

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:

The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

DEVELOPMENT OF AN EXPERT SYSTEM
FOR PHYSICAL PROPERTY PREDICTIONS

by

R. Banares-Alcantara \ A.W. Westerberg

December, 1983

DRC-06-«49-83

**DEVELOPMENT OF AN EXPERT SYSTEM
FOR PHYSICAL PROPERTY PREDICTIONS**

by

René Benares-Alcántara

Arthur W. Westerberg

Department of Chemical Engineering

Michael D. Rychener

*Design Research Center & Robotics Institute
Carnegie-Mellon University, Pittsburgh, PA 15213*

ABSTRACT

In this article we describe the steps followed to develop a prototype *expert system* called CONPHYDE (CONSultant for PHYsical property DEcisions). Using the framework of the existent expert system PROSPECTOR, CONPHYDE is designed to aid an engineer in the selection of an appropriate vapor-liquid equilibrium method when performing various process calculations. We also present the structure of the *inference networks* in CONPHYDE which capture the intent of the 37 heuristic rules that form the *knowledge base* on which the current version is based. An example session with CONPHYDE demonstrates its ability to ask questions in an "intelligent" sequence and provide "expert" advice based on them. The session also illustrates CONPHYDE's (i.e. PROSPECTOR's) ability to explain on request why it is asking certain questions and why it is offering the advice it is.

SCOPE

The main objective of this study has been to evaluate the feasibility of applying techniques from the area of Artificial Intelligence to Chemical Engineering design problems. In particular we examine the use of Expert System concepts to develop a consultant for selecting among the plethora of vapor-liquid equilibrium methods available to an engineer as he/she prepares to set up a process simulation.

Human expertise is used during different phases of problem solving: in the synthesis and analysis steps, while making decisions, and while evaluating alternatives and forming judgments (especially when the available information is uncertain and/or incomplete). This expertise is acquired both by theoretical studies and by practical experience. *The study of Expert Systems emphasizes the practical side of expertise, and it tries to formalize and encode the concepts, heuristics and strategies that form it, thus not having to depend completely on the more fragile and necessarily limited expertise of the engineer every time the solution of an engineering problem is sought.*

Expert Systems or Knowledge-Based Systems are programs containing large amounts of encoded task-specific knowledge. They have been successfully used in moderately narrow areas such as medical diagnosis, computer-system configuration and mineral prospecting (for a review of most of the systems see [Rychener 81], [Barr&Feigenbaum 81] and [Feigenbaum 77]). CONPHYDE is based on PROSPECTOR [Reboh 81], a computer consultant system intended to aid geologists in the evaluation of ore deposits. It accepts volunteered information and asks pertinent questions about the problem at hand. It deals with uncertain information. Upon arriving at a conclusion, PROSPECTOR can explain how the intermediate and final conclusions were established from the evidence supplied. Given its modular structure, it is possible to modify and improve with relative ease.

CONPHYDE has inherited these properties and applies them to the selection among available alternative methods to predict vapor-liquid equilibrium coefficients for process calculations given information about the expected concentrations, temperature and pressure, ranges, accuracy required, etc. Examples are given of CONPHYDE's performance on case studies. Based on our experience we suggest improvements required of expert systems if they are to be more useful solving engineering problems.

CONCLUSIONS AND SIGNIFICANCE

A prototype has been developed of an interactive program that acts like a consultant for the selection among Vapor-Liquid Equilibrium related properties prediction methods. The nature and size of the problem seems to be well suited for an application of an Expert System approach in its solution. Experience with the program suggest that even as it currently stands it can be very useful to aid inexperienced engineers in selecting the better methods and in gaining an understanding as to why the methods selected are thought to be better. The experience acquired in the development process of the system and the results obtained in the example runs indicate that many areas in Chemical Engineering may benefit from the application of Knowledge-Based Expert Systems.

1. INTRODUCTION

The main objective in an ongoing study is to understand the basic principles of the area of Knowledge-Based Expert Systems and the feasibility of successfully applying them to aid in the various steps of chemical engineering design.

As a first step, we describe here the creation of a prototype system, CONPHYDE, which can aid a Chemical Engineer in the proper selection of physical property estimation methods, in particular those related to vapor-liquid equilibrium. Obviously this selection is critical for a good design and simulation of a chemical process (e.g. [Zudkevitch 80]). Unfortunately, not only are there different approaches with several distinct equations respectively for each one of the various properties, but it is never possible to say that a given method is better for all possible combinations of chemicals and at all possible combinations of concentrations, temperatures and pressures. Furthermore, the reasons to prefer one method over another include not only an assessment of accuracy, but also of such factors as simplicity, availability and numerical convergence properties.

The problem at hand is a problem of *diagnosis*, in which a major part of the solution consists of informing the user about the category into which his chemical system falls. In order to be useful, the system should be flexible, interactive, natural and transparent, and have self-explanatory functions that can justify its results to the user.

PROSPECTOR was chosen to be the Expert System on which CONPHYDE would be based because not only does it fulfill the mentioned requirements, but its inference mechanisms and auxiliary systems seemed to be general enough to deal with a discipline different from the one for which it was originally created. In addition it is a system that has been commercially operated.

It should be mentioned that what could be considered as a solution to a diagnosis problem is by no means unique, since different ways to arrange and interrelate knowledge can yield practically the same behavior for the system. The area is not formalized enough to provide a proof of uniqueness or global optimality of the solution.

2. STRUCTURE

An Expert System is composed of two main parts : the Knowledge Base, where all the domain-specific knowledge is encoded and the Inference Mechanism that determines how the stored knowledge is going to be used. For flexibility reasons it is always convenient to have the Knowledge Base and the Inference Mechanism independently encoded inside the system.

Expert Systems can be classified in a number of different ways; according to the knowledge representation method that they use (e.g. Predicate Calculus, Production Systems, etc.), whether or not they deal with uncertain knowledge, according with the type of function they perform (e.g. search, planning, learning, diagnosis, etc.) or their domain of application (e.g. Medicine, Chemistry, Geology, etc.). The main differentiation among the wide assortment of Expert Systems is in the way the knowledge is represented, to the extent that this fact determines the general type of inference method used. A complete discussion and description about Knowledge Representation can be found in [Nilsson 80], or in standard textbooks on the subject. References to some of the most important Expert Systems can be found in [Rychener 81] and [Barr&Feigenbaur81].

2.1. Production Systems

We will concentrate on the description of Production Systems (RuleBased Systems) which encode knowledge in the form of *production rules*.

A production rule is a connection between two assertions specifying how the presence (or the absence) of an evidence affects the truth value of an hypothesis, that is:

IF <evidence> is present/absent

THEN <hypothesis> is {true, false} (to a degree LS/LN)

where LS is the *sufficiency measure* and LN is the *necessity measure** both of them can be referred also as *likelihood ratios*.

Because the production rules are pieces of modular and stylized form of knowledge representation, they make it possible to construct a flexible program with respect to the addition and expansion of the Knowledge Base. Furthermore, they seem to be a natural way to express knowledge, therefore making the knowledge acquisition task easier [Davis&Buchanan&Shoififfe 77].

2.2. Inference Networks

We can define a *model* as a body of knowledge about a particular *domain* of expertise [Reboh 81]. Models can be encoded as *Inference Networks*, which are tree structures consisting of nodes and arcs (we shall speak of the tree as if it were upside down). The nodes represent assertions (evidences or hypotheses) and the arcs the relation between nodes, namely a *production rule*. The bottom or leaf nodes correspond to the field evidence (direct observations) obtained from the user during a consultation, and the top node is the ultimate hypothesis to demonstrate. In general nodes can be defined as *askable*, corresponding to direct evidences or assertions that the system is going to ask about during a consultation, and *unaskable*, corresponding to assertions that can be established only by the propagation from supporting nodes which are attached below that node.

Initially the truth value of all the assertions is unknown, but, as information is gathered during the consultation, evidences become established and their probabilities are propagated to the nodes to which they are connected. The top-level assertion is that the available evidence matches the particular model. It is determined by several second-level assertions which in turn are determined by third-level assertions terminating when a node can be established from field evidence. This tree traversal is said to be executed in *consequent mode* (also known as *backward chaining* or *top to bottom*). The program can operate in a *mixed initiative dialogue mode*, making it possible for the user to volunteer information at any time during a consultation. Also, the use of production rules makes it easier to provide explanations of the program's behavior and knowledge. In addition, the system has the capability of working with uncertain information, since sometimes the evidence is not known to be certainly true or certainly false; in these cases we associate a degree of confidence with the observation of the evidence.

In PROSPECTOR there are different ways to connect the nodes in an inference network, each one defining a different relation:

1. Logical relations
 - a. Conjunction (AND)
 - b. Disjunction (OR)
 - c. Negation (NOT)
2. Plausible relations
3. Contextual relations

The first two determine the updating and propagation of probabilities. Logical relations are used when the evidences are statistically dependent, and the probability updating is made according to the Fuzzy Set Theory of Zadeh [Duda&Hart&Nilsson 76]. Plausible relations are production rules, extended to the general case where it is possible for an assertion to be supported by several evidences and in turn to support several hypotheses. The mechanism for the probability handling is based on the Bayesian Decision Theory. The derivation of the formulas can be found in [Duda&Hart&Nilsson 76], the general algorithm and some numerical examples in [Reboh 81] and [Bañares 82]. As an example of how the rules can be organized into inference networks refer to figure 2-1 where a segment of one model is depicted. Notice that the graphical representation using superposing rectangles is used only for space reasons, it is expanded into its equivalent tree structure at the top. The top-node (ACTIVITY-COEF) is supported by three independent pieces of evidence linked by plausible rules, and one of them (NON-IDEAL) is in turn connected through logical connections to three other nodes. The nodes HIGH-NON-DUM and POLY-DUM are dummy nodes representing HIGHLY-NONIDEAL and POLYMERS respectively. They are used because because a

node may be linked either by a logical relation or a plausible relation but not by both, and the original nodes are connected by means of plausible rules somewhere else in the network.

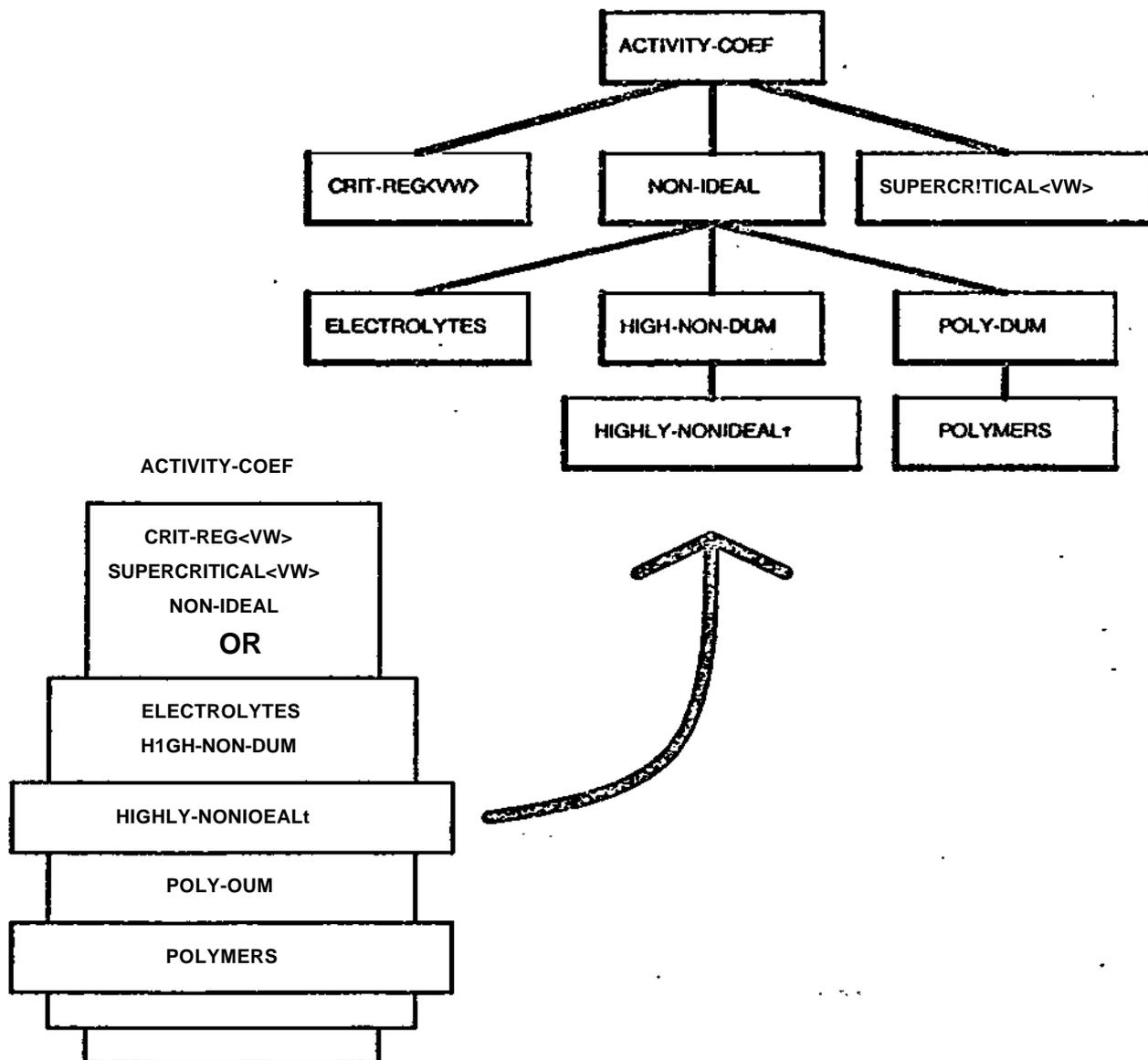


Figure 2-1: Example of an Inference Network.

Contextual relations deal with the sequencing of the questions asked in a consultation. We say that an hypothesis H_1 is a context of an hypothesis H_2 if H_1 must be established to a given extent (the extent determined by the inference network designer) before the system proceeds to investigate H_2 . The use of contexts assures that the questioning will not aimlessly jump from one subject to an unrelated*one.

In order to use the program it is necessary to input the expert's knowledge as production rules and part of the representational problem is the choice of the likelihood ratios LS and LN and a *prior probability* $P(H)$ for each hypothesis in the network (more on the inference network construction later). The user does not need to worry about these details though, and all the information he must have is about the evidences that are going to be asked. The user and the system communicate by means of *certainty measures*: $C(E|E^i)$ for the certainty the user has about an evidence given the observations, and $C(H|E^f)$ for the certainty the system has about a particular hypothesis. The range of values that a certainty can have go from -5.0 to 5.0, where 5 means that the assertion is true (present), -5 that the assertion is false (absent), and 0 means there is no information about it. The system also has the capability of asking questions that can be answered with a YES/NO reply and questions that must be responded to with a numerical value and its corresponding certainty (e.g. value of a polarity, a temperature, etc.).

2.3. Taxonomies

Taxonomies are another way to encode information in the Knowledge Base. They represent the classification of primitive concepts, organized in hierarchical structures. These structures can be taken as trees, where the nodes are the primitive concepts and the arcs the relationship between them. There are different kinds of relationships, *element relationship* e (set membership) and *subset relationship* s (set inclusion), with the possibility of making them more particular by making it possible to specify *disjoint subsets* ds and *distinct elements* de .

The advantage of using taxonomies as part of the representation arises from the fact that they naturally accommodate the hierarchical structure of some of the knowledge. Such information can be used at any time during the consultation, but it can play a very important part during the volunteering stage by permitting the system to interrelate the volunteered information given by the user and to infer properties of the primitive concepts (see initial part of the example session on section 5.2).

For illustration purposes, figure 2-2 shows a fraction of a proposed taxonomy of chemicals with its corresponding tree-structure representation.

Notice that in it would be possible to attach properties to each one of the nodes in the tree, and recognize that such properties can be inherited by its subcategories. That is, we could attach a polarity value equal to 0.0 for the node *alkane*, which has as its subcategories the nodes *methane* and *ethane*, and after declaring the presence of either methane or ethane in the chemical system the

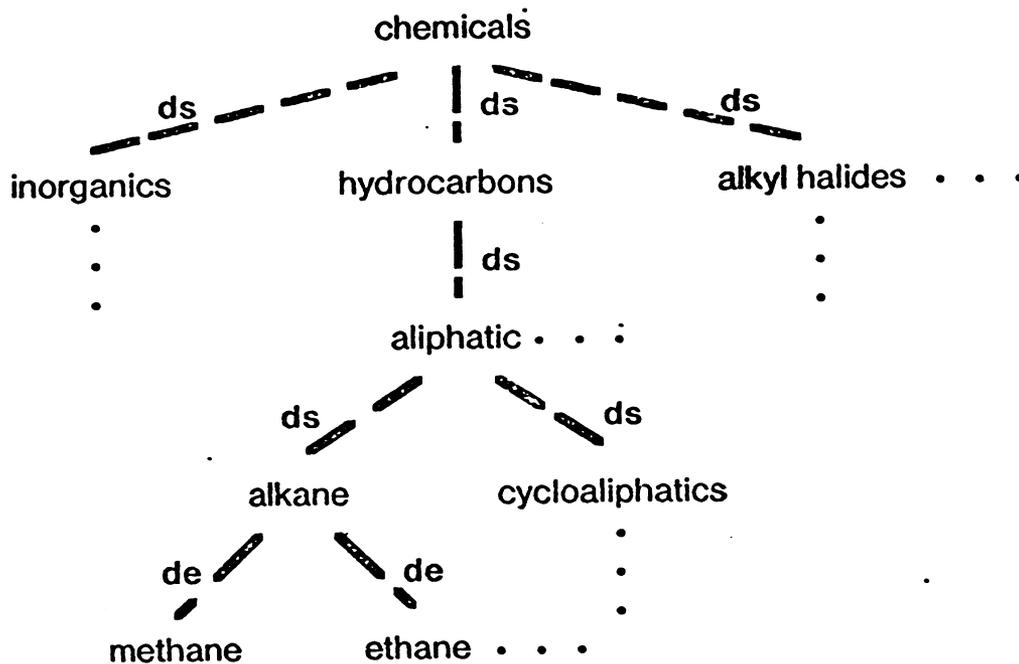


Figure 2-2: Example of a taxonomy.

taxonomy would establish that it is an alkane (and therefore all the conditional parts of the production rules of the form "IF an alkane is present/absent THEN ..." would apply), and that therefore it is non-polar.

3. VLE METHODS REPRESENTED

One of the most common doubts of a designer is the selection of prediction methods for vapor-liquid equilibria related properties. Such properties are used in the calculation of separation equipment. His decision is often a priori, even though an improper choice can lead to very serious miscalculations.

The criterion for equilibrium for the system is the equality of fugacities for each one of the components in the mixture in all the phases:

$$f_i^V = f_i^L \quad (1)$$

for this reason a method for the calculation of fugacity must be specified prior to the calculation of equilibrium.

For the vapor phase the fugacity is expressed in terms of the *fugacity coefficient* (Φ_i):

$$f_i^V = y_i \Phi_i^{VP} \quad (2)$$

For an ideal gas $\gamma_i^* = 1$ and the fugacity for the vapor phase is equal for the vapor pressure of component i .

In the liquid phase the fugacity might be represented in terms of the fugacity coefficient for the liquid phase:

$$f_i = x_i \gamma_i^* P_i^* \quad (3)$$

or in terms of the *activity coefficient* (γ_i):

$$f_i = x_i \gamma_i f_i^* \quad (4)$$

where f_i^* is the *standard state fugacity* (the fugacity of component i at the temperature of the system and at some arbitrarily chosen pressure and composition). For vapor liquid equilibrium two main kinds of approaches can be used. We can use a model for the activity coefficient in the liquid phase coupled with an equation of state for the vapor phase [Chao&Seader 61], or use an equation of state for both phases.

3.1. Activity coefficients.

The advantage of using activity coefficients in the liquid phase is that it is possible to represent highly non-ideal mixtures (mixtures that contain strongly polar or hydrogen bonded components, polymers or electrolytes) using relatively simple models. Their use has disadvantages also, their main limitation being that it is necessary to specify a standard state fugacity f_i^* for each component in the mixture. Such a specification is always arbitrary when supercritical components are present, and makes it difficult to use activity coefficients near the critical region.

There are two alternatives for the acquisition of the activity coefficients for mixtures without electrolytes: interpolating/extrapolating experimental data and estimating those coefficients.

3.1.1. Interpolation/Extrapolation of experimental data.

From the definition of the total excess Gibbs free energy (G^E) for a binary solution in terms of the activity coefficients and the Gibbs-Duhem equation it is possible to obtain a thermodynamically consistent relationship between an activity coefficient and G^E . Using this relation our problem is reduced to establishing a mathematical expression for G^E as a function of composition, and the obtaining of numerical constants in that function from experimental data. The above mentioned mathematical expressions are the set of models from which we must select the one which is most appropriate for our situation. The model can be empirical or semiempirical, and all of them contain adjustable parameters. These parameters have a dependence on temperature, but if the working temperature range is small enough, we can consider them temperature independent (the size of an acceptably small range depends on the equation).

As expected the accuracy in representing the experimental data generally increases with the number of parameters used, but so will then the number of experimental measurements needed to fit them. In general the maximum number of parameters used is three, because the uncertainties of the experimental data are often greater than the suspected gain in accuracy for the use of one more parameter.

One family of models is formed by the so called Margules equations (or Redlich-Kister expansions). There are two-suffix, three-suffix and four-suffix Margules equations (called Margules-1, Margules-2 and Margules-3 in this work since they have one, two and three parameters respectively) which are quadratic, third order and fourth order equations in mole fraction. Slightly more complex and with two adjustable parameters, the Van Laar equation is much easier to handle than the newer models: Wilson, UNIQUAC and NRTL.

Equations with one parameter are able to predict activity coefficients only for mixtures of similar components, while the two parameter equations can handle moderately nonideal mixtures. The Wilson equation is more useful than the Van Laar equation for strongly nonideal liquids, but it fails to predict the liquid separation into two phases, which both UNIQUAC and NRTL can predict.

UNIQUAC is more complex than NRTL but it uses only two parameters (while NRTL uses three) and has a better theoretical basis that enables it to predict the behavior of polymer solutions. The UNIQUAC equation distinguishes two kinds of molecular interactions and has a term for each:

- its combinatorial part deals with differences of size and shape of the molecules
- its residual part deals with differences in intermolecular forces

We will find this same pattern in more advanced models.

3.1.2. Estimation of coefficients.

We can estimate activity coefficients by means of correlations based on limited experimental data. The difference in this case is that the system to predict is not in general the same for which experimental data was obtained, but at most similar.

The Regular Solution Theory or Scatchard-Hildebrand equation assumes the excess entropy equal to zero. The model is good enough for hydrocarbon mixtures, but it cannot predict polar mixtures property.

Another alternative is the use of Group Contribution Methods. They correlate properties in terms of functional groups rather than molecules. Since there are a relatively small number of different

functional groups, it is simpler to obtain experimental data that represents the contribution of individual groups rather than for entire molecules. Like the UNIQUAC equation, both group contribution methods, ASOG and UNIFAC, account for two kinds of interactions. In the case of ASOG, the first term is represented by the athermal Flory-Huggins equation, whereas the second term can be considered to be the Wilson equation applied to functional groups. UNIFAC uses the combinatorial part of UNIQUAC for the first part and a specific equation for the solution of groups to account for molecular interactions.

3.2. Equations of state.

Equations of state are applicable to normal fluids (rare gases, nitrogen, oxygen, carbon monoxide and hydrocarbons), carbon dioxide, hydrogen sulfide, hydrogen and slightly polar substances and they work even at high pressures. They can be used while working with supercritical components since the specification of liquid standard states is avoided. Other advantages over the use of activity coefficients are:

1. Continuity in the critical region is assured (although convergence may be hard to achieve).
2. A wide variety of properties can be derived from them [Reid&Prausnitz&Sherwood 77],

Some of the most important properties that can be predicted using equations of state are:

- compressibility factor: $z = PV/RT$
- fugacities: $\ln(f/P)$
- vapor pressure: P_{vop}^{\wedge}
- density: ρ (inverse of the specific volume)
- departure functions:
 - $A - A^{\circ}$. Helmholtz free energy
 - $S - S^{\circ}$. Entropy
 - $H - H^{\circ}$. Enthalpy
 - $U - U^{\circ}$. Internal energy
 - $G - G^{\circ}$. Gibbs free energy
- Heat capacities: C_p

The disadvantage to their use is that they cannot deal with a number of situations which coincide with the cases where the Corresponding States Theory also does not apply, namely dealing with:

- polar compounds
- large molecules (polymers), and
- electrolytes

Besides no equation of state works for all densities with the same set of coefficients. They generally fail to predict the liquid phase very accurately since they are fitted to represent the vapor phase. In addition, they are very sensitive to the selection of mixing rules and binary parameters.

It should be clear from equations (1), (2) and (3) that the solution of a vapor liquid equilibrium problem is equivalent to the calculation of the K_i factors since:

$$\Lambda = \sum_i f_i r_i \quad (5)$$

and

$$K_i = \frac{y_i}{x_i} = \frac{\phi_i^L}{\phi_i^V} \quad (6)$$

Several classifications of equations of state have been proposed, according to their theoretical basis [Barnes&Bazii 77], according to their historical development [Fredenslund&Rasmussen&Mollerup 81], according to their degree of generalization [Oellrich&Plocker&Prausnitz&Knapp 81] and many more. The one that has been adopted in this work is a mixture of the above mentioned ones. We group the equations of state into:

1. Van der Waals family (semiempirical equations).

A large number of equations belong to the Van der Waals family. These equations are of the semiempirical kind, and they represent the pressure as a sum of two terms, one accounts for repulsion and the other for attraction [Barnes&Bazii 77]. Van der Waals proposed his equation in 1873, and it contains two constants that account for deviations from ideality; the attraction forces (a_j) and the molecular size (b). These simple considerations permit, for the first time, prediction of the existence of two phases. Its most important modification has been the equation proposed by Redlich and Kwong in 1949 from which more than one hundred modifications have been made. Among them are the ones by Wilson [Wilson 66], Barnes and King [Barnes 73], Soave [Soave 72], and Peng and Robinson [Peng&Robinson 76]. All these equations are computationally easy to handle. They are also called cubic equations of state because the compressibility factor z is expressed as a cubic equation in terms of the adjustable constants a and b , the universal gas constant R , the temperature and the pressure. The Soave and the Peng and Robinson equations are particularly accurate and have a wide range of application. They yield very similar results although it seems that the Soave equation has larger deviations near the critical region and in the liquid phase [Barnes&Bazii 77],

2. Benedict-Webb-Rubin family (multiparametric).

Empirical equations of state have, in general, many adjustable constants that in the cases of some equations like the one by Benedict Webb and Rubin have been later generalized in terms of the critical constants [Yamada 73]. The Benedict Webb and Rubin equation has eight constants and predicts properly paraffin and olefin systems, but for more complex systems it needs binary interaction parameters which are difficult to determine given the complexity of the equation [Zudkevitch&Joffe 70]. These parameters have a

strong dependence on pressure and temperature and the whole equation is not mathematically stable at high pressures. Its validity range is relatively small.

Starling proposed an equation with eleven constants [Starling 71] that works for larger ranges of temperature and density. It is very accurate for the liquid phase and is appropriate for the calculation of bulk properties (density, residual enthalpy and heat capacity) but inadequate for the prediction of partial molar properties, where the derivatives with respect to composition are involved [Fredenslund&Rasmussen&Mollerup81].

Equations like the one proposed by Bender, with twenty constants for a pure component and three constants per pair in the mixture, are very accurate but restricted in their use to those systems for which enough data exists. * They are also very time consuming and a proper extension to take into account multicomponent systems is not a simple problem.

The Lee and Kesler equation of state [Lee&Kesler 75] is an analytical BWR version of the acentric factor (w) correlation by Pitzer (Correspondent States Theory). It proposes that the compressibility factor (z) consists of a combination of a term for spherical molecules ($z^{(0)}$) and a correction term ($z^{(1)}$):

$$z = z^{(0)}(T_R, V_R) + \omega z^{(1)}(T_R, V_R)$$

The equation extends the range of application of Pitzer's correlation. Its coefficients are general, and therefore mixing rules are only used for the pseudocritical variables. Binary interaction parameters have been correlated in terms of the product of critical temperature and volume. This correlation applies even for wide boiling range mixtures (e.g. mixtures containing hydrogen, hydrogen sulfide, carbon dioxide). Its most important modification deals especially well with asymmetric mixtures [Plocker&Knapp&Prausnitz 78].

3. Virial family (theoretical).

The virial equation of state calculates the compressibility factor of a gas (z) as a power series in terms of density. At constant composition the coefficients of the series are exclusively a function of temperature. Since the equation has a solid base in statistical mechanics its coefficients can be calculated from functions related to intermolecular potentials. For engineering purposes its use is practical only when its convergence is fast (i.e. only when two to three terms of the series are used), this only occurs at moderate pressures. As a rule of thumb, the virial equation of state is inaccurate when the system density is larger than half its critical density.

Numerous review articles dealing with equations of state have been published, but new equations

appear constantly, making very difficult to keep up to date. Some fine reviews can be found in [Chao&Robinson 79], [Fredenslund&Rasmussen&Mollerup 81], [Leland 80], [Martin 79], and [Oellrich&Plocker&Prausnitz&Knapp 81].

4. DEVELOPMENT OF AN EXPERT SYSTEM.

The process of constructing, modifying, testing and maintaining the Knowledge Base is referred to as the *Knowledge Acquisition Process*. It has been observed that this process is the critical bottleneck in the development of an Expert System. During this process domain-related knowledge (e.g. conceptual primitives, production rules) must be obtained from the *Domain Expert* and encoded into the system. A domain expert is an authority on the subject for which the model is being developed, in the case of PROSPECTOR an exploration geologist.

Ideally it would be desirable to have a tool that would permit a natural language interaction (i.e. in English) between the domain expert and the program, but such a stage has not been reached. Therefore, a collaboration with a *Knowledge Engineer* (KE), a computer scientist that translates the DE knowledge to a computer representation, is necessary. The knowledge engineer works with support tools to make this task more efficient and fast. Such tools play a very important role in the development step, taking into account that the problem involves large representation structures, and that such structures interact, overlap (sometimes several DE's and KE's work on the same program) and change in time (since the knowledge base should be able to be updated).

4.1. Methodology of the knowledge engineering process

A useful methodology to encode the expert's knowledge is based on various elements:

- interviewing techniques
- principles for determining the overall structure of the model
- tools for the interactive construction of the model
- modification and testing of the models
- use of performance analysis procedures

A logical sequence of steps for the knowledge acquisition would include:

1. Initial preparation

- familiarization with the subject of study, in which both experts -the domain and the knowledge engineer- learn about each other's terminologies, since in general the KE knows as little about the domain as the DE knows about expert systems.
- preparation of the case studies to be covered by the system.

2. Initial Design of the model

- development of a hierarchical network structure of the model not taking in account the future numerical values
 - o identifying relevant factors that either establish or rule out different hypotheses
 - o organizing factors as field observations (defined as *askable* nodes) and indirect observations (defined as *unaskable* nodes)
- design of the inference network, based on considerations such as the importance of the spaces, independence from each other, etc., with basically two kinds of decisions:
 - o choice of the type of connections betw/een the spaces (OR, NOT, AND, RULES or some combination of these depending on their interaction)
 - o choice of the question sequencing and control (taking in account that questions are going to be asked in a depth-first fashion (Top-to-bottom and left-to-right), and that to change this sequence contexts can be used)
- selection of numerical values for parameters
- specifying how questions are going to be asked (their format)
- specifying what kind of answer is expected from the user (YES/NO, a certainty C or answer to a numerical rule -range of compositions, temperatures, etc.-)

3. Installation of the model in the consultation system

- creation of a computer file containing the model. This transfer is made by RENE (REsident Network Editor).
- performance of the first test runs. Test of the basic flow of control. In general the behavior is not equal to the one predicted, therefore it is changed by the KE by
 - o adding contexts
 - o using the *block* structure (defining a set of nodes as a unit)
 - o modifying some prior probabilities and askabilities
- test of the model in known cases
 - o creating a questionnaire for the model and getting it solved by the OE
 - o running the test cases in batch mode, changing the numerical values and network structure so it matches the expert's conclusions
- completion of the description of the model. This will allow the Semantic Network Matcher to relate statements, find probabilistic constraints in the inference network and uncover interaction and overlap of the models, as well as allow the user to volunteer information
 - o creating a semantic representation of several spaces (statements in the knowledge base, not used in the current version of CONPHYDE)
 - o expanding the taxonomies
 - o completing the explanation and rephrasing texts of the spaces

4. Performance and sensitivity analysis.

- assessment of accuracy and consistency in the prediction of the domain expert's judgment.
- test of *robustness*, i.e. sensitivity with respect to uncertainties in the observations, and with respect to variations in the parameter values.

5. Revision of the model. Identification of the sections of the model that need improvement.

4.2. Tools for encoding the models (Features of PROSPECTOR).

KAS (Knowledge Acquisition System) has three kinds of tools for the execution of the above mentioned tasks:

1. Tools to maintain the external descriptions of the Knowledge Base

- a. A formal language for the representation of taxonomies, inference and semantic networks
- b. A parser that interprets the mentioned external representation and generates internal network structures

2. Tools to assist in the model building task

- a. RENE (REsident Network Editor)
- b. Semantic Network Matcher

3. Tools to evaluate model performance

- a. Automatic questionnaire generation. It lists all the questions that might be asked during consultation
- b. Control in batch and save mode. The first one allows the use of an external file as input for a consultation, whereas the second one saves the user's answers in an external file. These features can be used simultaneously.

5. CONPHYDE

This area of application of Knowledge Based Systems was chosen not only because of the great importance that separation has in all industrial plants, but also because it has the characteristic features that make the use of Expert System techniques feasible and convenient:

- designers of chemical plants consult experts on the best choice of vapor-liquid equilibria methods for design and simulation tasks.
- the area is structured enough. It is possible to connect observable characteristics of a chemical system to the expected performance of an estimation method. This allows a systematic search for a preferred method.
- there are uncertainties both in the observable evidences and in their quantitative

influence on the use of the suspected optimal prediction methods; therefore, it is not possible to make the choice automatically.

Several models have been constructed representing some of the available knowledge regarding the use of estimation methods in the prediction of vapor-liquid equilibrium. In order to construct such models, modular segments of knowledge have been represented by means of production rules. A few dozens of these rules are present in CONPHYDE, and a list of them can be found in the appendix. Several others not yet encoded into the system can be found in [Bafiares 82].

The production rules are ordered into several inference networks, each one of them representing a different model. For a given level of expertise encoded in the system, the choice of numerical values for the likelihood ratios together with the ordering of the rules to constitute networks are the key factors in the development of a consistent Expert System, and many versions have been tried. Unfortunately the only way to find the numerical values and construct the inference networks within the PROSPECTOR framework is by trial and error. These are adjusted using test cases for which reliable advice exists. It should be obvious that the ultimate choice is subjective and that there is no proof of optimality.

Although CONPHYDE only represents a few of the many prediction methods available, one of its advantages is the easiness with which it can be expanded by adding more methods. In view of possible expansions, it was considered inconvenient to represent each equation with a model. Instead each model represents a set of similar methods.

Five models have been proposed. The first three deal with equations of state and the remaining two with activity coefficient formulas. Table 5-1 lists the models and the prediction methods contained in each one of them.

All the models have analogous structure (shown in figure 5-1), the top node inherits the certainty of the most promising method contained in the model. This is achieved by connecting all the methods to the top node by means of an OR relation. Each prediction method then has three different kinds of evidence supporting it. On the first level, some evidence is considered that notes that the method is either an equation of state or an activity coefficient method. It thus inherits the general characteristics of either approach (see Section 3 on VLE Methods Represented). On a second level, evidence related to the use of a general type of method is considered. On the third and last level, very specific rules point uniquely to a method according to its individual characteristics. In the following section the structure of the VANDERWAALS models is sketched and explained in detail. The exact

MODEL	PREDICTION METHODS
VANDERWAALS	Soave
	Peng and Robinson
BENEDICT	Benedict Webb and Rubin
	Starling
	Lee Kesler (Prausnitz, Knapp and Plocker modification)
VIRIAL-TOP	Virial
ADJUSTABLE	Margules-1 (one parameter)
	Margules-2 (two parameters)
	Van Laar
	Wilson
	UNIQUAC
	NRTL
	PREDICTABLE
	UNIFAC
	Regular Solution Theory

Table 5-1: Table of prediction methods included in the models.

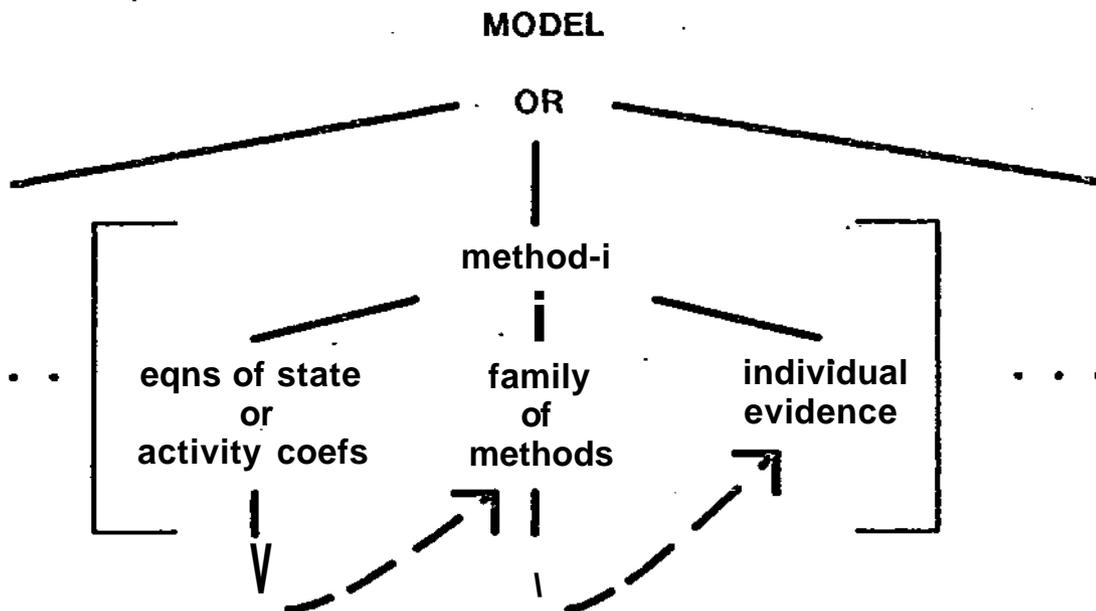


Figure 5-1: General Structure of the Models.

structures of the rest of the models together with the computer representation and the specification of the numerical values for the *rule strengths* and *prior probabilities* is given in [Bañares&Rychener&Westerberg 83]. ;

5.1. VANDERWAALS model.

Let us analyze the VANDERWAALS model in more detail. This first model groups the cubic equations of state. It contains the Soave and the Peng and Robinson equations. See figure 5-2.

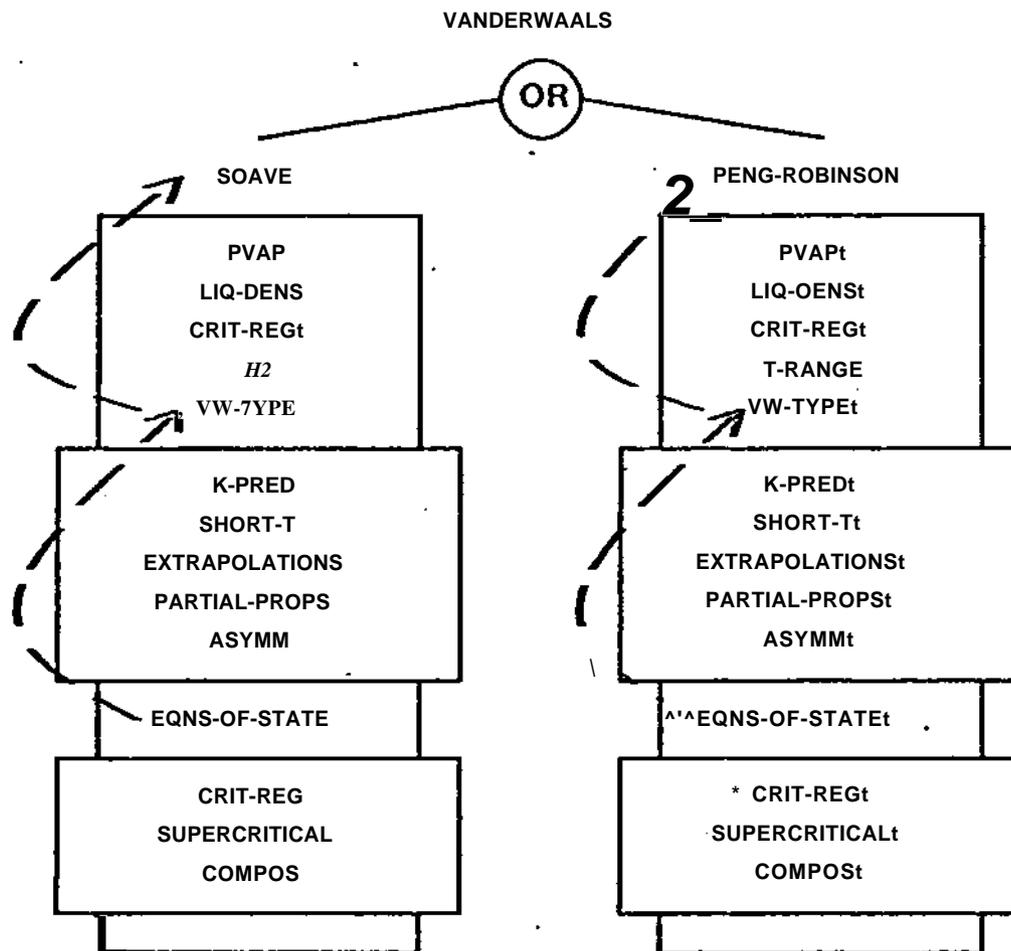


Figure 5-2: VANDERWAALS model.

Words in upper-case letters are names of nodes, and they are chosen so as to give an idea of their function. If a certain node has appeared before in the model an <t> sign is appended inside angle brackets. The representation is the same one explained in figure 2-1. The context relations are represented by broken lines and they control the question sequencing during a consultation.

CONPHYDE starts by trying to establish the EQNS-OF-STATE node (whether an equation of state or an activity coefficient method seems more promising). If the user has no information about this node, CONPHYDE asks lower level evidence: is the chemical system near the critical region? (CRIT-REG); is there is reason to suspect the presence of a supercritical component? (SUPERCRITICAL);

whether or not the system is entirely composed of a mixture of nonpolar or slightly polar compounds, normal fluids, carbon dioxide, sulfhydic acid or hydrogen (COMPOS). Once EQNS-OF-STATE has been established, the feasibility of application of a cubic equation of state is investigated (VW-TYPE). The VW-TYPE node is supported by evidence that is related to the general properties of the cubic equations: their semiempirical nature (EXTRAPOLATIONS, PARTIAL-PROPS); limitations in the prediction of asymmetric mixtures (ASYMM); algebraic simplicity (SHORT-T); encouraging experimental results (K-PRED). At last CONPHYDE investigates how the individual properties of each method apply to the problem at hand. Notice that some nodes are present in both equations, but they affect them with a different strength.

The certainty of the most successful method in the model is propagated to the top node (VANDERWAALS) by means of the OR relation.

5.2. Example session using CONPHYDE.

To illustrate some of the ideas exposed in this article we present an example run of CONPHYDE. The chemical system used consists of a mixture of light products obtained from the cracking of a heavy fraction. The mixture may contain hydrogen and water in relatively important concentrations and the conditions could be near the critical region, but they can be considered constant. The user has computer time limitations.

The example is edited but only to make the output shorter, all explanations and questions have been unchanged. In the example the answers and command given by the user are underlined, comments about specific features of the system are enclosed between curled parenthesis.

» **RUN**

This version of CONPHYDE contains models for the following subsystems:

- 1 A physical property prediction method that uses activity coefficients in the liquid phase, (the parameters of the model are estimated, it contains the ASOG, UNIFAC and Regular Solution Theory)
- 2 A physical property prediction method that uses activity coefficients in the liquid phase with adjustable parameters, (contains the Margules, Van Laar, Wilson, UNIQUAC and NRTL methods)
- 3 A physical property prediction method that uses equations of state of the virial type (contains the Virial equation of state)
- 4 A physical property prediction method that uses equations of state of the multiparametric type (contains the Benedict Webb & Rubin,

its Starling modification and the Lee Kesler equations)

5 A physical property prediction method that uses equations of state of the Van der Waals type (two parameters).(contains the Soave and Peng & Robinson modifications of the Redlich-Kwong equation)

Different components have properties that make the use of a prediction method appropriate or not. To help in selecting a subsystem to focus on, it would be useful for me to know any characteristics of the chemical system that are immediately available.

{Volunteering stage}

Do you have information about the conditions? (Not in operation now). NO

Do you have information about components present in the mixture?

(This can include items that are uncertain or vague.) YES

Please give a component and its associated certainty value.

(If you need instructions, type HELP; terminate by typing DONE.)

1 system contains hydrogen. (RULE 57). TAXONOMY: HYDROGEN

2 the mixture has a wide boiling temperature range (e.g. mixtures that contain quantum gases etc.). (RULE 56). TAXONOMY: QUANTUM GASES.

3 There are electrolytes present in the mixture. (RULE 14). TAXONOMY: ELECTROLYTE.

4 Polymers are present in the mixture. (RULE 13). TAXONOMY: POLYMER.

V,TRACE {It is possible to trace which nodes are affected by the user's answers }

OK

1 J 2_{Item 1 is present with a certainty of 2 (i.e. we are more or less sure that hydrogen is present)}

Changing the certainty of (* system contains hydrogen. (RULE 57). TAXONOMY: HYDROGEN)

from 0.0 to 2.0

Changing the score for: SPACE-BENEDICT.BENEDICT from 0.0 to 2.0

Changing the certainty of (* the use of the Soave eqn. of state is advisable)

from 0.0 to -.798864

Changing the certainty of (*
the use of the Benedict Webb and Rubin equation of state is advised)
from 0.0 to -.8

Changing the certainty of (•
the mixture has a wide boiling temperature range (e.g. mixtures
that contain quantum gases etc.). (RULE 56). TAXONOMY: QUANTUM GASES.)
from 0.0 to 2.0

{and so on...}

2:NOTRACE {stop tracing}

OK

2:2 J_ {Quantum Gases present with a certainty of 1, remember the
fact that hydrogen is a subset of quantum gases according to
the taxonomy in figure 2-2. CONPHYDE checks
consistency and informs the user of his error.}

** This certainty seems inconsistent with your previous statements:

1 - 2) system contains hydrogen. (RULE 57). TAXONOMY: HYDROGEN
(certainty 2.0)

Do you want to changes the certainty for Statement 2 ? YES

New certainty = ?2

2) the mixture has a wide boiling temperature range (e.g. mixtures
that contain quantum gases etc.). (RULE 56). TAXONOMY: QUANTUM GASES.
(certainty 2.0)

3:DONE

.....

Presently, the subsystems in CONPHYDE and their certainties are:

- 1) BENEDICT model (certainty 1.6)
- 2) VANDERWAALS model (certainty 0.0)
- 3) PREDICTABLE model (certainty 0.0)
- 4) VIRIAL-TOP model (certainty 0.0)
- 5) ADJUSTABLE model (certainty 0.0)

On the basis of the information you gave me, I would work first on
the first hypothesis.

Which (if any) of these hypotheses would you prefer to pursue? ±
OK. (Use the STAY or SWITCH commands to intervene.)

3 -- To what degree do you believe that the use of equations of state is preferred over the use of activity coefficient methods ? WHY

{Use of explanatory functions during questioning.}

If you have any reason to believe that either method is better than the other, it would not be necessary to investigate whether the use of equations of state is more appropriate than the activity coefficient methods given the nature of the system.

3 - To what degree do you believe that the use of equations of state is preferred over the use of activity coefficient methods ? p

{Since the user has no direct information CONPHYDE tries to establish it from more basic evidences (see VANDERWAALS model).}

4 - To what degree do you believe that the system is entirely composed of a mixture of nonpolar or slightly polar compounds, normal-fluids (rare gases, N₂, O₂, CO or hydrocarbons), CO₂, H₂S or H₂. ? ^f

5 - To what degree do you believe that supercritical components are present in the mixture (e.g. N₂, CH₄, C₂H₆, CO₂ at ordinary temperatures) (RULE 3) ? 2

6 - To what degree do you believe that the operating conditions are near the critical region. (RULES 1, 25, 61) 14

7 - To what degree do you believe that short computing times are necessary. (RULES 33,89) ? 5

I am inclined to consider a different hypothesis.

{CONPHYDE switches its current hypothesis and continues.}

18 - If polar components are present in the mixture, what is the polarity of the most polar of the components? (Answer with a range in polarity values, e.g. for slightly polar components the answer would be: 0.0 0.1) (RULE 12) ? 1.3 1.5 *{Range of polarities of most polar component.}*

What is your confidence in this estimate? 2^5

20 « Polymers are present in the mixture. (RULE 13). TAXONOMY: POLYMER.
? NO {Some nodes can be asked with a YES/NO question.}

The following models have been considered:

- 1) A physical property prediction method that uses equations of state of the Van der Waals type (two parameters).(contains the Soave and Peng & Robinson modifications of the Redlich-Kwong equation) (certainty 4.91847)
- 2) A physical property prediction method that uses activity coefficients in the liquid phase with adjustable parameters, (contains the Margules, Van Laar, Wilson, UNIQUAC and NRTL methods) (certainty 4.44114)
- 3) A physical property prediction method that uses equations of state of the multiparametric type (contains the Benedict Webb & Rubin, its Starling modification and the Lee Kesler equations) (certainty 3.90707)
- 4) A physical property prediction method that uses equations of state of the virial type (contains the Virial equation of state) (certainty -2.58095)
- 5) A physical property prediction method that uses activity coefficients in the liquid phase, (the parameters of the model are estimated, it contains the ASOG, UNIFAC and Regular Solution Theory) (certainty -3.220875) •• Current hypothesis *#

For which of the above do you wish to see additional information?

(Type ? for available options) ±

{Explanatory functions can justify CONPHYDE's choice.}

On a scale from -5 to 5, my certainty that

- 1) A physical property prediction method that uses equations of state of the Van der Waals type (two parameters).(contains the Soave and Peng & Robinson modifications of the Redlich-Kwong equation) is now 4.91847.

There are two favorable factors; in order of importance:

1:1) the use of the Peng & Robinson equation of state is advisable
(certainty 4.91847) •* establishes 1 *•

1: 2) the use of the Soave eqn. of state is advisable (certainty 4.648715)

For which of the above do you wish to see additional information? ±

.....

I suspect that

1.1) the use of the Peng & Robinson equation of state is advisable
(certainty 4.91847).

There are several favorable factors; in order of importance:

1.1:1) the use of equations of state is preferred over the use of
activity coefficient methods (certainty 3.080829)

1.1:2) the set of properties that confirm or rule out the use of a Van
der Waals equation of state (certainty 4.646151)

1.1:3) You were sure that the operating conditions are near the
critical region. (RULES 1, 25,61) (certainty 4.0)

There is one negative factor with neutral effect that, if positive, could have
been significant:

1.1:4) You denied that working conditions are in large temperature
range. (RULES 40,63) (certainty -4.0) ** would have helped if positive **

There are two factors that have not yet been considered; in order of importance:

1.1:5) the accuracy in the prediction of liquid densities (or properties
in the liquid phase) is important. (RULE 36) (certainty 0.0)

1.1:6) the accuracy in the calculation of vapor pressures should be
stressed. (RULE 35) (certainty 0.0)

For which of the above do you wish to see additional information? ±

.....

I suspect that

1.1.1) the use of equations of state is preferred over the use of activity
coefficient methods (certainty 3.080829).

There are two favorable factors; in order of importance:

1.1.1:1) You were sure that the operating conditions are near the
critical region. (RULES 1, 25, 61) (certainty 4.0)

1.1.1:2) You suspected that supercritical components are present in the mixture (e.g. N₂, CH₄, C₂H₆, CO₂ at ordinary temperatures) (RULE 3) (certainty 2.0)

There is one unfavorable factor:

1.1.1:3) the system is entirely composed of a mixture of nonpolar or slightly polar compounds, normal-fluids (rare gases, N₂, O₂, CO or hydrocarbons), CO₂, H₂S or H₂. (certainty -1.0)

For which of the above do you wish to see additional information? NONE

There are several favorable factors; in order of importance:

2.1:1) the use of the UNIQUAC (Universal Quasi Chemical) model is advisable (certainty 4.44114) ** controlling factor •*

2.1: 2) the use of the Wilson model is advisable (certainty 2.926204)

2.1:3) the use of the Margules equation with two adjustable parameters is advisable (certainty 2.511833)

2.1:4) the use of the Van Laar equation is advisable (certainty 2.511833)

For which of the above do you wish to see additional information? ±

I suspect that

2.1.1) the use of the UNIQUAC (Universal Quasi Chemical) model is advisable (certainty 4.44114).

There are two favorable factors; in order of importance:

2.1.1:1) You were sure that short computing times are necessary. (RULES 33,89) (certainty 5.0)

2.1.1:2) the system has important molecular interactions. (RULES 92,94) (certainty 4.193979)

There are two negative factors with neutral effect that, if positive, could have been significant; in order of importance:

2.1.1:3) You denied that Polymers are present in the mixture. (RULE 13). TAXONOMY: POLYMER, (certainty -5.0) •* would have helped if positive **

2.1.1:4) You denied that working conditions are in large temperature range. (RULES 40,63) (certainty -4.0) *• would have helped if positive ••

There is one uncertain factor whose score may be subject to change:

2.1.1:5) the use of activity coefficient methods is preferred over the use of equations of state (certainty .5205634)

For which of the above do you wish to see additional information? 2

.....
I suspect that

2.1.1.2) the system has important molecular interactions. (RULES 92,94) (certainty 4.193979).

There are tv/o favorable factors; in order of importance:

2.1.1.2:1) You suspected that the liquid splits in two phases. (Vapor-liquid-liquid equilibrium) (certainty 2.0)

2.1.1.2:2) You suspected that If polar components are present in the mixture, what is the polarity of the most polar of the components? (Answer with a range in polarity values, e.g. for slightly polar components the answer would be: 0.0 0.1) (RULE .12) (certainty-2.5)

{Summary of CONPHYDE's certainties for the example problem.}

PENG-ROBINSON<VANDERWAALS> 4.91847
 SOAVE<VANDERWAALS> 4.648715
 UNIQUAC<ADJUSTABLE> 4.44114
 LKPKP<BENEDICT> 3.90707
 MARGULES-1 <ADJUSTABLE> 3.069536
 WILSON<ADJUSTABLE> 2.926204
 MARGULES-2<ADJUSTABLE> 2.511833
 VAN-LAAR <ADJUSTABLE> 2.511833
 NRTL<ADJUSTABLE> 0.2518605
 STARLING<BENEDICT> -.498039
 VIRIAL<VIRIAL-TOP> -2.58095
 ASOG<PREOICTABLE> -3.22088
 UNIFAC<PREDICTABLE> -3.22088
 REGULAR-SOLN<PREDICTABLE>-3.27887
 BWR<BENEDICT> -3.80039

6. FUTURE DEVELOPMENTS

The single most important factor that would improve the operational capability of CONPHYDE would be the expansion of its inference networks and taxonomies, both in depth and breadth. This would imply the cooperation of an expert in physical property prediction methods to provide the kind of knowledge that is not possible to find in books and articles. This expansion would be entirely contained within the original capabilities of PROSPECTOR, but further modifications and additions to its inference mechanism and knowledge base would give CONPHYDE an industrial applicability. This need had been recognized in the initial thesis work, and in a later independent publication [Motard 83]. The possibility of substantial numerical computation during a CONPHYDE consultation seems to be basic not only for the present system, but for further applications of Expert Systems in Chemical Engineering.

Given its characteristics, CONPHYDE could be used as an educational tool in the future. It could be the first step towards the creation of a powerful and very useful set of programs which would aid the chemical engineer in the design of a chemical plant. It could be one of the many "experts" needed in the design problem (similar to the interaction described in [Lenat 75]), or provide guidance to accomplish the first depth-first step suggested by [Westerberg 81] in the understanding and evaluation of expert systems for the creation of an "intelligent" interface for the solution of complex design problems [Westerberg&Rychener 82].

I. APPENDIX: Production Rules

The production rules that were used in building the system are listed. In addition there are rules that were not encoded in the knowledge base but will be useful in refining the system. The complete list is given in [Bañares 82]. The production rules are ordered according to the prediction method they support, and the kind of criteria on which they are based.

Number	Kind of rule
1 - 10	Related to the approach to use for the liquid phase and that advise the use of equations of state.
11-20	Related to the approach to use for the liquid phase and that advise the use of activity coefficients.
21 - 30	Related to the operating conditions and their relation to the equations of state.
31 - 50	Related to the degree of accuracy/complexity of the equations of state.

- 51 - 60 Related to the operating compositions of the equations of state.
- 61 - 80 Related to the operating conditions and their relation to the activity coefficients.
- 81 - 90 Related to the degree of accuracy/complexity of the activity coefficients.
- 91 - 110 Related to the operating compositions of the activity coefficients.

Notice that the numbering has been chosen so as to coincide with the original reference (i.e. [Bañares82j]).

- (I) IF the operating conditions are near the critical region

THEN use of equations of state in the liquid phase is recommended
AND retrograde condensation may be found
AND convergence may be hard to achieve.
[Prausnitz 77]

- (3) IF supercritical components are present

THEN use of equations of state in the liquid phase is recommended.
[Reid&Prausnitz&Sherwood 77]

- (II) IF working with a highly non-ideal liquid mixture

THEN use activity coefficients in the liquid phase.
[Prausnitz 77]

- (12) IF polar components are present in the mixture

THEN use activity coefficients for the liquid phase.
[Prausnitz 77]

(13) IF polymer components are present in the mixture

THEN use activity coefficients for the liquid phase.
[Reid&Prausnitz&Sherwood 77]

(14) IF electrolytes are present in the mixture

THEN use activity coefficients for the liquid phase.
[Prausnitz 77]

(25) IF operating conditions are near the critical region

THEN the Peng-Robinson equation of state is advised due to
computational reasons (easier convergence).
[Juárez&Bañares 80]

(26) IF the density of the mixture is less than or equal to one half
or less than the critical density

THEN the Virial equation of state is able to predict behavior.
[Tsonopoulos&Prausnitz 69]

(27) IF working at low temperatures
AND working at high densities ($\rho_M > 1.5 \rho^*$)

THEN the Benedict-Webb-Rubin original equation of state (BWR) is
not recommended.
[Stotler&Benedict 53]

(28) IF working at low temperatures

THEN the Starling modification of the BWR is recommended among the equations of the virial type.

[Fredenslund&Rasmussen&Mollerup 81]

(29) IF working at $0.3 < Tr < 4$, and, $0 < Pr < 10$

THEN the modification of Plocker, Knapp and Prausnitz of the Lee Kesler equation of state (LKPKP) yields reliable results.

[Oellrich&Plocker&Prausnitz&Knapp 81]

(32) IF calculations are going to require extrapolations

THEN avoid the use of empirical equations of state
AND avoid the use of complex equations of state.

(33) IF fast calculations are required

THEN use equation of state of the VW type.

[Fredenslund&Rasmussen&Mollerup 81]

(34) IF K factors predictions are desired

THEN use a VW type equation of state.

[Fredenslund&Rasmussen&Mollerup 81]

(35) IF vapor pressure calculations are going to be performed

THEN use Soave equation of state.

[Reid&Prausnitz&Sherwood 77]

(36) IF liquid densities are going to be calculated

THEN use of Soave equation of state is discouraged.

[Reid&Prausnitz&Sherwood 77]

(37) IF partial molar properties are going to be calculated

THEN use of BWR type equation of state is discouraged.

[Fredenslund&Rasmussen&Mollerup 81]

(38) IF the use of theoretical mixing rules is recommended

THEN the use of the Virial equation of state is advised.

[Reid&Prausnitz&Sherwood 77]

(40) IF working conditions are over a large temperature range

THEN use of the Peng-Robinson equation of state is recommended

(because the binary adjustable parameters are less dependent
on the temperature).

[Juárez&Bañares 80]

(42) IF the binary interaction parameter is not available

THEN the use of the Lee-Kesler equation of state is advised since
the parameter can be correlated as a function of

$$\frac{T_{Ci}V_{Ci}}{Vq}$$

Vq

[Oellrich&Plocker&Prausnitz&Knapp 81]

(54) IF system is an asymmetric mixture

THEN use of the Plocker modification of the Lee-Kesler equation of state works slightly better than a VW type equation of state.

[Fredenslund&Rasmussen&Mollerup 81]

(55) IF the system consists of the following compounds:

- normal fluids (rare gases, N₂, O₂, CO, hydrocarbons)
- CO₂, H₂S, H₂
- slightly polar substances

THEN the equations of state are applicable.

[Oellrich&Plocker&Prausnitz&Knapp 81]

(56) IF working with mixtures with wide boiling temperature (mixtures that contain H₂, H₂S, CO₂, etc.)

THEN the Lee-Kesler equation of state has adequate mixing rules and the BWR equation of state yields large deviations.

[Oellrich&Plocker&Prausnitz&Knapp 81]

(57) IF the system contains hydrogen

THEN the BWR and Soave equations of state deviate from real behavior.

[Oellrich&Plocker&Prausnitz&Knapp 81]

(61) IF the operating conditions are near the critical region

THEN the use of the Regular Solution Theory is not advised since it does not have a dependence on pressure.

[Reid&Prausnitz&Sherwood 77]

(63) IF working conditions are in large temperature range

THEN prefer the UNIQUAC over the NRTL method (because the first has a better theoretical basis, therefore its parameters have a smaller dependence on temperature).

[Reid&Prausnitz&Sherwood 77]

(66) IF vapor-liquid-liquid (VLL) equilibrium calculations in a binary system are being performed

THEN either one of the following methods is recommended: NRTL or UNIQUAC.

[Reid&Prausnitz&Sherwood 77]

(68) IF phase splitting occurs and the working conditions are not confined inside the one phase region

THEN do not use the Wilson method (because it is unable to predict phase splitting).

[Reid&Prausnitz&Sherwood 77]

(69) IF azeotropes are suspected to be present in the system

THEN use the UNIFAC method (since it is capable of predicting them).

[Fredenslund&Rasmussen&Molterup 81]

(81) IF method with least number of parameters is desired

THEN the UNIQUAC method is advised over the NRTL method.

(NRTL UNIQUAC)

(binary 3 2)

(ternary 5 6)

[Reid&Prausnitz&Sherwood 77]

(89) IF fast calculations are desired

**THEN in general, the methods with least number of parameters
consume less time.**

[Fredenslund&Rasmussen&Mollerup 81]

(92) IF system is a strong nonideal mixture

(e.g. alcohols + hydrocarbons)

**AND operating conditions are in the dilute region with respect
to alcohol**

**THEN do not use Van Laar or Margules(2), use: Wilson, NRTL or
UNIQUAC.**

[Reid&Prausnitz&Sherwood 77]

(93) IF system is a symmetric mixture ($CJ = 0, 5 = 0$)

THEN Margules(1) works.

[Reid&Prausnitz&Sherwood 77]

(94) IF system is an asymmetric mixture

**THEN choose method with at least two adjustable parameters (3.
do not use Margules(1))**

[Reid&Prausnitz&Sherwood 77]

<98) IF electrolytes are present in the system

**THEN the following models are not applicable: Margules(all),
Van Laar, Wilson, NRTL, UNIQUAC, ASOG or UNIFAC.**

[Prausnitz 77]

(101) IF the system consists of a large range of components

THEN use the UNIQUAC method because it is applicable to the following components for example: hydrocarbons, ketones, esters, water, amines, alcohols, nitrites, etc.

[Reid&Prausnitz&Sherwood 77]

(102) IF working with polymers

THEN use the UNIQUAC method over the NRTL since it has a better theoretical basis, is applicable to small or large molecules (including polymers).

[Reid&Prausnitz&Sherwood 77]

7. BIBLIOGRAPHY

[Rychener81]

M.D. Rychener,

Knowledge-Based Expert Systems: A Brief Bibliography.

Department of Computer Science, Carnegie-Mellon University. Technical Report CMU-CS-81-127 (1981).

[Barr&Feigenbaum 81]

A. Barr, E. A. Feigenbaum, editors.

The Handbook of Artificial Intelligence.

Volumes I and II. Heuristech Press, Stanford, California (1981).

[Feigenbaum 77]

E.A. Feigenbaum,

The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering. Proc. Fifth International Joint Conference on Artificial Intelligence, 1014-1029 (1977).

[Reboh81]

R. Reboh,

Knowledge Engineering Techniques and Tools in the PROSPECTOR Environment. SRI International, Technical Note 243 (1981).

[Zudkevitch 80]

D. Zudkevitch,
Forensic Thermodynamics. Erroneous Decisions on Thermodynamic Data Can Cause Plant Failures. Paper presented in the AIChE symposium *Thermodynamics and Technology in Today's World* (1980).

[Nilsson 80]

N. J. Nilsson,
Principles of Artificial Intelligence.
Tioga Publishing Company, Palo Alto, California (1980).

[Davis&Buchanan&Shortliffe 77]

R. Davis, B. Buchanan & E. Shortliffe,
Production Rules as a Representation for a Knowledge Based Consultation Program. *Artificial Intelligence* 8,15-45 (1977).

[Duda&Hart&Nilsson 76]

R.O. Duda, P.E. Hart & N.J. Nilsson,
Subjective Bayesian Methods for Rule-Based Inference Systems.
Proc.AFIPS 1976 National Computer Conference 47,1075-1082 (1976).

[Bañares 82]

R. Bañares-Alcantara,
Development of a Consultant for Physical Property Predictions.
Master's Thesis, department of Chemical Engineering,
Carnegie-Mellon University (1982).

[Chao&Seader61]

K.C. Chao & J.D. Seader,
A General Correlation of Vapor-Liquid Equilibria in Hydrocarbon Mixtures.
AIChE Journal, 7 (4), pp 598-605, (1961).

[Reid&Prausnitz&Sherwood 77]

Reid,R.C.,Prausnitz,J.M. and Sherwood.T.K.
The Properties of Gases and Liquids. Third Edition.
McGraw-Hill, U.S.A., (1977).

[Barnés&Bazúa 77]

Barnés.F.J. and Bazúa.E.R.

Ecuaciones de Estado en la Ingeniería de Procesos.

In *Primer Coloquio sobre la Fisicoquímica de Fluidos*.

Publicación No. 77EF/224, Instituto Mexicano del Petróleo, (1977).

[Fredenslund&Rasmussen&Mollerup81]

Fredenslund,A.,Rasmussen,P. and Mollerup.J.

Thermophysical and transport properties for chemical process design.

In Mah.R.S.H. and Seider.W.D., editors,

Foundations of Computer-Aided Chemical Process Design. Volume II.,

Engineering Foundation, N.Y., pp 1-28., (1981).

[Oellrich&Plocker&Prausnitz&Knapp81]

Oellrich,L.,Plocker,U,Prausnitz,J.M. and Knapp.H.

Equation-of state methods for computing phase equilibria and enthalpies.

International Chemical Engineering 21(1), pp 1-16, (1981).

[Wilson 66]

Wilson.G.M.

Advances in Cryogenic Engineering 9, p 168, (1964).

and

Calculation of Enthalpy Data from a Modified Redlich-Kwong Equation of State.

Advances in Cryogenic Engineering 11, p 392, (1966).

[Barnés 73]

Barnés.F.J.

Ph.D. thesis, Department of Chemical Engineering, University of California, Berkeley, (1973).

[Soave72]

Soave.G.

Equilibrium constants from a modified Redlich-Kwong equation of state.

Chemical Engineering Science 27(6), pp 1197-1203, (1972).

[Peng&Robinson 76]

Peng,D. and Robinson,D.B.

A New Two-Constant Equation of State.

Industrial and Engineering Chemistry, Fundamentals 15(1),
pp 59-64, (1976).

[Yamada 73]

Yamada,T.

An Improved generalized Equation of State.

AIChE Journal 19(2), pp 286-291, (1973).

[Zudkevitch&Joffe 70]

Zudkevitch,D. and Joffe,J.

Correlation and Prediction of Vapor-Liquid Equilibria with the
Redlich-Kwong Equation of State.

AIChE Journal 16(1), pp 112-119, (1970).

[Starling 71]

Starling,K.

Thermo Data Refined for LPG. Part I: Equation of State and
Computer Prediction.

Hydrocarbon Processing 50(3), pp 101-104, (1971).

[Lee&Kesler 75]

B.I. Lee & M.G. Kesler,

A Generalized Thermodynamic Correlation Based on Three-Parameter
Corresponding States. *AIChE Journal* 21 (3), pp 510-527, (1975).

[Plocker&Knapp&Prausnitz 78]

Plocker,U.,Knapp,H. and Prausnitz,J.

Calculation of High-Pressure Vapor-Liquid Equilibria from a
Corresponding-States Correlation with Emphasis on Asymmetric
Mixtures.

*Industrial and Engineering Chemistry, Process Design and
Development* 17(3), pp 324-332, (1978).

[Chao&Robinson 79]

Chao.K.C. and Robinson R.L. (editors).
Equations of State in Engineering and Research
Advances in Chemistry Series 182, (1979).

[Leland 80]

Leland.T.
Equations of state for phase equilibrium computations.
In *Proc. of Second International Conference in Phase Equilibria
and Fluid Properties in the Chemical Industry*. DECHEMA.
EFCE Publication Series No. 11, pp 283-333, (1980).

[Martin 79]

Martin.J.J.
Cubic Equations of State-Which?
Industrial and Engineering Chemistry, Fundamentals 18(2),
pp 81-97, (1979).

[Bañares&Rychener&Westerberg 83]

Bañares R., Rychener M. and Westerberg A.
Internal Design Research Center Report.
In preparation.

[Motard83]

Computer horizons in process-systems engineering. Part II.
Chemical Engineering, 90 (5), pp 183-184, (1983).

[Lenat75]

Lenat.D.B.
BEINGS: Knowledge as interacting experts.
In *Proc. Fourth International Joint Conference on Artificial
Intelligence*, pp 126-133.
Tbilisi, (1975).

[Westerberg81]

Westerberg.A.W.

Design Research - Both Theory and Strategy.

From presentation prepared for ChE Division Award Lecture,
American Society for Engineering Education, Univ. So. Calif., L.A.
June (1981).

[Westerberg&Rychener 82]

Westerberg.A.W. and Rychener.M.D.

*An Intelligent Interface for a Complex Design Program, and
FEAT: an Intelligent User Interface to ASCEND-II.*

Unpublished, March (1982).

[Prausnitz 77]

Prausnitz,J.M.

State-of-the-Art Review of Phase Equilibria.

In Storvick,T.S. and Sandler.S.S., editors, *Phase Equilibria and
Fluid Properties in the Chemical Industry.*

ACS Symposium Series 60, pp 11-61, (1977).

[Juárez&Bañares 80]

Juárez,D. and Bañares R.

*Análisis de la importancia de la regla de mezclado para las
ecuaciones de estado.*

Tesis Profesional, Facultad de Química, Universidad Nacional
Autónoma de México, (1980).

[Tsonopoulos&Prausnitz 69]

Tsonopoulos,C. and Prausnitz,J.M.

Equations of State. A Review for Engineering Applications.

Cryogenics 9(5), pp 315-327, (1969).

[Stotler&Benedict 53]

Stotler.H.H. and Benedict.M.

Chemical Engineering Progress Symposium Series 49(6),
p 25, (1953).