6.1.2 HYDROID PROJECT

HYDROID - Studies in Distributed Processing and Problem Solving

Prof. Gio Wiederhold
Computer Science and Electrical Engineering
Stanford University

I. Summary of Research Program

A. Technical Goals

The objective of this research is the development of a methodology for the analysis and implementation of alternatives in distributed processing and problem solving. One of the primary reasons for interest in this area is its potential to break through the speed limitation barriers imposed by uniprocessing systems. If such a breakthrough can be achieved then the viability of the methods being developed by other projects using the SUMEX-AIM resource will be enhanced.

The rapid development of microprocessor and communications technology has given rise to a large number of proposed implementations of networks employing multiple processors. The computations to which these distributed systems are to be applied include heuristic decision-making problems, mathematical modelling, data reduction, and database search, as well as general purpose multi-access computing. There is however a lack of an adequate global understanding of the computational tradeoffs implied by network architectures.

In order to complement the experimental results of other investigators and broaden their applicability to the system-design decision-making process, we are developing a general framework for the study of processor interaction in distributed processing systems. The framework consists of rules to obtain parameters from programs which specify the computations, rules to parameterize descriptions of networks of processors, and procedures to calculate expected system performance from these parameter sets. The framework is to be sufficiently powerful so that, when it is validated, the methods will be able to assist in the a priori assessment of the potential performance of new system alternatives or of systems with improved system components.

One of the primary tools we are using to analyze the interaction between computations and distributed processor networks is simulation. The behaviour of processor network nodes, interprocessor control and task flow, and problem decomposition all require simulation at different levels of abstraction. Analytic queuing models may provide insight into relationships in networks, but are not adequate to provide quantitative results. Simulation is not seen as the end product of the study, but as a means to develop and assess the validity of our model of the interaction of computations and processor network architecture. Where possible, mathematical results will be used to assess the validity of model simulations.

A number of large computational applications are being analyzed in order to assess their potential for decomposition into modules for distributed processing. The current candidate applications are:

a) Programs which use heuristic methods in decision-making. Heuristic programs frequently employ recursive decomposition of problems into subsidiary problems which themselves may be suitable for distributed processing.

b) Programs which use multi-faceted databases to retrieve and abstract information. The process of intelligent data retrieval and analysis often depends on data or knowledge sources which are being maintained at geographically distributed processing sites.

c) Programs which acquire data from multiple, possibly dissimilar, sensors and attempt to reduce this data to simpler hypotheses.

d) Programs which solve large numerical problems, such as those found in image processing applications.

Parameters which describe the computations to be simulated include:

a) The computational kernel size: the cycle and memory demand of a computational unit between interprocessor reference requirements.

b) The computation definition message size: the amount of data required to transmit sufficient information to initiate a computational kernel.

c) The database size: the amount of data or program text required to sustain a computational kernel, and its availability and residence in the network.

The behaviour of the system can be varied through the adjustment of other parameters. These parameters may be set to reflect the architecture of specific hardware systems, or may be varied to obtain optimum performance. In addition to obvious parameters (as the number and power of the processors), we expect the following parameter types to be important in developing an understanding of the spectrum of distributed processor architectures:

a) Interconnection density. As the density decreases, the message delay and congestion increase. This parameter will provide a high level abstraction of multi-processor connectivity schemes. Geographical distribution will increase message delay and transmission cost.

b) Computational locality. A high degree of locality (of database or procedural information in the network) will enhance the probability that relevant knowledge exists in closely linked nodes, thus counteracting the effects of a low interconnection density.

c) Database viscosity. A database, including the programs required to carry out the computations at a node, may be more or less fixed to one specific node. This therefore encourages the use of certain nodes for specific functions. Many current processor networks are completely rigid in this sense, and for these networks optimal initial program and database

allocations may be determined. However, we hypothesize that a greater degree of dynamic resource allocation is desirable to cope with changing loads and in order to enhance reliability. For this reason this parameter needs to be included.

d) Redundancy. In order to assess the cost and benefits in terms of responsiveness and reliability, the redundancy of database and computations will also be made a parameter. In order to utilize the redundancy well, the computational resources (programs or data) which effect system performance most must be identifiable.

e) Error rate. In order to test the effectiveness of reliability strategies, node and communications channel failures will be simulated.

An important aspect of this model is that we intend to keep the abstractions at a sufficiently high level to allow analytic and intuitive verification of the model behaviour when applied to well understood computations. Computations have been mapped into specific parallel machines, but these results are not easily transferred to new architectures. The distributed processor systems now being built may have characteristics with unpredicted effects on system behaviour. We expect to be able to use the model to find potential bottlenecks, which then will define areas where extra design attention has a high payoff.

We do not intend to build hardware which is based literally on the abstract model. We hope to verify results obtained from the model using existing distributed processor systems and, assuming that our model (with appropriate parameters describing the load and architecture) matches the given system, be able to advise on system utilization or development aspects. A local resource of this type may be the Stanford I processor, now being built under ERDA sponsorship. In addition, if we determine that a certain, yet untried, architecture is promising, we would like to encourage and participate in its implementation.

B. Medical Relevance and Collaboration

Many applications at SUMEX consume large quantities of computational resources. The use of multiple distributed processors may provide a means to gain the required processing capabilities in an economic manner. In this sense the medical relevance of this study is indirect. We are attempting to develop tools which will be of use in medical computation problems.

Our studies in distributed data base applications have a more direct medical relevance. To this end, we are maintaining contact with Dr. Jim Fries, whose ARAMIS database network collects data for the analysis of disease progress and treatment efficacy in rheumatoid arthritis from a variety of institutions. Sharing of data to provide a broader base for analysis is also a feature of programs in cardiology and oncology in which physicians at Stanford participate. In each of these instances the distributed nature of the data resources leads to differences in the meaning of data items, so that simple aggregation of the data may not be valid. Distributed processing may provide a powerful alternative.

C. Progress Summary

The HYDROID project got underway in the fall of 1976. We have been involved since that time in developing a basic understanding of important problem areas in distributed processing and problem solving.

A weekly research seminar, begun in Dec. 1976 has brought together members of the faculty and students from a variety of disciplines, and has included several speakers from application areas where distributed processing may be beneficial.

We have developed a formalism in which to express the control of distributed problem solving in loosely-coupled processor networks. This CONTRACT NET protocol makes the cost of interprocessor interactions explicit. It is this cost which appears to generate one of the performance boundaries for distributed processor systems.

We have written a basic simulator with which to investigate the merits of the formalism together with problem solving methods applicable in the distributed processing environment. To this end the simulator is currently being tested with small search problems as a means of determining the necessary information that must be transferred from node to node in a distributed processor system for such problems together with the advantages to be accrued via a distributed approach. The simulator is being developed to cover a greater variety of computational interactions.

D. Publications

1) H. Garcia-Molina and Gio Wiederhold, "Application of the Contract Net Protocol to Distributed Data Bases", HPP-77-21, Heuristic Programming Project, Stanford University, April 1977.

2) R. G. Smith, "The Contract Net: A Formalism for the Control of Distributed Problem Solving", HPP-77-12, Heuristic Programming Project, Stanford University, February 1977 (also submitted to the Fifth International Joint Conference on Artificial Intelligence).

E. Funding

The HYDROID project is currently funded as part of ARPA Contract DAHC 15-73-C-0435. Other potential funding sources are currently being contacted for support of the specific areas of Hydroid application and interest.

II. Interactions with SUMEX-AIM

SUMEX-AIM currently provides all computing resources for the project. We thus enjoy a high degree of interaction with other projects involved in the problems which result from construction of large programs. Other points of contact are related to the use of the same programming languages as well as the abundance of AI expertise residing around the resource. This latter point is

especially important considering that one of our aims is discovery of suitable mappings of well understood AI methods onto highly parallel asynchronous processor networks.

SUMEX-AIM is also an excellent medium for informal transmission of reports, recent results and bulletins to users with related interests and problems. The powerful screen-oriented editors available greatly enhance our capabilities for writing both text and programs.

Finally, the development of simulation programs generally requires a highly interactive computing environment - the sort of environment we feel is provided by SUMEX-AIM.

6.1.3 MOLGEN PROJECT

MOLGEN - An Experiment Planning System for Molecular Genetics

Prof. J. Lederberg (Genetics, Stanford)
Prof. N. Martin (Computer Science, U. of New Mexico)
Prof. E. Feigenbaum (Computer Science, Stanford)

I. Summary of Research Program

A. Technical Goals

The goal of the MOLGEN project is to develop an experiment planning system for the domain of molecular genetics. In order to accomplish this, we hope to create and apply innovative methods of knowledge management and hierarchical planning.

Experiments in molecular genetics are concerned with the study and manipulation of DNA molecules. The MOLGEN knowledge base will include both declarative and procedural information about such structures and the laboratory tools and techniques which experimental geneticists use. Also represented will be much of the strategic information required to join individual experimental steps into a meaningful whole. We are using the uniform method of schemata for representation of all types of knowledge within MOLGEN. We believe this will facilitate knowledge acquisition and explanation and provide a consistent means of storing hierarchical and other relations among objects and rules in the system. We hope to make the underlying knowledge base flexible enough to allow for experimentation with a wide variety of specific planning strategies.

B. Medical relevance and collaboration

Molecular genetics has at least two major connections to medical research. Learning about the basic mechanisms which control the operation and transmission of genetic information is necessary to understand and treat the wide range of diseases (and health conditions like aging) which are genetically controlled. Also, recent developments in molecular genetics offer the promise of using genetic mechanisms to produce essentially limitless amounts of drugs and other biomedical substances. The MOLGEN project will develop a system designed to aid the molecular geneticist in planning experiments of these types.

The MOLGEN project is a joint effort of the Computer Science Departments of Stanford and the University of New Mexico and the Genetics Department of Stanford. Major participants are Professor Nancy Martin of the University of New Mexico, Professor Edward Feigenbaum, Peter Friedland, Jonathan King, and Mark Stefik of Stanford Computer Science, and Professor Joshua Lederberg and Jerry Feitelson of Stanford Genetics.

C. Accomplishments

MOLGEN is in the first year of formal funding as an independent entity. We have devoted this year to learning and analyzing the basic knowledge of experimental molecular genetics and to building part of the central structure of the knowledge base management system. A wide variety of experiments have been studied with the aim of extracting knowledge about the genetic objects and operators used as well as the higher-level know-ledge used to form the overall experimental plan. The object level knowledge is currently being organized into the schemata formalism for an initial attempt at a molecular genetics knowledge base.

A representation method for DNA structures and an interactive structure editing and entry system (EDNA) has been built and tested successfully with geneticist users. Work is proceeding on the schemata storage and access routines and on routines for acquiring and editing the rules which describe the procedural knowledge of the domain. We plan to have the basic MOLGEN system operational for the purpose of testing object and operator knowledge (the practical goal of experiment checking) by the end of July 1977.

D. Publications

1) N. Martin, P. Friedland, J. King, M. Stefik, "Knowledge Base Management for Experiment Planning in Molecular Genetics," submitted to Fifth International Joint Conference on Artificial Intelligence

2) M. Stefik and N. Martin, "A Review of Knowledge Based Systems as a Basis for a Genetics Experiment Designing System," Feb. 1977 Stanford CS Report STAN-CS-77-596, HPP77-5

3) N. Martin, P. Friedland, M. Stefik, "MOLGEN Knowledge Base I: Object System" To appear as HPP Working Paper

4) N. Martin, P. Friedland, M. Stefik, "MOLGEN Knowledge Base II: Rule System" To appear as HPP Working Paper

E. Funding

MOLGEN research is supported by NSF grants MCS76-11649 and MCS76-11935 for the two year period from June 1976 - June 1978.

II. Interactions with SUMEX-AIM

All system development has taken place on the SUMEX-AIM facility. We have used the system not only for programming, but also as a major aid in writing and transmitting among ourselves the wide variety of formal and informal reports which are necessary in the MOLGEN design phase. We believe the availability of good interactive text editing facilities like TV-Edit increases our productivity significantly.

Active collaboration with remote users at the University of New Mexico will begin in September 1977 (Prof. Nancy Martin has been visiting at Stanford this year). We expect this collaboration to occur over the ARPA network. We hope also to maintain a collaboration with Dusko Ehrlich, formerly a Stanford geneticist and now doing research at The Institut de Biologie Moleculaire Faculte de Science in Paris over a TYMNET link to Sumex.

We have benefited enormously from the collected expertise in both knowledge-based systems and general programming and design problems available from other SUMEX-AIM projects. We have especially strong ties to the knowledge management expertise of the MYCIN project, but we also share common objectives with parts of the DENDRAL, SECS, and protein crystallography projects. We have also benefited from the intense interaction with many other projects at the AIM conferences.

Finally, we have provided small amounts of SUMEX resources to geneticist users as part of a quid pro quo relationship for helping us understand that subset of genetic knowledge necessary for our initial knowledge base. The most outstanding example of this sort of collaboration occurred with Prof. Larry Kedes' group at the VA hospital in Palo Alto who are using SUMEX to determine the feasibility of automated assistance in analyzing complex DNA base sequences.

6.1.4 MYCIN PROJECT

MYCIN - Computer-based Consultation in Clinical Therapeutics

S. N. Cohen, M.D. (Pharmacology) and
B. G. Buchanan, Ph.D. (Computer Science)
Stanford University

I) Summary of research

Technical goals

The Mycin project is aimed at the development of a computer program capable of functioning as an expert consultant on a range of medical decision making problems. In particular, we have been working on the construction of a system that provides consultative advice on the diagnosis and therapy selection for a number of infectious diseases. Current areas of competence of the system include bacteremia and meningitis, and work is currently underway to extend this to urinary tract infections, pulmonary infections, and prophylactic use of antibiotics.

Our work has been guided by three fundamental objectives:

(1) A major goal of the MYCIN system has been to provide a computer-based therapeutic tool designed to be clinically useful, one that would be used eventually in the clinical setting. This goal requires development of a system that has a medically sound knowledge base, and that displays a high level of clinical competence in its field. The program must first convince clinicians of the quality of the information it is providing before they will be willing to use it.

(2) Since many clinicians are not likely to accept the advice provided by a computer-based system unless they can understand why the recommended therapy has been selected, the system has to do more than just give advice dogmatically. It should have the ability to explain the reasoning behind its decisions, and should be able to do so in terms that suggest to the physician that the program approaches the problem in much the same way that he does. This permits the user to validate the program's reasoning, and modify (or reject) the advice if he believes that some step in the decision process is not justified. It also gives the program an inherent instructional capability that allows the physician to learn from each consultation session.

(3) A third major goal is to provide the program with capabilities that enable augmentation or modification of the knowledge base by clinical experts in infectious disease therapy, in order to improve the validity of future consultations. The system therefore requires some capability for acquiring knowledge by interacting with experts in the field, and for incorporating this knowledge into its knowledge base.

Three separate parts of the MYCIN system accomplish these goals. The consultation system uses the knowledge base, along with patient-related data entered by the physician to generate therapeutic advice. The explanation system has the ability to explain the reasoning used during the consultation, and to document the motivation for questions asked or the rationale for conclusions reached. Finally, the knowledge acquisition system enables experts in antimicrobial therapy to update MYCIN's knowledge base, without requiring that they know how to program a computer.

We have also sought to use Mycin as a framework for understanding the process of medical decision making and the nature of clinical judgment. Physicians are constantly faced with the necessity of making decisions based on information that is both incomplete (missing historical data or test results not yet available) and inexact (results are rarely definitive). In addition, those decisions are often based on rules that are only approximate (e.g., "a gram-negative aerobic rod in the blood is probably a bacteriodes"). But decisions are made despite these problems, and the results often proven later to be valid. We have attempted to understand how this is done by developing in our system a parallel set of capabilities. We have relied on the "production rule" encoding of information, in which individual decision rules are specified in an "if/then" format. For example, the rule indicated just above is encoded in the system as:

If 1) the gram stain of the organism is gram negative, and
2) the morphology of the organism is rod, and
3) the aerobicity of the organism is anaerobic,
Then there is suggestive evidence (.6) that the identity of the organism is Bacteroides.

This encoding of knowledge offers a number of advantages over some of the more traditional approaches to diagnosis like decision trees, Bayesian analysis, and utility theory. Unlike decision trees, it can deal with both inexact and incomplete information. Unlike the Bayesian and utility theory approaches, it does not need extensive amounts of conditional probability data. A collection of independent rules is also far easier to augment than a complex decision tree; the rules thus provide a much more flexible body of knowledge to which new information is more easily added. The rules also make possible an explanatory capability: the system can justify any of its actions or decisions by displaying the relevant rules it invoked in reaching that decision. This provides an explanation that is far more comprehensible than any we might be able to provide by recapping the actions of a program based solely on statistical considerations.

A more specific goal of our research involves understanding the process of infectious disease diagnosis and therapy selection. This process is not as yet well understood, and we believe that by dissecting it down to individual decision rules, we can gain insight into how it works. In addition, the resulting set of rules may prove to be a useful compendium of knowledge about the task.

Since we believe this set of rules will also be quite large, we are studying the problems of accumulating, managing, and using large stores of such task-specific knowledge. We are working on a range of techniques to provide capabilities like insuring the consistency of the set of rules and making it easy to modify existing rules or add new ones.

Finally, since computer consultants are designed for use by people who might not otherwise make use of computers, we have devoted a great deal of attention to the issue of human engineering, and the "habitability" of the system. This ranges from such minor items as the automatic correction of misspelled answers, to the range of sophisticated explanation capabilities available.

Medical relevance and collaboration

A number of recent studies indicate a major need to improve the quality of antimicrobial therapy. Almost one-half of the total cost of drugs spent in treating hospitalized patients is spend on antibiotics [1,2], and if results of a number of recent studies are to be believed, a significant part of this therapy is associated with serious misuse [2,3,4,5], Some of the inappropriate therapy involves incorrect selection of a therapeutic regimen [4], while another serious problem is the incorrect decision to administer any antibiotic [2,4,5]. One recent study concluded that one out of every four people in the United States was given penicillin during a recent year, and nearly 90% of these prescriptions were unnecessary [6]. Other studies have shown that physicians will often reach therapeutic decisions that differ significantly from the decisions that would have been suggested by experts in infectious disease therapy practicing at the same institution.

Nonexperts sometimes choose a drug regimen designed to cover for all possibilities, prescribing either several drugs or one of the so-called "broad spectrum" antibiotics, even though appropriate use of clinical data might have led to more rational and less toxic therapy. Within a hospital environment in which professional resources are often overburdened, and in environments where expert sources are not readily available, a computer-based consultant will be highly useful. Such a system will also have broad fringe benefits in its educational impact on staff physicians and in providing a framework for quality control and peer-review evaluations.

Antimicrobial therapy appears to be an especially suitable area for the initial development of a computer-based system to assist physicians with decisions in clinical therapeutics. The components of the decision making process in antimicrobial therapy are more readily definable than in many other areas of medicine, and the consequences of the physician's decision can usually be assessed in terms of the direct therapeutic action. Nevertheless, the general approach used here is applicable to other areas of clinical decision making. The basis of rational antimicrobial therapy decisions is identification of the microorganisms causing the infectious disease. Accurate identification is important because of the specificity of antibiotic action: drugs that are highly effective against certain organisms are often useless against others. The patient's clinical status and history (including information such as prior infections and treatments) provide data that may be valuable to the physician in identifying the disease-causing organisms. However, bacteriological cultures that use specimens taken from the site of the patient's infection usually provide the most definitive identifying information.

Initial culture reports from a microbiological laboratory may become available within 12 hours from the time a clinical specimen is obtained from the

patient. While the information in these early reports often serves to classify the organism in general terms, it does not often permit precise identification. It may be clinically unwise to postpone therapy until such identification can be made with certainty, a process that usually requires 24 to 48 hours, or longer. Thus it is commonly necessary for the physician to estimate the range of possible infecting organisms, and to start appropriate therapy even before the laboratory is able to identify the offending organism and its antibiotic sensitivities. In this setting MYCIN plays two roles: (a) providing consultative advice that will assist the physician in making the best therapeutic decision that can be made on the basis of available information, and, (b) by its questioning of the physician, pinpointing the items of clinical data that are necessary to increase the validity of the clinical decision.

Our project is an interdisciplinary effort involving the joint effort of computer scientists from the Stanford Computer Science Department, and clinicians from both the Department of Clinical Pharmacology at Stanford and the Department of Infectious Disease at the University of Arizona. The task of the clinicians has been to specify the decision rules necessary for diagnosis and therapy selection, while the computer scientists have been devising ways to represent and use this information in the computer. The system is then tested by the clinicians using real cases obtained from journals and medical records.

A complete listing of the staff is given below.

Stanley N. Cohen, MD, Clinical Pharmacology
Bruce G. Buchanan, PhD, Computer Science
Stanton Axline, MD, Infectious Disease (now at University of Arizona)
Randall Davis, PhD, Computer Science
Frank Rhame, MD, (to 9/75), Infectious Disease
Edward Shortliffe, MD PhD (to 6/76, returning 6/77), Infectious Disease
Victor Yu, MD, Infectious Disease
Rudolpho Chavez-Pardo, MD, (to 9/75), Clinical Pharmacology
A. Carlisle Scott, MS, Computer Science
Sharon Wraith, BS, Clinical Pharmacology
Jan Aikins, BS, Computer Science
Robert Blum, MD, presently in Computer Science
William Clancey, AB, Computer Science
Larry Fagan, AB, Computer Science
William van Melle, AB, Computer Science

Progress Report

Period covered: June 1, 1974 through September 30,1976

Summary

Over the past three years we have designed, built and partially evaluated a computer program capable of diagnosis and therapy selection for certain varieties of infectious diseases. The program is intended to function as a consultant, and "interviews" a doctor about his patient, requesting information on clinical findings and results of laboratory tests. It relies on a store of judgmental knowledge (obtained from experts in infectious disease) to determine the

conclusions which can be drawn from the answers it receives. This judgmental knowledge is in the form of some 400 decision rules dealing with the wide range of topics that must be considered in determining the likely identity of causative organisms and selecting appropriate antimicrobials.

MYCIN is composed of the three systems described earlier (the consultation, explanation, and knowledge acquisition systems), all of which reference the knowledge base of decision rules. The program is currently capable of dealing with bacteremia and meningitis infections. It can diagnose the likely presence of more than 35 different organisms and can recommend therapy for 100 organisms, selecting drugs from a "pharmacopoeia" of 30 antimicrobials. The system can tailor its therapy recommendations to a specific organism and infection, can adjust dosage levels and durations in response to impaired renal status, and can combine drugs to create combination therapies, giving it a wide range of clinical applicability.

Detailed Report

Our work in the past several years has been organized around five main areas of investigation. We have

a) increased the system's competence in existing areas of clinical expertise while expanding its scope

b) developed a number of user-oriented features to increase the program's attractiveness to clinicians

c) developed a range of knowledge acquisition capabilities to speed the process of expanding the system's clinical competence

d) solved a number of technical problems to insure that the program does not outgrow the computer resources available to it

e) evaluated the system's level of expertise.

Clinical Capabilities

Since the primary qualification for any clinical consultant is competence in the domain, we have devoted significant effort to expanding MYCIN's knowledge base and widening its scope of competence.

For instance, the system was directed initially at patients with positive blood cultures, the basic methodology was generalized to support a much broader approach to the problem. MYCIN has now gained the ability to deal with infections from which the causative pathogen hasn't been isolated (e.g., pneumonia), or which haven't even been cultured (e.g., brain abscess). With this broadening of scope, it has also become necessary to be able to evaluate the meaningfulness of isolates for cultures taken from sites other than blood. For urine and sputum isolates, for example, the system gained the ability to base its evaluation of sterility of an isolate on both the method of collection and the user's estimation of conscientiousness of collection.

An extensive review of the program's approach to drug selection has led to a major revision in the basis for therapy selection during the course of program development. The program was given the ability to consider both the infectious disease diagnosis and the significance of the organism as further determinants of therapy, in addition to organism identity. These three together have become the primary factors in drug selection, with drug toxicity and ecological factors as secondary considerations. The result is a more appropriate, more sharply focussed drug selection that also includes dose, route, and duration.

While the initial development of the knowledge base focussed on rules concerned with the diagnosis and therapy for blood infections (bacteremia), the complexity of infectious disease therapy and the frequent occurrence of multiple infections in a single patient requires a broader knowledge if the system is to be clinically useful.

In response we have extended MYCIN's knowledge base, while at the same time improving the degree of sophistication with which the system deals with bacteremia. The second major area has been the diagnosis and treatment of meningitis, and more than 100 rules were added to provide the ability to deal with it. In the processs the program was also extended beyond bacteria, as it gained the ability to consider and treat both fungi and viruses.

This area has proved to be an especially useful domain because it has presented several new challenges. In particular, meningitis requires the ability to deal with a disease that is often diagnosed on clinical grounds alone, before any specific microbiological evidence is available (by comparison, the diagnosis of bacteremia on clinical grounds alone is far less certain, and usually requires establishment of the fact that bacterial growth has occurred in blood cultures.) For this reason, extension of the project into the meningitis area has made it necessary for MYCIN to consider a larger range of clinical factors, and has resulted in a system which has a broader picture of the whole patient.

Other contributions to the system's competence have come from expansion of the knowledge base to include information about normal bacteriological flora for a wide range of culture sites. This enables the program to distinguish between normal and pathological flora, and it can as a result decide more precisely on whether to treat.

User Oriented Features

Clinicians traditionally shun computer programs, and we believe this is in large measure due to insufficient attention paid to user oriented features. As a result, we have devoted significant effort to insuring that MYCIN is responsive to its users in a number of unique ways. The development of the explanation and question answering capabilities have been a essential for this work, and both have grown extensively in power.

The system's ability to explain the motivations for its questions, for instance, underwent a major design revision. It is now based on a more powerful approach that relies on the program's knowledge of its own control structure and ability to examine its own rules. The user can now fully explore the system's current line of reasoning, rather than just a single level, as initially implemented.

The language understanding capabilities of the question answering system have also been extensively revised. They now allow a broader range of questions to be asked and offer more precise answers. The use of this feature was also simplified so that the user no longer needs to classify his questions.

A comprehensive review of the kinds of questions asked by users of the system has led to a number of important features. MYCIN can now answer a much wider range of questions, and can, in particular, explain why it did not take a specific action, as well as why positive conclusions were reached. It is our feeling that capabilities such as these are of great importance in enabling the project's staff and clinical experts to understand the program's rationale for its actions in instances where its recommendations do not appear to be the most appropriate and most correct. Thus, the line of reasoning of the program can be evaluated, and requirements for new or modified rules can be uncovered. These kinds of capabilities are also important in optimizing user acceptance of the system.

A substantial addition to the question-answering facility enables the system to explain the process of therapy selection. In comparison to the diagnostic process, therapy selection is complicated somewhat by the need to consider a range of different factors simultaneously, such as the total number of drugs recommended, the degree of sickness of the patient, possible interactions between drugs, toxicity and other side effects, etc. Despite this complexity, explanations of therapy selection are phrased at a conceptual level that makes them comprehensible to the physician. As before, this makes it possible for the physician to verify the validity of the system's decisions, and makes it clear to him that the system reaches its results in much the same way that he does.

The explanation consists of a step-by-step review of the reasoning which led to recommending a particular drug for a specific organism. It considers such issues as why a drug was first considered for an organism, why a drug may have been chosen as the best therapy for that organism, how the total number of drugs was reduced by considering common drug classes among the candidates, and consideration of possible contraindications based on the patient's allergies, age, and other factors. By characterizing each drug according to this scheme, the program can explain why a drug was or wasn't prescribed, as well as why one drug is to be preferred over another. This offers an important explanatory capability that will make the system more attractive and acceptable to clinicians.

Several capabilities have been added to make the program easy to use. The system is now more tolerant of erroneous or inappropriate responses, and is able to provide a reworded question, along with a list of acceptable answers. In addition, it has the ability to recognize responses which are not sufficiently precise, and can rephrase its questions accordingly.

We have recently added to the system the ability to modify drug dosage in cases of renal failure. Where, previously, the system only issued a warning to modify doses, it is now able to use either creatinine clearance or serum creatinine levels to compute the level of renal function. The program then uses drug-specific information (e.g., half-life, percent loss of the drug via renal excretion, etc.) to adjust the regimen. It can either (a) adjust dose levels downward and leave dosing interval unchanged, or (b) increase dosing interval and

leave levels unchanged, or (c) allow the physician to select a dose interval, for which it chooses an appropriate dose level.

Since the problem of determining renal status and the proper adjustment of drug dose is important in the use of aminoglycoside antibiotics, cephalosporins, and other antimicrobial agents, the customization of drug dosage recommendations will be an important addition to the power of the system.

We have found, in addition, that there is a substantial amount of information that is routinely collected in every consultation, like the date and site of each of the cultures, gramstain and morphology results for each of the organisms that grew out, etc. Currently, the program exhaustively analyzes each culture and all of its organisms in turn. Some users of the program appear to be impatient with this method, and would much prefer to enter all the relevant data on all the cultures and organisms at once. This is faster and easier, since the information can be gathered in a single review of the chart, instead of having to review it several times as each culture is processed. In response to this, we have reorganized the consultation slightly, so that it is possible to enter all of this data at once, at the beginning. This offers two other advantages in addition to improving the program's acceptability to its users. First, it provides a basis for our future efforts to write rules which deal with interactions between infections (see below, "Specific Aims"), and second, it suggests a mechanism for eventually merging our work with the product of existing efforts to organize and automate the recording and handling of medical record data. This latter development may in time make it possible for MYCIN to obtain a large part of the information it requires directly from such automated records, sharply reducing the number of questions it has to ask, and speeding up the consultation considerably.

Finally, several new capabilities make the system convenient to use, in anticipation of its evaluation in the clinical setting. Among these are the option of the user to type a comment about system performance at any time during the consultation. His comment is recorded in a special file which is reviewed periodically by our medical staff, and provides an on-going opportunity for users to offer feedback aimed at improving the usefulness of the system. The user can also indicate his belief that the system has "broken down" in some way and he is invited to describe the problem. His description is saved along with information about the current state of the program, so that our systems programmers can deal with the problem later.

Knowledge Acquisition

A preliminary knowledge acquisition program was completed in the middle of 1974, and demonstrated the feasibility of having a physician teach the system new rules using a rather stylized subset of English. Building on the experience gained here, work began on a revised program designed to allow the user to examine and modify the program's knowledge and behavior as a single, unified action. This program was designed to make the explanation and knowledge acquisition capabilities available together, to make use of the fact that the nature of the explanations requested can give a clear hint about the content of a new rule. The program was also designed to advise the user about the effect of his rule on the original deficiency, indicating, for instance, whether or not it corrects the problem he noticed.

Work on a preliminary version of this new program was completed in 1976, making available a broad range of useful features enabling our clinical experts to add rules to the system without requiring that they have a knowledge of programming. If the expert finds that MYCIN's handling of a particular problem is at variance with his own expert knowledge, he can use the explanation capabilities to discuss the line of reasoning in use at that time, can add or modify rules in the knowledge base, and can determine the effects of the changes on MYCIN's subsequent performance. (Quality control is maintained on the overall system by regular meetings of our clinical and pharmacological experts who determine the "official" MYCIN knowledge base.)

Technical Issues

As MYCIN's clinical capabilities have expanded, efficiency has improved as a result of a number of modifications to the system's technical capabilities. Early in our work, for instance, a comprehensive review and modification of the control structure was undertaken to improve efficiency and generality. The resulting program was both more direct, and faster.

More recently, modifications have been made so that the the large English dictionary can be kept on the disk and accessed only as needed, rather than keeping it in core, which slows down the system's response speed. The self documenting features of the program have also been improved to make them faster, and the system's interaction with the terminal has been made more uniform, to prepare for the time when different users of the system may have various different kinds of terminals.

Evaluation Activities

Since clinicians are likely to require documentation of MYCIN's competence and utility before seeking its advice, considerable time has been spent on evaluating the system and on implementing a range of program features to support these efforts.

In the past two years we have obtained many useful suggestions from clinicians when the system was presented to several different conferences. In February 1975 it was presented to the Western Society for Clinical Research, in September 1975 to the International Symposium on Clinical Pharmacy and Clinical Pharmacology, and more recently (June 1976), it was presented to the Drug Information Association.

A large scale formal study and evaluation of MYCIN's performance was begun in January 1976. The same set of clinical data was provided to both MYCIN and a set of experts in infectious disease therapy. [Five of the experts were nationally recognized authorities in the field, the other five were clinical fellows in the Infectious Disease Division at Stanford. A complete list of names, titles and affiliations is found in Appendix B.] The judgments of the program and the experts were compared, and the experts were asked to evaluate MYCIN's performance.

To do this, we first designed a form to allow us to separate the variables requiring analysis. The parameters evaluated include

A. the "quality" of the interaction - were any questions irrelevant or missing
B. the program's ability to determine organism identity
C. the program's ability to determine organism significance
D. the program's ability to select proper therapy
E. overall performance evaluation
F. potential impact as a clinical tool or teaching facility

The evaluation form was designed to be informative yet simple to complete. It was tested in a pre-evaluation trial run, then used for the formal study.

Consecutive patients with positive blood samples were evaluated for inclusion in the study by project personnel, until we obtained at least 10 patients for which MYCIN recommended therapy, and 15 patients overall (patients were rejected if they were outpatients when the sample was drawn, if they had a previous blood culture in the preceding seven days, or if they had a diagnosis of meningitis or infectious endocarditis.) For each of the patients accepted, a one to two page clinical summary was prepared and combined with a summary of the laboratory test data as of the time when the first blood culture was obtained. This information was then used to obtain a therapeutic evaluation from MYCIN.

Each of the participating experts received a set of fifteen evaluation forms (one for each patient). Each form contained: (a) the clinical summary and lab data; (b) space for the expert to record his conclusions about the nature of the infection, likely causative organisms, and appropriate therapy; and (c) a transcript of the MYCIN consultation along with space for the expert to record his opinion of various aspects of MYCIN's performance. By presenting the information in this order, we obtained a therapeutic regimen from the expert based on the same information supplied to MYCIN. This allowed us to compare the expert's answers to MYCIN's, and also gave us the expert's opinion of the system's performance.

In the past few months a sufficient number of the forms have been returned that we were able to do a preliminary analysis. The figures below are based on the nine (out of ten) which have been returned.

Since it is difficult to select a single number which summarizes performance, we have in general measured each of the parameters listed above in three ways: (i) the percent of instances in which the program was judged exactly correct, (ii) the percent of instances in which the program's performance was judged exactly correct or an acceptable alternative, and (iii) the percent of cases in which a majority of the experts judged its performance exactly correct or an acceptable alternative. By using all three measures, we obtain a range of figures which give a good picture of the program's performance.

All of these attempts to evaluate performance are complicated by the fact that (as expected) the experts' own choices about each patient were not unanimous. Thus, we cannot ask whether MYCIN's answers were "correct" in any absolute sense, since there was no agreement on what constitutes "correct". Instead, we ask how often each individual expert rated the program's responses as

correct. But given the variation among experts themselves, the program can never be expected to reach 100%, and depending on the extent of the intra-group variation, the absolute limit may in fact be much lower. Thus the ideal question to ask is "Do experts rate MYCIN's performance correct at least as often as they rate each other's performance correct?" This would give a good indication of how close the system's performance was to that of the group of experts as a whole.

We have been able to do this in a few isolated cases, but in general it requires more information than we were able to collect. This is discussed in more detail below, but in general terms the problem is that we were able to ask each expert for his choices for each patient, and ask him to rate MYCIN's choices. But, without a second round of questionnaires, which would ask each expert to rate the acceptability of the other 9 experts' responses, we lack direct information about intra-expert variability. The figures below should be reviewed with this caveat in mind.

A. "Quality" of the interaction

To measure the first item, the experts were instructed to mark any questions in the consultation which they felt were irrelevant, and to note any questions which they felt were omitted by the system. Overall MYCIN did quite well, as there were no consultations in which a majority of the experts felt that any particular question was irrelevant or omitted. On the average, there were 0.53 questions judged irrelevant and 0.55 indicated as omitted.

Table I summarizes the next four measurements.

	MYCIN 1st choice identical to an expert's 1st choice	MYCIN 1st choice identical to or an acceptable alternative to an expert's 1st choice	MYCIN 1st choice identical to or an acceptable alternative judged by a majority of experts
ORGANISM IDENTITY	56.3% N= 414	75.6% N= 414	81.8% N= 11
ORGANISM SIGNIFICANCE	91.7% N= 36	NA	100% N= 4
THERAPY SELECTION	12% N= 99	75% N= 99	91% N= 11
OVERALL PERFORMANCE	17.0% N= 135	59.3% N= 135	60.0% N= 15

Table I
Summary of nine experts' responses to MYCIN's performance on 15 cases

B. Organism Identity

For organism identity, the experts were asked to rate each of MYCIN's selections as exactly correct (they agreed that the organism was likely to be present), an acceptable alternative (they had not chosen that organism, but agreed it might be present), or an unacceptable choice (they disagreed with its selection). Since 11 of the cases were not contaminants, and there was a total of 46 organisms chosen by the system, with 9 experts rating each of those choices we have an N of 414 for the first two columns and 11 for the third.

In 56% of the instances the system's choices were identical to the experts', 75% of them were either identical or acceptable alternatives, and in 82% of the cases, its results were acceptable to a majority of the experts.

In addition, the experts were asked to indicate which organisms they felt MYCIN had overlooked in its diagnosis. For the 11 non-contaminant cases, the experts indicated an average of only 0.35 organism identities that were overlooked by the system. In no case did a majority of experts feel that any particular organism had been overlooked, suggesting that even the 0.35 figure is a result of intra-expert variation.

C. Organism Significance

The first question on the evaluation form gave the expert a chance to indicate that he felt the patient did not need to be treated. The first column of the second row indicates the number of times the expert indicated no treatment was necessary for a case in which MYCIN also judged the organism to be a contaminant. (There is no number in the second column since we did not ask about a "close call" on whether or not to treat. In addition, the measurement is based only on the contaminant cases, since in many of the cases where both MYCIN and the expert determined that treatment was necessary, they based that decision on different organisms. We felt that it would be misrepresentative to call these situations "agreements".)

As the figures show, in only three out of 36 instances was there any disagreement with the system's decision on whether or not to treat.

D. Therapy Selection

The expert was asked to select therapy for the organisms which he felt were likely to be present before looking at MYCIN's therapy recommendation. He was then asked to judge MYCIN's choice of therapy for that patient. Since MYCIN was selecting therapy for the organisms which it felt were present (which may have differed from those chosen by the expert), this provides a fundamental comparison of performance - it compares therapy selection performance of the two when they are faced with the same clinical situation.

This comparison is a difficult one to make, since it is complicated by the difficulty noted above, of variability in the experts' performance and the need to judge MYCIN with respect to that variability. Looking only at exact agreements (i.e., two identical therapies) produces the figure in the first column, which indicates that 12% of the time MYCIN's recommendation was identical to that of an expert. Comparing each expert's therapy choice with the other 8 indicates that 35% of the time (N= 396) any pair of experts chose identical regimens. The experts were also asked to judge whether MYCIN's therapy was an acceptable alternative (if it was not identical to their own), producing the figure in the second column. This indicates that it was either identical, or they felt it was an acceptable alternative 75% of the time. (Unfortunately, we have no reliable way of judging the intra-expert variability here, without a second round of questionnaires which asked each expert to rate the acceptability of the other experts' choices.) [As an alternative, we have attempted to develop a measure of how "far apart" two non-identical regimens are. But the problem is difficult: for example, for gram negative rods with salmonella most likely, is gentamycin and chloramphenicol "very different" from gentamycin and ampicillin? We have been working on a "drug metric" to solve this problem, attempting to base the "difference" between two drugs on factors like organism susceptibility, toxicity, and drug efficacy, but this work is still in progress.]

The figure in the third column gives a crude overall measure of therapy selection performance, and indicates that in 91% (10 out of 11 cases), a majority of the experts rated MYCIN's regimen as either identical to their own or an acceptable alternative.

[The evaluation form also asked each expert to choose a regimen for the organisms which MYCIN had selected. The intent here was to compare the system's performance against the expert when both were faced with the same set of organisms (rather than compared with the same clinical situation, as above). Unfortunately, inconsistent answers on the part of the experts indicated that they were not answering the question according to the instructions. It appeared that they were not able to suspend their own judgments about organism identity sufficiently to select a regimen based on MYCIN's organisms alone. For this reason, we believe the data to be unreliable, and have not included it here.]

E. Overall Performance

At the end of each evaluation form, the expert was asked to rate the system's overall performance as either excellent, good, fair, or poor. The first two columns of the last row indicate that 17% of these evaluations were "excellent", and almost 60% were either "excellent" or "good" (only 13% were "poor"). In 60% of the cases (9 out of 15), a majority of the experts felt that MYCIN's overall performance was either "excellent" or "good".

F. Present Utility and Future Potential

Finally, after completing the entire set of 15 patients, each expert was asked to rate MYCIN's present utility and future potential as a clinical tool and as an educational tool, rating it as having "considerable", "some", or "no" potential. The table below summarizes their response.

Evaluation of Present Utility

	"considerable"	"some"	"none"
clinical tool	11%	67%	22%
educational tool	11%	89%	0%

Evaluation of Future Potential

	"considerable"	"some"	"none"
clinical tool	11%	89%	0%
educational tool	67%	33%	0%

Table II
Opinions of 9 experts on MYCIN's present utility and future potential

To aid these evaluation efforts, we have also implemented a number of useful features in the system. For instance, MYCIN now keeps continuing

statistics of the use of rules in its knowledge base. This will help us to monitor its long term performance, to study the interrelationship between rules, and perhaps detect automatically any inconsistencies or gaps in the knowledge base.

We have also designed and implemented a mechanism for "on-line" evaluation. At the end of each consultation, the system asks a few questions about the quality of its performance from the clinicians who are using it. This interchange will be brief to avoid being a burden to the user, but it is expected to represent an important addition to the other evaluation efforts.

It will, for instance, make possible a new form of evaluation of the system. Rather than using a series of "prepackaged" cases as was done in our initial evaluation, the next stage will be carried out using information entered at a terminal by the evaluator. The participating panel of experts will be selecting patients in areas covered by the MYCIN knowledge base, and will engage in a dialogue with the system about those patients. Following completion of the session, the on-line evaluation feature will ask questions about system performance, and the responses will be tabulated and evaluated on-line by appropriate biostatistical programs. Specific recommendations which may point out problem areas in the consultation will be reviewed by our staff. By this process we expect to be able to maintain a continuing evaluation of MYCIN's capabilities in various areas, and pinpoint specific areas where performance is suboptimal.

MYCIN Project Publications

THESES

Davis R,
Applications of meta level knowledge to the construction, maintenance, and use of large knowledge bases, Thesis: PhD in Computer Science, AI Memo 283, 304 pp, Stanford University, July 1976.

Shortliffe E H,
MYCIN: A rule-based computer program for advising physicians regarding Antimicrobial therapy selection, Thesis: Ph.D. in Medical Information Sciences, Stanford University, Stanford CA, 409 pages, October 1974. Also, Computer-Based Medical Consultations: MYCIN, American Elsevier, New York, 1976.

PAPERS

Buchanan B G, Davis R, Yu V, Cohen S N,
Rule-based medical decision making by computer, Proc. MEDINFO 1977, to appear.

Clancey W.
Chronicler: an explanation system based on set-predicate representation of computational processes, submitted to 5th IJCAI.

Aikins J S.
Use of models in a rule-based consultation system, short paper submitted to 5th IJCAI.

Davis R.
Interactive transfer of expertise: acquisition of new inference rules, submitted to 5th IJCAI.

Davis R.
Knowledge acquisition in rule-based systems: knowledge about representations as a basis for system construction and maintenance, to appear in Pattern Directed Inference Systems, Waterman and Hayes-Roth (eds.), Academic Press, in press. Also to be presented at Pattern Directed Inference Systems Workshop, Honolulu, May 1977.

Davis R, Buchanan B G.
Meta-level knowledge: overview and applications, submitted to 5th IJCAI, Cambridge, MA, August 1977.

Davis R.
A decision support system for medical diagnosis and therapy selection, Data Base (SIGBDP newsletter), 8 (Winter 1977) pp 58-72.

Wraith S, Aikins J, Buchanan B G, Clancy W, Davis R, Fagan L, Scott A C, van Melle W, Yu V, Axline S, Cohen S,
Computerized consultation system for selection of antimicrobial therapy, American Journal of Hospital Pharmacy, 33 (December 1976) pp 1304-1308

Scott A C, Clancey W, Davis R, Shortliffe E H,
Explanation capabilities of knowledge based production systems, American Journal of Computational Linguistics, Microfiche 62, 1977. Also, HPP Memo 77-1, Stanford Computer Science Department, February 1977.

Shortliffe E H, Davis R,
Some considerations for the implementation of knowledge-based expert systems, SIGART Newsletter, 55:9-12, December 1975.

Davis R, Buchanan B, Shortliffe E H,
Production rules as a representation for a knowledge-based consultation system, Artificial Intelligence, 8 (Spring 1977) pp 15-45. (Also, AI Memo 266, Stanford University, October 1975).

Davis R, King J J,
An overview of production systems, in Elcock and Michie (Eds.), Machine Intelligence 8: Machine Representations of Knowledge, John Wylie, to appear, 1977. (Also AI Memo 271, Stanford University, October 1975).

Shortliffe E H,
Judgmental knowledge as a basis for computer-assisted clinical decision making, Proceedings of the 1975 International Conference on Cybernetics and Society, pp 256-7, September 1975.

Shortliffe E H, Axline S, Buchanan B G, Davis R, Cohen S,
A computer-based approach to the promotion of rational clinical use of antimicrobials, in Gouveia, Tognoni and Van der Kleijn (Eds.), Clinical Pharmacy and Clinical Pharmacology, pp 259-274, Elseiver/North Holland Biomedical Press, 1976.

E H Shortliffe, R Davis, S G Axline, B G Buchanan, C C Green, S N Cohen,
Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system, Computers and Biomedical Research, 8:303-320 (August 1975).

E H Shortliffe and B G Buchanan,
A Model of Inexact Reasoning in Medicine, Mathematical Biosciences 23:351-379, 1975.

Shortliffe E H, Rhame F S, Axline S G, Cohen S N, Buchanan B G, Davis R, Scott A C, Chavez-Pardo R, and van Melle W J
MYCIN: A computer program providing antimicrobial therapy recommendations (abstract only). Presented at the 28th Annual Meeting, Western Society For Clinical Research, Carmel, CA, 6 Feb 1975. Clin. Res. 23:107a (1975). Reproduced in Clinical Medicine, p. 34, August 1975.

Shortliffe E H
MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection (abstract only); Proceedings of the ACM National Congress (SIGBIO Session), p. 739, November 1974. Reproduced in Computing Reviews 16:331 (1975).

E H Shortliffe, S G Axline, B G Buchanan, S N Cohen,
Design considerations for a program to provide consultations in clinical therapeutics, Presented at San Diego Biomedical Symposium 1974 (February 6-8, 1974).

E H Shortliffe, S G Axline, B G Buchanan, T C Merigan, S N Cohen.
An artificial intelligence program to advise physicians regarding antimicrobial therapy, Computers and Biomedical Research, 6:544-560 (1973).

Articles About MYCIN

"Which Antibiotic?" Emergency Medicine, January 1977, pp 152-162.