Geophysical DP requirements could exceed the world's GP capacity by 1985

by CARL H. SAVIT
Western Geophysical Company of America
Houston, Texas

INTRODUCTION

The reflection seismic method of underground mapping is today the only practicable means of selecting a place to drill a new oil or gas well. As a consequence, the free world spends about two billion dollars a year for reflection seismic exploration.

All of this activity consists of data gathering, data reduction, and data analysis. In 1977 the industry recorded between $10^{14}$ and $10^{15}$ bits of primary data in digital form. All of these data was processed at least once. Indeed, it is probable that the geophysics industry processes more primary digital data than does any other activity. Worldwide, the industry dedicates computer power equivalent to a few dozen IBM 370/168's to processing seismic reflection data.

During 1977 new field techniques were introduced to improve the resolution of the entire process. These new techniques begin by recording about 20 times as much data per unit of survey as are being recorded in conventional, present-day practice. A further increase to about 50 times the present data rates will be needed to make the horizontal resolution of the system reasonably consistent with the vertical. Ultimately, algorithms that take full advantage of the increased recording density will require 1000 or more additional computation steps per input point. Consequently, we perceive a need for a 50-fold increase in internal memory and a 50,000-fold increase in arithmetic speed. In terms of "power" (where power is defined as the product of internal memory size and the number of operations per unit time), these requirements translate to a $3 \times 10^6$-fold increase.

It has been the experience of the past three decades that available computer power (as we have defined it) increases about one order of magnitude every 2.7 years. If, as seems reasonable today, this trend continues until 1985, our computers will then be about $10^4$ times as powerful as they were in 1977. Also, as seems reasonable today, reflection seismologists will by 1985 be ready for the $3 \times 10^4$-fold increase we have projected. To accomplish the same number of units of survey in 1985 as were done in 1977, we could be facing a requirement for 3000 times as many 1985 model computers as the number of 1977 model machines we used in 1977. It would, in addition, be reasonable to expect the market for surveys to have doubled in the next seven years so that we may need to increase our demand by an additional factor of two. (Figure 1)

DATA AND ALGORITHMS

To understand the nature of the geophysical computation requirement, an overview of the basic physics of reflection seismology will be of some help.

In reflection seismology a mechanical excitation is applied at or near the surface of the earth. Pick-ups also on or near the surface and relatively near the point of excitation detect the response of the earth to the excitation. From the complex pattern of energy returned to the surface, the exploration geophysicist attempts to select the longitudinal elastic waves (i.e., the sound waves) that have returned to the surface by reflection from underground interfaces between different rock layers. In the current state of the art, energy of all other kinds is rejected or suppressed as unwanted "noise."

Nearly all processing methods in general use into the first few years of this decade were based on the concepts of geometrical optics. In other words, sound waves were assumed to travel along rays and to be reflected or refracted at the point of intersection of each ray with an interface or discontinuity. (Figure 2)

As a consequence of these assumptions, deriving a representation of the configuration of underground layering from the received acoustic signals involves the use of little more than the algebraic and trigonometric equations of analytic geometry. One major intermediate problem, that of deriving the velocity of sound as a function of depth from the surface, appears to require the solution of an integral equation. Instead, however, of attempting an analytic solution in the face of noisy and incomplete data, the exploration seismologist has opted for the method of exhaustive search. He solves the forward problem of determining a suite of likely reflection times for selected parts of the data over an anticipated range of assumed velocities. The algebra required is primarily the Pythogorean equation followed by a vast number of correlation calculations to determine the degree to which reflection times derived from the searched velocities fit the observed signals.

63

From the collection of the Computer History Museum (www.computerhistory.org)
The power of typical commercially available computers has been growing about one order of magnitude every 2.7 years and should continue to do so through the 1980's. Requirements of the seismic exploration industry have been growing and should continue to grow at a substantially faster rate.

The necessary ray-path calculations are quite simple and so produce a relatively small computational load. Most of the computation time is devoted to the extraction of good acoustic "signals" from noisy data. The two approaches used to improve signal-to-noise ratios are the averaging of redundant data and filtering by means of a battery of sophisticated techniques.

By far the largest computing load stems directly from the high redundancy of data being taken. A typical field record of the response of the earth to one mechanical input consists of 3000 data values taken simultaneously at each of 48 equally spaced surface locations. Each input thus produces about 150,000 floating-point numbers. Floating point is required because the signal immediately after the instant of input can be four orders of magnitude larger than the signal six seconds later when the recording is finished.

Typically one signal input is applied every 30 to 60 meters along a line of survey so that a seismic survey produces some $2.5 \times 10^6$ to $5.0 \times 10^6$ floating-point data values per kilometer.

Such surveys produce data redundancies ranging from 12 to 48. When more redundancy is required to achieve useful anti-noise discrimination, more data are taken per kilometer of survey. Redundancies of 96 are not uncommon in special situations.

In nearly half of all surveys on land, the input consists of a swept-frequency (chirp) signal of substantial duration. Systems using this form of input require the recording of thirty times more data values per kilometer. Simple computation in the field compresses these data about 16-fold, but the result of this compression must be further processed before it reaches the computational level of unprocessed impulse data.

Before redundant data can be combined or averaged, several distinct data processing steps must be applied. Those redundant data that apply to, or represent, the same subsurface point are gathered from diverse input and pick-up points. The different sound waves reflected from the same subsurface point have travelled paths of different length and have encountered different inhomogeneities in their paths. Field recordings are, of necessity, made in real time so that data from different pick-ups must be recorded in time-multiplexed order.

Normally, the first computational step is to demultiplex the data into time-sequential value sets for each pick-up. Next, data are normalized both as to amplitude and time scale to approximate values that would have been produced by horizontal plane waves. The necessary transformations are non-linear but have simple geometric bases.

In most cases the first of several filtering steps is applied to the data after the amplitude normalization and before application of the time-scale normalization. Almost always the filter is in the nature of a deconvolution; that is, an attempt is made to undo the various spread functions that characterize the effective shape of the input signal. That the input function has substantial duration is the result both of the imperfect nature of the mechanical impulse itself and of the reverberatory character of the earth materials near the input source.

Designing the filter operators to be applied to the data to convert the input signal to a band-limited, simple impulse function represents a substantial computation load. While it is a relatively simple matter to use the Wiener-Levinson algorithm to invert a known (i.e., recorded-in-the-field) mechanical impulse shape to a desired simple impulse, it is somewhat more burdensome to derive the analogous operator to compensate for an unknown reverberation function. The usual method is to derive the amplitude spectrum of the reverberation function from autocorrelations of the unfiltered data and to assume (on physical grounds) that the function is minimum-phase. Other more complex schemes are in use to various extents.

All of the steps described to this point are necessarily applied to essentially all of the field-recorded data and, hence, represent the great bulk of the computation load. In those cases in which the velocity of seismic waves varies rapidly along the line of survey, the intermediate step of...
velocity determination, described earlier, can dominate the computation requirement.

In the last computer pass of these preliminary computation steps, the data are combined or averaged to produce the compacted, signal-enhanced data set for final processing. Essentially all of the redundancy has now been removed so that data volume is reduced by the original redundancy factor of 12 to 96. Furthermore, the dynamic range of the data has been sufficiently reduced that results may validly be written as 16-bit, fixed-point numbers.

Until quite recently, virtually all seismic reflection data were subjected to at most one or two additional filtering steps and were amplitude-conditioned in one of two alternative modes and formatted for presentation in graphical form as hard copy.

Such additional calculations as have been made within the geometric-optics framework have ranged from simple smoothing operations to elaborate ray-tracing routines but in the aggregate have had negligible impact on computer requirements.

**ACOUSTIC WAVE EQUATION**

We in the exploration-geophysics community have long been aware that the geometric-optics model of the seismic reflection process is, at best, a first approximation to reality. Even the fundamental assumption of the geometrical optics model—that the longest wavelengths involved are small compared to the dimensions of interacting elements in the medium—is at best weakly valid. Some of the wavelengths we use are indeed substantially greater than the dimensions of subsurface features of interest.

The simplest model that enables us to relax the dimensional assumption required in geometric optics is acoustical. In the acoustical model it is assumed that the propagation of seismic energy can be described by the acoustic wave equation

$$\frac{\partial^2 \phi}{\partial t^2} = c^2 \nabla^2 \phi.$$  

That equation is the mathematical expression of a physical model in which energy is propagated in a lossless medium solely by means of variations of pressure.

It is well established that many of the defects of the ray propagation assumption of the geometric-optic model are mitigated or effectively eliminated by use of the wave-equation model. Unfortunately, these advantages are gained at the cost of orders of magnitude increases in the required computations.

This increase in computational load is inherent in the difference in the physical models. In the geometric-optic model, the reflection from a point on an interface obeys the specular rule—the angle of incidence is equal to the angle of reflection. Thus, an impulse from a particular surface point of origin can reflect at a given point on the reflector to only one receiver point on the surface. To process the information relevant to one subsurface point, it is only necessary to treat a number of data values equal to the design redundancy of the system. (Figure 2)

**Figure 3**—The acoustic model of reflection seismology postulates that energy is radiated in all directions from every point on a reflecting surface. Accordingly, every detector receives some energy from each reflecting point in response to a given impulse.

In the acoustic model, however, the reflection point is a new source and retransmits energy in all directions. Hence, every receiver records energy “reflected” from every point on the reflector in response to an impulse. Conceptually, we should use all data from the entire survey to determine the characteristics of each reflection point. While the details differ, operationally the process is analogous to the creation of an image corresponding to a hologram. (Figure 3)

Obviously, an immediate and complete shift from the ray-theory to the wave-equation approach would have required more computer power than is available in the entire world. In any event, the necessary analytical foundations for such a change have not yet been developed.

What has happened is that a working mix of the two models has been evolving. In its earliest manifestation the mix consisted of ray-path processing until a final compressed, enhanced data set is produced, following by a single-application of a limited-parameter version of the Kirchhoff-integral solution to the wave equation. The number crunching required for this process when applied to 48-fold-compressed data was of the same order of magnitude as that needed for all the preceding steps.

In today’s world, a battery of programming techniques and analytical simplifications have made the wave-equation approach more economical and the need for better data has made funding increasingly available.

In the most sophisticated approach known to me to be in commercial use, the data are processed at least part way in the geometric-optic manner to derive model parameters which are then used for extensive acoustic-wave processing. The result of that processing is then further modified by a technique derived from geometric optics to compensate for an inadequacy in the wave-equation formulations available today. (Figure 4)

Another few years should see a fully mature acoustic-wave technique to carry the entire processing sequence, but it will almost certainly require ray-path processing as a preliminary step to determine initializing or first-approximation parameters. Much of the processing may be iterative.

For those who are inclined to worry about a possible static future, I would point out that the introduction of attenuation into the equations should suffice to keep everyone busy for some years beyond these projections. After that we could profitably spend a generation or two applying
the tensor equations that describe the elastic behavior of real solids.

HARDWARE REQUIREMENTS

Since the seismic reflection industry went digital in the mid 60's, it has been pressing hard at the frontiers of computer technology. The first group to undertake regular computer processing of seismic data designed and built a special purpose system because of the limitations of the then-available GP systems. Seismic industry requirements soon stimulated the development by computer manufacturers of specialized hardware to cut computing costs for some heavily used algorithms and thus bring overall costs to an acceptable level. In the initial years, correlation and convolution operations were the pacing items making up computer costs. For that reason the first hardware specialties were a fast multiply-add instruction, followed within a matter of months by hardwired correlators. In 1966 IBM announced the 2938 array processor and made its first installation at the beginning of 1968.

The 2938, designed in collaboration with Western Geophysical, had an instruction set of about 12 instructions with data arrays as operands. Pipelining and other hardware innovations resulted in great speed improvements. Geophysical processing could, as a consequence, proceed for almost a decade without serious hardware constraints by taking advantage of normal GP computer developments to complement the array processor.

Today's array processors come in a bewildering variety of styles and capabilities. Married to a mini, one of today's better array processors will, on seismic data, outperform a GP computer and do it at less than 25 percent of the cost of the GP machine.

Seismic reflection data today, and for some time to come, consist of observations arrayed in a three-dimensional matrix. One dimension of the matrix, the one along the line of survey, is essentially unlimited but, as a practical matter, is limited to a few thousand values. The vertical (or time) dimension of the matrix is usually about 3000 data values; and the third dimension, representing the simultaneous observations for each input pulse, is in the process of growing from a present typical value of 50 to about 500 or 1000 in the next few years. For some applications in which an area is surveyed in detail (so called 3-D surveys), a fourth dimension of a few tens or hundreds of values is appended.

It is becoming more common to use extended inputs such as coded or pseudo-random pulse sequences—a process which greatly increases raw data volumes.

The only way all of these data will ever be processed is by extending the technology of specialized devices exemplified by the array processor.

I visualize the processor of 1985 built around a gigabyte high-speed memory with multiple access modes. Only a handful of algorithms will have to be applied to the data. The processor will have hundreds or perhaps thousands of identical single-chip microprocessors to execute each of the required algorithms. The micros will be used in a combination of parallel and pipeline modes so that all are operating simultaneously. All of this activity will be under control of a relatively small CPU which can handle traffic control and the calculation of processing parameters.

In operation all or a sizable chunk of the raw-data matrix will be read into main memory from the compact storage medium on which they were recorded in the field. A succession of operators will be passed through the matrix in various directions and various orders by directing sequences of data from memory through the micros.

Technology visible today can make the whole project feasible; and as Jules Verne has so aptly said, "What one man can imagine another man can do."