Many techniques are known for speeding up these searches, but most either use more space (by redundantly storing information), or slow down the other searches significantly. Richard Rashid, a graduate student successfully developed a scheme which avoids these pitfalls. He has accomplished this by changing the hash coding algorithm so that the hash code (from hashing the attribute and object together) contains some information about its parents. He thus can perform searches of types (6) and (7) by iterating through the parts of the hash code normally supplied by the first component in search (7) or the second component in search (6). The size of this set of possibilities is on the order of sqrt(N) where N is the number of buckets as opposed to N in the current LEAP implementation.

We were concerned that this new hashing algorithm might not be as good as the old for searches of types (1) - (5). In preliminary test runs with complex programs involving a large use of LEAP, however, we have found more than a 10% improvement -- far more than we would have expected from the improvements in searches (6) and (7) alone. We have therefore hypothesized that the new hashing algorithm is actually better for the normal uses of LEAP than the old and hope to begin testing that theory in the near future. If our preliminary results are shown to be valid, we expect to incorporate this new algorithm into SAIL itself sometime in the near future.

C. Automatic Choice of Associative Data Structures

Many computer applications, especially in AI and information retrieval, deal with relational data and use associative retrieval techniques. The choice of a good associative data structure for a given program is often crucial; a poor choice could well lead to gross inefficiencies in storage space and search times. Furthermore, there is no "general purpose" associative data structure that does the job. Although many general schemes work, virtually every program that would use one has some particular behaviour pattern for which the general data structure is sub-optimal. Significant improvements in performance are usual after changing from a general scheme to one that is chosen to match the requirements of the program at hand.

We are using SUMEX to study ways to systematize the selection of such data structures. We are developing techniques to model the behaviour of programs, the structure of data bases, and the properties of an important class of representation techniques. The goal is to learn how to build a "smart compiler" that will automatically select an associative data structure for a given program. Such a system will analyze a given program, ask questions of the program's designer analyze examples of the execution of the program, and then compose a data structure package from a library of data structure techniques. The selection will be based on the cost (for a given program) of candidate data structures, given by a function of their expected storage space and execution time requirements.

We started using SUMEX in a serious way in December, 1974. Since then, we have

1. implemented a library of associative data structure techniques (for n-ary relations), and an interactive program for composing a data structure package from this library. The techniques include hash tables, property lists, records, partially and fully inverted files, and methods of sharing storage between hash tables and inverted files. The programs are written in BCPL.

2. modified Jim Low's system for analyzing a SAIL program to include:

   a. extensions to the SAIL syntax to allow n-ary relations and n-ary associative retrieval

   b. various extensions to the modules that model the ways in which the given program uses its relational data.

We are currently studying the properties of the various representation techniques in our library, and the model of program behaviour that is derived by the analysis programs. We expect to begin soon to formulate and implement heuristics for suggesting representation techniques from usage patterns that appear in the model.

Comments on SUMEX:

My overall impression of the combination of TYMNET and SUMEX is that "it is OK: I can get my work done". In response to your request, I will write down my most strongly held criticisms, though I really don't feel them very strongly, since I am able to work this way.

1. TYMNET is down too much of the time, and crashes too often.

2. SUMEX is down too much of the time, especially during the day on weekends.

3. TYMNET echoing is a mess. TYMNET echoing behaviour should be adjustable under program control. I would often rather wait for echoing (and type ahead) than be confused by messed-up prompting and echoing, for example.

4. Response to control C is painfully slow.

On the other hand,

1. the load average is never too high for my work,

2. the people with whom I have dealt have been helpful and courteous

D. Planning and Acting Under Uncertainty

Any attempt to apply Artificial Intelligence methods to medicine will have to deal with the uncertainties, risks and costs involved in a clinical situation. Traditional AI techniques have been developed for purely symbolic situations where these issues were not so important. Traditional automated diagnosis techniques have involved the use of decision theory to solve these problems but have foundered on the huge trees required for realistic problems. We are engaged in fundamental and applied research on methods for combining heuristic and decision-theoretic methods.

E. Plan for the Immediate Future

When the University of Rochester set up the Computer Science Department in 1974, interdisciplinary studies were one of the paramount goals. The Medical School is one of the strongest parts of the University and already has strong ties with the Computer Engineering program. These factors plus the interests of the Computer Science faculty make a concerted effort on AI and Medicine very attractive. It is likely that the academic year 1975-76 will see joint appointments with Radiology and with Obstetrics. This will enable us to greatly expand a small current program in intelligent processing of medical images. We also intend to apply the work on planning and acting (section D) in an appropriate clinical context. The ideal would be to find an area where our image processing and problem solving efforts would be symbiotic. When the staff gathers next fall, we will make a concerted effort to define a research plan in the AIM domain.

IV.B.1.d NATURAL LANGUAGE UNDERSTANDING

Investigator: Prof. R. Lindsay
University of Michigan

(Financial support from University of Michigan)

This SUMEX account has been active since January 1975. The user is Kathie Gourlay in Ann Arbor, Michigan. Mrs. Gourlay is an assistant to Professor Robert K. Lindsay, visiting Stanford this year.

In the three months of this project´s existence, our main objective has been to familiarize Mrs. Gourlay with SUMEX. She has been learning TENEX, SOS, and INTERLISP. Access is over TYMNET´s Detroit node. We have found SNDMSG and LINKing to be very useful devices.

No substantive research has yet resulted. It is hoped that we will continue next year, as a non-Stanford SUMEX project. That work, for which use to date has been groundwork, will be concerned with natural language processing, particularly the development of memory structures for word meanings, and (perhaps) languages for organizing large data bases. In addition, Professor Lindsay will continue to work, via SUMEX, with Dr. Engelmore and Dr. Freer (UCSD) on the protein crystallography project.

We have concluded that some means of obtaining remote listings at reasonable speeds is essential. We are attempting to use a Centronix model 308 teletype for this purpose, but have not yet worked out the details. If successful, other installations might be interested in our experience with this device, which should be capable of listing over phone lines (not TYMNET) at 120 cps; the price of the model 308 is between $3000 and $4000, depending on options. A high speed input device such as a cassette based terminal would also be useful as a means of reducing phone charges. We have not looked into this, but mention it in case others have a similar need. The communications features of the system have been very useful, as noted above.

IV.B.1.e QUANTUM CHEM. INVEST. OF HEME PROTEINS AND FERREDOXINS

Investigator: Dr. Gilda Loew (Genetics)

(Grant NSF GB-40105, 2 years, $18,000 this year)

I. Current projects and goals involving use of the PDP/KI-10 computer of SUMEX are:

A. Study of antiferromagnetic behavior of the 2-iron family of sulfur containing proteins such as ferredoxin using

1. The method of electric field gradient

2. The method of electron correlation

3. And the method of magnetic susceptibility;

B. Calculations of quantum electromagnetic properties including

1. Nuclear-electron spin-spin coupling

a. Fermi contact term which includes ligand contribution

b. Dipolar tensor which excludes ligand contribution

2. Electric field gradient tensor and quadrupole splitting

3. Zero-magnetic-field splitting as a result of spin-orbit coupling, and g(gyromagnetic ratio) values for the active site of such iron-containing compounds as hemoglobin, ferredoxins, ferrichrome-A´s, mycobactins and of ferrecines.

C. The goal for next year will be to complete the studies described above.

II. Summary of project accomplishment by means of SUMEX

A. Our investigation of ferredoxin using the method of EFG (I.A.1) strongly suggests the existence of antiferromagnetic coupling; hence, we shall pursue the subject further by means of I.A.2 & 3;

B. We have successfully accounted for the set of properties described in I.B for both of the oxidized and reduced state of the ferredoxin compounds at their active sites

C. We have obtained a picture of antiferromagnetic coupling between the oxygen substituent and the iron at the hemoglobin active site from the results of EFG calculations consistent with that previously postulated from electronic spectra;

D. We have greatly extended the flexibility of the 1-electron g-value program to calculate related properties for many-electron system using the hole-excitation concept;

E. Regarding the study of the ferrocenes, a sandwich charge-transfer complex with iron in the middle, we have shown that semi-empirical methods give parallel results to ab-initio methods in describing the ionization process of the compounds and our calculated EFG show the collapse of the quadrupole splitting upon ionization in agreement with experiment;

F. Regarding the study of the ferrochrome-A, a type of iron transport compounds, our preliminary results on g-values and spin-spin coupling indicated that the latter can be vastly improved by using spin-mixing wavefunction from the former as basis function;

III. Comments

A. The interactive nature of the SUMEX computer has been a great help in the course of our effort to probe the validity of our methods of calculations. It has also helped a great deal in the use of the 1-electron and 5-electron g-value programs where extensive parameterization is required and immediate feedback is extremely valuable owing to the uncertainty of the values of the parameters;

B. DDT has been a very convenient debugging tool for us, and so are the TENEX file-handling facilities;

C. We hope to take advantage of SAIL´s excellent capabilities for handling utilities to construct

1. An information retrieval system to keep track of and quick reference of our increasing amount of results;

2. Tape utilities programs for storage and transporting;

D. We suspect that there is a bug in one of the FORTRAN library routines such as CABS and there seems to be questions about the PA10/50´s interpretation of the UUO´s;

IV. We have noticed that our needs to use the SUMEX computer have sometimes exceeded our original expectations and hence our original agreement[*] with SUMEX due to pressure from our research schedule. We have also tried whenever possible to use the computer at nights and abide with our agreement. In the future, we shall also try to use batch mode whenever possible. Finally, we are very grateful for our privilege to use the SUMEX computer and other facilities

[*] These studies are part of a trial collaboration to

investigate ways of introducing heuristic approaches for more efficient structure determination. In view of the large potential of other aspects of Dr. Loew´s work for large "number-crunching" consumption of CPU resources, Dr. Loew has agreed to conduct here work on a non-interference basis. The great bulk of her computation is done on other machines.

IV.B.1.f AUTOMATIC INTERMACHINE PROGRAM TRANSLATION

PILOT PROJECT SUMMARY FOR SUMEX ANNUAL REPORT

Investigator: Mr. J. Warren (Elect. Engrng.)

(Research Assist under grant NSF GJ-41644, 2 years, $43,748 total, Prof. T. Bredt, Stanford EE, P.I.)

1. UPDATED ABSTRACT:

This work concerns the development of a methodology and set of tools to assist in automatic transporting of computer programs between computers in any given class of machines, initially a large variety of minicomputers. The approach being used is that of "automatic programming":

A translating system is under development that will accept a description of the instruction set of any machine in the target class. From that description, it will derive or "learn" instruction sequences that perform more complex functions. It will then apply what it has learned to the translation of high-level language programs, through an intermediate form that is phrased in terms of those "more complex functions", into object code for the desired target machine. The translator creates the translation algorithms, on its own, utilizing only the static instruction set description, and built-in knowledge of the semantics of the primitives in which both input and instruction sets are described.

Essentially, this project involves the application of the concepts and techniques of artificial intelligence to a very low-level intellectual activity, and is in hopes of obtaining results that will be of significant practical value, both within and beyond the area of biomedical computer applications.

2. SUMMARY OF ACCOMPLISHMENTS OVER THE PAST YEAR

(Use of SUMEX facilities only began in earnest around the middle of February.) This translation system is still in a relatively early stage of design. Only recently have parts of the design been sufficiently solidified to allow some entry of data into SUMEX facilities, and the initiation of construction of some of the programs that will ultimately assist the proposed translation system.

3. COMMENTS CONCERNING SUMEX FACILITIES

(Use of SUMEX facilities only began in earnest around the middle of February. It took several weeks of part-time activity to learn TECO sufficiently well to use it as a production editor. Current efforts are to learn SAIL sufficiently well to use it as a production compiler.)

To date, the majority of this investigator's time spent on SUMEX has been expended in familiarization with the system, and choosing system facilities to be used (e.g. choosing an editor, and choosing a high-level language in which to construct the translator system). The SUMEX facility is a delight to use; it is much preferable to any of the several other interactive systems with which this researcher has had experience, both in general and for this particular project. Further, the staff have been consistently cooperative and helpful concerning system facilities and problems. The only problems noted have concerned telephone communications and have been telephone company hardware problems; not SUMEX problems. Documentation of some software has been somewhat cumbersome (e.g. SAIL documentation [*]), however, this is a problem that is virtually universal among general-purpose computing facilities and is by no means unique to SUMEX. Further, it is fairly evident that significant efforts are being made to upgrade the weaker documentation.

[*] "This is universally agreed. We are currently working on a TENEX-oriented revision of the NIH-DCRT "Beginner's SAIL Manual".

IV.B.2 NATIONAL PILOT PROJECTS

There are no pilot projects charged to the allocation of the AIM Executive Committee resource at this time. The management committees are considering a number of projects which may be enabled in this category in the future.

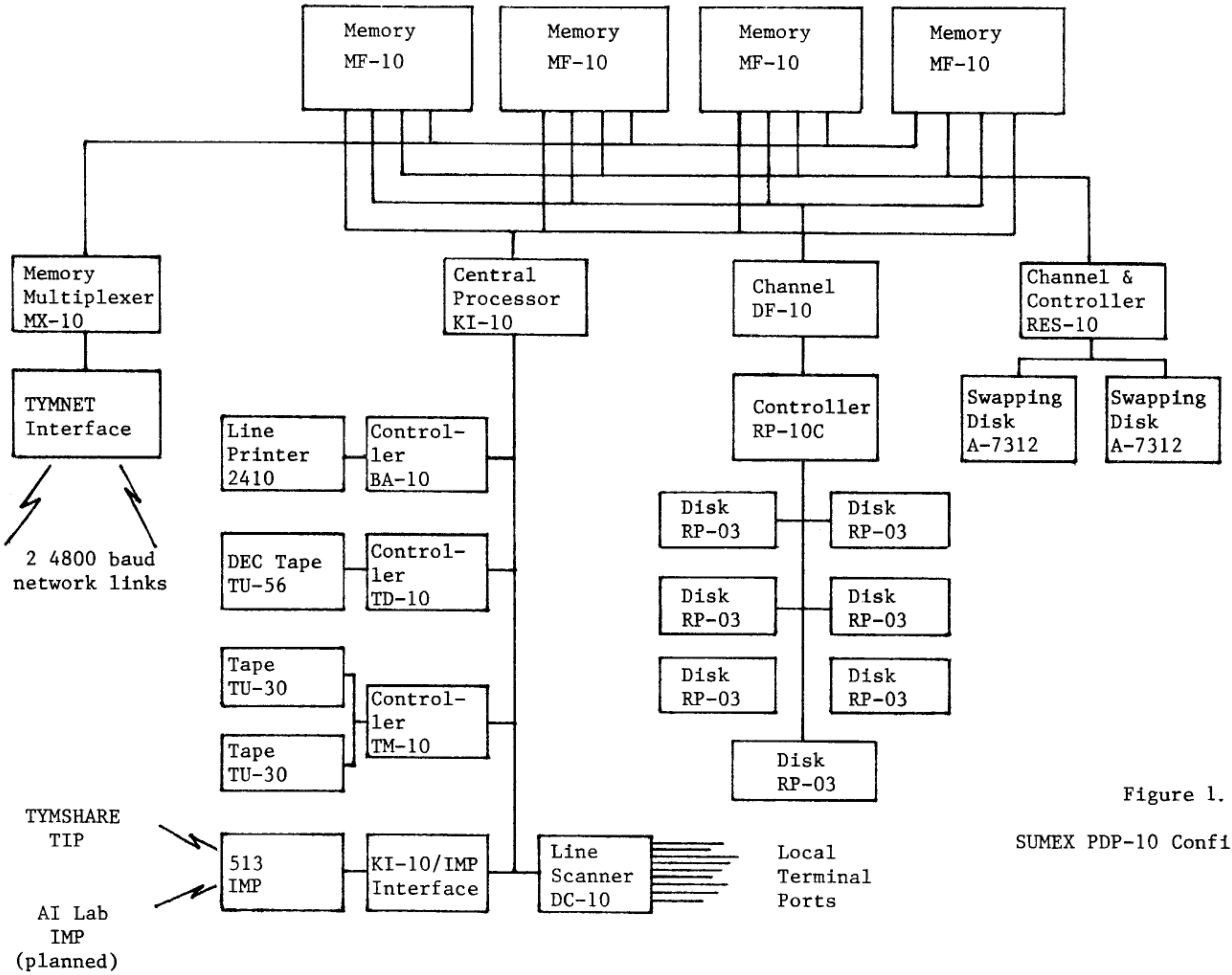


Figure 1.

SUMEX PDP-10 Configuration

Figure 2. TYMNET Network Map

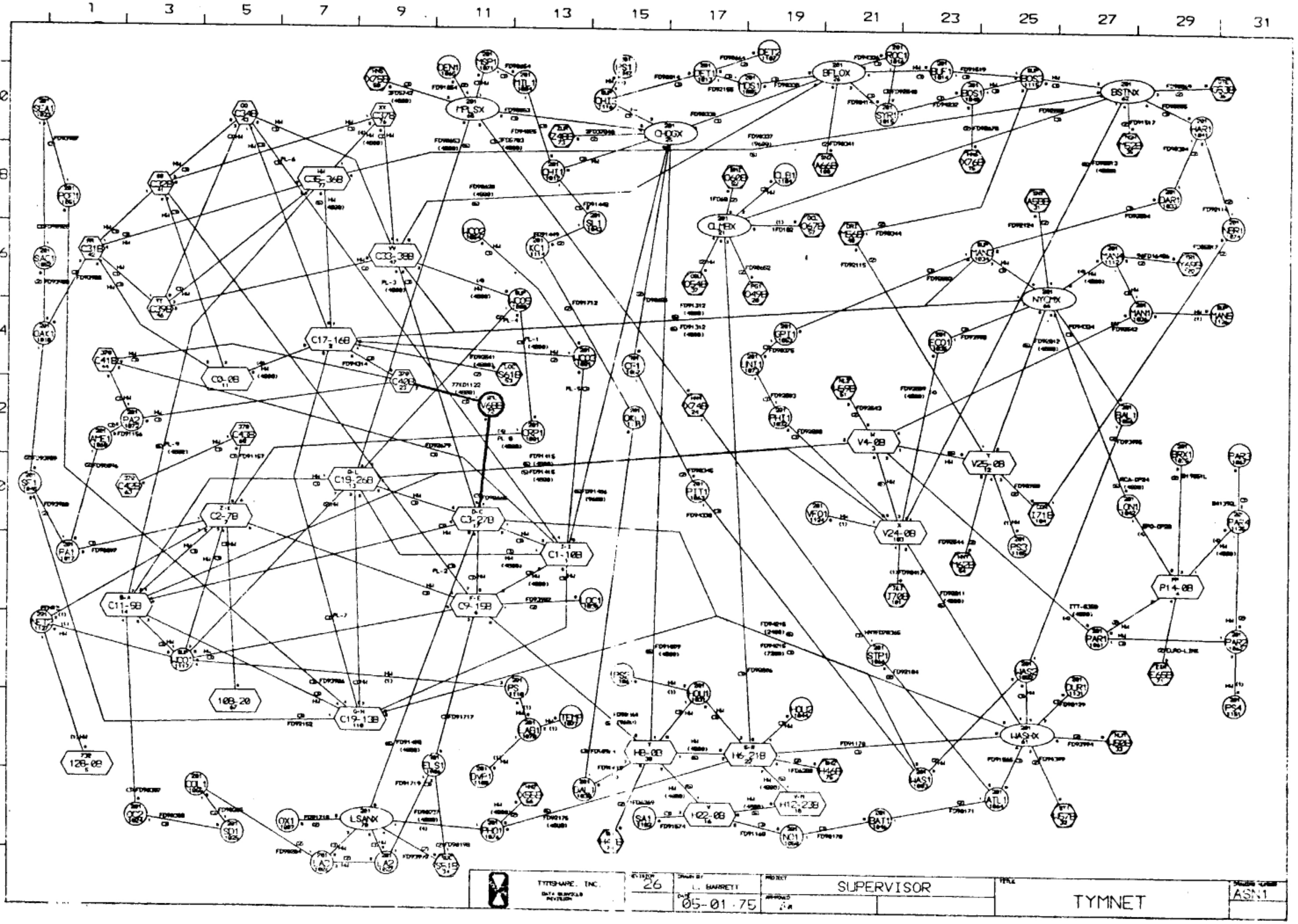
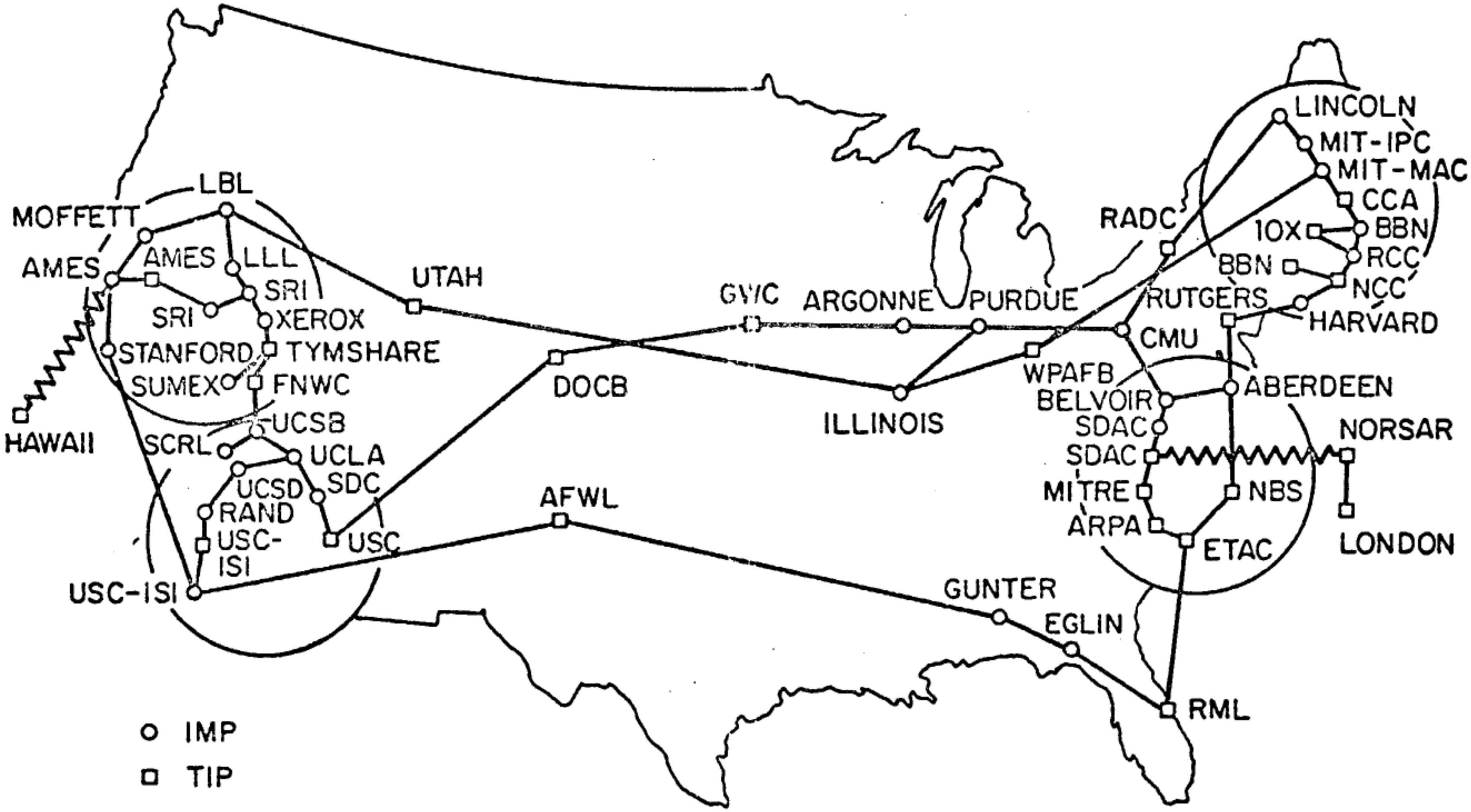


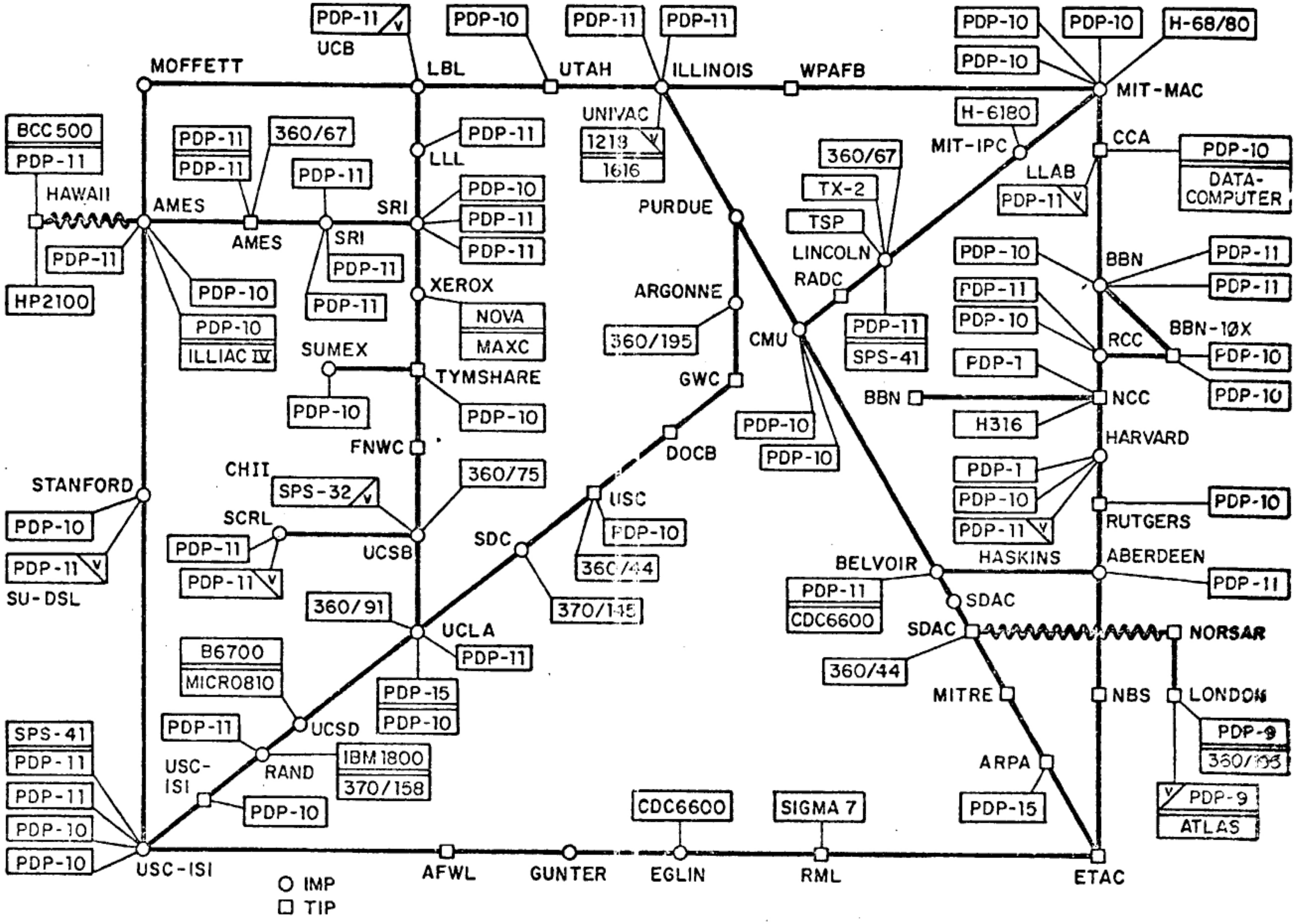
Figure 3.

ARPA NETWORK, GEOGRAPHIC MAP

APRIL 1975



ARPA NETWORK, LOGICAL MAP, APRIL 1975



APPENDIX A

AI Overview by E. A. Feigenbaum

ARTIFICIAL INTELLIGENCE RESEARCH

What is it? What has it achieved? Where is it going?

Excerpt from a report by
Professor Edward A. Feigenbaum
Stanford University

INTRODUCTION

In this briefing, these questions will be discussed as succinctly as possible:

I. What is the scientific field of artificial intelligence research, as seen from various viewpoints? What are the general goals of the field?

II. What are its practical working goals? What are some achievements relative to these goals (circa 1973)?

III. What steps (new goals, problems, potential achievements) seem to lie ahead, within a five year horizon?

ARTIFICIAL INTELLIGENCE (alias INTELLIGENT COMPUTER SYSTEMS):

General View;

Artificial Intelligence research is that part of Computer Science that is concerned with the symbol-manipulation processes that produce intelligent action. By "intelligent action" is meant an act or decision that is goal-oriented, arrived at by an understandable chain of symbolic analysis and reasoning steps, and is one in which knowledge of the world informs and guides the reasoning.

Some scientists view the performance of complex symbolic reasoning acts by computer programs as the sine qua non for artificial intelligence programs, but this is necessarily a limited view.

Yet another view unifies AI research with the rest of Computer Science. It is an oversimplified view, but worthy of consideration. The potential uses of computers by people to accomplish tasks can be "one-dimensionalized" into a spectrum representing the nature of instruction that must be given the computer to do its job. Call it the WHAT-TO-HOW spectrum. At one extreme of the spectrum, the user supplies his intelligence to instruct the machine with precision exactly HOW to do his job, step-by-step. Progress in Computer Science can be seen as steps away from that extreme "HOW" point on the spectrum: the familiar panoply of assembly languages, subroutine libraries, compilers, extensible languages, etc. At the other extreme of the spectrum is the user with his real problem (WHAT he wishes the computer, as his instrument, to do for him). He aspires to communicate WHAT he wants done in a language that is comfortable to him (perhaps English); via communication modes that are convenient for him (including perhaps, speech or pictures); with some generality, some abstractness, perhaps some vagueness, imprecision, even error; without having to lay out in detail all necessary subgoals for adequate performance - with reasonable assurance that he is addressing an intelligent agent that is using knowledge of his world to understand his intent, to fill in his vagueness, to make specific his abstractions, to correct his errors, to discover appropriate subgoals, and ultimately to translate WHAT he really wants done into processing steps that define HOW it shall be done by a real computer. The research activity aimed at creating computer programs that act as "intelligent agents" near the WHAT end of the WHAT-TO-HOW spectrum can be viewed as the long-range goal of AI research. Historically, AI research has always been the primary vehicle for progress toward this end, though science as a whole is largely unaware of the role, the goals, and the progress.

HISTORICAL TRACE

The Working Goals of the Science;
Progress toward those goals;

The root concepts of AI as a science are 1) the conception of the digital computer as a symbol-processing device (rather than as merely a number calculator); 2) the conception that all intelligent activity can be precisely described as symbol-manipulation. (The latter is the fundamental working hypothesis of the AI field, but is controversial outside of the field.) The first inference to be drawn therefrom is that the symbol-manipulations which constitute intelligent activity can be modeled in the medium of the symbol-processing capabilities of the digital computer.

This intellectual advance--which gives realization in a physical system, the digital computer, to the complex symbolic processes of intelligent action and decision--with detailed case studies of how the realization can be accomplished, and with bodies of methods and techniques for creating new demonstrations--ranks as one of the great intellectual achievements of Science, allowing us finally to understand how a physical system can also embody mind. The fact that large segments of the intellectual community do not yet understand that this advance has been made does not change its truth or its fundamental nature.

Three global "working goals" have dominated the AI field for the 17 years of its existence. These are:

1. Understanding heuristic search as a processing scheme sufficient to account for much intelligent problem solving behavior; and exploring the scope and pervasiveness of heuristic problem solving.

2. Semantic information processing: developing precise formulations of "understanding" by programs, and "meaning" of symbols that are input or stored; the acquisition, storage, and deployment of knowledge of the world in the service of symbolic problem solving.

3. Information Processing Psychology: developing precise models of human behavior in symbolic-processing tasks.

The first two goals represent the fundamental paradigms that have dominated the field. The third cuts across these orthogonally, and involves intense interdisciplinary contact with Psychology, and Linguistics.

GOAL 1. HEURISTIC SEARCH, HEURISTIC PROGRAMMING, SYMBOLIC PROBLEM SOLVING PROGRAMS

In the first decade, the dominant paradigm of AI research was heuristic search. In this paradigm, problem solving is conceived as follows: A tree of "tries" (aliases: subproblems, reductions, candidates, solution attempts, alternatives-and-consequences, etc.) is sprouted (or sproutable) by a generator. Solutions (variously defined) exist at particular (unknown) depths along particular (unknown) paths. To find one is a "problem". For any task regarded as nontrivial, the search space is very large. Rules and procedures called heuristics are applied to direct search, to limit search, to constrain the sprouting of the tree, etc. While some of this tree-searching machinery is entirely task-specific, other parts can be made quite general over the domain of designs employing the heuristic search paradigm. Two notions are critical. The first is that problem solvers generally face a "maze" of alternative courses of decision and action that is huge compared with their processing resources. The second is the use of heuristic knowledge to steer carefully through large mazes toward a solution seeking the plausible and potentially fruitful avenues, avoiding the absurdities and the high-risk paths. Heuristic knowledge is usually informal knowledge--to be distinguished from formal knowledge that is assertable with the rigor of proof. Polya, the famous mathematician who wrote Patterns of Plausible Inference and other books on problem solving, calls heuristic reasoning "the art of good guessing." Heuristic knowledge is often "common sense" knowledge of the world, rules-of-thumb for generally acceptable performance, or rules of good practice in specific situations. When we speak of the "expertise" of an expert, and the "good judgment" he brings to bear on complex problems in his domain, we often are speaking of the heuristics he has developed to search effectively.

Provocative essays by Polya notwithstanding, the first serious and detailed studies of heuristic problem solving ever done by Science were done as AI research in its first decade. As with any other science, progress came by the detailed examination of specific cases, from which gradually emerged both a broad picture of the nature of the phenomena being studied and, within this, more formal theories for specific parts.

Three sub-goals of heuristic programming are discernable.

SUBGOAL 1A. Demonstrate sufficiency of heuristic search for tasks of intellectual difficulty.

These heuristic programming efforts dealt with almost "pure" symbolic reasoning tasks (i.e., tasks not requiring much coupling to real-world knowledge), and used inference schemes that were either ad-hoc or of limited scope. Notable successes during this "prove-the concept" phase were: the Logic Theory Program, that proved theorems in Whitehead & Russel's propositional calculus; the Geometry Theorem Proving program, that proved theorems in Euclidean geometry at a level of competence exceeding that of the excellent high school geometry student; the Symbolic Integration program, that solved college freshman symbolic integration problems about as well as MIT freshmen; chess-playing programs that play respectable "club player" C or B Class chess; a checker playing program that was virtually unbeatable, except by the country's top few players (notable also for remarkable self-improvement in performance by analysis of its own play and "book-move" good play); and a number of competent management science applications (assembly-line balancing, warehouse location, job-shop scheduling, etc.).

To recapitulate briefly: the key concepts are: search in problem solving; and the use of generally informal knowledge to guide search effectively. The AI community was the first to devote serious scientific effort to developing the idea of the use of informal knowledge in problem solving, with notable successes. Few in Science recognize that this achievement has been made and is ready for exploitation.

SUBGOAL 1B. Generality in Problem Solving Programs

Generality here means the use of a small set of problem solving methods of wide applicability to solve problems of many different types. Each of the problems posed is stated to the program in a particular representation (or framework) with which the set of methods is constructed to handle.

The subgoal of generality arises first as a reaction to the array of "specialty" programs mentioned above; second, from the general observation that the ability to do a wide range of tasks is a special touchstone of intelligence; third, from a direct assessment that as the diversity and heterogeneity of the tasks handled by an agent increases, the likelihood that it can do them all without intelligent action decreases; and fourth, from the argument that any ultimate intelligent agent must have wide generality, since it must take the world and its problems as they come without any intermediary, making generality an important independent desideratum.

This subgoal was pursued with vigor for ten years in a number of projects, was important for its feedback value in clarifying issues for the AI field, and has temporarily (at least) been put back on the shelf as the field begins to explore knowledge-based problem solvers and issues in the representation of knowledge.

There were two discernable subthemes. The first was an attempt to create abstract heuristic search methods that were divorced from any particular content. Examples were: the General Problem Solver, which used a variant of heuristic search known as means-ends analysis; MULTIPLE, which introduced adaptivity in the selection of what subproblem to choose "next" in a search; and REF-ARF, which extended the generality of ordinary procedural programming languages to include the embedding of non-procedural problems of constraint satisfaction.

The second subtheme was the construction of theorem provers that take problems expressed as theorems to be proved in the first-order predicate calculus. This line of work was motivated by the (correct) observation that the scope for representing real-world facts and situations in first-order predicate calculus is very great; and by the invention of the resolution method, a computational method for finding proofs for theorems in this calculus. There has been continuous improvement on the basic method, taking the form of proposing more powerful inference techniques, rather than the form of specific ways for programs to adapt to particular problems. The very strength of the formulation in terms of generality, namely its complete homogenization of the particular task (all tasks are seen and dealt with in the same logical formalism) turns effort away from how to exploit the particularities of special classes of tasks. But it appears that only by exploiting the particularities can significant reduction in search be achieved. From a practical point of view the only proofs produced by such problem solvers were "shallow" proofs.

Much of this line of research has been temporarily "shelved", awaiting further knowledge on how best to represent knowledge for computer processing. Problems that are essentially simple when represented in their "natural" representation appear extraordinarily complicated when translated into first-order predicate calculus. The current search for theorem provers using higher-order logics is based not on the attempt to increase the raw expressive power, so to speak, of first-order logic, but on the belief that naturalness of expression will ultimately pay off.

SUBGOAL 1C: High-Performance Programs that perform at near-human level in specialized areas

As the heuristic programming area matured to the point where the practitioners felt comfortable with their tools, and adventuresome in their use; as the need to explore the varieties of problems posed by the real-world was more keenly felt; and as the concern with knowledge-driven programs (to be discussed later) intensified, specific projects arose which aimed at and achieved levels of problem solving performance that equalled, and in some cases exceeded, the best human performance in the tasks being studied. The example of such a program most often cited in the Heuristic DENDRAL program, which solves the scientific induction problem of analyzing the mass spectrum of an organic molecule to produce a hypothesis about the molecule's total structure. This is a serious and difficult problem in a relatively new area of analytical chemistry. The program's performance has been generally very competent and in "world's champion" class for certain specialized families of molecules. Similar levels of successful performance have been achieved by some of the MATHLAB programs that assist scientists in doing symbolic mathematics. The effectiveness of MATHLAB's procedures for doing symbolic integration in calculus is virtually unexcelled. Yet another example, with great potential economic significance, involves a program for planning complex organic chemical syntheses from substances available in chemical catalogs. The program is currently being used as an "intelligent assistant" in a new and complex organic synthesis.

GOAL 2. SEMANTIC INFORMATION PROCESSING (S.I.P.)

The use of the term "semantic" above is intended to connote, in familiar terms, something like: "What is the meaning of..." or "How is that to be understood..." or "What knowledge about the world must be brought to bear to solve the particular problem that has just come up?" The research deals with the problem of extracting the meaning of: utterances in English; spoken versions of these; visual scenes; and other real-world symbolic and signal data. It aims toward the computer understanding of these as evidenced by the computer's subsequent linguistic, decision-making, question-answering, or motor behavior.

Thus, for example, we will know that our "intelligent agent" understood the meaning of the English command we spoke to it if: a) the command was in itself ambiguous; b) but was not ambiguous in context; and c) the agent performed under the appropriate interpretation and ignored the interpretation that was irrelevant in context.

In this goal of AI research, there are foci upon the encoding of knowledge about the world in symbolic expressions so that this knowledge can be manipulated by programs; and the retrieval of these symbolic expressions, as appropriate, in response to demands of various tasks. S.I.P. has sometimes been called "applied epistemology" or "knowledge engineering".

To summarize: the AI field has come increasingly to view as its main line of endeavor: knowledge representation and use, and an exploration of understanding (how symbols inside a computer, which are in themselves essentially abstract and contentless, come to acquire a meaning).

To classify all of the current work into a relatively simple set of subgoals is a formidable and hazardous undertaking. Nevertheless, here is one rough cut (stated for convenience as questions).

A. How is the knowledge acquired, that is needed for understanding and problem solving; and how can it be most effectively used?

B. How is knowledge of the world to be represented symbolically in the memory of a computer?

B1. What symbolic data structures in memory make the retrieval of this information in response to task demands easy?

C. How is knowledge to be put at the service of programs for understanding English?

D. How is sensory knowledge, particularly visual and speech, to be acquired and understood? How is knowledge to be applied to intelligent action of effectors, such as arms, wheels, instrument controls, etc.

Significant advances on all of these fronts have been made in the last decade. The area has a rather remarkable coherence--with individual projects threading through a number of the goals stated above (this makes excellent science and difficult exposition!)

GOAL 2A. Knowledge Acquisition and Deployment for Understanding and Problem Solving

The paradigm for this goal is, very generally sketched, as follows:

a. a situation is to be described or understood; a signal input is to be interpreted; or a decision in a problem-solution path is to be made.

> Examples: A speech signal is received and the question is, "What was said?" The TV camera system sends a quarter-million bits to the computer and the question is, "What is out there on that table and in what configuration?" The molecule structure-generator must choose a chemical functional group for the "active center" of the molecular structure it is trying to hypothesize, and the question is, "What does the mass spectrum indicate is the ´best guess´?"

b. Specialized collections of facts about the various particular task domains, suitably represented in the computer memory (call these Experts) can recognize situations, analyze situations, and make decisions or take actions within the domain of their specialized knowledge.

> Examples: In the CMU Hear-Say Speech Understanding System, currently the Experts that contribute to the Current Best Hypothesis are an Acoustic-Phonetic Expert, a Grammar Expert, and a Chess Expert (since chess-playing is the semantic domain of discourse). In Heuristic DENDRAL, the Experts are those that know about stability of organic molecules in general, mass spectrometer fragmentation processes in particular, nuclear magnetic resonance phenomena, etc.

For each of the sources of knowledge that can be delineated, schemes must be created for bringing that knowledge to bear at some place in the ongoing analysis or understanding process. The view is held that programs should take advantage of a wide range of knowledge, creating islands of certainty as targets of opportunity arise, and using these as anchors for further uncertainty reduction. It is an expectation that always some different aspect provides the toe-hold for making headway--that is , that unless a rather large amount of knowledge is available and ready for application, this paradigmatic scheme will not work at all.

Within this paradigm lie a number of important problems to which AI research has addressed itself:

a. Since it is now widely recognized that detailed specific knowledge of task domains is necessary for power in problem solving programs, how is this knowledge to be imparted to, or acquired by, the programs?

a1. By interaction between human expert and program, made ever more smooth by careful design of interaction techniques, languages "tuned" to the task domain, flexible internal representations. The considerable effort invested by the AI community on interactive time-sharing and interactive graphic display was aimed toward this end. So is the current work on situation-action tableaus (production systems) for flexibly transmitting from expert to machine details of a body of knowledge.

a2. "Custom-crafting" the knowledge in a field by the painstaking day-after-day process of an AI scientist working together With an expert in another field, eliciting from that expert the theories, facts, rules, and heuristics applicable to reasoning in his field. This was the process by which Heuristic DENDRAL´s "Expert" knowledge was built. It is being successfully used in AI application programs to: diagnosis of glaucoma eye disease, to treatment planning for infectious disease using antibiotics, to protein structure determination using X-ray crystallography, to organic chemical synthesis planning, to a military application involving sonar signals, perhaps to other areas, and of course to chess.

a3. By inductive inference done by programs to extract facts, regularities, and good heuristics directly from naturally-occurring data. This is obviously the path to pursue if AI research is not to spend all of its effort, well into the 21st Century, building knowledge-bases in the various fields of human endeavor in the custom-crafted manner referred to above. The most notable successes in this area have been:

> ...the Meta-DENDRAL program which, for example, has discovered the mass spectrum fragmentation rules for aromatic acids from observation of numerous spectra of these molecules--rules previously not explicated by the DENDRAL chemists.
>
> ...a draw-poker playing program that inferred the heuristics of good play in the game by induction (as well as by other modes, including the aforementioned interaction with experts).

a4. By processes of analogical reasoning, by which knowledge acquired about one area can be used to solve problems in a another area if a suitable analogy can be drawn. Our human experience tells us that this approach is rich in possibilities. One successful project can be cited (and that is a limited success); a program that discovers an analogy (in full-blown detail) between a theorem-to-be-proved in modern algebra and another theorem in algebra whose proof is known. The analogy is used to pinpoint from a large set of facts those few that will indeed be relevant to proving the new theorem.

GOAL 2B. Representation of Knowledge

The problem of representation of knowledge for AI systems is this: if the user has a fact about the world, or a problem to be stated, in what form does this become represented symbolically in the computer for immediate or later use? Three approaches are being pursued:

B1. the approach via formal logic. As mentioned before, first-order predicate calculus was tried, but was found to be too cumbersome to represent ordinary situations and common-sense knowledge. Set theory and higher-order logics are currently under examination as better candidates to be a medium for homogeneous representation.

B2. The ad-hoc approach. Most problem domains have a "natural" representation that human experts use when operating in the domain. Translate that representation fairly directly for the computer, and tailor the information processes to work with it. This is the approach commonly taken, in DENDRAL, MATHLAB, in chess playing programs, visual scene analysis, and so on (almost everywhere). Though it gets the job done, it creates serious problems for the cumulation of knowledge, techniques, and programs in the science because of the inhomogeneity that arises therefrom throughout the collection of AI projects undertaken. One way out of the dilemma is to do research on the problem of translation (by program) from one ad-hoc representation into another (the so-called "shift of representation" problem). Little work has been done on this problem, except one excellent "pencil-and-paper" exercise in connection with a simple puzzle, and one subprogram in DENDRAL (the Planning Rule Generator, that translates mass spectral knowledge from its form as fragmentation processes to a form useful for pattern matching).

B3. the approach via a "computable" semantic theory. In this approach, computational linguists attempt to analyze the full range of actions, actors, objects, and their relations, of which the common-sense world is composed; then refine and formalize these into a useable computational theory for representing facts, utterances, problems, etc. The most successful of these efforts is the Conceptual Parser (and its follow-on, MARGIE, which successfully accomplishes English paraphrase and common-sense inference).

In lieu of a tight, parsimonious computable semantic theory, other more ad-hoc systems, known as semantic-net-memory models, have developed experimenting with various sorts of actor-action-object-relation data structures. Semantic-net-memory models have a ten year history relating particularly to intelligent question-answering. Perhaps most successful of these is the HAM program which combines ideas from semantic theory, semantic-net-memory structures, and more traditional linguistic analysis (all in the context of a rather good model of human sentence comprehension, validated with dozens of careful laboratory experiments).

GOAL 2C. Programs for Understanding English

One can readily observe that it will be almost impossible to disentangle the skein of research on understanding natural language (English) from the coordinate efforts in representation and deployment of knowledge. Most of the state-of-the-art programs for understanding English employ, in one form or another, the basic S.I.P. paradigm outlined previously. These systems have substantial linguistic components that are highly sophisticated compared with anything done in the past. All of them incorporate