Proceedings of the 65th Annual Meeting of the Society of Exploration Geophysicists, Houston, TX, October 1995

High-performance seismic trace compression

Paul L. Donoho, Raymond A. Ergas, Chevron Petroleum Technology Co.; and John D. Villasenor, University of California, Los Angeles*

SUMMARY

We have developed an algorithm for the compression of seismic data capable of compression ratios in excess of 100:1 for a wide range of datasets representing both prestack and post-stack data. Although the compression algorithm is lossy, in the sense that its use introduces noise into the final reconstituted dataset, the compression noise leads to no observable loss of geophysical information at compression ratios substantially greater than 100:1. This compression algorithm employs the wavelet transform to represent the data in a basis characterized by a number of subbands having differing temporal and spatial frequency content. The coherency and redundancy of multidimensional seismic datasets, ranging from prestack field data to migrated 3D stack cubes, permits efficient quantization of the data in each wavelet-transform subband, such that a very small average number of bits per data sample will accurately represent the data. The ability to compress data by such large factors leads to the possibility of transmitting entire field datasets from a multistreamer marine operation directly to the processing center in nearly real time. Complete 3D prestack datasets can be maintained online with realizable disk storage at 100:1 compression ratio. Network transmission and disk storage costs for other applications can also be reduced by large factors. The algorithm described here has been optimized for speed when compressing large 3D datasets. We expect even larger compression ratios, perhaps as high as 300:1 with acceptable noise, for datasets organized in a 4D fashion, such as field data from multistreamer acquisition with five or more streamers.

INTRODUCTION

The current emphasis on 3D seismic acquisition has led to the generation of datasets capable of taxing the storage and I/O capacities of large mainframe computers. The trend toward distributed, networked computing environments has exacerbated most problems of dealing with datasets which can approach or even exceed terabyte levels in pre-stack representations. Despite the rapid increase in available disk storage at ever lower cost, management of large datasets remains a formidable problem. Furthermore, in the distributed computing environment, network transmission of even stacked datasets is usually prohibitively expensive. A data compression technology which can provide at least 100:1 compression ratio reduces the terabyte-sized dataset to a manageable 10 gigabytes, and it permits transmission of a dataset normally requiring hours for network transmission in a matter of minutes.

It is, however, impossible to attain such a large compression ratio without introducing a certain degree of noise into the dataset. We have found perfectly noiseless compression algorithms applied to seismic data can achieve compression ratios no greater than a factor of about 2:1. It is possible, nevertheless, to devise a so-called lossy compression algorithm capable of large compression ratios at the expense of introducing noise into the dataset. A principal goal of the algorithm is to constrain the algorithm-induced noise such that there is no loss of geophysically significant information in the dataset. Lossy compression schemes have been employed for some time for still photographic images, leading, for example, to the JPEG standard (Wallace, 1990), which is now widely used. But JPEG introduces artifacts even at low compression ratios that would be totally unacceptable for seismic data.

More recently, the wavelet transform (Daubechies, 1992), a relatively new mathematical concept, has been employed to develop data-compression algorithms for many purposes, with noise characteristics superior to those of JPEG and other similar compression methods. Results for seismic data sets have been reported earlier (Luo and Schuster, 1992; Bosman and Reiter, 1993; Reiter and Heller, 1994).

We have developed a wavelet-transform based compression algorithm capable of attaining compression ratios greate than 100:1 (0.32 bits per sample) for both prestack and post-stack seismic data, with the introduction of negligible noise leading to no apparent loss of geophysical information in the final uncompressed dataset. The availability of large compression ratios, with acceptable noise characteristics, leads to the ability to carry out data-processing and transmission tasks that would otherwise be impractical and that promises to bring significant benefits to the seismic industry. For example, at a compression ratio of 100:1 it becomes possible to transmit a complete seismic field dataset to the process ing center in nearly real time during field acquisition oper ations. We are carrying out a field trial during a real 3I survey, transmitting data from a marine 5-streamer vess via satellite to a onshore processing center (Stigant *et al* 1995). With nearly instantaneous transmission of both sei mic and positioning data enhanced QC possibilities becom possible without major personnel or equipment deploymen Even more important is the probability of a significant in provement in turnaround time from the start of acquisitic to delivery of a final migrated stack.

In addition, compression technology can be integrated into many areas of seismic data processing, including archival storage. A major goal in this work is the development of a set of standard seismic trace compression tools. We are carrying out a statistical test of the effects of compression on interpretation to determine, with a high degree of confidence, the geophysical validity of the algorithm for large compression ratios. We further see the possibility of the development, in the near future, of an industry standard for seismic data compression (sponsored by SEG?) to provide uniform benefits to all segments of the industry.

Wavelet-Transform Based Compression

Most of the lossy data-compression algorithms reported in the literature follow similar approaches: first, the data are mathematically transformed to a new representation in which they are better organized for data-compression procedures than in the normal spatial-temporal domain. Most geophysicists are quite familiar with the use of the Fourier transform to reorganize data to facilitate a particular objective. The f-k transform is a prime example. The wavelet transform (Daubechies, 1992; Wickerhauser, 1994) organizes data into subbands, each of which exhibits a different scale of temporal and spatial frequency characteristics. Biorthogonal wavelet filters were employed along each dimension after a study of a wide range of different wavelet filters (Villasenor *et al.*, 1995). For example, the largest subbands consist of high-frequency data in temporal and all spatial dimensions. Because of the coherency of the data along the spatial dimensions, virtually all of the data in this subband represent noise of no geophysical significance. Many of the subbands at lower frequencies contain the important information that must not be contaminated with significant noise. But these important subbands contain far fewer data samples.

Thus, the data are organized in groups which contain mostly noise that can be discarded, representing most of the volume of data, and groups which contain most of the geophysical information, but which represent a small fraction of the total data volume. Such a decomposition of the data into distinctly different groups cannot be accomplished with, for example, the Fourier transform. It is important to organize the data prior to carrying out the wavelet transform in such a way as to maximize this separation into important and unimportant subbands. It is, therefore, necessary to examine the geophysical data initially to determine how best to take advantage of redundancy and to perform reversable processing such as programmed gain and NMO to increase statistical stationarity.

The second step leading to a valid data-compression algorithm consists of quantizing the data samples in the wavelet representation in such a way that important samples are preserved with the precision necessary to preserve their information content while unimportant samples are quantized to either zero values or very small integers requiring few bits for their representation. It is this quantization step which introduces all of the loss of the compression algorithm. Consequently, careful design of the quantization procedure is extremely important in obtaining good compression results. An acceptable quantization procedure requires statistical analysis of the data to be quantized, as well as an *a priori* understanding of what features of the data are geophysically important.

In the final step, the quantized data are subjected to further encoding, using run-length coding to compress strings of identical values, and entropy encoding to represent frequently occurring values by fewer bits than less frequent values.

Decompression just reverses the above steps. The ultimate goal of the overall compression/decompression procedure is to reconstitute a dataset as faithful as possible to the original dataset, but one which has been stored in highly compressed form.

Application to Seismic Data

Although the compression could be applied to individual seismic traces (Luo and Schuster, 1992), this does not take advantage of the trace-to-trace spatial coherence in two or more dimensions, and therefore limits acceptable data-compression ratios no greater than about 10:1. For a two-dimensional dataset (*e. g.* a shot gather, a CDP gather, or a stacked 2D section), the spatial coherency along the added dimension leads to greater compression possibilities. Reiter and Heller (1994) have, for example, reported compression ratios as high as 50:1 for CDP gathers to which NMO has been applied. A penalty is paid, of course, when datasets of higher dimensionality are compressed: individual traces can no longer be reconstructed from the compressed dataset without difficulty. In general, the entire compressed ensemble of traces must be uncompressed to obtain any individual trace.

The algorithm employed in our work has been applied primarily to 3D datasets. We define a 3D dataset, however, not in the same terms as used when referring to a 3D survey. Rather, we mean a dataset characterized by one temporal and at least two spatial dimensions. Thus a collection of sequential shot records would constitute a 3D dataset in our context. Likewise, in this context, the field data from a multistreamer marine acquisition survey would constitute a 4D dataset. Other 3D examples, would include a 3D stack cube, or a collection of CDP gathers from a 2D line. For the datasets of higher dimensionality that we have considered, compression ratios of at least 100:1 can be attained with negligible addition of noise. Higher levels of compression can be achieved when the data are oversampled in time, as frequently occurs. Much higher compression ratios have been realized in many cases with noise at a level where loss of geophysical information may occur. A formal validation test, based on statistical principles used in the radiology field, is underway to determine the compression-ratio limit which can be attained where no significant information is sacrificed by the compression/decompression process.

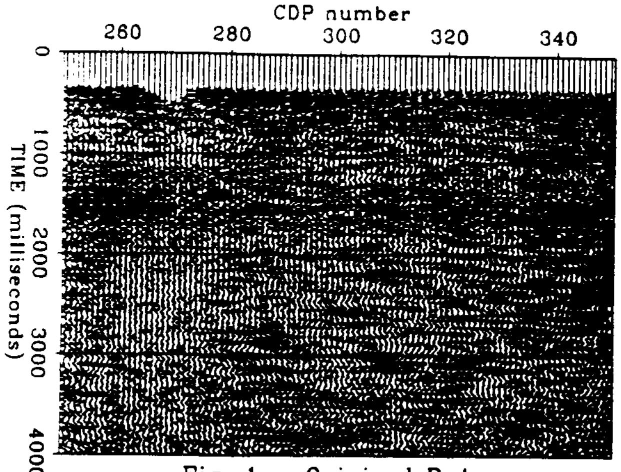

Fig. 1 – Original Data

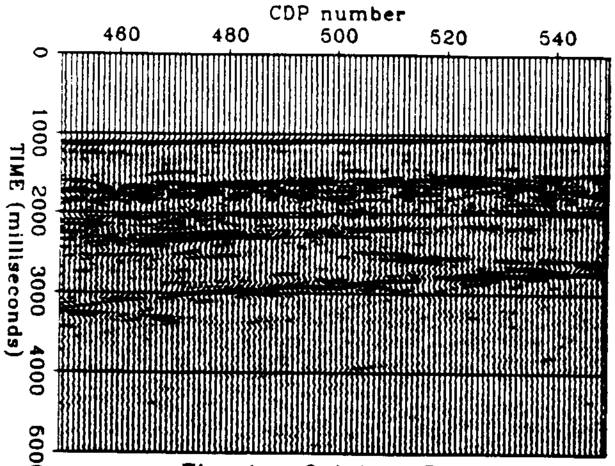

Fig. 4 – Original Data

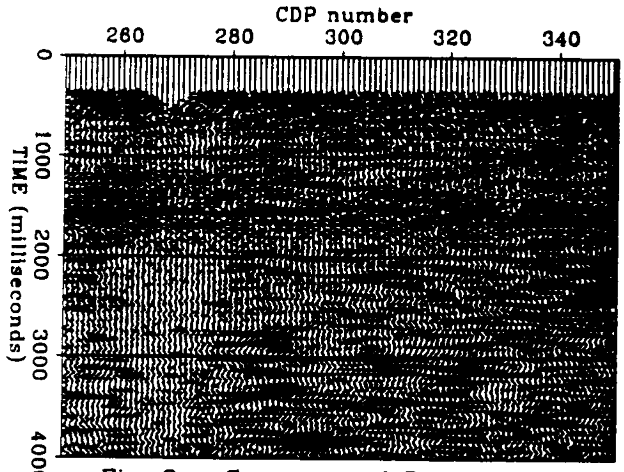

Fig. 2 – Compressed Data at 94:1

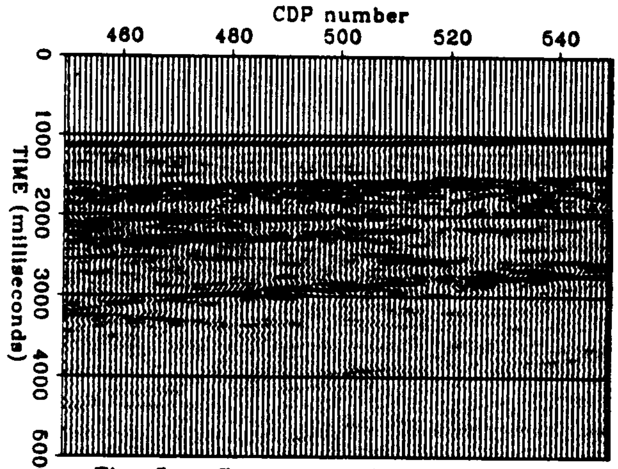

Fig. 5 – Compressed Data at 162:1

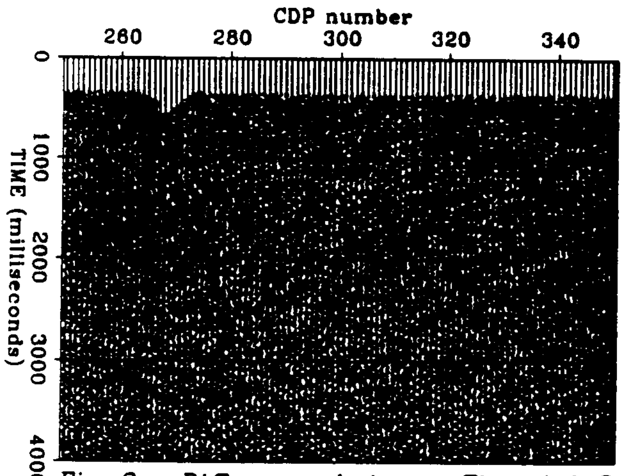

Fig. 3 – Difference between Figs. 1 & 2

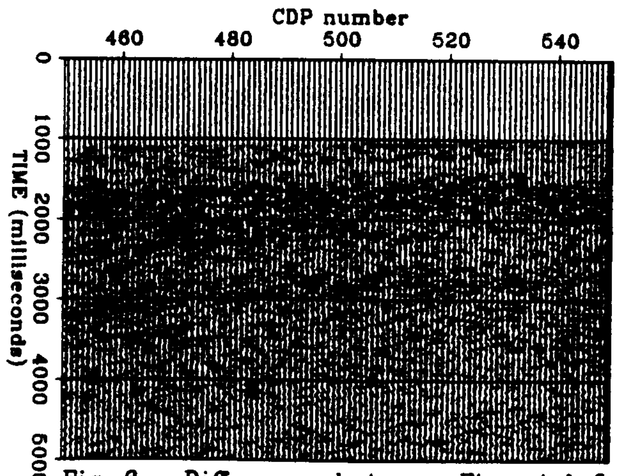

Fig. 6 – Difference between Figs. 4 & 5

Two examples, are shown in the accompanying figures. Figures 1-3 depict 2-D marine data data which were compressed and decompressed as shot gathers, and then stacked. Figure 1 is stacked data that has not been compressed, figure 2 is data previously compressed at a ratio of 94:1 and then stacked, and figure 3 the difference between 1 and 2, scaled by about 2 to show detail.

Figures 4-6 depict a part of one typical 2D section from a 3D stacked data cube, which was sampled at 4 ms. The 3D cube of 200 CDPs by 140 lines by 1251 samples required a total of 4 minutes to go through both compression and decompression, combined with the required input and output from disk, on an IBM RS6000 Model 990 workstation using the ProMAX system from Advance Geophysical. The data in Figure 4 have not been compressed, in figure 5, they were compressed at a ratio of 162:1 (0.2 bits per sample) and uncompressed, and again figure 6 represents the difference between the 4 and 5, scaled up by a factor of approximately 2 to show detail.

In both examples, it is apparent that very little data of any significance is lost in the compression process. The primary difference between the before and after sections is the absence of much of the higher-frequency noise seen in the before sections. The stacked data example (figures 4-6) may be missing a few events of interest at this high compression ratio. Practical values for 3D stacked data are closer to 100:1.

CONCLUSIONS

The seismic data-compression algorithm described here attains geophysically acceptable results at compression ratios exceeding 100:1. The maximum acceptable compression ratio for many datasets may exceed this level by a considerable margin, but a level of 100:1 represents a threshold at which certain otherwise infeasible data-management operational goals become attainable. The real-time transmission of a complete 3D multistreamer dataset from the field to the processing center discussed by Stigant *et al.* (1995) was made possible through our robust and fast algorithm, which can be utilized on most currently available workstations without requiring special capabilities. Other significant application potential includes the ability to increase the online volume of data (or slow the acceleration of needed disk space!), to reduce archival storage volume, and to speed the network transmission of stacked and unstacked seismic data. We foresee the likelihood that seismic data-compression technology will reach widespread acceptance in the near future, and it is our expectation that this technology will evolve to become an industry standard within a reasonable time.

ACKNOWLEDGEMENTS

We have a number of our colleagues to thank, including Jon Stigant, Steve Cole, and the management of Chevron for allowing us to publish. We also thank Prof. David Donoho, Statistics Department, Stanford University, for providing insight through many stimulating discussions and for providing very useful tutorial software.

REFERENCES

Bosman, C. and Reiter, E., 1993, Seismic Data Compression Using Wavelet Transforms, 1993, Presented at the 63th Ann. Internat. Mtg., Soc. Expl. Geophys.

Daubechies, I., 1992, *Ten Lectures on Wavelets*: Philadelphia, SIAM

Foster, D. J., Mosher, C. C., and Hassanzadeh, S., 1994, Wavelet transform methods for geophysical applications: Presented at the 64th Ann. Internat. Mtg., Soc. Expl. Geophys.

LeBras, R., Mellman, G. and Peters, M., 1992, Wavelet Transform Method for Downward Continuation: Presented at the 62th Ann. Internat. Mtg., Soc. Expl. Geophys.

Luo, Y. and Schuster, G. T., 1992, Wave Packet Transform and Data Compression: Presented at the 62th Ann. Internat. Mtg., Soc. Expl. Geophys.

Reiter, E. C., and Heller, P. N., 1994, Wavelet transform-based compression of NMO-corrected CDP gathers: Presented at the 64th Ann. Internat. Mtg., Soc. Expl. Geophys.

Stigant, J. P., Ergas, R. A., Donoho, P. L., Regnault, A., and Meignie, Y., 1995, Field Trial of Seismic Compression for Real Time Transmission: Submitted to the 65th Ann. Internat. Mtg., Soc. Expl. Geophys.

Villasenor, J., Belzer, B., and Liao, J., 1995, Wavelet Filter Evaluation for Efficient Image Compression: To appear in *IEEE Trans. Image Proc.*, August 1995.

Wallace, G. K., 1990, The JPEG Still Picture Compression Standard: *Communications of the ACM 34*, 31-44

Wickerhauser, M. V., 1994, *Adapted Wavelet Analysis from Theory to Software*, Wellesley, Mass., A. K. Peters