AOS radio-wave spectrum.

Compression Study: The large dynamic range and the variations within each trace present a challenge to the compression algorithm. Default prediction between samples is ineffective when the odd and even channels have different ADC gains. Three prediction modes are studied; the results are given in Table 4.9.

Table 4.9 SWAS Compression Result

Predictor	USES
previous sample	1.58
previous trace (external predictor mode)	1.61
previous trace (multispectral mode)	2.32

The result shows that even with the similarity in spectral traces, a predictor using an adjacent trace in a direct manner will not improve the compression performance caused by the uneven background offset. The multispectral predictor mode is especially effective in dealing with spatially registered multiple data sources with background offsets. In contrast, the LZW compression on the same spectral trace achieves no compression.

4.10 Gamma-Ray Spectrometer on Mars Observer

Mission Purpose: The Mars Observer was launched in September 1992. The Observer will collect data through several instruments to help the scientists understand the Martian surface, atmospheric properties and the interactions between the various elements involved.

Gamma-Ray Spectrometer (GRS): The spectrometer uses a high-purity germanium detector for gamma rays. The flight spectrum is collected over sixteen thousand channels, each corresponding to a gamma-ray energy range. The total energy range of a spectrum extends from 0.2 Mev to 10 Mev. Typical spectra for a 5-second and a 50-second collection time are given in Figure 4.10.1. These spectra show the random nature of the count; some are actually zero over several bins. The spectral count dynamic range is 8-bit.

Compression Study: Two different schemes using the same compression algorithm have been simulated. One scheme applies the USES algorithm directly to the spectrum, and the other implements a two-layer coding structure.

Direct application of USES (one-pass scheme): In this single pass implementation, the block size J is set at 16, and the entropy-coding mode is selected to bypass the predictor and the mapping function; the algorithm achieves over 20:1 lossless compression for a 5-second spectrum. As gamma-ray collection time increases, the achievable compression decreases

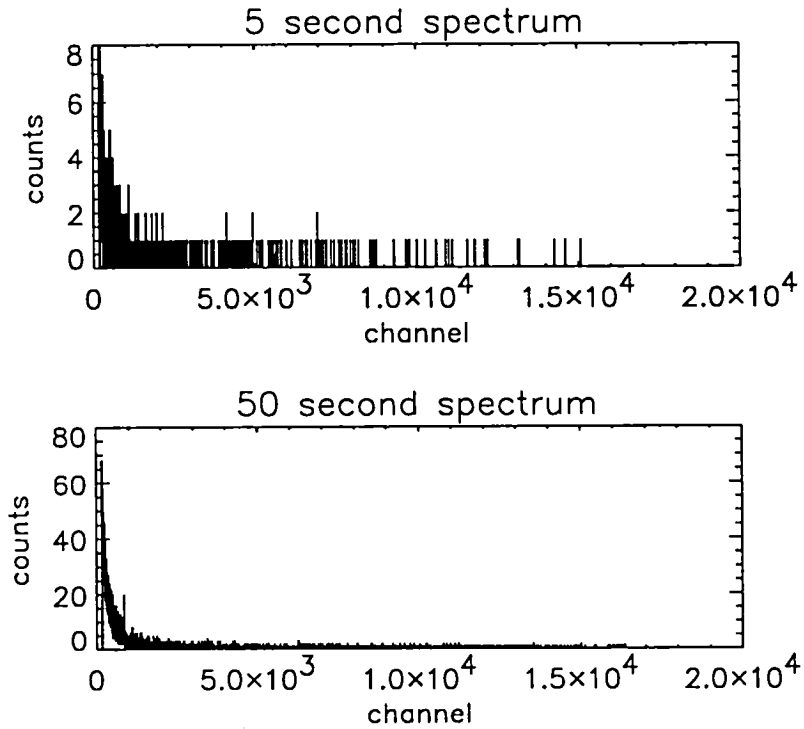


Figure 4.10.1. Gamma-ray spectrum.

Two-layer scheme: In a two-layer scheme, two passes are needed to compress the data. The first pass finds the channel numbers that have valid counts, and generates run-length of channels between them. Meanwhile, a count file is created which holds only the valid counts as data in the file. In the second pass, both the channel run-length file and the count file are compressed using the lossless compression algorithm at a block size of 16 and using the 1D default prediction mode.

The results of both schemes are plotted in Figure 4.10.2 and compared to the projected results from the actual implementation on the GRS instrument on the Mars Observer. The actual implementation used a different layered structure to code the valid channels and counts and also employed a different low-entropy coding option.

The GRS spectrum offers a specific example that requires an efficient source coding technique for low entropy data. For the 5-second spectrum, over 90% of the data are coded with the low-entropy option.

Lossless compression using LZW algorithm on the gamma-ray spectrum data produces compression ratios slightly lower than those from the one-pass application of USES at all different gamma-ray collection time.

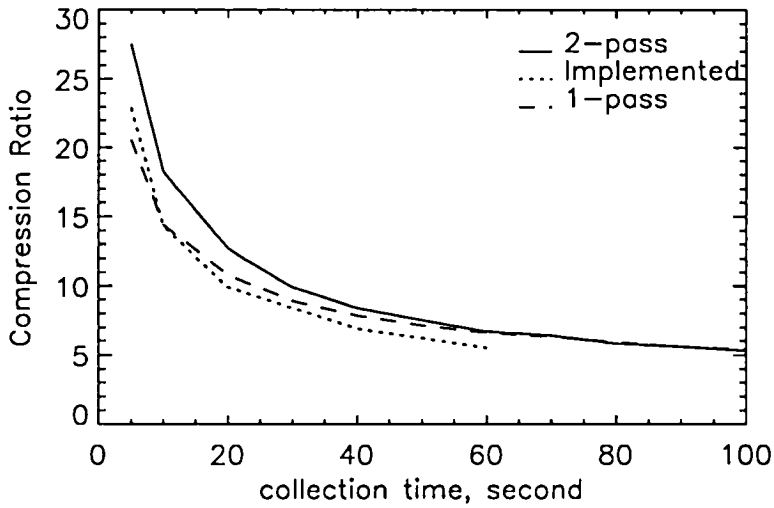
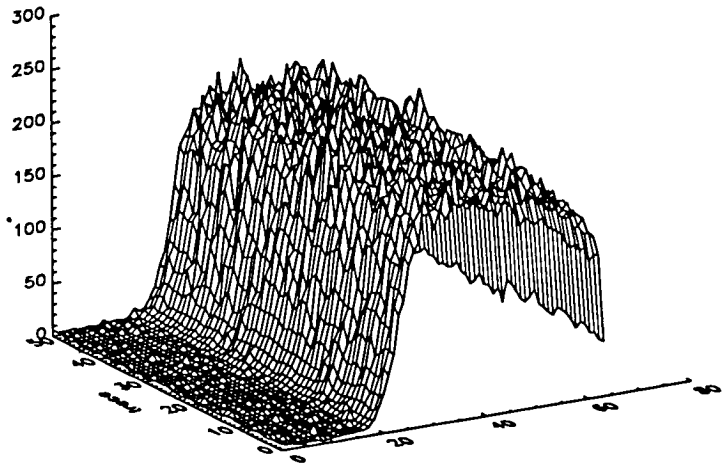


Figure 4.10.2. Compression result on GRS data.

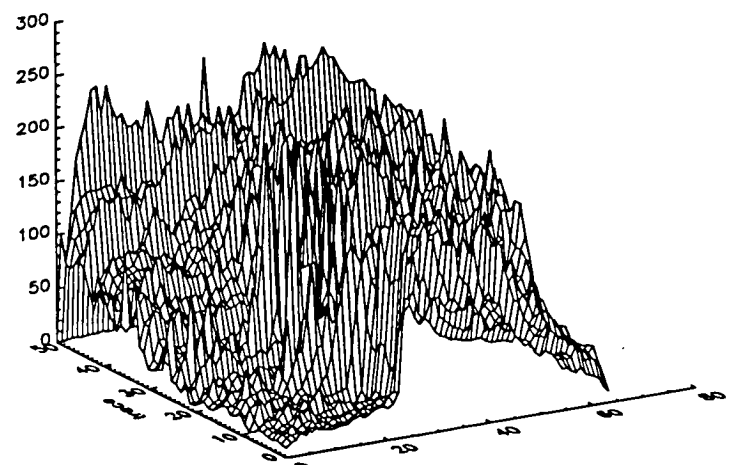
4.11 Radar Altimeter on TOPEX/POSEIDON

Mission Purpose: Launched in August 1992, the TOPEX/POSEIDON satellite will make the most accurate measurements ever of the sea surface using the satellite's radar altimeter, one of the six instruments on the spacecraft. The mission will help the scientists understand the interaction between ocean, atmosphere, and global climate.

Radar Altimeter: The dual-frequency radar altimeter measures the distance to the ocean surface by measuring the time it takes for a radar signal to reflect off the water surface and return to the spacecraft. Subsequent computations and corrections permit determination of the sea level to within an accuracy of a few centimeters. The altimeter data traces are inherently very noisy. A typical set of traces over water and land is given in Figure 4.11. These are all 8-bit data, with 64 samples for every trace.



Altimeter traces over water



Altimeter traces over land

Figure 4.11. Typical radar altimeter traces.

Compression Study: Lossless compression study was performed on these two data sets by setting a default prediction mode that uses the previous sample as predictor. The results are summarized in the following table:

Table 4.11. Compression Ratio for Altimeter Data

File	USES	LZW
water	1.43	1.08
land	1.40	< 1

Using a predictor across the trace does not provide consistent improvement over using the default predictor within trace.

4.12 Gas Chromatograph-Mass Spectrometer on Cassini

Mission Purpose: The Cassini mission includes a Saturn Orbiter and the Huygens Probe which will descend into the atmosphere of Titan, one of Saturn's moons. The launch is scheduled for 1997 to allow a Saturn encounter in 2004. The objectives of this mission include more detailed studies of Saturn's atmosphere, rings, and magnetosphere, close-up studies of Saturn's satellites, and investigations on Titan's atmosphere and surface.

Gas Chromatograph-Mass Spectrometer (GCMS): The instrument uses chromatography to separate gas molecules of different structures and the mass spectrometer to obtain mass distribution. Some typical GCMS traces are shown in Figure 4.12 in which the time-axis denotes different traces and the mass axis shows the mass distribution within a trace. Onboard data integration and other processing may be needed for this instrument besides data compression. This particular test set has a dynamic range of 16 bits.

Compression Study: Simulation was performed with several different configurations. The first one uses a predictor in the same trace, that is, by using a mass channel as predictor for the next mass channel; the second uses a predictor across the trace. The latter case has two data configurations for compression: when a block of 16 data is taken in the mass direction, then a buffer of one trace is needed to hold the previous trace as predictor for the next trace; when a block of data is taken in the time direction, a buffer large enough to hold 16 traces is needed. The results obtained are tabulated in Table 4.12.

Table 4.12. Compression Ratio on GCMS Data

Predictor Mode	Compression Direction	USES
1D default	mass axis	1.65
previous trace	mass axis	1.93
previous trace	time axis	1.96

Using LZW compression on the same file would achieve a compression ratio of 1.22.

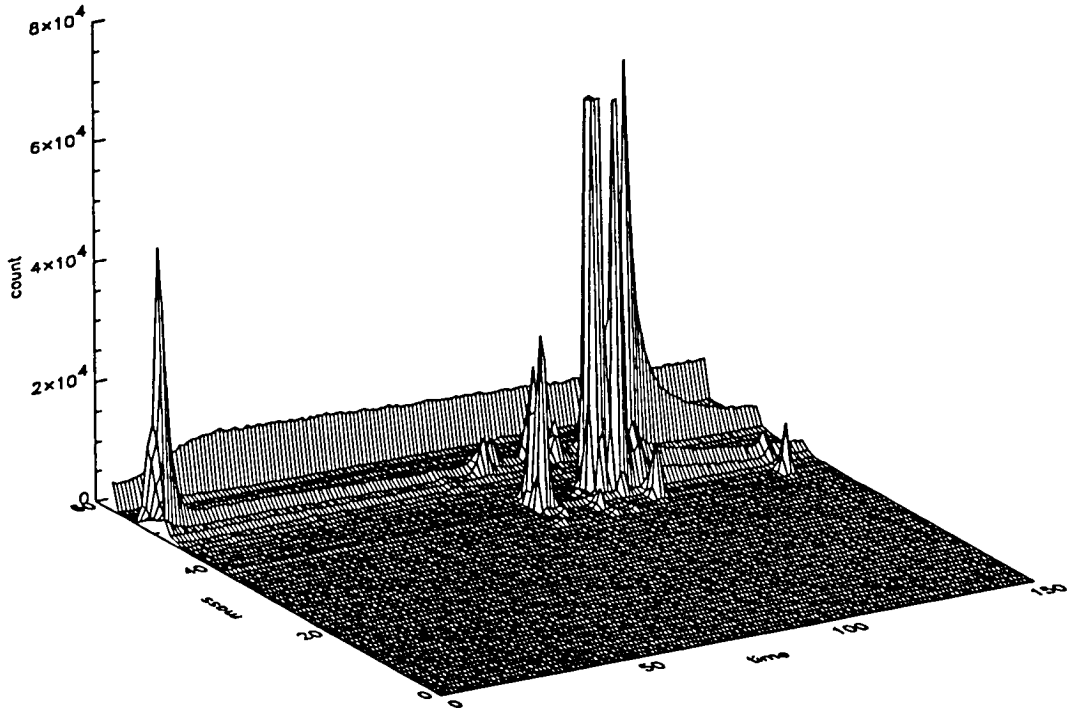


Figure 4.12. Typical GCMS test data.

5.0 References

1. Rice, Robert F. , Pen-Shu Yeh and Warner H. Miller, "Algorithms for a Very High Speed Universal Noiseless Coding Module," *JPL Publication 91-1*, 1991. A shortened version "Algorithms for High-Speed Universal Noiseless Coding," is published in the *Proceedings of the AIAA Computing in Aerospace 9 Conference,* San Diego, CA, Oct. 19-21, 1993.

2. Yeh, Pen-Shu, Robert F. Rice and Warner H. Miller, "On the Optimality of Code Options for a Universal Noiseless Coder," Revised version, *JPL Publication 91-2,* 1991. A shortened version "On the Optimality of a Universal Noiseless Coder," is published in the *Proceedings of the AIAA Computing in Aerospace 9 Conference*, San Diego, CA, Oct. 19-21, 1993.

3. Yeh, Pen-Shu, and Warner H. Miller, "The Implementation of A Lossless Data Compression Module in An Advanced Orbiting System: Issues And Tradeoffs," in *Proceedings of the First ESA Workshop on Computer Vision and Image Processing for*

Spaceborne Applications, ESTEC, Noordwijk, The Netherlands, June 10-12, 1991. Post-print available from the authors at Goddard Space Flight Center, Greenbelt, MD, 20771, USA.

4. Venbrux, Jack, Pen-Shu Yeh and Muye N. Liu, "A VLSI Chip Set for High-Speed Lossless Data Compression," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 2, No. 4, Dec. 1992.
5. "Universal Source Encoder - USE" Product Specification, Version 4.1, Microelectronics Research Center, University of New Mexico, 1991, and "Universal Source Encoder for Space - USES," Preliminary Product Specification, Version 2.0, MRC, University of New Mexico, 1993.
6. Software simulator, available from the Microelectronics Research Center, University of New Mexico.
7. "Advanced Orbiting Systems: Networks and Data Links: Architectural Specification," CCSDS 701.0-B-1 Blue Book, Oct. 1989.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1993	Technical Paper

4. TITLE AND SUBTITLE

Application Guide for Universal Source Encoding for Space

5. FUNDING NUMBERS

730
310-20-46

6. AUTHOR(S)

Pen-Shu Yeh and Warner H. Miller

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS (ES)

Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PEFORMING ORGANIZATION REPORT NUMBER

94B00028

9. SPONSORING / MONITORING ADGENCY NAME(S) AND ADDRESS (ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING ADGENCY REPORT NUMBER

NASA TP-3441

11. SUPPLEMENTARY NOTES

Yeh and Miller: NASA/Goddard Space Flight Center, Greenbelt, Maryland

12a. DISTRIBUTION / AVAILABILITY STATMENT

Unclassified - Unlimited
Subject Category 32

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Lossless data compression has been studied for many NASA missions. The Rice algorithm has been demonstrated to provide better performance than other available techniques on most scientific data. A top-level description of the Rice algorithm is first given, along with some new capabilities implemented in both software and hardware forms. The document then addresses systems issues important for onboard implementation, including sensor calibration, error propagation, and data pecketization. The latter part of the guide provides twelve case study examples drawn from a broad spectrum of science instruments.

14. SUBJECT TERMS

Source coding, Universal source encoding, data compression

15. NUMBER OF PAGES

42

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL