11/30/95

SANDIA REPORT

SAND95–2470 • UC–706
Unlimited Release
Printed November 1995

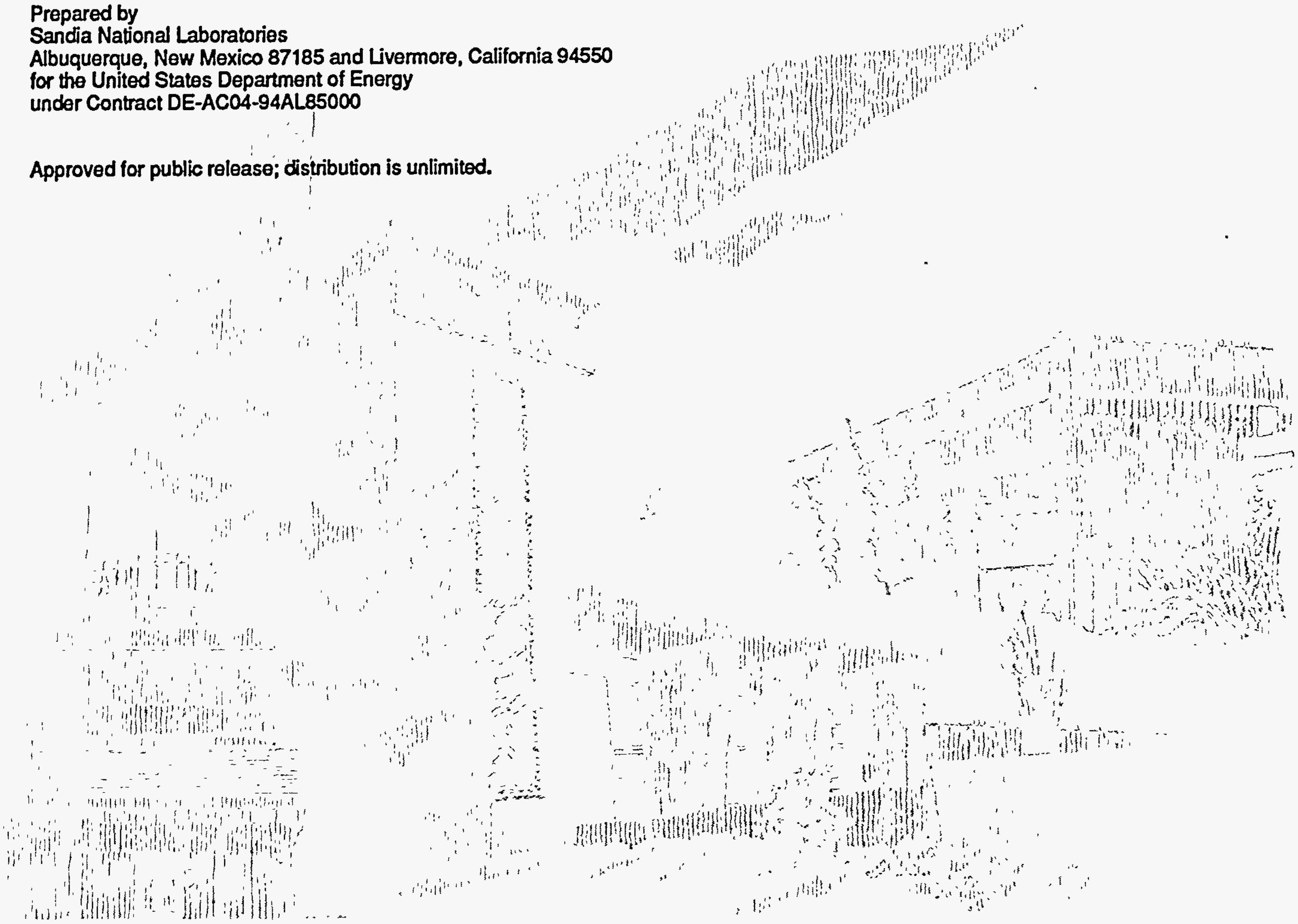
RECEIVED
DEC 12 1995
OSTI

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.



SF2900Q(8-81)

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
US Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A01

SAND95-2470
Unlimited Release
Printed November 1995

Distribution
Category UC-706

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.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

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 Tan 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

Figures

(blank page)

FINAL REPORT:
LOSSLESS COMPRESSION
OF INSTRUMENTATION DATA

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.

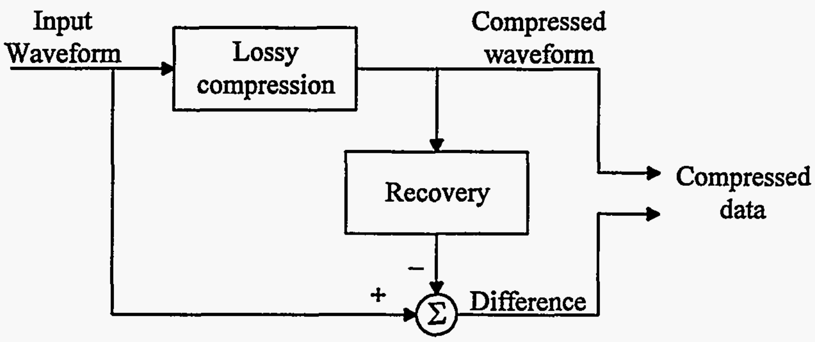


Figure 1. Lossy to lossless conversion.

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 Personnel: S.D. Stearns, J.A. Davis, G.J. Simmons, J. Pearcy, T. Marking, R.M. Clancy, D.M. Honea, G.R. Elliott, M.V. Bredemann.

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 Personnel: Profs. N. Magotra and N. Ahmed, students L.Z. Tan, J.W. McCoy, G. Mandyam, and F. Livingston.
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 <u>CRC Industrial Electronics Handbook</u> [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 $ix(0{:}K-1) = [ix_k]$, 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.

In the literature on waveform compression, the number of bits in ix 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 ix. Thus, if $\lceil\bullet\rceil$ stands for "the least integer equal to or greater than $(\bullet)$", then

$$\text{Number of bits in } ix \equiv K\,bps_{ix} = K\left(1 + \lceil \log_2 |ix_k|_{max} \rceil\right) \tag{1}$$

The number of bits per sample in ix, bps_{ix}, is found simply by dividing (1) by K, and the number of bits per sample in the compressed vector, bps_{iy}, is N_y/K. With these definitions, the compression ratio (CR) is expressed as

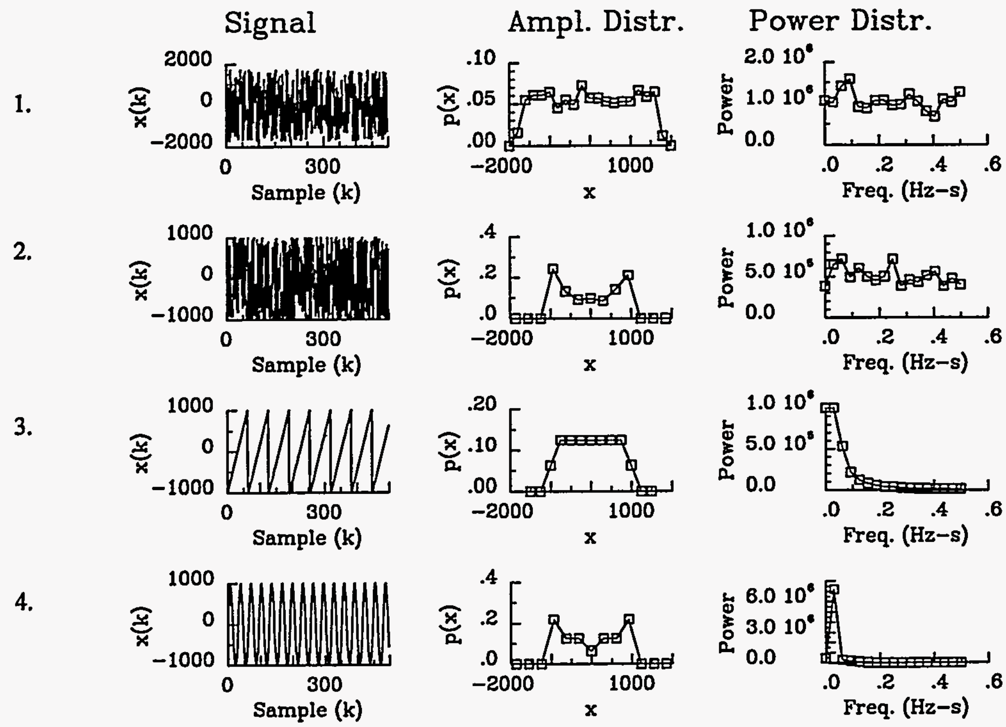
$$CR = \frac{bps_{ix}}{bps_{iy}} = \frac{1 + \lceil \log_2 |ix_k|_{max} \rceil}{N_y/K} \tag{2}$$

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 ix, 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.



Waveform	Amplitude	Spectrum	Lossless Compression
1.	Uniform	White	Almost none
2.	Nonuniform	White	Some
3.	Uniform	Nonwhite	More
4.	Nonuniform	Nonwhite	Lots

Figure 2. Examples of waveforms with different compressibility.

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.

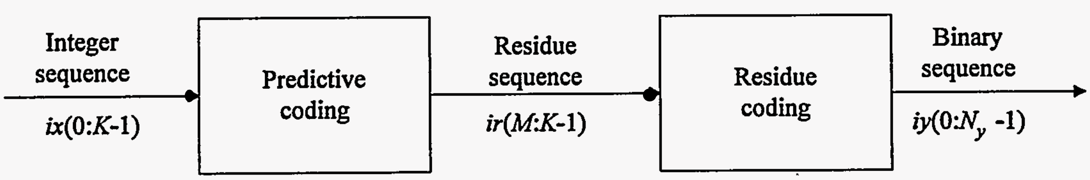


Figure 3. Lossless waveform compression in two stages.

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, σ_{ix}/σ_{ir}, is generally greater than one. In fact, we usually take σ_{ix}/σ_{ir} as a measure of success of the first compression stage. With reversible linear predictive coding, for some types of data, σ_{ix}/σ_{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 $iy(0{:}N_y-1)$. 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

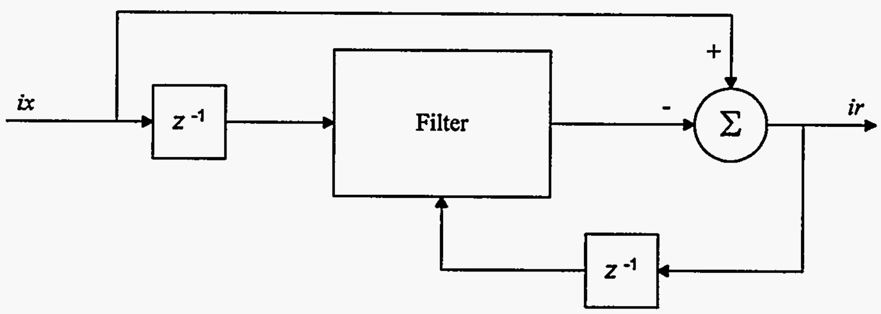


Figure 4. General form of the prediction stage in Fig. 3.

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, σ_{ix}/σ_{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].

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, *ir*, the objective of the second coding stage is to produce a binary sequence (*iy* 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, *ir*. The entropy of *ir* depends entirely on the amplitude distribution of *ir* and is given by

$$\text{Entropy of } ir = -\sum_{ir=ir_{\min}}^{ir_{\max}} f(ir)\log_2 f(ir) \tag{3}$$

where $f(ir)$ is the frequency of occurrence of *ir* in the sequence $ir(M{:}K-1)$, such that the frequencies sum to one.

As suggested by the examples in Fig. 2, a *uniform* distribution of *ir* amplitudes, where all frequencies are equal, produces maximum entropy. That is, if the range of *ir* from $ir_{\min}$ through $ir_{\max}$ is R, and if all values of *ir* appear equally often in $ir(M{:}K-1)$, then the entropy in (3) is $\log_2 R$ bits per symbol. When the values of *ir* 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 *ir* appears in the residue sequence nearly always, so that $f(ir)$ is close to one for this value and near zero for other values of *ir*, 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 *ir* 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, $ix(0{:}K-1)$, 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 *ix* 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 *ix* in Fig. 5. These represent statistics for the entire waveform. The amplitude distribution is from $-5{\bullet}10^5$ to $+5{\bullet}10^5$ 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 *ix* 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 *ix* in Fig. 5.

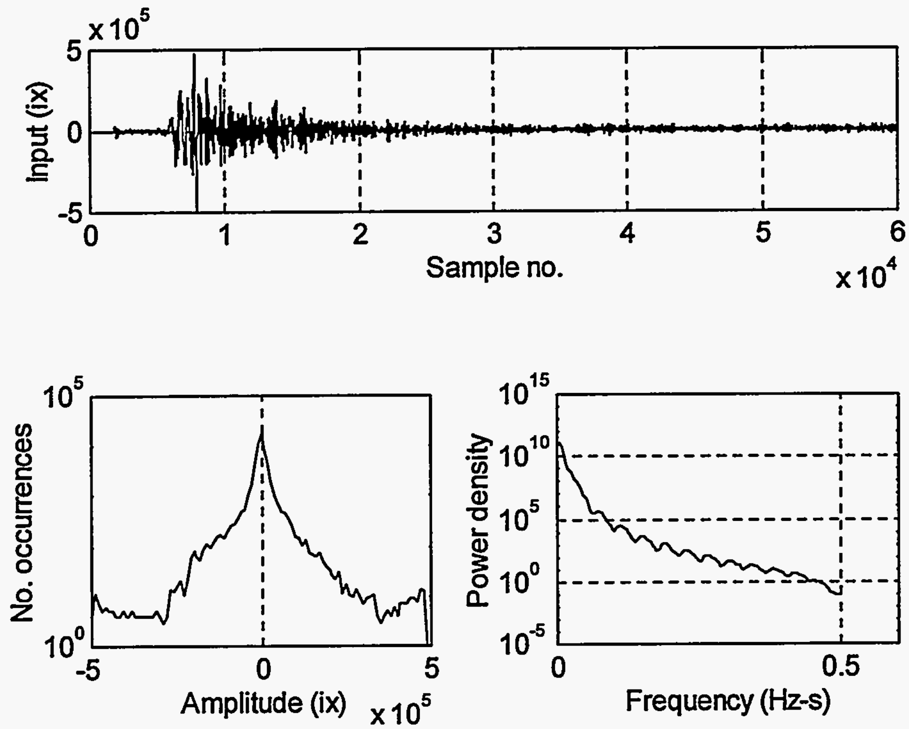


Figure 5. Seismic waveform, *ix*(0:59999), with amplitude and power distributions.

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, σ_{ix}/σ_{ir}, 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 <u>exactly</u> 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_{ir}, which, in accordance with (1), is measured by

$$bps_{ir} = 1 + \lceil \log_2 |ir_k|_{max} \rceil \qquad (4)$$

With arithmetic coding it is possible to reduce bps_{ir} to bps_{iy}, 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.

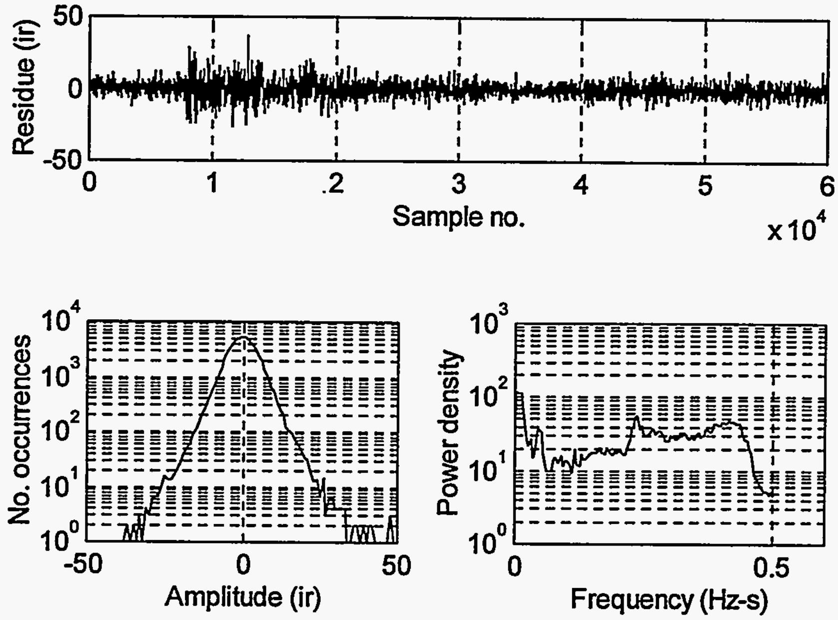


Figure 6. Residues of *ix*(0:59999) in Fig. 5 after predictive coding, with amplitude and power distributions. Frame size was K=1000; predictor size ranged from M=2 through M=8.

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 σ_{ir}=3.43 is the measured standard deviation. Since $|ir_k|_{\max}$ 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 σ_{ir}, which is given approximately by Woodward [J6, Ref. 9] as

$$H_{ir} \approx \log_2\sqrt{2\pi e\,\sigma_{ir}^2} = \log_2(4.133\sigma_{ir}) = 3.825 \qquad (5)$$

Therefore, we would expect the compression of *ir* in the first frame to be from bps_{ir}=5 to around bps_{iy}=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_{iy} = 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 σ_{ir} [D6,J6]. These results for the first data frame are typical of the other 59 frames of *ir* in Fig. 6.

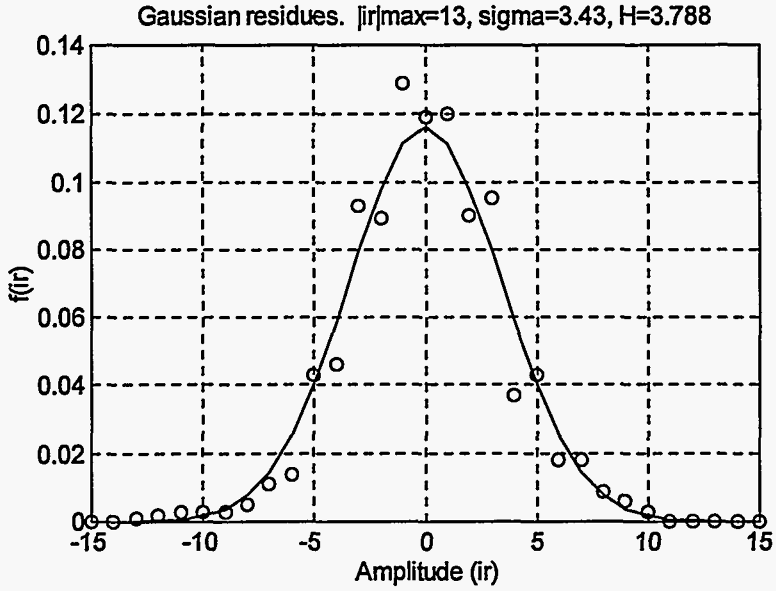


Figure 7. Relative frequency of occurrence of residues in the first frame of *ir* in Fig. 6; K=1000. Solid curve is the Gaussian envelope, $N(0,\sigma_{ir})$, where σ_{ir}=3.43 is the measured standard deviation. H=3.788 is the entropy, measured as in (3).

Finally, compression results for this example are summarized in Fig. 8 by comparing bits per sample, bps_{ix}, bps_{ir}, and bps_{iy}, 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_{ix}=20, the number of bits per sample in (1) needed to store *ix* 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:

Original data file:	ANMBHZ89.ASC	542,994 bytes
PKZIP compression:	ANMBHZ89.ZIP	223,534 bytes
WINZIP compression:	ANMBHZ89.WZP	195,888 bytes
Sandia compression (ENCOD5):	ANMBHZ89.CMP	39,379 bytes

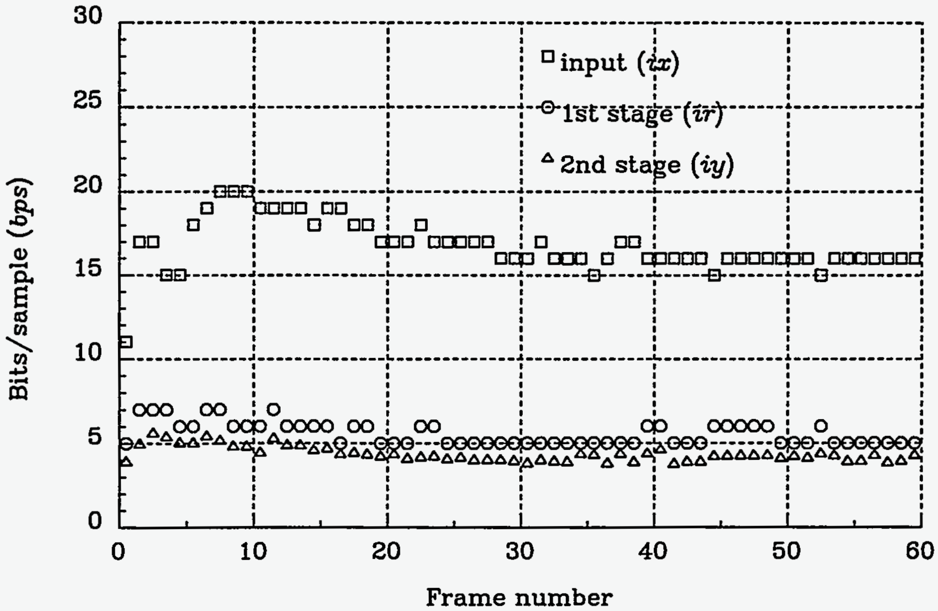


Figure 8. Bits per sample for each of 60 seismic data frames, showing compression in the first (linear prediction) stage, from bps_{ix} to bps_{ir}, and in the second (arithmetic coding) stage, from bps_{ir} to bps_{iy}.

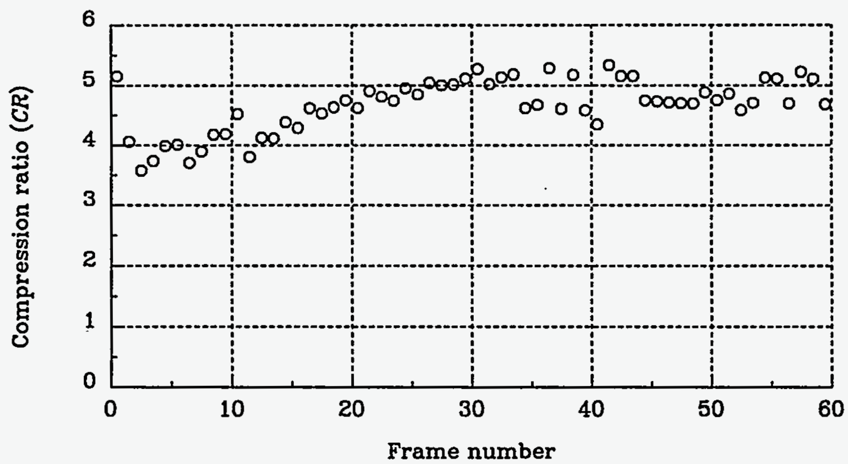


Figure 9. Compression ratio for each of 60 seismic data frames, with bps_{ix}=20.

If we divide the original file size by the compressed file size, we get the compression ratio

$$CR = \frac{\text{Original file size}}{\text{Compressed file size}} = 13.8 \tag{6}$$

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.WZP 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

D1. A.P. Berg, *Efficient Techniques for Compression and Classification of Signals*, Ph.D. dissertation, University of Central Florida, expected summer 1996.

D2. J.W. McCoy, *An Analysis of Lossless Waveform Compression*, Ph.D. dissertation, University of New Mexico, expected fall 1995.
D3. Y.W. Nijim, *Lossless Compression of Seismic Signals*, Ph.D. dissertation, University of Central Florida, expected fall 1995.
D4. H. Yu, *Multidimensional Transform Domain System Modeling and Signal Representation with Applications*, Ph.D. dissertation, University of Central Florida, January 1995.
D5. A. Ramaswamy, *Multidimensional Signal Compression Using Wavelets and Mixed Transforms*, Ph.D. dissertation, University of Central Florida, November 1994.
D6. L.Z. Tan, *Theory and Techniques for Lossless Waveform Compression*, Ph.D. dissertation, University of New Mexico, May 1992.

II. M.SC. THESIS

M1. M.F. Barsoum, *Digital Signal Processing Techniques for Automated Detection and Characterization of Doppler Bubbles in Man*, M.Sc. thesis, University of Central Florida, July 1994.

III. BOOK CHAPTER

B1. G. Mandyam and N. Magotra, and S.D. Stearns, "Lossless Waveform Compression", in Chapter 3, "Measurement System Architecture", Industrial Electronics Handbook, CRC Press, 1996 (J.D. Irwin, ed.).

IV. JOURNAL PAPERS

J1. A. Ramaswamy and W. B. Mikhael, "A Multi-Transform Approach for Medical Data Compression, " IEEE Trans. on Medical Imaging, submitted.

J2. Y. W. Nijim, S. D. Stearns, and W. B. Mikhael, "Differentiation Applied for the Lossless Compression of Medical Images," IEEE Trans. on Medical Imaging, submitted.

J3. W. B. Mikhael and H. Yu, "Adaptive Frequency Domain 2D Modeling Using Spatiotemporal Signals," Journal of Circuits, Systems and Computers, accepted 1995.

J4. Y. W. Nijim, S. D. Stearns, and W. B. Mikhael, "Lossless Compression of Seismic Signals Using Differentiation," IEEE Trans. on Geoscience and Remote Sensing accepted 1995.

J5. A. Ramaswamy and W. B. Mikhael, "Compression of Images Represented Using Multitransforms," Journal of Circuits, Systems and Computers, accepted 1995.

J6. S.D. Stearns, "Arithmetic Coding in Lossless Waveform Compression", IEEE Trans. on Signal Processing, vol. 43, no. 8, pp 1874-1879, Aug. 1995.

J7. W. B. Mikhael and S. M. Ghosh, "Two-Dimensional Block Adaptive Filtering Algorithms with Optimum Convergence Factors," IEEE Trans. on Circuits and Systems II, Aug. 1995.

J8. W. B. Mikhael and A. Ramaswamy, "An Efficient Representation of Non-Stationary Signals Using Mixed-Transforms with Applications to Speech," IEEE Trans. on Circuits and Systems II, vol. 42, no. 6, pp. 393-401, June 1995.

J9. W. B. Mikhael and H. Yu, "Two-Dimensional, Frequency Domain, Adaptive System Modeling Using 3-D Spatiotemporal Inputs," IEEE Trans. on Circuits and Systems II, vol. 42, no. 5, pp. 317-325, May 1995.

J10. W. B. Mikhael and H. Yu, "A Linear Approach for Two-Dimensional Frequency Domain, Least Square Signal and System Modeling," IEEE Trans. on Circuits and Systems II, vol. 41, no. 12, pp. 786-795, Dec. 1994.

J11. S.D. Stearns, L.Z. Tan and N. Magotra, "Lossless Compression of Waveform Data for Efficient Storage and Transmission", IEEE Trans. on Geophysics and Remote Sensing, May 1993, pp. 645-654.

J12. S.D. Stearns, L. Tan, and N. Magotra, "A bi-level Coding Technique for Compressing Broadband Residue Sequences", Digital Signal Processing Journal, July 1992, pp 146-156.

V. JOURNAL CORRESPONDENCE

C1. W.B. Mikhael and A. Ramaswamy, "Application of Multitransforms for Lossy Image Representation", IEEE Trans. on Circuits and Systems, June 1994, pp 431-434.

VI. CONFERENCE PAPERS

P1. H. Yu and W. B. Mikhael, "Two Dimensional Frequency Domain Motion Compensation for Spatiotemporal Signals," Proc. 29th Asilomar Conference on Signals, Systems, and Computers, Oct. 30-Nov. 1, 1995.

P2. A. Ramaswamy and W. B. Mikhael, "Multitransform Applications for Representing Three Dimensional Spatial and Spatiotemporal Signal," Proc. 29th Asilomar Conference on Signals, Systems, and Computers, Oct. 30-Nov. 1, 1995.

P3. Y.W. Nijim, S.D. Stearns, and W.B. Mikhael, "Lossless Compression of Images Using A Hybrid Transform/Differentiation Technique", accepted for 29th Asilomar Conf. on Signals, Systems, and Computers, Oct. 30-Nov. 1, 1995.

P4. Y. W. Nijim, S. D. Stearns, and W. B. Mikhael, "Lossless Compression of Images Using Two Dimensional Differentiators," Proc. 38th Midwest Symp. on Circuits and Systems, Rio de Janeiro, Brazil, August 13-16, 1995.

P5. A. Berg and W. B. Mikhael, "Signal Representation Using Adaptive Parallel Mixed Transforms," Proc. 38th Midwest Symp. on Circuits and Systems, Rio de Janeiro, Brazil, August 13-16, 1995.

P6. G. Mandyam, N. Ahmed, and S.D. Stearns, "Efficient Codebook Design for Lossless Image Compression", Proc. IEEE Ideas in Science and Electronics Symp., Albuquerque, NM May 9-11, 1995, pp. 31-34.

P7. G. Mandyam, N. Ahmed, and S.D. Stearns, "A Two-Stage Scheme for Lossless Compression of Images", Proc. ISCAS-95, Seattle, May 1-4, 1995, pp 1102-1105.

P8. Y.W. Nijim, S.D. Stearns, and W.B. Mikhael, "Lossless Compression of Seismic Signals Using Least Square, Frequency Domain Pole Zero Modeling", Proc. ISCAS-95, Seattle, May 1-3, 1995, pp 1106-1109.

P9. M. Coffey and D. Etter, "Sea Surface Height Data Compression Using Wavelet Decomposition", Proc. ISCAS-95, Seattle, May 1-3, 1995, pp 1110-1113.

P10. G. Coutu and M. Fargues, "Lossless Compression of Ultra Wideband Radar Using Adaptive Filters", Proc. ISCAS-95, Seattle, May 1-3, 1995, pp 1259-1262.

P11. A. Ramaswamy and W.B. Mikhael, "Development of A Near Lossless Image Compression Technique Using Mixed Transforms", Proc. ISCAS-95, Seattle, May 1-4, 1995, pp 1263-1266.

P12. F. Livingston, N. Magotra, S.D. Stearns, and W. McCoy, "Real-Time Implementation Concerns for Lossless Waveform Compression", Proc. ISCAS-95, Seattle, May 1-4, 1995, pp 1267-1270.

P13. G. Mandyam, N. Ahmed, and N. Magotra, "A DCT-Based Scheme for Lossless Image Compression", Digital Video Compression: Algorithms and Technologies 1995, IS&T/SPIE Electronic Imaging Conference, San Jose, CA, February, 1995.

P14. G. Coutu, S.D. Stearns, M. Fargues, and D.J. Plaza, "Adaptive Systems with Coefficients Indicating Frequency", Proc. 28th Asilomar Conf. on Signals, Systems, and Computers, Oct. 31-Nov. 2, 1994.

P15. F. Livingston, N. Magotra, W. McCoy, and S.D. Stearns, "Lossless Data Compression in Real Time", Proc. 28th Asilomar Conf. on Signals, Systems, and Computers, Oct. 31-Nov. 2, 1994.

P16. A. Ramaswamy and W. B. Mikhael, "Applications of Multitransforms to 3D Spatial Signal Representation," Proc. 28th Asilomar Conference on Signals, Systems, and Computers, pp. 749-753, Oct. 30- Nov. 2, 1994.

P17. J.W. McCoy, N. Magotra, and S.D. Stearns, "Lossless Predictive Coding", Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, Aug. 3-5, 1994.

P18. M. Fargues, G. Coutu, and S.D. Stearns, "Seismic Data Compression Using Adaptive Filters", Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, Aug. 3-5, 1994.

P19. M. F. Barsoum and W. B. Mikhael, "A Conjugate-Gradient-Based Block Adaptive Algorithm, its Applications and Performance Evaluation," Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, pp. 903-906, Aug 3-5, 1994.

P20. Y. W. Nijim, S. D. Stearns, and W. B. Mikhael, "Evaluation of Differentiation for Lossless Compression of Several Signal Classes," Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, pp. 939-942, Aug 3-5, 1994.

P21. A. P. Berg and W. B. Mikhael, "A Survey of Techniques for Lossless Compression of Signals," Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, pp. 943-946, Aug 3-5, 1994.

P22. H. Yu and W. B. Mikhael, "2D and 3D Adaptive Signal Processing in the Frequency Domain," Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, pp. 917-920, Aug 3-5, 1994.

P23. S. Ghosh and W. B. Mikhael, "Comparative Evaluations of 2D Predictor Implementations for Lossless Compression of Images," Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, pp. 947-9950, Aug 3-5, 1994.

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

Transforms," Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, pp. 793-796, Aug 3-5, 1994.

P25. J.W. Haines and D.M. Etter, "Analysis of Lossless Data Compression Using Multirate Filter Banks", Proc. 37th Midwest Symp. on Circuits and Systems, Lafayette, LA, Aug. 3-5, 1994.

P26. H. Yu and W. B. Mikhael, "Two Dimensional Spectral Representation with Applications," Proc. IEEE International Symp. on Circuits and Systems (ISCAS 94), London, England, pp. 589-592, May 30-June 2, 1994.

P27. W. B. Mikhael and A. Ramaswamy, "Constrained Image Representation with Applications," Proc. IEEE International Symp. on Circuits and Systems (ISCAS 94), London, England, pp. 297-300, May 30-June 2, 1994.

P28. S. Ghosh and W. B. Mikhael, "Two-Dimensional Optimal Algorithms for Image Compression Using ARMA Predictors," Proc. IEEE International Symp. on Circuits and Systems (ISCAS 94), London, England, pp. 613-616, May 30-June 2, 1994.

P29. A. Ramaswamy and W. B. Mikhael, "An Efficient Coding Technique for Multi-Transform Image Representation," Proc. IEEE International Symp. on Circuits and Systems (ISCAS 94), London, England, pp. 71-74, May 30-June 2, 1994.

P30. D.M. Honea and S.D. Stearns, "Lossless Waveform Compression: A Case Study", Proc. 27th Asilomar Conf. on Signals, Systems, and Computers, Nov. 1-3, 1993.

P31. M. Halfin, D. Sturim, G. Coutu, and S.D. Stearns, "Adaptive Frequency Selective Filters", Proc. 27th Asilomar Conf. on Signals, Systems, and Computers, Nov. 1-3, 1993.

P32. S.D. Stearns, R.L. Kirlin and J. Fan, "Techniques for Geophysical Data Compression", Proc. 1993 SPIE Aerospace & Remote Sensing Symp.(Conf. 1941: Ground Sensing), April 1993.

P33. C.V. Peterson and C.R. Hutt, "Lossless Compression of Seismic Data", Proc. 28th Asilomar Conf. on Signals, Systems, and Computers, October 26-28, 1992.

P34. S.D. Stearns, L. Tan, and N. Magotra, "A Technique for Lossless Compression of Seismic Data", Proc. 1992 International Geoscience and Remote Sensing Symp., Houston, May 26-29, 1992

VII. PRESENTATIONS

"Lossless Waveform Compression: Recent Developments and Current Research", Seminar, Univ. of Central Florida, Orlando, FL, April 19, 1995.

"Lossless Waveform Compression", Grad. Seminar, Michigan State Univ., March 30, 1995.

Invited Lectures on Signal Processing and Waveform Compression, Darmstadt Technical University, Darmstadt, Germany, June 22-28, 1994.

" Progress in Lossless Waveform Compression", Seminar, Univ. of Central Florida, Orlando, FL, April 13, 1994.

"Seismic Data Compression", talk at Jet Propulsion Laboratory, Pasadena, CA, March 2, 1994.

"Lossless Waveform Compression", Grad. Seminar, Univ. of Colorado, Boulder, CO, Nov. 17, 1993.

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

" Lossless Waveform Compression", Grad. Seminar, Kansas Univ., Lawrence, KA, March 16, 1993.

(blank page)

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

1	MS 1436	D.L. Chavez	1007
1	MS 0572	T.M. Calloway	5912
1	MS 0750	M.C. Walck	6116
1	MS 0151	G. Yonas	9000
1	MS 0972	M.V. Bredemann	9222
1	MS 0980	G.R. Elliott	9225
1	MS 1165	J.E. Powell	9300
1	MS 1165	L.E. Larsen	9300
1	MS 1170	P.L. Nelson, Jr.	9305
1	MS 1170	B.C. Bedeaux	9305
1	MS 1165	F.L. Nutt	9309
5	MS 1159	M.A. Hedemann	9311
5	MS 1159	S.D. Stearns	9311
1	MS 1160	H.D. Garbin	9312
1	MS 1106	R.E. Pepping	9342

1	MS 9018	Central Technical Files	8523-2
5	MS 0899	Technical Library	4414
1	MS 0619	Print Media	12615
2	MS 0100	Document Processing for DOE/OSTI	7613-2

Org.	Bldg.	Name	Rec'd by	Org.	Bldg.	Name	Rec'd by

Sandia National Laboratories