1994

Design And Implementation Of Expert Systems In Trace Metal Analysis

Sharbari Lahiri

DESIGN AND IMPLEMENTATION OF EXPERT SYSTEMS IN TRACE METAL ANALYSIS

by

Sharbari Lahiri

Department of Chemistry

Submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

April 1994

ISBN 0-315-93180-9

Abstract

Atomic Absorption Expert (AAexpert) is an expert system dealing with automated analysis for trace metals by atomic absorption spectrometry (AAS). This program consists of several stand-alone expert systems, each performing a specific task associated with the analysis.

In this thesis, the design and implementation of three important modules of AAexpert have been studied. These are (a) AAdiagnosis - detects problems associated with analytical data of poor quality and instrument malfunction, (b) AA-Quality Control (AA-QC) - provides a real-time assessment of the quality of measured absorbance values, and (c) AAmethods - selects a method of analysis from a method selection database and a rulebase.

The focal points of the work presented here are (a) the transfer of human expertise to a computer program, and (b) the requirements for totally unattended automated analyses.

In the transfer of human expertise, knowledge acquisition has been stressed and the development of a knowledge table has been described. The developer of the expert system assembles the chemical knowledge as case histories. The knowledge table generates rules that represent the acquired knowledge. This approach of knowledge acquisition would greatly assist chemists in the generation, expansion, and portability of their knowledge bases. The implementation of the knowledge table has been demonstrated in AAdiagnosis in which the developer gathers knowledge as a matrix of symptoms (rows) and the underlying causes (columns).

The goal towards totally automated analysis has been described in AA-QC. This module closely interacts with AAcontrol, a module responsible for solution handling and acquisition of data by flame AAS (FAAS). It has been shown that by modelling the detector response (the absorption profile), it is possible to detect a few common and simple problems associated with analysis by FAAS. AA-QC uses a training set based on the numerical analysis of the

absorption profile for a set of standard solutions. AA-QC uses production rules to detect problems associated with the analysis.

The use of an expert database system has been studied in AAmethods. The use of rules to suggest a "closest match" method of analysis contained in an "expert database" has been demonstrated.

Dedication

I dedicate this thesis to Ma and Baba. For all their love and support, this thesis is as much theirs as it is mine.

Thank you.

Acknowledgements

I wish to thank my supervisor Dr. Martin J. Stillman for giving me the opportunity to work in the area of expert system applications in analytical chemistry. I thoroughly enjoyed working with him and appreciate his constant encouragement and support.

I would like to thank my advisory committee members, Dr. G. M. Bancroft and Dr. J. W. Lorimer. Special thanks to Dr. Lorimer for giving useful suggestions during the preparation of my research seminar.

I wish to acknowledge the financial support by the Ontario Ministry of Environment, a visa fee waiver by the Province of Ontario, and a fee bursary by the University of Western Ontario.

Thanks to all the friends I made during my stay here; my lab mates; the secretaries of the chemistry department, especially to Diana and Cheryl who I could approach without any hesitation; to Ron Maslen who made demonstrating third year analytical labs less stressful; to Leslie Tchorek who gave a lot of useful suggestions in my first year of demonstrating labs; to Marty and Sheri of chemistry stores, and to Glen, John, and Warren of the electronic shop.

There are no words that can express my gratitude for the support extended by my family. They have solidly stood behind me through the difficult times making life a lot easier. I am grateful to Anthony for being my best friend.

Table of Contents

Page

Certificate of Examination ii
Abstract iii
Dedication v
Acknowledgements vi
Table of Contents vii
List of Tables xi
List of Figures xii
Abbreviations xv
CHAPTER 1 INTRODUCTION
1.0 Introduction 1
1.1 Simulation of Human Thought Processes 2
1.2 Components in an Expert System 3
1.2.1 Comparison Between Conventional and Expert system Programming 3
1.2.2 Building an Expert System 5
1.3 Knowledge Acquisition 7
1.4 Knowledge Representation 8
1.4.1 Logic 8
1.4.2 Semantic Networks 8
1.4.3 Frames 11
1.4.4 Rule-based Systems 14
1.5 Inference Strategies 15
1.6 Applications of Expert Systems 20
1.6.1 Early Years of AI: DENDRAL and MYCIN 20
1.6.2 Expert Systems for Chemical Analysis (ESCA) 21
1.7 Automated Metal Analysis Using Atomic Absorption Expert (AAexpert) 22
1.7.1 The Atomic Absorption Spectrometer 22

1.7.2 Atomic Absorption Expert (AAexpert) 24
1.7.3 Structure of AAexpert 24
1 8 Scope of Thesis 26
1 9 References 27
CHAPTER 2 SOLUTION HANDLING AND QUALITY CONTROL OF ANALYTICAL DATA
2.0 Introduction 31
2.1 A Model for Totally Automated Metal Analyses 33
2.2 Hardware and Software 34
2 3 AAcontrol: A Process Control Module 37
2.4 A Model for AA-Quality Control: Characteristics of the Absorption Profile 40
2.5 A Flow Chart of AA-Quality Control 44
2.6 Calculation of Absorbances 44
2.7 Prediction of a Concentration 47
2.8 Fluctuations in the Absorption Profile 48
2.9 Implementation of AA-Quality Control 49
2.9.1 Testing that the Training Set Represents the Current Instrument Response 51
2.9.2 Effect of a Blocked Burner 54
2.9.3 Effect of a Viscous Solution 60
2.9.4 Effect of a Blocked Capillary Tube 60
2.9.5 Effect of a Long Capillary Tube 63
2.10 Automated Correction 63
2.11 Discussion 72
2.12 References 75
CHAPTER 3 DIAGNOSING CAUSES OF PROBLEM DATA
3.0 Introduction 77
3.1 Model for a Diagnostic Expert System 78
3.2 The Knowledge Domain 79

3.3 Choice of the Building Tool 84
3.3.1 Use of KDS: A Commercially Available Expert System Shell 84
3.3.2 Use of EAshell: A Shell Designed and Coded in our Laboratory 85
3.4 The User Interface 86
3.5 Rule Generation in Rule-based Expert Systems 87
3.5.1 Rule Generation in KDS Shell 88
3.5.2 Rule Generation in EAshell 90
3.6 Inference Strategies in AAdiagnosis based on KDS Shell 92
3.7 Inference Strategies in AAdiagnosis based on EAshell 93
3.8 Acquisition of Knowledge using KDS Shell 97
3.9 Encoding Chemical Knowledge using KDS Shell 106
3.10 Acquisition of Knowledge using Table_Generator of EAshell 109
3.11 Acquisition of Knowledge using Rule_Editor of EAshell 110
3.12 Encoding Chemical Knowledge using EAshell 112
3.13 Discussion 118
3.14 References 120
CHAPTER 4 METHOD SELECTION
4.0 Introduction 122
4.1 AAmethods: An Expert Database System for Method Selection 123
4.2 The Knowledge Domain in AAmethods 123
4.3 The User Interface in AAmethods 124
4.4 Use of Database in AAmethods 127
4.5 Use of Rulebase in AAmethods 129
4.5.1 Inference Strategy in AAmethods 134
4.6 Implementation of AAmethods 134
4.7 Discussion 138
4.8 References 140

CHAPTER 5 CONCLUSIONS
5.0 Encoding Chemical Knowledge 141
5.1 Summary 143
5.1 1 AA-Quality Control and AAcontrol 145
5.1.2 AAdiagnosis 146
5.1.3 AAmethods 146
5.2 Thoughts 147
VITA 48

List of Tables

Table	Description	Page
2.1	Error conditions and interpretations determined in AA-Quality Control	70
2.2	Error messages and corrections suggested by AA-Quality Control	71
3.1	List of symptoms and causes encoded in AAdiagnosis	80
3.2	Solutions offered by EAshell's forward chaining, depth-first search strategy and the interpretations	95
3.3	The knowledge matrix in AAdiagnosis using the KDS shell	102
3.4	The knowledge matrix in AAdiagnosis using EAshell	113
4.1	Knowledge representation in AAmethods (Microsoft Access environment)	128
5.1	The knowledge to be transformed into rules	142
5.2	Rules generated from the chemical knowledge present in the knowledge table	143

List of Figures

Figure	Description	Page
1.1	Components in an expert system	4
1.2	Components in an expert system shell	6
1.3	Representation of knowledge using a semantic network	10
1.4	An exception in inheritance in a semantic network	12
1.5	Representation of knowledge using frames	13
1.6	Forward chaining with depth-first inference strategy	18
1.7	Backward chaining with depth-first inference strategy	19
1.8	A block diagram of the Varian 875 atomic absorption spectrometer	23
1.9	The structure of AAexpert	25
2.1	Real-time quality control of the atomic absorption spectrometer	35
2.2	Communication: instrument to computer	36
2.3	The modules of AAcontrol	38
2.4	The user interface in AAcontrol	39
2.5	The components of an absorption profile	41
2.6	Flow chart to calculate the FFT of raw absorbance values	43
2.7	Flow chart to calculate criteria of data quality in AA-Quality Control	45
2.8	Flow chart of AA-Quality Control	46
2.9	Flow chart to calculate the number of peaks in the derivative plot of the absorption profile	50
2.10	Absorption profiles (raw, smooth) of a 4.0 ppm Cu^{2+} solution aspirated under different conditions	52
2.11	Derivative plots of a 4.0 ppm Cu^{2+} solution aspirated under different conditions	53
2.12	Step 1 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (no instrument faults)	55
2.13	Step 2 in the calculation of assessment criteria of a 4.0 ppm	

Cu^{2+} solution (no instrument faults) 56
2.14 A scanned photograph showing the burner head blockage after aspiration of a saturated NaCl solution 57
2.15 Step 1 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a blocked burner) 58
2.16 Step 2 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a blocked burner) 59
2.17 Step 1 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a viscous solution) 61
2.18 Step 2 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a viscous solution) 62
2.19 Step 1 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a blocked capillary tube) 64
2.20 Step 2 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a blocked capillary tube) 65
2.21 Step 1 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a long capillary tube) 66
2.22 Step 2 in the calculation of assessment criteria of a 4.0 ppm Cu^{2+} solution (effect of a long capillary tube) 67
2.23 Modes of operation of an automated solution handler 69
3.1 Logic developed at the primary level using the KDS shell 98
3.2 Development of logic in AAdiagnosis using the KDS shell 107
3.3 Encoding knowledge in AAdiagnosis using the KDS shell 108
3.4 Encoding knowledge in AAdiagnosis using EAshell 117
4.1 Representation of knowledge in AAmethods 125
4.2 The structure of the user interface in AAmethods 126
4.3 The user interface in AAmethods: analysis for cadmium in lake water 130
4.4 The user interface in AAmethods: use of the rulebase (cadmium in lake water) 131
4.5 The user interface in AAmethods: closest match method

of analysis (cadmium in lake water) 132
4.6 The user interface in AAmethods: analysis for nickel in drinking water 135
4.7 The user interface in AAmethods: use of the rulebase (nickel in drinking water) 136
4.8 The user interface in AAmethods: closest match method of analysis (nickel in drinking water) 137
4.9 The structure of AAmethods 139
5.1 Acquisition of chemical knowledge using the Table_Generator 142

Abbreviations

AAS	Atomic Absorption Spectrometry
AAexpert	Atomic Absorption expert
AA-QC	Atomic Absorption Quality Control
AI	Artificial Intelligence
DBMS	DataBase Management System
DLL	Dynamic Link Library
EAengine	Environmental Analytical engine
EAexpert	Environmental Analytical expert
EAshell	Environmental Analytical shell
EEC	European Economic Community
ESCA	Expert Systems for Chemical Analysis
FAAS	Flame AAS
FFT	Fast Fourier Transform
GFAAS	Graphite Furnace AAS
GPS	General Problem Solver
HGAAS	Hydride Generation AAS
KDS	Knowledge Delivery System
SQL	Structured Query Language

CHAPTER 1

INTRODUCTION

1.0 Introduction

Artificial intelligence (AI) has been described by Winograd in the following way (1):

> *There are two quite different starting points to define artificial intelligence - the dream and the technology. As a dream, there is a unified goal of duplicating human intelligence in its entirety. As a technology, there is a fairly coherent body of techniques (such as heuristic search and propositional representations) that distinguish the field from others in computer science .*

The research of Newell, Shaw and Simon on the Logic Theory Machine, a chess playing program, and the General Problem Solver (GPS) dominated the first decade of AI (from the mid 1950s to the mid 1960s). Their research in areas like heuristic search, problem solving, planning and knowledge representation are still considered important areas of AI (2).

The development of knowledge based systems began in 1965 when researchers (3) started work on DENDRAL, a system to interpret mass spectra; the findings from this work were published in 1970 (4) In 1973, researchers at Stanford University (5) reported their results for MYCIN, a knowledge based system that was used to treat bacterial infections in blood. In 1974, PROSPECTOR, an expert system to aid geologists in ore exploration was developed by the Stanford Research Institute (6). DENDRAL, MYCIN and PROSPECTOR are landmarks in the history of the application of artificial intelligence techniques aimed at solving complicated real world problems. One of the major projects in the application of expert systems in analytical chemistry was started in May 1987 by a group carrying out the Expert Systems for Chemical Analysis (ESCA) project funded by the EEC (7). By this time, expert systems had generated a great deal of interest among researchers from all

disciplines of chemistry. The work on Atomic Absorption expert (AAexpert) began at UWO in 1986 (8-10).

1.1 Simulation of Human Thought Processes

The General Problem Solver (GPS), coded in IPL-V (Information Processing Language, a list processing language) and run on an IBM 7090, was one of the first programs created to simulate human thought processes (11). In the GPS:

- the problem was expressed in an external representation,
- a translator converted the problem into an internal representation,
- problem solving techniques provided the solution by processing the internal representation.

The power of the GPS was based on the effectiveness of the problem solving techniques used and the generality was demonstrated by showing that the program could solve problems associated with different knowledge domains Problem solving was based on a heuristic search that was guided by a technique called "means-end analysis". In means-end analysis, the approach to analysis was based on the desired solution (end to be reached). The GPS consists of a group of methods capable of solving one or more problems. A "big switch" model of generality was proposed if there was a need for a diagnostic routine to relate a given problem with a particular method.

However, the concept of generality was found to give rise to systems that were inefficient in terms of performance. The early consultation systems DENDRAL and MYCIN which were developed in the 1970s, showed that knowledge based systems offered high levels of performance because the emphasis was on knowledge accumulation in a particular area. These programs were called expert systems and a new discipline of artificial intelligence was born. An expert system has been defined by Jackson (12) as.

> *A computer program that represents and reasons with knowledge of some specialist subject with a view to solving problems or giving advice.*

1.2 Components in an Expert System

The components of an expert system are represented in Figure 1.1. These components involve quite different design specifications for the developer of an expert system. The components are:

a. Domain Knowledge: contains the facts, rules, databases, and goals of the problem domain as described by the expert;
b. Knowledge Acquisition: chemical knowledge may be assembled by the developer as a matrix of facts in which rows represent the conditions or symptoms and the columns contain the underlying conclusions or causes;
c. Inference Engine: code that has problem solving ability and that can deduce results based on the input by the user and the knowledge contained in the knowledge base;
d. Explanation System: provides the line of reasoning associated with conclusions offered by the inference engine, and
e. User Interface: code that channels information between the inference engine and the expert system user, can be represented by menus, icons, images, sound, and video-clips. It displays solution (advice, conclusions, method) offered by the inference engine.

1.2.1 Comparison Between Conventional and Expert System Programming

Although computers can perform knowledge intensive tasks according to the decision making logic of conventional programs, it is difficult to encode heuristic human knowledge in procedural programs. Procedural programs consist of two parts an algorithm and a database. The algorithm determines how best to solve a specific kind of problem, while the database characterizes the parameters of the problem at hand. However, human expertise does not fit this model. Expert systems are capable of applying heuristic knowledge to problem solving and can comment on the solution or advice given to the user. It is important to remember that an expert system should generally be developed to solve complicated problems for which experts are in short supply and for which the domain of expertise is well focused.

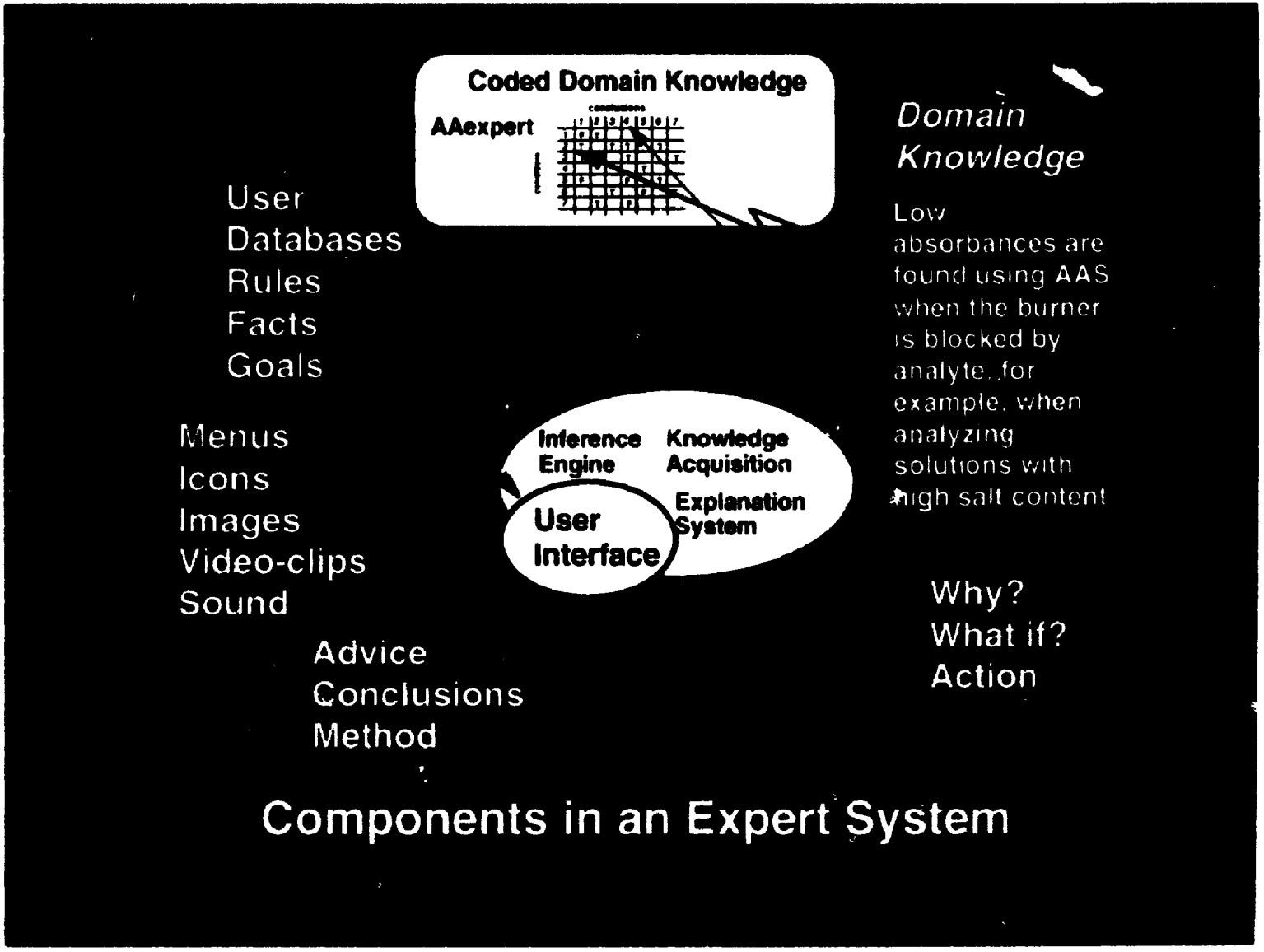


Figure 1 1 Components in an expert system

While it is possible to construct an expert system in a conventional language such as Pascal or C, the development of an extensive, unambiguous database, and the subsequent access to this knowledge base, is difficult. However, several expert systems used in the chemical laboratory have been written in procedural languages. For example, the expert system for organic synthesis (SYNLMA) uses a theorem prover as its inference engine which is based on Pascal routines (13). Expert system programs differ from conventional data processing programs in that extensive use is made of heuristics and inference mechanisms to represent and extract knowledge from a knowledge base. In expert systems, the separation of the knowledge base from both the inference engine and the user interface allows the knowledge base to be maintained and even expanded without changing the control structure. Settle and Pleva have discussed the expert system development tools available to the chemist (14).

Expert system shells can greatly reduce the task of developing a new expert system. An expert system shell contains an inference engine, various interface and knowledge acquisition aids, but lacks rules and facts in the knowledge base (15). A well described use of expert system shell is in the selection of HPLC conditions for the analysis of pharmaceuticals (16). We illustrate the structure of an expert system shell and its components in Figure 1.2.

1.2.2 Building an Expert System

The problem domain is the most important criterion in the development of expert systems. The crucial factors that determine whether an expert system will solve a particular problem depend on the nature of the problem, the availability of human expertise, and the possibility of analysing the expertise and the problem in such a way that the knowledge can be coded into an expert system (12). Full scale development of an expert system should be preceded by development of a prototype that can be used to indicate the appropriateness of using an expert system to solve the problem (17).

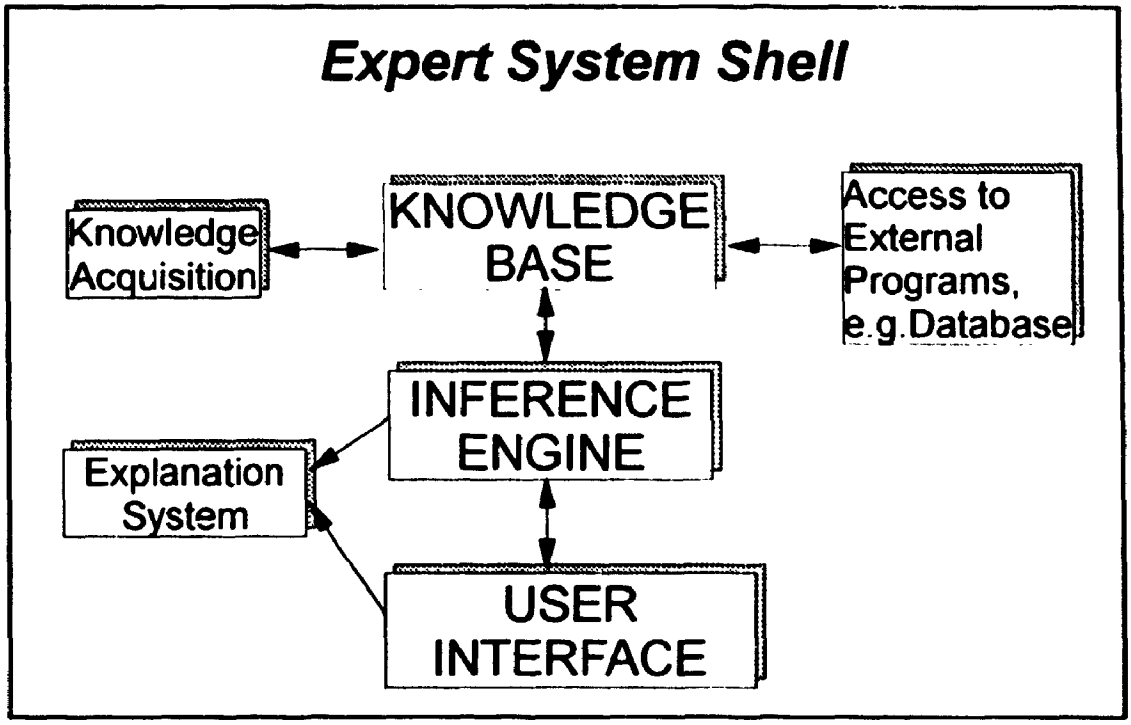


Figure 1.2 Components in an expert system shell.

1.3 Knowledge Acquisition

The identification of the problem domain is followed by knowledge acquisition, this step is often referred to as "the bottleneck problem" in the generation of expert systems. Knowledge acquisition has been defined as the transfer and transformation of potential problem solving expertise from some knowledge source to a program (12).

Approaches to knowledge acquisition can be grouped as follows (18):

a. Traditional knowledge engineering.
b. Interactive knowledge acquisition tools.
c. Machine learning.

In traditional knowledge engineering, the transfer of knowledge involves classification of a domain expert's information by a knowledge engineer. The information that is extracted is then transferred to a form usable by the computer. This approach has also been referred to as the basic model of knowledge engineering. The major problem associated with this basic model is the assumption that knowledge engineers are available who can successfully transfer the specialized knowledge (19).

The interactive model of knowledge acquisition involves the expert directly transferring knowledge into a knowledge base and the knowledge engineer then collaborating with the expert to validate the knowledge base (19). Often, the knowledge acquisition tools require the knowledge to be represented in a form that is difficult to extract from the expert. For example, when the knowledge acquisition tool requires the expert to provide weighted links associating data with hypotheses, the expert may find it hard to supply the numbers (18). Repertory grid centred tools for knowledge acquisition have been used to build knowledge based systems (20). In a repertory grid, also known as rating grid, solutions, referred to as elements, are placed in the column and the traits, referred to as constructs, are placed as rows of the grid. These tools interact with the expert and perform tasks such as interviewing the expert, testing and refining the resultant knowledge base. AQUINAS, an expanded version of the

Expertise Transfer System (ETS) can automatically generate production rules from rating grids, which can be reformatted for use in various expert system building tools

The third approach of knowledge acquisition, machine learning, involves production of rules from examples. An application of machine learning requires setting up an induction algorithm to transform examples into rules (18). Quinlan (21) developed an inductive algorithm (the ID3 algorithm) which has been applied in commercially available expert system shells, e.g. the KDS (Knowledge Delivery System) shell.

1.4 Knowledge Representation

Knowledge representation can be grouped into four categories (22):

a. Logic;
b. Semantic Networks;
c. Frames, and
d Rule-based Systems.

1.4.1 Logic

First order logic (FOL) expresses features of deductive reasoning as propositions. Propositions are defined as statements that are either TRUE or FALSE. Complex expressions are generated by using a set of symbols called connectives. Commonly used connectives are AND, OR, NOT and IMPLIES.

The drawback of representing knowledge using FOL is that propositions in the real world cannot always be expressed as TRUE or FALSE. Multi-valued logic like fuzzy logic has been developed to express uncertainty (23).

1.4.2 Semantic Networks

Quillian was the first to apply semantic network ideas to the field of natural language translation and understanding. He proposed an associational model of human memory, called semantic memory, which attempted to capture the meanings of words (similar to capabilities of human memory) and implement it in a computer. Winston's idea of structured descriptions was based on

generalizations from examples, similar to representation of human thought processes. The problem of using semantic networks to represent knowledge arises from the fact that definitions of meanings, also known as concepts, are subjective and difficult to incorporate into an program (24).

A semantic network is defined as a labelled directed graph that consists of vertices and labelled arcs between vertices. Each vertex represents a concept, also known as word meanings. The arcs represent binary relations between concepts. Relations commonly used in a semantic network are as follows:

a. A "part-of" relation: a relation in which the first concept is a "part of" the second concept. For example, halogens are a "part of" the Periodic Table, and
b. An "is-a" relation: there are two types of is-a relation between concepts. These are as follows:

i. Set inclusion relation: a relation in which a concept "is a" subclass of another concept. For example, a transition metal "is a" metal, and
ii. Membership relation: a relation in which a concept "is a" member of a certain class of objects. For example, copper "is a" transition metal.

Figure 1.3 shows the representation of knowledge regarding the periodic table of elements in a semantic network.

The subset-of and member-of links can be used to derive new information and may form the basis for an inference engine. The use of links in a reasoning mechanism called inheritance is explained in the following example.

Example: The statements represented in Figure 1.3 are:

a. Copper is a transition metal;
b. Transition metals are metals, and
c. Metals are elements of the Periodic Table.

From the above statements, the following statement can be derived.

Transition metals are elements of the Periodic Table.

The following statement can be derived using subset-of and member-of links.

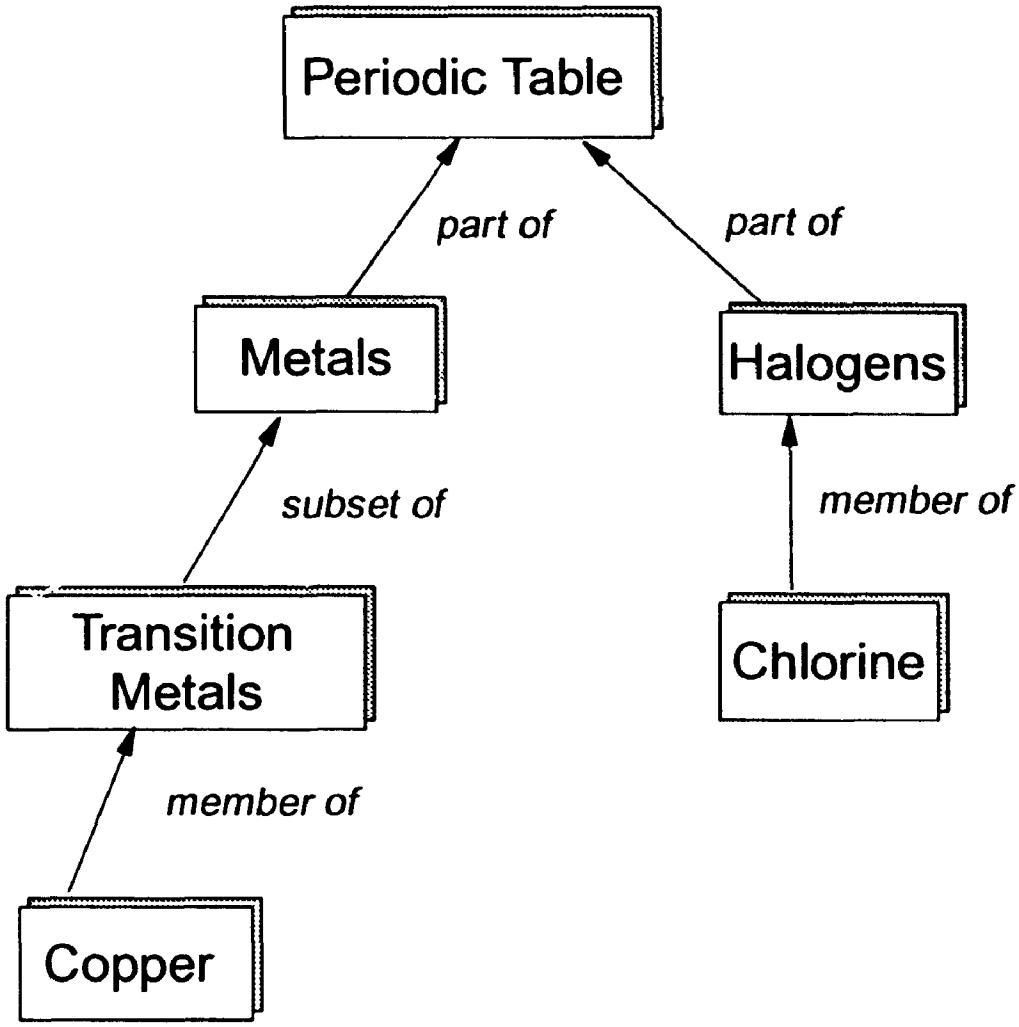


Figure 1.3 Representation of knowledge using a semantic network.

Copper is a metal.

Therefore, in an inheritance mechanism, the concept inherits properties of concepts higher in the semantic network through links. However, the hierarchical representation gives rise to problems in knowledge representation. Figure 1.4 depicts the problems associated with inheritance. The network shows that metals are solids and conduct electricity. Therefore, mercury, a metal, should inherit both properties of metals. Although mercury conducts electricity, it is not a solid. Flexibility has been introduced with regards to the inheritance of properties, by using knowledge representation in terms of frames (25).

1.4.3 Frames

In a frame-based representation, knowledge relevant to a concept is stored in entities called frames. A frame has been defined as a network of nodes and relations organised in a hierarchy, where the topmost nodes represent general concepts and the lower nodes represent more specific instances of the concepts (26). This mode of knowledge representation is known as a frame hierarchy or frame taxonomy in which frames are represented by vertices and arcs denote "is-a" links between two frames.

Class frames, also known as generic frames, represent knowledge concerning classes of objects. Knowledge concerning individual objects is represented by instance frames. A frame indicates its relative position in a taxonomy by using two types of is-a link. These links are:

a. An instance-of link: a link between an instance frame and a class frame, and
b. A superclass link: a link between two class frames.

Figure 1.5 shows an example of frame taxonomy. Metals represent a superclass and copper is a specialisation, a transition metal. Representation of knowledge in a frame takes the form:

instance *copper* is
instance-of *transition metals*;
group = I B;
atomic number = 29;

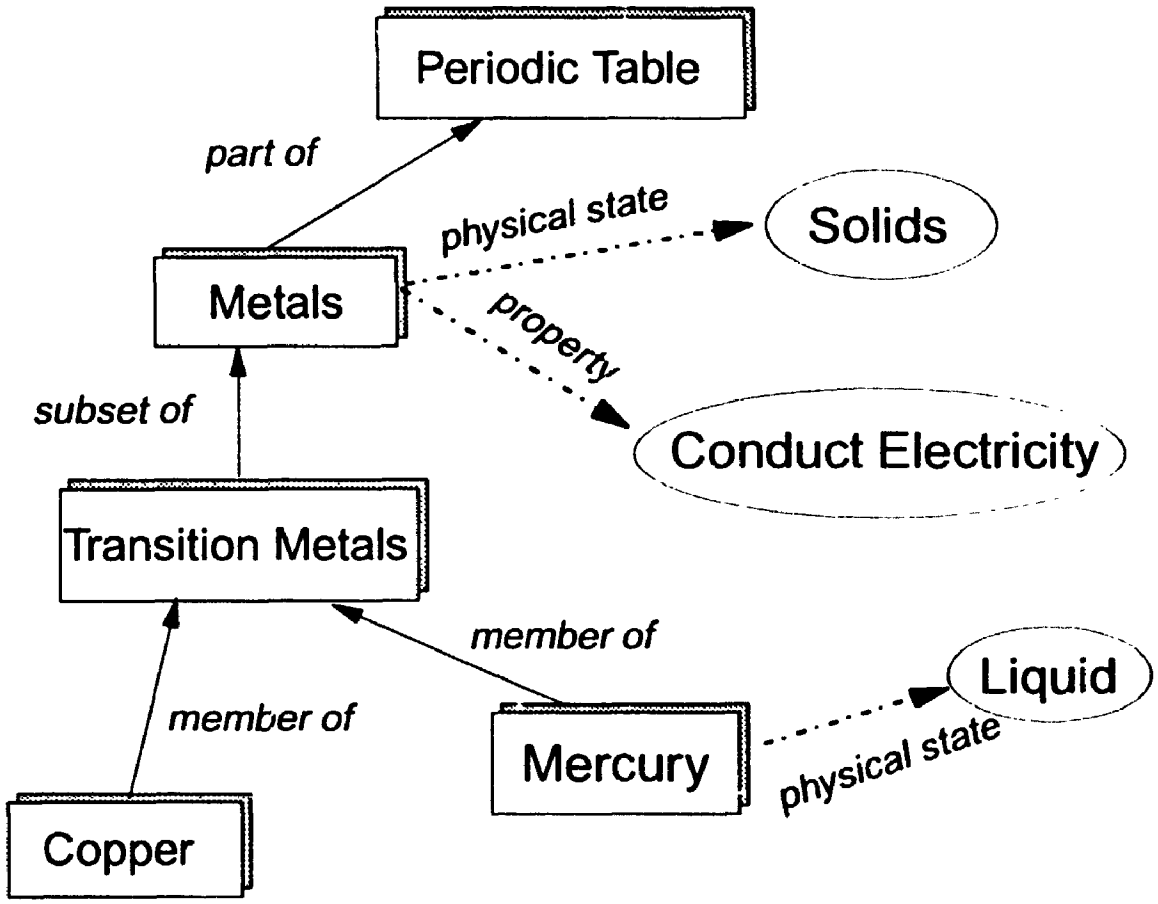


Figure 1.4 An exception in inheritance in a semantic network.

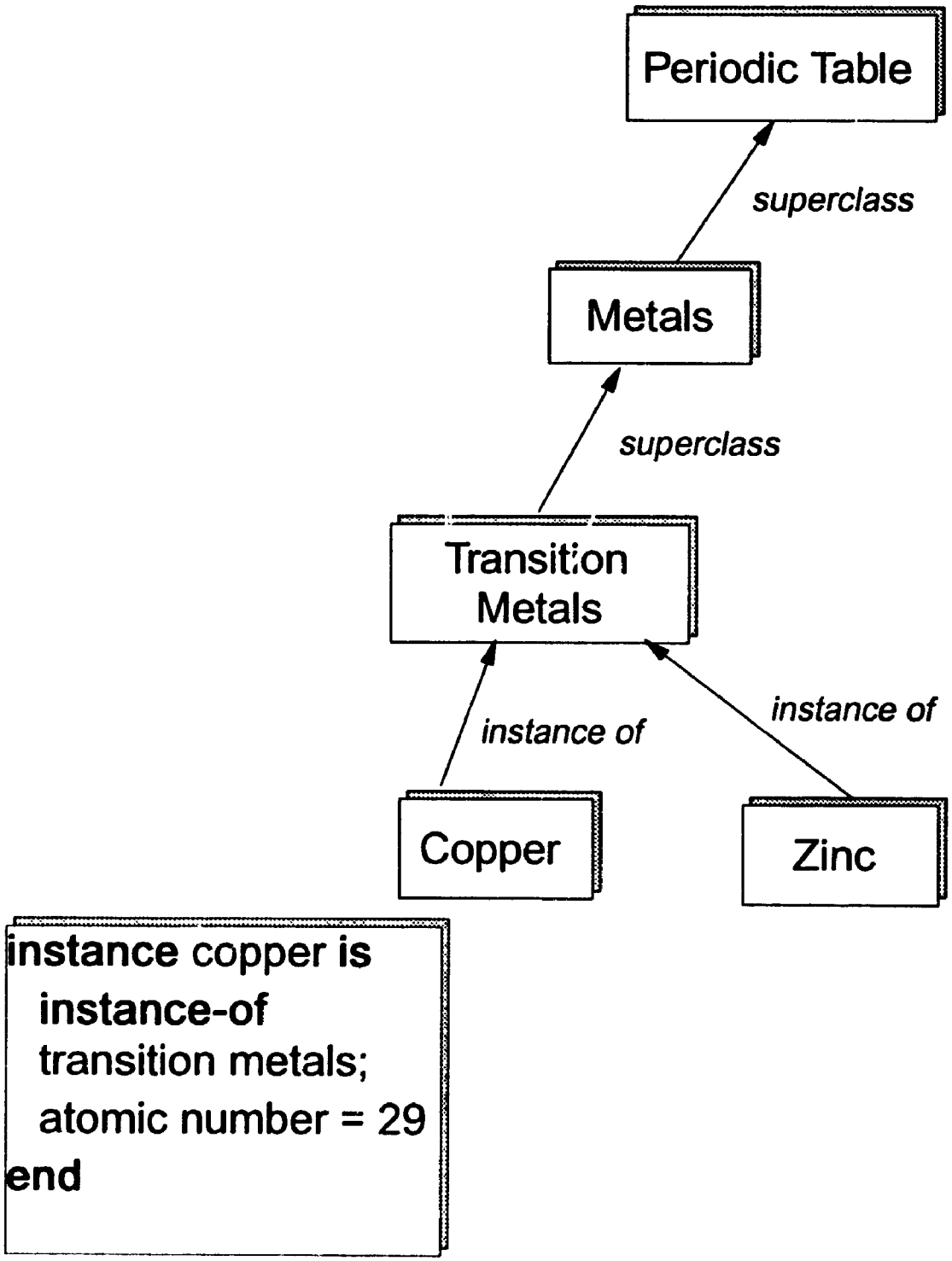


Figure 1.5 Representation of knowledge using frames (tree-like taxonomy).

physical state = *solid*;
property = *conducts electricity*
end

Procedures present in frames are called demons that can be activated at a particular time during the manipulation of the frame. Commonly used demons are *if-needed*, *if-added*, and *if-removed*. Demons can also be used in knowledge based systems that use frames and production rules. Knowledge in CENTAUR, a LISP-based expert system used to assist in the treatment of pulmonary diseases, is represented in frames and production rules. It has been suggested that descriptive knowledge can be successfully represented in a frame-based system (25).

1.4.4 Rule-based Systems

The concept of rules to represent knowledge was introduced by researchers of the DENDRAL project (3). Rules represent knowledge in the form of IF-THEN statements. The IF part contains the *premise* of the rule and the THEN part contains the *action* or *conclusion* of the rule.

Example: IF pH of a solution = 3.0 THEN the solution is acidic.

The premise of a rule is a Boolean expression that must be satisfied for the rule to be executed or *fired*. The conclusion of a rule can either be a list of commands to be carried out when the rule fires or be evaluated to true when the premise does. The set of rules that describe a particular knowledge domain is referred to as the rule base. In an expert system, the rule base along with facts associated with the knowledge domain forms the knowledge base (27).

Rules that are grouped in sub-areas of the problem domain are referred to as well-written rules. Well-written rules are said to be transparent in that the developer of the rule base is able to see through the syntax to the meaning. It is easy to modify knowledge in such a rule-based system because one part of the rule base can be changed without affecting the other parts of the rule base. Well written rules should have the following features:

a. Organization of rules: the maintainability of a rule based system improves if rules that have the same conclusion are grouped together;
b. Ordering of rules: higher performance is achieved for an expert system when in a particular group, rules are placed in order of most likely to least likely, and
c. Sequence of rules in a rule base: the order of rules in a given group should be based on the primary inference strategy used by the inference engine (28).

The drawback of a rule-based system is that it cannot represent structural knowledge. In a rule-based representation, it is not possible to represent knowledge describing a particular entity in the form of clusters which are the characteristic feature of frame-based systems (27).

Commercially available expert system shells help developers unfamiliar with AI techniques to build expert systems. However, each shell comes with a particular knowledge representation scheme which may be suited for only a particular type of application. While the idea of a standard knowledge representation scheme is appealing to a user, it is discouraged by vendors. The possible ways in which a particular knowledge base can be transformed from one representation to another, in particular, transformation between frames and rules has been discussed by Thuraisingham (27). The author suggests that two methods of knowledge transformation are possible, namely

a. Direct translation between different knowledge representation schemes involving the use of a translator, and
b. An intermediate representation, generic representation scheme, in which a given knowledge representation is translated into an intermediate representation; the intermediate representation is transformed into the second knowledge representation scheme.

1.5 Inference Strategies

Inference strategies in an expert system are based on either forward chaining or backward chaining. In a forward chaining system, data are placed in

the dynamic database and the inference engine reports on any conclusions it can make by taking the premise (the ***if*** part) of an ***if ... then*** rule and inferring that the conclusion or action (the ***then*** part) of the rule is true. A forward chaining system is said to be data driven. In a backward chaining system, the inference engine works backwards through the conclusions of rules trying to determine whether the subject of a rule is true for a particular goal Backward chaining systems are said to be goal driven.

The forward chaining or backward chaining system can be combined with either a depth-first or a breadth-first search. Depth-first searches require a search through all levels of a branch of a tree until a conclusion or a contradiction is reached. In a depth-first search the inference engine continues to pass from level to level based on a single affirmative answer. Breadth-first searches require that at each level all the branches must be searched thoroughly, followed by a new search of the next level of only the positive responses. This process is repeated until a conclusion is reached.

A depth-first approach can be much faster than a breadth-first approach. However, if the system gets stuck while going along a path on which there is no solution, it will never find a solution since the search strategy involves following that path as far as possible from level to level, and not across each level. It is possible to combine a depth-first search with other heuristic methods for choosing which path to explore, and which rules to use to solve subgoals. This is an effective method because solutions are obtained that the users want solved without exhausting the system's resources. The breadth-first search is a complete strategy: if the problem posed by the user has a solution, a breadth-first search will find it unless the system runs out of resources (8, 29). A significant advantage of the breadth-first search is that the user can be given a very complete description of the rules that led to a particular conclusion being reached. This can allow the user to select alternative conclusions based on data not available to the inference engine.

EAshell, an expert system shell developed and coded in our laboratory, uses rule-based knowledge representation. In the rule base, similar rules are grouped together giving rise to major areas of knowledge. In a given sub-group of rules, the rules are arranged in an order of most likely to least. *Necessary facts* are information or data required to fire any rule. *Necessary facts* may be present in the knowledge base, when *necessary facts* are not present in the knowledge base the inference engine requests the user to supply the information referred to as the *missing facts*. Facts available in the knowledge base are known as *available facts*, these facts may or may not be *necessary facts*. The inference strategies used by EAshell include:

a. Forward chaining with depth-first search, and
b. Backward chaining with depth-first search.

Figure 1.6 depicts forward chaining with depth-first search strategy in EAshell. The inference engine gathers information in order to satisfy the premise of a rule so that it can fire a rule. The depth-first search implies that the search moves from one level of knowledge to another, or from one rule to another, only after all the facts in this level have been tested. The search begins with level 4, and concludes at level 1 of the knowledge base with the inference engine offering the goal as the solution to the problems indicated by the user. The *missing facts* in levels 4, 2, and 1 are *necessary facts* and therefore in order to reach the goal or conclusion must be supplied by the user.

Figure 1.7 shows the backward chaining with depth-first search strategy in EAshell. The goal or the problem is selected by the user. In order to satisfy the goal, the inference engine requires the user to supply the *missing facts* which are *necessary facts*. The depth first search strategy requires the knowledge base to have a mechanism to avoid a dead-end situation. A possible solution is incorporation of rules that prompt the user to verify the responses that were supplied as *missing facts*.

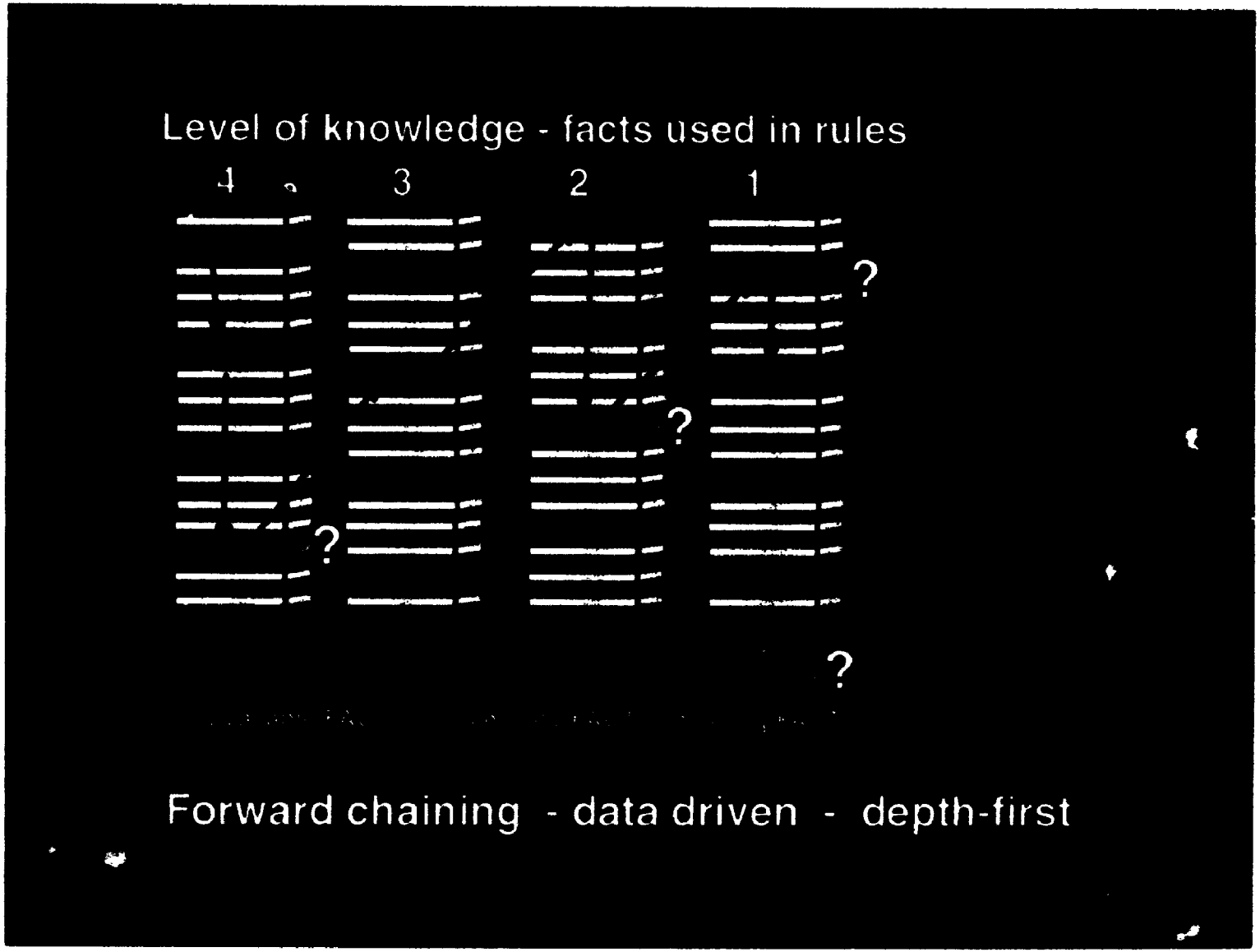


Figure 1 6 Forward chaining with depth-first inference strategy.

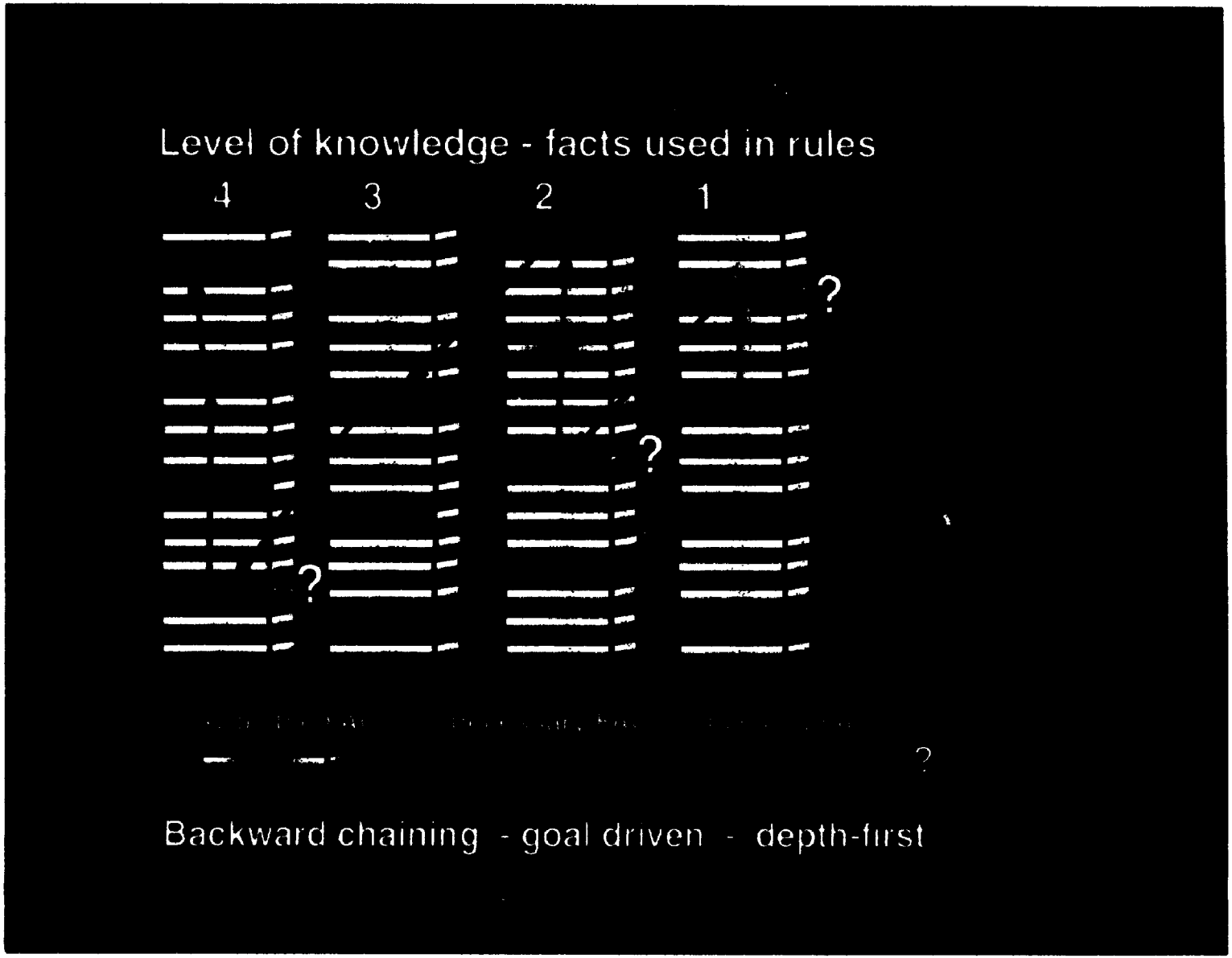


Figure 1.7 Backward chaining with depth-first inference strategy.

1.6 Applications of Expert Systems

Applications of expert systems to problems in analytical chemistry include (i) diagnosis, in which the cause of chemical problems are inferred from a list of symptoms, (ii) interpretation, in which a possible chemical structure is inferred from spectroscopic data, (iii) planning, in which a method of analysis is designed, and (iv) instruction, in which operators and technicians are taught how to correctly operate equipment.

1.6.1 Early Years of AI: DENDRAL and MYCIN

The DENDRAL project has been hailed as the first major application of heuristic programming in an empirical science. DENDRAL originally stood for DENDRitic ALgorithm, a program that searched for topologically distinct arrangements of atoms based on chemical valency. The aim of DENDRAL was to suggest likely structures based on mass spectra. *Knowledge engineering*, the process of transferring knowledge from an expert in the problem domain to a programmer, was used extensively in the project. Prof. Djerassi's group provided the chemical information about fragmentation patterns in mass spectrometry. The researchers of this project introduced the concept of *knowledge is power*, a program whose knowledge belongs to a well-defined narrow domain, and reported that the separation of knowledge of the other components of the computer program made knowledge maintenance an easier task. Knowledge was represented in the form of IF/THEN statements. The DENDRAL project consists of several programs, these include CONGEN, generation of connectivity isomers, PREDICTOR, likely structures, and MSPRUNE, generation of structures consistent with a spectrum. Knowledge acquisition and representation were important to the developers of the DENDRAL project (30).

The major contribution of the MYCIN project was the understanding of the use of production rules to represent knowledge. MYCIN, an expert system to provide advice on treatment of bacterial blood infections, was written in INTERLISP. The program consisted of 200 production rules. The premise of

these production rules consist of conjunctions of clauses, but contain conjunctions or disjunctions nested within each clause. Rules that could be written as disjunctions of clauses were written as separate rules. The inference strategy used involved backward chaining together with a depth-first search. The researchers suggested that a premise containing more than six clauses was not practical since it was unlikely that a large number of clauses would interact and result in one action (31).

1.6.2 Expert Systems for Chemical Analysis (ESCA)

This project has made significant contributions in the field of expert systems in the area of method development in liquid chromatography. The salient features of this project can be summarized as follows (7):

a. Commercially available expert system shells were chosen in order to reduce the time taken for generating the expert systems. The tools were selected after the researchers tested the suitability of the software on a test knowledge base;
b. The area of method development was sub-divided into four areas, method selection and retention optimization, selectivity optimization, system optimization, and method validation;
c. The project comprised a number of stand-alone expert systems, each of which performed a specific task in the method development process, and
d. Validation and evaluation of the different expert systems was considered to be a very important aspect of the project.

Specific examples of expert systems that are used in analytical chemistry include sample preparation by microwave digestion (32, 33), analysis of environmentally hazardous mixtures (34-36), identification of compounds from mass spectra (37-40), detection of petroleum based volatiles for suspected arson samples (41), voltammetric determination of metals (42, 43), solvent extraction of rare-earths (44), structural interpretation using nuclear magnetic resonance spectroscopy, and infra red spectroscopy (45, 46).

1.7 Automated Metal Analysis using Atomic Absorption Expert (AAexpert)

1.7.1 The Atomic Absorption Spectrometer

Atomic absorption spectroscopy is based on absorption of radiation of a fixed energy by metal atoms of interest. The block diagram of the Varian 875 AAS is shown in Figure 1.8.

Formation of metal atoms in the optical path occurs in the sequence (a) aspiration, (b) nebulization - the solution strikes a glass bead resulting in the formation of uniform size droplets (mist), and (c) mixing of the mist with the gas mixture (air-acetylene or air-nitrous oxide). This mixing, along with the heat generated at the burner, gives rise to gas phase atoms.

The hollow cathode lamp has a cathode surface made of the metal whose absorption measurements are desired. The lamp is filled with an inert gas (Ne or Ar) that aids in the formation of sputtered metal atoms from the cathode Some of the sputtered atoms are in the excited state and emit radiation characteristic of the metal on the surface of the cathode.

The Varian 875 atomic absorption spectrometer has a double beam design in which the beam from the hollow cathode lamp is split, one half going through the flame (the sample beam), while the other half (the reference beam) passes around the flame. The sample beam is focused on the flame and a part of the radiant energy is absorbed by the analyte atoms present in the flame The radiation leaving the flame and the radiation from the reference beam are recombined by a rotating chopper mirror and the resulting beam passes through a monochromator. Light emerging from the monochromator slit is directed onto the photomultiplier tube where it is converted into electric current. The signal is amplified and sent to the processor for internal calculations. Inputs synchronized from the chopper are accepted by the processor. The signals received from the reference beam are used to calculate the hollow cathode lamp drift while the signals received from the sample beam are used to calculate the absorbance values.

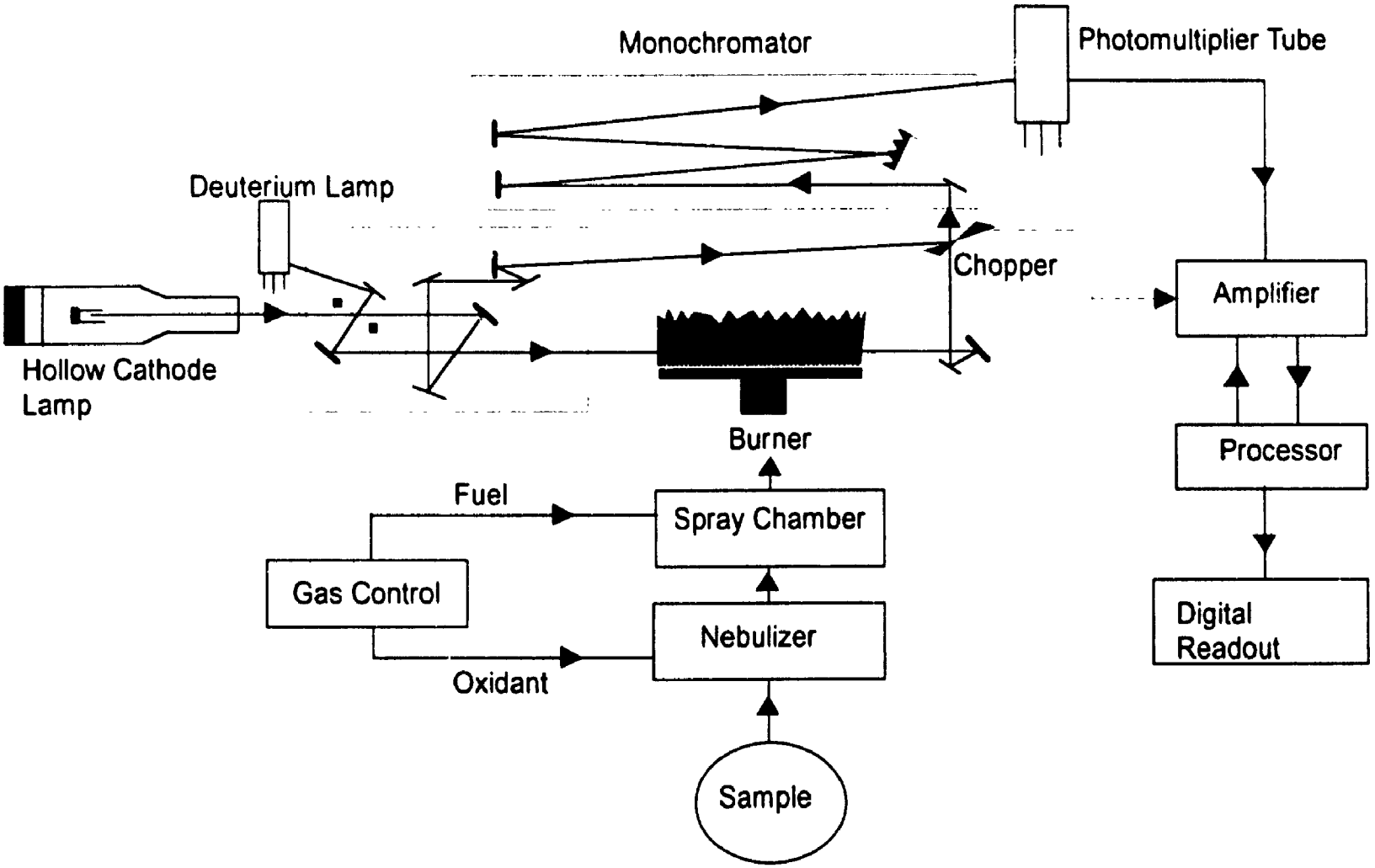


Figure 1.8 A block diagram of the Varian 875 atomic absorption spectrometer.

1.7.2 Atomic Absorption Expert (AAexpert)

The AAexpert project is part of a large, multicomponent expert system named ACexpert, a system which is concerned with all aspects of instrumental analysis (8-10). In its present design, ACexpert acts as an adviser in selecting appropriate analytical procedures. The system is subdivided into a set of individual expert systems designed to perform specific tasks. The structure of AAexpert is shown in Figure 1.9.

The expert system AAexpert has been designed so that each module addresses a single, limited domain of expertise within the overall domain of analysis by flame atomic absorption spectrometry. The individual areas identified include method selection (AAmethods), control of the solution handling and measurement steps (AAcontrol), real-time assessment of analytical data (AA-Quality Control), and diagnosis of errors due to instrumental malfunction and interfering sample chemistry (AAdiagnosis). These four examples each involve different applications of expert system technology.

1.7.3 Structure of AAexpert

On the receipt of the sample the MANAGER's task is to consult with both the CUSTOMER and the REGULATORY AGENCY to determine the criteria to be used in the analysis. AAassurance is the quality assurance expert system that is used by the MANAGER and the ANALYST to assist in the execution of a laboratory quality assurance program. The ANALYST's role is supervision of the expert system. AAanalyst is the process control and quality control expert system. AAanalyst, will complete the required analyses using the modules: AAmethods for method selection, AAcontrol for sample scheduling and handling, AAdiagnosis for diagnosing faults associated with both instrumental operation and quality of data, and AC-QC for quality control. AAteach is the instruction expert system which gives advice and examples on how to operate the instrument.

AAexpert

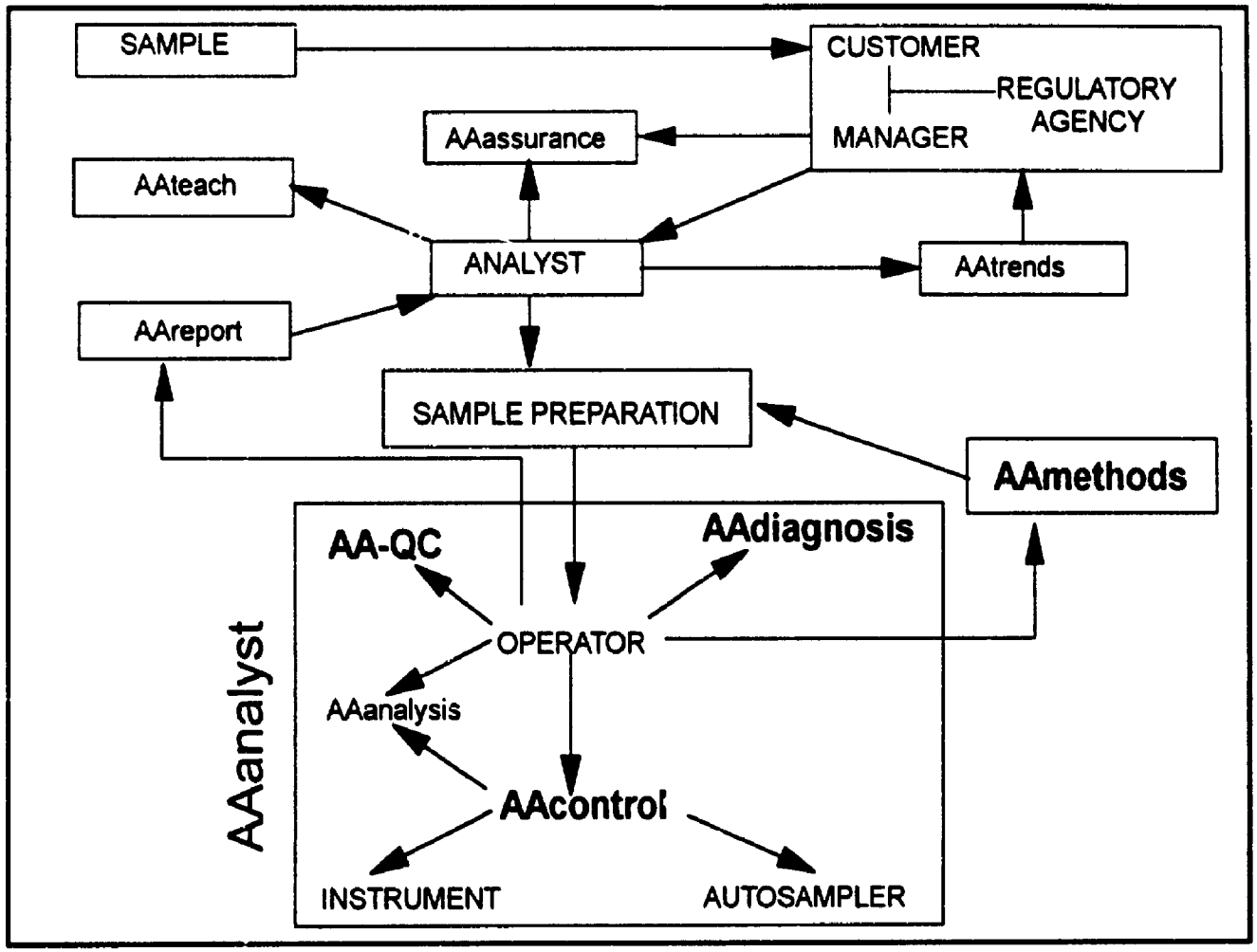


Figure 1.9 The structure of AAexpert.
This expert system provides assistance to the user in the analysis for metals by flame atomic absorption spectrometry.

1.8 Scope of Thesis

The work in this thesis is aimed at understanding the process of encoding chemical knowledge in three knowledge based expert systems. Transfer of human knowledge into a computer requires (a) acquisition of the knowledge, and (b) representation of the acquired knowledge. The emphasis of the work is on encoding chemical knowledge to build expert systems. Rule based representation was chosen to describe the chemical knowledge. This is the first time that such a study has been undertaken here, therefore this work represents much of the learning curve necessary for the generation of future expert systems in Dr. Stillman's laboratory. The knowledge based systems studied in this thesis are outlined in the following chapters:

- **Solution Handling and Quality Control of Analytical Data** AAcontrol and AA-Quality Control (AA-QC) are modules involved in measuring real-time data from the atomic absorption spectrometer and providing an assessment of the quality of the data. The focus of this chapter is on whether the detector response can be modelled in such a way that AA-QC can detect simple faults associated with analysis for metals by flame atomic absorption spectrometry (FAAS).
- **Diagnosing Causes of Problem Data** Knowledge based systems concerned with problems associated with data of poor quality and instrument malfunction are discussed. A method of encoding chemical knowledge to assist chemists in generating expert systems is described.
- **Method Selection** The steps involved in generating an expert database system are reported.
- **Conclusions**.

1.9 References

1. Winograd, T., in Bobrow, D. G., Hayes, P. J., (Eds.), *Artif. Intell.*, 1985, **25**, 375.
2. Klahr, P., Waterman, D. A., *Expert Systems Techniques, Tools & Applications*, Addison-Wesley Publishing Co., Reading, Massachusetts, 1986, ch. 1, pp 3-23.
3. Feigenbaum, E. A., Buchanan, B. G., *Artif. Intell.*, 1993, **59**, 233.
4. Buchs, A., Delfino, A. B., Duffield, A. M., Djerassi, C., Buchanan, B. G., *Helv Chim. Acta*, 1970, **53**, 1394.
5. Shortliffe, E. H., Axline, S. G., Buchanan, B. G., Merigan, T. C., Cohen, S. N., *Comput. Biomed. Res.*, 1973, **6**, 544.
6. Waterman, D., *A Guide to Expert systems*, Addison-Wesley Publishing Company, Reading, Masachusetts, 1986, ch. 6, pp 49-60.
7. Buydens, L., Schoenmakers, P., Maris, F., Hindriks, H., *Anal. Chim. Acta*, 1993, **272**, 41 and references cited therein.
8. W. R. Browett, and M. J. Stillman, *Prog. Analyt. Spectrosc.*, 1989, **12**, 73.
9. Browett, W. R., Cox, T. A., Stillman, M. J., in *Expert System Applications in Chemistry*, Hohne, B. A., Pierce, T. H., Eds., American Chemical Society, Washington, DC, 1989, ch. 17, pp 210-234.
10. Lahiri, S., Stillman, M. J., *Anal. Chem.*, 1992, **64**, 283A.
11. Ernst, G. W., Newell, A in *GPS: A Case Study in Generality and Problem Solving*, Academic Press, New York, 1969, pp 279-293.
12. Jackson, P., *Introduction to Expert Systems* (2nd Edition) Addison-Wesley Publishing Co., Reading, Massachusetts 1990, ch. 1, pp 3-14.
13. P. Y. Johnson, I. Burnstein, J. Crary, M. Evans, and T. Wang in *Expert System Applications in Chemistry*, B. A. Hohne, T. H. Pierce, Eds., American Chemical Society, Washington, 1989, ch 9 pp 102-123.

14. F. A. Settle, Jr. and M. A. Pleva, *Chemometrics and Intelligent Laboratory Systems*, 1991, **11**, 13.

15. Van Leeuwen, J. A., Buydens, L., Vandeginste, B. G. M., Kateman, G., *Trends Anal. Chem.*, 1990, **9**, 49.

16. Hindriks, R., Maris, F., Vink, J., Peeters. A., De Smet, M., Massart, D. L., Buydens, L., *J. Chromatogr.*, 1989, **485**, 255.

17. Waterman, D., *A Guide to Expert systems*, Addison-Wesley Publishing Company, Reading, Masachusetts, 1986, ch. 12, pp 135-141.

18. Gruber, T. R., in *The Foundations in Knowledge Acquisition*, Ed. Boose, J. H., Gaines, B. R., Academic Press, San Diego, USA, 1990, pp 115-133.

19. Gaines, B. R., in *The Foundations in Knowledge Acquisition*, Ed. Boose, J. H., Gaines, B. R., Academic Press, San Diego, USA, 1990, pp 1-21.

20. Boose, J. H., in *The Foundations in Knowledge Acquisition*, Ed. Boose, J. H., Gaines, B. R., Academic Press, San Diego, USA, 1990, pp 61-84.

21. Quinlan, R. J. in *Machine Learning-An Artificial Intelligence Approach*, Eds. Michalski, R. S., Carbonell, J. G., Mitchell, T.M., Tioga Publishing Co., Palo Alto, California, 1983, ch.15, pp 463-482.

22. Duce, D., Ringland, G., in *Approaches to Knowledge Representation-an Introduction*, Ringland, G. A., Duce, D. A., Eds., Research Studies Press Ltd., Letchworth, Hertfordshire, England, 1988, ch 1, pp 1-11.

23. Pavelin, C., in *Approaches to Knowledge Representation-an Introduction*, Ringland, G. A., Duce, D. A., Eds., Research Studies Press Ltd., Letchworth, Hertfordshire, England, 1988, ch. 2, pp13-43.

24. Mac Randal, D., in *Approaches to Knowledge Representation-an Introduction*, Ringland, G. A., Duce, D. A., Eds., Research Studies Press Ltd., Letchworth, Hertfordshire, England, 1988, ch. 3, 45-79.

25. Lucas, P., Van der Gaag, L., *Principles of Expert Systems* (2nd Edition) Addison-Wesley Publishing Co., Wokingham, England, 1991, ch. 4, pp 173-252.

26. Waterman, D. A., in *A Guide to Expert Systems*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1986, ch. 7, pp 63-79.

27. Thuraisingham, B., *AI Expert*, October 1989, 30.

28. Pederson, K., in *Expert Systems Programming: Practical Techniques for Rule-Based Systems*, John Wiley & Sons, New York, 1989.

29. L. Henschen in *Encyclopedia of Artificial Intelligence (Vol 2)*, S. C. Shapiro and D. Eckroth, Eds., John Wiley & Sons Inc., New York, 1987, 822.

30. Lindsay, R. K., Buchanan. B. G., Feigenbaum, E. A., Lederberg, J., *Artif. Intell.,* 1993, **61**, 209.

31. Davis, R., Buchanan, B. G., Shortliffe, E., *Artif. Intell.*, 1977, **8**, 15.

32. Settle, F. A. Jr., Diamondstone, B. I., Kingston, H. M., Pleva, M. A., *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 11.

33. Settle, F. A. Jr., Walter, P. J., Kingston, H. M., Pleva, M. A., Snider, T., Boute, W., *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 349.

34. Wytoff, B. J., Tomellini, S. A., *Anal. Chim. Acta*, 1989, **227**, 51.

35. Ying, L., Levine, S. P., Tomellini, S. A., Lowry, S. R., *Anal. Chim. Acta*, 1988, **210**, 51.

36. Ying, L., Levine, S. P., Tomellini, S. A., Lowry, S. R., *Anal. Chem.*, 1987, **59**, 2197.

37. Sarkar, M., Glen, W. G., Yin, L., Dunn III, W. J., Scott, D. R., Swanson, S., *Anal. Chim. Acta*, 1992, **257**, 229.

38. Scott, D. R., *Anal. Chim. Acta*, 1992, **265**, 43.

39 Scott, D. R., *Anal. Chim. Acta*, 1991, **246**, 391.

40. Scott, D. R., *Anal. Chim. Acta*, 1989, **223**, 105.

41. Holzer, G., Bertsch, W., Zhang, Q. W., *Anal. Chim. Acta*, 1992, **259**, 225.

42. Esteban, M., Ruisanchez, I., Larrechi, M. S., Rius, F. X., *Anal. Chim. Acta*, 1992, **268**, 95.

43. Esteban, M., Ruisanchez, I., Larrechi, M. S., Rius, F. X., *Anal. Chim. Acta*, 1992, **268**, 107.

44. Xu, L., Xiao, Y., Li, D., *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 437.

45. T. J. Harner, G. C. Levy, E. J. Dudewicz, F. Delaglio, and A. Kumar in *AI Applications in Chemistry*, T. H. Pierce, and B. A. Hohne, Eds., American Chemical Society, Washington, 1986, ch. 26, 337-349.

46. H. J. Luinge, G. J. Kleywegt, H. A. Van't Klooster, and J. H. Van Der Maas, *J. Chem. Inf. Comput. Sci.*, 1987, **27**, 95.

CHAPTER 2

SOLUTION HANDLING AND QUALITY CONTROL OF ANALYTICAL DATA

2.0 Introduction

Over the last two decades laboratory automation has received a great deal of attention by analytical chemists (1-5). The need for higher productivity including reducing overall costs per sample and improving quality control for repetitive tasks has prompted research that will lead to development of totally automated, unattended instrument analysis. Before an instrument can carry out unattended, automated analyses, the controlling system must be able to interrupt or modify the operation in real-time to correct instrument malfunctions and to change sample measurement procedures. The capability of an automated system to measure, record and control a greater number of parameters than a human operator on a continuous basis makes an automated laboratory system an important goal in the analysis of environmental samples (6).

The first pioneering work in the area of automated organic synthesis was performed by Merrifield et al. (7). Lindsey and coworkers have discussed the importance of automation with the capability of parallelism in laboratory synthetic chemistry (8). These authors showed that multitasking capabilities of the computer are crucial to the success of automation of chemical work stations. Other examples of automated systems include systems for neutron activation analyses (9, 10), determination of Ca in human serum (11), and NMR structure refinement (12). The use of robots in the area of environmental chemistry include a system for monitoring nutrient levels in the North American Great Lakes (13) and analysis of organic micropollutants in water (14). The use of an automated system for pharmaceutical formulation pretreatment, analysis by HPLC, and report generation has been discussed by van der Voorden et al. (15). Robots have also had a significant impact in the area of health care, the benefits

of using surgical robots include lower exposure of staff to communicable diseases and lower labour costs (16). The use of automation for the assay of glycohemoglobin (useful for monitoring long term glucose control in diabetic patients) has been discussed by Herold et al. (17).

Crook (18) has suggested that the choice between dedicated automation (autosamplers) and flexible automation (robotics) is dependent on the application. In making the choice, the major criteria involve sample volume and the degree of difficulty of the analyses. Routine measurements on a large number of samples can be handled efficiently by an autosampler. However, a robotic system can deal with changes in analytical procedure (e.g. sample preparation) better than an autosampler. Vogelsanger (19) reported that economics, and the need for real-time communication with the instrument when performing repetitive tasks, are some of the reasons for using robots in the laboratory environment. The nature of error trapping in automated analyses requires that errors that occur during unattended analyses must be detected, identified and corrected, during the analysis. Techniques used to detect the presence of instrument or sample handling malfunctions are (a) visual records, (b) video records, and (c) computer generated audit trails. Although a visual record may be the best method to determine many instrumental failures, the objective in automation is to remove the technical attendant. A video recording can be an invaluable tool with which to monitor instruments for malfunctions when the instrumentation is visually accessible; for example the quality of the burner during analysis by flame atomic absorption spectrometry (FAAS) can be easily determined visually. However, in many instrumental techniques, for example, gas chromatography, no visual signature is available from which to assess the condition of the instrument during the measurement Although a computer-generated log file will provide an audit trail from which to reconstruct activities in unattended operation, this type of error trapping and recording for automated or robotic systems can only be used after the event to help a remote

or absent operator to locate the cause of a failure (7). This mode of corrective action greatly reduces the productivity of automated analyses.

The AAexpert program has been designed to implement totally automated analysis by FAAS in a number of stages. The program currently comprises an interconnected set of stand alone expert systems each performing a specific task connected with metal analysis by FAAS (1, 20, 21). AAcontrol provides real-time sample handling capabilities with a robotic solution handling station and AA-QC provides automated real-time assessment of the quality of the analytical data. The AA-QC (AA-Quality Control) module of AAexpert, which carries out the error trapping for AAexpert, must be closely associated with AAcontrol.

2.1 A Model for Totally Automated Metal Analyses

Automation of any analytical technique, in our case metal analyses by flame AAS, calls for real-time communication between (a) the instrument and its control software, (b) the autosampler and its control software, and (c) the quality control module. In a first step of error trapping, a quality control module can alert the operator of potentially poor quality data, and of instrument failures, and then halt the analytical process. A more complex second step in automation introduces the facility of intervention. Total automation requires that the quality control module can change the course of the analytical schedule so that new sample preparation and instrumental parameters can be requested, and a new sequence of events initiated. AAcontrol provides the interactive control necessary to accommodate real-time requests by AA-QC for changes in the sequence. Close connection between AA-QC and AAcontrol is necessary for unattended operation.

The trace of absorbance versus time, the absorption profile, has been identified as the best real-time indicator of data and instrumental quality of a single analytical measurement by FAAS. A series of features seen in the absorption profile following certain instrumental or measurement failures have been characterised using a number of real sample conditions. From this evaluation a number of assessment criteria have been extracted that become

part of a training set for AA-QC. AA-QC receives the absorption profile in real-time from AAcontrol as each sample is analysed. AA-QC must transmit messages back to AAcontrol to allow continuation of the analysis in the autosampler mode or to make modifications to the sequence in the robotic mode of AAcontrol. Real-time communication between AA-QC and AAcontrol is an essential step required to achieve total automation.

Figure 2.1 shows the sequence of events necessary in metal analysis by FAAS as implemented in the software described here. Following identification of samples, blanks, and standards by the user, AAcontrol establishes the sequence of tubes and beakers to be used for analysis. With this complete the analysis is initiated. The absorbance profile of the solution being aspirated is transmitted to AA-QC. AA-QC uses the parameters extracted from the training set applied by a number of production rules, to determine whether the recent measurement meets pre-set criteria. The appropriate message is transmitted back to AAcontrol which continues with the analysis. In the event that the method of analysis has to be changed, the autosampler functions as a robot in order to perform tasks such as dilution, addition of modifiers, and remeasurement of blanks and samples.

2.2 Hardware and Software

AAcontrol controls and acquires data from the Varian 875 FAAS, controls a Gilson 222 autosampler, and a Gilson 401 diluter (Figure 2.2). This liquid sample handler includes locations for samples, blanks, standards and modifiers used in the autosampler mode, and also additional modifiers and solvents that can be used in the robotic mode. AAcontrol has been updated (20) to run on an Intel 80486 based 33 MHz ALR computer with source code written in Microsoft C (version 6.0) using the 386 enhanced mode of the Microsoft Windows 3.1 environment for multitasking. The Varian 875 transmits the raw absorbance data in real-time at 60 Hz to the computer through a serial communications port. AA-QC was written using Microsoft FORTRAN (version 5.1) and Microsoft Visual Basic (version 3.0).

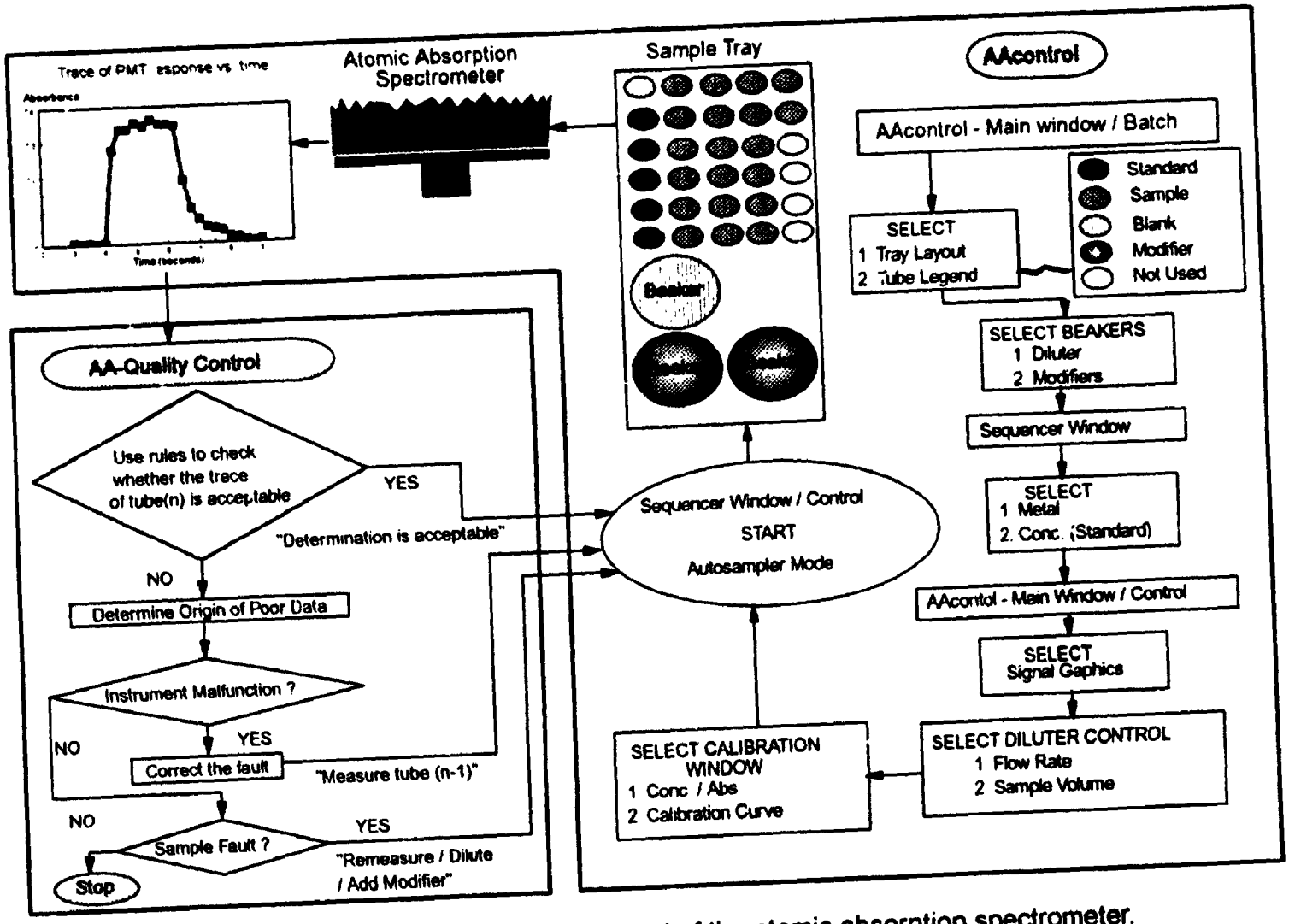


Figure 2.1 Real-time quality control of the atomic absorption spectrometer. Interaction between AAcontrol and AA-Quality Control.

93sep7a9.pre /sl

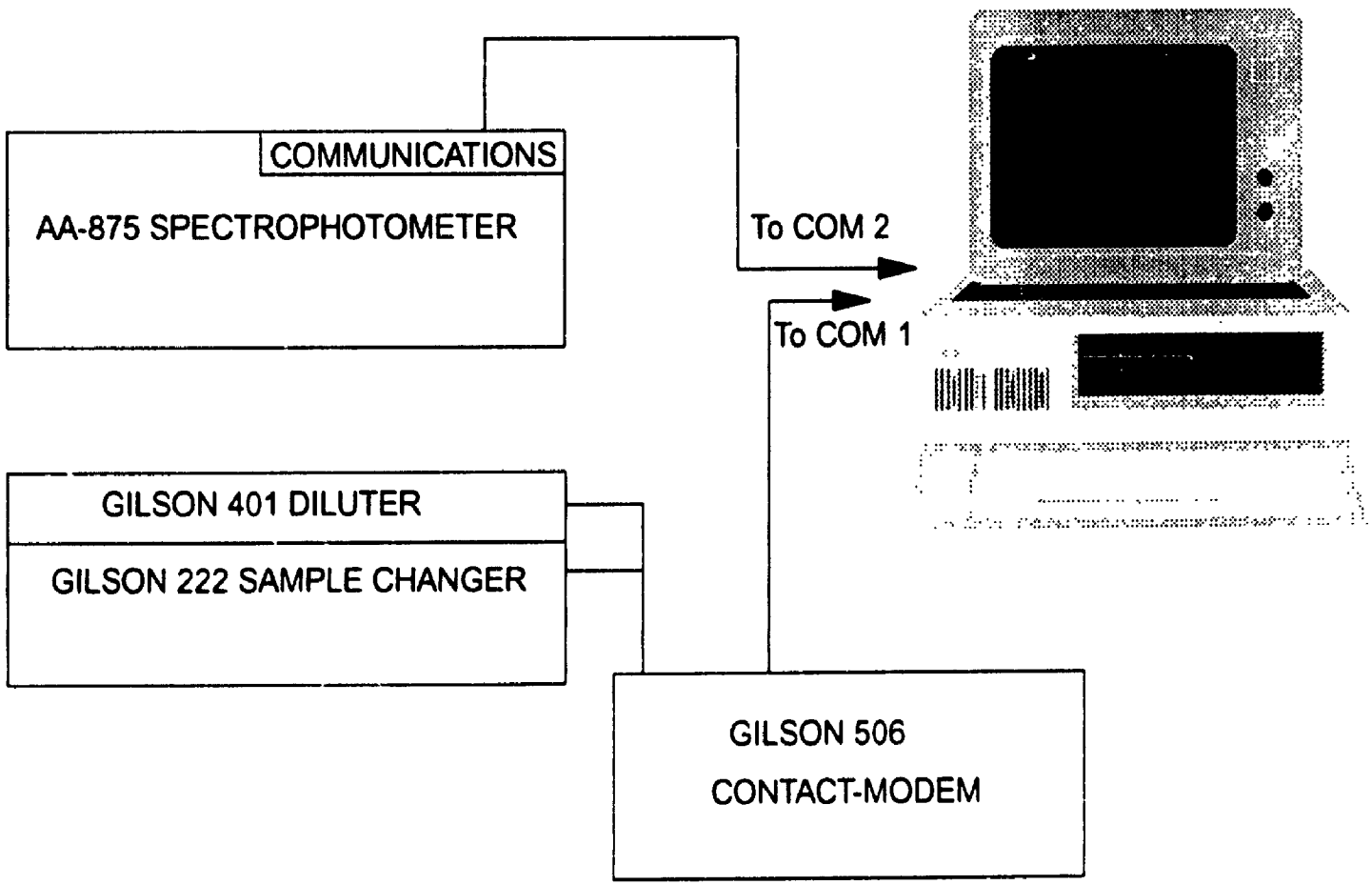


Figure 2.2 Communication: instrument to computer.
The Varian 875 atomic absorption spectrometer transmits raw absorbance measurements via a serial communication port to the computer. The Gilson 222 autosampler and the Gilson 401 diluter are connected via a second serial communication port to the computer.

2.3 AAcontrol: A Process Control Module

AAcontrol is responsible for all aspects of robotic solution handling, sample scheduling, instrument control, and analytical data acquisition from the AAS. The solution handling is carried out by a cartesian programmable autosampler and a diluter. To achieve the necessary flexibility, AAcontrol allows the user to assign a variety of tube racks and beakers using a set up file that includes the sizes of assorted racks and beakers. Figure 2.3 shows the modules that form AAcontrol During a run, AAcontrol presents the user with the "AAcontrol-Instrument Control" screen from which the tray layout is selected. The "Tube Legend" is used to select standards and samples in the tray layout. Next, the user selects the "Sequencer", chooses the element for which the solutions are being analysed and enters the concentration of the standards. The user returns to the "Instrument Control" window and selects Absorption Profile from the "Control" menu, this initiates the autosampler. The "Diluter Control" menu is used to select the flow rate and the sample volume. The "Calibration Window" menu reports the absorbance of the standard solutions, plots the calibration curve and calculates the concentration of the sample. The AAS sequence is initiated from AAcontrol by selecting Start from the "Control" menu in "Sequencer" (Figure 2.4). An absorption profile is captured following aspiration of each solution.

Each of the tasks (a-h) listed below operates as an independent module of the program. The tasks are.

a Tray-layout: Selects the tubes to be analysed and identifies blanks, standards and samples;

b. Analyte-selection. Selects the metal to be analysed and assigns the concentration values to the standards;

c AAS-control: Provides an image of the control panel on the Varian 875 AAS, allowing the user full access to all modes of operation;

d. Pump-control: Controls the pump used to inject the sample; allows the user to select the flow rate and aspiration volume of sample;

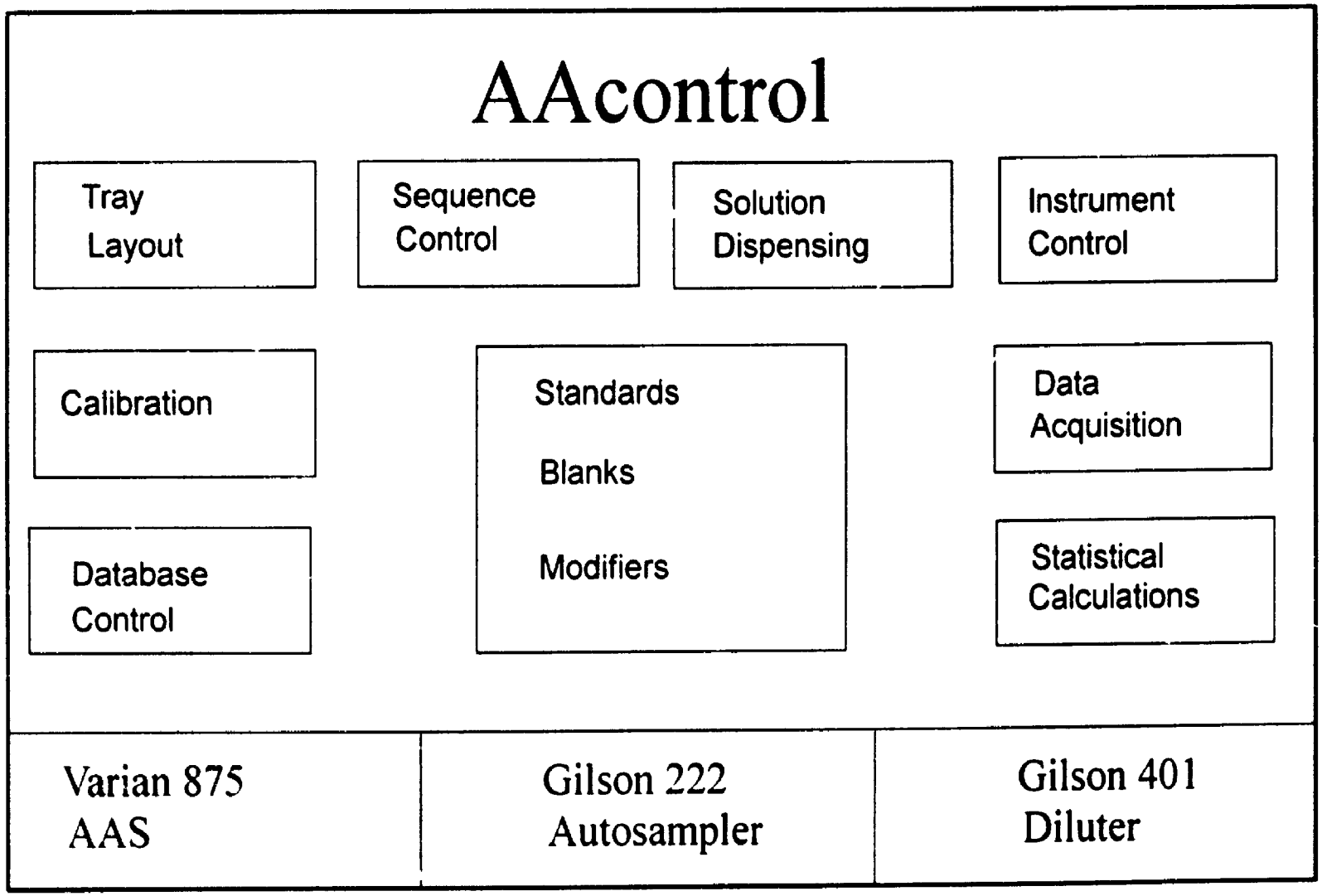


Figure 2.3 The modules of AAcontrol.

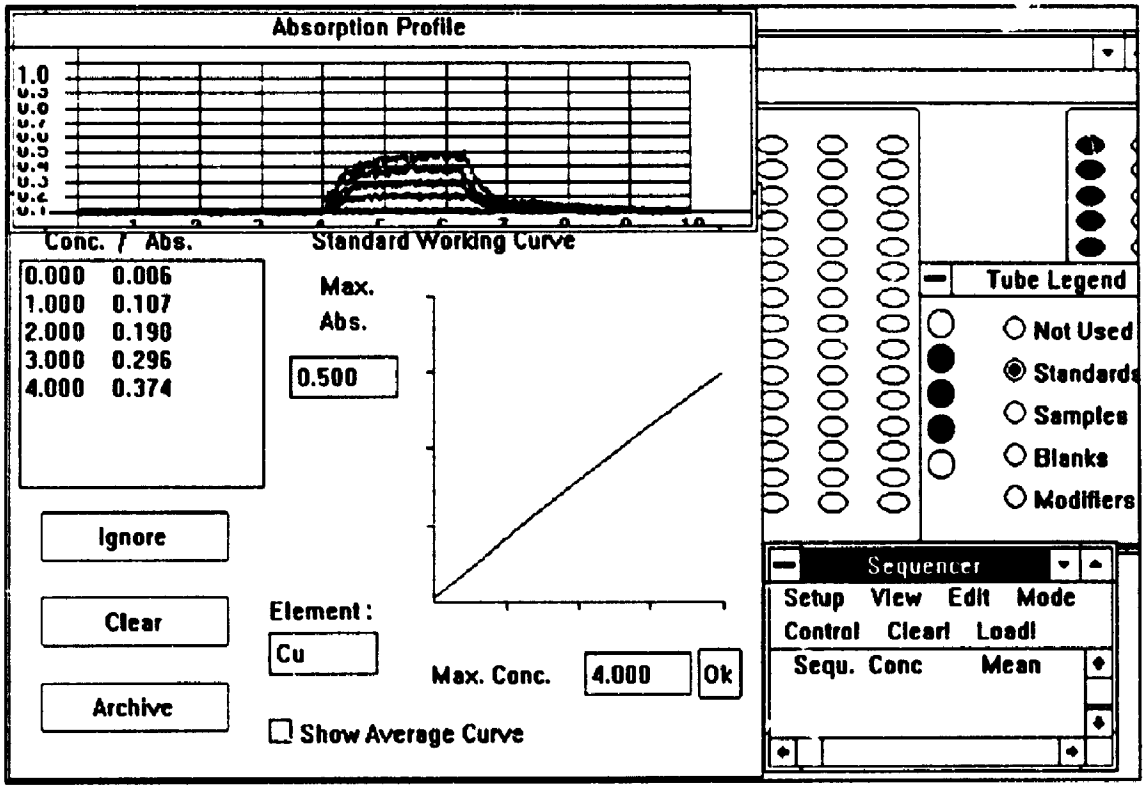


Figure 2.4 The user interface in AAcontrol.

e. Sequence-control: Starts and prematurely ends the analysis (if required);
f. Data-capture: Captures the absorption profile as a series of absorbances at a rate of 60 Hz following injection of the sample aliquot;
g. Data-communication: Transmits measured data to the AA-QC computer, and
h. Data-calculation: Provides a calibration graph for the standard solutions and concentration for samples.

2.4 A Model for AA-Quality Control: Characteristics of the Absorption Profile

AA-QC is an expert system that determines the quality of the analytical data by assessing the quality of the time profile of the absorption obtained in real-time during aspiration of a small volume of the sample solution. The chromophores in the atomic absorption experiment are isolated (gaseous) atoms. Typically, when a flame is used for atomisation, the process begins with the aspiration of the sample into the spray chamber (22). The quality of the measured absorption signal depends on (i) the stability of the flame, (ii) the homogeneity of the atoms in the flame, and (iii) the absence or presence of chemical interferences. Clearly, anything that interrupts the formation of atoms or interferes with the flow of the atoms in the flame in the optical path will change the absorption profile. The temporal distribution of the absorbance following introduction of a small volume of analyte, typically 300 μL, into the flame, reflects the hydraulic properties of the capillary uptake tube, the spray chamber, and the burner assemblies, together with the performance of the burner. Figure 2.5a shows the absorption profile obtained following aspiration of 300 μL of a 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm. There are three distinct parts, namely (a) a rise in the absorbance as the aspirated metal ions form metal atoms, (b) a peak region when the number of ground state metal atoms reaches an equilibrium concentration in the region of the optical path resulting in maximum absorption, and (c) a decay of the absorbance as the aliquot is consumed. Three parameters that reflect the status of the instrument, sample preparation, and sample uptake conditions have been extracted for use

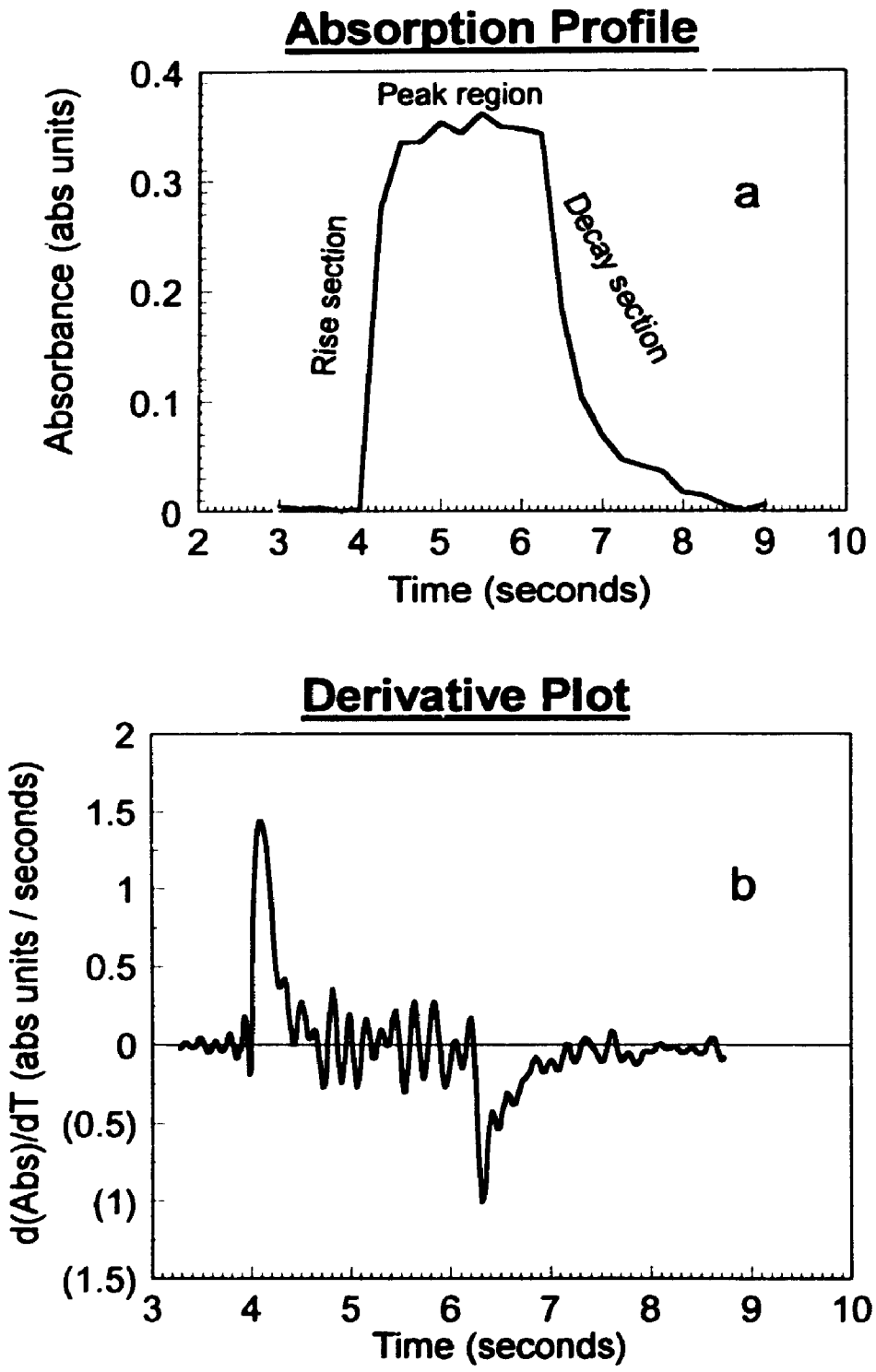


Figure 2.5 The components of an absorption profile.

(a) The absorption profile of a 4.0 ppm Cu(2+) solution monitored at 324.8 nm with a spectral band width of 0.5 nm.

(b) The derivative plot of the absorption profile of a 4.0 ppm Cu(2+) solution monitored at 324.8 nm with a spectral band width of 0.5 nm.

as assessment criteria, namely (a) the maximum positive slope, (b) the maximum negative slope, and (c) the mean absorbance of the peak region of the absorption profile. By experiment it has been determined that for a standard solution (a solution of known and defined matrix) the derivative of the absorption profile comprises (a) one maximum positive slope (corresponding to the increase in atomic concentration), (b) a flat region (corresponding to the equilibrium concentration), and (c) one maximum negative slope (corresponding to the reduction in atomic concentration), Figure 2.5b.

AA-QC uses a number of production rules to carry out real-time quality control. These rules are based on the analysis of problems that can occur during measurements. Training sets were generated by specially preparing the spectrometer to create faulty conditions.

A 1000 ppm Cu stock solution in 5% HNO_3 (BDH Inc.) was used to prepare four standard solutions. These were:

a. A 1.0 ppm Cu^{2+} solution;
b. A 2.0 ppm Cu^{2+} solution;
c. A 3.0 ppm Cu^{2+} solution, and
d. A 4.0 ppm Cu^{2+} solution.

The abosrbance was measured at 324.8 nm with the spectral band width set at 0.5 nm.

On aspirating a solution, AAcontrol determines the detection limit (DL) of the instrument. Any absorbance measuring less than 3*DL is considered noise and absorption profile data values are not transmitted to AA-QC. The raw absorption data obtained from AAcontrol are smoothed using a Fast Fourier Transform (FFT) routine written in FORTRAN (Figure 2.6). The inverse FFT is calculated by treating the first 22% and the last 22% of the Fourier transformed data as data points, the remaining data are treated as zero. This level of smoothing was chosen to allow for meaningful calculations involving the data stream without altering the essential characteristics of the trace. The parameters extracted from the absorption profile are calculated in a FORTRAN

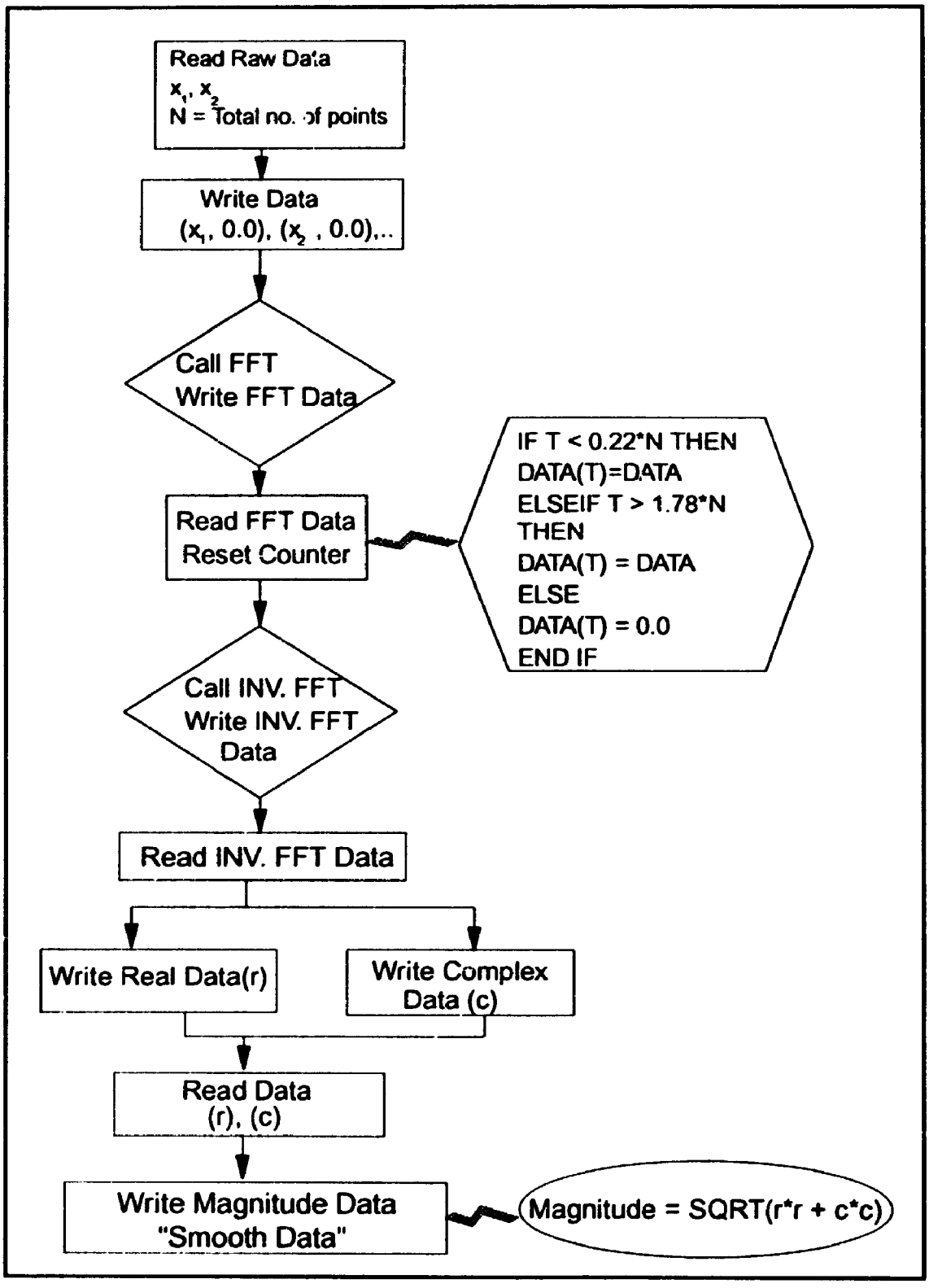


Figure 2.6 Flow chart to calculate the FFT of raw absorbance values.

dll routine (Figure 2.7). Based on assumptions that for standard solutions there are no instrument malfunctions, the calculated parameters, along with the name of the element being analysed and concentration of the standard solution, is archived. If a better set of values are obtained during subsequent analyses of the standard, the archived parameters are updated.

2.5 Flow Chart of AA-Quality Control

The flow chart outlining the structure of AA-QC is shown in Figure 2.8. On opening the data file of interest, the user can select Absorption Profile Parameters (Recent File). Recent criteria, namely (i) maximum positive slope, (ii) maximum negative slope, and (iii) mean peak absorbance of the measured data values are calculated. Next, the user can view either the absorbance trace of the current measurement by selecting Graph or compare the most recent criteria with an archival copy. In "Comparison with an Archival Copy", the user must first find the concentration of the sample in order to compare the current criteria with an appropriate archival copy. The calculation of the sample concentration provides information regarding the nature of the absorbance profile and the possibility of comparison with an archival copy. On entering the value of the corresponding archived concentration, the user selects "Find Archival Copy". The user is offered criteria of the archived copy and a comment regarding the comparison of the archived slopes with the current maximum positive and maximum negative slopes. The selection of "Sample Conc." for a second time results in further information about the comparison of the absorbance values. On the other hand, the user can view the absorbance trace after viewing the data and proceed to calculate the recent criteria.

2.6 Calculation of Absorbances

The absorbance value of a particular solution is calculated using the flow chart in Figure 2.7. From the set of data belonging to the peak region, a mean absorbance is reported. From the study of the absorption profile of measured data values, it was found that in an ideal case (measurement of a standard solution with no instrument faults) the "separation" in terms of data points

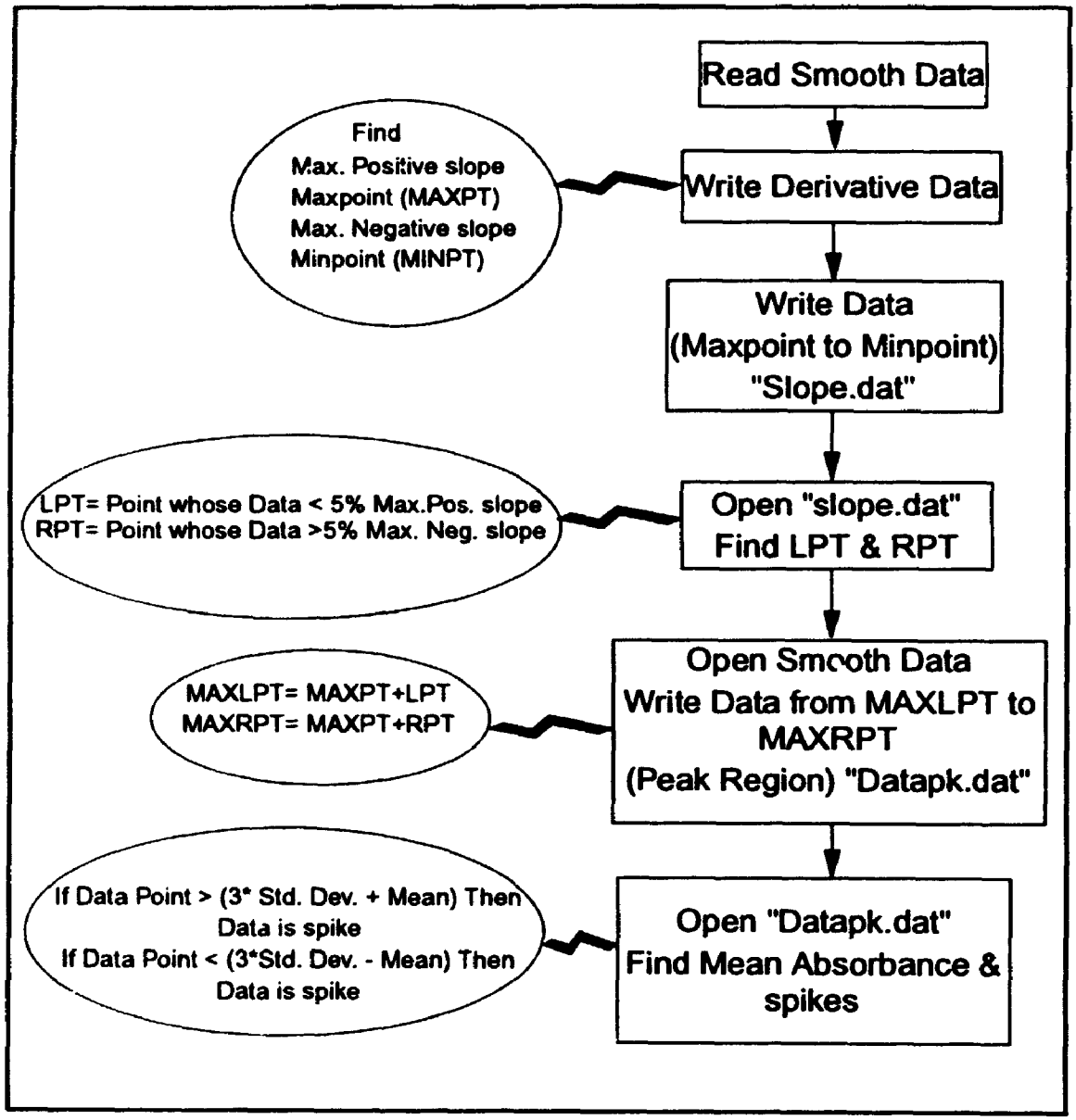


Figure 2.7 Flow chart to calculate criteria of data quality in AA-Quality Control.

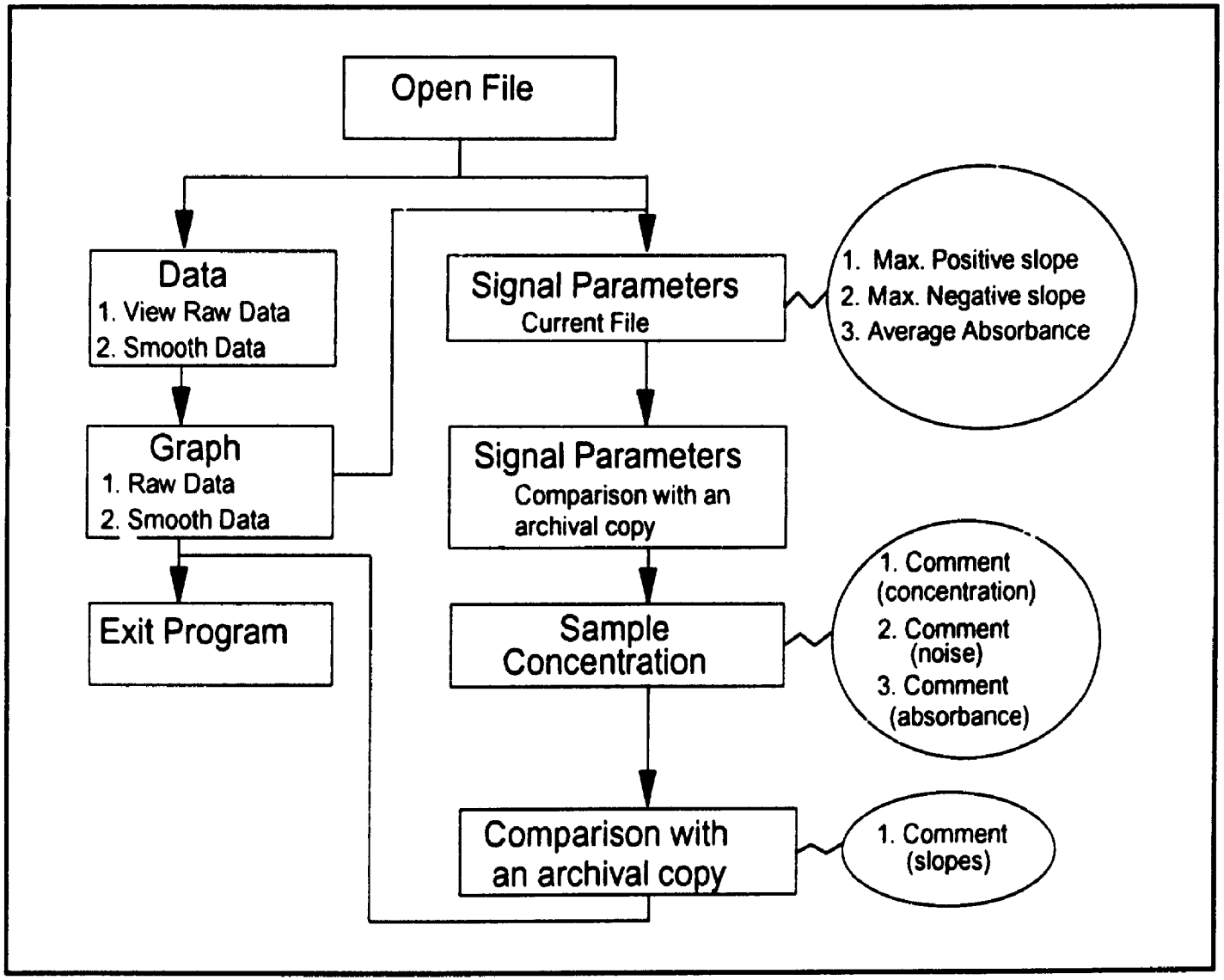


Figure 2.8 Flow chart of AA-Quality Control.

between the maximum positive slope and maximum negative slope is about 2.2 seconds. Although the timings will differ from instrument to instrument the overall pattern of absorbance versus time will be the same. The following rule is used in order to decide whether the reported absorbance value is a meaningful number.

IF separation > 1.98 And separation < 2.42 THEN

report the mean absorbance of the current data stream (Print the value)

ELSE do not report the mean absorbance (Print "No comment"). (rule 1)

2.7 Prediction of a Concentration

The approximate concentration of a sample aliquot is based on the mean absorbance of the peak region in the absorption profile. The following equation calculates concentration (approx_conc) of an unknown :

approx_conc= (Value of mean absorbance of the unknown) / (ratio)

where ratio is calculated from the set of standard solutions that have been archived, ratio = (Sum of Recent ratio) / (Number of Standard solutions) and Recent ratio = (Value of an archived mean absorbance) / (Conc. of the corresponding solution).

However, in some cases an approximate concentration is not reported because (a) the separation between the maximum positive slope and the maximum negative slope is not in the expected range, and/or (b) the point of maximum negative slope of the absorption profile (Minpoint) appears before the point of maximum positive slope (Maxpoint). In such instances a "No comment" is printed. The following rule is used in reporting the approximate concentration of an unknown:

IF Minpoint > Maxpoint AND current_absorbance <> "No comment" THEN calculate approx_conc (Print the value; Comment [conc] "Concentration can be calculated, compare with archival copy."

ELSE do not report approx_conc (Print "No Comment"; Comment [Conc.] "Concentration error [Absorbance error]). (rule 2)

2.8 Fluctuations in the Absorption Profile

Two kinds of systematic fluctuation are found: (a) spikes, which are defined here as being systematic, significant and rapid increases in absorbance that can occur at various times in the absorption profile itself, and (b) spikes occurring at various times and in differing numbers in the derivative plot of the absorption profile. Fluctuations in the absorption profile can be further subdivided into three regions, namely (a) spikes at the beginning of the peak region, (b) fluctuations in the middle of the peak region, and (c) spikes at the end of the peak region. The derivative plot of a standard's absorption profile will produce just two peaks, which correspond to the maximum positive slope and maximum negative slope.

Fluctuations in the Peak Region

Although the type of peak fluctuation is calculated for every trace, it is not used as a selective criterion. A data point from the peak region is identified as a spike when a data point lies outside of the mean peak region absorbance +/- three standard deviations of the peak region absorbance. A file consisting of the time of the detected spikes is written in terms of the percentage of the total elapsed time of the absorption profile using (time where the spike occurs/total number of points in the peak region)*100. Simple rules to categorize the fluctuations in the absorbance values were constructed of the type:

IF percentage of profile passed when spike occurs >0 AND <20% THEN "Spike at the beginning of the peak region". (rule 3)

IF percentage of profile passed when spike occurs >20% AND < 80%THEN "Fluctuation in the peak region". (rule 4)

IF percentage of profile passed when spike occurs >80% AND <= 100% THEN "Spike at the end of the peak region". (rule 5)

The user interface is able to display combinations of messages in cases where the fluctuations belong to more than one region. The following rule is used when AA-QC does not report the fluctuations in the peak region:

IF recent_absorbance="No Comment" Or sample_conc="No Comment" THEN "Cannot comment on fluctuations in the absorption profile, refer to the graph". (rule 6)

Fluctuations in the Derivative Plot of the Absorption Profile

The flowchart for the calculation of the number of maxima from the maximum positive slope to the maximum negative slope in a derivative plot is shown in Figure 2.9. The following rules are used to predict the nature of the derivative graph of the absorption profile.

IF separation >1.98 AND separation < 2.42 AND peak_number=2 THEN "Derivative plot shows one maximum positive slope and one maximum negative slope." (rule 7)

IF separation >1.98 AND separation < 2.42 AND peak_number<> 2 THEN "Derivative plot shows the separation between maximum positive slope and maximum negative slope is within range. However the number of maxima: (Input calculated peak_number)." (rule 8)

IF separation < 1.98 OR separation > 2.42 THEN "Poor derivative plot - number of maxima: (Input calculated peak_number)." (rule 9)

2.9 Implementation of AA-Quality Control

Problems encountered during analyses will be reflected in changes in the absorption profile. We chose three common problems involved in analyses by atomic absorption spectrometry to test the automated system. The problems chosen were:

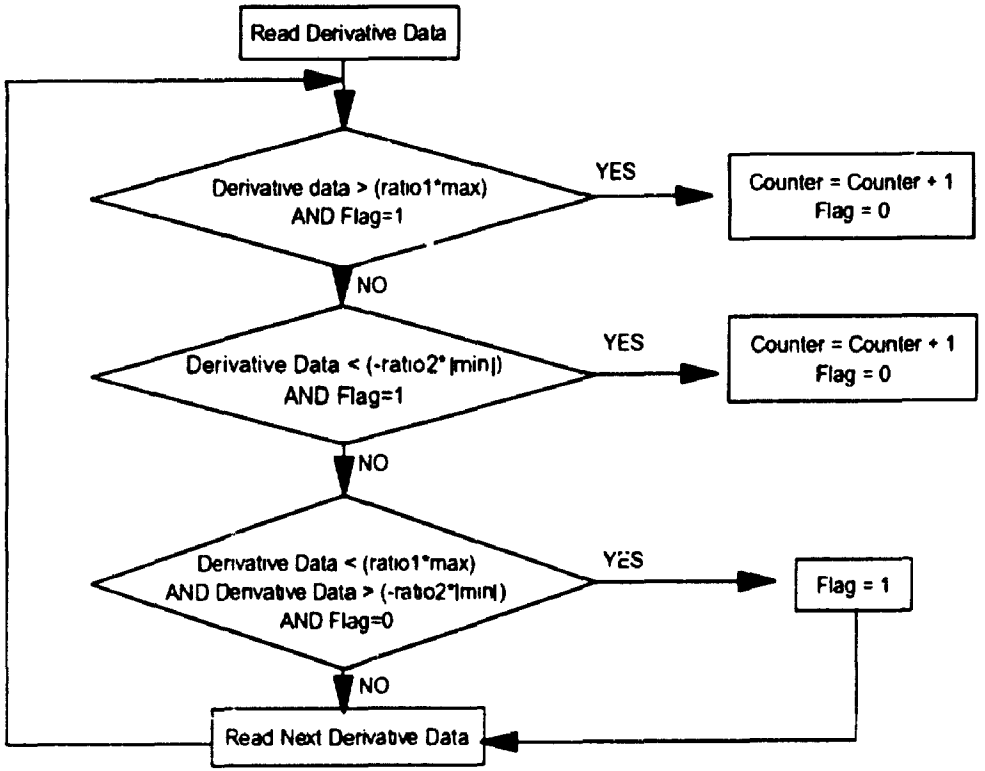
a. a blocked burner due to a high salt content in the matrix;

b. the use of a viscous solution, and

c. a blocked capillary tube.

In addition, we also tested an instrumental problem,

d. the selection of a capillary tube for aspiration that was too long for a specific solution.



$$\text{ratio1} = \frac{\sum_{\text{ppm}}^{\text{ppm}} 3 * \text{std dev. (peak region)}}{\sum_{\text{ppm}}^{\text{ppm}} \text{max}}$$

$$\text{ratio2} = \frac{\sum_{\text{ppm}}^{\text{ppm}} 3 * \text{std dev (peak region)}}{\sum_{\text{ppm}}^{\text{ppm}} |\text{min}|}$$

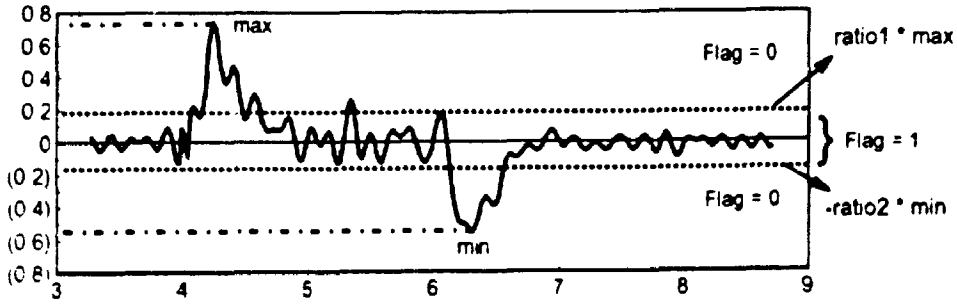


Figure 2.9 Flow chart to calculate the number of peaks in the derivative plot of the absorption profile

Figure 2.10 shows the absorption profile of a 4.0 ppm Cu^{2+} solution along with those generated from the chosen problems. The derivative plots of the absorption profiles are shown in Figure 2.11

2.9.1 Testing that the Training Set Represents the Current Instrument Response

Before AA-QC can be used to provide quality control for unknown samples, the appropriateness of the training set for this specific analysis must be confirmed. This is carried out by measuring the absorption profiles of standard solutions. The quality of the most recent measurement of a standard solution is compared with an archival copy using the assessment criteria. The information regarding the most recent slopes is obtained by comparing both the most recent maximum positive slope (recent_positive_slope) and maximum negative slope (recent_negative_slope) of a standard solution with its archival copy (archive_positive_slope, archive_negative_slope). The rules used in both comparisons are identical in that if either of the most recent slopes fall in a range outside 90% to 110% of the stored value then the most recent solution needs to be remeasured. The following rule compares the most recent maximum positive slope with an archival copy.

IF recent_positive_slope < 0.9*archive_positive_slope THEN "Most recent data is not satisfactory, remeasure"

ELSEIF recent_positive_slope >1.1*archive_positive_slope THEN

"Most recent data is not satisfactory, remeasure"

ELSE

recent_positive_slope > 0.9*archive_max_slope AND

recent_positive_slope <1.1*archive max slope THEN "You may replace the archived data ". (rule 10)

The most recent mean absorbance (recent_absorbance) is compared with the archived value (archive_absorbance) using the following rule:

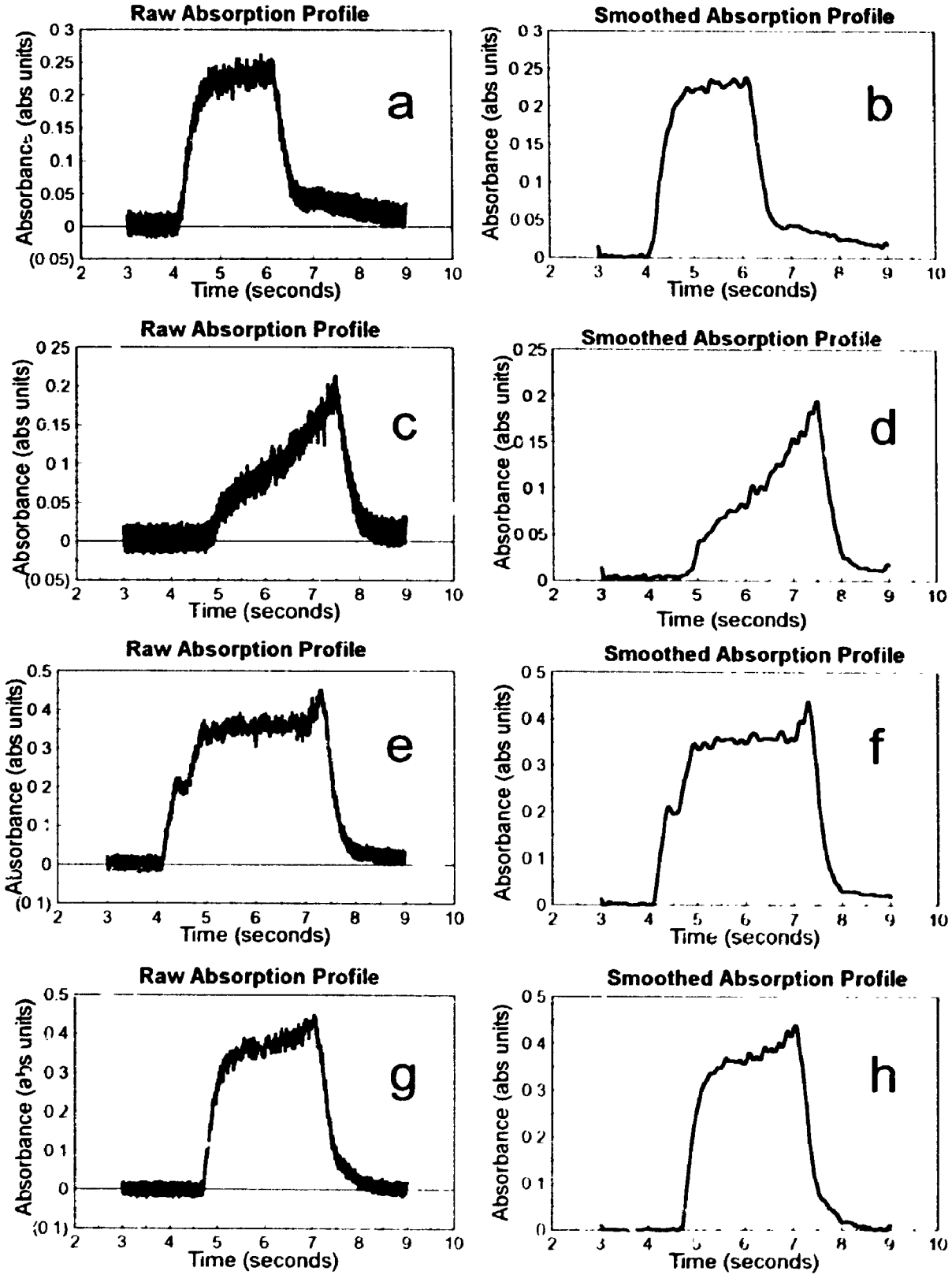


Figure 2.10 Absorption profiles (raw, smooth) of a 4.0 ppm Cu(2+) solution aspirated under different conditions. A blocked burner (a,b) a viscous solution (c,d), a blocked capillary tube (e,f), and a long capillary tube (g, h).

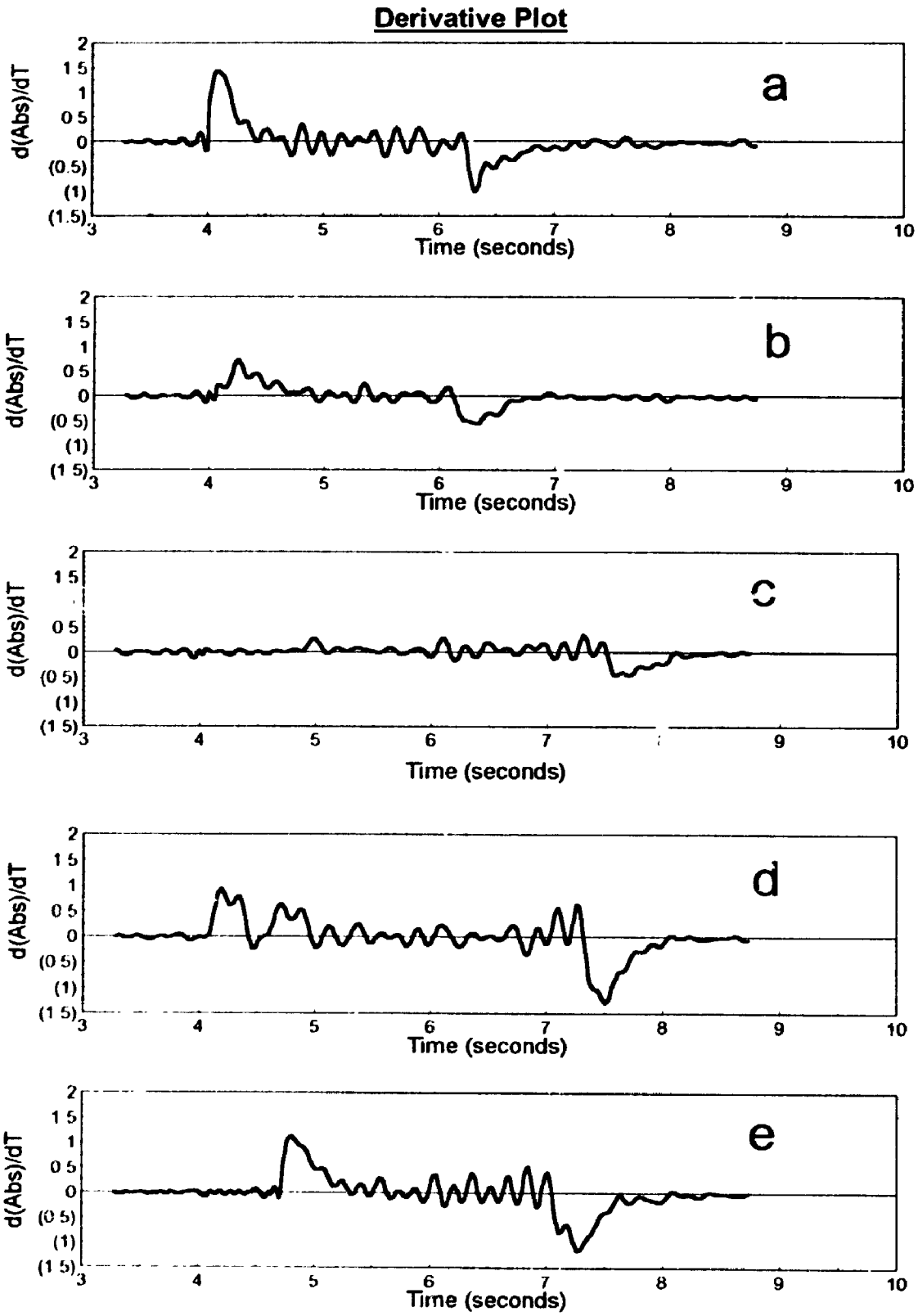


Figure 2.11 Derivative plots of a 4.0 ppm Cu(2+) solution aspirated under different conditions. No instrument fault (a), a blocked burner (b), a viscous solution (c), a blocked capillary tube (d), and a long capillary tube (e).

94ma23b2.pre/sl

IF recent_absorbance <0.9*archive_absorbance OR recent_absorbance >1.1*archive_absorbance THEN "There is a concentration error (absorbance considerations)"

ELSE " Present concentration is within range (absorbance considerations)". (rule 11)

Figures 2.12 and 2.13 show the assessment of the quality of data of a 4.0 ppm Cu^{2+} solution.

2.9.2 Effect of a Blocked Burner

When the burner is blocked by salt, encrustation of salt around the orifices in the nebuliser results in poor transport efficiency of the analyte (22,23). The presence of the non-volatile matrix will affect the vaporisation rate of the analyte, which will give rise to a loss in the absorption signal. The nebuliser performance may also be erratic leading to a noisy absorption profile.

A saturated solution of NaCl (25%) was aspirated until the obvious result of a blocked burner was noticed, namely a ragged flame with a hissing sound (Figure 2.14). The quality of the burner head led us to believe that we would observe major fluctuations in the peak region of the absorption profile. A 4.0 ppm Cu^{2+} solution was aspirated into the burner and the data values captured. Figures 2.15 and 2.16 show the assessment of the quality of data carried out in two steps. In Step 1, the approximate concentration of the solution is calculated and it was found to be 2.6 ppm. The blockage lowers the absorbance of the 4.0 ppm Cu^{2+} solution. To our surprise, the only effect observed from the effect of the blocked burner was a lowering of the absorbance values. However, AA-QC was able to detect two additional maxima in the derivative plot (Figure 2.11b). The presence of two additional maxima in the derivative plot suggests irregularities in the transport processes of the analyte. It is also possible that the presence of salt in the burner slot contributes to the poor derivative plot.

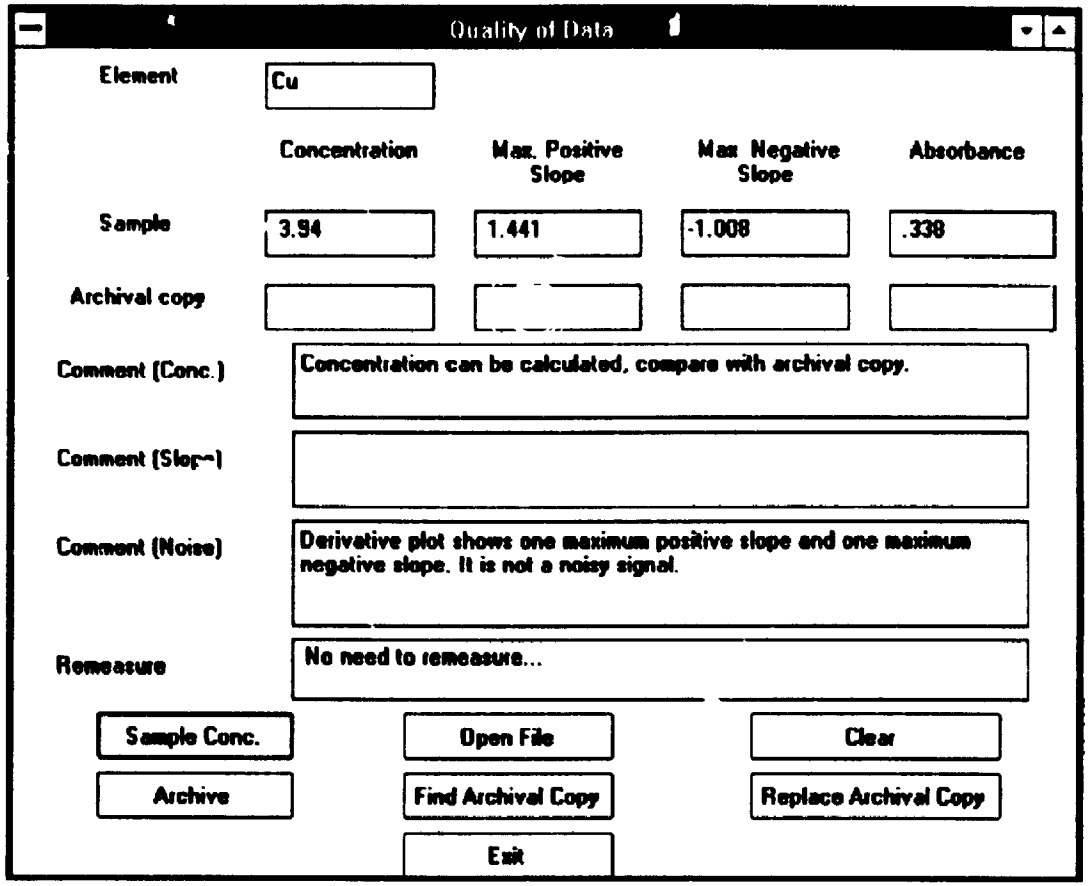


Figure 2.12 Step 1 in the calculation of the quality of analytical data (aspiration of 300 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm with no instrument faults). On clicking "Sample Conc.", concentration of the most recent absorbance values is reported.

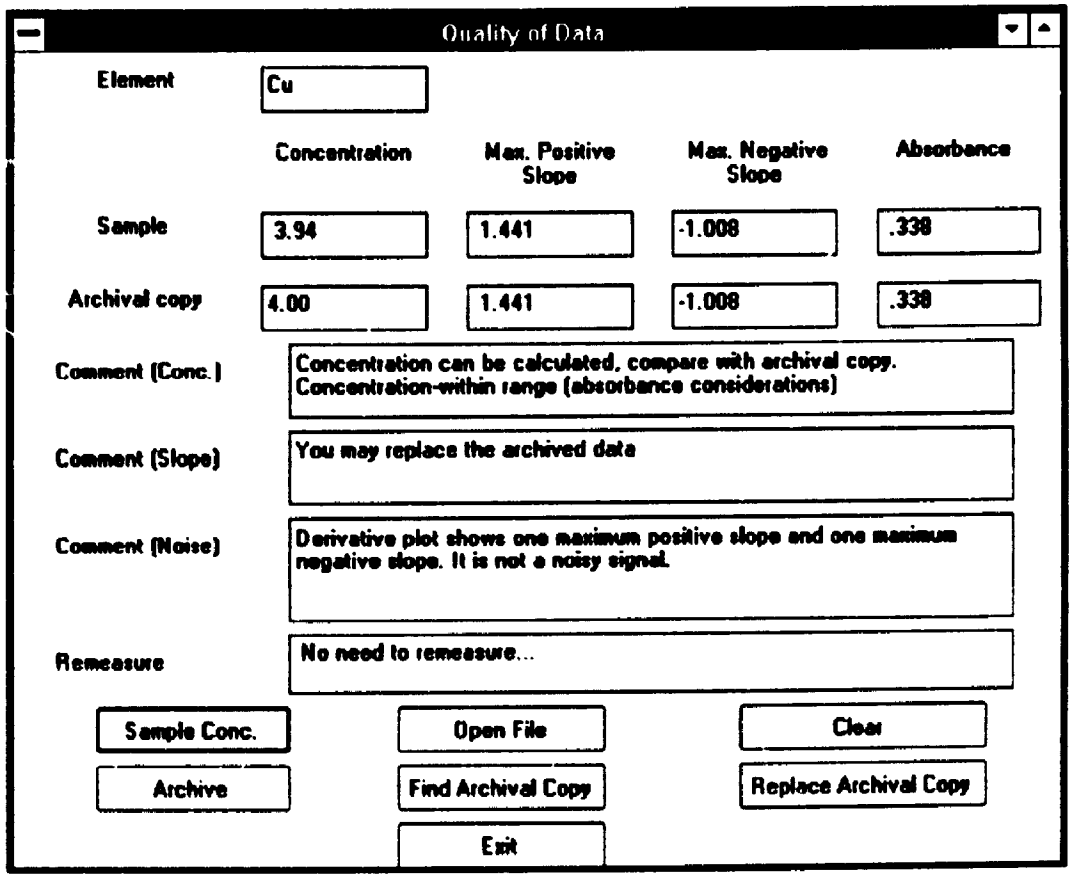


Figure 2 13 Step 2 in the calculation of the quality of analytical data (aspiration of 300 µL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm with no instrument faults). On entering the value of the archived concentration in the box against archival copy and clicking "Find Archival Copy", the values of the archived criteria appear in the appropriate boxes.

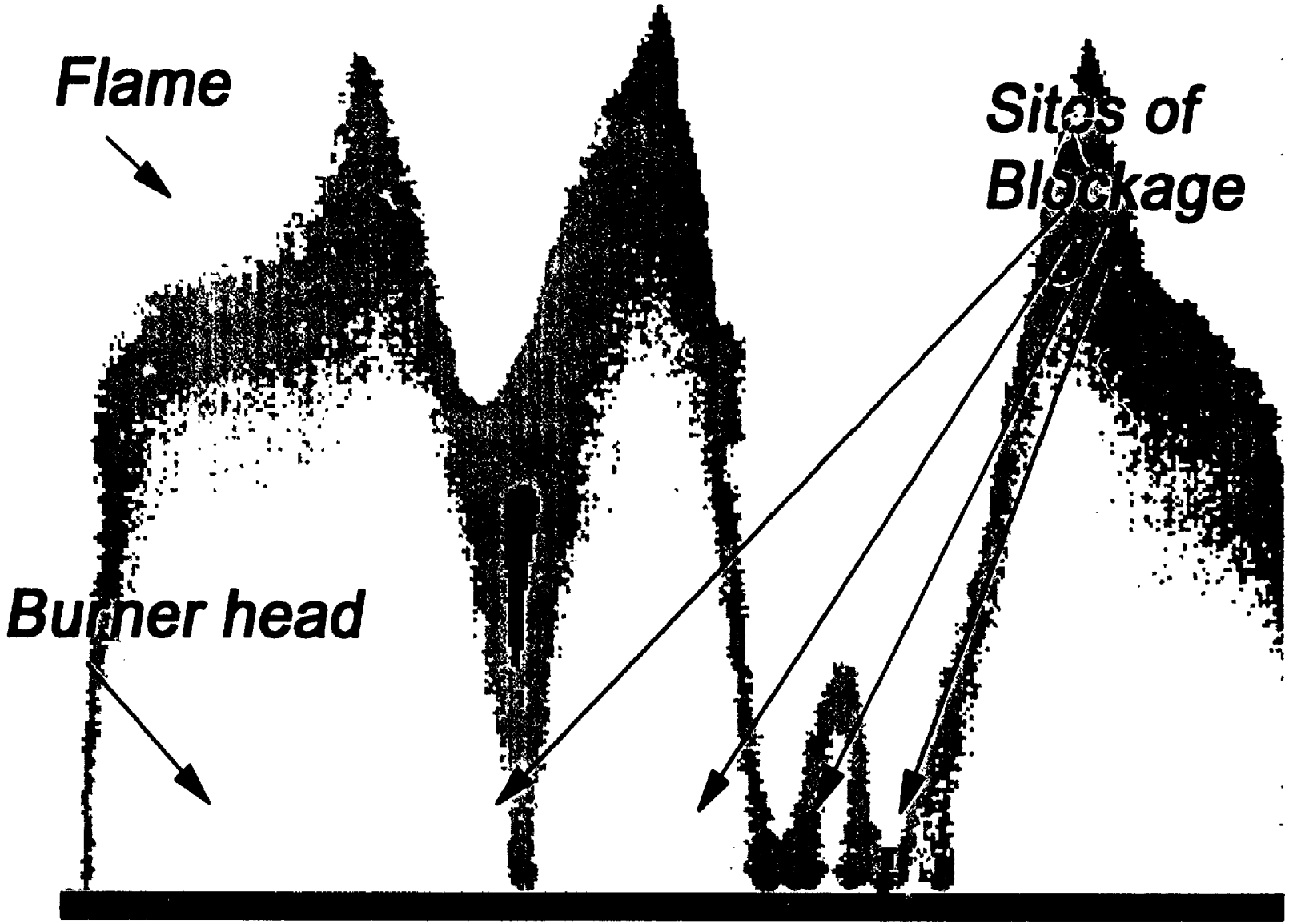


Figure 2.14 Scanned photograph showing the burner head blockage after aspiration of a saturated solution of NaCl.

94ap10ap4.pre/sl

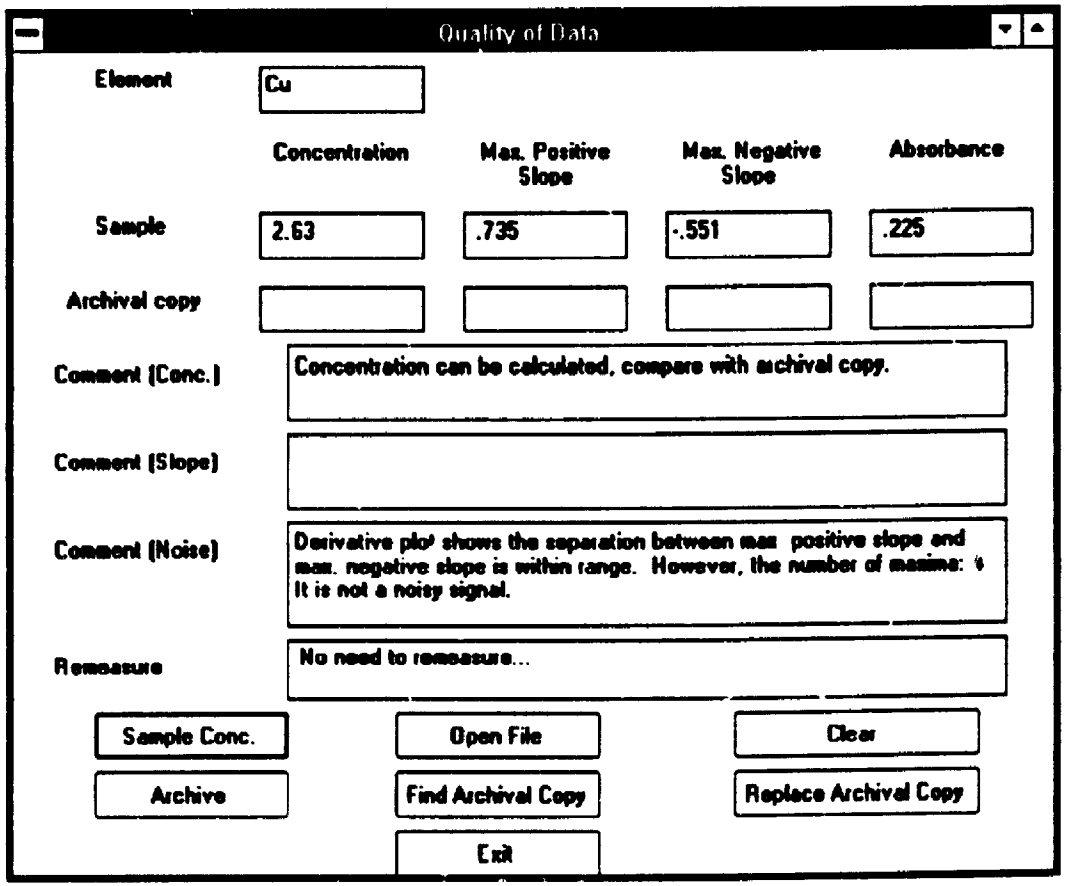


Figure 2.15 Step 1 in the calculation of the quality of analytical data (effect of a blocked burner on the aspiration of 300 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On clicking "Sample Conc.", concentration of the most recent absorbance values is reported.

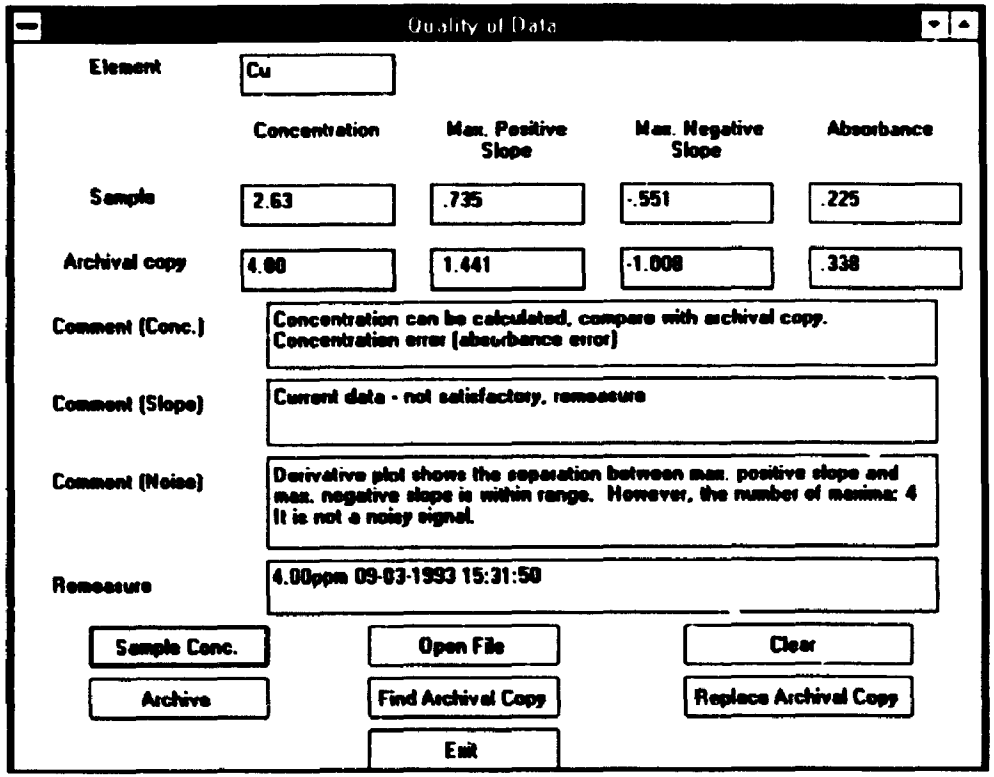


Figure 2.16 Step 2 in the calculation of the quality of analytical data (effect of a blocked burner on the aspiration of 300 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On entering the value of the archived concentration in the box against archival copy and clicking "Find Archival Copy", the values of the archived criteria appear in the appropriate boxes.

2.9.3 Effect of a Viscous Solution

On replacing water with a more viscous, less volatile organic solvent, we expect to observe a reduction in the absorption signal. This occurs primarily because the aspiration rate is lower giving rise to lesser amount of primary aerosol. In addition, the lower volatility of the solvent reduces its evaporation from the aerosol leading to the formation of larger droplets (24). The combination of these effects will give rise to a noisy absorbance trace with a long rise time, together with reduced absorbance values.

A 4.0 ppm Cu^{2+} solution in 70% ethylene glycol was used to study the effect of a viscous, less volatile organic solvent. Figures 2.17 and 2.18 show the quality of data. Because the separation between the maximum positive slope and the maximum negative slope is not in the expected range (rules 1 & 2), the program does not calculate the approximate concentration of the 4.0 ppm Cu^{2+} solution. Comment (Noise) shows that the absorption profile might be characterised as having a long rise time. The smoothed absorption profile shows that indeed the signal has a long rise time (Figure 2.10d). The derivative plot is not as expected for a normal trace as it exhibits ten maxima (Figure 2.11c). The absorption profile suggests that the aspiration and nebulisation processes result in the irregular formation of fewer analyte atoms in the flame.

2.9.4 Effect of a Blocked Capillary Tube

The reduced flow rate from a blocked capillary tube results in a broadening of the aliquot of sample as it passes along the capillary tube. This should be seen as a larger separation between the maximum positive slope and the maximum negative slope. The effect of this broadening will be reflected in a spike at the end of the peak region. The blockage should also give rise to irregularities in the transport phenomena.

The effect of a blocked capillary tube was to reduce the diameter of the tube using a clamp. A 4.0 ppm Cu^{2+} solution was aspirated into the burner to study this effect on the absorption profile. The smoothed absorption profile shows that there is a spike at the end of the peak region (Figure 2.10f). Figures

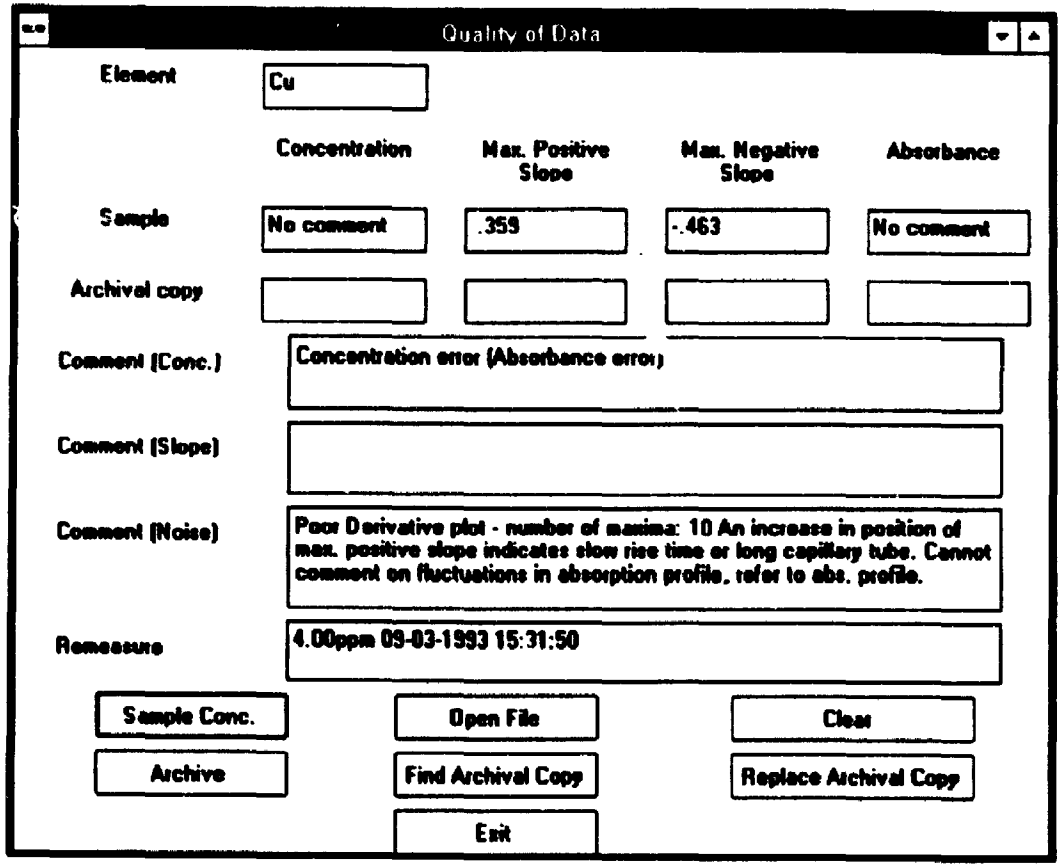


Figure 2.17 Step 1 in the calculation of the quality of analytical data (effect of a viscous solution on the aspiration of 200 µL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On clicking "Sample Conc.", it is found that the concentration of the most recent absorbance values is not reported.

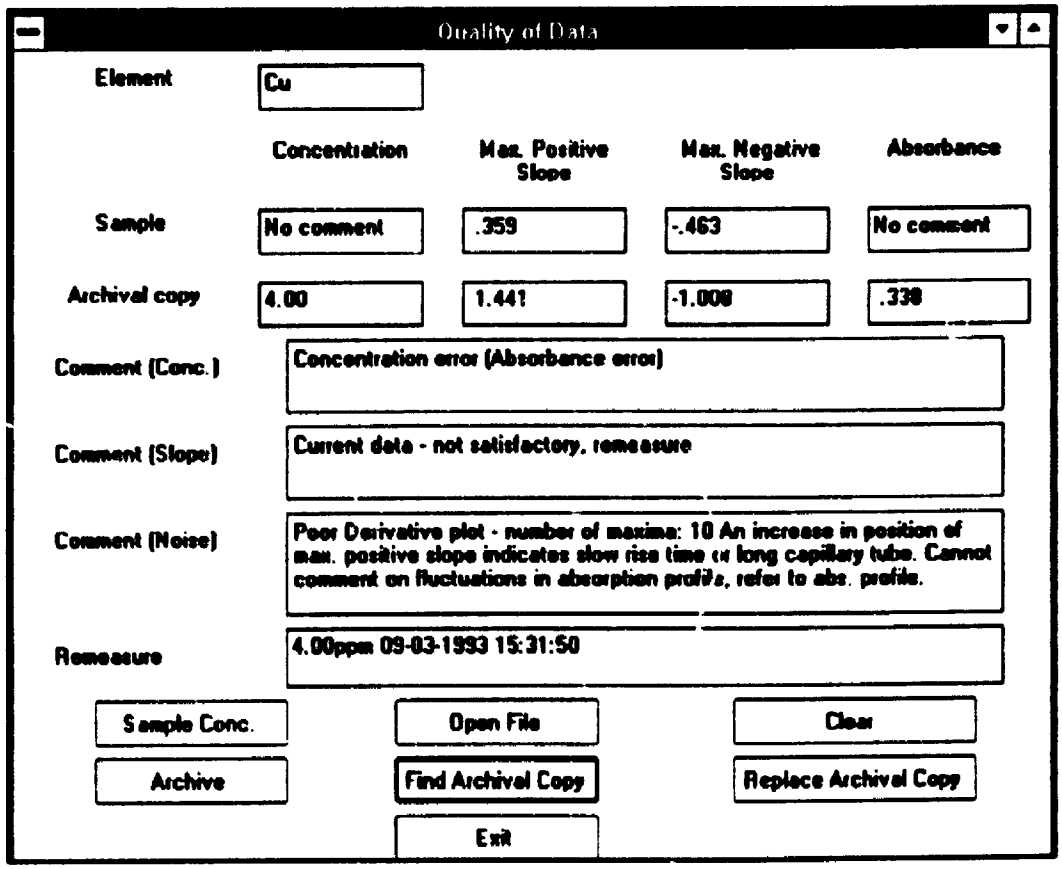


Figure 2.18 Step 2 in the calculation of the quality of analytical data (effect of a viscous solution on the aspiration of 200 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On entering the value of the archived concentration in the box against archival copy and clicking "Find Archival Copy", the values of the archived criteria appear in the appropriate boxes.

2.19 and 2.20 show the quality of data from the interpretative data. The greater number of peaks, in the derivative plot, (Figure 2.11d), is due to problems in the aspiration of the sample aliquot through the narrow tube and to the increased suction that affects the last portion of the sample.

2.9.5 Effect of a Long Capillary Tube

If the capillary tube of the normal length (L= 50 cm) is replaced by a longer capillary tube (L=90 cm), the increased time taken for aspiration will be reflected in a shift in the position of the maximum positive slope. Also, the increased broadening of the microsample in the capillary tube will result in greater separation between the maximum positive slope and maximum negative slope. It was found that the position of maximum slope (Max slope pt) was approximately 0.83 seconds. The following rule was written in order to account for problems of a long capillary tube or a signal with a very slow rise time.

> IF most_recent_positive_slope > 1.0 THEN "An increase in the position of the maximum slope indicates either slow rise time or a long capillary tube." (rule 12)

A long capillary tube was used to test the effect of a long capillary tube. A 4.0 ppm Cu^{2+} solution was aspirated using a 90 cm long capillary tube. Figures 2.21 and 2.22 show the quality of a 4.0 ppm Cu^{2+} solution using the 90 cm long capillary tube. The separation between the maximum positive slope and the maximum negative slope is larger due to the broadening of the microsample in the capillary tube. There is a delay in the occurrence of the maximum slope (rule 12). The derivative plot (Figure 2.11e) shows the presence of four maxima due to the presence of a small spike at the end of the peak region of the smoothed absorption profile (Figure 2.10h). The presence of the spike can be explained by the increased suction for the last portion of the microsample.

2.10 Automated Correction

AA-Quality Control provides an assessment of the data and suggests corrections to rectify simple problems. The following correctional messages, not

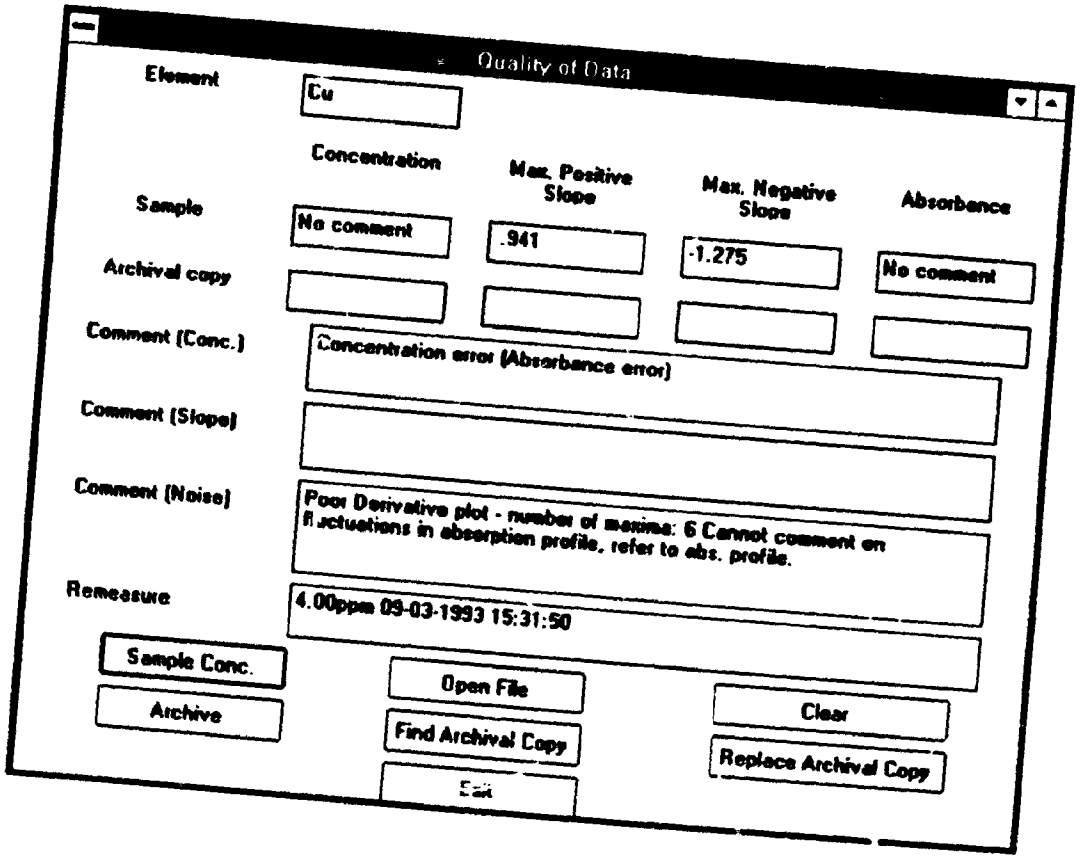


Figure 2.19 Step 1 in the calculation of the quality of analytical data (effect of a blocked capillary tube on the aspiration of 300 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On clicking "Sample Conc.", it is found that the concentration of the most recent absorbance values is not reported.

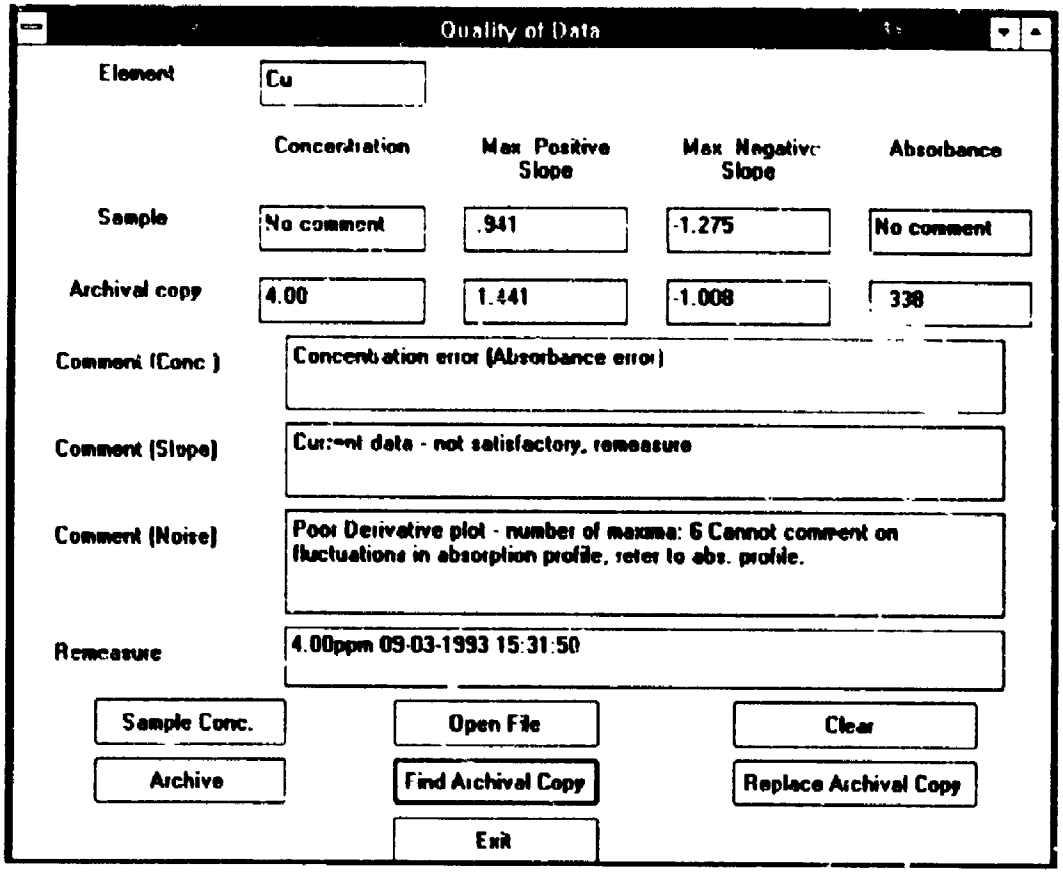


Figure 2.20 Step 2 in the calculation of the quality of analytical data (effect of a blocked capillary tube on the aspiration of 300 µL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On entering the value of the archived concentration in the box against archival copy and clicking "Find Archival Copy", the values of the archived criteria appear in the appropriate boxes.

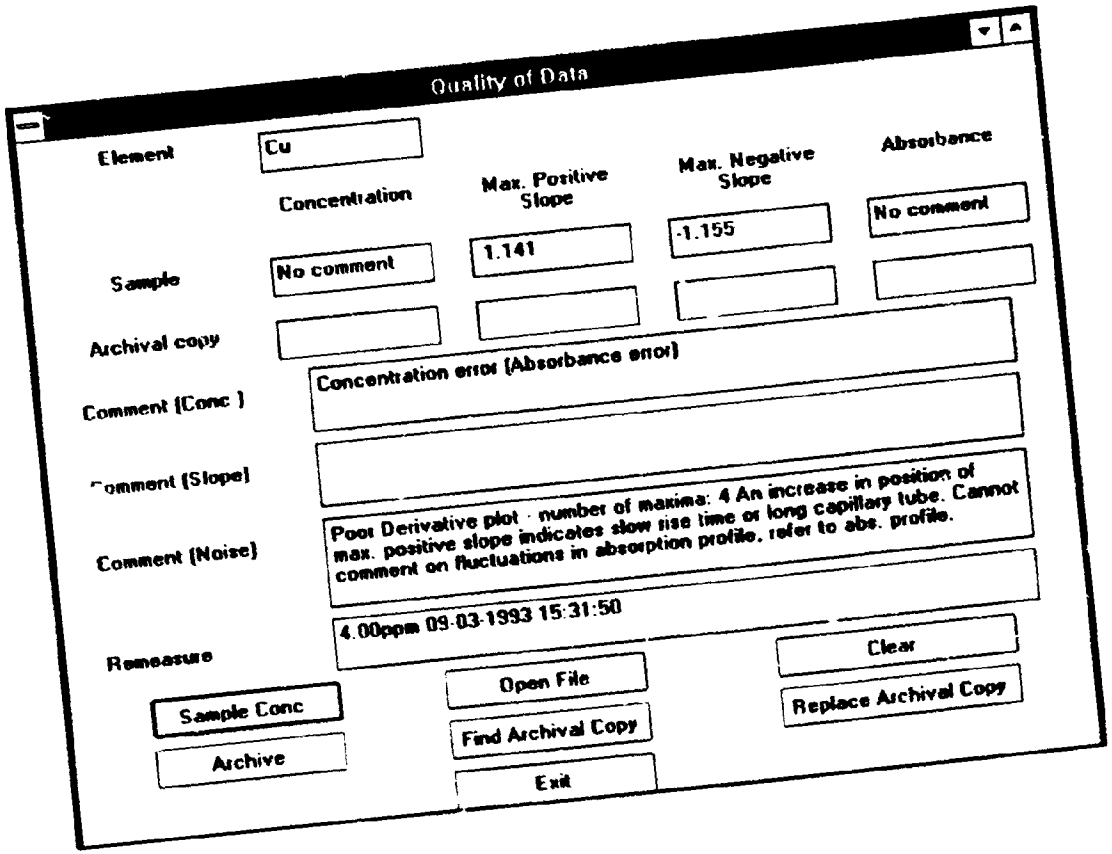


Figure 2.21 Step 1 in the calculation of the quality of analytical data (effect of a long capillary tube on the aspiration of 300 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On clicking "Sample Conc.", it is found that the concentration of the most recent absorbance values is not reported.

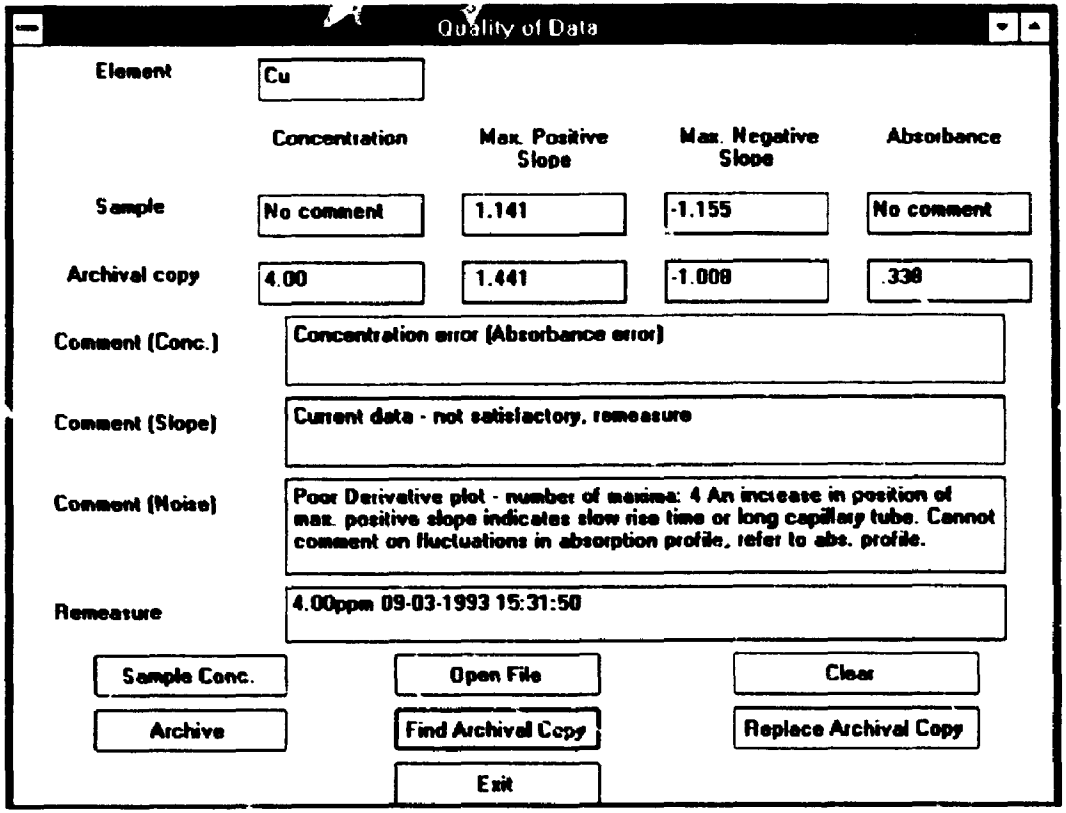


Figure 2.22 Step 2 in the calculation of the quality of analytical data (effect of a long capillary tube on the aspiration of 300 μL of 4.0 ppm Cu^{2+} solution monitored at 324.8 nm with a spectral band width of 0.5 nm). On entering the value of the archived concentration in the box against archival copy and clicking "Find Archival Copy", the values of the archived criteria appear in the appropriate boxes.

in any particular order, can be transmitted from AA-QC to AAcontrol depending on the cause identified from the analytical data. These are:

a. Remeasure, tube number;
b. Dilute, tube number;
c. Add modifiers, tube number;
d. Aspirate blank, and
e. Stop analysis.

In order to carry out the instructions, AAcontrol switches from an autosampler mode to a robotic mode of operation (Figure 2.23). Following this corrective action, AAcontrol switches back to the autosampler mode and continues with sample analysis as scheduled.

The problems studied in this paper gave rise to four kinds of error messages. Table 2.1 lists the error conditions and the relevant interpretations. Error messages along with the suggested corrective action are listed in Table 2.2.

Problems with a blocked burner generate error 1. AA-QC attempts to solve problems associated with blockage by issuing instructions to aspirate blank to dissolve solids and to remeasure a standard solution in order to check the quality of the instrument. Analysis of a viscous sample produces an absorption profile with a long rise time (error 2) therefore dilution of the sample should remove the problem. The dilution has to be carried out in stages with repetitive measurements, just as one would in a manual operation. If AA-QC encountered error 3 while analysing a standard solution, the suggested correction is to stop the analysis. The operator is required to check the length of the capillary tube and shorten the tube, if necessary. Aspiration of solution through a blocked capillary tube generates error 4. The correction suggested by AA-QC is to stop the analysis. The analysis can be restarted after cleaning the capillary tube with the cleaning wire.

Automated correction should be attempted for a maximum of two trials. If the correction does not produce data of acceptable quality, the analysis should

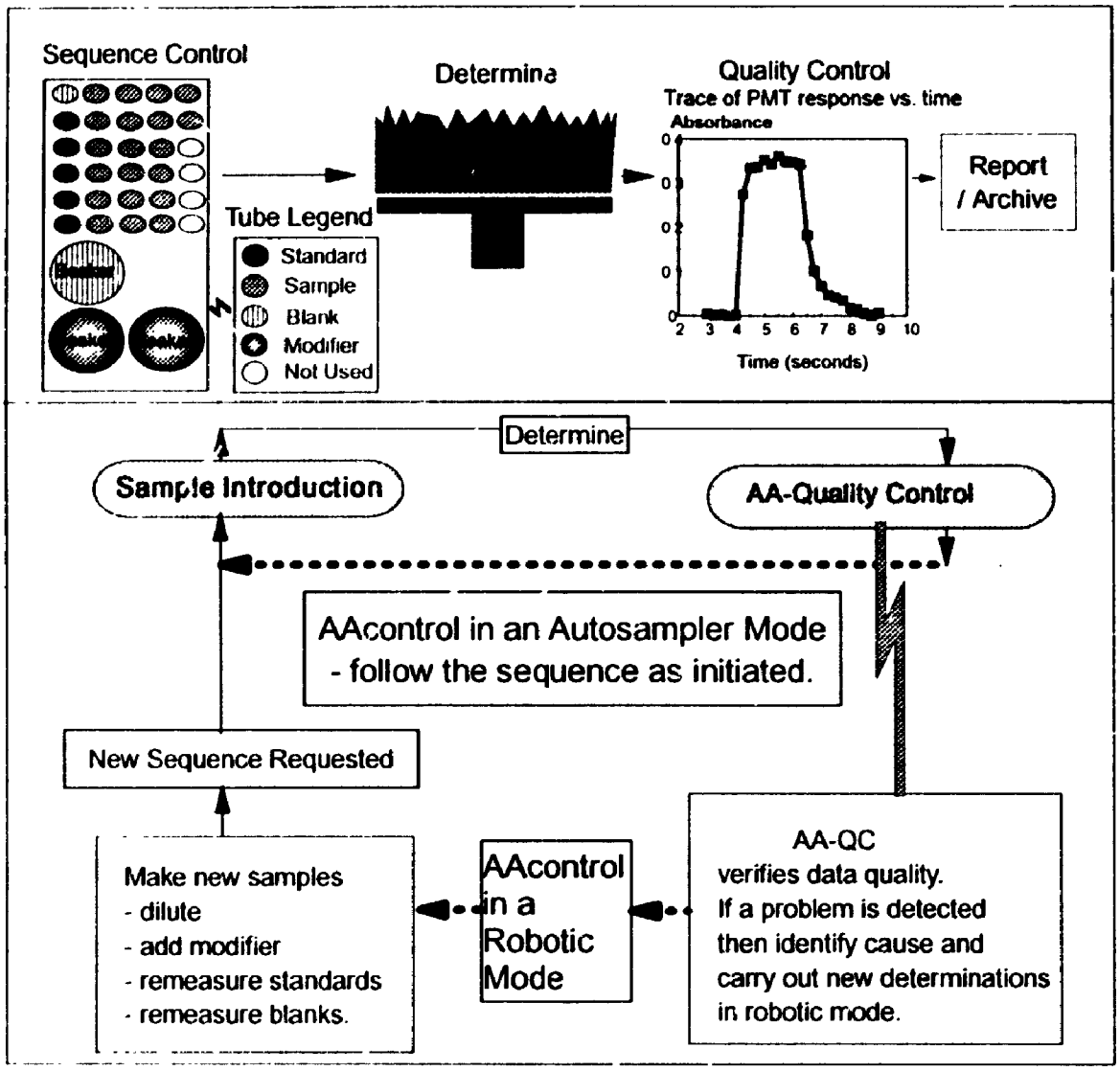


Figure 2.23 Modes of operation of an automated solution handler.

Table 2.1 Error conditions and interpretations determined in AA-Quality Control.

Error condition	Interpretation
Separation between the maximum positive slope and the maximum negative slope is in the expected range but number of maxima is greater than 2.	Irregularities in the hydraulic transport phenomena of the analyte results in deviation from a standard derivative plot containing two maxima as shown in Figure 2.11b.
Poor derivative plot; the number of maxima = n. Separation between the maximum positive and maximum negative slope is less than 90% of the expected range.	Problems associated with aspiration and nebilization of samples give rise to multiple maxima whose separation is less than the expected range.
Poor derivative plot; the number of maxima = n. Separation between the maximum positive and maximum negative slope is greater than 110% of the expected range.	Aspiration of samples through restrictions associated with blocked capillary tubes, or use of an excessively long capillary tube, leads to an increased suction for the last portion of the sample. This effect is observed as an increase in the separation between the maximum positive and the maximum negative slope.
There is a delay in the appearance of the maximum positive slope.	Increased time taken for aspiration of samples to form metal atoms causes a delay in the appearance of the maximum positive slope.

Table 2.2 Error messages and corrections suggested by AA-Quality Control.

Error Message	Interpretation	Auto-Correction
Error 1: Separation between the maximum positive and the maximum negative slope is in the expected range but number of maxima is greater than 2.	Blockage in nebulizer or burner slot result in irregularities in the transport phenomena.	**Aspirate blank** for 15 minutes, **remeasure** standards.
Error 2: (i) Poor derivative plot; the number of maxima = n. Separation between the maximum positive and maximum negative slopes is less than 90% of the expected range. (ii) There is a delay in the appearance of the maximum positive slope.	Aspiration of a viscous sample gives rise to an absorption profile with a long rise time. Poor derivative plot with a large number of maxima and with less separation than expected indicates problems involved with the aspiration and nebulisation processes.	**Dilute tube (n), remeasure tube (n).**
Error 3: Poor derivative plot; the number of maxima = n. Separation between the maximum positive and maximum negative slopes is more than 110% of the expected range.	Increased separation between the maximum positive and maximum negative slopes indicates increased suction in the aspiration o the last portion of the sample aliquot. When this problem is associated with problems with transport phenomena, it indicates blockage in the capillary tube.	**Stop analysis**, remove capillary tube, clean tube.
Error 4: (i) Poor derivative plot; the number of maxima = n. Separation between the maximum positive and maximum negative slopes is more than 110% of the expected range. (ii) There is a delay in the appearance of the maximum positive slope.	Greater separation between the maximum positive and maximum negative slopes indicates increased suction in the aspiration of the last part of the sample aliquot. This problem in combination with a delay in the appearance of the maximum positive slope indicates an increased time taken for aspiration of the metal ions to form metal atoms. The increased suction at the end of aspiration gives rise to a greater number of maxima in the derivative plot.	**Stop analysis**. Check the length of capillary tube, shorten capillary tube, if necessary.

be stopped and the corrective software, AAdiagnosis, now a Windows-based expert system, should be consulted to find the cause of the problem.

2.11 Discussion

Continuous operation of the atomic absorption spectrometer calls for three kinds of software (i) real-time control of the instrument and solution handling robot, (ii) quality control software, and (iii) software that is able to correct faults detected prior to the next analysis. Our work shows that if appropriate real-time data can be captured from the instrument and estimates made about the quality of the analysis then possible causes of problem data can then be identified. Diagnostic software that acts as an advisor to the analyst has been described in our earlier paper (1). AAdiagnosis uses symptoms provided by the user and suggests specific corrective action to be taken in order to fix the cause of the problem. However, AAdiagnosis requires that the analyst provide the symptoms of faulty analysis in terms of descriptions of the spectrometer (e.g. the condition of the flame) or sequences in the set of data (e.g. slowly rising baseline) or quality of the analytical data (e.g. very noisy). While this type of diagnostic expert system works well "off line" as an advisor it requires operator intervention.

In this chapter, whether completely automated analysis is possible based only on a mathematical analysis of the analytical data is explored Can a rule based expert system like AA-QC recognize a few common but significant problems associated with, in this case, metal analyses by FAAS, by carrying out tests of the analytical data? In AA-QC numerical analysis of the absorption profile is used to allow AA-QC to recognize a limited number of problems shortly after the sample aliquot passes through the flame It is believed that the absorption profile is the most sensitive indicator of possible problems that occur during the analysis. The changes in this trace can be interpreted in terms of the build up of atoms in the flame, the establishment of equilibrium, and finally the reduction in atoms as the sample volume is exhausted. To analyse this temporal distribution of the absorbance we have defined a set of criteria that the data must meet to be defined as a successful analysis.

In the case of a blocked burner, irregularities in the transport phenomena of the analyte are reflected in the derivative plot as an increased number of maxima. For viscous samples, irregular aspiration and nebulisation processes are detected from the delay in the position of the maximum positive slope resulting from a slowly rising absorption signal associated with a slow establishment of an equilibrium concentration of atoms in the flame. Blockage of the capillary tube causes a broadening of the sample aliquot which gives rise to an increased separation between the maximum positive slope and the maximum negative slope. The effect of poor transport of the sample is reflected in a greater number of peaks in the derivative plot. In cases where a long capillary tube is used, the program determines that there is a delay in the occurrence of the maximum positive slope and that the separation between the maximum positive and maximum negative slopes in not in the expected range of 1.98 to 2.42 seconds. Incorrect separations in the above three cases are interpreted by AA-QC as absorbances for which (a) the mean absorbance of the solution, (b) the approximate concentration, and (c) the type of fluctuation in the peak region (rules 1, 2, 6) cannot be used usefully.

This study has emphasized the need for a dual mode of solution handling in real time automation of analytical instruments. Solution handling in AAcontrol implements these two modes: (a) an autosampler mode that is able to carry out routine analysis, and (b) a robotic mode that is able to modify the method of analysis. The robotic mode of analysis is prompted by the AA-Quality Control on encountering measured data values that do not qualify as acceptable data. Necessary corrections are made and a new sequence of measurement is requested. In this new sequence of determination, the analysis is initiated at the point of the last acceptable measurement. Figure 2.23 summarizes the switching of modes in AAcontrol which is a necessary option in automated solution handling.

It has been found that the critical steps involved in realization of real-time corrective control of analytical measurements are (i) obtaining measured data

values that reflect the instantaneous condition in the sample chamber and (ii) modelling the measured data values in terms of analyte quality. Multitasking abilities are essential for simultaneous communication with the instrument and processing the raw data. Modelling the detector response implies understanding the physical and chemical processes involved in the production of the analytical signal.

2.12 References

1. Lahiri, S; Stillman, M. J. *Anal. Chem.* 1992, **64**, 238A.
2. Frazer, J. W. *Accounts of Chemical Research* **1974**, **7**, 141.
3. Buydens, L; Schoenmakers, P; Maris, F.; Hendriks, H. *Anal. Chim. Acta* 1993, **272**, 41.
4. Settle, F. A., Jr.; Walter, P J.; Kingston, H. M.; Pleva, M. A.; Snider, T; Boute, W. *J. Chem. Inf. Comput. Sci.* 1992, **32**, 349.
5. Zhou, T; Isenhour, T. L.; Marshall, J. C. *J. Chem. Inf. Comput. Sci.* 1993, **33**, 569.
6. Guette, J. P.; Crenne, N; Bulliot, H; Desmurs, J. R; Igersheim, F. *Pure & Appl Chem.* 1988, **60**, 1669.
7. Merrifield, R. B.; Stewart, J. M.; Jernberg, N. *Anal. Chem.* 1966, **38**, 1905.
8. Corkan, L A.; Lindsey, J. S. *Chemom. Intell. Lab. Syst.: Lab. Info. Management* 1992, **17**, 47.
9. Munoz, L ; Gras, N.; Guzman, C.; Thieck, M.; Vanska, L. *J. Radioanal. & Nucl. Chem.* 1993, **167**, 97.
10. Reis, M. F ; Martinho, E. *J. Radicanal & Nucl. Chem.* 1993, **167**, 39.
11. Merrer, R. J.; Berthrong. P.; Claffey, K. *Amer. Lab.* 1992, **24**, No. 17, 15.
12. Kim, S. G.; Reid, B. R. *J. Magn. Res.* 1992, **100**, 382.
13. Agemian, H.; Ford, J. S.; Madsen, N. K. *Chemom. Intell. Lab. Syst.: Lab. Info. Management* 1992, **17**, 145.
14. Millier, A.; Vallet, G. *Chemom. Intell. Lab. Syst.: Lab. Info. Management* 1992, **17**, 153.
15. van der Voordern, E. C., Gout, E.: van Houlten, O. L. T.; Vendrig, D. E. M. M. *Chemom. Intell Lab Syst.: Lab. Info. Management* 1992, **17**, 129.
16. Felder, R. A , Boyd, J. C.; Margrey, K. S.; Holman, W.; Roberts, J.; Savory, J *Anal. Chem.* 1991, **63**, 741A.

17. Herold, C. D.; Andree, K.; Herold, D. A.; Felder, R. A. *Clin. Chem.* 1993, **39**, 143.

18. Crook, M. *Chemom. Intell. Lab. Syst.: Lab. Info. Management* 1992, **17**, 3.

19. Vogelsanger, M. *Chemom. Intell Lab. Syst.: Lab Info. Management* 1992, **17**, 107.

20. Browett, W. R.; Cox, T. A.; Stillman, M. J. In *Expert System Applications in Chemistry*; Hohne, B. A.; Pierce, T. H., Eds.; American Chemical Society: Washington DC, 1989; pp 210-235.

21. Browett, W. R.; Stillman, M. J. *Prog. Anal. Spectrosc.* 1989, **12**, 73.

22. Browner, R. F.; Boorn, A W. *Anal. Chem.* 1984, **56**, 786A.

23. Posta, J.; Berndt, H. *Spectrochimica Acta* 1992, **47B**, 993

24. Cresser, M. S. *Prog. Analyt. Atom Spectrosc.* 1982, **5**, 35

CHAPTER 3

DIAGNOSING CAUSES OF PROBLEM DATA

3.0 Introduction

Rule based expert systems have been widely adopted as a method of representing knowledge in diagnostic expert systems. Expert systems have been written using declarative languages, such as PROLOG (1-3) and conventional languages, such as FORTRAN (4), which are based on knowledge expressed in the form of rules. An expert system for method development in liquid chromatography, ESCA, developed by Buydens and coworkers (5) is one of the predominant rule based expert systems in the field of analytical chemistry because it provides an understanding of the steps involved in building an expert system. The authors evaluated several expert system shells and used medium sized shells (KES, Nexpert and Goldworks) for their expert system. In KES, the knowledge is represented in the form of attributes and rules (6), while in Goldworks knowledge representation is in the form of frames and rules (7). Nexpert is written in the C language; it has a rule based and an object based inference engine. Other examples of rule based expert systems used in analytical chemistry include determination of trace metals by voltammetry. Esteban et al. (8) have encoded knowledge related to the appearance of peaks, their characteristic potentials, and shape and width of peaks in the form of ***if..then*** rules. KES was used as the expert system shell to encode the chemical knowledge.

However, there are limitations when expert systems are developed based on the input of expert knowledge in the form of rules. The developers of PIRExS, an expert system designed to predict inorganic reactions, which was written in PROLOG, discovered that formulating rules to cover different types of reactions was not an easy task (9). On the other hand, there are expert system tools that generate a set of rules based on a number of examples (10). These tools are referred to as inductive expert system shells. Inductive systems simplify

development of expert systems by generating rules from information assembled in a tabular form.

AAdiagnosis solves instrumental and chemical problems associated with flame atomic absorption spectrometry (FAAS). Knowledge acquisition in AAdiagnosis is based on machine learning, the generation of rules from examples. This expert system, which is a module of AAexpert, is capable of determining causes of "unacceptable data" following assessment of the absorption profile by the quality control module AA-Quality Control. If the problem detected by AAdiagnosis is associated with tasks that the autosampler can rectify without bringing the analysis to a halt, the automated analysis is continued after making the necessary changes. On the other hand, the problem detected by AAdiagnosis may call for a shut down of the analysis. In such a case, the appropriate actions must be taken prior to remeasurement of blanks, standards and samples.

3.1 Model for a Diagnostic Expert System

Knowledge required to diagnose problems associated with an analytical technique is often presented in the form of rules It is not easy to expand this type of rulebase because the order of rules often governs the efficiency of the expert system. On the other hand, if the knowledge can be acquired in the form of a matrix that can subsequently be coded into rules, handling modification and expansion of the knowledge base becomes an easier task

Manipulation of the knowledge matrix will then lead to an improved set of rules in the knowledge base. This technique of encoding knowledge in the form of case histories enables the analyst to write out the necessary knowledge in an organised form on paper during development of the expert system As part of this interaction, the analyst can carry out further experimental tests which are used to reduce ambiguities in the knowledge The value of this approach is that the knowledge base can hold the cumulative knowledge of several experts and this knowledge is offered in a concise and efficient way

3.2 The Knowledge Domain

The role of AAdiagnosis in AAexpert is to diagnose problems associated with the measurement of analytical data. The list of symptoms and associated causes was generated (a) experimentally and (b) from the operation manual of the Varian AA-875 Atomic Absorption Spectrophotometer (11-13).

The symptoms and causes that form the knowledge base were initially verified by Jim Stanton in Dr. Stillman's laboratory using the same Varian AA-875 Atomic Absorption Spectrophotometer (4th year thesis, 1988). In this present work four problems were extensively studied. These were:

a. The effect of a blocked burner;
b. The effect of a viscous solution;
c. The effect of a blocked capillary tube, and
d. The effect of a long capillary tube.

These effects have been discussed in section 2.9 in chapter 2 of this thesis.

Table 3.1 lists the symptoms and the related causes encoded in AAdiagnosis. The symptoms are subdivided into the following areas:

a. Symptoms related to the quality of data;
b. Symptoms related to the absorption profile;
c. Symptoms that are observable by the user;
d. Symptoms related to problems with the status of the instrument, and
e. Symptoms related to problems with the set-up conditions of the instrument.

In our method, the chemical knowledge is acquired as a matrix of symptoms and causes. The matrix is created when the developer assembles the symptoms and identifies possible causes. The matrix serves as a platform to illustrate the chemical knowledge and the logical connections between symptoms and causes gives rise to the rule base. The method of encoding chemical knowledge in a matrix provides a robust technique that is generally applicable to encoding chemical knowledge for use in other advisory expert systems.

3.2 The Knowledge Domain

The role of AAdiagnosis in AAexpert is to diagnose problems associated with the measurement of analytical data. The list of symptoms and associated causes was generated (a) experimentally and (b) from the operation manual of the Varian AA-875 Atomic Absorption Spectrophotometer (11-13).

The symptoms and causes that form the knowledge base were initially verified by Jim Stanton in Dr. Stillman's laboratory using the same Varian AA-875 Atomic Absorption Spectrophotometer (4th year thesis, 1988) In this present work four problems were extensively studied. These were:

a. The effect of a blocked burner;
b. The effect of a viscous solution;
c. The effect of a blocked capillary tube, and
d. The effect of a long capillary tube.

These effects have been discussed in section 2.9 in chapter 2 of this thesis.

Table 3.1 lists the symptoms and the related causes encoded in AAdiagnosis. The symptoms are subdivided into the following areas:

a. Symptoms related to the quality of data;
b. Symptoms related to the absorption profile;
c. Symptoms that are observable by the user;
d. Symptoms related to problems with the status of the instrument, and
e. Symptoms related to problems with the set-up conditions of the instrument.

In our method, the chemical knowledge is acquired as a matrix of symptoms and causes. The matrix is created when the developer assembles the symptoms and identifies possible causes. The matrix serves as a platform to illustrate the chemical knowledge and the logical connections between symptoms and causes gives rise to the rule base. The method of encoding chemical knowledge in a matrix provides a robust technique that is generally applicable to encoding chemical knowledge for use in other advisory expert systems.

Table 3 1 The list of symptoms and causes encoded in AAdiagnosis. The * indicates symtoms associated with the instrument (cont'd on pages 81-83).

SYMPTOMS	CAUSES
High relative standard deviation of replicate measurements.	Contamination from previous samples. Improperly positioned glass bead. Contamination in the spray chamber or burner slot. Precipitation in the nebulizer. Incomplete mixing of sample and fuel. Too few atoms in the flame due to formation of precipitate over time. Background absorption too high-probably due to high salt concentration. Liquid build up in the spray chamber. Partially blocked burner slot. Blockage in the capillary tube. Formation of bubbles in the capillary tube. Use of viscous samples. Unstable lamp. Integration time too small.
Absorbance values are 50-80% of the expected values.	Improperly positioned glass bead. Contamination in the spray chamber or burner slot. Precipitation in the nebulizer. Incomplete mixing of sample and fuel. Background absorption too high-probably due to high salt concentration. Partially blocked burner slot. Blockage in the capillary tube. Use of viscous samples. Burner obstructing light path. Formation of bubbles in the capillary tube.

SYMPTOMS	CAUSES
Absorbance values are 20-40% of the expected values.	Burner rotated by 90 degrees. Too few atoms in the flame due to formation of precipitate over time. Unstable lamp. Wavelength error. Improper gas mixture.
Fluctuations observed in the absorption profile.	Scale expansion set high. Improperly positioned glass bead. Too few atoms in the flame due to formation of precipitate over time. Liquid build up in the spray chamber. Partially blocked burner slot. Background absorption too high-probably due to high salt concentration. Blockage in the capillary tube. Formation of bubbles in the capillary tube. Unstable lamp. Integration time too small. Low pressure in the fuel cylinder
Occasional pulse up or down observed in the absorption profile.	Contamination in the spray chamber or burner slot
Decaying signal observed in the absorption profile.	Precipitation in the nebulizer.
Pulsating signal observed in the absorption signal.	Incomplete mixing of sample and fuel. Liquid build up in the spray chamber.
Spike at the beginning of peak region.	Contamination from previous samples
Long rise time observed in the absorption profile.	Check the viscosity of the solution

SYMPTOMS	CAUSES
Calibration graph levels off at high concentrations.	Few atoms in the flame due to formation of precipitate over time. Contamination in the spray chamber. Precipitation in the nebulizer. Blockage in the capillary tube. Background absorption is high- probably due to high salt concentration. Use of viscous samples. Blockage in the capillary tube. Improper gas mixture.
The calibration curve is sigmoidal.	Ionization in sample.
There is a change in absorption sensitivity between recalibration.	Contamination from previous samples. Contamination in the spray chamber. Precipitation in the nebulizer. Incomplete mixing of sample and fuel. Liquid build up in the spray chamber. Partially blocked burner slot. Background absorption too high-probably due to high salt concentration. Too few atoms in the flame due to formation of precipitate over time. Blockage in the capillary tube. Use of viscous samples. Ionization in sample. Unstable lamp. Wavelength error.
Drift in baseline.	Liquid build up in the spray chamber. Unstable lamp. Low pressure in the fuel cylinder. N_2O valve icing up.
The flame has a ragged appearance.	Burner slot is partially blocked. Contamination in the spray chamber or burner slot. Precipitation in the nebulizer. Liquid build up in the spray chamber.

SYMPTOMS	CAUSES
Bubbles form in the capillary tube.	Presence of dissolved gases in the sample. Improper insertion of the capillary tube in the sample being aspirated.
*Integration time is more than 3 seconds.	Integration time is too small.
* An oxidising flame is being used.	Improper gas mixture. Element affecting absorbance.
*Pressure is higher than 70 psi in the fuel cylinder.	Check pressure in the fuel cylinder.
*The burner is set below 8 on the vertical scale.	Check to make sure that the burner is not obstructing light.
* A N_2O-C_2H_2 flame is being used.	N_2O valve icing up.
A refractory element is present in the solution.	Element affecting absorbance.
The uptake rate is more than 5 mL/min.	Blockage in the capillary tube. Precipitation in the nebulizer. Too few atoms in the flame due to formation of precipitate over time.
Fluctuating display of lamp intensity while optimising lamp.	Unstable lamp.
* The hollow cathode lamp is on but there is no display of lamp intensity while optimizing lamp.	Wavelength error. Burner obstructing light path

3.3 Choice of the Building Tool

In the design of a diagnostic expert system, the developer may choose either a software system that requires that the expert knowledge is entered in the form of rules or in the form of case histories. The software systems chosen by our group require knowledge to be encoded in the form of conditions and case conclusions (essentially case histories). The conditions may be thought of as symptoms of a problem and the conclusions can be related to the likely cause of the problem.

The knowledge base of symptoms and causes, and the resultant rulebase, is general for all flame AAS analyses and so this knowledge base can be applied to other systems. In our approach we connect the instrumental set-up conditions and sample chemical properties, to the analytical symptoms through logical relationships established at the primary and secondary level of knowledge development.

3.3.1 Use of KDS : A Commercially Available Expert System Shell

The expert system shell KDS (from KDS Corporation, 934 Hunter Road, Wilmette, IL 60091, USA) was chosen as the shell within which our expert system was first developed. KDS is an inductive expert system shell comprising a development component and a run time component. The development component allows information to be entered in form of case histories broken down into conditions and conclusions. The run-time component permits the distribution of a compiled version to users. KDS works in the sequence of facts...rules...knowledge base. This may be contrasted with subjective systems that evolve as rules...knowledge base. Facts are generally objective, whereas explicitly entered rules are subjective, since they consist of the opinions of one or more people. It is often difficult to input a large number of logically correct rules in a knowledge base that requires chemical information to be entered in the form of rules. The KDS shell has the following characteristics:

a. Human knowledge is encoded in the form of a matrix of facts comprising conditions and conclusions. Compilation of the facts results in conditional if-then rules;

b. The problem-solving skill of the program increases at a rate proportional to the enlargement of the knowledge base;

c. Solutions to a wide range of complex problems are possible by the use of the inductive nature of the matrix of facts and the deductive nature of the rules, and

d. Conclusions are explained by retracing the lines of reasoning

3.3.2 Use of EAshell: A Shell Designed and Coded in our Laboratory

EAshell is a Windows (3.1) based expert system shell comprising three separate modules; namely, EAengine, Rule_Editor, and Table_Generator EAshell was developed in our laboratory using Microsoft C (version 7.0). Microsoft Visual Basic (version 3.0) can serve as the programming language and the user interface for EAshell. The Table_Generator is used to express the chemical knowledge in the form of a matrix of symptoms (rows) and causes (causes). The logical connections that are used to encode knowledge are true (T), false (F), or unknown (?). A blank cell represents no connection between a cause and a symptom. We have encoded knowledge as 100% true and 100% false. The rulebase is generated by the program Table_Generator. The chemical knowledge is expressed in the form of ***if..then*** rules connecting the symptoms (the premise in the ***if*** part of the rule) associated with a particular cause (the conclusion or action in the ***then*** part of the rule). The Table_Generator can only use AND as a logical connector in the premise of the rule. In addition, the Table_Generator cannot create any rules containing the AND connector in the conclusion part of the rule. The Rule_Editor converts the knowledge base created by the Table_Generator into a form that can be used by EAengine, the inference engine. The Rule_Editor can be used to write rules containing logical connectors (AND/OR) and relational operators in the premise

of the rule. Rule_Editor can also generate rules containing AND as a connector in the conclusion of the rule.

EAshell contains functions and subroutines that are used as dynamic link libraries. Some of the important functions of EAshell are as follows:

a. LoadKBF: loads the knowledge base file to be used for inference processes;
b. WELoadFactFile: loads the fact from the disk (input from the user);
c. WELoadFactData: loads the facts from the application (via the user interface);
d. WESetGoal: used to set the goal for goal driven inference processes;
e. WEinfProcess: performs inference processes;
f. WEGuess: performs inference based on insufficient information;
g. WEDisplayInfResult: displays the results of inference processes, and
h. WEGetLogFileName: returns the contents of the log file (that contains the logic behind the inference process).

3.4 The User Interface

The role of a user interface is very important in these expert systems. Graphical images help the user understand the symptoms described by the developer of the program. Graphical images are more crucial in resolving the ambiguity often associated with the symptoms related to problems encountered during analyses. The interface in the run-time module of the KDS shell is not user friendly and forces the user to rely more on textual information. Incorporation of graphical images into the KDS shell, especially those related to the quality of the absorption profile, is difficult. Visual Basic serves as a user interface for the expert system based on EAshell. The uniformity of a Windows-based user interface in all the modules of AAexpert makes the front end of the expert system more accessible as well as more attractive and, therefore, likely to succeed as a commercially available expert system.

3.5 Rule Generation in Rule-based Expert Systems

The simplest rule-based languages are similar in expressive power to propositional logic (14). The inference engine of an expert system uses a theorem from propositional logic called Modus Ponens which states that if we know A implies B and we are given that A is true, we can conclude B is true Thus, the inference engine repeatedly applies Modus Ponens to a set of rules to solve a particular problem (15). Rules in conventional languages like C, PASCAL and the artificial intelligence language LISP, are based on propositional logic. PROLOG and OPS5 express knowledge in predicate logic where the facts can be more complex than in a propositional logic system. Predicate logic has in addition to the connectivities of propositional logic the ability to decompose a rule into a predicate and a number of arguments Predicate logic includes propositional logic, therefore a predicate logic system can run rules from a propositional logic system (14).

In contrast to conventional programming environments that support only the Boolean logic operations, expert system shells offer various models for performing logical operations with uncertain information The mineral prospecting expert system PROSPECTOR used the probabilistic model based on Bayes's conditional probability theorem of combining evidence (16). MYCIN uses probability-like measures of certainty and reliability and a specialized inexact reasoning mechanism for inference (17).

The order of rules in an expert system where the knowledge base is formed directly from rules must be governed by the inference strategy used by the inference engine (18). In forward chaining strategy, conclusions that can be inferred from fewer number of facts should be written prior to rules that require a greater number of facts. Also, commonly occurring symptoms should appear early in the rulebase since the inference engine queries the rules in the order given by the developer. In the backward chaining strategy, highest level rules

should be written first since the inference engine works backwards to lower level rules in order to reach the goal requested by the user.

3.5.1 Rule Generation in KDS shell

The KDS shell handles uncertainty by providing the option of entering "?" (meaning "don't know") when the end user has doubts as to what the correct answer should be. The KDS shell uses a proprietary uncertainty resolving algorithm to provide logically reliable results.

The KDS inference engine analyzes the matrix of facts and generates a set of rules. We have chosen the conservative rules mode for the generation of rules. The following types of rules are generated by the system:

a. If the subject (*is true*) then the predicate (*is false*);

b. If the subject (*is true*) then the predicate (*is true*), and

c. If the subject (*is false*) then the predicate (*is false*).

A rule of the type **if the subject (*is true*) then the predicate (*is false*)** is generated when the two symptoms, namely the subject and the predicate of the rule, are not associated with any common cause. For example, consider the following rule.

> **Rule 1. If the absorbance values are 20-40% of the expected values in the calibration graph (*is true*) then the signal is pulsating (as observed from absorption profile) (*is false*).**

Explanation The causes associated with a pulsating signal namely (i) improper mixing of the sample and fuel and (ii) liquid build up in the spray chamber, do not specifically result in absorbance values that are 20-40% less than the expected values from the calibration graph.

A rule of the type **if subject (*is true*) then the predicate (*is true*)** is generated when the subject and the predicate of the rule are associated with the same set of causes Consider the following rules.

Rule 2. If the flame has a ragged appearance (*is true*) then the relative standard deviations are high (>5%) and absorbance values are not precise (*is true*).

Explanation Contamination in the spray chamber or burner slot and/or precipitation or blockage in the capillary tube or the nebulizer, will give rise to a flame with a ragged appearance and high relative standard deviations will be measured from the absorbance values.

Rule 3. If bubbles form in the capillary tube (*is true*) then abnormally noisy signal (observed from absorption profile) (*is true*).

Explanation The presence of dissolved gases and/or the improper insertion of the capillary tube into the solution being aspirated, will also give rise to an abnormally noisy absorption signal.

Rule 4. If low pressure in the fuel cylinder (*is true*) then abnormally noisy absorption signal (*is true*).

Explanation Pressure lower than 70 psi in the acetylene cylinder will result in acetone being delivered to the flame giving rise to an abnormally noisy absorption signal.

In the conservative mode of rule generation, any symptom that is not related to every cause cannot be the subject of a rule. In this mode a rule of the type **if subject (*is false*) then predicate (*is false*)** is generated having the same logic as **if the subject (*is true*) then predicate (*is true*)** provided the predicate of this rule is completely filled in for all the cases Consider the following rule.

Rule 5. If the relative standard deviation of the measurements are high (>5%) and the absorbance values are not precise (*is false*) then occasional pulse up or down in the signal (as observed from the absorption profile) (*is false*).

The above rule is the same as "if occasional pulse up or down in the signal (as observed from absorption profile) (*is true*) then the relative standard

deviation of the measurements are high (>5%) and the absorbance values are not precise (*is true*)" This rule with **"occasional pulse up or down in the signal (as observed from absorption profile)"** as the subject of the rule is not generated because this symptom is not related to the cause **"nitrous oxide valve icing up"**.

3.5.2 Rule Generation in EAshell

The inference engine in EAshell is able to identify logical and relational operators in the premise of the rule. In a rule whose premise contains both OR and AND, OR takes precedence over AND. In the rule shown below, in order to reach the result with a confidence factor of 100, the user has to respond correctly to one of the symptoms connected by OR followed by a correct response to ***"fluctuations are observed in the absorption signal"***.

Rule: If the abs values are 50-80% of the expected values =True OR the abs values are 20-40% of the expected values =True AND fluctuations are observed in the absorption profile=True THEN Contamn_Blockage= Background absorbance is high-probably due to high salt content.

The developer often codes the same cause as giving rise to more than one combination of symptoms. It is more likely that the user will reach a conclusion with a confidence factor of 100 when multiple sets of symptoms are associated with the same cause

For example, consider the following rule:

Rule: IF RSD high=True AND Spike at beginning of peak region=True AND Ragged flame=True THEN Contamn_Blockage=Contamination from previous samples.

The above rule is generated because the knowledge in the matrix indicates that "contamination from samples" results in the following symptoms:

a. Relative standard deviation (RSD) of the measurements are high (> 5%) and the absorbance values are not precise,
b. A spike at the beginning of the peak region of the absorption profile, and
c. The flame has a ragged appearance.

Other combinations of symptoms arising from ***"contamination from previous samples"*** are listed below.

Rule: IF RSD high=False AND Spike at beginning of peak region=True AND Ragged flame=True THEN Contamn_Blockage=Contamination from previous samples.

Rule: IF RSD high=True AND Spike at beginning of peak region=True AND Ragged flame=False THEN Contamn_Blockage=Contamination from previous samples.

Rule: IF RSD high=False AND Spike at beginning of peak region=True AND Ragged flame=False THEN Contamn_Blockage=Contamination from previous samples.

The rules related to the cause "contamination from previous samples" show that of the three symptoms associated with it, the symptom "a spike at the beginning of the peak region of the absorption profile" is the characteristic symptom. A knowledge subdomain like this can be further expanded if other symptoms are found to be associated with the same cause The rule base was thus generated using the Table_Generator keeping in mind the various combinations of symptoms that could arise from different causes

The clicking of " ? (UNKNOWN)" (which is offered as a button in EAshell along with "OK" and "WHY") does not require the developer to incorporate UNKNOWN as a option for variables in the rule base If the variable being prompted by the engine only has TRUE and FALSE in the OPTION statement and the user clicks the " ? (UNKNOWN) " all the rules containing the variable fail. If the user chooses UNKNOWN from the OPTION section of the query and

if the value of the variable has been encoded as UNKNOWN by the developer, the rule is fired.

3.6 Inference Strategies in AAdiagnosis based on KDS Shell

In the run-time module, the compiled form of the fact matrix is evaluated inductively against the facts supplied by the end user. Each time the system learns a fact from the user's response to a question, the entire rule set is searched and all applicable rules are fired and the best possible answer is offered to the user. The expert system offers a conclusion based on the responses given by user and presents the line of reasoning to the user at the end of each consultation. Inference strategies in AAdiagnosis are based on either one of the following:

Forward chaining combined with a breadth-first search

The user's response starts with an answer to the watershed condition question then the symptoms are presented in the same order as compiled by the developer. For example, the conclusion ***burner slot is partially blocked*** (see Figure 2.14) is reached based on the following input by the user :

a. Relative standard deviation of the measurements is high (>5%) and the absorbance values are not precise (is true);
b. Abnormally noisy absorption signal (observed from absorption profile) (is false);
c. Absorbance values are 50-80% of the expected values in the calibration graph (is true);
d. The uptake rate is less than 5mL/min. (is true), and
e. The calibration graph levels off at high concentrations (is false).

User-directed chaining with breadth-first search

The user is able to select the conditions that most closely matches the problem which means that the solution is offered after fewer responses. The conclusion ***"the burner slot is partially blocked"*** is reached based on the user's responses which are as follows:

a. Absorbance values are 50-80% of the expected values in the calibration graph (is true);
b. The uptake rate is less than 5mL/min (is true);
c. The calibration graph levels off at high concentrations (is false), and
d. The flame has a ragged appearance (is true).

3.7 Inference Strategies in AAdiagnosis based on EAshell

EAshell can use three inference strategies to search the AAdiagnosis rule base, namely

1. Forward Chaining, Depth-First Search.
2. Mixed Chaining, Depth-First Search.
3. Backward Chaining, Depth-First Search.

Forward Chaining, Depth-First Search

In this strategy, the user offers no preconceived notion regarding the area of problem. All the observable symptoms are selected (true, false) by the user The inference engine gathers the facts and fires all the rules that contain at least one correct response. The confidence factor (cf) associated with every solution indicates the percent of symptoms (for the premise of the particular rule) correctly answered by the user. The advantage of encoding knowledge by combining only a few symptoms to give a rule becomes evident in that the solution offered is associated with a high confidence factor and the engine can handle incomplete information better.

Example

Observations:

a. The absorbance values are 50-80% of the expected values;
b. The flame has a ragged appearance,
c. There are no fluctuations observed in the absorption profile,
d. There is no occasional pulse observed in the absorption profile, and
e. No pulses are observed in the peak region of the absorption profile

The forward chaining, depth-first search strategy results in the following information. The solutions offered by EAshell using the forward chaining, depth-first search and the associated interpretation are shown in Table 3.2.

Backward Chaining, Depth-First Search

In this strategy, the rules are tested in a chronological order based on the goal requested by the user. The engine prompts the user for a response to the primary variable appearing in the first rule. If the user's response does not result in the rule being satisfied, the inference engine requires the user to respond to a different variable in order to satisfy the premise of another rule. A rule is said to fire when the premise of any rule is correctly responded to by the user. Therefore, the only answer presented to the user is the cause whose symptoms are 100% satisfied (all correct responses) by the user. It is important to avoid dead-end situation by incorporating rules of the following type at the end of each goal.

Rule: IF Do you have any more information=False THEN

Contamn_Blockage= Cannot detect any problems.

Rule: IF Do you have any more information=False THEN

Contamn_Blockage= Try again-verify symptoms.

Example:

Knowledge: A problem has been detected and it is related to Contamination/Blockage of the atomic absorption spectrometer.

Therefore, Contamination/Blockage is chosen as the problem domain. The goal (Contamn_Blockage) is chosen by the user. Next, the engine prompts the user for input and the following symptoms are selected by the user:

a. The absorbance values are 50-80% of the expected values;
b. The flame has a ragged appearance;
c. There are no fluctuations observed in the absorption profile;
d. There is no occasional pulse observed in the absorption profile, and

Table 3.2 Solutions offered by EAshell's forward chaining, depth-first search strategy and the interpretations. The cf values indicate the percent of symptoms (in the presmise of a rule) correctly answered by the user.

Solution offered by EAshell	Interpretation
cf=100, Contamn_Blockage=Burner slot is partially blocked.	Aspiration of a solution into a partially blocked burner slot gives rise to (a) a ragged flame, and (b) low (50-80%) absorbance values.
cf=66, Contamn_Blockage=Contaminated spray chamber or burner slot AND formation of precipitate in nebulizer.	The symptoms of (a) a ragged flame, (b) low (50-80%) absorbance values and (c) high relative standard deviation values, arise (i) from contamination in spray chamber or burner slot, and (ii) formation of precipitate over time.
cf=50, Sample_soln_prob=Incomplete mixing of sample and fuel.	Incomplete mixing of sample and fuel is associated with (a) pulsating signal, (b) ragged flame, (c) high relative standard deviation values, and (d) low (50-80%) absorbance values.
cf=50, Contamn_Blockage=Liquid build up in the spray chamber.	Liquid build up in the spray chamber gives rise to (a) a pulsating signal, and (b) a ragged flame.
cf=33, Sample_soln_prob=Few atoms in the flame due to the formation of precipitate over time.	Few atoms in the flame due to the formation of precipitate over time gives rise to (a) low absorbance (20-40%) values, (b) a ragged flame, and (c) a calibration graph that levels off at high concentrations.
cf=33, Optimize_prob=Improperly positioned glass bead.	An improperly positioned glass bead could result in (a) high relative standard deviation values of replicate measurements, (b) low (50-80%) absorbance values, and (c) fluctuations in absorption signal.
cf=33, Optimize_prob=Improperly positioned glass bead.	An improperly positioned glass bead could result in (a) high relative standard deviation values of replicate measurements, (b) low (50-80%) absorbance values and (c) change in absorption sensitivity on recalibration.
cf=25, Contamn_Blockage=Element present in solution that affects absorbance.	An effect in absorbance due to the presence of an element is associated with (a) low (50-80%) absorbance values, (b) change in absorption sensitivity on recalibration, (c) a calibration graph that levels off at high concentrations, and (d) presence of a refractory element in the solution.

e. No pulses are observed in the peak region of the absorption profile.

The inference engine performs a backward chaining with depth-first search and the solution offered to the user is shown below.

cf=100, Contamn_Blockage=Burner slot is partially blocked.

Mixed Chaining, Depth-First Search

This is a combination of both forward and backward chaining strategies used for a single inference process. The inference engine begins a forward chaining strategy and if the needed switches to backward chaining strategy. On completion of the backward chaining search, the engine switches back to the forward chaining strategy.

Chemical Knowledge: Aspiration of a viscous solution results in a uptake rate less than 5.0 mL/min and produces an absorption profile with a long rise time.

Rule 1: If Absorption profile has a long rise time=True AND Uptake rate is less than 5.0 mL/min=True THEN sample_soln_prob=Check viscosity of the solution.

Rule 1 can be expressed as follows:

Rule 2: If Absorption profile has a long rise time=True AND Uptake rate=Low THEN sample_soln_prob=Check viscosity of the solution.

The inference process begins in the forward chaining, depth-first search mode starting with the first variable in the initial rule appearing in the rulebase. If the above rule is the first rule in the rulebase, the system requires the user to respond to ***"Absorption profile has a long rise time"***. If the response is correct, the system continues the same inference strategy and requires a response to ***"Uptake rate"***. At this stage, the inference engine switches to backward chaining, depth first search because ***"Uptake rate=Low"*** appears as a conclusion in the following rules.

Rule 3: If Aspiration rate > 5.0 THEN Uptake rate=Normal

Rule 4: If Aspiration rate < 5 .0 THEN Uptake rate=Low.

ASK "Aspiration rate (mL/min) ?"

If the user's response is less than 5.0 (mL/min) then the engine concludes that ***"uptake rate=Low"***. The engine switches back to forward chaining, depth-first search mode and the solution ***"check viscosity of the solution"*** is offered to the user.

3.8 Acquisition of Knowledge using KDS shell

The KDS knowledge base is represented as the logical connections between facts labeled as conditions and conclusions.

The first condition that is entered is called the watershed condition. It is very important that the user finds this symptom easy to understand and can answer the question concerning this condition without uncertainty. The watershed symptom must be able to divide the causes that have been identified by the developer into major categories. We have chosen ***"the relative standard deviations of the measurements are high (>5%), the absorbance values are not precise"*** as our watershed condition (symptom). On entering the watershed symptom or condition, the system requires the developer to specify a cause or case conclusion that could be true when the watershed condition is true. Next, the developer specifies a conclusion that could be true when the watershed condition is false. Knowledge is added when the developer adds a new cause followed by a symptom in order to discriminate between the existing causes and the new cause. This step in knowledge encoding is referred to as the primary level.

Figure 3.1 shows the logical connections between conditions and case conclusions at this primary level. This figure describes the order in which the logical connectivities between the symptoms or conditions (on the left) and the causes or case conclusions (on the right) were set up Symptoms that describe poor analytical quality (symptoms 1, 2, 3, 8 and 11) relate to the numerical data. Symptoms 4, 5, 6 and 7 directly relate to the analytical measurement process

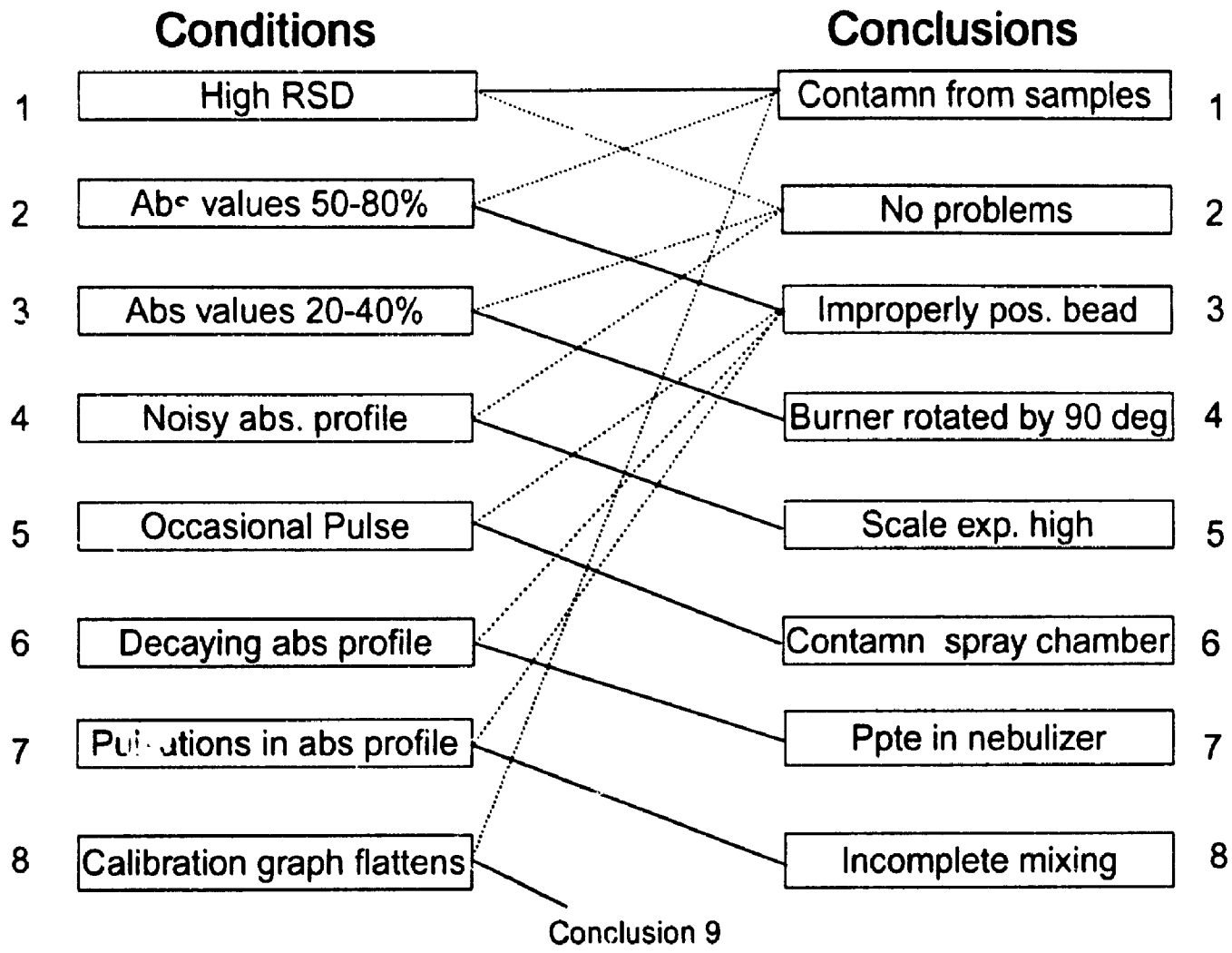
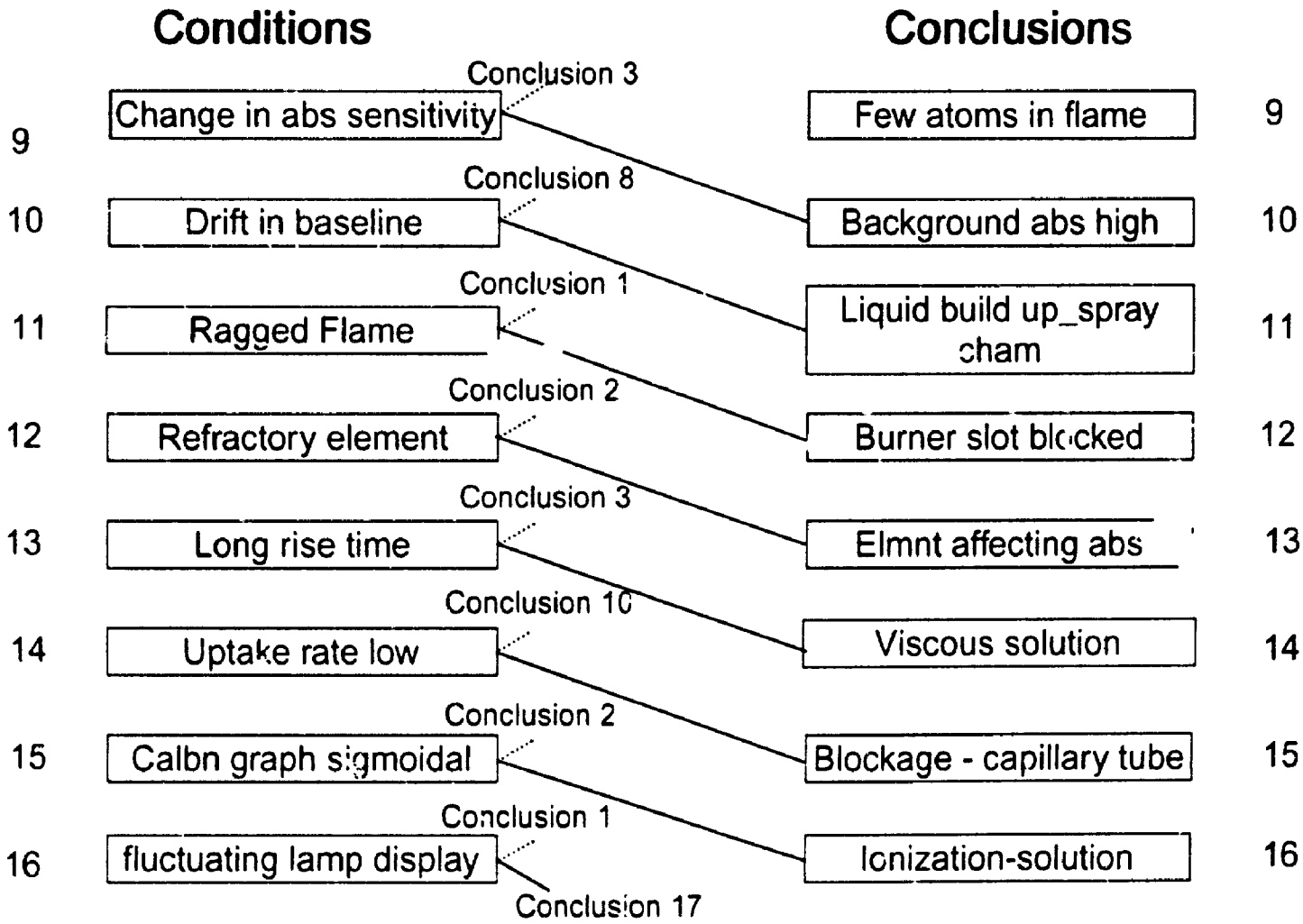
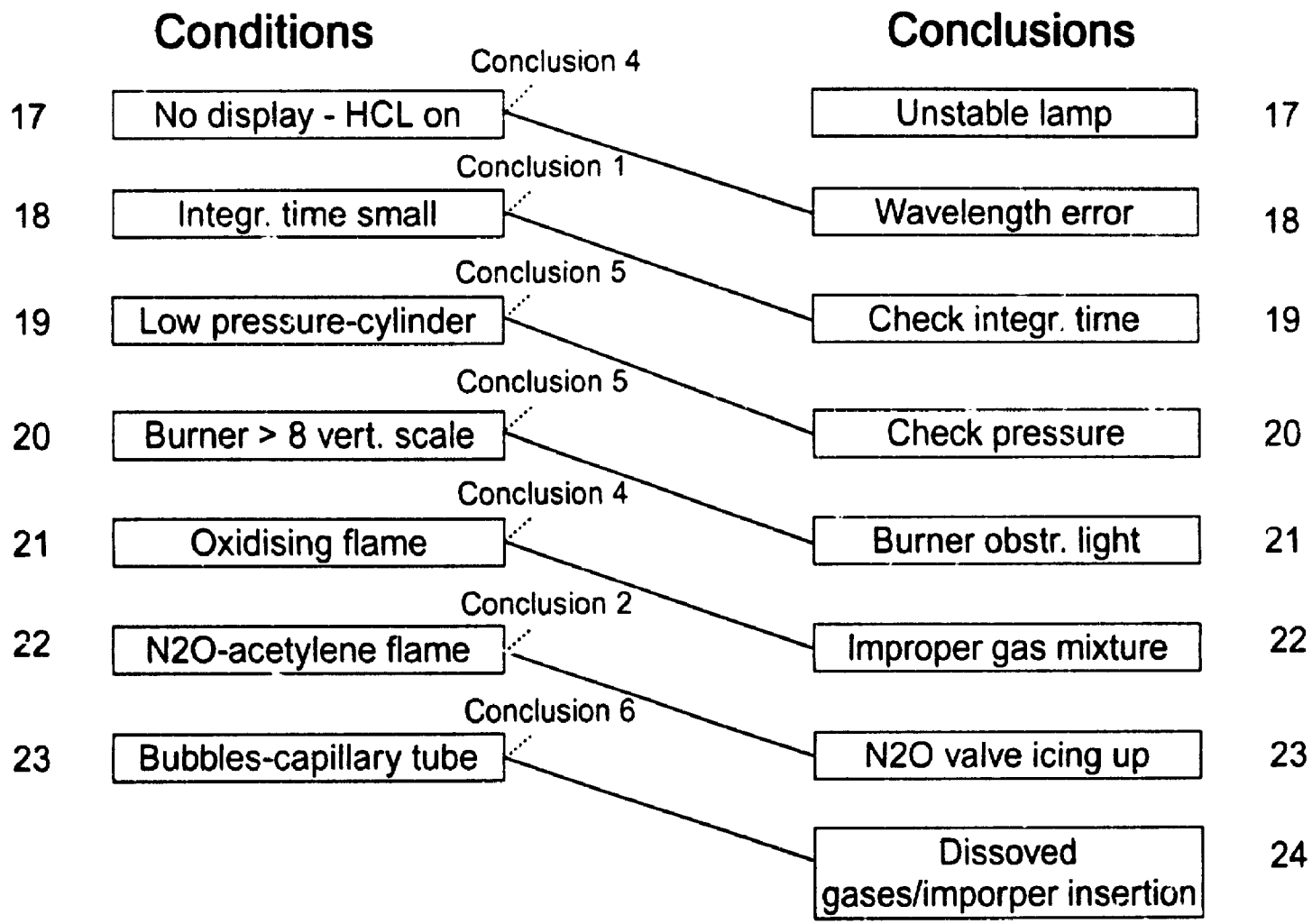


Figure 3.1 Logic developed at the primary level. A "true" relationship is indicated by a solid line and a false relationship is indicated by a dashed line (cont'd on pages 99-100).


Conditions
Conclusions
Conclusion 3
Change in abs sensitivity
Few atoms in flame
9
9
Conclusion 8
Drift in baseline
Background abs high
10
10
Conclusion 1
Ragged Flame
Liquid build up_spray cham
11
11
Conclusion 2
Refractory element
Burner slot blocked
12
12
Conclusion 3
Long rise time
Elmnt affecting abs
13
13
Conclusion 10
Uptake rate low
Viscous solution
14
14
Conclusion 2
Calbn graph sigmoidal
Blockage - capillary tube
15
15
Conclusion 1
fluctuating lamp display
Ionization-solution
16
16
Conclusion 17


Conditions
Conclusions
Conclusion 4
17
No display - HCL on
Unstable lamp
17
Conclusion 1
18
Integr. time small
Wavelength error
18
Conclusion 5
19
Low pressure-cylinder
Check integr. time
19
Conclusion 5
20
Burner > 8 vert. scale
Check pressure
20
Conclusion 4
21
Oxidising flame
Burner obstr. light
21
Conclusion 2
22
N2O-acetylene flame
Improper gas mixture
22
Conclusion 6
23
Bubbles-capillary tube
N2O valve icing up
23
Dissoved gases/imporper insertion
24

and are not traditionally treated statistically as these data represent the detector signal recorded in real time as the sample enters the flame. Processing these data is essential in a totally automated system because these measurements replace input from an operator who will usually be aware of the overall system performance from experience that is based on both visual and aural inputs during sample introduction.

The secondary level of knowledge is completed when conditions and case conclusions not related at the primary level are connected logically. Thus knowledge is represented in the form of a matrix of facts in which columns are represented by case conclusions and rows are represented by conditions. The knowledge in the matrix is said to be completely **"filled"** when the developer does not leave any condition unrelated to a conclusion The secondary level of encoding knowledge is crucial in the functioning of a diagnostic expert system as the knowledge is expanded by this step. Rules are produced only after compiling the secondary level of knowledge.

The final level of knowledge is added to the module when the developer encodes case histories that include causes that already exist in the knowledge base. The combination of symptoms that result from the same cause is different from that encoded at the primary level. These **"parallel cases"** depend on the previous experience of the expert Table 3.3 shows the matrix of knowledge in AAdiagnosis. Unexpected results may arise due to the lack of a particular combination of symptoms at the tertiary level. In such instances, the user is expected to report the new case history to the developer of the expert system who will encode the information. The tertiary level of knowledge helps in creating an expanded knowledge base which takes into account more case histories than is possible based only on the primary and secondary levels of knowledge

	Contamination from samples	Improperly positioned glass	Burner rotated by 90 degrees	Scale expansion set too high	Contaminated spray chamber	Precipitation in the nebulizer	Incomplete mix sample & fuel	Too few atoms in the flame	Background absorption high	Liquid build up-spray chamber	Burner slot blocked
DATA QUALITY CONTROL											
Relative standard dev high	T	T			T	T	T	T	T	T	T
Noisy absorption signal		T		T				T	T	T	
Abs values 50-80%		T			T	T	T		T	T	T
Drift in baseline										T	
Change in absorption sensitivity	T				T	T	T	T	T	T	T
Calbn graph levels off					T	T		T			
Abs values 20-40%			T					T			
Calibration graph sigmoidal											
Flame - ragged appearance					T	T				T	T
Pulsating signal							T			T	
Bubbles form in capillary tube											
Occasional pulse					T						
Decaying signal						T					
Long rise time (absorption profile)										?	
INSTRUMENTAL CONDITIONS											
Pressure <70 psi											
Oxidizing flame	?	?		?	?	?	?	?		?	?
Integration time too small		?	?	?	?				?		?
N2O-acetylene flame	?	?	?	?	?	?	?			?	?
Burner>8 on vertical scale											
No display - HCL on											
Fluctuating display- lamp intensity											
CHEMICAL CONDITIONS											
Uptake rate less than 5 mL/min	?					T		T			
Refractory element present											

Table 3.3 The knowledge matrix in AAdiagnosis using the KDS shell (contd. on pages 103-105).

	Element affecting absorbance	Viscous solution	Blockage in capillary tube	Ionization in sample	Unstable lamp	Wavelength error	Integration time too small	Check pressure in cylinder	Burner obstructing light	Improper gas mixture	Nitrous oxide valve icing up	Bubbles_ cap. tube
DATA QUALITY CONTROL												
Relative standard dev high		T	T		T		T					T
Noisy absorption signal			T		T		T	T		T		T
Abs values 50-80%	T	T	T						T			T
Drift in baseline					T			T			T	
Change in absorption sensitivity	T	T	T	T	T	T		T				
Calbn graph levels off	T	T	T			?				T		
Abs values 20-40%					T	T				T		
Calibration graph sigmoidal				T								
Flame - ragged appearance												
Pulsating signal												
Bubbles form in capillary tube												T
Occasional pulse											?	
Decaying signal												
Long rise time (absorption profile)		T							?			
INSTRUMENTAL CONDITIONS												
Pressure <70 psi								T				
Oxidizing flame	T	?	?	?	?	?	?		?	T		?
Integration time too small	?	?	?	?	?		T		?	?	?	?
N2O-acetylene flame		?	?		?	?	?	?	?	?	T	?
Burner>8 on vertical scale									T			
No display - HCL on						T			T			
Fluctuating display -lamp intensity					T							
CHEMICAL CONDITIONS												
Uptake rate less than 5 mL/min		T	T									T
Refractory element present	T	?	?						?			

	Ionization in sample	Ionization in sample	Ionization in sample	Ionization in sample	Ionization in sample	Check pressure in cylinder	Check pressure in cylinder	Check pressure in cylinder	Check pressure in cylinder
DATA QUALITY CONTROL									
Relative standard dev high									
Noisy absorption signal						T	T	T	T
Abs values 50-80%									
Drift in baseline	T	T				T	T	T	T
Change in absorption sensitivity	T	T	T	T	T		T	T	T
Calbn graph levels off									
Abs values 20-40%									
Calibration graph sigmoidal	T	T	T	T	T				
Flame - ragged appearance									
Pulsating signal						T	T		
Bubbles form in capillary tube									
Occasional pulse									
Decaying signal									
Long rise time (absorption profile)				T	T		T	T	
INSTRUMENTAL CONDITIONS									
Pressure <70 psi						T	T	T	T
Oxidizing flame	?	?	?	?	?		T	T	
Integration time too small	?	?	?	?	?			T	T
N2O-acetylene flame	T					?	?	?	?
Burner>8 on vertical scale									
No display - HCL on									
Fluctuating display- lamp intensity			T	T					
CHEMICAL CONDITIONS									
Uptake rate less than 5 mL/min		T	T	T	T		T	T	
Refractory element present									T

	Bubbles_ cap. tube	Bubbles_ cap. tube	Bubbles_ cap. tube	Contamination from samples	Incomplete mix sample & fuel	Too few atoms in the flame	Too few atoms in the flame	Integration time small	Integration time too small	Improper gas mixture
DATA QUALITY CONTROL										
Relative standard dev high	T	T	T	T	?	T	T	T	T	T
Noisy absorption signal	T	T	T			T	T	T	T	T
Abs values 50-80%	T	T	T		T					
Drift in baseline	T	T								
Change in absorption sensitivity	T	T	T	T	T	T	T			
Calbn graph levels off	T					T	T			T
Abs values 20-40%						T	T			T
Calibration graph sigmoidal		T								
Flame - ragged appearance				T	T	T				
Pulsating signal					T					
Bubbles in capillary tube	T	T	T							
Occasional pulse				T						
Decaying signal										
Long rise time (absorption profile)									T	
INSTRUMENTAL CONDITIONS										
Pressure <70 psi										
Oxidizing flame	?	?	?	?	?	?	?	?	?	T
Integration time too small	?	?	?				T	T	T	?
N2O-acetylene flame	?	?	?	?	?		?	?	?	?
Burner>8 on vertical scale										
No display - HCL on										
Fluctuating display -lamp intensity										
CHEMICAL CONDITIONS										
Uptake rate less than 5 mL/min				?		T	T	T	T	
Refractory element present			T				T	T		

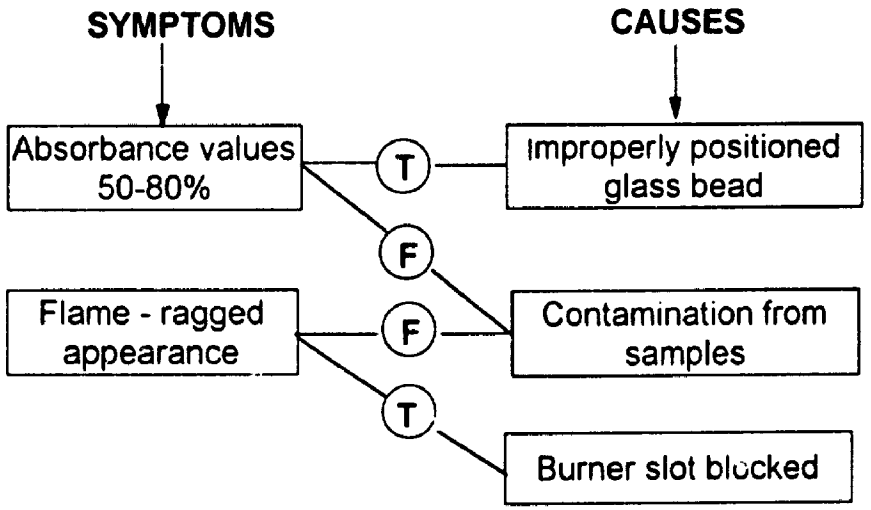
3.9 Encoding Chemical Knowledge using KDS Shell

The steps involved in encoding chemical knowledge can be understood by considering Figure 3.2. At the primary level, the cause ***"improperly positioned glass bead"*** gives rise to the symptom ***"absorbance values are between 50-80% of the expected values in the calibration graph".*** Also, the cause ***"contamination from previous samples"*** did not result in ***"absorbance values between 50-80% of the expected values in the calibration graph"***. Next, the symptom ***"the flame has a ragged appearance"*** was encoded for the cause ***"the burner slot is partially blocked"***.

In the secondary level, a relationship between (a) ***"absorbance values are between 50-80% of the expected values in the calibration graph"*** and ***"the burner slot is partially blocked"*** and (b) ***"the flame has a ragged appearance"*** and ***"improperly positioned glass bead"*** was established.

The table in Figure 3.2 depicts part of the knowledge matrix. In this table, the cause ***"burner slot is partially blocked"*** is associated with the symptom ***"absorbance values are between 50-80% of the expected values in the calibration graph"*** and ***"contamination from previous samples"*** is not connected with low absorbance values.

At the primary level, the symptom ***"the flame has a ragged appearance"*** results from a combination of causes. It is possible that ***"the flame has a ragged appearance"*** can arise from a different combination of causes in which only the logic of one cause differs from that in the primary level. Thus, ***"the flame has a ragged appearance"*** can be associated with a set of causes which includes ***"contamination from previous samples"*** being true. Such causes (cases) are called **parallel cases** and are encoded at the tertiary level. The importance of parallel cases lies in the fact that they enable the user to arrive at the correct answer in situations where it is possible to have one cause giving rise to more than one possible combination of symptoms. Figure 3.3 shows how chemical information can be encoded in diagnostic expert systems in general.



	Improperly positioned glass bead	Contamination from samples	Burner slot blocked	Contamination from samples
Absorbance values 50-80%	* T	* F	** F	*** F
Flame - ragged appearance	** F	* F	* T	*** T
Absorbance values 20-40%	** F	** F	* T	*** F

Logic developed at { * Primary level, ** Secondary Level, *** Tertiary Level

Figure 3.2 Development of logic in AAdiagnosis using the KDS shell.

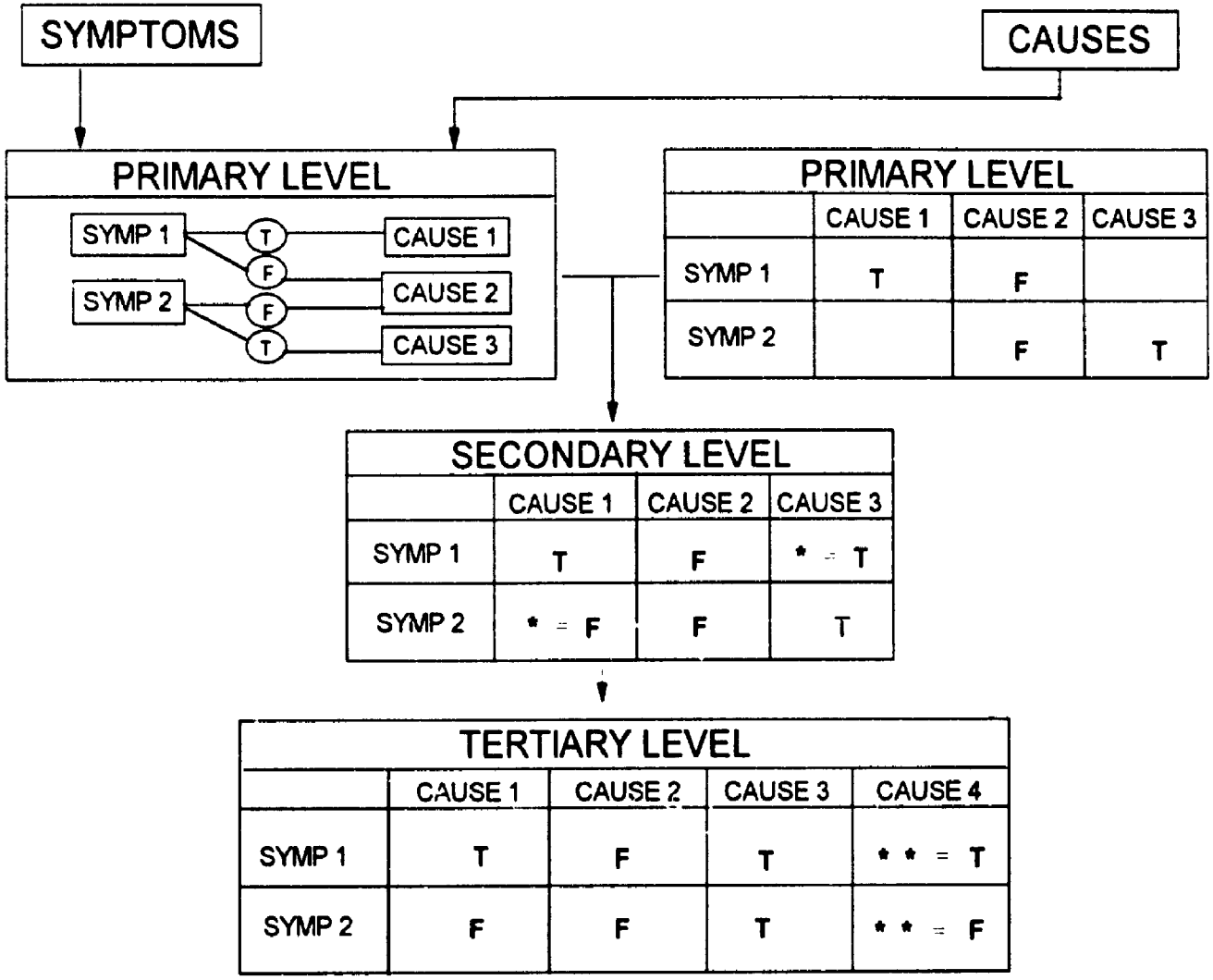


Figure 3.3 Encoding knowledge in AAdiagnosis using the KDS shell. The figure shows the sequence of entering information in AAdiagnosis. This method can be used to generate the knowledge base in any diagnostic expert system that is based on the case history approach.

3.10 Acquisition of Knowledge using Table_Generator of EAshell

The knowledge matrix for EAshell is created when the chemical information is organized as symptoms (rows) and causes (columns) in the Table Generator. However, in this scheme, there is no distinction between primary, secondary and tertiary levels of knowledge. The cells in the matrix are logically connected using T(true) and F (false). In some cases, it was found that a single cause could give rise to different combinations of symptoms. The combination of logical connections leading to a cause (n1) gives rise to a rule of the following type:

Rule: *If* (symptom (n1) =T/F AND symptom (n2)=T/F...) *Then* goal = Cause (n1)

The causes in the knowledge domain have been subdivided into a number of major areas. This breakdown into areas is necessary when goal-directed inference is requested by the user. The chemical knowledge in AAdiagnosis was divided into the following goals:

1 Contamination/Blockage problems (Contamn_Blockage);

2. Sample/solution related problems (Sample_soln_prob);

3 Optimization problems (Optimize_prob), and

4. Fuel problems (Fuel_prob).

The success in firing a rule depends on the user's response to the symptoms that make up the premise of the rule. Often, more than one rule has to be written to express the following type of knowledge.

Rule: If symptom(1)=True THEN cause(1)

Rule: If symptom(2)=True THEN cause(1)

The alternative is to write the information as follows:

Rule: If symptom(1)=True OR symptom(2)=True THEN cause(1)

For example the two individual rules written below,

Rule: If the abs values are 50-80% of the expected values =True AND fluctuations are observed in the absorption profile=True THEN Contamn_Blockage= Background absorbance is high-probably due to high salt content.

Rule: If the abs values are 20-40% of the expected values =True AND fluctuations are observed in the absorption profile=True THEN Contamn_Blockage= Background absorption is high-probably due to high salt content.

can be combined and expressed as a single rule as follows:

Rule: If the abs values are 50-80% of the expected values =True OR the abs values are 20-40% of the expected values =True AND fluctuations are observed in the noisy absorption profile=True THEN Contamn_Blockage= Background absorbance is high-probably due to high salt content.

The Table_Generator has no provisions for writing a rule which has OR as a connector in the premise of the rule. The Rule_Editor was used to encode the rules containing the OR connector.

It was found that in some cases more than one cause gives rise to the same combination of symptoms. In such instances, the rule requires the use of AND as a logical connector in the conclusion of the rule. Such rules were coded using the Rule_Editor.

3.11 Acquisition of Knowledge using Rule_Editor of EAshell

The Rule_Editor has more features than the Table_Generator. The main features of the Rule_Editor are listed below:

a. Logical connectors AND, OR can be used in the premise of the rule;
b. AND can be used in the conclusion of the rule;
c. Relational operators (=, <>, >, <, >=, <=) can be used in the premise of the rule;

d. The ASK statement presents the symptom to the user;

e. The OPTION statement is a prompt for the ASK statement;

f. The FIND statement lists variables which have been set as "goal" by the user. Thus, if the developer wishes to incorporate goal-driven chaining (backward chaining and mixed chaining), use of the FIND statement is mandatory;

g. Lists words that are reserved for use in EAshell and hence cannot be used by the developer to write rules;

h. The keywords IF, THEN, ASK, OPTION are made available to the user, so that the user is only required to fill in the information, and

i. Requests the developer to set "goal/s" that are necessary in backward and mixed chaining strategies.

Example:

Chemical Knowledge: Liquid build up in the spray chamber gives rise to a pulsating signal and a ragged flame.

FIND

Contamn_Blockage,

Rule: **IF** Variable7=True AND Variable11=True **THEN**

Contamn_Blockage=Liquid build up in the spray chamber.

ASK Variable 7

"Pulsating signal observed in the absorption profile?"

OPTION

True, False;

ASK Variable11

"Ragged Flame ?"

OPTION

True, False;

Rules generated using the Rule_Editor are loaded as a rule base in the form of a EAshell's *dll* function. Table 3.4 represents the knowledge matrix that was used to make the rule base in AAdiagnosis.

3.12 Encoding Chemical Knowledge using EAshell

Figure 3.4 shows the steps involved in encoding chemical knowledge using EAshell.

Example:

Aspiration of a viscous solution is associated with the following symptoms:

a. An uptake rate of less than 5 mL/ min (this symptom is related to the set-up conditions);
b. An absorption profile that has a long rise time (this symptom is related to the real-time detector response), and
c. Absorbance values that are 20-40% of the expected values from the calibration graph (this symptom is related to the quality of the data).

The symptoms and the associated cause ***"check viscosity of the solution"*** were assembled in the Table_Generator.

It was found that the cause ***"viscosity of the solution is high"*** gives rise to the symptoms listed above and the knowledge can be expressed in the following rule.

Rule: If uptake rate is less than 5 mL/min=True AND Absorption profile has a long rise time=True AND Absorbance values are 20-40% of the expected values=True THEN sample_soln_prob=Check viscosity of the solution.

The cause ***"viscosity of the solution is high"*** was always associated with ***"uptake rate of less than 5 mL/ min"*** and may also have been related with ***"absorption profile has a long rise time"*** or ***"absorbance values are 20-40% of expected values"***. A rule of the following type is used to express the knowledge. The Rule_Editor is used to write the rule containing OR as the

Table 3.4 The knowledge matrix in AAdiagnosis using EAshell. * indicates the use of OR as a connector in the premise of a rule (contd. on pages 114-116).

	1 Contamination from samples	2 Contamination from samples	3 Contamination from samples	4 Contamination from samples	5 Contamination from samples	6 Contaminated spray chamber	7 Contam. spray cham./ ppte nebulizer	8 Contam. spray cham./ ppte nebulizer	9 Contam. spray cham./ ppte nebulizer	10 Ppte in nebulizer	11 Background abs high - salt conc	12 Background abs high - salt conc	13 Background abs high - salt conc	14 Background abs high - salt conc	15 Liquid build up in spray chamber	16 Liquid build up in spray chamber
High RSD	T	F	F	T			T	T	T							
Abs value 50-80%					F	T	T				T	T*	T*	T		
Abs value 20-40%												T*	T*			
Noisy absorption signal											T	T	T	T	T	
Spike at begin_of_ peak	T	T	T	T												
Occasional pulse					T	T										
Decaying-abs. profile										T						
Pulsations-abs. proflie															T	T
Calbn graph levels off								T					T	F		
Change in abs sensitivity								T	T		T			T		
Drift in baseline																
Ragged Flame	F	F	T	T	T		T		T						T	T
Refractory element																
Long rise time-abs.profile																
Uptake rate < 5 mL/min																
Calbn graph sigmoidal																
Fluctuating display																
No display -HCL on																
Integration time too small																
Pressure < 70 psi																
Burner > 8 on vert. scale																
Oxidizing Flame																
N2O-acetylene flame																
Bubbles form in cap tube																
More information																

	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	Burner slot is blocked	Element affects absorbance	Element affects absorbance	Element affects absorbance	Blocked capillary tube	Blocked capillary tube	Blocked capillary tube	Check bubbles in capillary tube	Check bubbles in capillary tube	Check bubbles in capillary tube	Check bubbles in capillary tube	Check bubbles in capillary tube	Cannot detect any new problems	Try again-verify symptoms
High RSD				F		T			T					
Abs value 50-80%	T		T											
Abs value 20-40%														
Noisy absorption signal					T	T	T	T	T					
Spike at begin_of_ peak														
Occasional pulse														
Decaying-abs. profile														
Pulsations-abs. proflie														
Calbn graph levels off			T	T			T					T		
Change in abs sensitivity			T							F	T			
Drift in baseline														
Ragged Flame	T													
Refractory element		T	T	T										
Long rise time-abs.profile														
Uptake rate < 5 mL/min					T	T								
Calbn graph sigmoidal														
Fluctuating display														
No display -HCL on														
Integration time too small														
Pressure < 70 psi														
Burner > 8 on vert. scale														
Oxidizing Flame														
N2O-acetylene flame														
Bubbles form in cap tube								T	T	T	T	T		
More information													F	T

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
	Incomplete mixing of sample and fuel	Incomplete mixing of sample and fuel	Incomplete mixing of sample and fuel	Few atoms due to formn of ppte	Few atoms due to formn of ppte	Few atoms due to formn of ppte	Few atoms due to formn of ppte	Check viscosity of sample	Check viscosity of sample	Ionization in solution	Cannot detect any new problems	Try again-verify symptoms	Improperly positioned glass bead	Improperly positioned glass bead	Improperly positioned glass bead	Burner rotated by 90 degrees
High RSD		T	T										T	T	T	T
Abs value 50-80%	T	T	T										T	T	T	
Abs value 20-40%				T	T	T	T									T
Noisy absorption signal				T					T						T	
Spike at begin_of_ peak																
Occasional pulse																
Decaying signal																
Pulsating signal	T	T	T													
Calbn graph levels off					T	T	T		T							
Change in abs sensitivity													F	T		
Drift in baseline																
Ragged Flame		F	T			F	T									
Refractory element																
Slow rising signal								T*	T							
Uptake rate < 5 mL/min								T*								
Calbn graph sigmoidal										T						
Fluctuating display																
No display -HCL on																
Integration time too small																
Pressure < 70 psi																
Burner > 8 on vert. scale																
Oxidizing Flame																
N2O-acetylene flame																
Bubbles form in cap tube																
More information											F	T				

	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
	Scale expansion set high	Scale expansion set high	Check lamp intensity	Check the wavelength	Check integration time	Check integration time	Burner set high	Cannot detect any new problems	Try again-verify symptoms	Check fuel cylinder pressure	Check fuel cylinder pressure	Improper gas mixture	N2O valve icing up	Cannot detect any new problems	Try again-verify symptoms
High RSD			T		T										
Abs value 50-80%	T	F													
Abs value 20-40%															
Noisy absorption signal	T	T				T				T					
Spike at begin_of_ peak															
Occasional pulse															
Decaying signal															
Pulsating signal															
Calbn graph levels off															
Change in abs sensitivity															
Drift in baseline										T			T		
Ragged Flame															
Refractory element															
Slow rising signal															
Uptake rate < 5 mL/min															
Calbn graph sigmoidal															
Fluctuating display			T												
No display -HCL on				T											
Integration time too small					T	T									
Pressure < 70 psi											T				
Burner > 8 on vert. scale							T								
Oxidizing Flame												T			
N2O-acetylene flame													T		
Bubbles form in cap tube															
More information								F	T					F	T

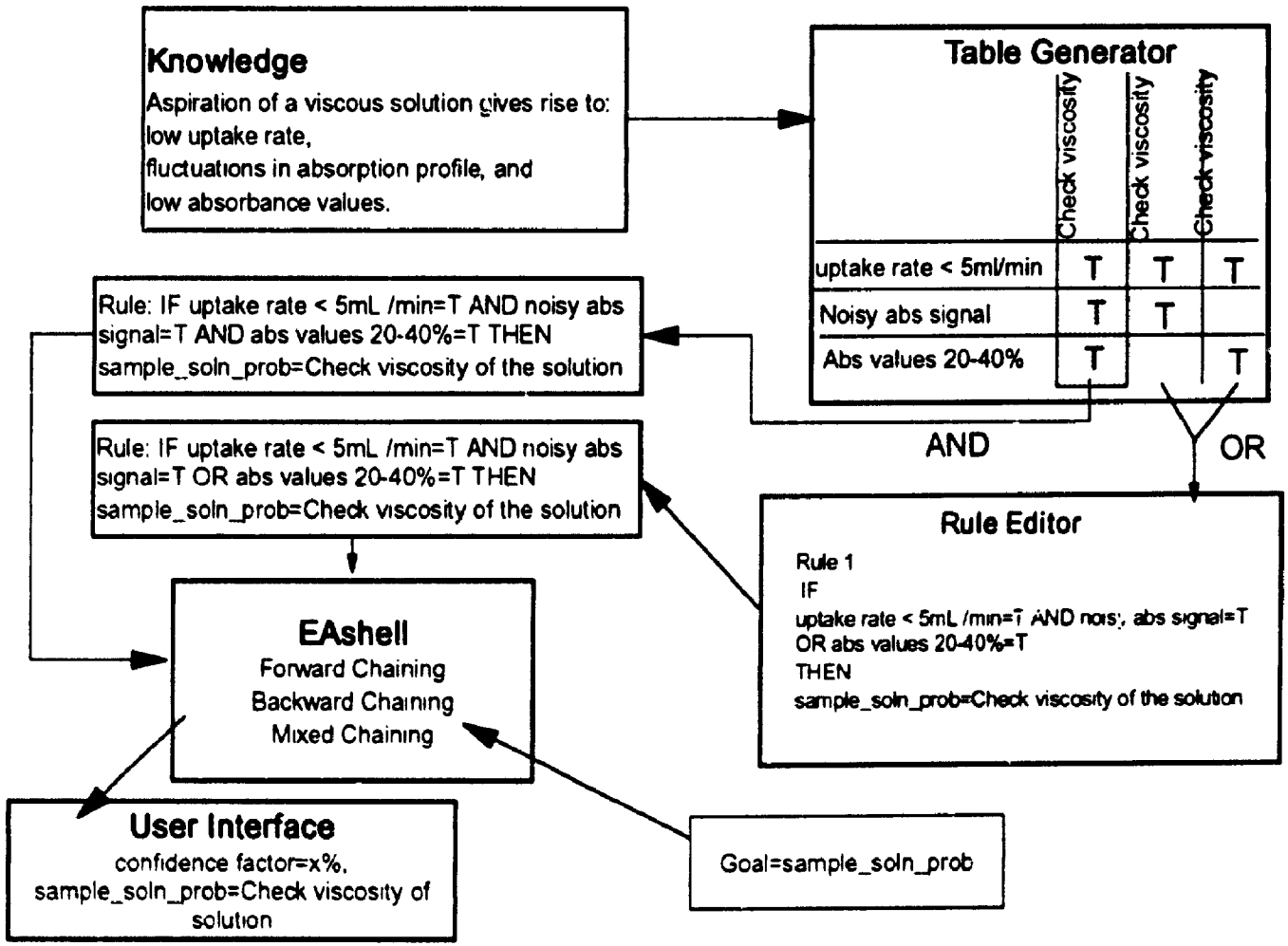


Figure 3.4 Encoding knowledge in AAdiagnosis using EAshell. Use of Table_Generator and Rule_Editor to acquire the chemical knowledge.

logical connector in the premise of the rule.

Rule: If Uptake rate is less than 5 mL/min=True AND Absorption profile has a long rise time=True OR Absorbance values are 20-40% of the expected values=True THEN sample_soln_prob=Check viscosity of the solution.

At run-time, the rules are loaded in EAshell and the inference is completed based on the available inference strategies. The final result is offered to the user together with the confidence factor.

3.13 Discussion

Expert system applications in the field of analytical chemistry fall into two broad categories (19): (i) systems that are developed to interpret experimental results, and (ii) systems that provide guidance as to the course of action the analyst should take. We believe that at the present time, there are highly productive roles for the use of expert systems in the analytical laboratory as advisers on method selection, instrument control, and diagnosing problems encountered during the analysis.

In this chapter, a method has been described for encoding chemical information dealing with problems encountered during metal analysis by flame atomic absorption spectrometry. The constraints involved in adding new knowledge to an expert system based on conventional languages led us to the use of an expert system shell for the development of AAdiagnosis. A rule based expert system shell was chosen that requires the chemical knowledge to be encoded in the form of case histories. The poor user interface in the KDS based expert system led to development of a Windows (version 3.1) based expert system shell in our laboratory.

The slowest step in the development of an expert system is the encoding of expert knowledge. This problem has been tackled and a novel method of encoding the necessary chemical information has been described. This is a general method for chemical information that is coded for advisory expert

systems. AAdiagnosis treats the symptoms and the related causes as data The symptoms are universal and together with the causes represent our experimental results. We have identified three levels that are necessary in encoding chemical knowledge in AAdiagnosis (using the KDS expert system shell), namely (a) the primary level, (b) the secondary level, and (c) the tertiary level. The primary level of knowledge relates symptoms directly with causes. The secondary level of knowledge relates symptoms with symptoms The rules in AAdiagnosis are generated after compiling the secondary level of knowledge and therefore relate symptoms. The tertiary level deals with those situations where one cause can give rise to more than one combination of symptoms In our method of encoding chemical knowledge, causes give rise to symptoms. Therefore, if a set of causes are the same for two symptoms, then the symptoms are related, and a rule of the type "if the (subject) is true then the (predicate) is true" is generated.

AAdiagnosis, using EAshell, was coded in the form of a matrix in which the developer assembles knowledge as causes that give rise to combinations of symptoms. There was no formal distinction between levels of knowledge. Logical connectors (AND, OR) and relational connectors were employed in the generation of rules. In a rule containing both the logical connectors in the *if* part of the rule, OR takes precedence over AND

3.14 References

1 Huixiao, H.; Xinquan, X. *J Chem. Inf. Comput. Sci*, 1990, **30**, 203.

2. Takeuchi, K.; Kuroda, C.; Ishida, M. *J. Chem. Inf. Comput. Sci*, 1990, **30**, 22.

3. Luinge, H-J., van der Mass, J. H. *Anal Chim Acta* 1989, **223**, 135.

4. Woodruff, H. B., Munk, M. E. *Anal. Chim. Acta*, 1977, **95**, 13.

5 Buydens, L.; Schoenmakers, P.; Maris, F.; Hendriks, H. *Anal. Chim. Acta*, 1993, **272**, 41 and references therein.

6 DeSmet, M.; Musch, G.; Peeters, A.; Buydens, L.; Massart D. L. *J. Chrom.* 1989, **485**, 237.

7. Mulholland, M.; Dunand, N.; Cleland, A.; Van Leeuwen, J. A.; Vandeginste B. G. M. *J. Chrom*, 1989, **485**, 283.

8 Esteban, M ; Ruissanchez, I.; Larrechi, M. S.; Rius, F. X. *Anal. Chim. Acta*, 1992, **268**, 95.

9. Birk, J. P in *Expert System Applications in Chemistry*, B. A. Hohne, T. H. Pierce, Eds., American Chemical Society, Washington, 1989, ch. 3, pp 20-33

10. Settle, F. A., Jr.; Pleva, M. A. *Chemom. and Intell. Lab. Sys.*, 1991, **11**, 13.

11. Browett, W. R.; Cox, T A.; Stillman, M. J. in *Expert System Applications in Chemistry*, B A Hohne, and T. A. Pierce, American Chemical Society, Washington, 1989, ch. 17, pp 210-235.

12. Lahiri, S., Zhu, Q.; Stillman, M. J. *Proceedings of the MOE Technology Transfer Conference*, 1990, 672.

13. Lahiri, S , Stillman, M. J. *Anal Chem* 1992, 64, 283A.

14 Leaman, C. M *AI Expert*, 1989, **4**, no. 5, 28.

15 Brown, D. E *AI Expert*, 1990, **5**, no 2, 32.

16 Furse, G in *Expert Systems Principles and Case Studies*, R. Forsyth Ed., Chapman and Hall Computing, London, 1989, ch. 10, pp 184-196

17. Bowerman, R. G ; Glover, D. E. in *Putting Expert Systems into Practice*, Van Nostrand Reinhold Co., New York, **1988**, ch. 4, pp 94-140.

18. Pederson, K. in *Practical Techniques for Rule-Based Systems*, John Wiley & Sons, Inc., New York, 1989, ch. 5, pp 70-88.

19. Bridge, T. P.; Williams, M. H.; Fell, A. F. *Chem. Brit.*, 1987, **23**, 1085.

CHAPTER 4
METHOD SELECTION

4.0 Introduction

Automation in the analytical laboratory has focused on sample preparation, solution handling, diagnostic software to aid in quality control, and selection of a method of analysis. Method selection is based on knowledge about the sample, the matrix and possible interferences. The knowledge used can be represented in a database, a rulebase, or in a combination of rulebase and database.

The concept of "intelligent databases" has been discussed by Schur (1). Schur has reported that the combination of database and expert system requires structural compatibility, since the stress in databases is on syntax over semantics and the reverse is true for expert systems. The relational model of data is the commonly used technique in database systems. In a relational database, data is accessed using a Structured Query Language (SQL). The use of expert systems in association with database management system (DBMS) can provide:

a. Training to novice users on how to interact with the DBMS;
b. An intelligent user interface, and
c. Help to users in generating run-time queries.

Several authors have reported method development expert systems in the area of liquid chromatography; a research area that is a part of the ESCA project (2). The authors described the steps involved in selection of a method of analysis which includes retention optimisation, system optimisation, and method validation. The route to method selection in the ESCA project is based solely on a search of the appropriate database, e.g. a database for the selection of the detector and another database for the selection of a column.

Settle and co-workers have developed an expert system for method selection for sample preparation by microwave dissolution (3,4). The authors have delineated the knowledge necessary in a method selection expert system

in terms of analytical descriptors. In their system, the expert system aids the user in selecting the matrix in which the analysis is required. On completion of the inference process, the expert system consults the database and reports the detailed analysis information to the user. PCplus, an expert system shell, was used to encode the knowledge and dBASE III was used as the database program.

4.1 AAmethods: An Expert Database System for Method Selection

AAmethods provides the sample preparation methods and the appropriate method of atomic absorption spectrometric analysis. The knowledge is contained in a methods selection database and a rulebase. The database contains methods of analysis based on metals in specific matrices. On the other hand, rulebase knowledge is represented in the form of the composition of the matrix and the chemical nature of the metal. The user first interacts with the database through the user interface. If sufficient facts are present, the database search yields the method of analysis. If the database does not contain the method of analysis, the rulebase is consulted by the user.

4.2 The Knowledge Domain in AAmethods

Ten metals and two groups of matrices were chosen as the knowledge domain in AAmethods. The two general matrices namely, water and biological, were further sub-divided into specialised groups. The knowledge encoded in the database is subdivided into the following categories:

a. Analyte of interest;
b. Matrix which contains the analyte;
c. Possible interferences present in the sample, and
d. Method of analysis.

The knowledge encoded in the rulebase was based on specific characteristics of the matrix and metal. For example, in the analysis of metals in drinking water, the matrix is described in terms of the ions present in the water. In terms of the metal, it is important to know the method to be used for matrix

separation and/or the preconcentration steps in the pretreatment process. The methods employed for pretreatment include chelation, followed by a liquid-liquid extraction, and ion exchange chromatography. Methods of analysis were subdivided into major areas that served as goals in the inference process. The goals encoded in AAmethods are:

a. Analysis for metals in drinking water (DWmatrix);
b. Analysis for metals in natural waters (Wmatrix), and
c. Analysis for metals in a biological matrix (Bmatrix).

Contrary to the approach adopted in the literature (3,4) the approach taken in the design of AAmethods involves searching a database followed by consultation of the rulebase, if necessary. Figure 4.1 shows the representation of knowledge in AAmethods.

4.3 The User Interface in AAmethods

Visual Basic (version 3.0) was used as the programming language. Microsoft Access (version 1.0) was used as the database program. EAengine provided the inference processes for the rulebase written with the help of the Table_Generator and the Rule_Editor. Figure 4.2 shows the user interface in the AAmethods module. On selecting "Metal" from the main window ("AAmethods-A method selection expert system"), the periodic table of elements is presented to the user. The user selects the metal for which the analysis is to be carried out. The message box in this window, "Selection of a metal", indicates the name of the metal selected by the user. When the user selects "OK" the control of the program is transferred to the main window. At this stage, the user selects "Matrix" and also the desired matrix from "Selection of a matrix". Since analysis is matrix-specific, it is essential that the user can identify the matrix containing the analyte of interest. When the user identifies the matrix, the message box indicates both the metal and the matrix in which it is contained. On selecting "OK", the AAmethods database is opened. If there is a specific method of analysis for the particular combination chosen by the user, the main window reports the analysis number. The menu "Analysis Info" contains detailed

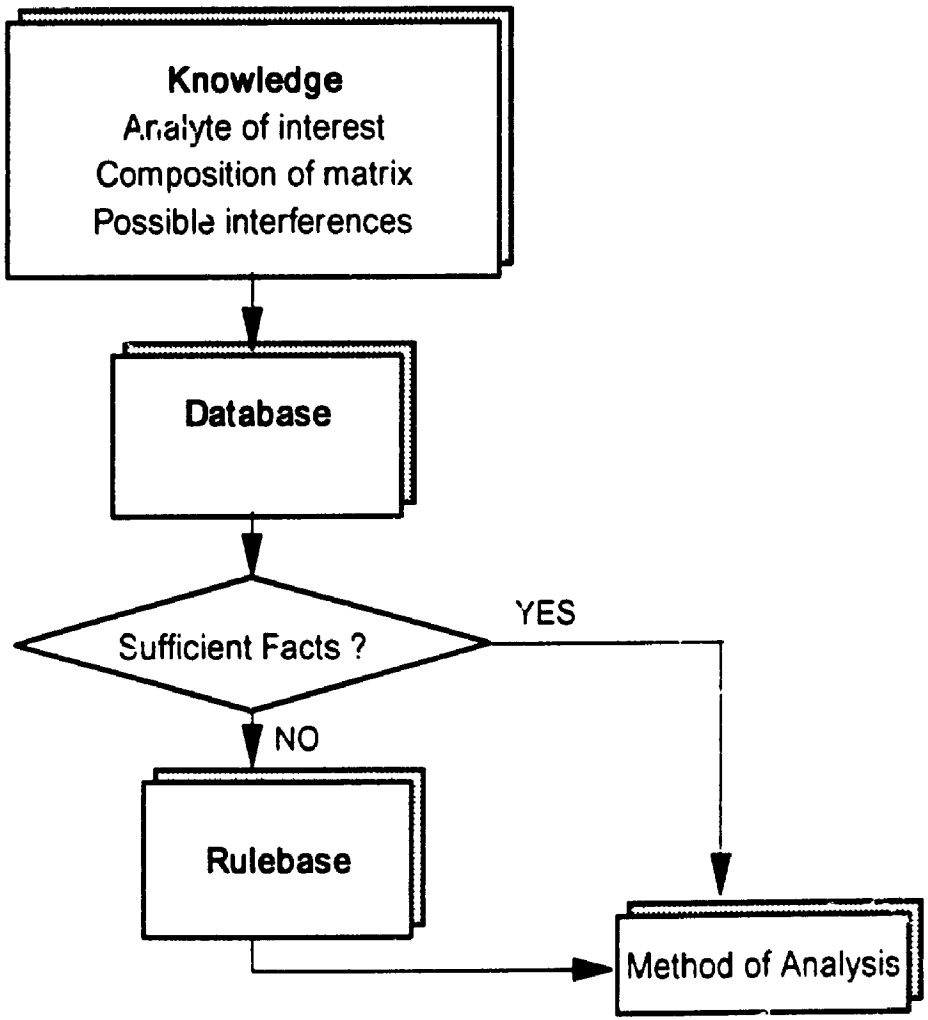


Figure 4.1 Representation of knowledge in AAmethods.

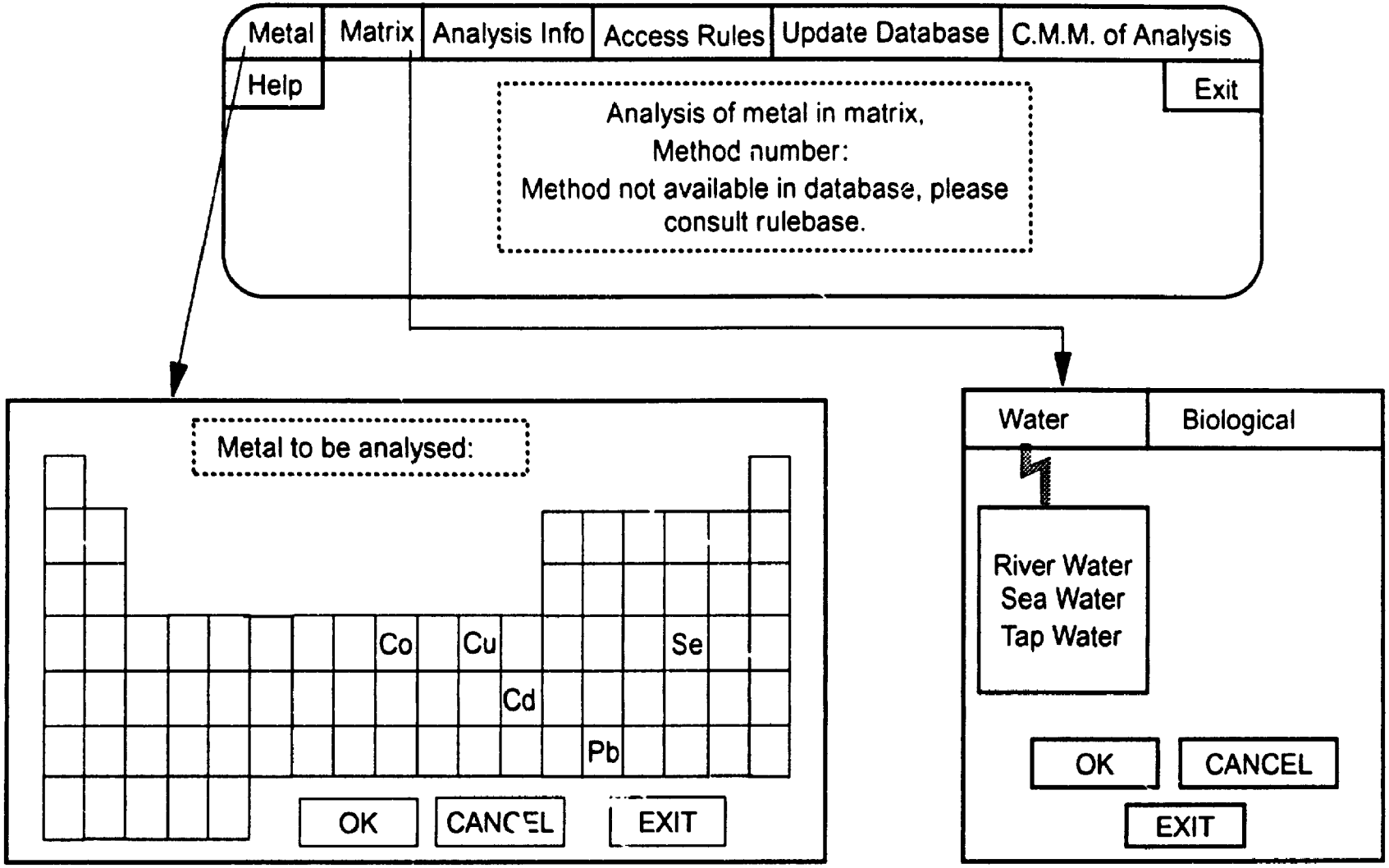


Figure 4.2 Structure of the user interface in AAmethods. C.M.M.= closest matched method.

information about the selected method of analysis. If a method of analysis cannot be found in the database, the user is requested to consult the rulebase by selecting "Access Rules" which provides the closest match to a method contained in the database.

4.4 Use of the Database in AAmethods

Microsoft Access (version 1.0) was used to encode the database for AAmethods. Table 4.1 shows the database constructed for AAmethods using the Microsoft Access environment. The knowledge encoded in the database is acquired from published methods of analysis for metals in the specified matrices. Only one method of analysis has been encoded for a particular combination of metal and matrix. The entries were extracted from the available literature, which explains the absence of methods of analysis for each metal in every matrix.

AAmethods contains two copies of the methods selection database. **"Expert Database"** searches for a method of analysis and cannot be modified by the user. When the **"Expert Database"** does not contain the method of analysis for the metal in the matrix requested by the user, it is possible that the user will find the **"closest match"** method of analysis obtained from the rulebase acceptable. In such a situation, the user is required to validate the method by experimentation and verification using literature. However, if the user wishes to update a validated method, the user version of the database, **"User Database"**, can be updated. The developer can update the expert version of the database at a later date.

In addition to the major categories in which the knowledge was delineated, four additional categories were found to enhance the intelligence of the AAmethods module. These new categories are as follows:

a. Analysis Number: Methods of analysis belonging to a similar matrix and having similar sample preparation procedure are allocated numbers in the same range. For example, in the analysis of chromium in drinking water (analysis number 1900), lake water (analysis number 1910), and sea water

Metal ID	Metal	Matrix	Matrix_class	Interference	Anal	Comments	Anal_Num	ANALID
ARSENIC_RIVER_WATER	Arsenic	River_Water	Wmatrix		HG-AAS		1000	A
BARIUM_SEA	Barium	Sea_Water	Wmatrix	High Salt Content	GF-AAS		1200	B
CADMIUM_BLOOD	Cadmium	Blood	Bmatrix		GF-AAS	Pd, NH4NO3 modifier	5000	C
CADMIUM_SEA	Cadmium	Sea_Water	Wmatrix	High Salt Content	GF-AAS	APDC, pH > =4	1500	D
CADMIUM_URINE	Cadmium	Urine	Bmatrix	PO4(3-), SO4(2-)	GF-AAS	Pd, NH4NO3 modifier	3100	E
CADMIUM_WATER_COMPLEX_HUMIC	Cadmium	Humic_Natural_Waters	Wmatrix	Fe(3+) absorptio	GF-AAS		1600	F
CHROMIUM_DRINKING_WATER	Chromium	Drinking_Water	Wmatrix		GF-AAS	NaDDTC, ethanol	1900	J
CHROMIUM_LAKE_WATER	Chromium	Lake_Water	Wmatrix		GF-AAS	NaDDTC, ethanol	1910	J
CHROMIUM_SEA	Chromium	Sea_Water	Wmatrix	High Salt Content	GF-AAS	NaDDTC, ethanol	1920	J
CHROMIUM_WATER_SEDIMENT	Chromium	Sediment_Natural_Waters	Wmatrix		GF-AAS	Pd, HNO3 modifier	1810	I
COBALT_SEA	Cobalt	Sea_Water	Wmatrix	High Salt Content	GF-AAS	NaDDTC, CCl4, pH=4.5	1700	G
COBALT_URINE	Cobalt	Urine	Bmatrix	PO4(3-),SO4(2-)	GF-AAS	Pd modifier	3200	E
COBALT_WATER_SEDIMENT	Cobalt	Sediment_Natural_Waters	Wmatrix		GF-AAS	Pd, HNO3 modifier	1800	I
COPPER_BOTANICAL	Copper	Botanical	Bmatrix		GF-AAS	solid sampling ?	5100	K
COPPER_SEA	Copper	Sea_Water	Wmatrix	High Salt Content	GF-AAS	APDC complex, pH 2-8	1510	D
COPPER_WATER_COMPLEX_HUMIC	Copper	Humic_Natural_Waters	Wmatrix	Fe(3+) absorptio	GF-AAS		1610	F
COPPER_WATER_SEDIMENT	Copper	Sediment_Natural_Waters	Wmatrix		GF-AAS	Pd, HNO3 modifier	1820	I
IRON_SEA	Iron	Sea_Water	Wmatrix	High Salt Content	GF-AAS	NaDDTC, CCl4, pH=4.5	1710	G
LEAD_BLOOD	Lead	Blood	Bmatrix		GF-AAS	NH4H2PO4 + (MgNO3)2	6100	P
LEAD_DRINKING_WATER	Lead	Drinking_Water	Wmatrix	Sulfate, Chloride	GF-AAS	La, HNO3 modifiers	2000	N
LEAD_SEA	Lead	Sea_Water	Wmatrix	High Salt Content	GF-AAS	NaDDTC, CCl4, pH=4.5	1730	G
LEAD_WATER_COMPLEX_HUMIC	Lead	Humic_Natural_Waters	Wmatrix	Fe(3+) absorptio	GF-AAS		1620	F
LEAD_WATER_SEDIMENT	Lead	Sediment_Natural_Waters	Wmatrix		GF-AAS	Pd, HNO3 modifier	1840	I
NICKEL_RAIN_WATER	Nickel	Rain_Water	Wmatrix	Ca(2+), Fe, Na(+	GF-AAS	Complex with DDTC, pH	1400	L
NICKEL_SEA	Nickel	Sea_Water	Wmatrix	High Salt Content	GF-AAS	NaDDTC, CCl4, pH=4.5	1720	G
NICKEL_WATER_SEDIMENT	Nickel	Sediment_Natural_Waters	Wmatrix		GF-AAS	Pd, HNO3 modifier	1830	I
SELENIUM_RIVER_WATER	Selenium	River_Water	Wmatrix	Cu(2+), Ni(2+)	HG-AAS		1010	A
SELENIUM_SERUM	Selenium	Serum	Bmatrix		GF-AAS	Cu(NO3)2 + Mg(NO3)2	7710	M
SELENIUM_TISSUE	Selenium	Tissue	Bmatrix		GF-AAS	Pd: thermal stabilization	8000	O

Table 4.1 Knowledge representation in AAmethods (Microsoft Access environment).

(analysis number 1920), chromium is chelated using NaDDTC and extracted with ethanol. The method of analysis is by graphite furnace atomic absorption spectrometry in all the three matrices, and these analyses are therefore given analysis numbers in the same range;

b. Analysis Identifier: Every method of analysis is associated with an analysis identifier. This feature is used to find the closest match in the database. If a method of analysis is not present in the database, the user consults the rulebase. The method of analysis offered by the rulebase includes a method number and an method identifier. This identifier provided by the rulebase is matched with the analysis identifier and the user is provided with the **"closest match"** present in the database, and

c. Matrix_class: Matrices in the database are subdivided into major areas. If the database does not contain a method of analysis for the metal in the matrix chosen by the user, a search in terms of matrix_class arrives at methods of analysis for the particular metal in other matrices belonging to the same area.

Example: The user wishes to analyse for cadmium in lake water.

The metal (Cd) and the matrix (lake water) are selected by the user. The **"Expert Database"** is opened, the search does not produce the required method of analysis (see Figures 4.3-4.5). However, this database lists the methods of analysis for cadmium in water related matrices. Therefore, the following closest matched methods of analysis are offered by the **"Expert Database"**. These are:

a. Analysis for cadmium in sea water, method D1500, and

b. Analysis for cadmium in natural waters (humic complexed), method F1600.

4.5 Use of Rulebase in the AAmethods

Production rules related to assigning a method of analysis based on the presence of interferences was encoded in the rulebase. EAshell (version 1.0) was used to encode the chemical knowledge. The identifier associated with every method of analysis offered as a solution is same as the analysis identifier

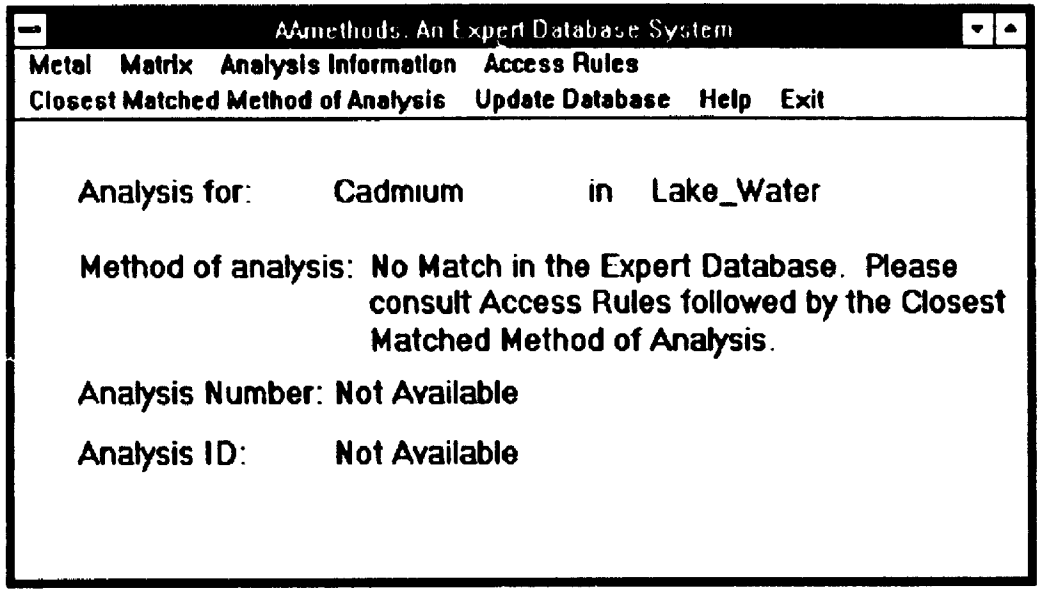


Figure 4.3 The user interface in AAmethods: analysis for cadmium in lake water. The "Expert Database" does not contain a method of analysis for cadmium in lake water.

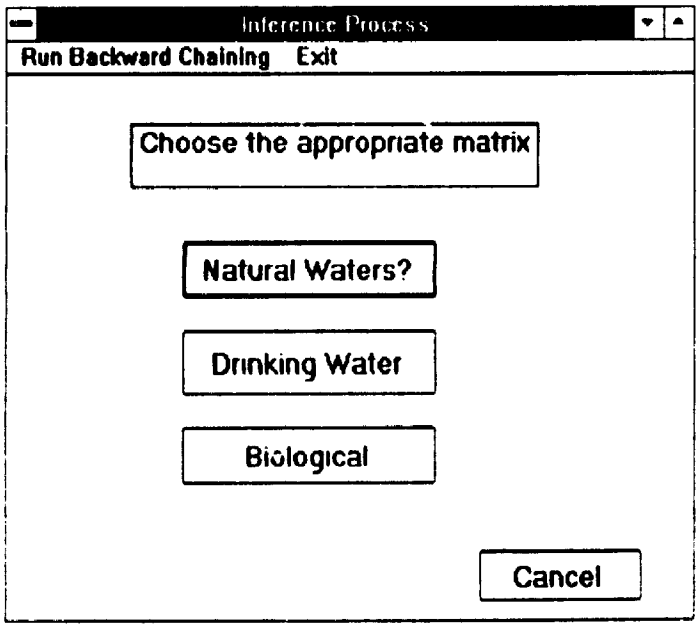


Figure 4.4 The user interface in AAmethods: use of the rulebase (cadmium in lake water). The rulebase helps the user find a "closest match" method of analysis for cadmium in lake water.

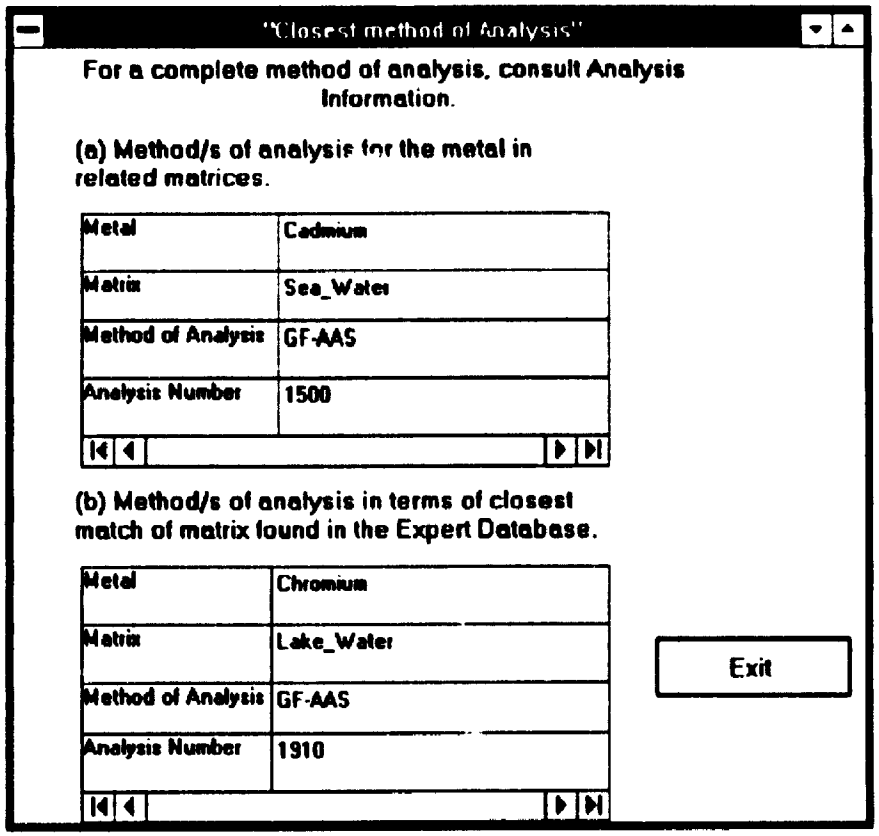


Figure 4.5 The user interface in AAmethods: closest match method of analysis (cadmium in lake water). The available methods of analysis are (a) cadmium in water related matrices, and (b) chromium in lake water, this search is obtained by using the rulebase and the "Expert Database".

connected to the same matrix (in the database).

Example:

Rule: IF variable2=Lake_Water AND Metal can complex with dithiocarbamates=True AND metal_extractable at pH < 4.0=False THEN Wmatrix=GFAAS method J1910*

The method of analysis in the same matrix, lake water, available in the **"Expert Database"** belongs to analysis of chromium. The analysis number is 1910 and J is the analysis identifier. Therefore, for metals such as cadmium, cobalt, chromium, iron and lead the above rule leads to method J1910 Alternatively, the above rule can be subdivided into a number of rules, each representing a metal for which the method of analysis has been requested by the user. For example the following rule can be used to find the method of analysis for cadmium in lake water.

Rule: IF variable2=Lake_Water AND variable1=Cadmium AND metal_extractable at pH < 4.0=False THEN Wmatrix=GFAAS method J1910*

The user can use method J1910 to learn the procedure of analysis (including extraction and chelation methods) for determination of cadmium in lake water. In addition, the user consults methods D1500 and F1600 to obtain the analysis parameters for the determination of cadmium by graphite furnace atomic absorption spectrometry.

For metals, such as barium, that cannot be analysed by chelation/extraction the following rule can lead to the closest match present in the database.

Rule: If variable1=Barium or analysing for alkaline earth metals=True Then Wmatrix= GFAAS method B1200*.

The method B1200 is connected with analysis of barium in sea water in the AAmethods database.

4.5.1 Inference Strategy in AAmethods

Backward chaining with depth-first search was used as the inference strategy in the AAmethods rulebase. Rules were classified in terms of goals related to the matrices. In order to avoid the inference engine enquiring about the metal and the matrix for the method of analysis requested by the user, a factbase containing this information can be loaded prior to the search of the rulebase.

4.6 Implementation of AAmethods

The success of a methods selection expert system depends on the program being able to understand analyses in terms of matrices. The database provides the quickest route to obtaining a method of analysis based on available literature. The rulebase is based on heuristic knowledge, for example metals capable of forming complex with dithiocarbamates can be analysed by chelation with NaDDTC, extraction with ethanol, followed by transferring back into the aqueous phase by shaking with HNO_3.

Example: The user wishes to analyse Ni in drinking water.

The user does not find the method of analysis in the database (Figure 4.6). However, the database search yields the following methods of analysis available for nickel in water related matrices. The available methods are:

a. Analysis of nickel in rain water, method L1400;
b. Analysis of nickel in sea water, method G1720, and
c Analysis of nickel in sediments and suspended matter of natural waters, method I1830.

In order to know how to remove the inferences present in drinking water, the user consults the rulebase (Figure 4.7). If either chloride or sulphate are present in the matrix, the following rule suggests that the closest method of analysis, in terms of interferences, present in the database is analysis of lead in drinking water, method N2000 (Figure 4.8). At this stage, the user can consult "Analysis Information" for the detailed procedure to remove interferences from drinking water. On consulting a method associated with analysis for nickel in

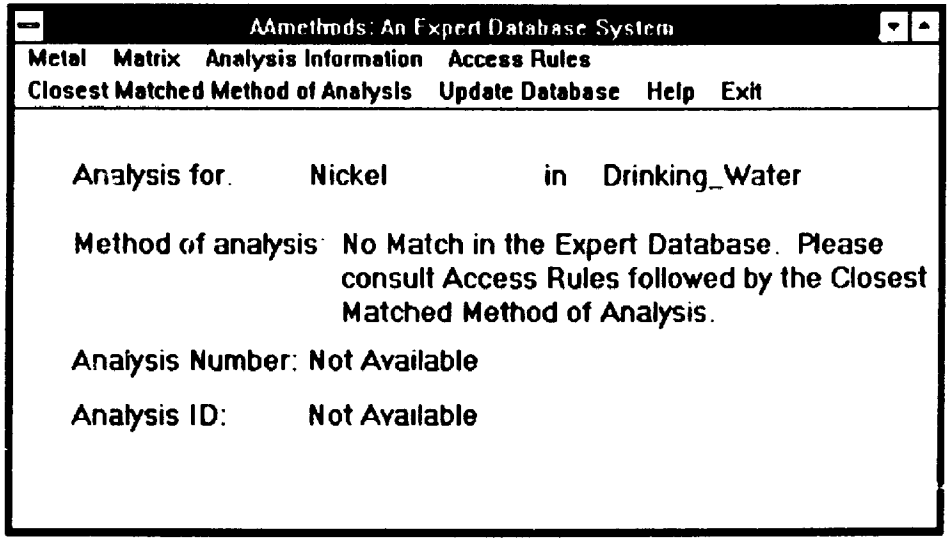


Figure 4.6 The user interface in AAmethods: analysis for nickel in drinking water. The "Expert Database" does not contain a method of analysis for nickel in drinking water.

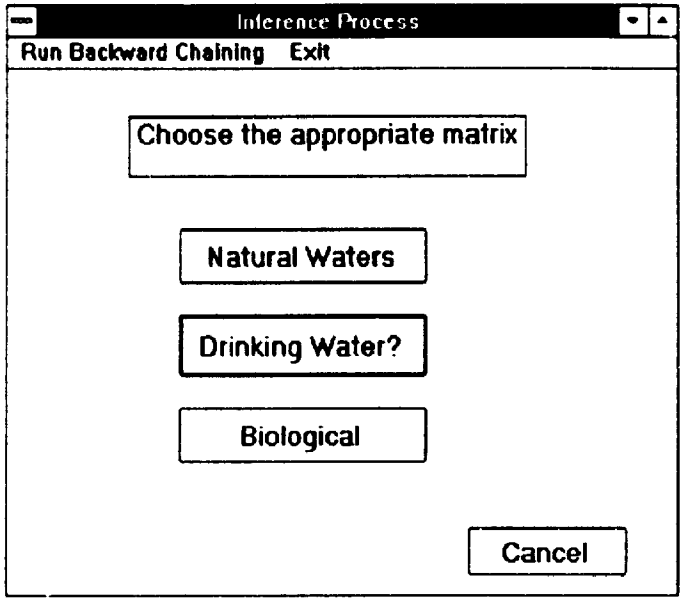


Figure 4.7 The user interface in AAmethods: use of the rulebase (nickel in drinking water). The rulebase helps the user find a "closest match" method of analysis for nickel in drinking water.

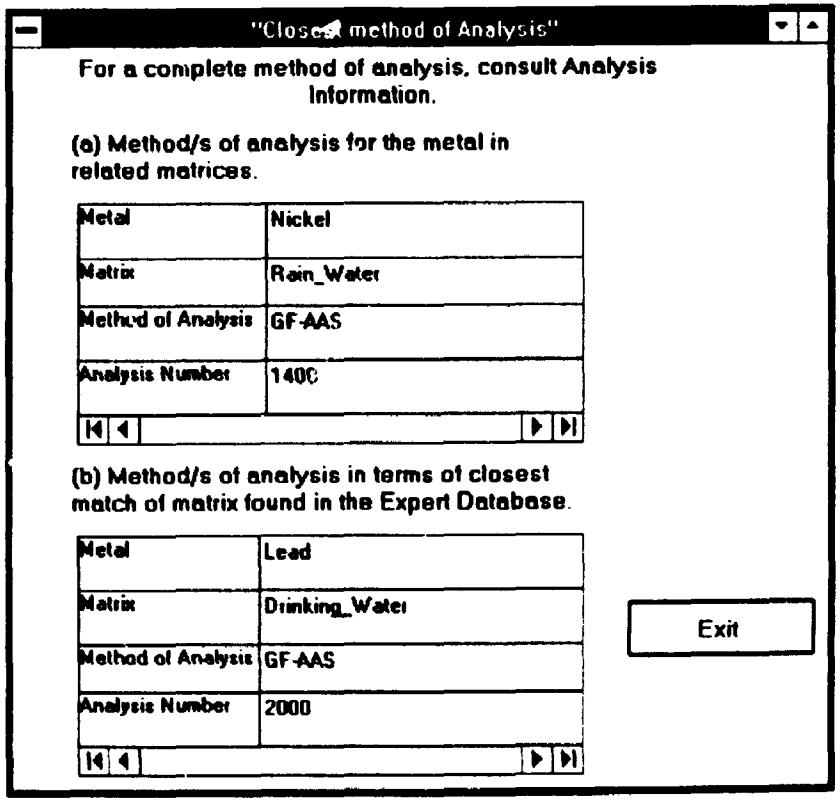


Figure 4.8 The user interface in AAmethods: closest match method of analysis (nickel in drinking water). The available methods of analysis are (a) nickel in water related matrices and (b) lead in drinking water, this search is obtained by using the rulebase and the "Expert Database".

water, the user can obtain information about analysis parameters for determination of nickel. The method of analysis including removal of interferences and subsequent analysis of nickel have to be validated by the user.

Figure 4.9 shows the structure of AAmethods database expert system. If the method of analysis is found in the **"Expert Database"** the consultation is complete, the user then selects the analysis parameters and performs the analysis. If the **"Expert Database"** does not contain a method of analysis, the rulebase is suggests a **"closest-match"** method of analysis. Following validation of the **"closest-match"**, the user updates the **"User Database"**.

4.7 Discussion

The development and implementation of a method selection expert system requires the use of a database and a rulebase. The search of the database prior to use of the rulebase allows the user to obtain a validated method of analysis. The rulebase advises the user the best possible method of analysis not contained in the **"Expert Database"**. The knowledge encoded in the rulebase pertains to the chemical properties of the metal and the presence of interferences. The most important role of the expert system in an expert database system is to be able to suggest a method of analysis. The suggested method of analysis must be validated by the user .

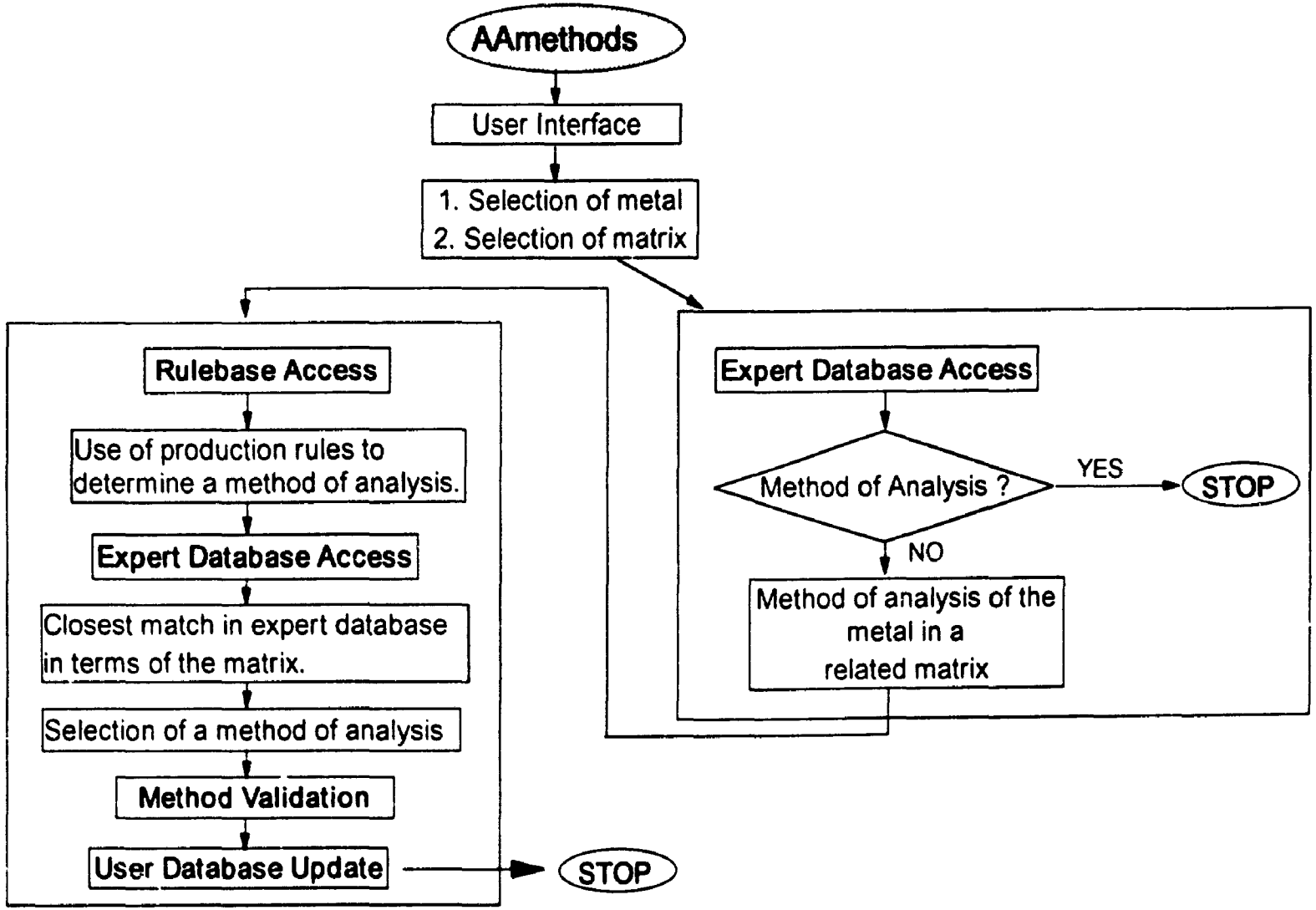


Figure 4.9 The structure of AAmethods.

4.8 References

1. Schur, S., *AI Expert*, January 1988, **3**, 26.
2. Buydens, L., Schoenmakers, P., Maris, F., Hindriks, H., *Anal. Chim. Acta*, 1993, **272**, 41 and references cited therein.
3. Settle, F. A. Jr., Diamondstone, B. I., Kingston, H. M., Pleva, M. A., *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 11.
4. Settle, F. A. Jr., Walter, P. J., Kingston, H. M., Pleva, M. A., Snider, T., Boute, W., *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 349.

CHAPTER 5

CONCLUSIONS

5.0 Encoding Chemical Knowledge

There are two approaches in the design of a rule-based representation of knowledge: (a) the system will require a direct input of knowledge in the form of rules, and (b) the system will generate rules from the available chemical knowledge. We have stressed during this thesis that the maintainability of the knowledge base is important and believe that adding knowledge to a system that requires input as rules is not an easy task. Therefore, in our work, we have opted for the second approach to build rule-based expert systems. As chemists, we treat the chemical knowledge as facts, and we are more able to modify facts rather than a rule-based structure of the same knowledge.

In our approach to encoding knowledge, we have designed a knowledge acquisition tool called the Table_Generator which builds tables of knowledge encoded directly by experts. The table format of knowledge acquisition greatly aids in the verification, portability, and expansion of the knowledge base. The requirements for encoding knowledge into the table format are that:

a. The observables should occupy the rows and the solutions or advice should occupy the columns of the matrix;
b. Knowledge belonging to the same sub-domain must be grouped together;
c. The expert should treat the knowledge as a case history and should not focus on the resultant rules, and
d. The values that can be used to connect observables and solutions are "true" (T), "false" (F), and "unknown" (?). The expert should attempt to avoid the use of the variable "unknown" in situations where it is possible to assign a definite value to a connection.

The following example shows how the Table_Generator can be used to give rise to rules. The available knowledge is shown in Table 5.1.

Table 5.1. The knowledge to be transformed into rules.

Knowledge Domain	Observable	Action/Advice/Solution
Contamination_Blockage	The flame has a ragged appearance.	The burner slot may be partially blocked.
Contamination_Blockage	Occasional pulse observed from the absorption profile.	Contaminated spray chamber.
Contamination_Blockage; Solution_Problem	Lower than expected absorbance values.	The burner slot may be partially blocked. There may be contamination in the spray chamber. The solution may be too viscous.
Solution_Problem	Long rise time observed from the absorption profile.	The solution is too viscous.

The chemical knowledge is assembled as case histories in the Table_Generator (Figure 5.1).

	Contamn_Blockage = Burner slot is blocked	Contamn_Blockage = Contaminated spray chamber	Soln_Problem = Check viscosity of solution
The flame has a ragged appearance	T		
Lower than expected absorbance values	T	T	T
Long rise time observed from the absorption profile			T
Occasional pulse observed from the absorption profile		T	

Figure 5.1 Acquisition of chemical knowledge using the Table_Generator

When the input of knowledge is completed, the Table_Generator creates the rules (Table 5.2). These rules are edited by the Rule_Editor which incorporates features that make the knowledge base usable by the inference engine.

Table 5.2. Rules generated from the chemical knowledge present in the knowledge table.

Rule 1. IF The flame has a ragged appearance=True AND Lower than expected absorbance values=True THEN Contamn_Blockage=Burner slot is blocked.
Rule 2. IF Lower than expected absorbance values=True AND Occasional pulse observed from the absorption profile=True THEN Contamn_Blockage=Contaminated spray chamber.
Rule 3. IF Lower than expected absorbance values=True AND Long rise time observed from the absorption profile=True THEN Soln_Problem=Check viscosity of the solution.

Although, it has been found that this method of rule generation is very efficient, there is room for further improvement in the operation of these software tools. The drawbacks of the present system are that:

a. The confidence factor (cf) associated with a rule assigns equal value to each symptom. A more realistic approach would involve assigning a weighted value for each symptom of a particular cause. For example, in the present case, in rule 1, for the cause "burner slot is blocked", if the symptom "lower than expected absorbance values" is observed a greater number of times than the symptom "the flame has a ragged appearance" then the former symptom deserves a higher weighting. However, this system of weighted symptoms requires the expert system to assign weight(s) after each consultation;
b. the Rule_Editor can check the syntax and format of rules to be used only by EAshell, the expert system shell developed in our laboratory. Further work is required to make the Rule_Editor universally capable of verifying formats for other rule-based expert systems.

5.1 Summary

Expert systems are knowledge based computer programs that attempt to apply the experience of an expert in a particular area of knowledge. In studying expert system applications in chemistry, the hardware and software components

serve only as tools to represent and apply the chemical knowledge. The role of expert systems is not to replace chemists but to aid them with advice. Researchers interested in expert system applications in analytical chemistry have mainly focused on a single area of expertise, refined the knowledge base, and written prototypes. Their objectives are to distribute these applications to industry.

An expert system application in the area of analytical chemistry should be able to:

a. Express human expertise as heuristic knowledge;
b. Store and retrieve information;
c. Control instrument(s), and
d. Provide a real-time assessment of the quality of analytical data.

There are two important stages involved in the transfer of human expertise into an expert system. These are: (i) encoding the chemical knowledge, and (ii) representing the chemical knowledge. There has been little work reported about how knowledge in general can be coded for use in an expert system. As computational tools becomes more available, the coding of chemical knowledge becomes more and more of a *bottleneck*. Once coded, the availability of different forms of knowledge representation introduces a *knowledge transportation bottleneck* because not only can rules not be readily transferred between systems, the knowledge encoded in a rule-based representation cannot be used in other representations, for example knowledge coded for a rule-based system cannot be transferred to a frame-based system.

A second use of expert system technology is in the field of total automation. The goal of total automation for chemical analysis has not been discussed in the literature previously. An understanding of the different aspects of automation will lead to appreciation of the complexities involved in total unattended automation of chemical analyses.

5.1.1 AA-Quality Control and AAcontrol

The modules AA-Quality Control and AAcontrol deal with the possibility of unattended automation. AAcontrol is responsible for automated solution handling, scheduling, instrument control, and data acquisition from the atomic absorption spectrometer. AA-QC provides an assessment on the quality of data transmitted by AAcontrol. The trace of absorbance versus time, the absorption profile, was found to be the best real-time indicator of the data and instrument quality of a single measurement by FAAS. AA-QC uses assessment criteria to determine if a measured set of data are acceptable. The assessment criteria calculated are the maximum positive slope of the absorption profile, the maximum negative slope of the absorption profile, and mean absorbance of the peak region, reflect the quality of the data and the status of the instrument. AA-QC uses production rules to identify four simple problems associated with analysis by FAAS. These are (a) a blocked burner due to excessive amounts of salt, (b) a viscous solution, (c) a blocked capillary tube, and (d) a long capillary tube. The contour of the derivative plot provides information regarding the likely problem associated with the analysis. Test samples activate the AA-QC routines. The module alerts the user through a combination of error messages indicating the cause of the problem. These derivative plots provide a unique interpretation for a single analytical measurement. Therefore, the detector response can be used to model the quality of analytical data and instrument performance. However, in order to model the detector response, it is important to understand the physical and chemical processes that give rise to the analytical signal. Automated correction is possible if AA-QC can transmit messages back to AAcontrol. The interaction between AAcontrol and AA-QC emphasises the need for a dual mode of operation for real-time quality control of analytical instruments. These are: (a) the autosampler mode of AAcontrol to carry out routine analysis, and (b) the robotic mode of AAcontrol to take necessary corrective action and adopt a new sequence of determination.

5.1.2 AAdiagnosis

KDS 3, a commercially available expert system shell was used to encode the chemical knowledge associated with problems encountered during analysis for metals by FAAS. The necessary knowledge was assembled as a matrix in which the symptoms or conditions represent the rows of the matrix and the causes or conclusions represent the columns of the matrix. It was found that the knowledge can be acquired in three stages: (a) the primary level, (b) the secondary level, and (c) the tertiary level. At the primary level, symptoms were related with underlying causes. The secondary level of knowledge associates symptoms with other symptoms. Rules are generated after compiling the secondary level of knowledge. The tertiary level deals with case histories where a cause can give rise to more than one combination of symptoms. The need for a better user interface led to the development of a Windows based expert system shell, EAshell, in our laboratory. In EAshell, knowledge is acquired as a matrix of facts. The knowledge is then represented as rules which relate symptoms with causes. Knowledge encoded in the form of a matrix improves the maintainability of an expert system.

5.1.3 AAmethods

Automated selection of a method of analysis has been studied in AAmethods. The concept of an expert database system has been developed to aid the user. The "expert database" contains validated methods of analysis for metals in different matrices. The user is required to have prior information regarding the matrix for which a method of analysis is selected. The user interacts with the "expert database" in order to obtain a method of analysis. In the event that a method is not present in the "expert database", the user consults the rule base. The rule base contains knowledge as characteristics of metals and matrices. This systems attempts to find a "closest match" of a method of analysis found in the "expert database". On finding a "closest match"

method of analysis, the user is required to validate the method and update the "user database".

5.2 Thoughts

It is hoped that the acquisition of knowledge in the form of a matrix will lead to a common method for encoding chemical knowledge for rule-based expert systems. This will allow such knowledge bases to be transported from one rule-based expert system tool to another. Also, the knowledge gained in the area of modelling the detector response to assess the quality of analytical data can be applied to other analytical instruments.

For economic and safety reasons, the goal of unattended automated analysis must be fully pursued. A route to this goal requires interfacing robotics with expert systems. Applications such as cleaning up nuclear wastes and analysis of hazardous compounds require this technology urgently.