A KNOWLEDGE-BASED APPROACH TO FULL WAVE DATA PROCESSING

by

Marc H. Larrère

Ingénieur IST–Géophysique de l'Université Paris 6
Paris France
(1984)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE IN
EARTH AND PLANETARY SCIENCES

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1987

Signature of Author ....
Department of Earth, Atmospheric, and Planetary Science
January 16, 1987

Certified by ....
Prof. M. Nafi Toksöz
Thesis Supervisor

Accepted by ....
Prof. Theodore R. Madden
Chairman, Department Committee

A Knowledge-Based Approach
to Full Wave Data Processing

by
Marc H. Larrère

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
on January 16, 1987 in partial fulfillment of the requirements for the
Degree of Master of Science in Geophysics.

Abstract

A new approach to the processing of sequences of full waveform acoustic logs is investigated. The rationale for this approach is primarily based on the observation that processing and interpretation tasks strongly depend on each other. Hence, a system that incorporates geologic knowledge in data processing naturally and uses processing results for petrophysical evaluation can improve the overall geological interpretation. The implementation of such ideas requires the use of a versatile computer environment, allowing numeric and symbolic processing. The new generation of *Lisp machines* satisfies these characteristics.

An interactive environment for the processing of sequences of acoustic signals was designed using *object-oriented* programming. The package includes a novel method for acoustic full waveform signal matching that uses *dynamic time warping*. The system is tested on synthetic data and field data are processed.

The feasibility of an expert consultant system for full waveform processing and interpretation was investigated by implementing a prototype *knowledge-based* system performing qualitative reasoning in the rock physics domain. Examples of reasoning processes are presented and discussed.

Thesis Supervisor: M. Nafi Toksöz
Title: Professor of Geophysics

Acknowledgements

I would like to thank my advisor Nafi Toksöz who aroused my interest for artificial intelligence. Nafi certainly is an excellent ***patron*** and knows — almost — as much about wines as about geophysics.

I am very grateful to Denis Schmitt and Gilles Garcia for all their help during this thesis. The synthetic seismograms were generated with Denis' over-optimized and ultra-stable programs. Denis also proofread full waveforms related topics and helped me for some text editor translations even when he was working 28 hours a day to write papers — with too many figures. Not satisfied with consulting all the articles in Gilles' library and benefiting from his luminous explanations, I stole his colored pencils, squatted his office and finally destroyed his terminal. May he forgive me.

Yves Bernabe taught me a lot, including that research is still possible without a computer, provided there is observation and intelligence.

Thanks to Stephen Gildea, who implemented nice text editing facilities, and despite Pierre Tarif, who tried hard to discourage me with ubiquitous red corrections, typesetting this text was a sinecure.

I must acknowledge my Turkish officemates Cengiz Esmersoy, Sadiye and Fatih "*Şoför*" Güler for contributing to the enrichment of my foreign language skills. Fatih also mastered very quickly basic French five letter words. Jarry's six letter version, however, was more difficult to catch.

I am indebted to Patricia Arditty who made my stay in Cambridge possible and to Jean Pierre Panziera for bringing me efficient anti-homesickness cassettes.

The dynamic time warping algorithm and the Lisp function for plotting warping paths were implemented by David Lineman.

This work was supported by the Full Waveform Acoustic Logging Consortium at MIT and by an Elf Aquitaine fellowship.

Contents

Chapter 1

Full Waveform Acoustic Logging

1.1 Introduction

Well logging measurements are records of the variation with depth of a physical property of rocks in boreholes. The spatial resolution in depth is very high (order of 10 cm) compared to classical resolution for surface geophysical methods (ranging from 10 m to 500 m). Classical surface seismic methods give a geometric picture of the subsurface showing large stratigraphic units. However, a precise knowledge of the lithology and of the petrophysical properties of formations, such as porosity, permeability, fluid saturation, necessitates in-situ measurements. Thus, downhole measurements are devoted to provide detailed geologic characterization. One of the most widely used property for this purpose is the travel time of compressional acoustic waves. In classical sonic logging, the onsets of P-waves refracted through the formations are picked in real time, allowing the computation of compressional velocities. In full waveform logging, the entire wavetrain propagating in the borehole-formation medium is recorded with a multi-receiver acoustic tool. Figure 1-1 shows a typical tool configuration during data acquisition. The sampled wave field theoretically enables to recover P- and S-wave velocities and attenuations, mechanical parameters of the formation, and furthermore estimates of other petrophysical properties such as porosity and permeability. Since a more accurate and reliable lithological identification is gained through the additional knowledge of S-wave velocities and P- and S-waves attenuations (see Winkler and Nur, 1982; Newman and Worthington, 1982) full waveform logging is a powerful tool for formation evaluation. The full waveform logging technique presents also the important advantage of giving results in cased boreholes.

The drawback of the method is the large amount of collected data that precludes manual interpretation and calls for automatic methods to determine the elastic parameters. Thus, digital signal processing techniques must be developed to extract primitive information from the digitized waveforms. A brief review of the physics of wave propagation in fluid-filled borehole is necessary before describing classic processing techniques for the determination of formation velocities and discussing their main limitations.

1.2 Modeling the borehole waveguide

The borehole-formation medium composes a waveguide in which different types of waves can propagate. Their existence, energy, and frequency content depend on the geometrical characteristics of the borehole environment, the mechanical properties of the formation, and the frequency band of the source. The superposition of these diverse waves associated with different propagations of elastic energy forms an overall complex arrival.

In the simplest configuration, the borehole environment can be represented as a fluid-filled cylinder axially infinite surrounded by a radially semi-infinite homogeneous formation. More realistic and complex geometries consider radially layered formations, which enables to represent well- or poorly-bonded cased-holes as well as for the presence of a damaged zone in the vicinity of the well bore. Figure 1-2 shows the principal geometrical configurations representing common real-world situations. A cylindrical elastic tool in the center of the well can also be taken into account.

A complete modeling of wave propagation in a borehole requires also the description of the mechanical behavior of the formation. The simplest model assumes that formations are elastic and isotropic. More complex formulations consider multi-phase porous media. In the elastic case, for an open hole and provided that P- and S-wave velocities in the formation are greater than the P-wave velocity in the borehole fluid (the so-called "fast" formation case), the signal received at the center of the hole is composed of 4 types of waves; two body waves (compressional and shear [1]) and two guided waves (pseudo-Rayleigh and Stoneley). Properties of these waves in the elastic case have been discussed by many authors (Biot, 1952; White and Zechman, 1968; Tsang and Rader, 1979; Cheng and Toksöz, 1981). When the S-wave velocity in the formation is lower than the compressional fluid velocity (for a "slow" formation), the S-wave and pseudo-Rayleigh wavetrains are no longer recorded, but the Stoneley wave carries information about the shear-wave characteristics of the formation.

When radially non-homogeneous formations (succession of concentric annuli) are described, the situation becomes more complex (see Tubman, 1984 ; Schmitt and Bouchon, 1985). In particular, the presence of a casing can generate ringing arrivals if velocities in the casing and in the formation are comparable. The bonding conditions between the different layers (casing-cement and cement-formation) and their respective thicknesses determinate the relative amplitudes of waves. The case of porous formations was modeled using Biot's theory (1956 a,b) by Rosenbaum (1974) and Schmitt (1985). These studies show that waves' propagation, and especially guided waves' propagation, is also dependent on non-elastic parameters such as porosity and permeability, fluid formation characteristics, and borehole wall conditions (see also Burns, 1986).

The various theoretical forward models show the complexity of wave propagation in the borehole environment. The trend in modeling has been towards describing more realistic and complex geometries accompanied with more detailed and specific physical representations of the media surrounding the borehole. These studies indicate that the existence of waves and their responses to the different formation properties depend on geometrical parameters, elastic properties, as well as on petrophysical determinants (such as porosity and permeability values, saturant fluid). In fact, the choice of the adequate wave propagation model (i.e. geometri-

[1] This wave corresponds in fact to the S- to P-wave conversion of the shear wave refracted through the borehole wall. Its phase velocity is equal to the velocity of the S-wave in the formation.

cal configuration plus mechanical type of the formation) is a requisite before a more refined estimation of in-situ properties. This determination can be obtained in most cases with the knowledge of a set of highly determinant parameters and thresholds such as presence of casing, the quality of cement-formation bonding, the type of formation (fast or slow), and the porosity. Reciprocally, these thresholds have a strong influence upon the waveforms' main features so that model characteristics can be recognized when looking at the waveforms.

1.3 Data processing

The primary purpose of processing is to obtain an automatic, accurate and robust determination of P-, S-, and Stoneley wave velocities for all borehole environments and for all formations. This goal is reasonably well achieved by digital processing methods in the simple configuration of an open borehole and for a "fast" formation. Most techniques provide good results in well-bonded cased holes. Difficulties arise when logging in poorly bonded cased boreholes and in any borehole when the shear wave is weak or absent ("slow" formation). The diverse methods estimate the formation shear wave velocity with the first cycles of the pseudo-Rayleigh arrival. The estimation of other wave properties (attenuation, amplitude, frequency) is possible only when the position of compressional and shear arrivals in the wavetrain is located.

Automatic velocity determination

Statistical techniques are the basic approach for velocity determination since they are robust. Two strategies were adopted to obtain velocities from full waveform tools with arrays of receivers — the choice of a particular technique is motivated in part by the tool characteristics, i.e. the number of receivers and the offset range.

1. The first approach consists in computing move-outs on a *couple* of waveforms, this operation being repeated (if required) for the whole array at a given depth. In the time domain, the different methods use some type of cross-correlation (with or without normalization)

for an accurate move-out computation. The frequency band can be narrowed and tuned to the expected frequency range of each wave if the correlation is done in the Fourier domain (Ingram et al., 1985). The rationale for this approach lies on the expectation that the separation of arrivals is better in the frequency domain. The separation in the time domain, however, is certainly very good when considering medium to large offsets, which limits the need for such techniques.

All approaches suppose an a priori model of the different waves propagating in the borehole. In practice, the first energy arrival picked by a threshold detection is granted to be a P-wave, then it is assumed there exists an S-wave. In the P-correlated S-method (Willlis and Toksöz, 1983), the S-wave arrival time is estimated by correlating the P-wavetrain with the entire wave. The assumption is that "the shear arrival is similar or greater in amplitude than the P-wave." The frequency domain phase determination method described by Ingram et al. (1985) also assumes there is a S-wave propagating with a sufficient energy. Furthermore, it makes strong assumptions about the frequency band location of the correct phase velocity [2] that may not be correct in all situations.

2. The second approach consists in estimating average velocities from an array of waveforms. In the time domain, this can be done with the *semblance* technique (Kimball et al., 1984), which is a normalized linear slant-stacking of the sequence of waveforms. The algorithm's output is a map of the relative energy of coherent arrivals in the plane (*time, slowness*). A recognition problem is then to be solved, since slownesses corresponding to maxima of energy have to be associated with a given type of wave. In the case of interfering arrivals, resolution can be improved with the Maximum Likelihood Method, but computational cost is then too high for routine applications (Hsu and Baggeroer, 1986).

   These techniques make no assumption about the generation of waves. The only assumption is that the existence of a coherent arrival will give a coherent stacking of energy at a given slowness. Thus, velocity analysis methods can detect the existence of any number of arrivals. This later characteristic makes these methods appropriate in a complex borehole environment where the prediction of the existing waves is difficult. It was shown to be

[2] For the P-wave, the band chosen corresponds to the maximum amplitude; for the S-wave, the phase lag is picked at the central frequency of the S arrival band.

effective for well-bonded cased hole data (Kimball et al., 1984) and in some cases when there is a poor bonding between cement and formation (Block et al., 1986).

Pitfalls in full wave data processing

All processing techniques face an important trade-off between *resolution* in depth and *precision*. Summations are required to improve the signal-to-noise ratio, hence to gain accuracy on velocity estimates and robustness. For instance, semblance analysis must use an array of at least 4 waveforms to give a sufficient precision . This involves averaging measurements, hence a loss of spatial resolution. An efficient solution to overcome this problem consists in making velocity analysis or signal matching on waveforms corresponding to overlapping tool positions (Arditty et al., 1981). This affords to maintain a high spatial sampling rate while stacking signals at the same time, at the price, however, of a costly array processing.

All processing schemes presuppose that arrivals are not dispersed. When arrivals are even slightly dispersive (for instance in formations where intrinsic attenuation is high), we are no longer certain we are estimating the phase velocity. In fact, for average to large spacings between source and receiver, correlation-like and stacking methods certainly provide an estimate somewhere between phase and group velocity. This is particularly true for S-wave velocities that are estimated from the first cycles of the pseudo-Rayleigh arrival. Thus, we obtain a correct phase velocity estimate for hard formations with little attenuation and very likely an underestimate for softer formations.

Difficulties also arise for cased-holes when the P-wave velocity is close to casing velocity or when the bonding is of poor quality. In these cases, correlation methods generally fail to pick the true compressional velocity of the formation. Semblance analysis methods, however, because of the absence of assumptions about wave's existences give better results.

The other weakness that all techniques suffer is the determination of S-wave velocities when S-wave energy is low. Semblance analysis methods cannot resolve the S-arrival in this case. Correlation methods may give an estimate of shear wave velocity, because of the presence of correlated noise, even when the arrival is absent (i.e. in slow formations). In most cases, the uncertainty about the S-wave existence is not resolved, and additional processing and interpretation must be taken into account to obtain decisive information.

The general pitfall in automatic velocity determination is the large variability of signals. There exists not a unique, but several possible configurations of waves, depending on a somewhat limitated set of geometrical, mechanical and physical parameters. An ideal processing scheme should be context-dependent to behave correctly in all field situations. In practice, when a priori constraints are given, accuracy and computational cost are improved, at the depend of robustness. Without a priori assumptions, unconstrained algorithms may give answers in almost any cases. In fact, the important problem not addressed by the different processing techniques is the primordial choice of the model corresponding to the actual physical situation. The interpretation phase and the processing of signals must not be totally independent in order to make possible the adaptation of processing operations to diverse real world configurations.

1.4 Full waveform and formation evaluation

The main contribution of full waveform processing is to bring information about shear wave properties. Recording the entire wavetrain enables one to recover the characteristics of the S-wave from the low frequency portion of the pseudo-Rayleigh wave for all configurations (see Burns, 1986). The knowledge of the S-wave velocity makes possible the development of novel interpretative methodologies based on the conjoint utilization of compressional and shear waves characteristics. Information about the pore structure of rocks can be deduced from the P- and S-wave velocities (Cheng and Toksöz, 1979). For most lithologies, P- and S- waves velocity ratios (or Poisson's ratios) are powerful indicators. Moreover for clastic rocks V_p and V_s were shown to be linearly proportional (Castagna et al., 1985). The additional knowledge of the P- and S- wave attenuations has also a high potential for lithology determination and formation evaluation, in particular to characterize the saturant fluid (see Newman and Worthington, 1982; Winkler and Nur, 1982).

The estimation of formation permeability is another important application of full waveform logging. Theoretical studies (Schmitt, 1985) show that relative values can be deduced from the characteristics of the Stoneley wave (velocity and attenuation). However, more accurate analyses necessitate to take into account P- and S-waves velocities and attenuations and to compare the respective attenuations of the pseudo-Rayleigh and Stoneley waves.

1.5 The knowledge-based approach

Sources of knowledge

In the following chapters a new approach for the digital processing of sequences of full waveform data is investigated. This approach will allow the integration of relevant knowledge in the treatment, and therefore the ability to make processing operation dependent from the geologic context. The knowledge necessary to reach this goal comes from three primary sources:

1. Description of the geologic and physical environment, i.e. rough information about the lithology (especially shale content) and the main physical characteristics of the borehole (diameter, presence of casing, etc.).

2. Knowledge of the physics of wave propagation in fluid-filled boreholes, gained by theoretical modeling and studies. This includes the description of the relationships between physical properties of formations surrounding the borehole and properties of waves.

3. Knowledge about the signal itself, i.e. about waveforms' attributes and principal features (presence and location of energy peaks, frequency content, etc.).

These three types of knowledge are required for the processing and interpretation tasks and come forth at different stages. Initially, basic and even crude information about the borehole and the formation can provide a general insight for the choice of the wave propagation model. Let us consider a simple example to illustrate how the different types of knowledge are taken into account for the understanding of signals. If we deal with a shallow zone in an open borehole, knowing from a γ-ray log that the lithology is mainly shale, we may want to check first if there is some energy at the reasonable arrival time for a S-wave in such a formation, instead of directly applying a correlation-like method that will give no solution or a wrong one. This decision involves two types of knowledge:

- First, from a simple geologic description of the zone of interest we are able to infer a qualitative estimate of petrophysical parameters; this first step tells us that the S-wave velocity may happen to be lower than the borehole fluid velocity — which is known to be about 1500 m/s.
- Second, we know that if the S-wave velocity in the formation is less than the compressional fluid velocity, the S-wave cannot be recorded. This is the physical understanding of the wave propagation problem.

Given this knowledge, we can determine that S-waves may or may not be present and test for the waves' existence, i.e. estimate the energy in a reasonable S-wave time-window. Clearly, by considering geologic and physical knowledge, we improve the adequacy of processing actions and furthermore improve their efficiency and accuracy, since we are able to add constraints [3].

Drawing inferences about the geologic environment and wave propagation will be generally unsufficient. After all, once we have picked the P-wave arrival time, we are able to infer more about the other components of the wavetrain, hence we can put more constraints on the following processing actions. For instance, for a clastic rock, a linear relationships between P- and S-wave velocities holds (Castagna et al., 1985). This knowledge provides an initial estimate of the S-wave arrival time. This type of feedback between sources of information involves using newly acquired knowledge about data as well as about characteristics of the signal.

A knowledge-based approach to geophysical processing certainly necessitates an adequate representation of relevant geologic knowledge and of physical knowledge about the wave propagation problem. The conjunction of these two sources of knowledge enables to draw inferences

[3] In the shale formation example, the qualitative geologic information also provide information about the range of the S-wave arrival time — if the S-wave is recorded at all.

which determine a model of propagation and constrain the choice of the processing operators and of their parameters. The limitation of such an approach is that the final output of the reasoning process will be a rather rigid automatic planning of (hopefully) optimal processing operations. However, this may not be sufficient in the full wave logging domain for two reasons. First, processing results are not the ultimate goal, but rather a step toward a complete lithological identification and an estimation of the physical properties of formations. Second, an intelligent system must allow for different processing choices based on intermediary results, i.e. provide enough flexibility for changing the initial planning when new information is acquired. Therefore, a more general and useful structure should favor the manipulation of processing operators, so that they can be readily embedded in the overall interpretative system.

1.6 Overview

Building a system that naturally favors symbolic manipulations of procedures and arguments, hence the integration of relevant knowledge during the processing, requires flexible processing tools with three main design requisites:

- The first requirement is the ability to handle and apply operators independently one by one, without presupposed ordering or interdictions except mathematical incompatibilities.
- The second is being able to start over an operation, which involves that previously transformed signals and intermediary results are in any situation readily available.
- Finally, an efficient system for user interaction calls for an easy access to accumulated information and efficient displays of results.

An interactive system for the processing of arrays of waveforms that satisfies these specifications was designed and is described in chapter 2. The modular architecture due to *object-oriented* programming is outlined and a new processing tool for accurate velocity analysis and signal matching with *dynamic time warping* is described. Applications to synthetic and real full waveform data are presented and discussed.

In chapter 3, *knowledge-based* reasoning techniques are used to draw inferences in the petrophysical domain. The reasoning system is intended to provide a more complete description of the geologic environment based on an initial set of qualitative field data. A frame-like representation for lithologies is used as a foundation for rule-based reasoning about the physical properties of rocks. Examples of reasoning processes are presented and discussed.

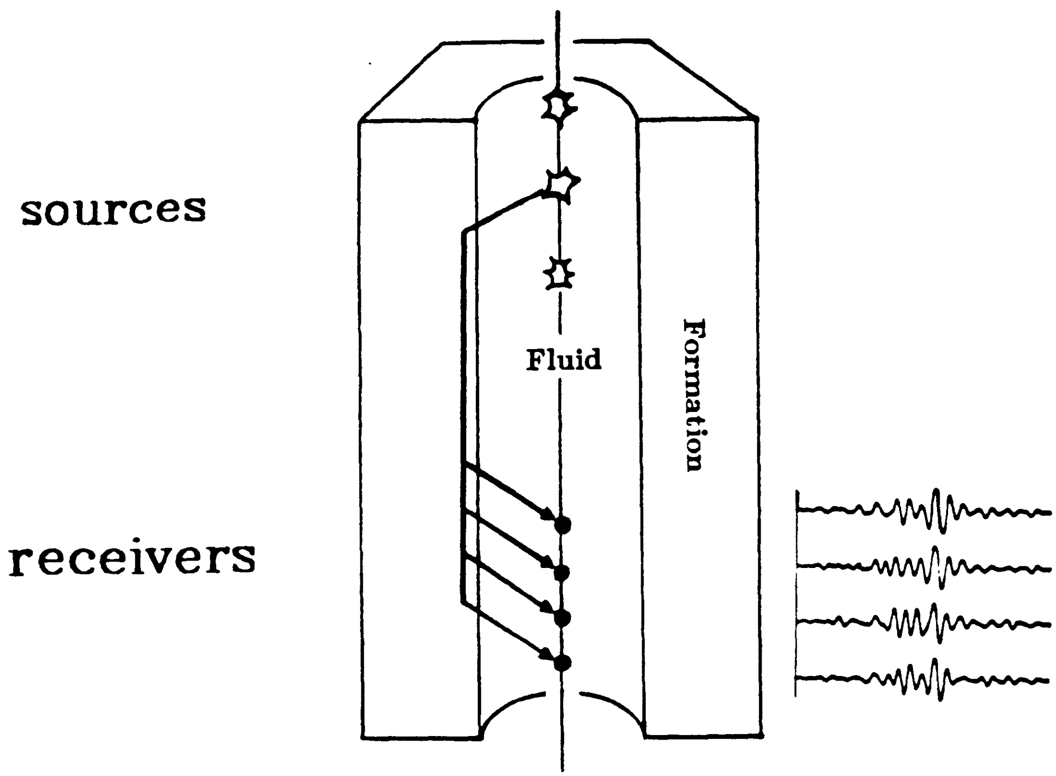


Figure 1-1: Full Waveform Logging with a multi-transmitter, multi-receiver tool.

After Garcia and Cheng, 1985.

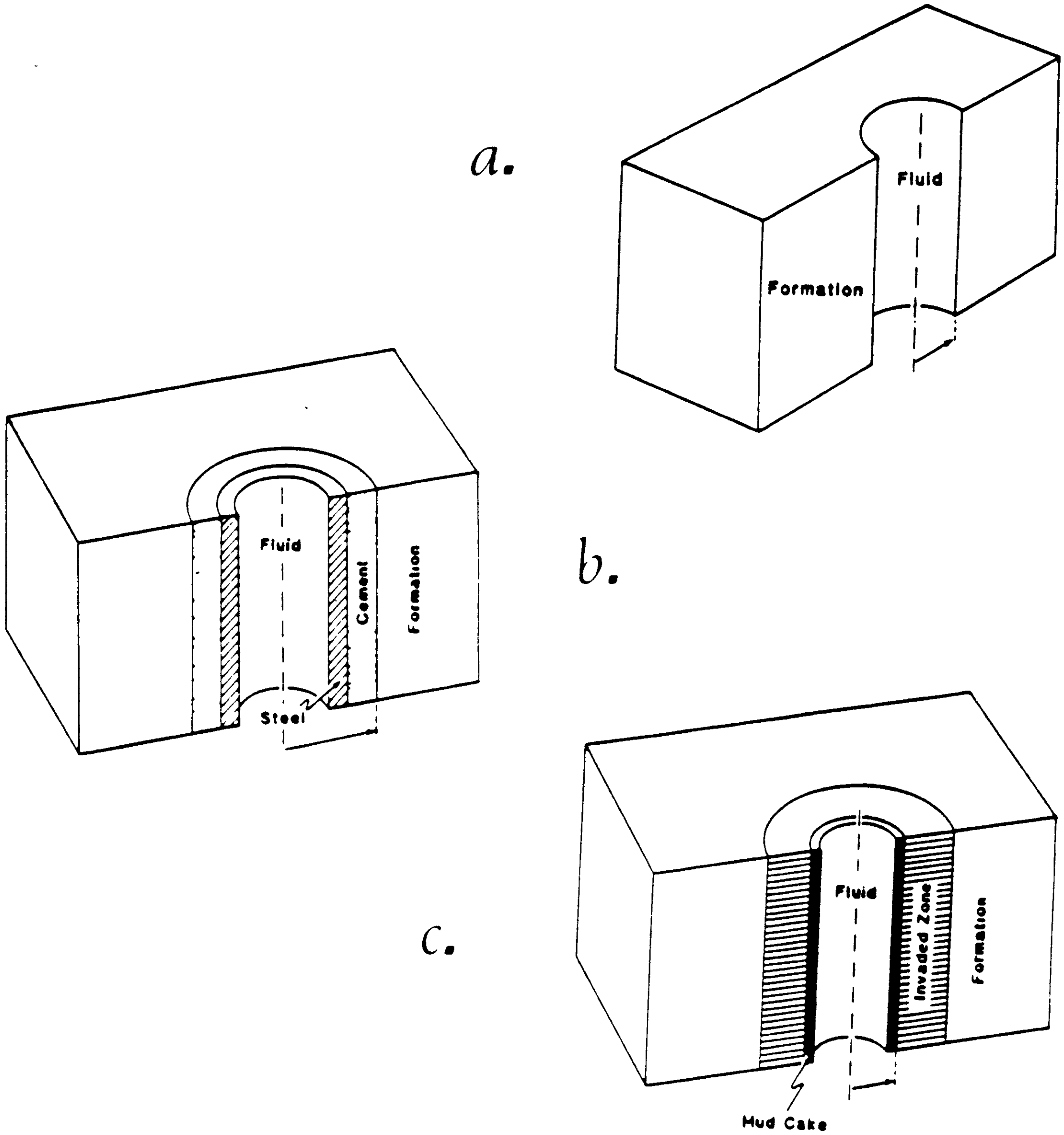


Figure 1-2: Different Borehole Geometries:

a - The open borehole. The formation can be elastic or porous.

b - Well bonded cased borehole with layers of steel and cement. Poor bonding situations can be modeled by inserting an additional fluid layer between the steel and cement or between the cement and the formation.

c - Open borehole including a mud-cake and an invaded zone; this zone is created by the filtration of drilling mud in a porous formation.

After Tubman, 1984.

Chapter 2

An Object-Oriented System for Full Waveform Data Processing

2.1 Introduction

The motivation for the ANIS [1] system is twofold: first, provide an interactive processing environment for *sequences* of full waveforms (as well as single waveforms), second, enable to combine numeric and symbolic operations on signals. These two goals are complementary since achieving these tasks require a flexible structure oriented toward an easy manipulation of arrays of signals, elementary signals, and segments of signals. The general philosophy of the system is to make few assumptions about the specific methodologies of processing. The implementation on *Lisp machine* uses object-oriented programming and takes advantage of the powerful programming environment — especially for graphical applications.

A determinant design choice was to define the concept of *sequence* of waveforms as the elementary object, as opposed to more general-purpose data processing systems that represent isolated signals (see for instance Kopec, 1984; Dove et al., 1984). This choice is essential in

[1] Acronym for A Nice Interactive System

full wave acoustic data processing where the principal processing operations concern arrays of waveforms. It would be awkward to implement a velocity analysis or a controlled threshold detection technique if the elementary concept were a single signal. User interaction is an important facet of the system; Appendix A illustrates the "style" of interaction and demonstrates the use of some operators and the geophysical applications of the ANIS system.

Since the system's philosophy and performance are strongly influenced by LISP programming and more specifically object-oriented programming, the main characteristics of these programming techniques are briefly described before describing the system's structure, the processing operators and presenting applications to full waveform acoustic data.

2.2 LISP and object-oriented programming

LISP [2] is a language primarilily devoted to symbol manipulation that originated at the same period as FORTRAN — late fifties. It is being widely used since suitable hardware has became available. LISP is a functional language, i.e. most programming is done by combining existing functions at different levels of specialization rather than by describing a sequence of operations. This process, called *procedural abstraction*, favors the partition of the task to more manageable subtasks and therefore makes incremental programming easy. LISP's structure encourages this type of programming characterized by an "applicative" style close to the composition of functions in mathematics. Furthermore, recursive applications of functions are possible and commonly used to describe procedures. In fact, the representation of programs is done with the same data structure (lists) as any other data. This enables to handle complex programs with the same ease as elementary data. Despite its orientation toward symbolic operations, compiled versions of LISP are also suitable for arithmetic computations (Winston, 1984).

Another strategy to augment abstraction in programming is to encourage the organization of data. Suppose we are to write processing operators for single seismic waveforms; a simple and useful data structure that describes a general concept WAVEFORM will be composed of a

[2] Acronym for List Programming.

time-serie, a sampling rate and some identification. These components are also called the *slots* (or *attributes*) of the abstract data type (or *object*) WAVEFORM. The strength of *data-abstraction* is that pieces of related data can be treated as a unique entity. In addition, specific procedures are defined to construct, access and modify the elementary slots. This suppresses the burden of retrieving and organizing the diverse components for every specific task and enables better programming since, according to Winston (1984), "keeping track of such details can cause brain damage". Data abstraction aids concentrating on high-level concepts and makes programs easier to modify since information is concentrated in well-defined compound structures.

A systematic recourse to data abstraction where objects are also responsible for the management of functions is called *object-oriented* (or object-centered) programming. In object-oriented programming, procedures are attached to objects in much the same way as any other attribute. LISP makes it easy to handle since data and programs are represented with the same basic structure: a list. *Message-centered* languages are subspecies of object-oriented languages characterized by an original syntactic feature: a given procedure attached to an object is executed in response to a message sent by another object. Suppose we have two data types WAVEFORM and SEQUENCE (representing sequences of waveforms). We can associate a procedure "`draw-self`" to both objects that operates differently for a single waveform and for a sequence of waveforms. The adequate response is given when an instance of SEQUENCE or WAVEFORM receives the message "`draw-self`". The specific details for the actual execution of the task, however, are transparent for the higher level operations.

ZetaLisp is a dialect of LISP that includes a *message-centered* language called the *flavor system*. Flavors are non-hierarchically structured objects that can be mixed together to form a new concept. The "mixed" flavor inherits the attributes of each parent flavor as well as attached messages (also called *methods*). Invoking the application of a method is called message passing. A thorough description of the principles of message-centered programming and flavors can be found in Winston (1984).

In pure message-centered programming, an operation can occur only when an object sends a message to another object. In practice, every operation does not need to be initiated by message passing and low level procedures are performed via basic LISP functions. Operators on arrays are also written in LISP since compiled *ZetaLisp* is as fast as FORTRAN for arithmetic operations and includes powerful built-in function for the description and manipulation of arrays. [3]

2.3 The ANIS system

The ANIS system owes much to object-oriented programming concepts as described in the section below. The description of a few basic data types forms the core of the processing environment. This confers the ability of a fast and simple access to data, operators, and results. Another advantage of an object-oriented design for signal processing is that the development of processing algorithms and the practical utilization on real data are done with a unique language (Kopec, 1984). Thus, the tasks of development and utilization can be tackled within the same environment, which allows incremental improvement of the processing operators. Also, this type of structure is very well suited for encapsulating the processing operators into a knowledge-based system, both from the standpoints of data and result description and of the planning of processing operations.

Since data and results are represented as abstract objects, they can be easily manipulated and accessed. The only drawback of the actual implementation may be that instances of objects and attributes have no memory of their past values, i.e. application of an operator twice leads to loss of the first result. This type of bookkeeping can be handled by a higher level object structure. Since processing operators are defined as messages attached to data structures, they are manipulated with the same ease as data. Their applications can be easily controlled by high level constructor procedures.

The basic data-types for signals in the ANIS system are TRACE and SEQUENCE representing respectively waveforms and arrays of waveforms. Processing operators are attached to sequences

[3]Basic mathematical operators on arrays could also be implemented with FORTRAN subroutines, using either network links or the *Symbolics* FORTRAN.

and/or traces depending on the nature of the task they perform. Some operators applied on certain classes of data can produce side effects, i.e. related processing results are assigned to the adequate slots of traces. Initial arrays of waveforms can be transformed with operators or they can be segmented: this leads to the creation of new instances of specialized sequences, respectively `TRANSFORMED-SEQUENCE` and `SUB-SEQUENCE`. Thanks to the object-centered structure, any creation of more specific instances confers also the ability of invoking the same collection of operators and of accessing all relevant information. In particular, "images" (i.e. graphical representation of objects in windows), stay physically present and are readily available in the environment. Again, an illustration of the possibilities of the system, with the help of practical sessions, is given in Appendix A.

General structure

The objects in `ANIS` are structured hierarchically. Two main concepts are defined at the root of the tree, representing for one part waveforms and arrays of waveforms (`DATA-TYPE`), for another part abstract data types useful for results and for the practical implementation (`ABSTRACT-TYPE`). A concept subsumed by other objects inherits their slots. Figure 2-1 shows a portion of the hierarchical relations between `ANIS` objects subsumed by `DATA-TYPE`. The data type `SUB-SEQUENCE` has an *instance* `sub-sequence-03` that is an actual piece of data. It is subsumed by the object `SEQUENCE` which is a specific `DATA-TYPE`. The list of the definitions of objects is presented in Appendix B. The central concept is the data type `SEQUENCE` that embodies two important slots, `image` and `list-of-traces`.

`Image` represents the abstracted part of the object `SEQUENCE`, related to graphical representations. Processing operations are primarily attached to `images` of sequences rather than to sequences themselves.

`list-of-traces` relates a sequence to its primary components i.e. the individual waveforms, represented by the data type `TRACE`.

Since every data type slot is assumed to be an instance of some defined object in the environment, the overal structure forms a description of the *semantic* of the domain, i.e. of the *meaning* of links between the various concepts. A piece of this network is shown on Figure 2-2. This network shows, in particular, that a SEQUENCE has an image, which is an instance of the particular object SEK-IMAGE, that is itself an ABSTRACT-TYPE containing other members of ABSTRACT-TYPE called REGIONS that contain DATA-TYPE objects.

Processing operators

Processing operators in ANIS are messages that can be sent to instances of SEQUENCE or TRACES. The listing of theses messages is given in Appendix C. Most operators are built on lower level array processing functions. The most important operators on SEQUENCE are:

1. Operators for single signals, generalized for sequences of traces, including normalization, interpolation with cubic splines, estimation of maxima in time-windows and computation of envelopes via moving average.
2. Two methods for picking:
   - Automatic threshold detection for all kind of waves. The legal time-band for picking is restricted by taking into account the nature of the wave.
   - Manual picking with the mouse [4]. A minimum of two picked points is required. The values for other waveforms in the sequence are linearly extra- or interpolated.
3. An accurate computation of move-out between traces using *dynamic time warping* [5] with the possibility of interactively adapting and optimizing the parameters — i.e. the position of windows and the number of points.
4. Waves' dispersion study in the time domain — computation of phase velocity variations as a function of the length of the path of propagation.

[4]This technique is not intended to give precise arrival time estimates since there is the limitation of the initial sampling rate. Nevertheless, it can provide high manual precision picking if done after spline interpolation.

[5]The technique is described in the next section.

Some operators are very general and can be invoked for any type of sequences and traces, other are restricted to certain data types. The two following examples illustrate why the field of application of operators is sometimes restricted:

- The message "`envelope`" can be sent to any type of traces and sequences, including fragments of sequences and already transformed sequences. The operator is not task-dependent, hence messages for traces and for sequences are built on the same very general LISP function.
- The message "handpick-arrival-time" is only defined for instances of `RAW-SEQUENCE` of at least two traces, and does not make sense for an instance of `SUB-SEQUENCE` except when the value of the type slot of `SUB-SEQUENCE` is P-, S-, or Stoneley waves.

Processing results are described by two objects linked with the instances of `TRACE`. These objects are `INITIAL-PROCESSING-VALUES` and `FINAL-PROCESSING-VALUES`. Picking methods (i.e. automatic threshold detection and manual picking) fill the "initial-values" slots of waveforms. The initial arrival time values are then used to compute velocities with signal matching and the results are affected to the "final-values" slots of waveforms. All results, as well as the history of operations applied to a given instance of `SEQUENCE` can be retrieved with the help of specific messages (see the list of operators in Appendix C).

2.4 Signal matching with dynamic time warping

Dynamic time warping

Dynamic time warping can be regarded as a generalization of cross-correlation that allows not only shifting but also stretching and squeezing of one signal with respect to the other. A general signal matching problem consists of estimating a mapping function between two time series at any point in time. This mapping function must be such that it minimizes a given measure of dissimilarity or *distance* between the two signals. Therefore, the problem can be formulated as an optimization problem, and was tackled with two different approaches:

- A non-linear least-square inversion for estimating the mapping function as a sum of simple analytical functions. Martinson et al. (1982) used truncated Fourier series and applied successfully this technique to geophysical data.

- The problem can also be formulated as a search in the two-dimensional discrete space of accumulated distances between the two signals. Sakoe and Chiba (1971) proposed an algorithm using dynamic programming. The technique, called ***dynamic time warping***, was widely used and developed for speech recognition problems (see Rabiner et al., 1978). Anderson (1983) reviews some possible applications in the geophysical domain, among which waveform classification and well-to-well log correlation (Lineman, 1986) were developed. He suggested the use of dynamic time warping for the processing of entire sonic waveforms.

Figure 2-3 shows the dynamic time warping problem for two discrete signals a_i and b_j of respective lengths N and M. We need to determine a discrete mapping function $c_k = [i(k), j(k)]$ that corresponds to a minimum distance between each couple of samples. Choosing a local cost function $d(c(k))$, we are to minimize the overall cost function $D(c) = \sum d(c(k))$. This problem is equivalent to a path finding problem in the $N \times M$ discrete domain of accumulated costs (see Figure 2-3). Given constraints on endpoints and with the definition of the legal local moves (the set of allowed moves from any point $[i, j]$ to its neighbors), it can be shown that the minimum cost path from the origin $[0, 0]$ to any point $[i, j]$ is independent of what happens beyond this point. The minimum cost path is determined recursively by minimizing more and more local costs. An optimal path finding algorithm using dynamic programming can be applied to evaluate the mapping function c_k (Sakoe and Chiba, 1971). A complete description of the algorithm can be found in Parson (1986).

Application to full waveform processing

The mapping function $c_k = [i(k), j(k)]$ is a representation of time shifts between the two input signals for all samples. The time-shifts are $\delta t_k = | i(k) - j(k) |$. Figure 2-4-a illustrates the case where the two signals are identical: we have $i(k) = j(k)$ for all k, hence c_k is a straight line between the initial and final tie-points, and $\delta t_k = 0$, for all k. Figure 2-4-b shows that if two identical signals are shifted by a constant number of samples s (corresponding to a time move-out t_s), the theoretical mapping function is a straight line beginning at $c_1 = [0, s]$. If the time shift between the two signals increases with time, as shown in Figure 2-5, the mapping function departs from the constant slope. The dynamic time warping technique presents two interests for full waveform matching applications in the context of the ANIS system.

- The method is potentially very accurate for recovering the variation of move-out with time due to wave's dispersion between a *couple* of waveforms.
- The nature of the algorithm enables to control the mapping and to set constraints in order to limit the space of possible matches. These constraints can drastically prune the search tree, hence make the matching computationally effective.

This signal matching technique was applied to two slightly distinct purposes: first, to perform fast correlations on short windows to estimate travel times (and therefore waves velocities); second, to make detailed analyses of the dispersion of arrivals in the time domain. The first task could be addressed with traditional cross-correlation techniques since the very beginning of arrivals is in general not dispersed. However, for dispersed waves (i.e. PL modes, pseudo-Rayleigh and Stoneley) determining the phase velocity as a function of time can be done by non-linear matching techniques.

Velocity determination The determination of a wave velocity with dynamic signal matching involves five steps:

1. Determination of the arrival-time t_0 of the wave for each waveform in the sequence using a fast picking method.

2. Windowing the arrival around t_0; the window length depends on the dominant frequency — for instance, the window is longer for the S wavetrain for the P-wave.

3. Interpolation of the windowed waveform with cubic spline. The interpolation factor depends on the precision required.

4. Correlation by dynamic time warping with a maximum time shift constraint. The maximum time shift is the result of a trade-off between confidence in the first estimate and computational cost.

5. Computation of the wave velocity from the initial move-out and the dynamic time warping time shift.

The process is repeated for every couple of waveforms in the sequence. The first step is essential for the reliability of results, especially in the case of a S-wave detection. For an interactive process, P- and S-wave first arrival estimates are obtained with automatic or manual picking and errors can be easily and quickly corrected. For an automated process, a priori assumptions must be made (choice of a model) and control procedures must be set in order to check the validity of the picks.

Dispersion studies Signal matching with dynamic time warping allows the study of the dispersion of arrivals in the time domain for two waveforms. This information could also be obtained in the frequency domain or via τ-p transform but these methods require to work on arrays of waveforms. The study of dispersion is restricted to a part of the waveform for two reasons:

- The measure of similarity between signals emphasizes the resemblance of the prominent waves. Because high amplitude arrivals have a prevailing contribution on the cost function, weak arrival are matched less accurately. For instance, in a hard formation and for short offsets, the method cannot resolve the P-wave time delays because of the dominant energy in the pseudo-Rayleigh wave.

- The computational cost is too high when two entire waveforms are matched [6] with the high sampling-rate required for a sufficient precision.

2.5 Results

Synthetic microseismograms

Resultsusing synthetic waveforms are presented first in order to test the accuracy of the method. Synthetic microseismograms were generated with the discrete wavenumber method in the case of an open borehole surrounded by an homogeneous formation. The mechanical properties are: $V_P = 4000$ m/s, $V_S = 2310$ m/s, $Q_P = 50$, $Q_S = 25$, $\rho = 2.4 \times 10^3$ kg/m^3. The time sampling rate is 11.84 microseconds. The borehole radius is 10 cm and the center frequency of the source is 5 kHz. Figure 2-6 shows the complete sequence of synthetic waveforms. The source to receiver distances range from 2.50 m to 8.00 m in increments of 0.50 m. The energy in the Stoneley wave is preponderant for all the waveforms.

Velocity determination A P-wave velocity analysis was performed on the synthetic data shown on Figure 2-6. For the sake of illustration, the initial automatic threshold detection step was ill-done. There is a cycle-skipping for the 7.00 m offset. This cycle-skipping was the result of a relatively poor signal-to-noise ratio for large offsets due to the attenuation. Figure 2-7 and 2-8 show the time-windowed P-waves after spline interpolation. The window length is about two and a half cycle and the time sampling is less than a microsecond.

As shown on figures 2-9 and 2-10, the mapping functions are nearly perfect straight lines for offsets less than 5.00 m. The quality of match decreases for larger offsets, as the level of

[6] The computational cost of dynamic time warping is theoretically proportional to the product of the two signal lengths. In fact, the internal building of the recursion slows down the computation when signal lengths pass a given threshold.

numerical noise increases. Note also that the mapping function between the 6.50 m and 7.00 m offsets — involving a cycle-skipping — shows a linear part that corresponds to the maximum shift constraint on the match. This indicates that paths with lower cost could be found if larger move-outs were tolerated. The move-out value is estimated by taking the average of the three most common time-shift values, excluding the extremities of the mapping function. This estimate was found to be more robust than the straightforward average value.

Since the mapping function is constrained to stay in a diagonal band defined by a maximum time shift, the matching with the offset 7.00 m (with initial cycle-skipping) corresponds to a wrong estimate. In order to make the proper correlation the initial window length must be larger so that it includes the first skipped cycle and the constraint on the maximum shift must be relaxed. These increase the computational cost significantly.

Excluding results corresponding to the arrival detected with a cycle skipping, all final velocity values are within 2.0 % of the theoretical value. The average value for the entire array is 3960 m/s (the theoretical value being 4000 m/s). The determination of the S-wave velocity (not shown here) was done with the same relative error. This example is representative of the order of precision of the method as applied to a few sequences of synthetic seismograms. Tests for formations without attenuation showed less deviation. All results, including for P-wave velocity determination, have a systematic negative bias, i.e. an underestimation of velocities, in the order of 0.5 % to 1 %. This is consistent with the fact that body waves velocities are in all cases upper bounds to the phase velocities of guided arrivals and leaky waves.

Dispersion study A study of dispersion in the time domain was done on the same set of microseismograms, for the first cycles of the pseudo-Rayleigh arrival. Figure 2-11 shows the initial sequence, from which the beginning of the pseudo-Rayleigh wavetrains corresponding to the offsets 5.00 m and 5.50 m were correlated with dynamic time warping. The two signals are normalized and interpolated before signal matching.

The result is shown on Figure 2-12. As expected, the general trend is a decrease of the phase velocity with time. The very beginning of the arrivals is weak in amplitude and contains a small component of P-wave arrival which explain the scatter in the results. For the first 500 microseconds, the average velocity is about 2300 m/s (the theoretical S-wave velocity being 2310 m/s). The last 300 microseconds correspond to a decreasing phase velocity, from 2300 m/s to about 2100 m/s. Thus, the underestimation of the S-wave velocity depends on the length of the window estimate. Nevertheless, taking the average over a 1000 microsecond window still provides a good estimate of the S-wave phase velocity (2250 m/s).

Field data

The sequence displayed on Figure 2-13 is twelve traces of field data. The first receiver is ten feet from the source and the distance between successive traces is a half foot. Each trace contains a P-wave, a pseudo-Rayleigh wavetrain and a Stoneley arrival. The relative amplitude of the pseudo-Rayleigh arrival is low.

A velocity analysis with signal matching was performed for each couple of traces for the P-, S- and Stoneley waves. The respective average values for the velocities are 4100 m/s, 2440 m/s and 1460 m/s. These values agree very well with results obtained with the semblance method and the maximum likelyhood method (Block, 1986). (Ellefsen, personal communication). The velocities between every successive slices of formation, however, show important variations. For the P-wave, velocities vary between 3400 m/s and 4500 m/s, for the S-wave between 2140 m/s and 2900 m/s. Accurate signal matching gave no significant trend for the variation of P- and pseudo-Rayleigh wave phase velocities. As shown on Figure 2-14, the pseudo-Rayleigh arrival seems to be fairly non dispersive.

2.6 Conclusions

The ANIS system proved to be well-suited for an interactive processing of sequences of waveforms. The primary advantage over other types of structure is that the core concept of *sequence* makes possible the easy manipulation of complex two-dimensional objects.

The system is adequate for a moderate amount of data, i.e. for sets of a few dozens of traces for the present implementation. The operators are flexible and accurate, as demonstrated by tests on synthetic data. These tests also confirm that the S-wave velocity is in general well-estimated from the characteristics of the pseudo-Rayleigh arrival. Nevertheless, the general trend is to underestimate velocities, by an amount that depends on the mechanical properties of the formation.

Working with graphical representations of signals provides an instantaneous understanding of the effects of operators. Thus the user has the ability to redo operations easily, until the processing results are satisfactory. This type of approach is very useful for development and for testing tasks.

The *message-oriented* style of programming allows modularity. Operators can be easily encapsulated in more complex and general structures. This later characteristics is essential for further development and provides a wide range of applicability. Basic operators form a very top-level language that can serve as a basis to construct more specific tools. The ability to treat the operators as abstract structures is also essential for the integration in a knowledge-based system for full waveform interpretation.

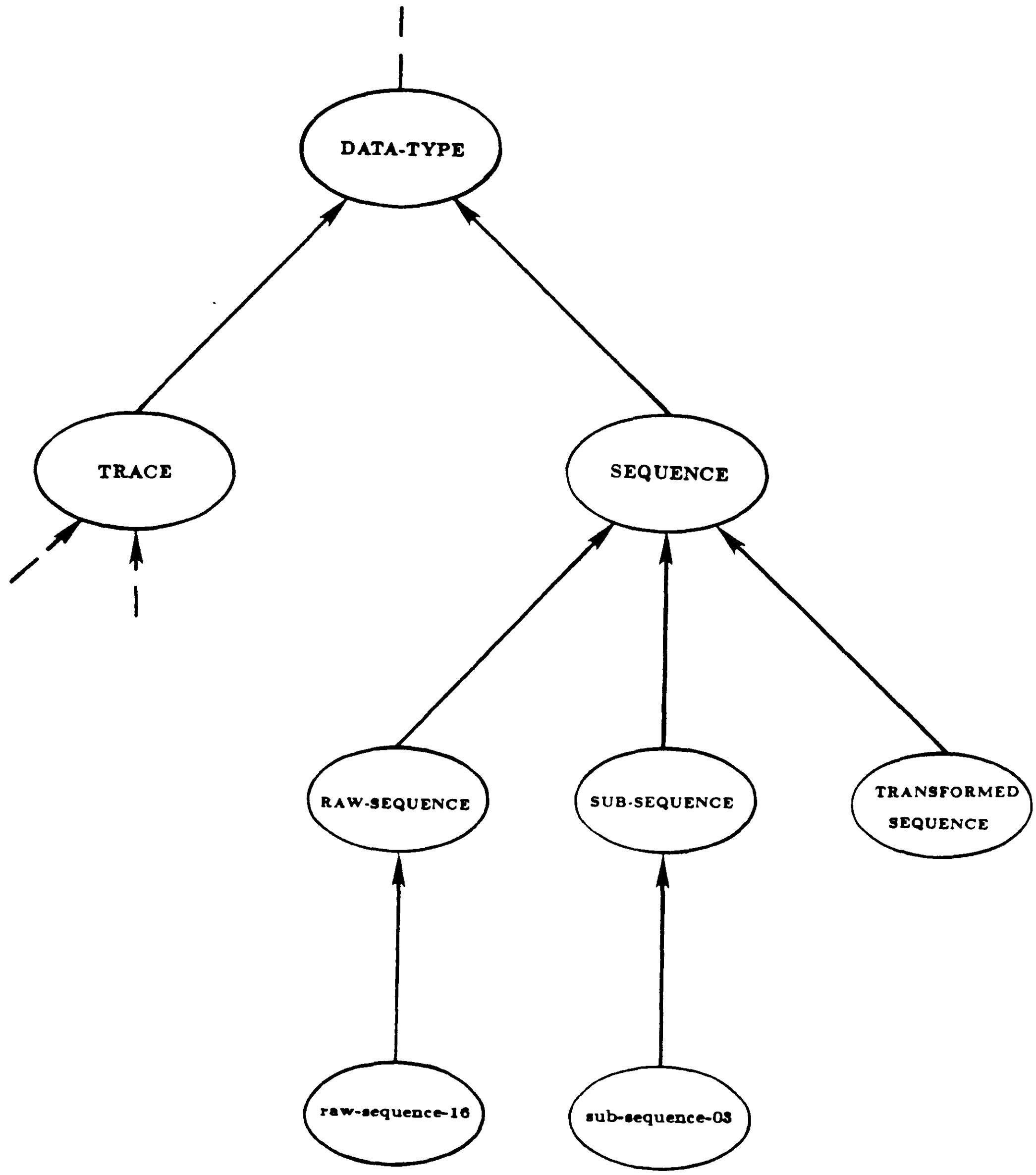


Figure 2-1: Hierarchical relations between ANIS objects: raw-sequence-16 and sub-sequence-03 are instances of the abstract data types RAW-SEQUENCE and SUB-SEQUENCE.

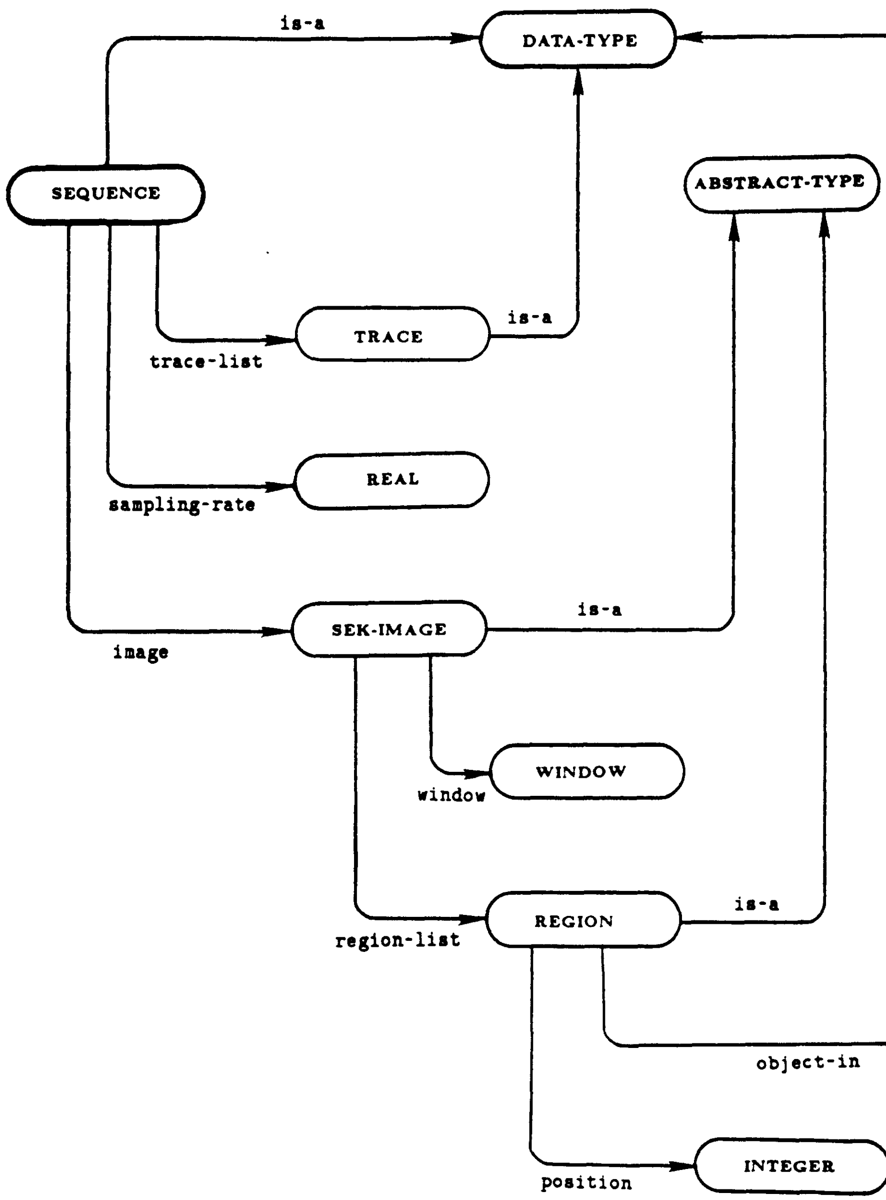


Figure 2-2: A partial representation of the network of relations between ANIS objects and attributes. IS-A links represent subset-set relations between concepts.

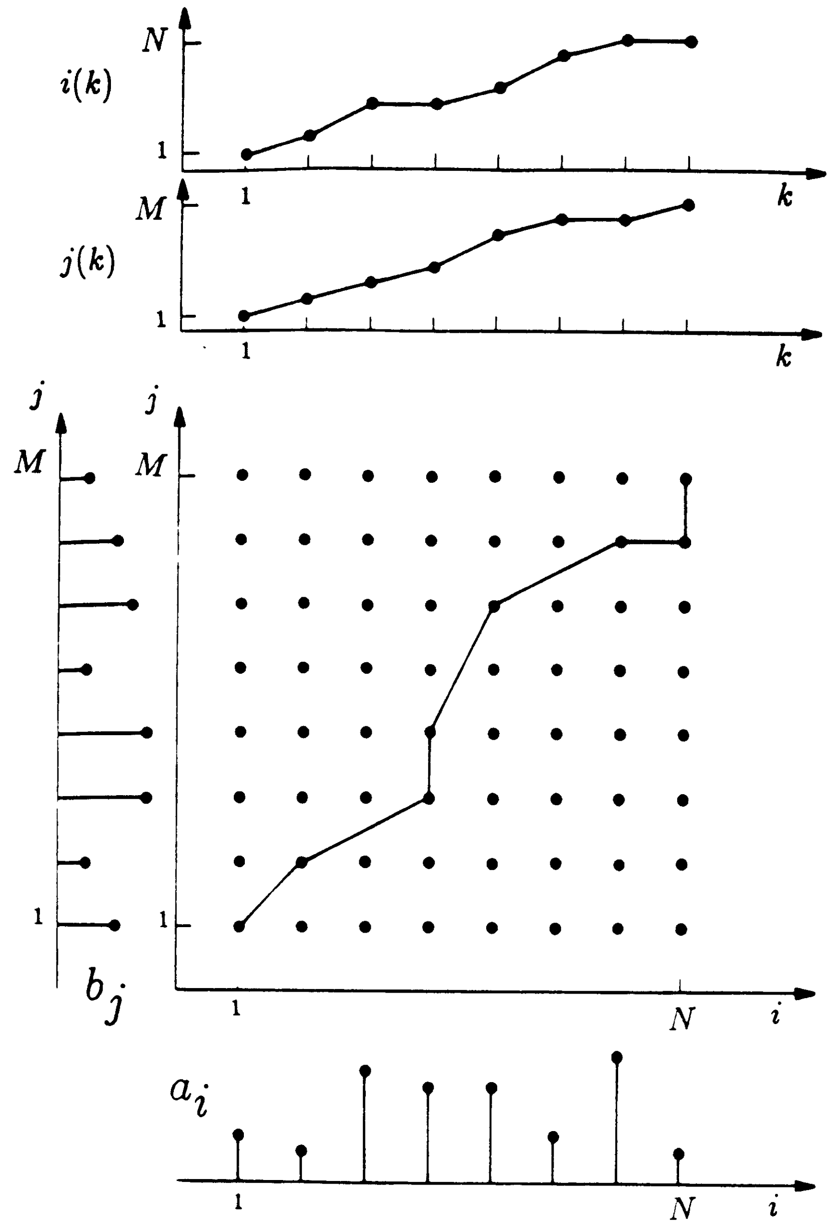
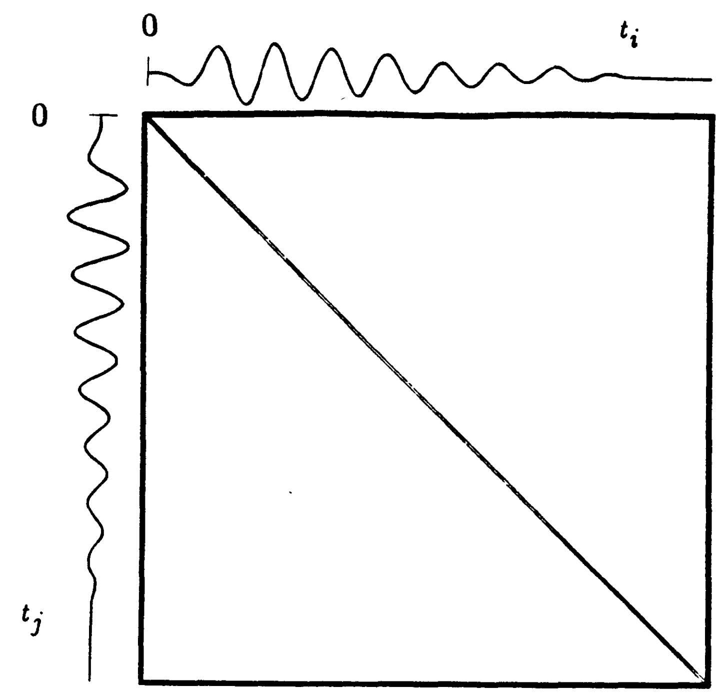


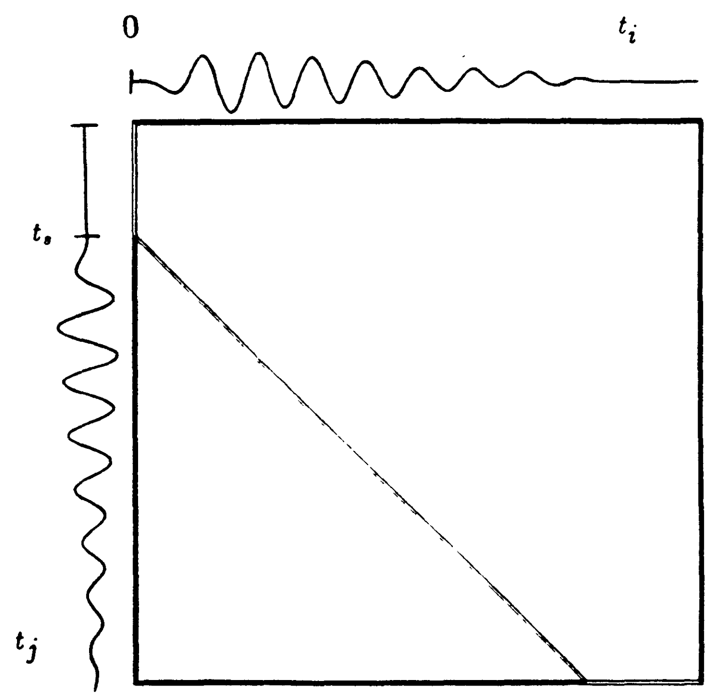
Figure 2-3: Matching of two discrete signals with dynamic time warping.

The mapping function is $c_k = [i(k), j(k)]$.

After Myers, 1980.



(a)



(b)

Figure 2-4: Mapping functions corresponding to:

a- Two identical signals.

b- Two identical signals with a time shift t_s.

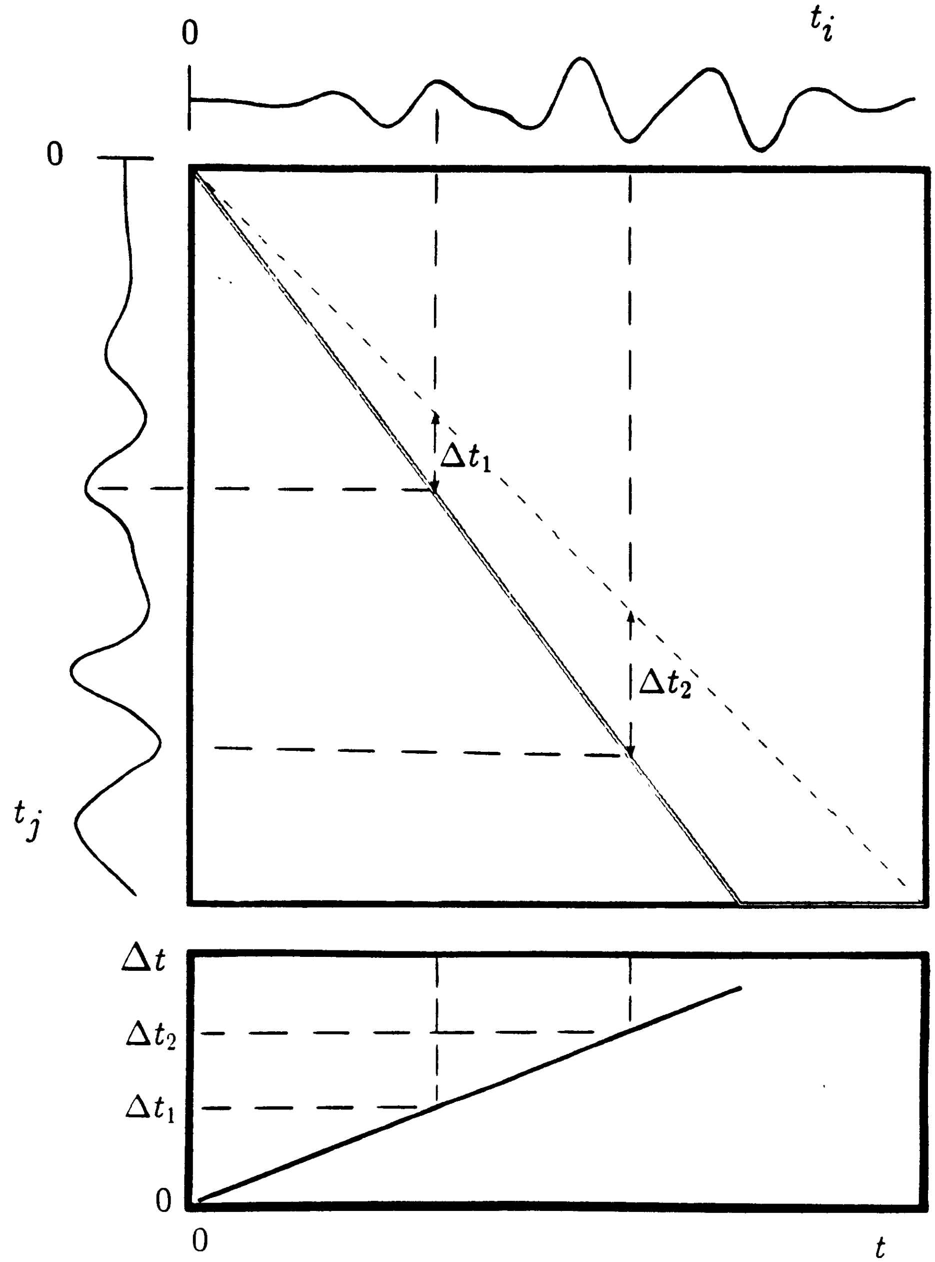


Figure 2-5: The mapping function between two signals (nearly identical but one — t_j — is stretched relative to the other — t_i) and the corresponding time shifts $\Delta t(t)$.

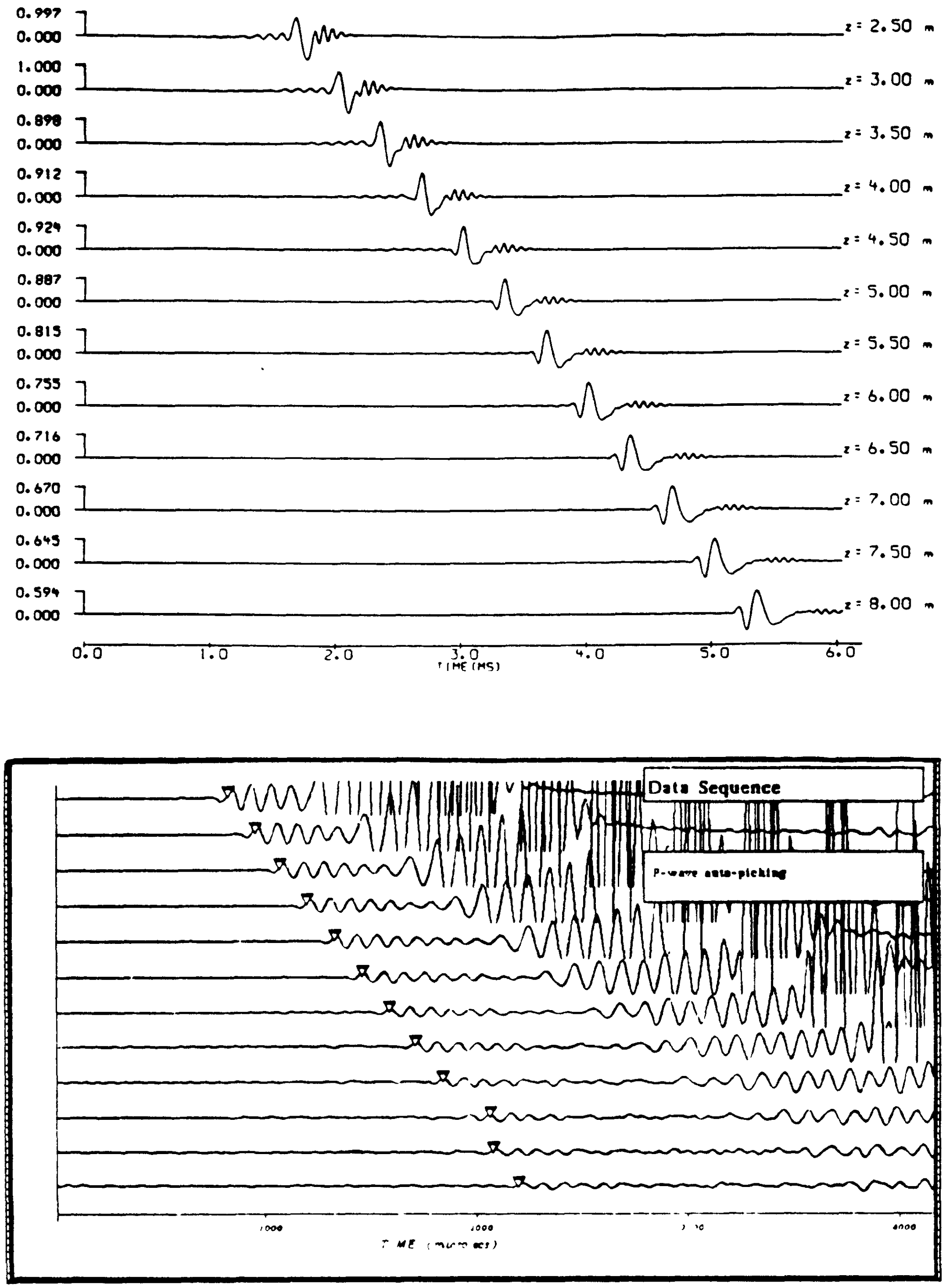


Figure 2-6: Synthetic microseismograms and automatic threshold detection of the P-wave arrival. The amplitudes are magnified in the lower diagram to show the P-waves. Note the cycle-skipping at 7.00 m.

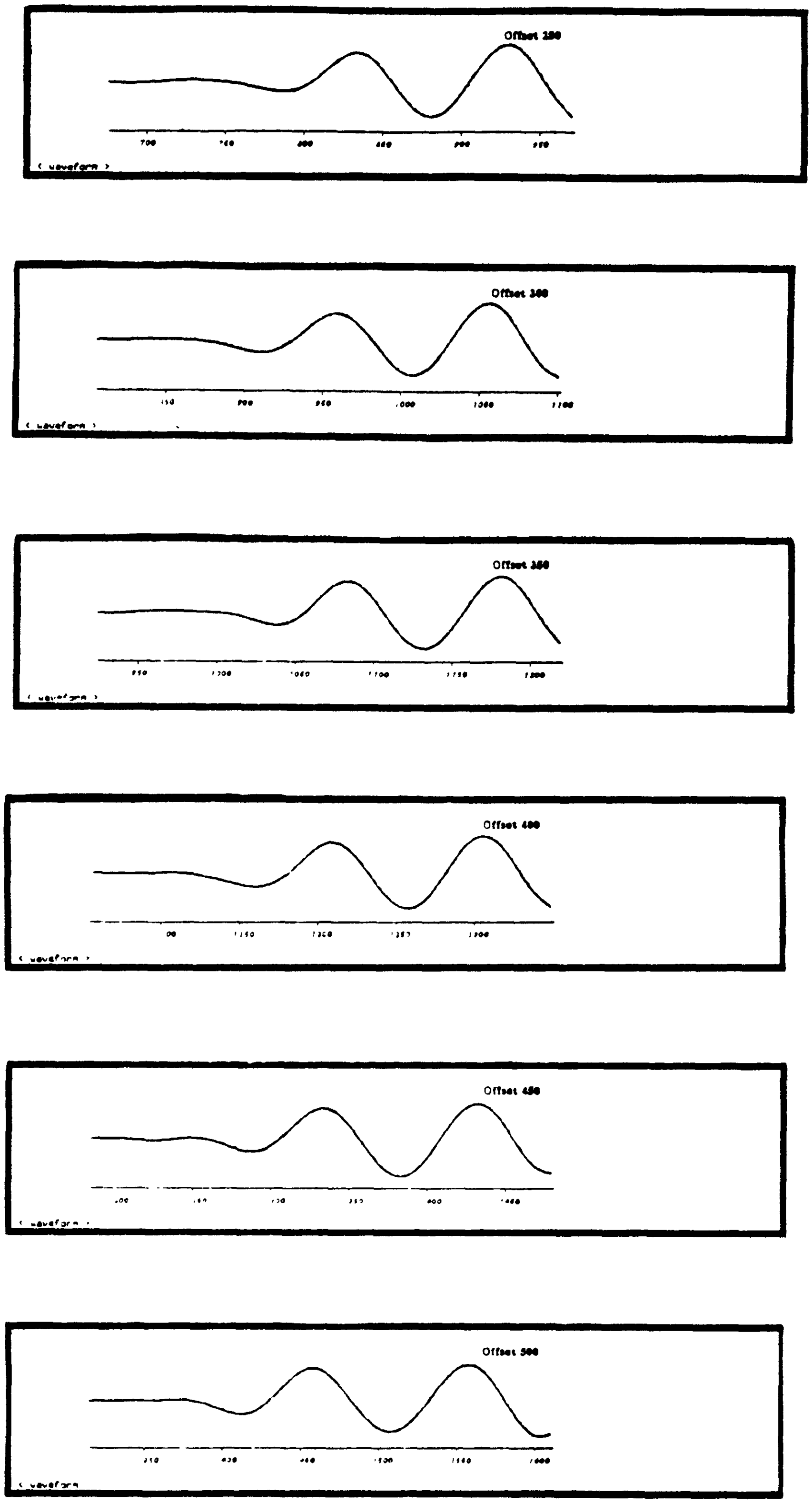


Figure 2-7: Windowed P-waves for signal matching. Offsets from 2.50 m to 5.00 m.

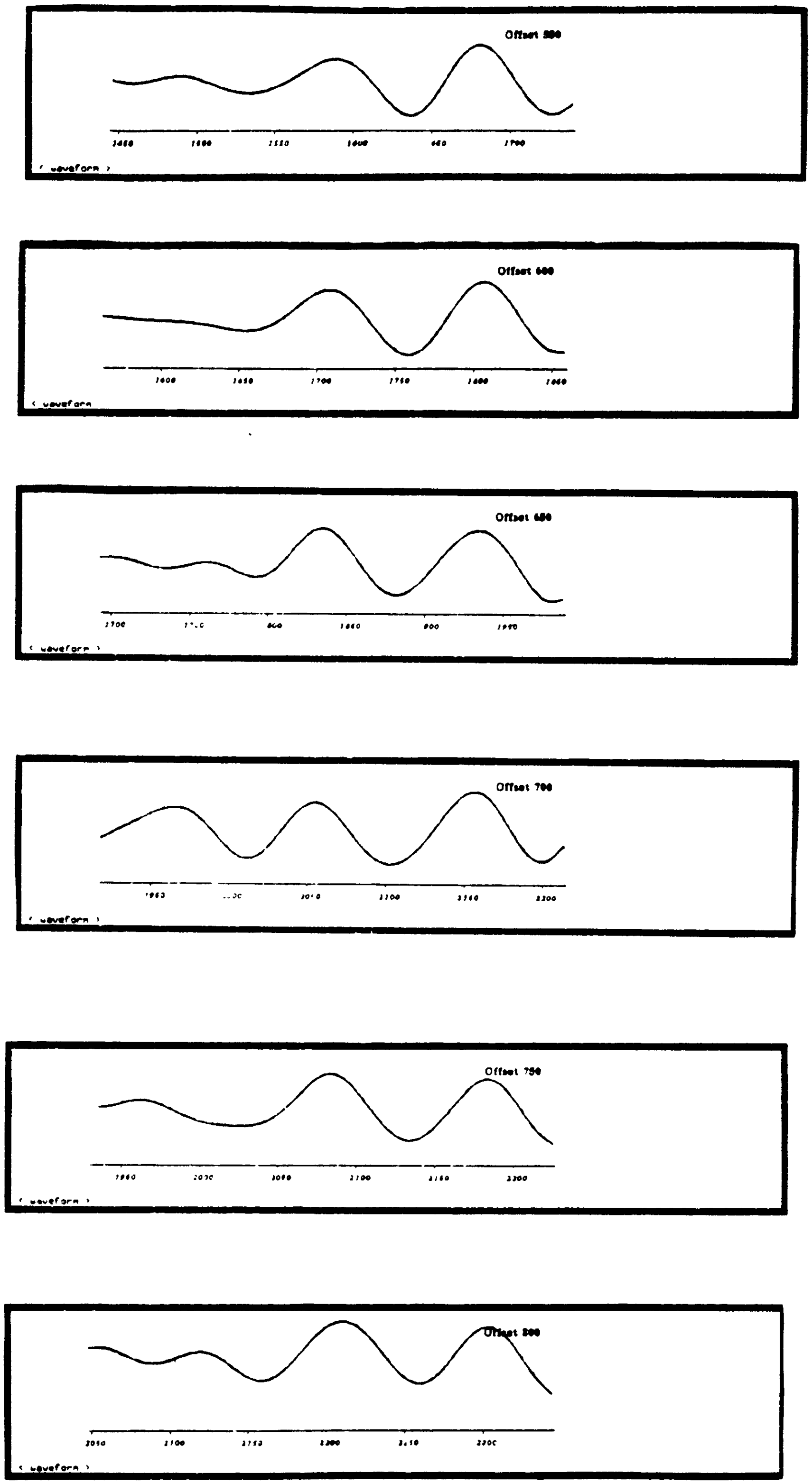


Figure 2-8: Windowed P-waves for signal matching. Offsets from 5.50 m to 8.00 m.

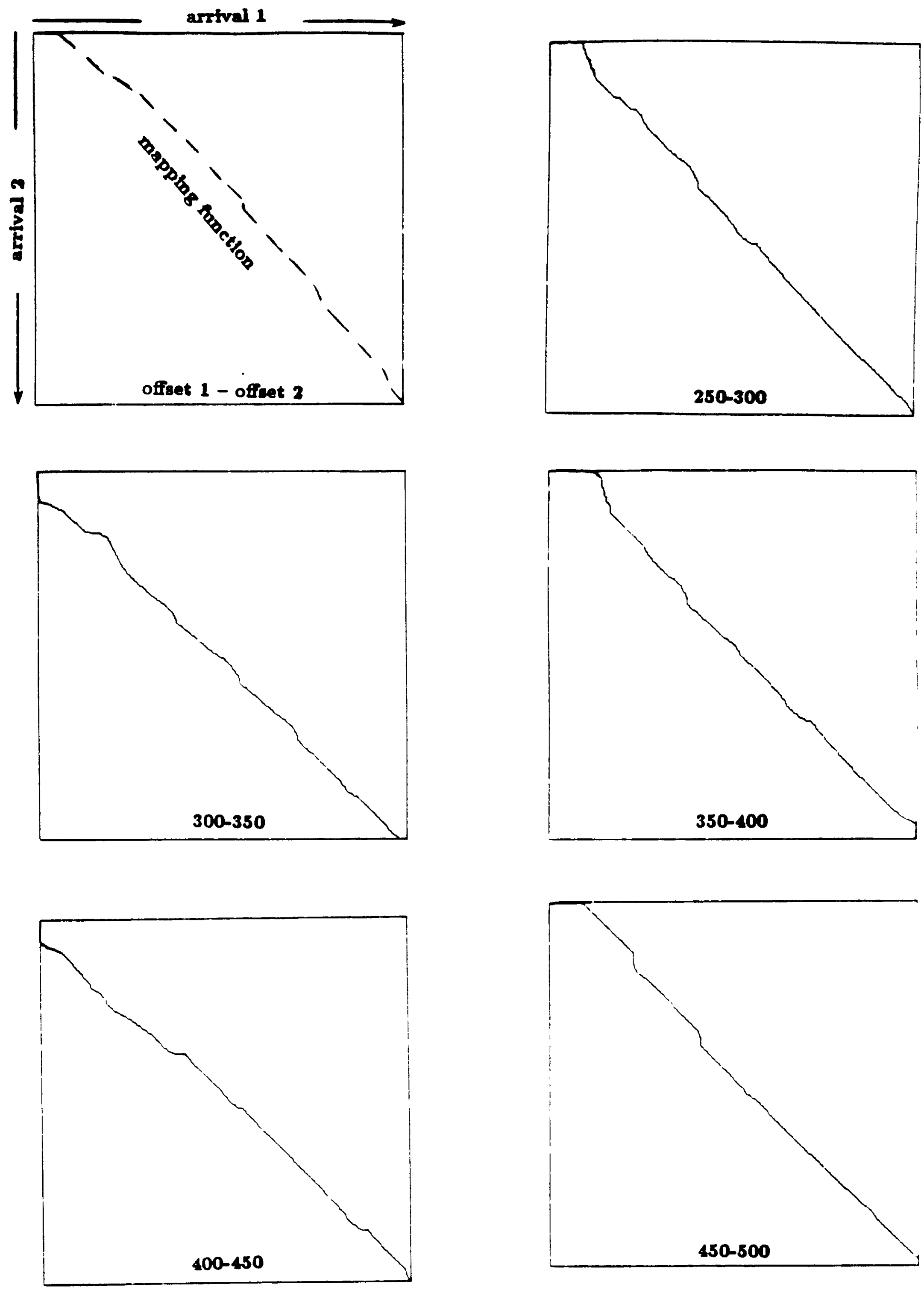


Figure 2-9: The mapping functions for adjacent waveform pairs. Offsets between 2.50 m and 5.00 m.

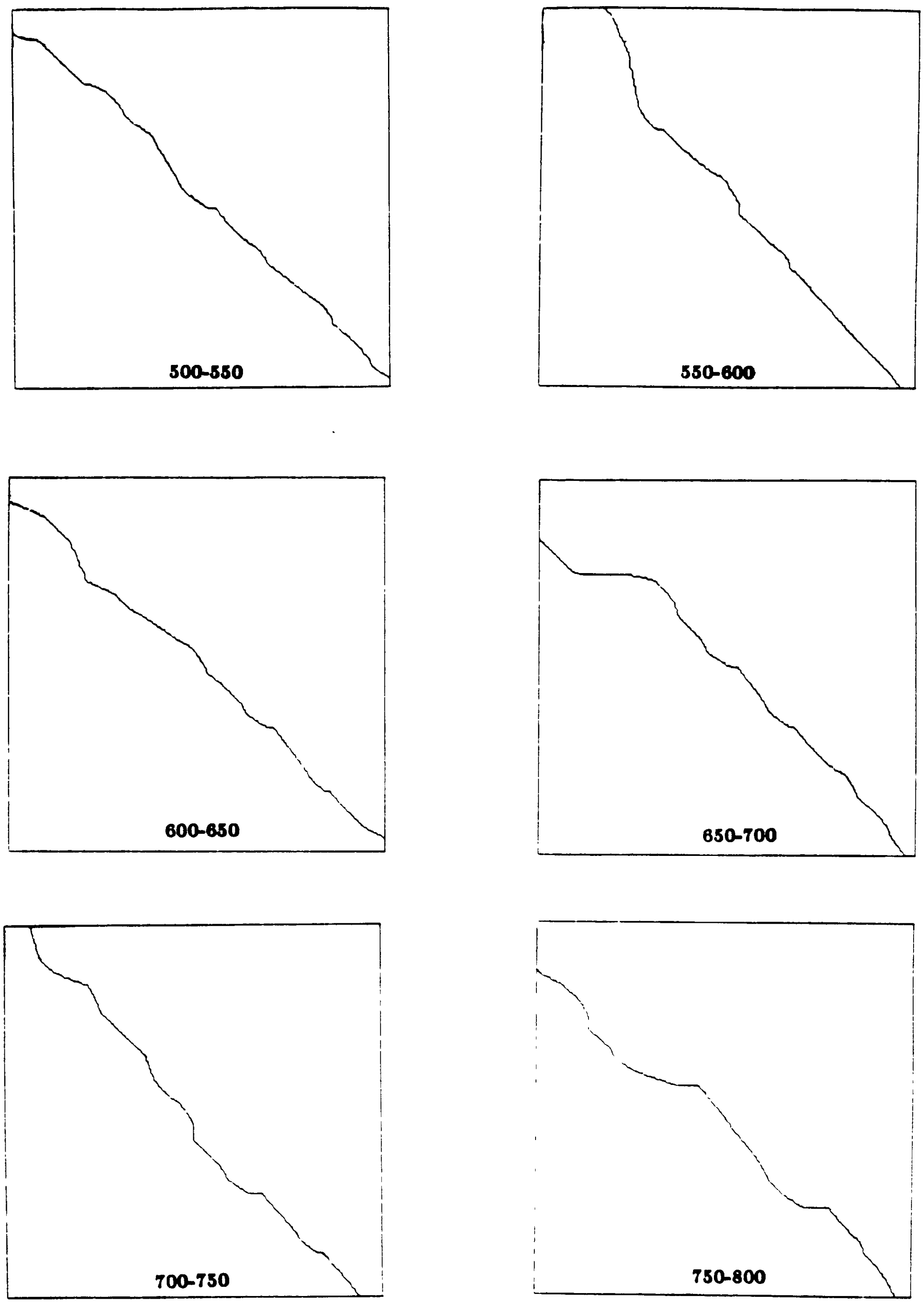


Figure 2-10: The mapping functions for adjacent waveform pairs. Offsets between 5.00 m and 8.00 m. As the signal-to-noise ratio decreases the mapping function becomes more noisy Note the effect of the maximum time shift constraint on the beginning of the mapping function at 650-700.

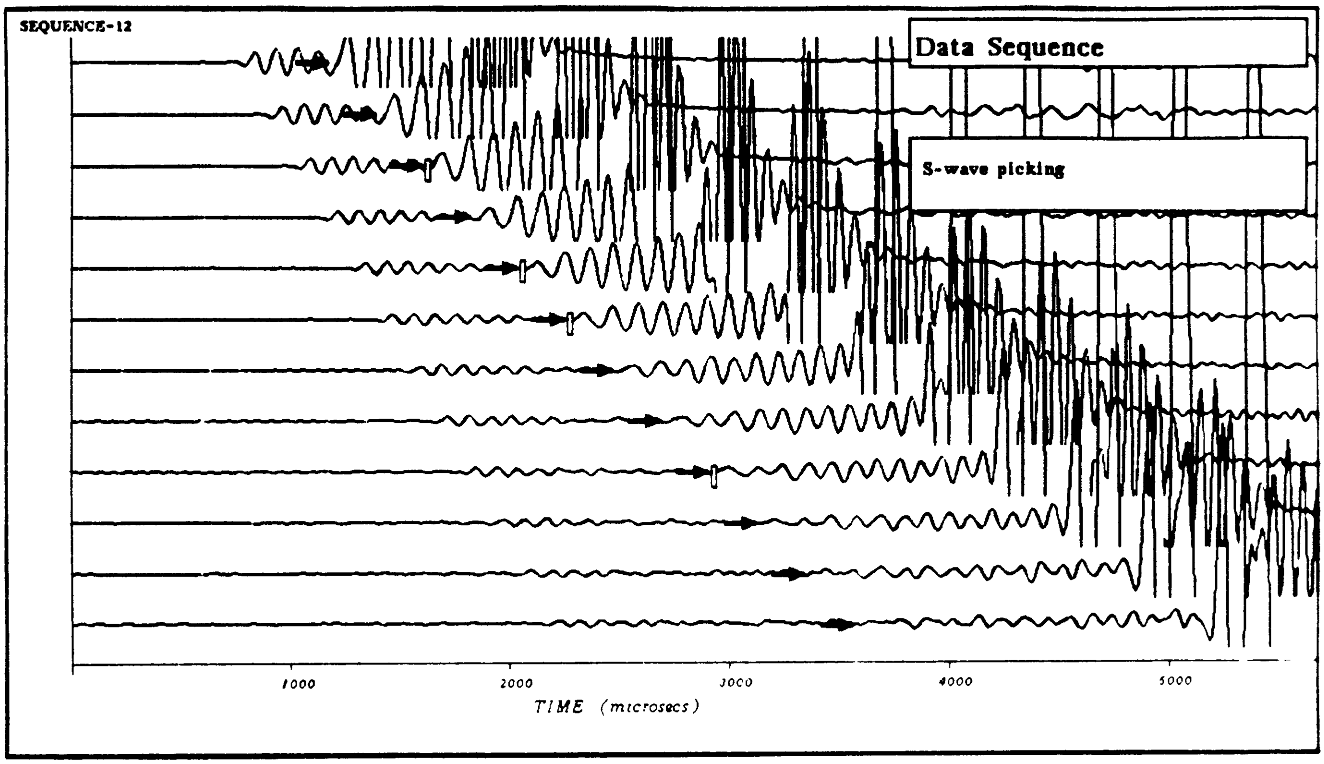


Figure 2-11: Picking the S-wave onset with the mouse. Small rectangles correspond to the arrival times picked by the user. Arrows point at interpolated time values for all traces.

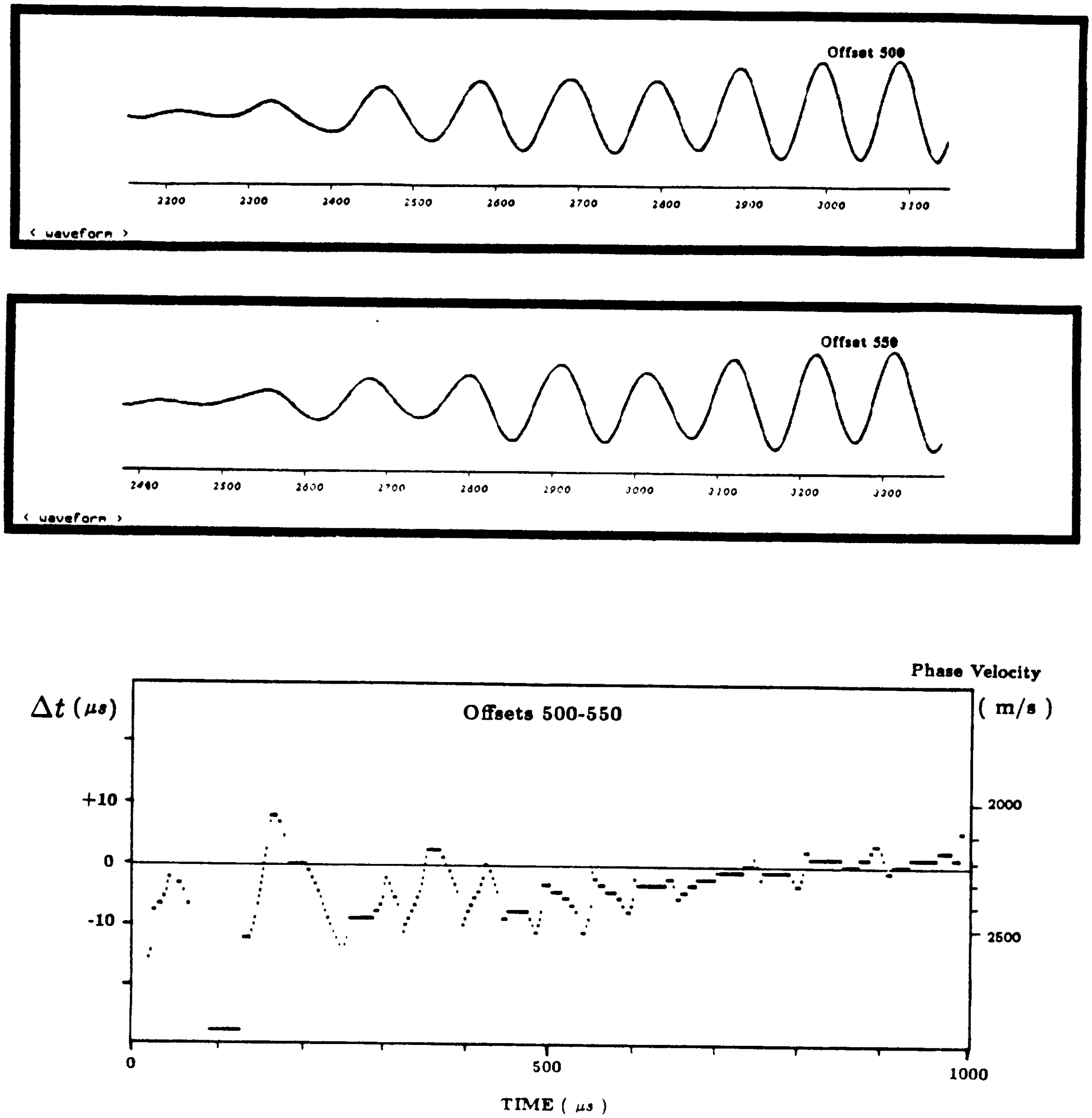


Figure 2-12: Two pseudo-Rayleigh arrivals for the offsets 5.00 m and 5.50 m and the corresponding variation of time delay with time.

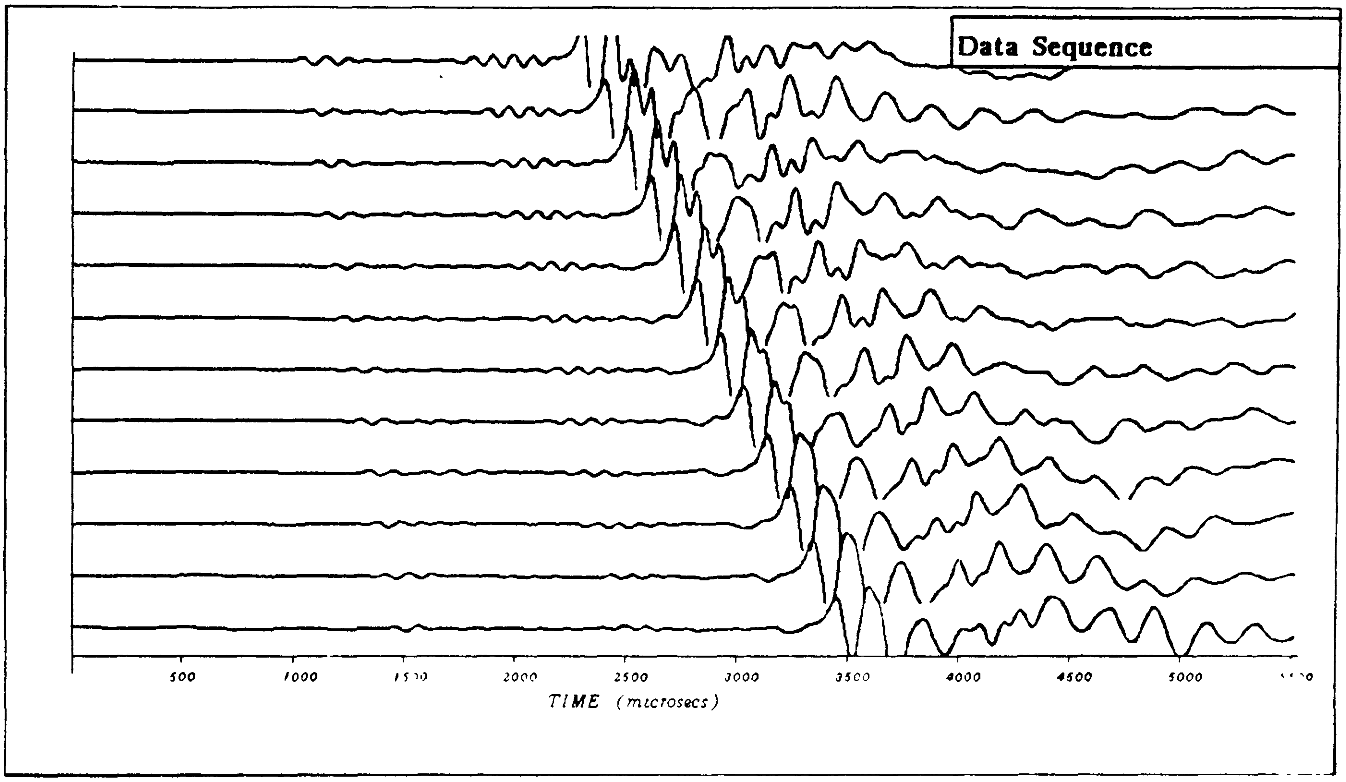


Figure 2-13: Raw array of full waveforms. The first receiver is 10 feet from the source and the distance between two successive traces is 0.5 foot.

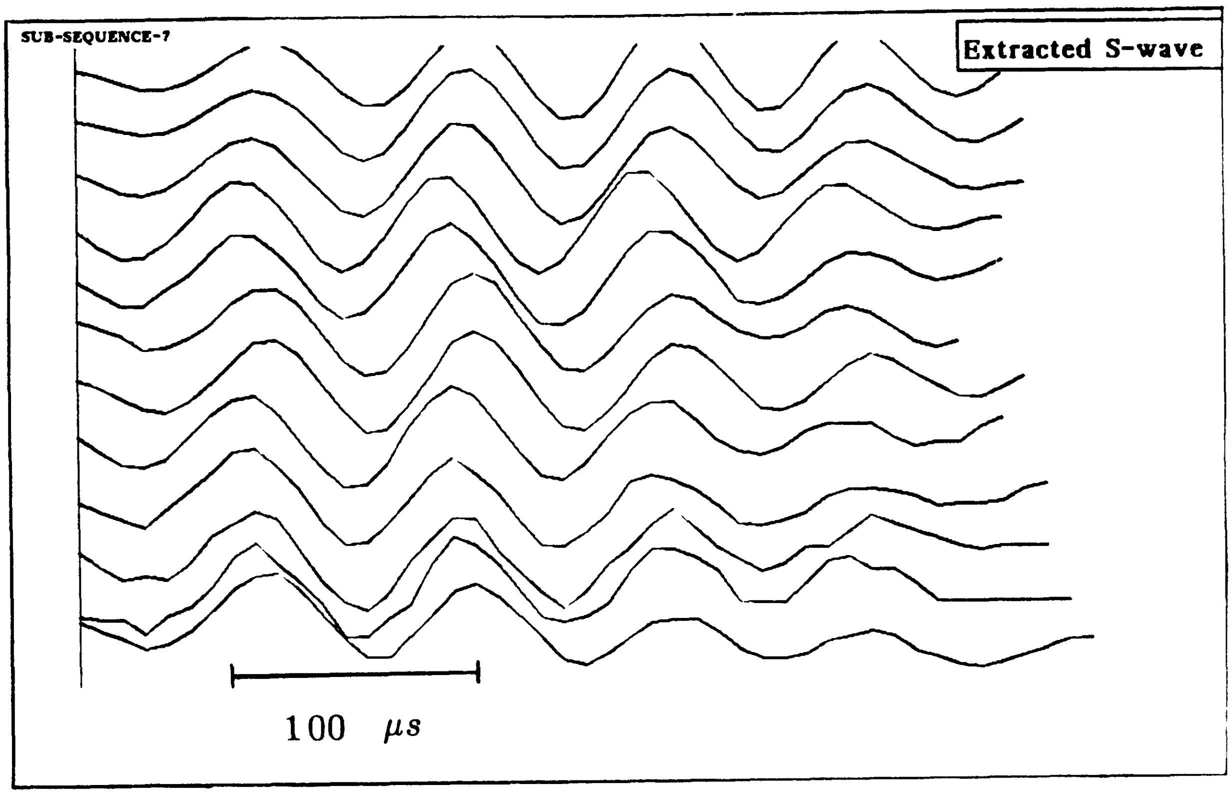


Figure 2-14: Extracted S (and pseudo-Rayleigh ?) wavetrains from the sequence presented on Figure 2-13.

Chapter 3

Qualitative Reasoning in Rock Physics

3.1 Introduction

The aim of this chapter is to investigate the application of knowledge-based techniques in the rock physics domain. Qualitative reasoning about rock properties provides a foundation for symbolic reasoning about the physics of wave propagation in boreholes. The problem addressed is the following: given a set of qualitative information about the properties of a rather homogeneous zone of a formation, infer other qualitative values that are of interest for the wave propagation. For instance, an interesting qualitative information is the *porosity* of the formation. If we are in a non-porous formation, the right model for wave propagation is the elastic model. Since waveforms are highly sensitive to changes in lithological properties and in borehole parameters, even qualitative information in these domains is of importance to adapt proper processing operations. Some petrophysical properties are generally known, at least qualitatively, and can be used during the interpretation process. Other basic information needed for a qualitative modeling of the wavetrain include the range of wave velocities and the presence of fractures. Knowledge of the borehole environment is also required to choose the adequate model (diameter, casing, presence of mud-cake, etc.).

A reasoning module performing qualitative reasoning in the rocks physics domain was implemented. In the next section, basic principles of knowledge-based systems are briefly reviewed.

3.2 Principles of knowledge-based systems

Knowledge-based techniques have already been developed for interpretative tasks in the geologic and geophysical domain. Well-logging interpretation is one of the most frequently addressed problems. In particular, the *Dipmeter Advisor* system (Davis et al., 1981) was the precursor in this field and demonstrated the basic possibilities of knowledge-based techniques as well as the principal pitfalls and limitations. Appendix D reviews the diverse applications in the geophysical domain. Before describing the prototype developed for qualitative reasoning in rock physics, some general principles of knowledge-based programming techniques must be outlined.

The general hypothesis of knowledge-based programs is that it is possible to represent adequately with some language the knowledge about a domain, without having to specify in advance the actual way in which this knowledge is to be used to perform a given task.

The first important type of representation is based on associational networks that attempts to describe the relations between the concepts involved in the description of the domain. When these networks are structured and include a mechanism for an automatic inheritance of properties based on the hierarchical relations between objects, they are called *frame*-based representations. As introduced originally by Minsky (1975), the rationale for frames was based on the consideration that a good deal of reasoning uses the recognition of some prototype situations rather than performing inferences based on very primitive evidences. Inheritance mechanisms reduce the need for theorem proving and allow to recognize the truth of a given statement with a limitated amount of search.

The second important representation tool captures the relevant pieces of knowledge in simple autonomous structures called *rules*. These rules, in their simplest form, consists in an antecedent clause (or a conjunction of clauses) and a consequent clause. More knowledge about the domain is then obtained when all facts in the antecedent part of a rule are true, which causes the consequent part to be true. The actual mechanism of inference can be controlled, basically either by initiating the reasoning process by a query (*backward-chaining* strategy) or by letting any data in the knowledge base trigger the application of every relevant rule (*forward-chaining*). Rule-based (or *production* system) were the first successfully developed system for diagnosis tasks, using a backward-chaining (or *goal-driven*) control strategy.

One of the general principles of knowledge representation is that declarative knowledge, i.e. "static" knowledge about the domain, must be kept clearly separated from deduced or evolutive knowledge. In fact, languages tend to be developed separately for these two tasks.

3.3 Qualitative petrophysical reasoning

Since there is no exact and complete formulation of the inter-dependencies between properties, the interpretative process is often based on qualitative physical laws and relies on practical situations observed on the field and on laboratory experiments.

Two components were constructed for the representation of relevant petrophysical knowledge. First, the explicit knowledge about the domain is represented in a frame-like knowledge base. Second, implicit knowledge, i.e. knowledge that can be deduced from the facts embedded in the knowledge-base, is captured via a rule-based formalism.

The `Litho` knowledge-base

The declarative knowledge about rocks is described in a *frame*-like structure representing the materials useful for the full waveform interpretation problem. The hierarchical tree of this knowledge base (called `Litho`) is shown on Figure 3-1. The definition of slots alone does not constitute a complete description of the knowledge in the data base. A complete description of the *semantic* of the domain is more explicitly illustrated by the network of relations between the different concepts. Figure 3-2 shows a part of the declarative knowledge enbedded in the LITHO knowledge base.

An important characteristic of the geologic knowledge used in this problem is its hierarchical nature. The lithological taxonomy is adequately represented by a frame structure with a mechanism for an automated inheritance of properties. One can easily reason about clastic rocks or evaporites without explicitly saying all relevant information. The knowledge embedded in the associational network of frames enables to recognize that a sandstone is a type of clastic rock, and therefore to apply any rule about clastic rocks to sandstones.

The declarative knowledge embedded in the knowledge base is a straightforward representation of the classical taxonomy of rocks. This data base could be extended easily if necessary, but the nature of qualitative rules does not necessitate a more detailed structure. Most of relevant information is in fact embedded in the slots of rock individuals themselves. Once we know that a rock is a carbonate, its real petrophysical properties (fracturing, porosity) are more determinant than a more subtle geologic classification. The knowledge base covers the description of some useful materials in well logging. In particular, the presence of material such as mudcakes, casings, fluids allow a more complete description of the domain, if needed.

The other advantage of the frame-based structure is that the description of attributes is done at different levels of specialization. For example, a density slot will be attached to all materials, including fluids, whereas a porosity slot will exist only for rocks. Therefore, the natural structure of knowledge is clearly apparent in the implementation.

The geologic rules

The geologic rules were designed with the help of qualitative petrophysical principles that link properties of rocks. Detailed sample examples of reasoning processes are exposed in the next section. A complete listing of these rules is given in Appendix E. The translation of these principles into the rule formalism is in most cases simple and allows flexibility and incremental refinement. Nevertheless some remarks must be made:

- A typical reasoning tree is one or two levels deep. This is mostly due to the fact that the rule base is too sparse. Nothing is said, for instance, about the relations between the conductivity of rocks and the others properties. The reasoning process tends to invoke "dual" properties (fracturing and permeability for instance).

- The representation makes sometimes awkward to reason with *negations* or *exceptions*. We had to define artificial rules (see Appendix E, "rules x") to take into account assertions such as "this fact is true except for shales". The choice, is to define either a new frame concept such as a "non-shale" rock or to capture this fact with a rule.

- The last difficulty is to try to formalize a more *physical* knowledge of the problem. The rule-based formalism makes difficult to express, for instance, that if the porosity increases, the P-wave velocity will decrease (all other parameters remaining unchanged).

3.4 Examples of reasoning processes

The reasoning process of the system is described with the help of two simple examples. The strategy for these two cases study is *backward chaining* (or goal-driven) and leads to the depth first search of a tree. The depth of reasoning is shallow. Typical successful queries generally involve the application of one or two rules only. The basic object in the knowledge structure is a *zone*. A zone has an attribute *lithology* whose slot-filler is a given instance of any frame subsumed by the frame `ROCKS`. The declarative knowledge about rocks is embodied in the `LITHO` knowledge-base. The rules are listed in Appendix E.

Example 1

The complete search tree [1] for this example is displayed in Figures 3-3, 3-4. Given a zone whose lithology slot-filler is *halite*, deep in the borehole, what is the value of its attribute *permeability*?

[1] For illustration, the whole search tree is explored. Of course, for practical applications the search ends after the first successful match.

The first rule that has the datum ***permeability*** in its premises is **rule A-5**. The first assertion in the antecedent part of the rule is common to all petrophysical rules and doesn't interfere with the reasoning process. This clause takes into account the fact that we are dealing with the lithology related to a given instance of the frame ***zone***. Thus the slot lithology is bound to the given instance of rock (***rock 1*** in this example). The second clause is false since a halite *is* a sedimentary rock. Therefore, a second rule must be tried out.

Rule A-3 is not successful since one of its premises requires to try **rule A-9** which involves in turn permeability. Hence, this way is a dead-end.

The second clause of **rule A-2** (that is itself a conjunction of clauses) is satisfied thanks to the hierarchical structure of the knowledge — a halite being a kind of evaporite. The third clause cannot be matched in the actual database, thus two rules are invoked to try to bind the slot ***fracturation*** to some value.

- The second premise of **Rule C-5** cannot be matched in the actual database. Two rules are then invoked to find out about the porosity:
  - **Rule MC-1** can be successfully applied but this is worthless since the second clause of **rule-C-5** is about permeability — which is the initial query.
  - **Rule X-3** is irrelevant.
- **Rule C-4** allows to conclude since a halite is an evaporite. Taking into account the initial assertion "The *max.depth* of *rock 1* is *deep*", all the clauses of **rule A-2** are matched in the database. A solution is found.

Rule A-1 cannot be satisfied since its antecedent clause mentions null porosity and there is no rule in the rule-base to demonstrate that the porosity of a rock is nul. **Rule A-4** fails in finding any fact to infer that the *porosity* of the rock is high. Applying **rule A-7** leads to a loop and therefore the backtracking stops. **Rule A-6** does not help since the slotfiller ***max.depth*** of the rock is not ***shallow***.

The winning path that infers a null permeability which involves **rule A-2** and **rule C-4** is displayed again in Figure 3-5.

Example 2

The second example is about a zone where we know there exists a mud cake. This type of information can be obtained, for instance, by comparing two different caliper tool logs. The rock in the zone is a limestone with no permeability. Considering these assumptions, the *fracturation* of the rock must be *absent* since the mud cake indicates that the formation is porous and, since, for all rocks, the existence of fracture is not compatible with both near-null permeability and non-null porosity. Figures 3-6 and 3-7 show the backward chaining search tree. The rule base allows to find a solution, which involves **rule C-5** and **rule MC-1**. This example shows the decisive contribution of practical well-logging knowledge to obtain qualitative property values.

The first path involving **rule C-6** shows a weakness of our rule-based representation. Being told that the permeability of the rock is nul (the initial assertion), the system tries to match the fact "the permeability of the rock is high" because of the existence of **rule A-4**. **Rule A-4** states that if the permeability of a rock is not high, medium, or low, then it is null — which seems fairly straightforward since these qualitative values are theoretically the only allowed . The problem here is that the set of legal values has to be restricted either in the declarative part of the system (which implies multiple cross-checkings and hence loss in computational efficiency) or explicitly specified by rules. In our case, **rule-X4** enables us to write rules about "non-null" permeability. The drawback is that the theorem prover looses time in attempting to match the clauses of **rule-X4**, searching in branches that should be pruned.

3.5 Discussion

The actual limitations of the system come from the size of the rule base (about 25 rules). The conductivity, the saturant fluid characteristics and other attributes of the rock should be taken into account. This would relate more efficiently acoustic parameters to general petrophysical properties in the reasoning process. The borehole environment characteristics are another useful source of information to consider.

The rules about the mud cake existence show the importance of the well logging rules of thumb. Simple evidence about the borehole environment can provide decisive information about the qualitative properties of the formations. Such additional knowledge can be determinant in the estimate of properties and for reasoning about waves' characteristics. In fact, this type of knowledge, together with statistical heuristic knowledge about the effects of a given set of geological conditions on the recorded waveforms, could be as well encoded with the same formalism. This was, however, out of the scope of the present study.

Another direction of improvement would be to use a planning of reasoning in order to adapt the choice of rules to the general geologic context. For instance, one could use different sets of rules for sedimentary basins and igneous regions. This strategy will also help in pruning the search tree.

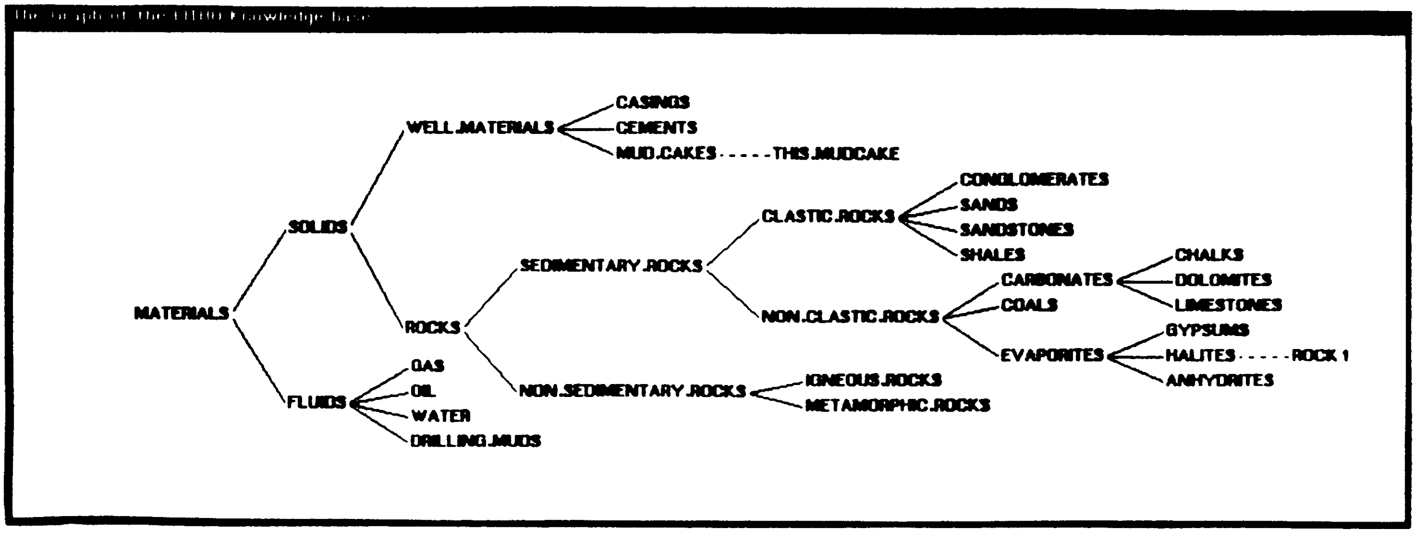


Figure 3-1: The hierarchical frame-based representation of materials for the full waveform interpretation problem. *This.mudcake* and *Rock.1* are specific instances of the frames *mudcakes* and *halites*.

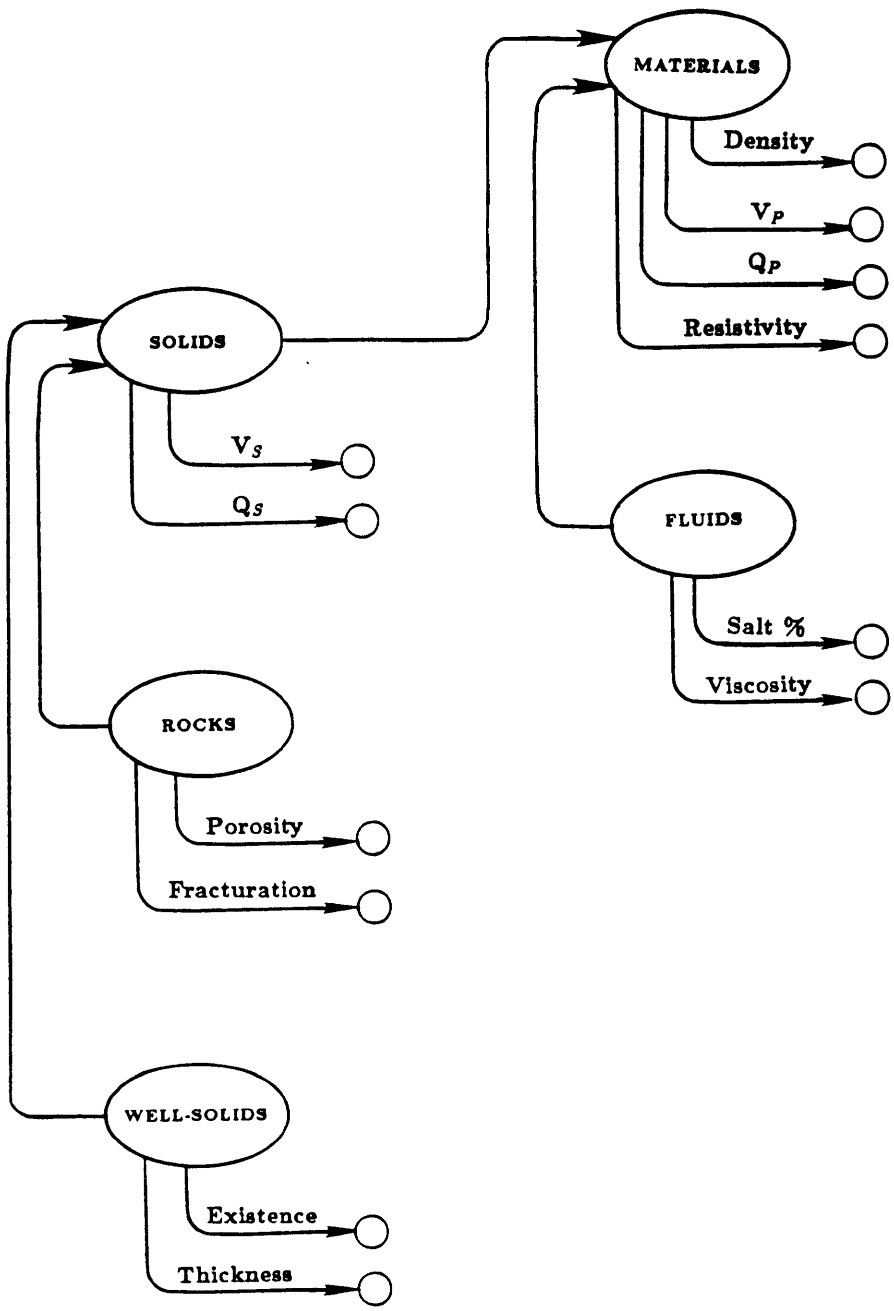


Figure 3-2: A piece of the network of concepts in the Litho knowledge-base.

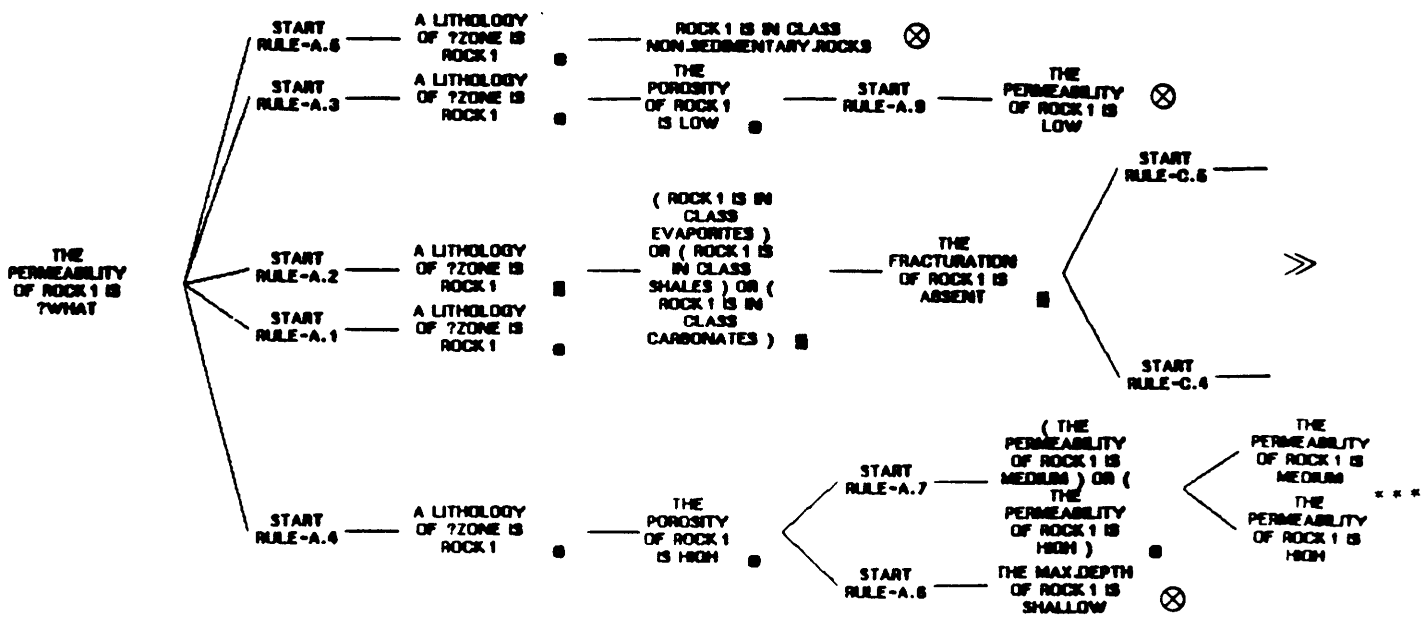


Figure 3-3: Example 1. Backward chaining tree, part 1. The initial assertions are:

The *lithology* of *this.zone* is *rock 1*

The *mudcake* of *this.zone* is *this.mudcake*

The *existence* of *this.mudcake* is *true*

Rock 1 is in class *halites*

The *max.depth* of *rock 1* is *deep*.

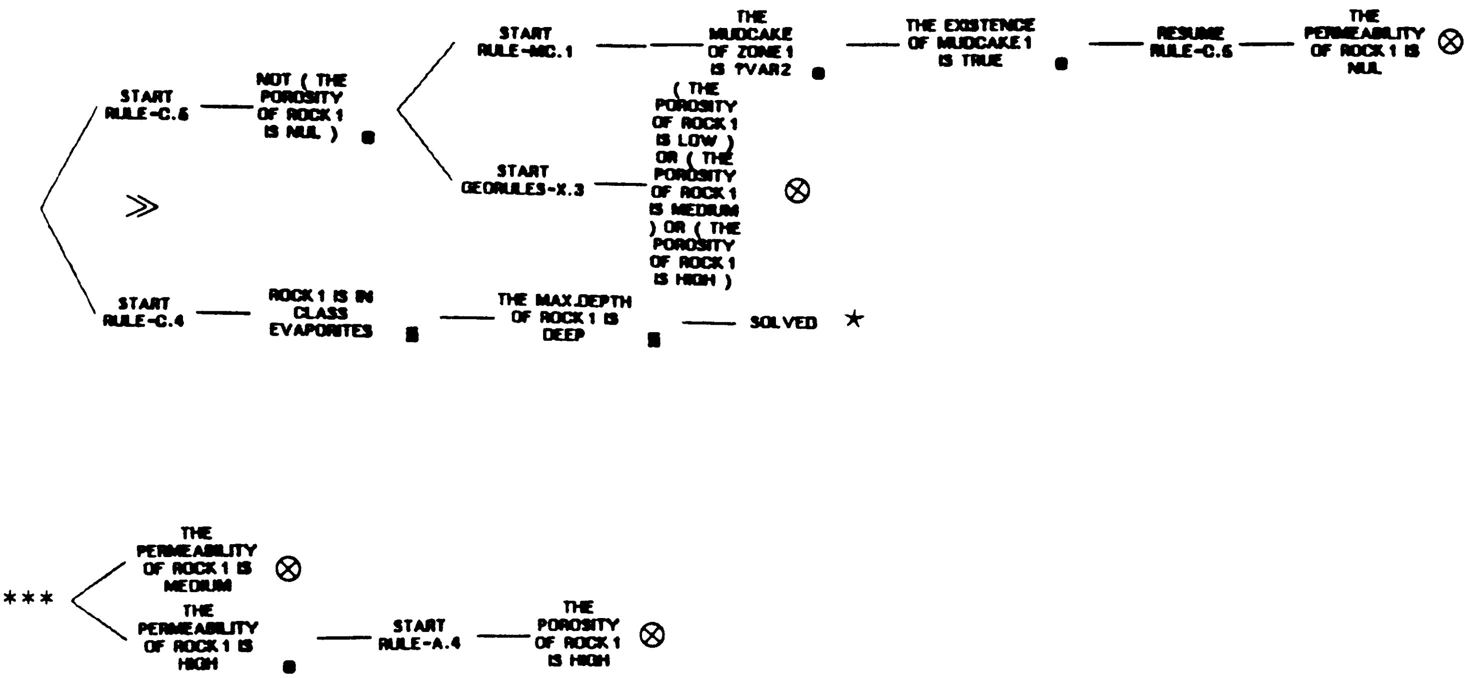


Figure 3-4: Example 1. Backward chaining search tree, part 2.

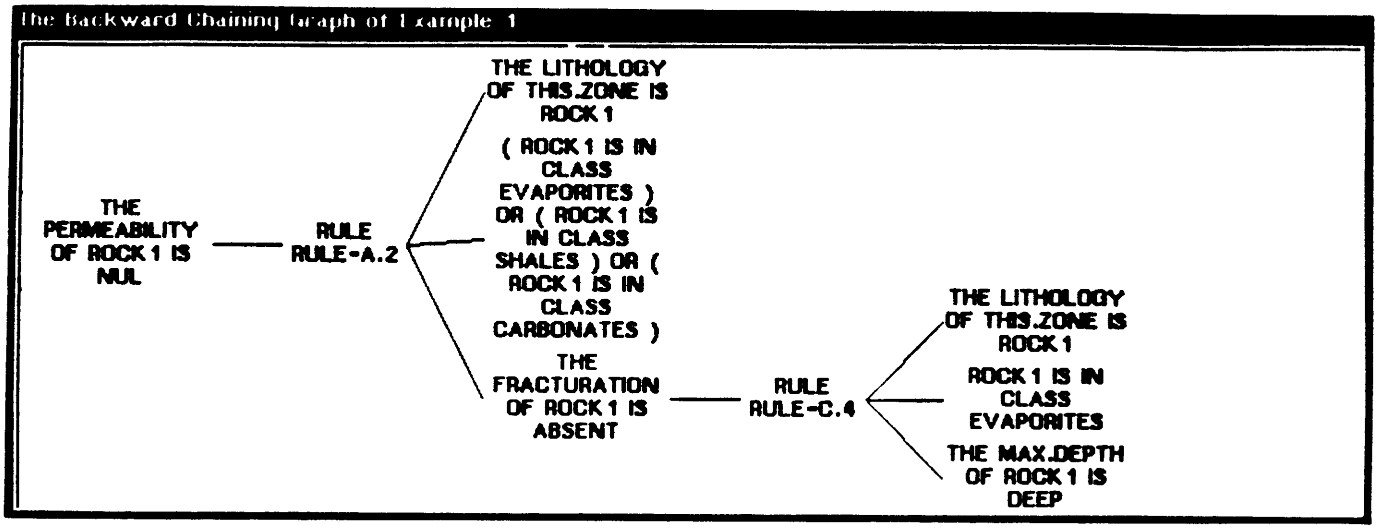


Figure 3-5: Example 1. The final backward chaining path.

The slot *permeability* was bound to *nul*, after applying **rule C-4** and **rule A-2**.

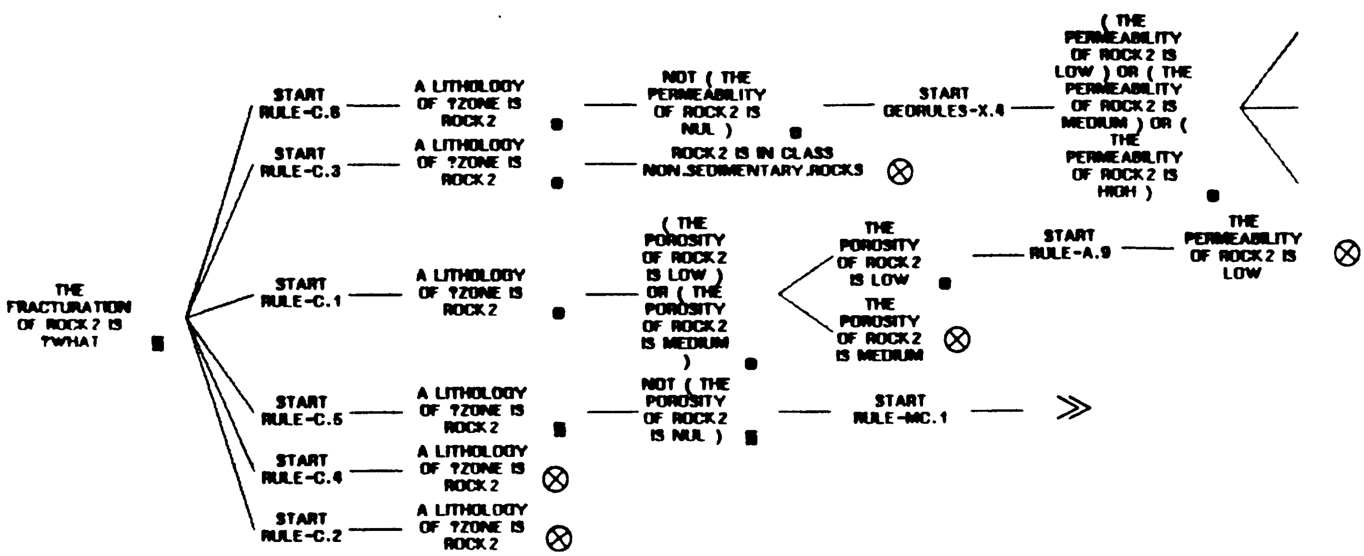


Figure 3-6: Example 2. Backward chaining tree, part1. The initial assertions are :

The *lithology* of *this.zone* is *rock 2*

The *mudcake* of *this.zone* is *this.mudcake*

The *existence* of *this.mudcake* is *true*

Rock 2 is in class *limestones*

The *permeability* of *rock 2* is nul.

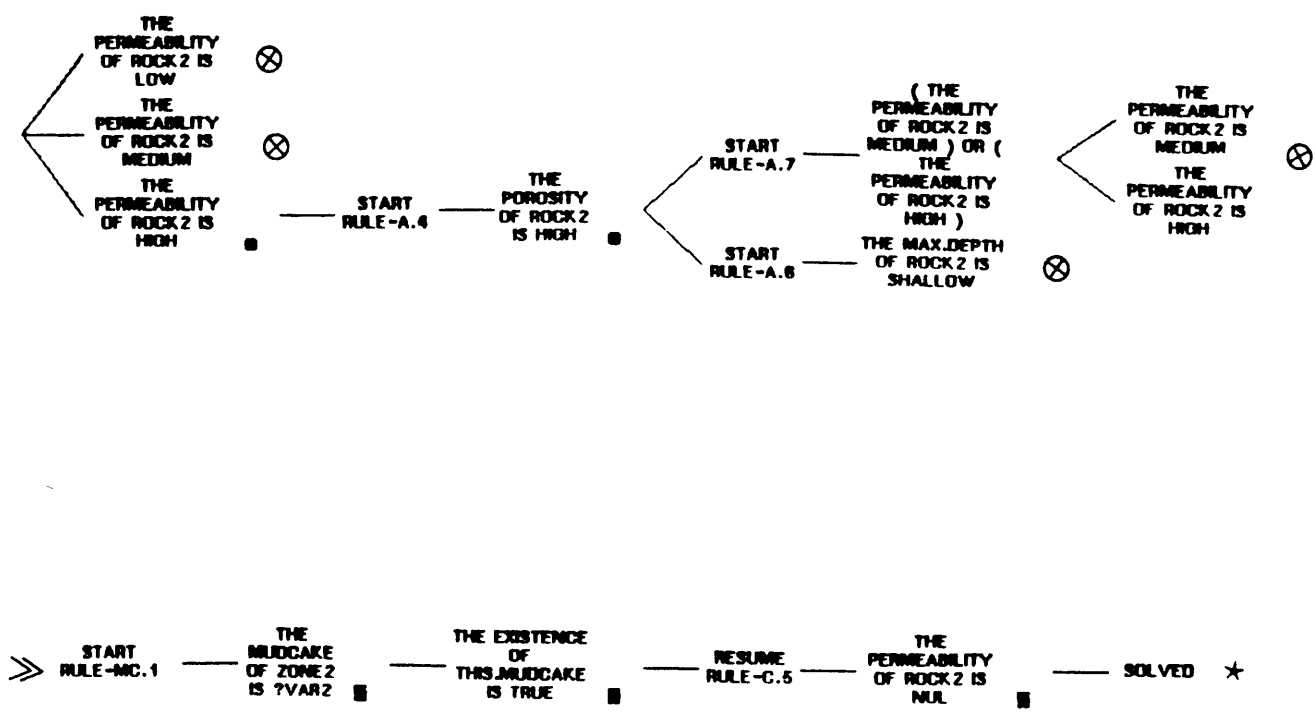


Figure 3-7: Example 2. Backward chaining search tree, part 2.

Chapter 4

Conclusions

In this work, a novel approach to the processing and interpretation of geophysical data was investigated. The ultimate goal, which is to incorporate geologic knowledge in the processing of full waveform acoustic data was tackled with the help of two primordial subtasks. First, for the processing part, operators suitable for symbolic manipulation were implemented. Second, in order to test the interest of geologic symbolic reasoning, a prototype knowledge-based system performing qualitative reasoning in the rock physics domain was designed.

In order to favor intelligent processing, a workstation-based system was developed using object-oriented programming. This style of implementation was found to be convenient, and furthermore necessary to provide the high flexibility necessary for manipulating the operators. The principal advantage of the processing package is to be built around the concept of sequence of waveforms. This favors the development of functions for complex two-dimensional signals.

Synthetic data were used to test the reliability and accuracy of the processing operators. In particular, a novel method for acoustic waveforms matching, using dynamic time warping was shown to be suitable for very accurate correlations. A set of field data was also processed with the system.

The prototype knowledge-based system demonstrated that the geologic knowledge can be adequately represented with a frame-like language. Moreover, additional qualitative evidence is gained by using rule-based techniques. The conjunction of these techniques enables one to maximize the amount of information (explicit and implicit) about the geologic environment.

Nevertheless, one limitation of the actual system is that the knowledge embedded in the rule-base is often too general in nature to provide decisive information. More effectiveness can be gained by capturing heuristic observations about the domain. In well-logging simple rules of the thumb are powerful ways to obtain qualitative results about rock properties. This type of development, however, was beyond the scope of the present study.

References

Anderson, K.R. and Gaby, J.E., 1983, Dynamic Waveform Matching: Information Sciences, 31, 221-242.

Arditty, P.C., Arens, G., Staron Ph., 1981, EVA: A Long-spacing Sonic Tool for Evaluation of Acoustic Velocities and Attenuations: Presented at the 51st Annual International SEG Meeting in Los Angeles.

Biot, M.A., 1952, Propagation of Elastic Waves in a Cylindrical Bore Containing a Fluid: J. Appl. Phys., 23, 997-1005.

Biot, M.A., 1956a, Theory of Propagation of Elastic Waves in a Fluid Saturated Porous Solid. I- Low Frequency Range : J. Acoust. Soc. Am., 28, 168-178.

Biot, M.A., 1956b, Theory of Propagation of Elastic Waves in a Fluid Saturated Porous Solid. II- Higher Frequency Range : J. Acoust. Soc. Am., 28, 179-191.

Block, L.V., Cheng, C.H. and Duckworth, G.L., 1986, Velocity Analysis of Multireceiver Full Waveform Acoustic Logging Data in Open and Cased Holes: Presented at the 56st Annual International SEG Meeting in Houston.

Bonnet, A. and Dahan, C., 1983, Oil-Well Data Interpretation Using Expert System and Pattern Recognition Technique: IJCAI 8, 185-189.

Burns, D.R., 1986, Formation Property Estimation from Guided Waves in a Borehole: Ph.D.

Thesis, MIT.

Castagna, J.P., Batzle, M.L., Eastwood, R.L., 1985, Relationships between Compressional-wave and Shear-wave Velocities in Clastic Silicate Rocks: Geophysics, 50, 571-581.

Cheng, C.H. and Toksöz, M.N., 1979, Inversion of Seismic Velocities for the Pore Aspect Ratio Spectrum of a Rock: J. Geophys. Res., 84, No. B13, 7533-7543.

Cheng, C.H. and Toksöz, M.N., 1981, Elastic Wave Propagation in a Fluid Filled Borehole and Synthetic Acoustic Logs: Geophysics, 46, 1042-1053.

Davis, R., Austin, H., Carlbom, I., Frawley, B., Pruchnik, P., Sneiderman, R. and Gilreath, A., 1981, The Dipmeter Advisor: Interpretation of Geologic Signals: IJCAI 7, 846-849.

Dove, W.P., Myers C., Milios, E.E., 1984, An Object-Oriented Signal Processing Environment: The Knowledge-Based Signal Processing Package : M.I.T., R.L.E., Technical Report No. 502.

Duda, R.O., Hart, P.E., Nilsson, N.J., Sutherland G.L., 1978, Semantic Network Representations in Rule-Based Inference Systems: in *Pattern Directed Inference System*, ed. D.A. Waterman and F. Hayes-Roth, Academic Press, New York.

Duda, R.O., Hart P.E., Konolige K., Reboh R., 1979, A Computer-Based Consultant for Mineral Exploration: Final Report, Artificial Intelligence Center, SRI International.

Duda, R.O., Gaschnig, J.G., Hart P.E., 1980, Model Design in the Prospector Consultant System for Mineral Exploration: pp. 153-167 in *Expert System in the Microelectronic Age*, ed. D. Michie, Edinburgh University Press, Edinburgh.

Fikes, R. and Kehler, T., 1985, The Role of Frame-based Representation in Reasoning: Communications of the ACM, 28, 904-920.

Garcia, G. and Cheng C.H., 1985, The Full Waveform Acoustic Log Inversion Problem: Full

Waveform Acoustic Logging Consortium, ERL-MIT, Cambridge, MA.

Hsu, K. and Baggeroer, A.B., 1986, Application of the Maximum-Likelihood Method (MLM) for Sonic Velocity Logging: Geophysics, 51, 780-787.

Ingram, J.D., Morris, C.F., MacKnight E.E., Parks T.W. B, 1985, Direct Phase Determination of S-wave Velocities from Acoustic Waveform Logs: Geophysics, 50, 1746-1755.

Kimball, C. V. and Marzetta, T.L., 1984, Semblance Processing of Borehole Acoustic Array Data: Geophysics, 49, 274-281.

Kopec, G. E., 1984, The Integrated Signal Processing System ISP: IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-32, No. 4.

Lineman, D.J., Mendelson, J.D., Toksöz, M.M., 1987, An Expert System for Well-to-Well Correlation: Submitted to AAPG Bull.

Martinson, D.G., Menke, W., Stoffa, P., 1982, An Inverse Approach to Signal Correlation: J. Geophys. Res., 78, No. B6, 4807–4818.

Minsky, M., 1975, A Framework for Representing Knowledge: in *The Psychology of Computer Vision*, edited by P. H. Winston.

Myers, C.S., 1980, A Comparative Study of Several Dynamic Time Warping Algorithms for Speech Recognition: M.S. Thesis, MIT, Cambridge, MA.

Newman, P.J., and Worthington, M.H., 1982, In-situ investigation of Seismic Body Wave Attenuation in Heterogeneous Media: Geophys. Prosp. 30, 377-400.

Parson, T.W., 1986, *Voice and Speech Processing*, 297-303, McGraw-Hill.

Rosenbaum, J.H., 1974, Synthetic Microseismograms: Logging in Porous Formations: Geophysics, 39, 14-32.

Rabiner, L.R., Rosenberg, A.E., Levinson, S.E., 1978, Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition: IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-26, No. 6.

Sakoe, H. and Chiba, S., 1971, A Dynamic Programming Approach to Continuous Speech Recognition: Proc. Int. Cong. Acoust., Budapest, Hungary, paper 20C-13.

Schmitt, D.P., 1985, Simulation Numérique de Diagraphies Acoustiques. Propagation d'Ondes dans des Formations Cylindriques Axisymétriques Radialement Stratifiées Incluant des Milieux Elastiques et/ou Poreux Saturés. Ph.D. Thesis, Grenoble Univ., France.

Schmitt, D.P. and Bouchon, M., 1985, Full-Wave Acoustic Logging: Synthetic Microseismograms and Frequency-wavenumber Analysis: Geophysics, 50, 1756-1778.

Simmons, R.G., 1983, Representing and Reasoning About Change in Geologic Interpretation: Technical Report 749, MIT Artificial Intelligence Laboratory.

Smith, R. G. and Young, R. L., 1984, The Design of the Dipmeter Advisor System: Proceedings ACM National Conference, 15-23.

Startzman, R.A. and Kuo, T.B., 1986, An Artificial Intelligence Approach to Well Log Correlation: SPWLA Twenty Seventh Annual Logging Symposium, paper WW.

Thadani, S., 1984, Artificial Intelligence as Applied to the SCAT Method of Dipmeter Interpretation. Part II. The SCAT Hybrid Expert System: Presented at the 54st Annual International SEG Meeting in Atlanta.

Tubman, K.M., 1984, Full Waveform Acoustic Logs in Radially Layered Boreholes: Ph.D. Thesis, MIT, Cambridge, MA.

Tsang, L. and Rader, D., 1979, Numerical Evaluation of the Transient Acoustic Waveform Due to a Point Source in a Fluid-filled Borehole: Geophysics, 44, 1706-1720.

White, J.E., Zechman, R.E., 1968, Computed Response of an Acoustic Logging Tool: Geophysics, 33, 302-310.

Willis, M.E., Toköz M.N., 1983, Automatic P and S Velocity Determination from Full Waveform Digital Acoustic Logs: Geophysics, 48, 1631-1644.

Winkler, K.W. and Nur, A, 1982, Seismic Attenuation: Effects of Pore Fluids and Frictional Sliding: Geophysics, 47, 1-15.

Winston, P.H., Horn, K.P., 1984, *Lisp*, Addison-Wesley, Reading, MA.

Wu, X. and Nyland, E., 1986, Well Log Data Interpretation Using Artificial Intelligence Technique: SPWLA Twenty Seventh Annual Logging Symposium, paper M.

Zhen Z. and Simaan, M., 1986, A Knowledge-Based System for Automatic Segmentation of a Seismic Section: Presented at the 56st Annual International SEG Meeting in Houston.

Appendix A

Sessions with ANIS

This section describes some important facets of the practical use of the `ANIS` system. An initial sequence of waveforms can be segmented in depth (by selecting a subset of offsets), and in time, by creating new subsequences. In order to extract subsignals, the user clicks on the *mouse*, that is, for this application, the computer analogue for scissors. As illustrated by the examples below, the ability of dissecting signals is not restricted by the nature of these signals. Any sequence can be transformed and cut apart without loosing track of the absolute arrival time and sampling rate values. The examples also illustrate the graphical facilities of the system.

Example 1

Figure A-1 shows a raw sequence of 12 synthetic microseismograms in a fast formation with offsets ranging from 2.50 m to 8.00 m, by steps of 0.50 m. The characteristics of the formation are: $V_P = 4000$ m/s, $V_S = 2310$ m/s, $Q_P = 50$, $Q_S = 25$, $\rho = 2.40 \times 10^3$ kg/m^3. The borehole radius is 10 cm, the central frequency of the source 5 kHz and the sampling rate is 11.84 microseconds.

The first weak arrival corresponds to the pseudo-Rayleigh wavetrain. The relative amplitude of the P-wave is too low so that the wave does not appear on the plot shown in figures A-1 and A-2. With the help of the mouse, the user picks arrival times on a few waveforms in the window attached to the sequence. These arrival times correspond to the onsets of the Stoneley wave. The small arrows point at the arrival time values as interpreted and interpolated by the procedure "`extract-wave`". In Figure A-2, ending points of arrivals were chosen by the user on some of the waveforms, so that the extraction of the Stoneley wave is completed automatically for the whole sequence. A new instance of the data type `TRANSFORMED-SEQUENCE` is created and its graphical representation is displayed. This new instance can be accessed for additional manipulations.

Example 2

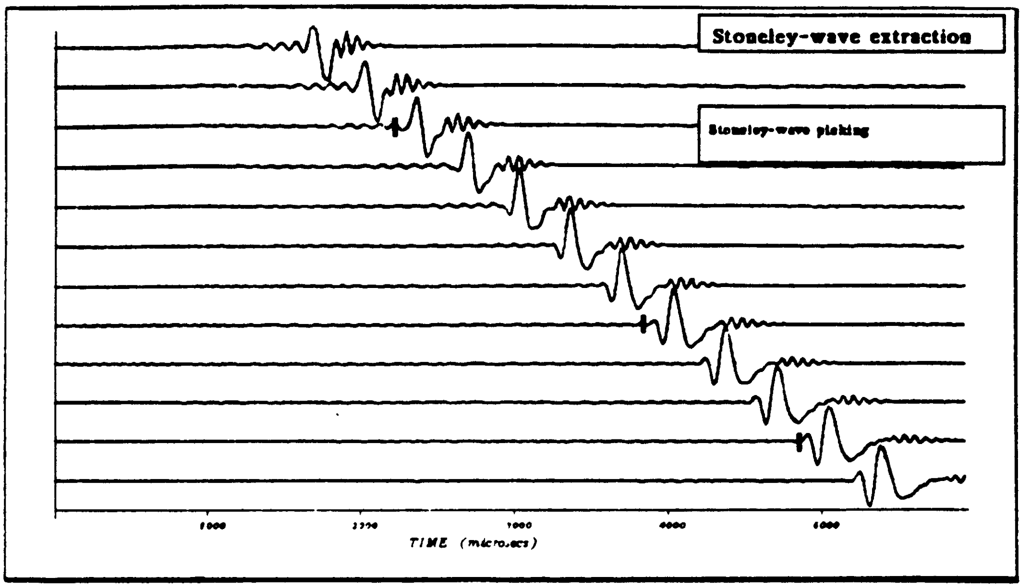
A second example of segmentation is shown on figures A-3, A-4, A-5, and A-6. The initial sequence of synthetic microseismograms displayed on Figure A-3 contains nine waveforms. The characteristics of the formation are: $V_P = 4500$ m/s, $V_S = 2670$ m/s, $\rho = 2.45 \times 10^3$ kg/m^3, $Q_P = Q_S = \infty$, $V_f = 1600$ m/s. The borehole radius is 10 cm, the central frequency of the source 5 kHz and the sampling rate is 11.84 microseconds. The last heigh traces were selected for analysis.

The S-wave arrival times are estimated by automatic threshold detection (see Figure A-4). Note that two attempts were necessary to pick correctly the onset of the S-wave for all traces. The first threshold given as argument to "`autopick-arrival`" was too low which leads to detect the onset of the P-wave arrival instead. As shown on Figure A-5, the S-wave extraction only requires picking the ending points of the arrival since the beginning of the S arrival is already determined. The envelopes of the initial waveforms were computed, which produced the creation of an instance of the data type `TRANSFORMED-SEQUENCE`, whose image is displayed on Figure A-6. Finally, the segments of the envelopes identified as the Airy phases of the arrivals are extracted (Figure A-6).

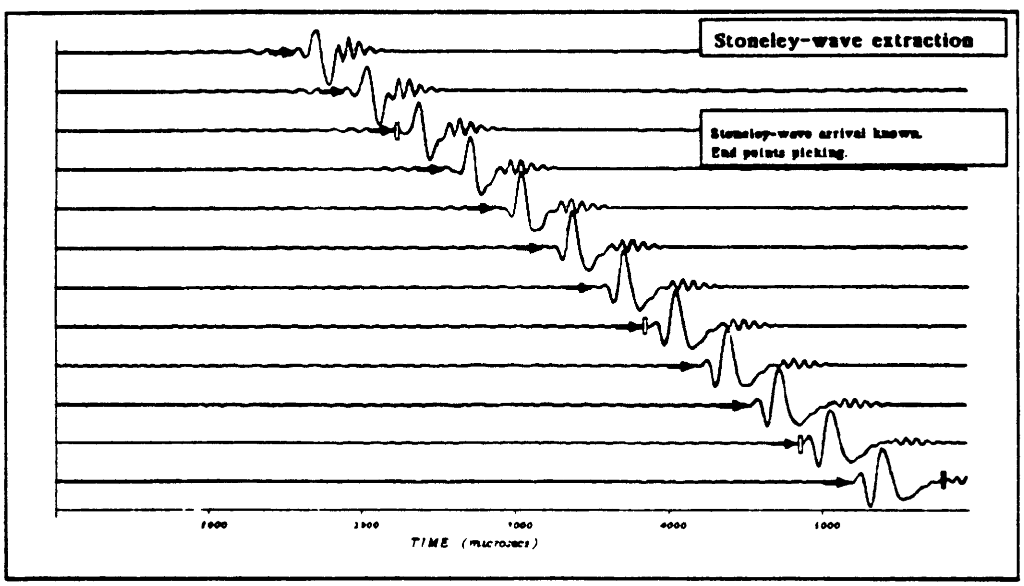
Example 3

Synthetic seismograms were generated for a slow formation with the following characteristics: $V_P = 2800$ m/s, $V_S = 1300$ m/s, $V_f = 1600$ m/s, $\rho = 2.45$, $Q_P = Q_S = Q_f = \infty$, $\rho = 2.45 \times 10^3$ kg/m^3, $\rho_f = 1.20 \times 10^3$ kg/m^3. The borehole radius is 10 cm. The sampling rate is about 11.84 microseconds and the center frequency of the source is 5 kHz.

The original data set, with offsets ranging from 1.00 m to 5.00 m by 50 cm steps, is displayed on Figure A-7. Note that some part of the screen are "mouse-sensitive", i.e., useful information is displayed when the mouse is on them. Envelopes are computed (Figure A-8) that show the respective well-separated energy packets of the P- and Stoneley waves. All traces are associated with mouse-sensitive regions in the window, so that the offset is indicated automatically when the mouse points to a given waveform.

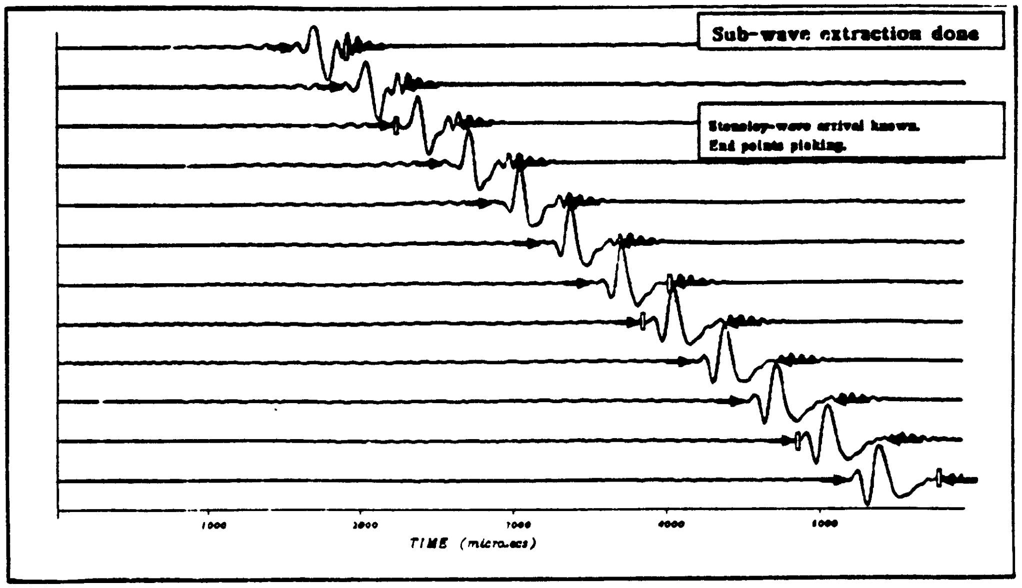


(a)

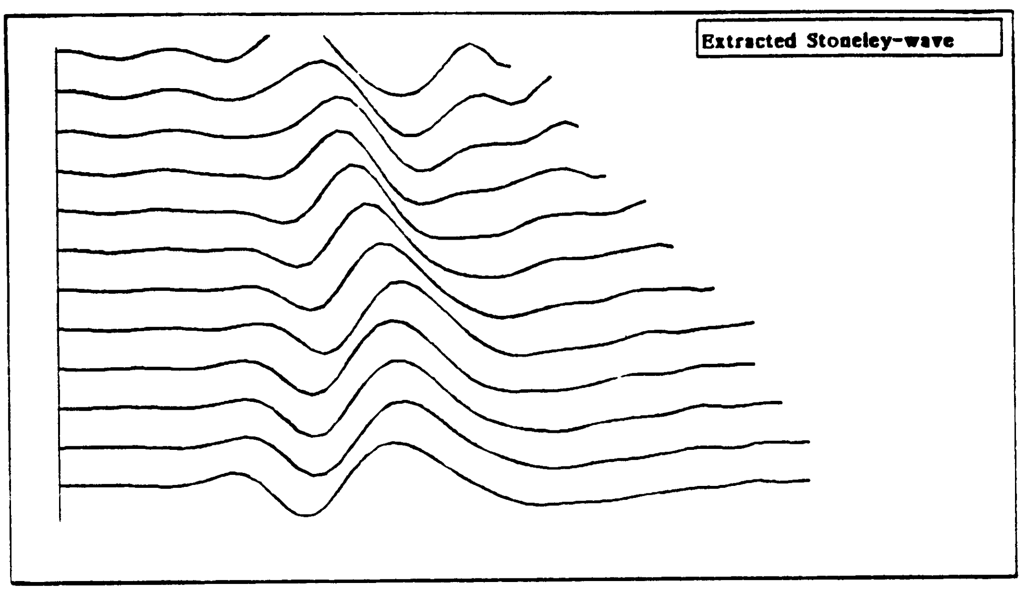


(b)

Figure A-1: Example 1. Extraction of a Stoneley arrival from a sequence of synthetic micro-seismograms. Small rectangles in (a) and (b) are echos of the times picked manually with the mouse. Arrows in (b) point at interpolated values for all traces.



(a)



(b)

Figure A-2: Example 1. Picking of the ending times for the Stoneley wave extraction.

(a) Leftward arrows point at the interpolated end-time values.

(b) The extracted Stoneley wave: a new instance of the data type SUB-SEQUENCE is created.

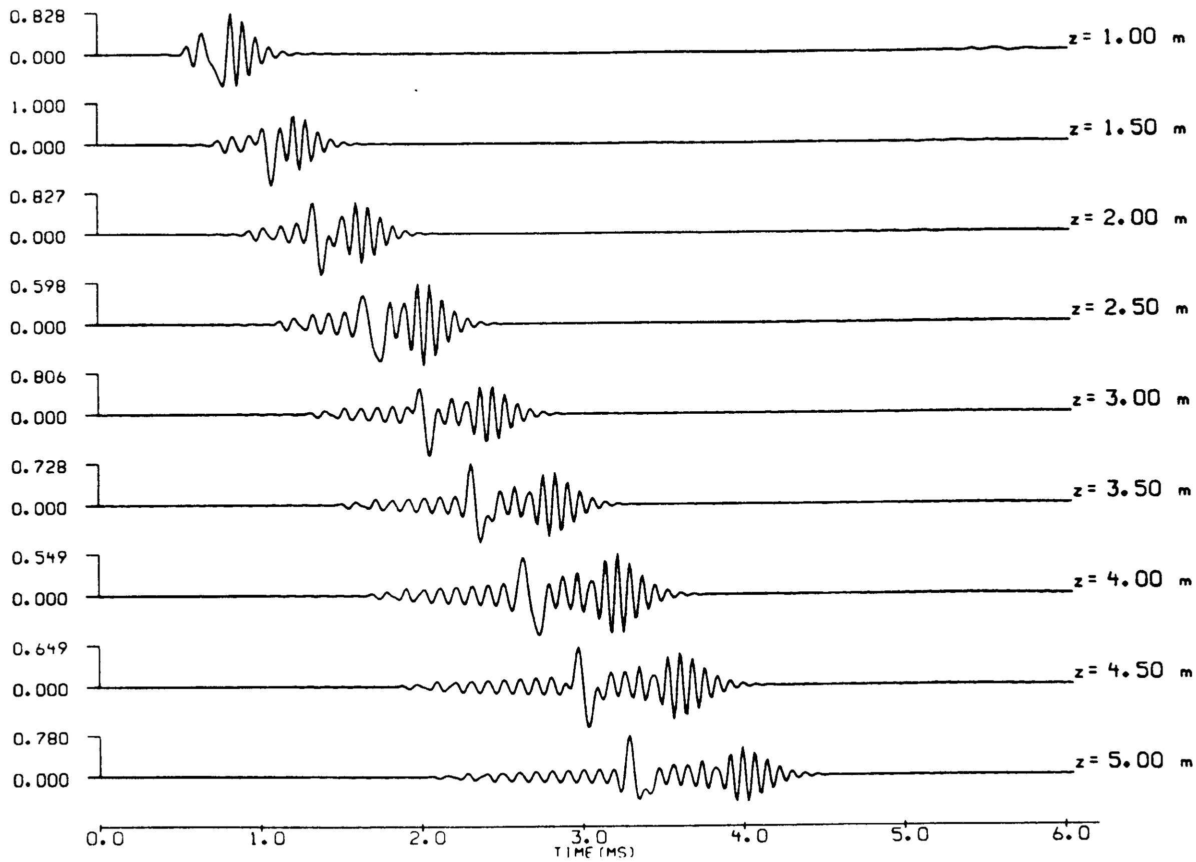


Figure A-3: <u>Example 2</u>. Synthetic microseismograms in an elastic formation with $V_P = 4500$ m/s, $V_S = 2670$ m/s, $\rho = 2.45 \times 10^3$ kg/m^3, $Q_P = Q_S = \infty$, $V_f = 1600$ m/s. The borehole radius is 10 cm, the central frequency of the source 5 kHz, the sampling rate is 11.84 microseconds.

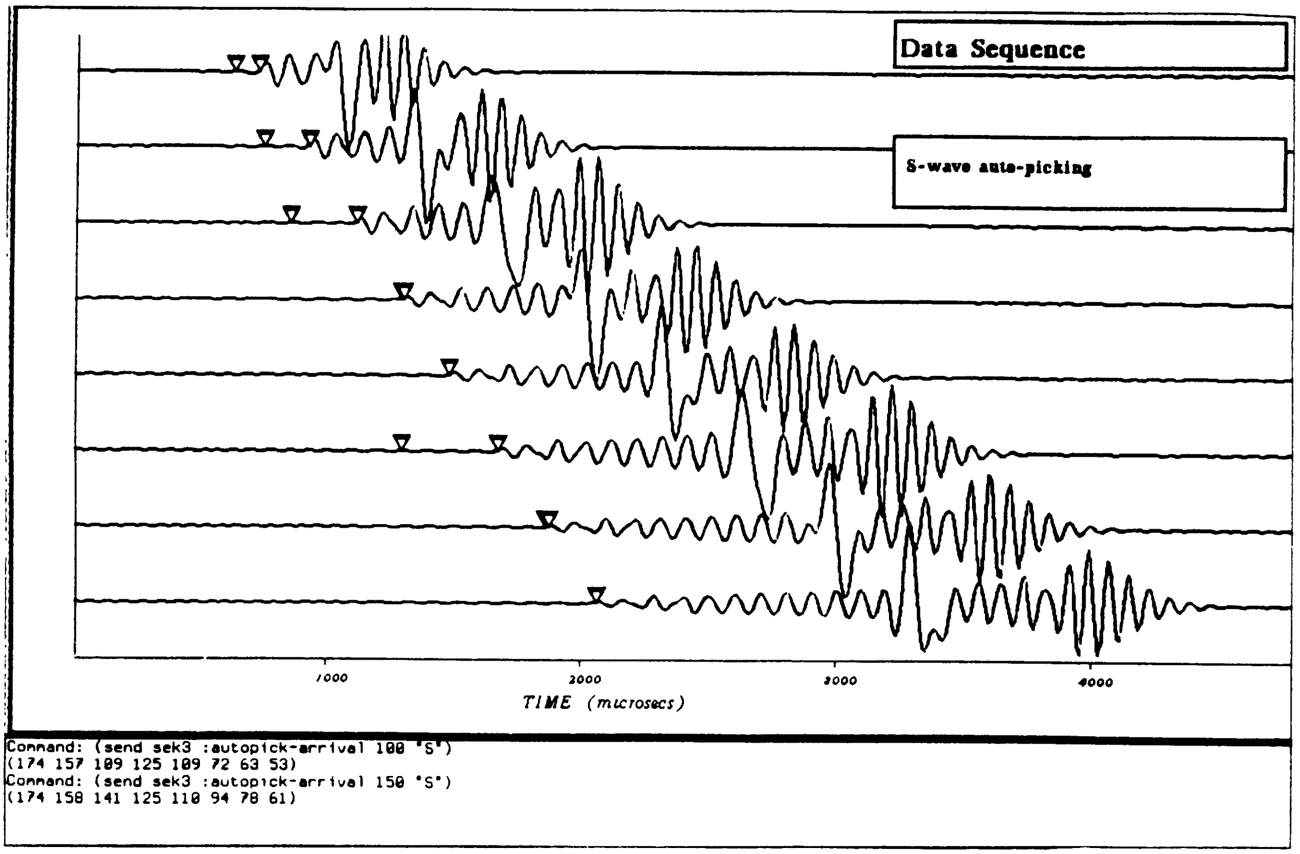
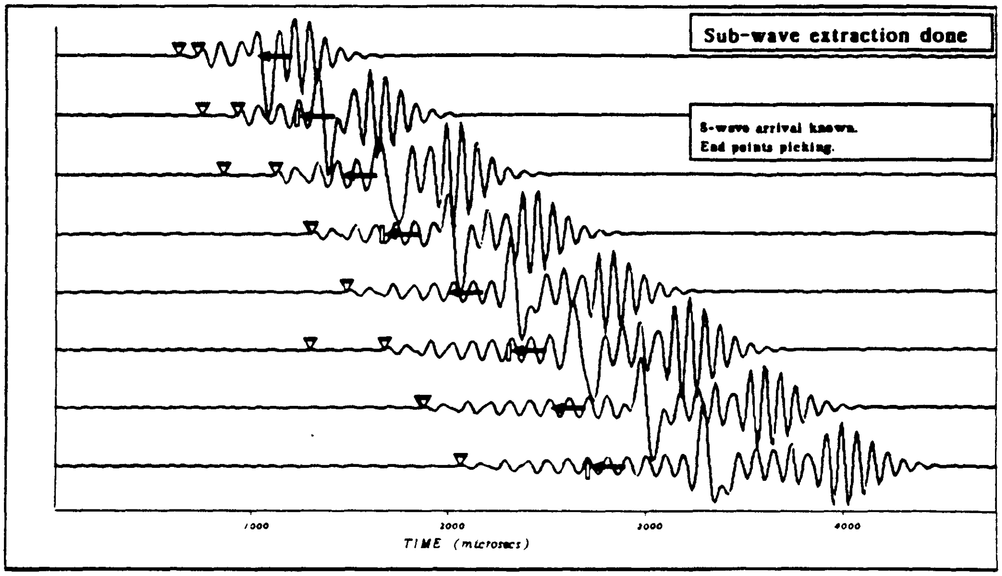
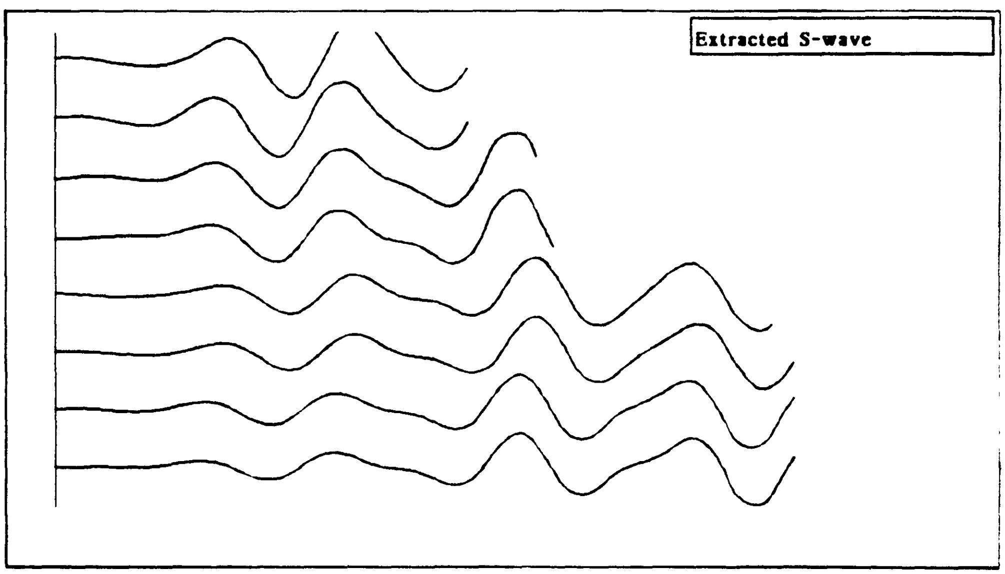


Figure A-4: Example 2. S-wave arrival time determination with threshold. The selected offsets range from 1.50 m to 5.00 m by increment of 50 cm.. Triangles mark the detected arrivals. Note that the first attempt detected in part the onset of the P-wave arrival. A higher threshold corresponds to the first cycle of the S-wave for all traces.



(a)



(b)

Figure A-5: Example 2.

(a) Extraction of the first cycles of the pseudo-Rayleigh arrival for the offsets ranging from 1.50 m to 5.00 m.

(b) The extracted subsequence.

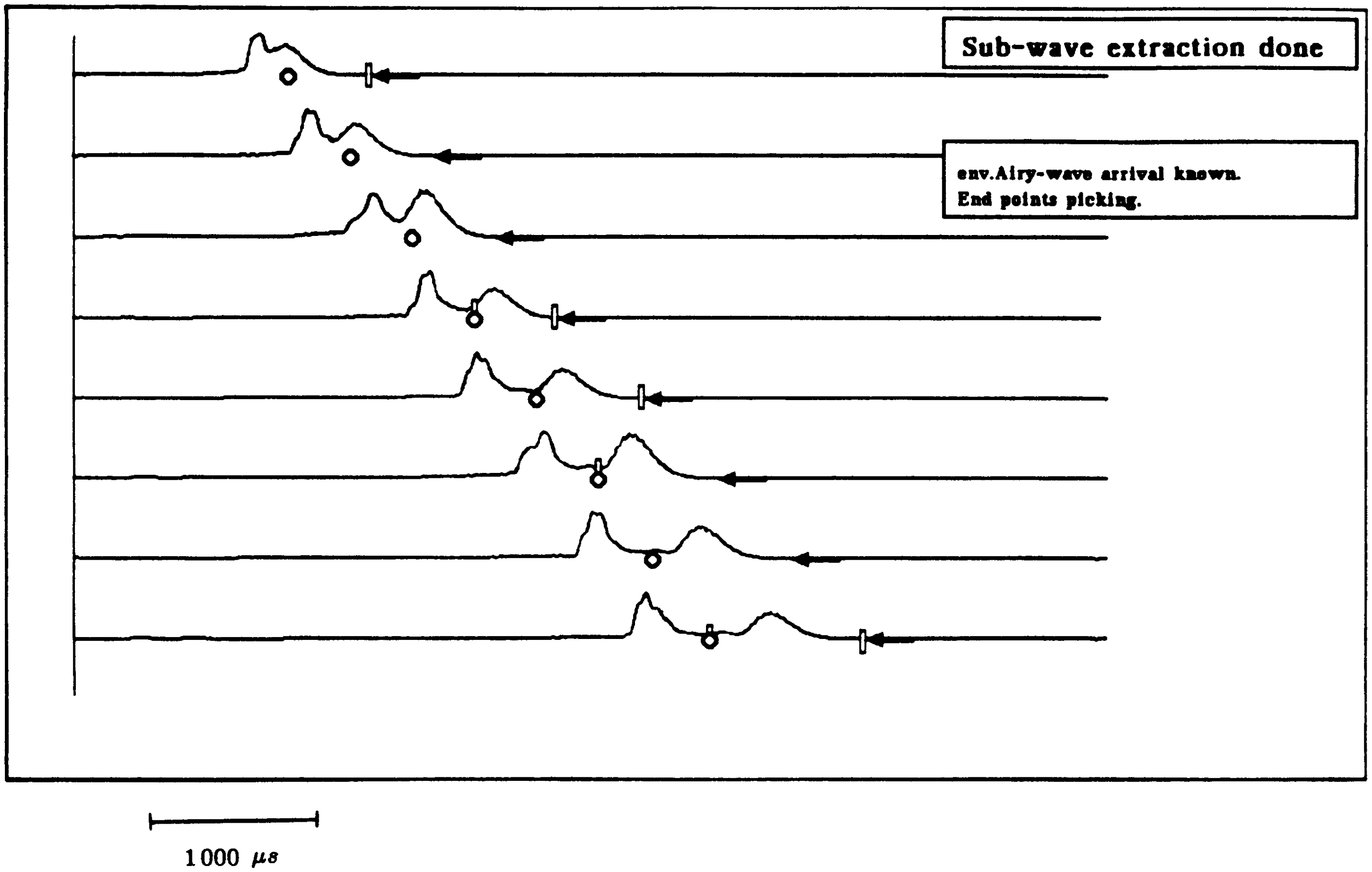


Figure A-6: Example 2. Extraction of a "sub-envelope", i.e. creation of an instance of SUB-SEQUENCE from an instance of TRANSFORMED-SEQUENCE, for the offsets ranging from 1.50 m to 5.00 m. Points actually picked by the user are marked by small hollow rectangles. Circles mark the interpolated arrival times for the first cut of the segmentation.

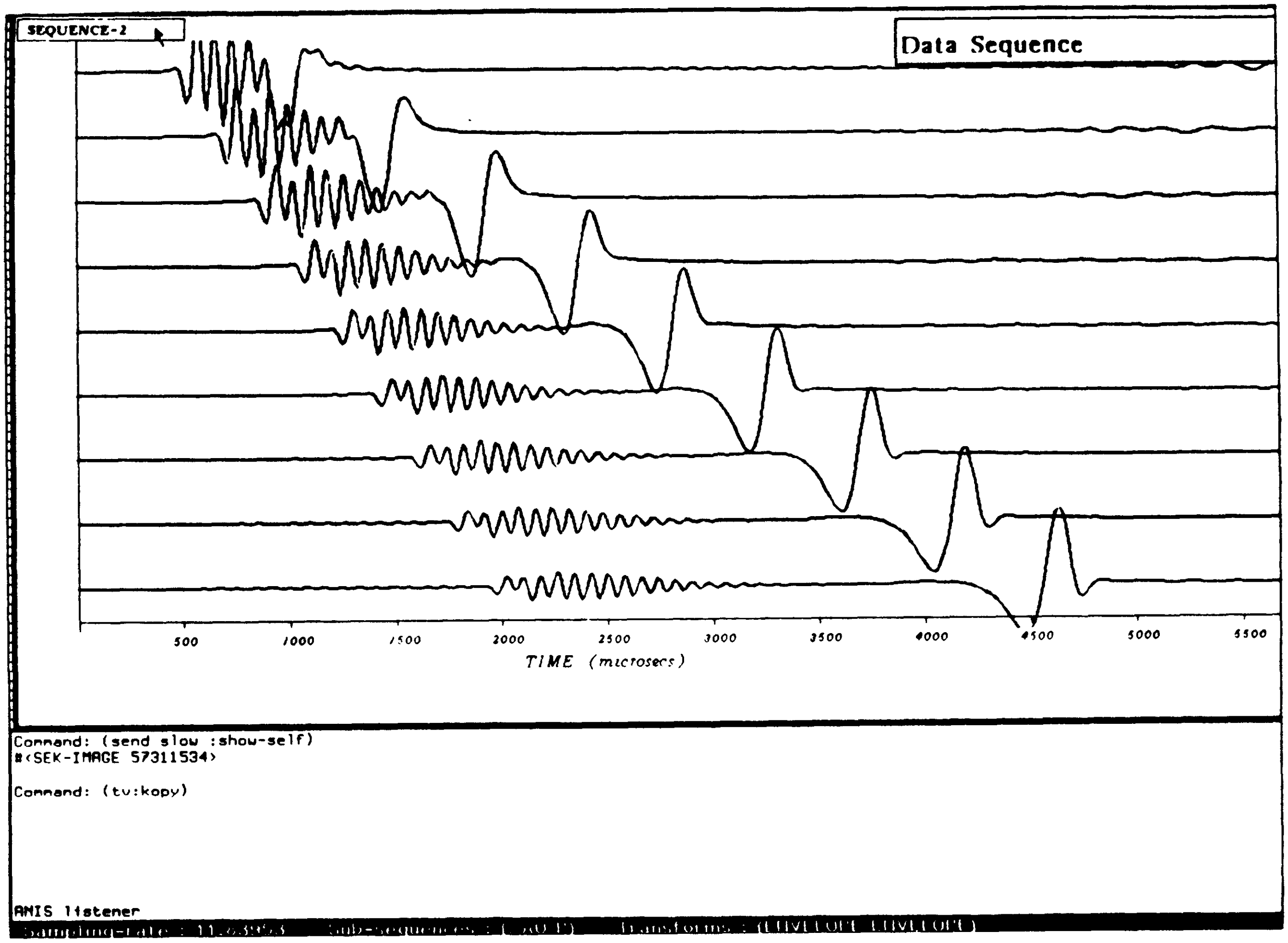


Figure A-7: <u>Example 3</u>. Initial sequence of synthetic microseismograms for a slow formation. Offsets ranging from 1.00 m to 5.00 m by increment of 0.50 m. The black stripe at the bottom of the screen (the "who-line") displays information about the sequence when the mouse arrow is on a "mouse-sensitive" region.

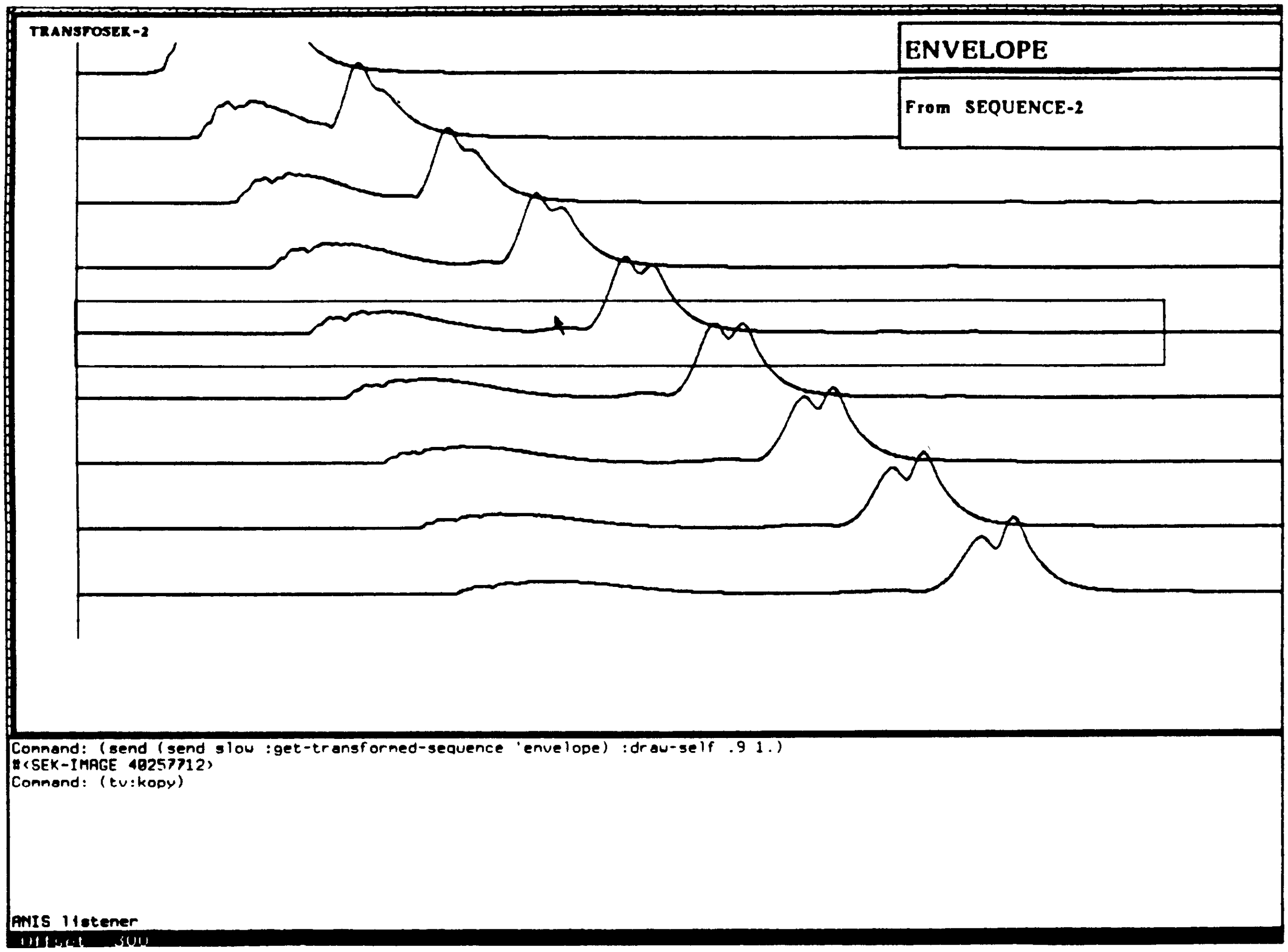


Figure A-8: Example 3. Envelopes of the waveforms shown in Figure A-7. Note the well-separated energy packets corresponding to the P and Stoneley wavetrains. Each region corresponding to a given trace on the screen is "mouse-sensitive"; the offset of the trace is displayed on the "who-line".

Appendix B

ANIS Objects

This appendix lists the descriptions of objects that form the skeleton of the `ANIS` system. The implementation uses *ZetaLisp* objects (the so-called *flavors*). The `ANIS` flavors are structured hierarchically [1] with root `ANIS` and two main top-level types: `DATA-TYPE` for the description of data (i.e. waveforms and sequences) and `ABSTRACT-TYPE` for the manipulation of information and processing results. An object inherits all the slots of its parents. Objects can be declared *abstract* when they are not supposed to have instances by themselves but rather serve as the basis for the definition of more specific concepts. The very top-level object (called `ANIS`) is an abstract object similar to the default *ZetaLisp* basic object [2]. Basic types (integer, real, array, character, etc.) are *ZetaLisp* data types.

The general form of descriptions is:

$$
\begin{gathered}
< \textbf{object} > \; \subset \; < \textbf{super} - \textbf{object} > \\
< \textbf{attribute} > \; \in \; < \textbf{type} > \\
< \textbf{attribute} > \; \in \; < \textbf{type} > \\
\ldots\ldots\ldots\ldots \quad \ldots \quad \ldots\ldots\ldots
\end{gathered}
$$

[1] Flavors are non-hierarchically structured objects but allow to build a hierarchical structure, if needed.
[2] That is "Vanilla flavor".

where < type > is generally a restriction of the form:

- member-of < *ZetaLisp* data type >
- one-of (list-of < *Lisp* object >)
- member-of < **anis** object >
- member-of (list-of < **anis** object >)

Data description

The two principal concepts are `TRACE-TYPE` and `SEQUENCE`.

`DATA-TYPE` $\subset$ **`anis`**

  name $\in$ variable

`SEQUENCE` $\subset$ **`data-type`**

  number-of-traces $\in$ integer

  list-of-traces $\in$ (list-of **`trace-type`**)

  selected-offsets $\in$ (list-of integer)

  sub-sequences-list $\in$ (list-of `SUB-SEQUENCE`)

  transformed-sequences-list $\in$ (list-of `TRANSFORMED-SEQUENCE`)

  image $\in$ **`sek-image`**

  sampling-rate $\in$ real

`RAW-SEQUENCE` $\subset$ **`sequence`**

  traces-in $\in$ (list-of symbol)

  matrix $\in$ 2D-array

`SUB-SEQUENCE` $\subset$ **`sequence`**

  from-sequence $\in$ `RAW-SEQUENCE`

  type $\in$ ('P' 'S' 'Stoneley' any-string)

`TRANSFORMED-SEQUENCE` $\subset$ `sequence`

from-sequence $\in$ (or `RAW-SEQUENCE` `SUB-SEQUENCE`)

transform $\in$ symbol

`TRACE-TYPE` $\subset$ `data-type`

parent-sequence $\in$ `sequence`

infos $\in$ `trace-info`

portrait $\in$ `simple-image`

graph-location $\in$ `region`

parameters $\in$ `temp-values`

transformed-traces-list $\in$ (list-of `transformed-trace`)

sub-traces-list $\in$ (list-of `sub-trace`)

time-serie $\in$ 1D-array

`TRACE` $\subset$ `trace-type`

qualitative-values $\in$ `qualitative-processing-values`

initial-values $\in$ `initial-processing-values`

final-values $\in$ `final-processing-values`

`SUB-TRACE` $\subset$ `trace-type`

parent-trace $\in$ `trace-type`

wave-type $\in$ (one-of 'P' 'S' 'Stoneley' 'X')

processing-values $\in$ `processing-values`

`TRANSFORMED-TRACE` $\subset$ `trace-type`

parent-trace $\in$ (or `trace` `sub-trace`)

processing-values $\in$ `processing-values`

transform $\in$ symbol

Representation of information and results

`ABSTRACT-TYPE` $\subset$ `anis`

owner $\in$ `anis`

`TRACE-INFO` $\subset$ **`abstract-type`**

offset $\in$ integer

number-of-points $\in$ integer

first-time-index $\in$ integer

sampling-rate $\in$ real

`SEK-IMAGE` $\subset$ **`abstract-type`**

name $\in$ symbol

window $\in$ tv:window

io-sheet $\in$ tv:window

region-list $\in$ (list-of **`region`**)

list-of-actions $\in$ (list-of string)

`REGION` $\subset$ **`abstract-type`**

object-in $\in$ **`data-type`**

scaling-factors $\in$ (real real)

upper-left-corner $\in$ (real real)

lower-right-corner $\in$ (real real)

`SIMPLE-IMAGE` $\subset$ **`abstract-type`**

window $\in$ tv:window

`TEMP-VALUES` $\subset$ **`abstract-type`**

time-value $\in$ (list-of real)

any-value $\in$ (list-of real)

`TRACE-PROCESSING-VALUES` $\subset$ **`abstract-type`**

P-velocity $\in$ real

S-velocity $\in$ real

St-velocity $\in$ real

Vp-Vs-ratio $\in$ real

P-amplitude $\in$ real

S-amplitude $\in$ real

St-amplitude $\in$ real

`QUALITATIVE-VALUES` $\subset$ **`trace-processing-values`**

S-existence $\in$ (one-of 'yes' 'no' 'unknown')

P-S-amp-ratio $\in$ real

`INITIAL-PROCESSING-VALUES` $\subset$ **`trace-processing-values`**

P-arrival-time $\in$ real

S-arrival-time $\in$ real

St-arrival-time $\in$ real

`FINAL-PROCESSING-VALUES` $\subset$ **`trace-processing-values`**

P-arrival-time $\in$ real

S-arrival-time $\in$ real

St-arrival-time $\in$ real

Appendix C

ANIS Operators

This appendix lists the main top-level messages (or *methods*) that can be sent to instances of objects (or *flavors*) `TRACE-TYPE` and `SEQUENCE` and to objects below in the hierarchy. Occasionally, a message may not make sense for a given subobject. For instance, the method "`compute-velocities`" can be invoked for instances of `SEQUENCE-I` only, i.e. for unmodified sequences of entire waveforms. The general form of a description is:

`MESSAGE` A1 A2 ... $\langle$ B1 B2 ...$\rangle$

where A1, A2 ... are required parameters and B1, B2 ... are optional parameters — whose default values are presumably reasonable.
The *Zetalisp* syntax to send messages is:

send *flavor-instance :message argument1 argument2* ...

In addition to listed methods, automatic methods are generated to get or to set any slot value — these are *access procedures* [1]. For instance, the *ZetaLisp* messages to get and set a new value for the slot `sampling-rate` are respectively "`sampling-rate`" and "`set-sampling-rate`". A few procedures for data initialization and affectation are defined as regular LISP

[1] ANIS objects and slots are listed in Appendix B.

functions rather than *ZetaLisp* messages.

Operators for sequences

Functions

`SET-SEQUENCE` file-pathname ⟨ sequence-name ⟩

`RESET-SEQUENCE` ⟨ sequence-name ⟩

`CLEAN-SEQUENCE` ⟨ sequence-name ⟩

Messages

Operators

`DRAW-SELF` ⟨ gain time-stretch ⟩

`SHOW-SELF`

`CLEAN-UP`

`DRAW-TRACE` offset ⟨ gain time-stretch ⟩

`GET-TRACE` offset

`GET-SUB-SEQUENCE` wave

`GET-FIRST-SUB-SEQUENCE`

`GET-SECOND-SUB-SEQUENCE`

`GET-TRANSFORMED-SEQUENCE` transform

`AUTOPICK-ARRIVAL` wave threshold

`HANDPICK-ARRIVAL` wave

`EXTRACT-WAVE` wave

`SET-INITIAL-VELOCITIES` wave

`COMPUTE-VELOCITIES` wave

`COMPUTE-AMPLITUDES` wave

`KORRELATION`

`INTERPOLATION` ⟨ factor ⟩

`NORMALIZATION` ⟨ reference-value ⟩

`ENVELOPE` ⟨ window-length ⟩

Information retrieval

`DESCRIBE`

`NUMBER-OF-TRACES`

`OFFSETS`

`SAMPLING-RATE`

`SUB-SEQUENCES`

`TRANSFORMED-SEQUENCES`

`WHAT-ACTIONS`

`INITIAL-RESULTS` ⟨ wave ⟩

`FINAL-RESULTS` ⟨ wave ⟩

`RESULTS` ⟨ wave ⟩

Operators for traces

Setting-up

`SET-TRACE` file-pathname ⟨ trace-name ⟩

Messages

`DRAW-SELF` ⟨ gain time-stretch ⟩

`ENVELOPE` ⟨ window-length ⟩

`SPLINE-SELF` ⟨ factor ⟩

`NORMALIZE-SELF` ⟨ maximum-value ⟩

`AMPLITUDE-MAX`

`MAX-IN-INTERVAL` ⟨ first-time-index last-time-index ⟩

`DESCRIBE`

`INFOS`

Appendix D

Knowledge-Based System Applications in Geophysics

D.1 Introduction

Among the applications of Artificial Intelligence (AI) techniques to real-world problems, there exist a few systems in the fields of geology and geophysics. Generally, these systems address geophysical interpretation problems in domains where a large amount of experience is required. These applications investigated mainly two types of problem; *Prospector* tackles a broad diagnosis problem, very close in nature to medical AI applications, whereas most of the recent applications described in the literature investigate the interpretation of well logging or reflection seismic data in narrower domains, but which often involve solving a pattern extraction and/or a recognition problem.

The most significant applications of artificial intelligence in the fields of geology and geophysics are briefly reviewed, with emphasis on the specific tasks to be achieved and on the structures chosen to solve the problem.

D.2 Prospector

Prospector (Duda et al. 1978, 1979, 1980) is a knowledge-based system intended to assist geologists in evaluating the hard-rock mineral resources of a potentially productive region (a prospect).

The task The system is designed to work in a consultant mode; the user possesses a set of field data in a region and wants the system to come up with a geologic diagnosis about the existence of a class of mineral deposit. The field evidences consist of regional information about the prospect (description of the structural environment) and more local data (existence and composition of a given type of rock). Most evidence is affected by *uncertainties.* The interpreter may have only presumptions or indirect signs that a type of rock is present.

A particular type of deposit is described with a specific geologic model that contains a large amount of probabilistic reasoning. Since evidences contribute with a different strength to the establishment of a new hypothesis, each inference step modifies the current probability of assertions. The task is to match the data against a model in the direction of the most promising hypothesis and ask for additional information when no more evidence is relevant. The final output of the reasoning process is of course weighted by a certainty coefficient. The quality of the conclusion depends heavily on the adequacy (was the model the good one?) and on the completeness of the model.

Internal structure Each model is encoded as an *inference network* that captures the geologic knowledge required to solve one of the deposit evaluation subtasks. The structure of models is hierarchical, from assertions which can be matched by field evidences to several higher-level assertions. The inference network is defined as the network of relations between pieces of evidences and geologic hypotheses. Figure D-1-a shows a simplified part of the inference network describing a massive sulfide deposit model. The relations in the network are mostly inference rules (an example of their internal representation is shown in Figure D-1-b). Due to the nature of the domain, these rules must take into account uncertainties at the same time:

field evidences are assumed to be inaccurate and incomplete and conclusions drawn from each piece of evidence are themselves uncertain.

Three different relations can be distinguished according to the nature of the uncertainty contained in the assertions: *logical* relations, *plausible* relations and *contextual* relations. Each type of relation is associated with a law of probability propagation. The probability of an hypothesis defined as the logical conjunction or disjunction of a set of evidences is computed with the Zadeh fuzzy-set formulae; there are respectively equal to the minimum and the maximum of the probability of the antecedent facts. When each evidence has a probability and contribute in a different way to the change of probability of the hypothesis (plausible relation), an ad-hoc Bayesian formalism is used to propagate the variations of probability in the network (see Duda et al., 1980). Contextual relations capture the fact that an evidence can be relevant for a conclusion only in a given context.

Relations are invoked in a backward-chaining mode, the initial direction being the one of the highest probability. Assertions are also connected by a taxonomic structure which allows the system to recognize subsumption between geologic concepts.

D.3 Well logging interpretation

In well-logging operations, tools are lowered into a borehole, then raised to the surface while they measure a variety of physical quantities (traveltime of compressional waves, resistivity, radioactivity, etc.). The problem of inferring a geologic description given these measurements and the responses of the different sensors, is a large inverse problem.

Two facts help in the geologic reconstruction task. First, the measurement of various independent parameters gives better constraints on the possible solutions together with sources of information such as cores, cuttings, etc.. Second, interpreters use an important amount of geologic knowledge. By applying general geologic laws, and more specifically, knowledge of the local geologic environment, experts can drastically reduce uncertainties. This later characteristic of interpretation explains why purely algorithmic and statistical approaches are not sufficient to solve the entire problem. Tentatives have been made to capture the interpretation process with knowledge-based systems.

The Dipmeter Advisor

The first system that tackles a log interpretation problem was the *Dipmeter Advisor* (Davis et al. 1981, Smith and Young 1984).

Nature of the task The *dipmeter* tool is composed of 4 coplanar sensors which measure the resistivity of rocks in different directions around the borehole. By computing correlations between the variations of resistivity it is possible to infer the magnitude and azimuth of the dip of lithological layers. The data are represented in the following way: at each depth where correlation is meaningful, the measure is plotted as a small arrow called *tadpole*. The position of the tadpole in the inclination scale indicate the magnitude of the dip. The direction of the dip is represented by the direction of the tail of the tadpole. Some patterns of tadpoles are related to structural and sedimentary features. An example of such data with the corresponding geologic interpretation is presented in Figure D-2. Interpreters first look for large-scale tilting linked with structural dip, then identify zones showing more local patterns characteristic of the stratigraphic environment. Combining these observations with other logs, they use geologic knowledge to describe the stratigraphic features.

System structure Three main components were built to capture the reasoning process of human interpreters:

- A set of pattern detection algorithms to identify lithology zones and tadpole patterns.
- About 90 production rules used to infer the presence of geologic features. The basic object in the clauses of the rules is a *zone* corresponding to a given layer of formation. Each zone is associated with two kinds of knowledge: the existence of significant patterns and their

geometry (derived from the pattern detection step) and geologic evidences that may be added by the human interpreter.

- An inference engine that invokes rules in a forward-chained manner. The control strategy is limited to a rule ordering.

The system is designed to allow the user's intervention at every stage of the process and the modification of conclusions reached by the inference engine.

Other well logs interpretation systems

Recently, a few systems were designed to solve parts of the well-log interpretation problem. Their structures are very close to the Dipmeter Advisor organization.

- The *SCAT* hybrid expert system (Thadani, 1984) also interprets dipmeter data but both inputs and outputs are different from the Dipmeter Advisor ones. Tilt information is represented as a set of curves (dip as a function of azimuth, dip versus depth, etc.) and outputs are diagrams (cross-sections and contour maps of the structure surrounding the well). Thus the pattern vocabulary is different, but the structure of the program and the scope of the rules are very similar to the Dipmeter Advisor ones.

- *LITHO* (Bonnet and Dahan, 1983) addresses the more general problem of building a complete lithology log given a set of measurements in a borehole. The system uses the same techniques: pattern detection and production rules to draw geologic inferences. The pattern recognition step consists here of a syntactic parsing of logs. Shape of the curves and geologic (external) knowledge are the two types of evidence used in the rule-base. A large amount of geologic knowledge is required — about 500 rules are necessary.

  This system uses certainty factors in order to assign a probability for the presence of each plausible lithology. The final lithology log is obtained by a classic clustering method but probabilities of existence are taken into account for the determination of lithology. Geologic knowledge is used to help in deciding whether a formation is more likely to exist in a given context when a statistical analysis cannot discriminate between plausible solutions. Symbolic reasoning appears as a solution to remove the ambiguities of a purely statistical approach.

- The *COREX* system (Lineman et al., 1987) contains a set of rules that embodies knowledge of the paleo-environment to guide an automatic correlation of logs. Geologic knowledge helps in optimizing the dynamic programming algorithm used for signal matching, and adding constraints on the solution space as well.

Other applications in well-logging interpretation were investigated by Wu and Nyland (1986), and Startzman and Kuo (1986). In seismic reflection interpretation, similar issues occur for pattern recognition and/or texture analysis. Addition of geologic constraints via a rule-based formalism is a way to improve the performances of pattern analysis algorithms (see for instance Zhen and Simaan, 1986).

D.4 Structural interpretation: the *Imagining* technique

The *imagining* technique (Simmons, 1983), has been developed to attack the problem of building an interpretation of simplified two-dimensional cross-sections of the subsurface.

The problem A cross-section represents a "cut" of the spatial configuration of geologic formations. The interpreter's task is to infer a sequence of events that explains the present situation (in agreement with a set of geologic laws and hypotheses). The imagining technique tackles a subproblem of the interpretation task: the system uses both quantitative and qualitative methods to *simulate* an hypothesized sequence of events that leads to the observed section. A sequence of events generated by quantitative simulation is presented in Figure D-3. The geologic interpretation model is the so-called *layer cake* model which assumes the initial depositions to be horizontal and laterally infinite. Two additional assumptions are made. First, the only structural modifications are due to sedimentation, erosion, faulting, homogeneous tilting and intrusions (no folding or "curved" structures). Second, interacting simultaneous events are not considered.

It should be noted that the success of the simulation does not guarantee the hypothesized sequence to be the "right" one. Imagining is a test method, part of the general *generate and test* paradigm. The goal is to check the validity of a particular solution; no indications are given on the range of possible alternate solutions.

System organization The simulation builds on the representation of geologic *objects* (rock-unit, formation, boundary, geologic point) and *processes* which cause the spatial change of these objects. The overall simulation is done in two steps:

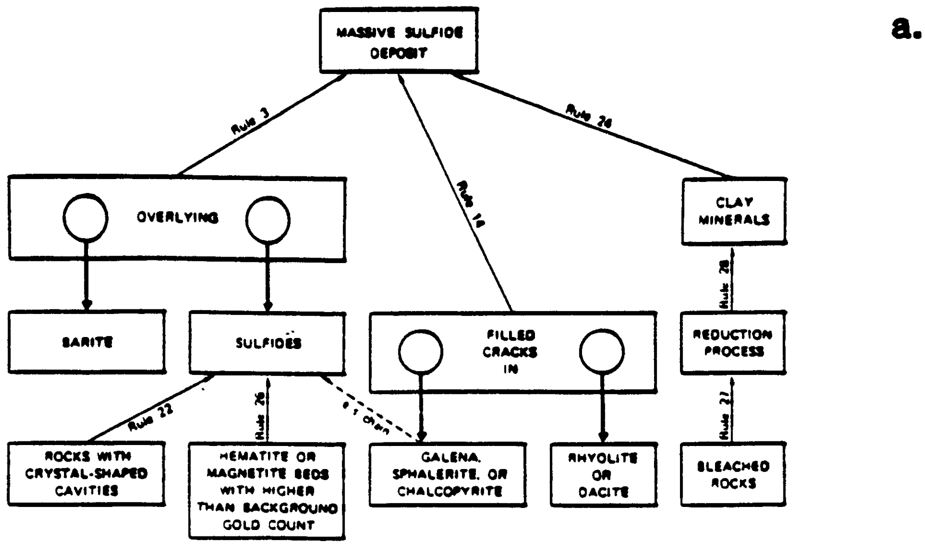
- The *qualitative* simulation reasons about objects (frame-like structures) whose existence and attributes change over time. This representation (called *history*) enables to simulate change qualitatively, checking constraints and determining the range of legal quantitative values for the attributes.
- The estimated values are then used to simulate *quantitatively* changes in *diagrams*. Diagrams are two-dimensional representations of cross-sections at any point in time. Geometric objects are described as a collection of vertices, edges and faces which enables to recover the spatial relations between units. Each process is a quantitative geometric operation applied to the current state of the section.

Given a set of geologic and geometric laws and constraints, and a representation of physical processes, the system simulates the hypothesized succession of events that explains the final cross-section. Representations are chosen to do spatial and temporal reasoning both in a quantitative and qualitative ways. A valuable side-effect of this method is to give the values of the geologic objects' attributes at any time during the geologic history.

D.5 Conclusion

In the geologic and geophysical domains, the first rule-based approaches to solve interpretation tasks were partially successful. Whereas these techniques were found suitable to design *consultant* systems, the expertize can fall abruptly when dealing with real-world situations. In order to maintain performances in new environments, the rule-base must grow and embody more specific and specialized pieces of knowledge. This observation motivated further research to represent a deeper knowledge of geologic principles. The *Imagining* technique shows the interest of using multiple representations for taking into account causal relationships. The rationale for this approach is that different classes of knowledge about the domain are useful for each subtask, and require different representational tools. The drawback is that a representation that attempts to describe very general physical processes can lead to untractability and inefficiency.

When geophysical signals are actual inputs of the system, the extraction of significant features is done by processing methods whose performance is controlled by the choice of a set of parameters. In the Dipmeter Advisor system, for instance, original data (resistivity curves) are processed to give the dip estimates. Then, the user specifies parameters to optimize the pattern detection algorithms. The general data-flow in the program is from the processing modules to symbolic translation and reasoning. The reverse path is not addressed. In fact, the choice and optimization of processing parameters often depend on the actual geologic context and could be constrained with the acquisition of geologic information.



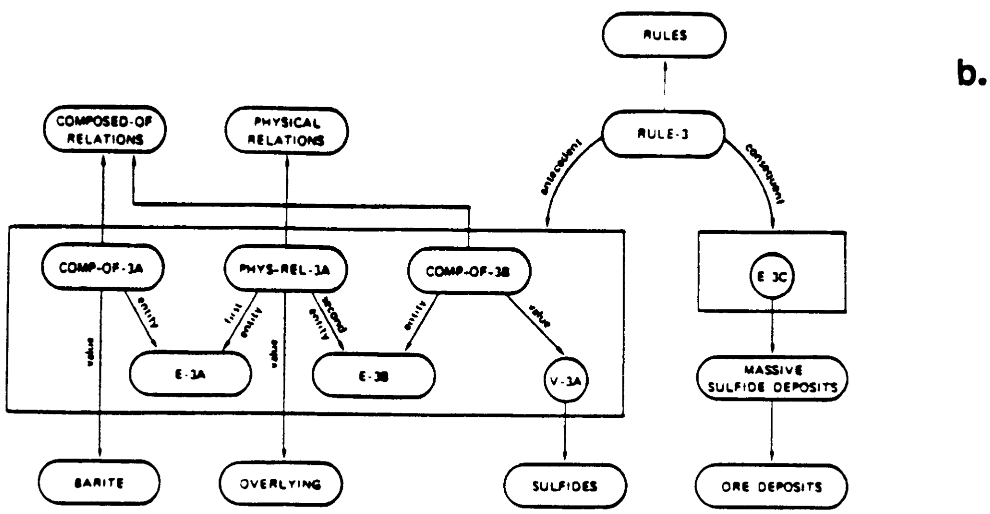


Figure D-1: *Prospector*:

a- Simplified representation of a *Prospector* model implemented as an inference network.

b- Representation of the rule "barite overlying sulfides suggests the possible presence of a massive sulfide deposit."

After Duda et al., 1978.

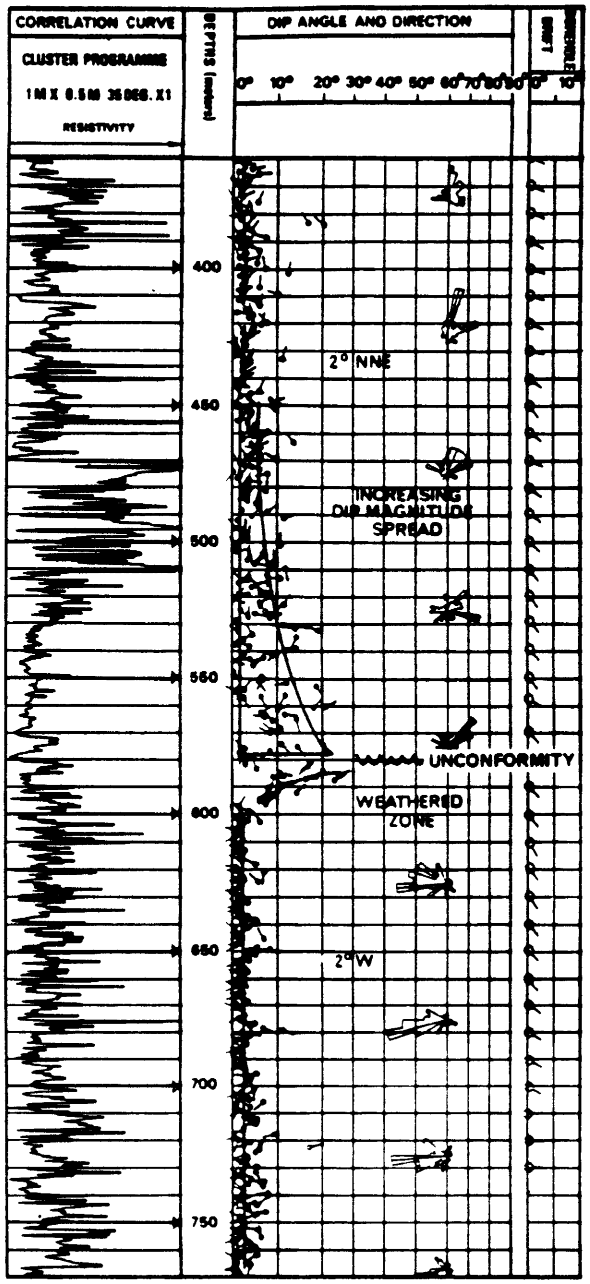
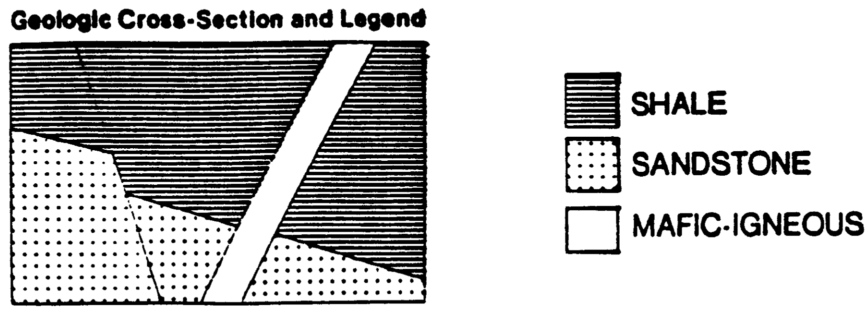


Figure D-2: An example of dipmeter data representation.

Patterns of "tadpoles" are interpreted in terms of geologic situations.



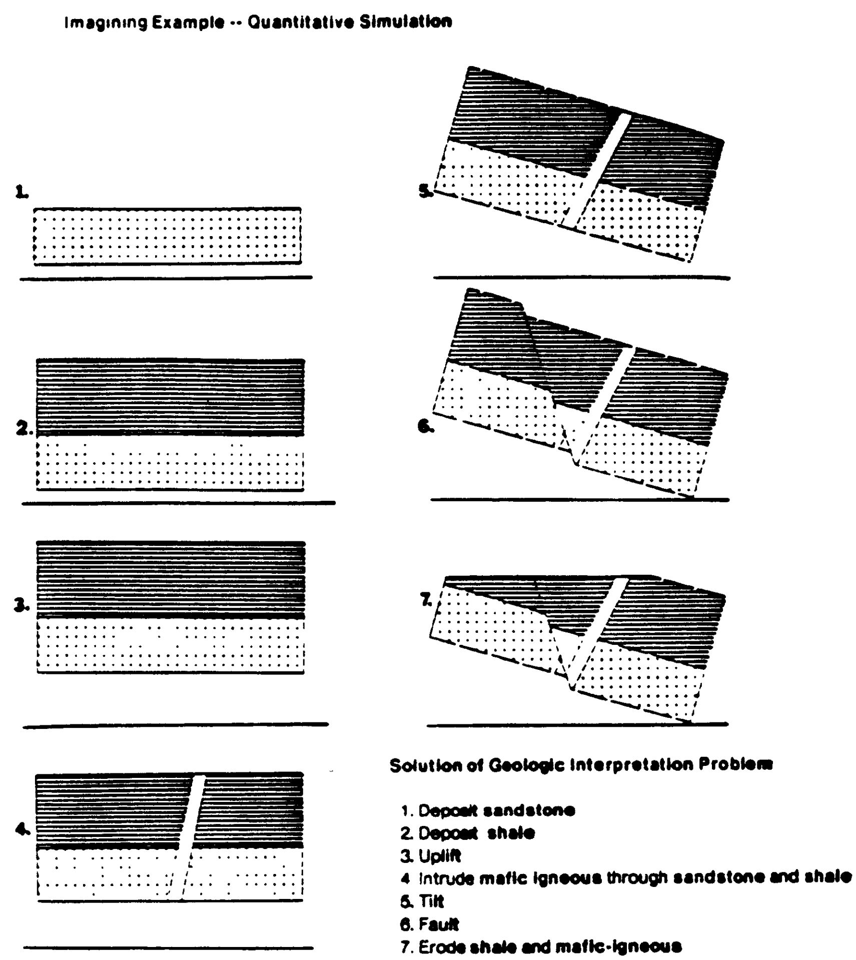


Figure D-3: Given a scenario of geologic events, the sequence of facts is quantitatively simulated with diagrams. The final diagram can be matched against the initial geologic cross-section at the top.

After Simmons, 1983.

Appendix E

Rules for Qualitative Reasoning

The rules that draw inferences in the petrophysical domain are listed bellow. These rules, implemented with the *KEE* shell (Fikes and Kehler, 1985), are translated in an English-like form — very close to the actual *KEE* syntax. A description of the LITHO knowledge-base that embodies the declarative part of the reasoning system can be found in Chapter 3, part 3. The syntax for rules is:

> **IF** < conjunction or disjunction of clauses >
> **THEN** < clause or conjunction of clauses >,

where a clause can have two basic forms:

- 3-uples (*attribute, object, value*). Ex: The *porosity* of the *rock* is *null*, where *object* must be an instance of the frame *rock* in the LITHO knowledge-base.
- Frame membership [1]. Ex: The *rock* is a *sandstone*.

[1] This form is not used in the consequent part of the rules. It would be for classification purpose, i.e. identification of the lithology.

The *KEE* syntax allows negation of clauses. In practice, negations are avoided whenever it is possible. The declarative knowledge in LITHO is structured in a way that limitates the need for negation. Attribute's values are restricted by the definition of frames; the legal values of qualitative petrophysical parameters (ϕ, k, V_p, etc.) are generally element of (null low medium high) or (low medium high). Rock ***fracturation*** can be ***present*** or ***absent***.

Petrophysical rules

This set of rules attempts to capture qualitatively the physical relationships between petrophysical parameters. The porosity is supposed to be the connected one. ***Shales*** and ***evaporites*** must often be excluded from the general behavior. Most rules capture general physical relationships. Some rules, however, should depend on the geological environment — for instance, when relating the depth of the rock, its lithology and the wave velocities.

```
* RULE-A.1
   IF          THE POROSITY OF THE ROCK IS NULL
   THEN        THE PERMEABILITY OF THE ROCK IS NULL

* RULE-A.2
   IF          THE ROCK IS AN EVAPORITE or A SHALE or A CARBONATE
                and THE FRACTURATION OF THE ROCK IS ABSENT
   THEN        THE PERMEABILITY OF THE ROCK IS NULL

* RULE-A.3
   IF          THE POROSITY OF THE ROCK IS LOW
                and  THE FRACTURATION OF THE ROCK IS ABSENT
   THEN        THE PERMEABILITY OF THE ROCK IS NULL

* RULE-A.4
```

```
IF          THE POROSITY OF THE ROCK IS HIGH
              and  THE FRACTURATION OF THE ROCK IS PRESENT
              and  THE ROCK IS not A SHALE or AN EVAPORITE
THEN        THE PERMEABILITY OF THE ROCK IS HIGH

* RULE-A.5
IF          THE ROCK IS A NON SEDIMENTARY ROCK
              and  THE FRACTURATION OF THE ROCK IS ABSENT
THEN        THE PERMEABILITY OF THE ROCK IS NULL

* RULE-A.6
IF          THE MAX DEPTH OF THE ROCK IS SHALLOW
              and  THE ROCK IS A CLASTIC ROCK
THEN        THE POROSITY OF THE ROCK IS HIGH
              and  THE ATTENUATION OF THE ROCK IS HIGH

* RULE-A.7
IF          THE PERMEABILITY OF THE ROCK IS MEDIUM or HIGH
              and  THE FRACTURATION OF THE ROCK IS ABSENT
THEN        THE POROSITY OF THE ROCK IS HIGH

* RULE-A.8
IF          THE PERMEABILITY OF THE ROCK IS NULL
              and  THE FRACTURATION OF THE ROCK IS PRESENT
THEN        THE POROSITY OF THE ROCK IS NULL

* RULE-A.9
IF          THE PERMEABILITY OF THE ROCK IS LOW
              and  THE FRACTURATION OF THE ROCK IS PRESENT
THEN        THE POROSITY OF THE ROCK IS LOW
```

* RULE-B.1

IF	THE ROCK IS A SEDIMENTARY ROCK
	and THE ROCK IS not AN EVAPORITE
	and THE MAX DEPTH OF THE ROCK IS SHALLOW
THEN	THE P- and S-VELOCITY OF THE ROCK ARE LOW

* RULE-B.2

IF	THE ROCK IS A NON CLASTIC ROCK
	and THE MAX DEPTH OF THE ROCK IS DEEP
THEN	THE P-VELOCITY OF THE ROCK IS HIGH
	and THE S-VELOCITY OF THE ROCK IS HIGH

* RULE-B.3

IF	THE P-VELOCITY or THE S-VELOCITY OF THE ROCK IS HIGH
	and THE MAX DEPTH OF THE ROCK IS MEDIUM or DEEP
THEN	THE ATTENUATION OF THE ROCK IS LOW

* RULE-B.4

IF	THE ROCK IS A NON SEDIMENTARY ROCK
	and THE FRACTURATION OF THE ROCK IS ABSENT
	and THE MAX DEPTH OF THE ROCK IS MEDIUM or DEEP
THEN	THE P-VELOCITY OF THE ROCK IS HIGH
	and THE S-VELOCITY OF THE ROCK IS HIGH

* RULE-B.5

IF	THE FRACTURATION OF THE ROCK IS ABSENT
	and THE ROCK IS A NON CLASTIC ROCK
	and THE MAX DEPTH OF THE ROCK IS MEDIUM or DEEP
	and THE POROSITY OF THE ROCK IS NULL or LOW

THEN THE P-VELOCITY OF THE ROCK IS HIGH

and THE S-VELOCITY OF THE ROCK IS HIGH

* RULE-C.1

IF THE POROSITY OF THE ROCK IS LOW or MEDIUM

and THE PERMEABILITY OF THE ROCK IS HIGH or MEDIUM

THEN THE FRACTURATION OF THE ROCK IS PRESENT

* RULE-C.2

IF THE ROCK IS not A SHALE

and THE ROCK IS not AN EVAPORITE

and THE PERMEABILITY OF THE ROCK IS LOW

and THE POROSITY OF THE ROCK IS MEDIUM or HIGH

THEN THE FRACTURATION OF THE ROCK IS ABSENT

* RULE-C.3

IF THE ROCK IS A NON SEDIMENTARY ROCK

and THE PERMEABILITY OF THE ROCK IS MEDIUM or HIGH

THEN THE FRACTURATION OF THE ROCK IS PRESENT

* RULE-C.4

IF THE ROCK IS AN EVAPORITE

and THE MAX DEPTH OF THE ROCK IS DEEP

THEN THE FRACTURATION OF THE ROCK IS ABSENT

* RULE-C.5

IF THE POROSITY OF THE ROCK IS not NULL

and THE PERMEABILITY OF THE ROCK IS NULL

THEN THE FRACTURATION OF THE ROCK IS ABSENT

```
* RULE-C.6
   IF          THE PERMEABILITY OF THE ROCK IS not NULL
                 and  THE ROCK IS A SHALE or AN EVAPORITE or A CHALK
   THEN        THE FRACTURATION OF THE ROCK IS PRESENT
```

The following rule illustrates the use of "well-logging knowledge" to get decisive information about rocks' properties.

```
* RULE-MC 1
   IF          THERE IS A MUD CAKE WITHIN THE ZONE
   THEN        THE POROSITY OF THE ROCK IS not NULL
```

Negation rules

Reasoning with exceptions has several drawbacks. In particular, there is often a trade-off between the transparency and expressive power of representation and the computational cost. The following sample rules illustrate this point in two ways:

- Rules must be written to state *explicitly* that an individual is *not* member of a given frame when this fact is relevant to draw inferences (see rules *X-1, X-2*).
- Similarly, when the legal slot values are not declared disjoint, one must explicit negations such as in rules *X-3* and *X-4*.

In the case of frame membership, the clauses required to express "non-membership" are determinated by the given hierarchical structure of the knowledge base. The form of rules *X-1* and *X-2* are clearly justified when considering the tree structure of LITHO (see Figure E-1).

```
* RULE-X.1
   IF          THE ROCK IS A NON SEDIMENTARY ROCK or A NON CLASTIC ROCK
```

```
                or A SANDSTONE or A SAND  or A CONGLOMERATE
THEN                  THE ROCK IS not A SHALE

* RULE-X.2
IF             THE ROCK IS A NON SEDIMENTARY ROCK or A CLASTIC ROCK
                or A CARBONATE or A COAL
THEN           THE  ROCK IS  not AN EVAPORITE

* RULE-X.3
IF             THE POROSITY OF THE ROCK IS LOW or MEDIUM or HIGH
THEN           THE POROSITY OF THE ROCK IS not NULL

* RULE-X.4
IF             THE PERMEABILITY OF THE ROCK IS LOW or MEDIUM or HIGH
THEN           THE PERMEABILITY OF THE ROCK IS not NULL
```

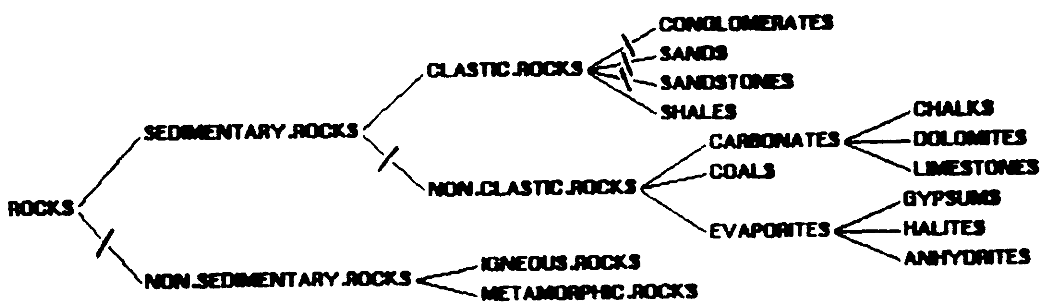


Figure E-1: The graph of the LITHO knowledge-base.

Expressing that a rock is *not* a shale, requires few negative clauses thanks to the hierarchical structure of the knowledge-base. Negations are needed only at the levels indicated by bars