Final Report: Lossless Compression of Instrumentation Data

Samuel D. Stearns

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-94AL85000

Approved for public release; distribution is unlimited.
FINAL REPORT:
LOSSLESS COMPRESSION
OF INSTRUMENTATION DATA

Samuel D. Stearns
Materials Radiation Science Department
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

This is our final report on Sandia National Laboratories Laboratory-Directed Research and Development (LDRD) project 3517.070. Its purpose has been to investigate lossless compression of digital waveform and image data, particularly the types of instrumentation data generated and processed at Sandia Labs. The three-year project period ran from October 1992 through September 1995.

This report begins with a descriptive overview of data compression, with and without loss, followed by a summary of the activities on the Sandia project, including research at several universities and the development of waveform compression software. Persons who participated in the project are also listed.

The next part of the report contains a general discussion of the principles of lossless compression. Two basic compression stages, decorrelation and entropy coding, are described and discussed. An example of seismic data compression is included.

Finally, there is a bibliography of published research. Taken together, the published papers contain the details of most of the work and accomplishments on the project. This final report is primarily an overview, without the technical details and results found in the publications listed in the bibliography.
Acknowledgement

The author gratefully acknowledges the support of all who have contributed to the project. These are mainly the personnel listed in the Summary of Activities. At Sandia, special recognition is also due to J.A. Davis and G.J. Simmons for solving a difficult problem in binary sequences early in the project. The solution was essential to the development of bi-level coding.

In the author's experience, funds spent on research at U.S. universities have always yielded rich rewards in terms of useful results as well as numbers of man-hours. In the present case, we have spent less than the cost of one Sandia FTE (that is, one Sandia man-year), and have obtained several man-years of research by talented researchers at three universities. The author is especially grateful to professors N. Magotra and N. Ahmed, and to students Giridhar Mandyam, Wes McCoy, and Dr. Li Zhe Taa at UNM, to professor W.B. Mikhael and students Yousef Nijim, Alan Berg, Shomit Ghosh, and Arun Ramaswamy at UCF, and to professor D.M. Etter and student Jonathan Haines at UC Boulder. The work and ideas of these individuals constitute the major portion of this project. The author is also grateful for the interest and contributions of other faculty, including professors R.L. Kirlin at the University of Victoria, G. Coutu at the Hartford Graduate Center, and M. Fargues at the Naval Postgraduate School, Monterey. In addition, the author is grateful to Dr. C.R. Hutt and Ms. C.V. Peterson at the USGS Albuquerque Seismological Laboratory for their contributions of data and compression studies.

Finally, in Organization 9300, the author thanks Frank Nutt for rescuing the project from financial confusion by providing timely and accurate financial data for the past year. At Sandia, such data has been difficult to obtain. Also, thanks go to Mark Hedemann, Manager, Dept. 9311, for providing leadership and support for the final phase of the project.
Contents

Introduction ............................................................... 1
Summary of Activities ..................................................... 2
Compressibility ............................................................. 3
Compression in Two Stages ................................................. 5
Predictive Coding ........................................................... 6
Residue Coding ............................................................... 7
Transform Coding ............................................................ 8
Example: Seismic Waveform Compression .............................. 8
Conclusions ................................................................. 13
Bibliography ................................................................. 13
Distribution ................................................................. 20

Figures

1. Lossy to lossless conversion ............................................. 2
2. Examples of waveforms with different compressibility .................. 5
3. Lossless waveform compression in two stages .......................... 5
4. General form of the prediction stage in Fig. 3 ......................... 6
5. Seismic waveform, $i_x(0:59999)$, with amplitude and power distributions .......... 9
6. Residues of $i_x(0:59999)$ in Fig. 5 after predictive coding, with amplitude and power distributions ............................................. 10
7. Relative frequency of occurrence of residues in the first frame of $i_r$ in Fig. 6 .......... 11
8. Bits per sample for each of 60 seismic data frames, showing compression in the first (linear prediction) stage, from $bps_{i_x}$ to $bps_{i_r}$, and in the second (arithmetic coding) stage, from $bps_{i_r}$ to $bps_{b}$ ............................................. 12
9. Compression ratio for each of 60 seismic data frames, with $bps_{i_x} = 20$ ................ 12
Introduction

Techniques for data compression have been in use for a long time. The oldest techniques are those used for the compression of text composed from a fixed alphabet of symbols. The Morse code, in which symbols used more frequently generally have shorter codes, is an example of text compression using fixed-symbol coding.

Compression of data is said to be lossless if, in an error-free environment, the original data can be recovered exactly from the compressed data. Thus, for our purposes, "loss" refers only to the compression process itself, and not to other types of loss due to communication channel noise, storage dropouts, or other causes. All fixed-symbol coding and other text compression methods are lossless in this sense. If the compression technique is such that the recovered (decompressed) data is only an approximation to the original data, then the technique is lossy.

This is the final report on a Laboratory-Directed Research and Development (LDRD) project to investigate lossless compression of digitized instrumentation data. Traditional waveform data compression techniques, that is, speech and video compression techniques, are lossy. The decompressed signal is not an exact replica of the original signal, even though it may sound the same to an average listener or look the same to an average viewer. The purpose of this project has been to develop and extend new data compression techniques that are lossless, that is, that allow exact recovery of the original data. The new lossless compression techniques are applicable specifically to the large field test and telemetry instrumentation data bases at Sandia, and generally to similar data bases where lossy compression is not acceptable. They are also applicable to real-time communications in space and global surveillance systems, where again exact recovery of the original signals is essential, and bandwidth is at a premium. Furthermore, they are applicable to image compression in areas such as medical x-rays, where exact data must be preserved for legal reasons, and in satellite and radar imaging cases wherever exact image recovery is required.

As one would expect, less compression is possible in general when exact recovery is required, that is, with lossless compression. Figure 1 shows a simple method for converting from lossy to lossless compression. If the "recovery" process shown in the figure is used on the compressed waveform as shown, and if the "difference" in the figure is added to the recovered version of the compressed waveform, the result will be an exact replica of the input waveform. Thus, to achieve exact recovery, the compressed data must contain the difference in addition to the compressed waveform, and so there is less overall compression. Lossless compression is not done exactly as in Fig. 1, but it is clear that less compression is the usual result.
Nevertheless, significant savings in bandwidth or storage space are possible with instrumentation data. For example, with seismic and ground motion data, which has comprised most of our experimental data, storage requirements have been reduced to less than 12% of the original amount.

More precise and complete experimental results have been reported, and some examples are included later in this report. First, we provide a summary of the activities on the project, and describe the principles and basic theory of lossless waveform compression.

Summary of Activities

Funded research on lossless compression has been conducted at four locations: Sandia Labs, the University of Colorado at Boulder, the University of Central Florida at Orlando, and the University of New Mexico at Albuquerque. These universities were selected partly because, in each case, there was a key faculty member who was already recognized nationally for work in digital signal processing and signal compression. In addition, contributions have been made by other faculty and students at several universities who have become interested in the work. A brief summary of the personnel and developments at each sponsored location follows. In most cases, the relevant references in the Bibliography are noted.

Sandia
Originated and developed two-stage lossless compression technique [J6,J11,J12].
Developed and tested lossless linear predictive coding [P30].
Developed and applied new bi-level and arithmetic coding techniques [J6,J12].
Implemented two-stage compression in a Fortran source library, CMPLIB.FOR (DOE copyright July 1995).
Implemented a companion technique for data authentication using cyclic codes.
Compressed and authenticated seismic data base for Test Information Program [P30].
Compressed and authenticated satellite telemetry data for Dept. 9225.
Sponsored faculty/student research at three universities. Six dissertations [D1-D6]. Organized sessions on compression at international conferences [P7-P13,P18-P26].

**UC**

Personnel: Prof. D.M. Etter, students J.W. Haines and M. Coffey.
Studied lossless waveform compression using wavelets and multirate filters [P19,P25].
Translated the Fortran source library, CMPLIB, into the C programming language.

**UCF**

Personnel: Prof. W.B. Mikhael, students Y.W. Nijim, A. Ramaswamy, A.P. Berg, and S.M. Ghosh
Surveyed techniques for lossless compression [D1,P21].
Developed new differentiation methods for lossless predictive coding [D3,J4,P3,P4,P20].
Applied differentiation methods to lossless image compression [P4,P24].
Studied lossless waveform and image compression using mixed transform coding [D4,D5,C1,P3,P11,P16,P24,P29].

**UNM**

Developed basic theory for optimizing lossless predictive coding [D2,D6,J11,P17].
Developed an adaptive lattice scheme for lossless prediction [D2,P17].
Studied codebook design and other methods for lossless image compression [P6].
Studied real-time implementations of lossless waveform compression [P12].
Developed a DCT-based scheme for lossless image compression [P13].
Wrote compression section for CRC Industrial Electronics Handbook [B1].

Contributions from universities not sponsored by the LDRD project, notably by R.L. Kirlin at the University of Victoria, G. Coutu at the Hartford Graduate Center, and M. Fargues at the Naval Postgraduate School, can be seen by referring to the conference papers in the Bibliography [P10,P14,P18,P31,P32].

**Compressibility**

In this section of the final report we begin a brief general discussion of lossless waveform compression. The purpose here is to provide some basic concepts that help us to decide which waveforms are compressible and which are not, and to decide theoretically how much we may be able to compress a given waveform.

Our principal measure of compressibility is the *compression ratio*, which is defined as the number of bits in the original data divided by the number of bits in the compressed data. Suppose the original waveform data consists of $i(x(0:K-1) = [ix_k]$, a vector with $K$ integer elements, and that the compressed version of $ix$ is $iy(0:N_y-1)$, a vector with $N_y$ binary elements (bits). Then the compression ratio is the number of bits in $ix$ divided by $N_y$. 

3
In the literature on waveform compression, the number of bits in $i_x$ is not measured consistently. A conservative measure, which we will use here, is $K$ times the number of bits needed to represent the sign and magnitude of the sample with the largest magnitude in $i_x$. Thus, if $[\cdot^*]$ stands for "the least integer equal to or greater than ($\cdot$)", then

$$\text{Number of bits in } i_x = K \text{bps}_{ix} = K \left(1 + \left\lceil \log_2|i_{x_{\max}}| \right\rceil \right)$$  \hspace{1cm} (1)$$

The number of bits per sample in $i_x$, $\text{bps}_{ix}$, is found simply by dividing (1) by $K$, and the number of bits per sample in the compressed vector, $\text{bps}_{y}$, is $N_y/K$. With these definitions, the compression ratio ($CR$) is expressed as

$$CR = \frac{\text{bps}_{y}}{\text{bps}_{ix}} = \frac{1 + \left\lceil \log_2|i_{x_{\max}}| \right\rceil}{N_y/K}$$ \hspace{1cm} (2)$$

When comparing results using the compression ratio, it is important to consider not only the data being compressed, but also the definition of the compression ratio. If, for example, the length in bits of an ASCII file is used in place of (1) for the length of $i_x$, the compression ratio will appear to be much larger than the value of $CR$ computed using (2).

Assuming a consistent measure, the value of $CR$ depends on the compression scheme and the data being compressed. In many practical cases, such as text, speech, and video coding, the maximum compression ratio is based on the amount of useful information in the signal, and the definition of "useful" can be very subjective. In lossless compression, on the other hand, objective statements are more easily made about the compressibility of signals in general, and about the maximum compression ratio that can be achieved with particular signals.

Some examples of different degrees of compressibility are shown in Fig. 2. In general, excluding special "codebook" schemes where there is a priori knowledge of the waveform structure [P6], there are two statistical properties that determine the compressibility of a waveform: the distribution of waveform amplitudes, and the distribution of the waveform power in the frequency domain. As suggested by Fig. 2, the degree of compressibility depends on the degree of nonuniformity of each of these distributions. The first waveform, which is uniform white noise, is essentially incompressible. The second, white noise with a nonuniform amplitude distribution, is somewhat compressible. The third and fourth waveforms, with nonuniform power distributions, are highly compressible.

Typical instrumentation and telemetry waveforms do not usually have uniform amplitude and power distributions. A uniform distribution of power in the frequency domain translates to an impulsive correlation function in the time domain, that is, to uncorrelated data, and instrumentation data is typically correlated, and therefore compressible.
Compression in Two Stages

In this and the next three sections we complete the general discussion of lossless waveform compression by discussing the two-stage process shown in Fig. 3. We have focused on the basic two-stage concept throughout the project, and have developed computer codes that implement the scheme in Fig. 3 with linear predictive coding and arithmetic residue coding.

![Figure 3. Lossless waveform compression in two stages.](image-url)
The purpose of the first stage in Fig. 3 is to decorrelate the input sequence, \( ix(0:K-1) \), and produce a residue sequence, \( ir(M:K-1) \), the latter having a more or less flat power density spectrum, in an exactly reversible manner. In this operation, predictive coding has the effect of removing power from \( ix \), so that the ratio of standard deviations, \( \sigma_{ix}/\sigma_{ir} \), is generally greater than one. In fact, we usually take \( \sigma_{ir}/\sigma_{ix} \) as a measure of success of the first compression stage. With reversible linear predictive coding, for some types of data, \( \sigma_{ix}/\sigma_{ir} \) can be several orders of magnitude.

The purpose of the second stage is to compress the residue sequence, \( ir(M:K-1) \), into the binary sequence \( iy(0:N_r-1) \). As suggested in Fig. 2, \( ir \) is compressible if its amplitude values are distributed nonuniformly, which is generally the case for the residue sequence from a predictive coder.

**Predictive Coding**

At Sandia we have implemented the first stage in the form of a recoverable linear predictive coding process based on traditional linear prediction. This form has been the standard of comparison for more recent developments during the course of the project, some of which have been more successful with given types of data. In general, the predictive coding stage in Fig. 3 must be in the form of Fig. 4. As long as the unit delays \( z^{-1} \) are used as shown, the process of producing \( ir \) form \( ix \) is predictive. Furthermore, it is not difficult to see that the process is reversible, that is, given an identical "filter" for recovery, \( ix \) can be recovered exactly from \( ir \). The "filter" may be linear or nonlinear, may be recursive or nonrecursive, may or may not use memory, and may be fixed or time-varying. The goal in lossless compression is to discover filters that best decorrelate \( ix \) and thereby maximize the power ratio, \( \sigma_{ix}/\sigma_{ir} \), in different situations.

Several filter schemes were studied during the course of the project. Various differentiation techniques were investigated by Mikhael, Nijim, and others [J4,P3,P4,P20].

![Figure 4. General form of the prediction stage in Fig. 3.](image-url)
Nijim and Mikhael also studied a recursive modeling scheme in the form of Fig. 4 [P8]. Various forms of adaptive filters were investigated by Coutu and Fargues [P10] and by McCoy and Magotra [P17]. An adaptive lattice structure by the latter appears promising at this time. The use of multirate filter banks, which amounts to a wavelet transformation, was studied by Etter, Haines, and Coffey [P9, P25], and some linear two-dimensional predictor forms were studied by Ghosh and Mikhael [P23]. Much of this work is still proceeding, and will continue past the end of the project at Sandia.

Residue Coding

Given that the first stage produces a decorrelated residue sequence, \( i_r \), the objective of the second coding stage is to produce a binary sequence (i.e., in Fig. 3) with a length, \( N_y \), approaching the residue sequence length, \( K-M \), times the entropy in bits per sample of the residue sequence, \( i_r \). The entropy of \( i_r \) depends entirely on the amplitude distribution of \( i_r \) and is given by

\[
\text{Entropy of } i_r = -\sum_{i_r_{\text{min}}}^{i_r_{\text{max}}} f(i_r) \log_2 f(i_r)
\]

where \( f(i_r) \) is the frequency of occurrence of \( i_r \) in the sequence \( i_r(M:K-1) \), such that the frequencies sum to one.

As suggested by the examples in Fig. 2, a uniform distribution of \( i_r \) amplitudes, where all frequencies are equal, produces maximum entropy. That is, if the range of \( i_r \) from \( i_r_{\text{min}} \) through \( i_r_{\text{max}} \) is \( R \), and if all values of \( i_r \) appear equally often in \( i_r(M:K-1) \), then the entropy in (3) is \( \log_2 R \) bits per symbol. When the values of \( i_r \) are distributed uniformly, second-stage coding cannot be expected to compress the residue sequence, no matter which scheme is chosen.

At the other extreme, if one value of \( i_r \) appears in the residue sequence nearly always, so that \( f(i_r) \) is close to one for this value and near zero for other values of \( i_r \), then the entropy in (3) is close to zero. In this case, almost any second-stage coding scheme can be expected to produce significant compression of the residue sequence.

In the practical applications we have studied, with linear predictive coding as the first compression stage, the distribution of \( i_r \) has been essentially Gaussian, which is a compromise between the two extremes just mentioned. At the beginning of the project we developed a bi-level coding technique [D6, J11, J12, P32, P34] which is relatively simple, but performs suboptimally with Gaussian residue data. More recently, we have developed a practical method to implement the second compression stage with a form of arithmetic coding [D2, D6, J6]. This method produces near-entropy coding, that is, nearly the minimum number of bits per sample given by (3), and represents an improvement over the bi-level method. It also appears to work well with non-Gaussian residue sequences.
Transform Coding

Transform coding is an alternative to the two-stage process in Fig. 3 that is often used in lossy waveform and image compression. In lossy transform coding the waveform, \(i(x:K-1)\), is first transformed using the Discrete Cosine Transform (DCT) or some other transform. Then the set of transform values is compressed, typically via quantization or by eliminating "unimportant" components. With narrow-band data, for example, DCT components outside the band can be eliminated without compromising the data. In some cases, the compression process may even increase the signal-to-noise ratio. In decompression, the approximation to \(i(x)\) is then reconstructed from the compressed transform data.

Applications of transform coding have been studied at UCF and UNM [D4,D5,C1,P3,P11, P13], but to date we have not seen superior results when the coding is made truly lossless. To make transform coding lossless we must revert essentially to the scheme in Fig. 1, and include a difference signal, which may itself be compressed. This added burden of data has the effect of increasing \(N_y\) and reducing the compression ratio in (2).

Example: Seismic Waveform Compression

We include here a single example of lossless seismic waveform compression to illustrate the operation of each of the two compression stages with real data. The papers listed in the Bibliography contain many examples of both waveform and image compression. Several papers contain statistics on the compression of waveforms in a large seismic data base [J6,P30,P33].

A waveform from a short-period vertical seismometer, courtesy of C.R. Hutt, USGS Albuquerque Seismological Laboratory, is shown in Fig. 5. The segment shown has a total duration of 3000 s and was recorded at Albuquerque during the Loma Prieta Earthquake near Santa Cruz, California in October, 1989. The recording rate was 20 samples/s.

Amplitude and power distributions are shown below the plot of \(i(x)\) in Fig. 5. These represent statistics for the entire waveform. The amplitude distribution is from \(-5\cdot10^2\) to \(+5\cdot10^2\) in intervals of \(10^4\). The power density spectrum was taken by averaging magnitudes of DFT's of half-overlapping segments of size 256. We note that neither distribution is uniform and therefore, as discussed in connection with Fig. 2, the seismic waveform here is compressible.

The waveform \(i(x)\) in Fig. 5 was partitioned into 60 frames, each of size \(K=1000\), and compressed via the two-stage process in Fig. 3, using non-adaptive linear prediction in the first stage and non-adaptive arithmetic residue coding in the second stage. The value of \(K\) used here is not critical and is typical for nonstationary waveforms such as \(i(x)\) in Fig. 5.
The output of the first compression stage, $ir(M-K)$, is plotted continuously for all 60 frames in Fig. 6. The amplitude and power distributions in Fig. 6 were obtained as in Fig. 5, except that the amplitude interval in Fig. 6 is 1 instead of $10^4$. For each frame, $ir(M)$ was extended back through $ir(0)$ to give a more continuous plot. The values of $M$ ranged from 2 through 8 in the 60 frames. In this case the residue sequence, $ir$, is seen to be on the order of four orders of magnitude smaller than the original sequence, $ix$. Thus the ratio of standard deviations, $\sigma_\mu/\sigma_\nu$, which measures the performance of the first stage as discussed previously in connection with Fig. 3, indicates significant first-stage compression in this example.

The cause of the large compression from $ix$ in Fig. 5 to $ir$ in Fig. 6 can be seen by comparing the power density spectra. The spectrum of $ix$ varies through about 12 orders of magnitude, indicating that the samples of $ix$ are far from independent. The spectrum of $ir$, on the other hand, varies through only about one order of magnitude, indicating that the predictor has successfully decorrelated each frame of $ix$. Note that, in our implementations of predictive coding, $ix$ is exactly recoverable from $ir$ without numerical noise, given the few initial values of $ix$ along with the predictor weights for each frame.

After the first stage has produced $ir$, the second stage attempts to reduce the number of bits per sample, $bps_\nu$, which, in accordance with (1), is measured by

$$bps_\nu = 1 + \lceil \log_2 |ir_{\nu}^\nu| \rceil$$

(4)
With arithmetic coding it is possible to reduce $bps_r$ to $bps_y$, the number of output bits per sample, with the latter close to the entropy of $ir$ given by (3). As an illustration we take the results from the first data frame, that is, the encoding of the first $K=1000$ samples of $ir$ in Fig. 6.

The amplitude distribution for this first data frame is plotted in Fig. 7. The vertical scale is $f(ir)$, the relative frequency of each value of $ir$. The circles represent the values of $f(ir)$ for the data frame and the solid curve is the Gaussian envelope $N(0,\sigma_{ir})$, where $\sigma_{ir}=3.43$ is the measured standard deviation. Since $|ir|_{max}$, in this case is 13, we have $bps_{ir}=5$ in accordance with (4).

In Fig. 7 the amplitude distribution of $ir$ is seen to be approximately Gaussian, and the entropy, $H=3.788$, is approximately that of a zero-mean Gaussian amplitude distribution with standard deviation $\sigma_{ir}$, which is given approximately by Woodward [J6, Ref. 9] as

$$H_p = \log_2\sqrt{2\pi e \sigma_{ir}^2} = \log_2(4.133\sigma_{ir}) = 3.825$$

Therefore, we would expect the compression of $ir$ in the first frame to be from $bps_{ir}=5$ to around $bps_y=3.8$ bits per sample. As a matter of fact, due to the finite frame size and the non-Gaussian points around $ir=0$, the actual value was even lower at $bps_y=3.6$ bits per sample.

In general, the compressibility of a Gaussian variate has been shown to be between one
and two bits per sample over a wide range of $\sigma_w$ [D6,J6]. These results for the first data frame are typical of the other 59 frames of $ir$ in Fig. 6.

![Gaussian residues. $\|r\|_{max}=13$, sigma=3.43, $H=3.788$.](image)

**Figure 7.** Relative frequency of occurrence of residues in the first frame of $ir$ in Fig. 6; $K=1000$. Solid curve is the Gaussian envelope, $N(0,\sigma_w)$, where $\sigma_w=3.43$ is the measured standard deviation. $H=3.788$ is the entropy, measured as in (3).

Finally, compression results for this example are summarized in Fig. 8 by comparing bits per sample, $bps_{ir}$, $bps_{w}$, and $bps_{p}$, for all 60 frames of the waveform. These results are typical for continuous instrumentation data in the sense that decorrelation in the first stage typically yields more compression than entropy coding in the second stage. The compression ratio for each frame, $CR$ in (2), is computed for each frame using $bps_{ir}=20$, the number of bits per sample in (1) needed to store $ir$ without compression, and plotted in Fig. 9.

As noted under Compressibility, the compression measure used in Fig. 9 is a conservative measure not always used in the literature. For example, actual file sizes are sometimes used to compute the compression ratio, leading to results which are format-dependent, but which nearly always look more impressive than the results in Fig. 9. In the present example, the ASCII file sizes of different versions of the seismic waveform in Fig. 5 are as follows:

<table>
<thead>
<tr>
<th></th>
<th>File Name</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original data file:</td>
<td>ANMBHZ89.ASC</td>
<td>542,994 bytes</td>
</tr>
<tr>
<td>PKZIP compression:</td>
<td>ANMBHZ89.ZIP</td>
<td>223,534 bytes</td>
</tr>
<tr>
<td>WINZIP compression:</td>
<td>ANMBHZ89.WZP</td>
<td>195,888 bytes</td>
</tr>
<tr>
<td>Sandia compression (ENCOD5):</td>
<td>ANMBHZ89.CMP</td>
<td>39,379 bytes</td>
</tr>
</tbody>
</table>
Figure 8. Bits per sample for each of 60 seismic data frames, showing compression in the first (linear prediction) stage, from $bps_{in}$ to $bps_{ir}$, and in the second (arithmetic coding) stage, from $bps_{ir}$ to $bps_{fy}$.

Figure 9. Compression ratio for each of 60 seismic data frames, with $bps_{in}$ = 20.
If we divide the original file size by the compressed file size, we get the compression ratio

$$CR = \frac{\text{Original file size}}{\text{Compressed file size}} = 13.8$$

(6)

This result looks more impressive than the results in Fig. 9, but it represents the same compression process.

As a matter of interest, the sizes of "zipped" files using PKZIP, copyrighted 1989-1990 by PKWARE, Inc., Glendale, WI, and WINZIP, copyright 1991-1994 by NicoMak Computing, Inc., Bristol, CT, both of which are widely used compression programs, are also given above. The comparison of the sizes of ANMBHZ89.ZIP and ANMBHZ89.WZIP with the size of ANMBHZ89.CMP is not meant to be a criticism of PKZIP or WINZIP, which are the best programs the author has found for general file compression. It is only meant to show that, for lossless compression of continuous waveforms, the two-stage process in Fig. 3 can offer real advantages over other methods.

**Conclusions**

Thus we conclude this short final report on the LDRD project to develop lossless waveform and image compression. All significant results are available in the literature listed in the Bibliography. Most of these papers, as well as compression and authentication codes in Fortran and a compression program in C, are available from the author.

**Bibliography**

Nearly all of the papers and presentations listed here were supported by the LDRD project. A few earlier items that predated and led up to the project are also listed for the sake of completion. The reader can identify these by noting that the LDRD project began in October 1992. Expected Ph.D. dissertations, not yet completed, are also listed for the sake of completion.

The bibliography is divided into seven sections: Ph.D. Dissertations, M.Sc. Thesis, Book Chapter, Journal Papers, Journal Correspondence, Conference Papers, and Presentations. Items within each section are ordered chronologically in reverse order.

We note that one paper, "Comparative Evaluations of 2D Predictor Implementations for Lossless Compression of Images," by S. Ghosh and W. B. Mikhael [P23], was one of two Best Paper Award recipients at the 37th Midwest Symposium.

**I. PH.D. DISSERTATIONS**


II. M.SC. THESIS


III. BOOK CHAPTER


IV. JOURNAL PAPERS


V. JOURNAL CORRESPONDENCE


VI. CONFERENCE PAPERS


P24. A. Ramaswamy and W. B. Mikhael, "Development and Verification of an Adaptive Algorithm to Efficiently Represent Three Dimensional Signals Using Mixed


VII. PRESENTATIONS


"Lossless Waveform Compression", Student Seminar, Univ. of Central Florida, Orlando, FL, April 15, 1993.

DISTRIBUTION:

1 Dr. T. Ahern
IRIS Data Management Center
1408 NE 45th St.
Seattle, WA 98105

1 Prof. N. Ahmed, Chair
Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

1 A.P. Berg
Electrical and Computer Engineering
University of Central Florida
Orlando, FL 32816-2450

1 Prof. G. Coutu
Hartford Graduate Center
275 Windsor ST.
Hartford, CT 06120-2991

1 Prof. D.M. Etter
Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309

1 Prof. M. Fargues
Electrical and Computer Engineering
U.S. Naval Postgraduate School
Monterey, CA 93943-5000

1 Dr. C.R. Hutt
Albuquerque Seismological Lab.
Bldg. 10002, KAFB East
Albuquerque, NM 87115-5000

1 Prof. R.L. Kirlin
Electrical and Computer Engineering
University of Victoria
Victoria, BC, Canada V8W3P6

1 Prof. N. Magotra
Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

1 Giridhar Mandyam
Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

1 Mr. Wes McCoy
Motorola, Inc. - Mail Stop S-243
5401 Beach
Fort Worth, TX 76137

1 Prof. W.B. Mikhael, Chair
Electrical and Computer Engineering
University of Central Florida
Orlando, FL 32816-2450

1 Yousef Nijim
Electrical and Computer Engineering
University of Central Florida
Orlando, FL 32816-2450

1 Mr. C.C. Stearns
S3 Inc.
2770 San Thomas Expressway
Santa Clara, CA 95051

1 Dr. J. Steim
Quanterra, Inc.
325 Ayer Road, Suite 16
Harvard, MA 01451

1 Dr. L.Z. Tan
Iterated Systems, Inc.
5550-A Peachtree Parkway - Suite 650
Norcross, GA 30092
<table>
<thead>
<tr>
<th></th>
<th>MS</th>
<th>Name</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MS 1436</td>
<td>D.L. Chavez</td>
<td>1007</td>
</tr>
<tr>
<td>1</td>
<td>MS 0572</td>
<td>T.M. Calloway</td>
<td>5912</td>
</tr>
<tr>
<td>1</td>
<td>MS 0750</td>
<td>M.C. Walck</td>
<td>6116</td>
</tr>
<tr>
<td>1</td>
<td>MS 0151</td>
<td>G. Yonas</td>
<td>9000</td>
</tr>
<tr>
<td>1</td>
<td>MS 0972</td>
<td>M.V. Bredemann</td>
<td>9222</td>
</tr>
<tr>
<td>1</td>
<td>MS 0980</td>
<td>G.R. Elliott</td>
<td>9225</td>
</tr>
<tr>
<td>1</td>
<td>MS 1165</td>
<td>J.E. Powell</td>
<td>9300</td>
</tr>
<tr>
<td>1</td>
<td>MS 1165</td>
<td>L.E. Larsen</td>
<td>9300</td>
</tr>
<tr>
<td>1</td>
<td>MS 1170</td>
<td>P.L. Nelson, Jr.</td>
<td>9305</td>
</tr>
<tr>
<td>1</td>
<td>MS 1170</td>
<td>B.C. Bedeaux</td>
<td>9305</td>
</tr>
<tr>
<td>1</td>
<td>MS 1165</td>
<td>F.L. Nutt</td>
<td>9309</td>
</tr>
<tr>
<td>5</td>
<td>MS 1159</td>
<td>M.A. Hedemann</td>
<td>9311</td>
</tr>
<tr>
<td>5</td>
<td>MS 1159</td>
<td>S.D. Stearns</td>
<td>9311</td>
</tr>
<tr>
<td>1</td>
<td>MS 1160</td>
<td>H.D. Garbin</td>
<td>9312</td>
</tr>
<tr>
<td>1</td>
<td>MS 1106</td>
<td>R.E. Pepping</td>
<td>9342</td>
</tr>
<tr>
<td>1</td>
<td>MS 9018</td>
<td>Central Technical Files</td>
<td>8523-2</td>
</tr>
<tr>
<td>5</td>
<td>MS 0899</td>
<td>Technical Library</td>
<td>4414</td>
</tr>
<tr>
<td>1</td>
<td>MS 0619</td>
<td>Print Media</td>
<td>12615</td>
</tr>
<tr>
<td>2</td>
<td>MS 0100</td>
<td>Document Processing</td>
<td>7613-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for DOE/OSTI</td>
<td></td>
</tr>
</tbody>
</table>

22