II. Description of Program Activities

This section corresponds to the predefined forms required by the Division of Research Resources to provide information about our resource activities for their computerized retrieval system. These forms have been submitted separately and are not reproduced here to avoid redundancy with the more extensive narrative information about our resource and progress provided in this report.

II.A. Scientific Subprojects

Our core research and development activities are described starting on page 18, our training activities are summarized starting on page 49, and the progress of our collaborating projects is detailed starting on page 85.

II.B. Books, Papers, and Abstracts

The list of recent publications for our core research and development work starts on page 45 and those for the collaborating projects are in the individual reports starting on page 85.

II.C. Resource Summary Table

The details of resource usage, including a breakdown by the various subprojects, is given in the tables starting on page 52.

III. Narrative Description

We are about to start the final grant year in the current SUMEX-AIM award. This annual report was prepared in parallel with a competing renewal proposal for continuing the resource beyond July 1986. The report is based on the comprehensive progress sections of the proposal for resource core research and development and for the collaborating scientific community.

III.A. Summary of Research Progress

III.A.1. Executive Summary

This summary provides an overview of our accomplishments. In the almost twelve years since the SUMEX-AIM resource was established, computing technology and biomedical artificial intelligence research have undergone a remarkable evolution. As we prepare to renew the resource through the remainder of the 1980's, we take pride in the realization that SUMEX has both influenced and responded to those changing technologies. It is widely recognized that our resource has fostered highly influential work in medical AI -- work from which it is generally acknowledged that the expert systems field emerged -- and that it has simultaneously helped define the technological base of applied AI research. The LISP machines to which we directed our attention in 1980 have now demonstrated their practicality as research tools and, increasingly, as potential mechanisms for disseminating AI systems as cost-effective decision aids in clinical settings such as private offices. We look forward to another half decade during which the era of centralized machines for AI research will come to an end, having been supplanted by networks of distributed and heterogeneous single-user machines sharing common resources such as file servers, printers, and gateways to other local and long-distance networks.

We continue to be motivated by three main goals:

1. to develop and provide impeccable computing resources and human assistance to scientists working on applications of artificial intelligence research in medicine and biology;

2. to demonstrate that it is feasible to provide resources and assistance to a national community of researchers from a central site, integrating distributed and centralized computing technology, local and national computer communication networks, and a staff oriented toward the special problems of individuals participating in AIM research at other institutions;

3. to develop the community of scientists interested in working on applications of AI to the biomedical sciences; facilitating the growth, health, and vigor of the community by providing electronic communications that link its members and by assisting with the dissemination of systems software and applications programs that are of use to the wider community of AIM researchers. One question we have been asking is, "Is there a new style of science that will emerge in a communications-enhanced setting of national, rather than institutional, scope?" Within a decade it was clear that the answer to this question was (and is) "yes"!

SUMEX's Success as a National Research Resource

The SUMEX Project has demonstrated that it is possible to operate a computing research resource with a national charter and that the services providable over networks were those that facilitate the growth of AI-in-Medicine. Many NIH computer RR's have been mostly institutional in scope, occasionally regional (like the UCLA resource). SUMEX now has the reputation of a model national resource, pulling together the best available interactive computing technology, software, and computer communications in the service of a national scientific community. Planning groups for national facilities in cognitive science, computer science, and biomathematical modeling have discussed and studied the SUMEX model and new resources, like the recently instituted BIONET resource for molecular biologists, are closely patterned after the SUMEX example.

A decade ago, when machines up to the task of supporting AI research cost $1M, some of the most notable projects in the history of Artificial Intelligence were done with terminal-and-network, without a computer on site. In human terms, this meant, of course, not having the headaches and energy drains of proposing a machine, installing it, maintaining it and its software, hiring its system programmers and operators, dealing with communication vendors, etc. The famous INTERNIST program was developed from Pittsburgh in this way. And the ACT computer model was begun at Michigan, continued at Yale, and later at Carnegie-Mellon, all without moving the program or losing a day's work because of machine transition problems. The GENET community of over 300 molecular biologists grew up in a year around SUMEX programs for analyzing DNA sequences. Their demand for these centralized capabilities ultimately swamped our machine and led to the initiation of a separate resource (BIONET) to meet their needs.

The projects SUMEX supports have generally required substantial computing resources with excellent interaction. Even today though, with the growing availability of Lisp workstations, this computing power is still hard to obtain in all but a few universities. SUMEX is, in a sense, a "great equalizer". A scientist gains access by virtue of the quality of his/her research ideas, not by the accident of where s/he happens to be situated. In other words, the resource follows the ethic of the scientific journal.

SUMEX has demonstrated that a computer resource is a useful "linking mechanism" for bringing together and holding together teams of experts from different disciplines who share a common problem focus. For example, computer scientists have been collaborating fruitfully with physical chemists, molecular biochemists, geneticists, crystallographers, internists, ophthalmologists, infectious disease specialists, intensive care specialists, oncologists, psychologists, biomedical engineers, and other expert practitioners. And in some of these cases, the interdisciplinary collaboration, usually so difficult to achieve in the best of circumstances, was achieved in spite of geographical distance between the participants, using the computer networks.

SUMEX has also achieved successes as a community builder. AI concepts and software are among the most complex products of computer science. Historically it has not been easy for scientists in other fields to gain access to and mastery of them. Yet the collaborative outreach and dissemination efforts of SUMEX have been able to bridge the gap in numerous cases. Over 36 biomedical AI application projects have developed in our national community and have been supported by SUMEX over the years. And 9 of these have matured to the point of now continuing their research on facilities outside of SUMEX. For example, the BIONET resource (named GENET while at SUMEX) is being operated by IntelliCorp; the Rutgers Computers in Biomedicine resource is centered at Rutgers University; the CADUCEUS project splits their research work between their own VAX computer and the SUMEX resource; and the Chemical Synthesis project now operates entirely on a VAX at U.C. Santa Cruz.

The SUMEX mission has been able to capture the contributions of some of the finest computers-in-medicine specialists and computer scientists in the country. For example, Professor Joshua Lederberg (SUMEX's first PI, now President of The Rockefeller University) is a member of SUMEX's Executive Committee; and Dr. Donald Lindberg, former Director of the University of Missouri's Medical Information Science group, and now Head of the National Library of Medicine, was until recently the Chairman of the AIM Advisory Group. Professor Herbert Simon of Carnegie-Mellon University, Professor Marvin Minsky of MIT, and many other distinguished scientists serve on that peer review committee.

SUMEX and Artificial Intelligence Research

The SUMEX Project is a relative latecomer to AI research. Yet its scope has given strong impetus to this historic development in applied computer science. AI research is that part of computer science that investigates symbolic reasoning processes, and the representation of symbolic knowledge for use in inference. It views heuristic or judgmental knowledge to be of equal importance with "factual" knowledge, indeed to be the essence of what we call "expertise". In its "Expert Systems" work, it seeks to capture the expertise of a field, and translate it into programs that will offer intelligent assistance to a practitioner in that field.

For computer applications in medicine and biology, this research path is crucial, indeed ineluctable. Medicine and biology are not presently mathematically-based sciences; unlike physics and engineering, they are seldom capable of exploiting the mathematical characteristics of computation. They are essentially inferential, not calculational, sciences. If the computer revolution is to affect biomedical scientists, computers will be used as inferential aids.

Perhaps the larger impact on medicine and biology will be the exposure and refinement of the hitherto largely private heuristic knowledge of the experts of the various fields studied. The ethic of science that calls for the public exposure and criticism of knowledge has traditionally been flawed for want of a methodology to evoke and give form to the heuristic knowledge of scientists. The AI methodology is beginning to fill that need. Heuristic knowledge can be elicited, studied, critiqued by peers, and taught to students.

The tide of AI research and application is rising. AI is one of the principal fronts along which university computer science groups are expanding. Federal and industrial support for AI research is vigorous and growing, although support specifically for biomedical applications continues to be limited. The pressure from student career-line choices is great: to cite an admittedly special case, approximately 80% of the students applying to Stanford's computer science Ph.D. program cite AI as a possible field of specialization (up from 30% 4 years ago). At Stanford, we have vigorous special programs for student training and research in AI -- a new graduate program in Medical Information Sciences and the two-year Masters Degree in AI program. All of these have many more applicants than available slots. Demand for our graduates, in both academic and industrial settings, is so high that students typically begin to receive solicitations one or two years before completing their degrees.

There is an explosion of interest in medical AI. The American Association for Artificial Intelligence (AAAI), the principal scientific membership organization for the AI field, has 7000 members, over 1000 of whom are members of the medical special interest group known as the AAAI-M. Speakers on medical AI are prominently featured at professional medical meetings, such as the American College of Pathology and American College of Physicians meetings; a decade ago, the words "artificial intelligence" were never heard at such conferences. And at medical computing meetings, such as the annual Symposium on Computer Applications in Medical Care and the international MEDINFO conferences, the growing interest in AI and the rapid increase in papers on AI and expert systems are further testimony to the impact that the field is having.

AI is beginning to have a similar effect on medical education. Such diverse organizations as the National Library of Medicine, the American College of Physicians, the Association of American Medical Colleges, and the Medical Library Association have all called for sweeping changes in medical education, increased educational use of computing technology, enhanced research in medical computer science, and career development for people working at the interface between medicine and computing. They all cite evolving computing technology and (SUMEX-AIM) AI research as key motivators.

In industry, AI is on an exponential growth path as well. In the USA alone, over 30 AI start-up companies have been formed in the past four years and many groups have been established in large companies as well. The list of names is long and includes Hewlett-Packard, Schlumberger (including Fairchild), Texas Instruments, Xerox, IBM, DEC, General Motors, General Electric, Boeing, Rockwell, FMC Corp, Ford-Aerospace, Apple Computer, Teknowledge, IntelliCorp, Syntelligence, Lucid, Inference Corp, Symbolics, LMI, and so on... Many of these firms are marketing hardware and software tools for expert system development, as well as custom system services. And Japan has mounted a long-term, well-funded "Fifth Generation" computing effort to broadly develop knowledge-based systems technology as part of their national economic base of the 1990's.

The AI tide is rising largely because of the development in the 1970's and early 1980's of methods and tools for the application of AI concepts to difficult professional-level problem solving. Their impact was heightened because of the demonstration in various areas of medicine and other life sciences that these methods and tools really work. Here SUMEX has played a key role, so much so that it is regarded as "the home of applied AI."

SUMEX has been the nursery, as well as the home, of such well-known AI systems as DENDRAL (chemical structure elucidation), MYCIN (infectious disease diagnosis and therapy), INTERNIST (differential diagnosis), ACT (human memory organization), ONCOCIN (cancer chemotherapy protocol advice), SECS (chemical synthesis), EMYCIN (rule-based expert system tool), and AGE (blackboard-based expert system tool). In the past four years, our community has published a dozen books that give a scholarly perspective on the scientific experiments we have been performing. These volumes, and other work done at SUMEX, have played a seminal role in structuring modern AI paradigms and methodology. First among these scientific directions has been a switch in AI's focus from inference procedures to knowledge representation and use. There is now a recognition that the power of problem solvers derives primarily from the knowledge that they contain -- of the elements of the problem domain, of the strategies for solving problems in that domain, and of the forms in which the knowledge is to be acquired. In 1977, Goldstein and Papert of MIT, writing in the journal Cognitive Science, described the change of focus as a "paradigm shift" in AI. This shift was induced largely (though, of course, not exclusively) by the work at SUMEX, beginning with the DENDRAL development in 1965.

Toward the '90s: the Future of SUMEX

Given this setting of success and vitality, what is the future need and course for SUMEX as a resource -- especially in view of the on-going revolution in computer technology and costs, with the emergence of powerful single-user workstations and local area networking? The answers remain clear.

At the deepest research level, despite our considerable success in working on medical and biological applications, the problems we can attack are still sharply limited. Our current ideas fall short in many ways against today's important health care and biomedical research problems brought on by the explosion in medical knowledge and for which AI should be of assistance. Just as the research work of the 70's and 80's in the SUMEX-AIM community fuels the current practical and commercial applications, our work of the late 80's will be the basis for the next decade's systems. Our growing knowledge is clearly attained in an incremental fashion; we build today on the results of the past decade, and we will build in the 1990's on the work we undertake today.

At the resource level, there is a growing, diverse, and active AIM research community with intense needs for computing resources to continue its work. Many of these groups still are dependent on the SUMEX-AIM resources. For those who have been able to take advantage of newly developed local computing facilities, SUMEX-AIM provides a central cross-roads for communications and the sharing of programs and knowledge. In its core research and development role, SUMEX-AIM has its sights set on the hardware and software systems of the next decade. We expect major changes in the distributed computing environments that are just now emerging in order to make effective use of their power and to adapt them to the development and dissemination of biomedical AI systems for professional user communities. In its training role, SUMEX is a crucial resource for the education of badly needed new researchers and professionals to continue the development of the biomedical AI field. The "critical mass" of the existing physical SUMEX resource, its development staff, and its intellectual ties with the Stanford Knowledge Systems Laboratory (previously called the Heuristic Programming Project), make this an ideal setting to integrate, experiment with, and export these methodologies for the rest of the AIM community.

III.A.2. Resource Goals and Definitions

SUMEX-AIM is a national computer resource with a multiple mission: a) promoting experimental applications of computer science research in artificial intelligence (AI) to biological and medical problems, b) studying methodologies for the dissemination of biomedical AI systems into target user communities, c) supporting the basic AI research that underlies applications, and d) facilitating network-based computer resource sharing, collaboration, and communication among a national scientific community of health research projects. The SUMEX-AIM resource is located physically in the Stanford University Medical School and serves as a nucleus for a community of medical AI projects at universities around the country. SUMEX provides computing facilities tuned to the needs of AI research and communication tools to facilitate remote access, inter- and intra-group contacts, and the demonstration of developing computer programs to biomedical research collaborators.

III.A.2.1. What is Artificial Intelligence?

Artificial Intelligence research is that part of Computer Science concerned with symbol manipulation processes that produce intelligent action [1, 26, 29, 35]. Here *intelligent action* means an act or decision that is goal-oriented, is arrived at by an understandable chain of symbolic analysis and reasoning steps, and utilizes knowledge of the world to inform and guide the reasoning.

Placing AI in Computer Science

A simplified view relates AI research with the rest of computer science. The manner of use of computers by people to accomplish tasks can be thought of as a one-dimensional spectrum representing the nature of the instructions that must be given the computer to do its job. At one extreme of the spectrum, representing early computer science, the user supplies his intelligence to instruct the machine precisely *how* to do the job, step-by-step.

At the other extreme of the spectrum, the user describes *what* he wishes the computer to do for him to solve a problem. He wants to communicate what is to be done without having to lay out in detail all necessary subgoals for adequate performance, yet with a reasonable assurance that he is addressing an intelligent agent that is using knowledge of his world to understand his intent, complain or fill in his vagueness, make specific his abstractions, correct his errors, discover appropriate subgoals, and ultimately translate what he wants done into detailed processing steps that define how it should be done by a real computer. The user wants to provide this specification of what to do in a language that is comfortable to him and the problem domain (perhaps English) and via communication modes that are convenient for him (including perhaps speech or pictures).

Progress in computer science may 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. illustrate this trend. The research activity aimed at creating computer programs that act as *intelligent agents* near the *what* end of the spectrum can be viewed as a long-range goal of AI research.

Expert Systems and Applications

The national SUMEX-AIM resource has in large part made possible a long, interdisciplinary line of artificial intelligence research at Stanford concerned with the development of concepts and techniques for building expert systems [15]. An *expert* *system* is an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. For some fields of work, the knowledge necessary to perform at such a level, plus the inference procedures used, can be thought of as a model of the expertise of the expert practitioners of that field.

The knowledge of an expert system consists of facts and heuristics. The *facts* constitute a body of information that is widely shared, publicly available, and generally agreed upon by experts in a field. The *heuristics* are the mostly-private, little-discussed rules of good judgment (rules of plausible reasoning, rules of good guessing) that characterize expert-level decision making in the field. The performance level of an expert system is primarily a function of the size and quality of the knowledge base that it possesses.

Projects in the SUMEX-AIM community are concerned in some way with the application of AI to biomedical research. Brief abstracts of the various projects currently using the SUMEX resource can be found in Appendix C on page 215 and more detailed progress summaries in Section IV on page 85. The most tangible objective of this approach is the development of computer programs that will be more general and effective consultative tools for the clinician and medical scientist. There have already been promising results in areas such as chemical structure elucidation and synthesis, diagnostic consultation, molecular biology, and modeling of psychological processes.

Needless to say, much is yet to be learned in the process of fashioning a coherent scientific discipline out of the assemblage of personal intuitions, mathematical procedures, and emerging theoretical structure comprising artificial intelligence research. State-of-the-art programs are far more narrowly specialized and inflexible than the corresponding aspects of human intelligence they emulate; however, in special domains they may be of comparable or greater power, e.g., in the solution of structure problems in organic chemistry or in the rigorous consideration of a large diagnostic knowledge base.

III.A.2.2. Resource Sharing

An equally important function of the SUMEX-AIM resource is an exploration of the use of computer communications as a means for interactions and sharing between geographically remote research groups engaged in biomedical computer science research and for the dissemination of AI technology. This facet of scientific interaction is becoming increasingly important with the explosion of complex information sources and the regional specialization of groups and facilities that might be shared by remote researchers [19, 5]. And, as projected earlier, we are seeing a growing decentralization of computing resources with the emerging technology in microelectronics and a correspondingly greater role for digital communications to facilitate scientific exchange.

Our community building effort is based upon the developing state of distributed computing and communications technology. While far from perfected, these capabilities offer highly desirable latitude for collaborative linkages, both within a given research project and among them. A number of the active projects on SUMEX are based upon the collaboration of computer and medical scientists at geographically separate institutions, separate both from each other and from the computer resource (see for example, the MENTOR and PathFinder projects).

In the early 1970's, the initial model for SUMEX-AIM as a centralized resource was based on the high cost of powerful computing facilities and the infeasibility of being able to duplicate them readily. As planned, this central role has already evolved significantly and continues to evolve with the introduction of more compact and inexpensive computing technology now available at many more research sites. At the same time, the number of active groups working on biomedical AI problems has grown and the established ones have increased in size. This has led to a growth in the demand for computing resources far beyond what SUMEX-AIM could reasonably and effectively provide on a national scale. We have actively supported efforts by the more mature AIM projects to develop or adapt additional computing facilities tailored to their particular needs and designed to free the main SUMEX resource for new, developing applications projects. To date, over 10 of the national projects have moved some or all of their work to local sites and several have begun resource communities of their own (see page 79). Thus, as more remotely available resources have become established, the balance of the use of the SUMEX-AIM resource has shifted toward supporting start-up pilot projects and the growing AI research community at Stanford.

III.A.2.3. Significance to Biomedicine

Artificial intelligence is the computer science of representations of symbolic knowledge and its use in symbolic inference and problem-solving processes. There is a certain inevitability to this branch of computer science and its applications, in particular, to medicine and biosciences. The cost of computers will continue to fall drastically during the coming two decades. As it does, many more of the practitioners of the world's professions will be persuaded to turn to economical automatic information processing for assistance in managing the increasing complexity of their daily tasks. They will find, from most of computer science, help only for those problems that have a mathematical or statistical core, or are of a routine data-processing nature. But such problems will be relatively rare, except in engineering and physical science. In medicine, biology, management, indeed in most of the world's work, the daily tasks are those requiring symbolic reasoning with detailed professional knowledge. The computers that will act as *intelligent assistants* for these professionals must be endowed with symbolic reasoning capabilities and knowledge.

The growth in medical knowledge has far surpassed the ability of a single practitioner to master it all, and the computer's superior information processing capacity thereby offers a natural appeal. Furthermore, the reasoning processes of medical experts are poorly understood; attempts to model expert decision-making necessarily require a degree of introspection and a structured experimentation that may, in turn, improve the quality of the physician's own clinical decisions, making them more reproducible and defensible. New insights that result may also allow us more adequately to teach medical students and house staff the techniques for reaching good decisions, rather than merely to offer a collection of facts which they must independently learn to utilize coherently.

The knowledge that must be used is a combination of factual knowledge and heuristic knowledge. The latter is especially hard to obtain and represent since the experts providing it are mostly unaware of the heuristic knowledge they are using. Medical and scientific communities currently face many widely-recognized problems relating to the rapid accumulation of knowledge, for example:

- codifying theoretical and heuristic knowledge
- effectively using the wealth of information implicitly available from textbooks, journal articles and other practitioners
- disseminating that knowledge beyond the intellectual centers where it is collected
- customizing the presentation of that knowledge to individual practitioners as well as customizing the application of the information to individual cases

We believe that computers are an inevitable technology for helping to overcome these problems. While recognizing the value of mathematical modeling, statistical classification, decision theory and other techniques, we believe that effective use of such methods depends on using them in conjunction with less formal knowledge, including contextual and strategic knowledge.

Artificial intelligence offers advantages for representing and using information that will allow physicians and scientists to use computers as intelligent assistants. In this way we envision a significant extension to the decision-making powers of specific practitioners without reducing the importance of those individuals in that process.

Knowledge is power, in the profession and in the intelligent agent. As we proceed to model expertise in medicine and its related sciences, we find that the power of our programs derives mainly from the knowledge that we are able to obtain from our collaborating practitioners, not from the sophistication of the inference processes we observe them using. Crucially, the knowledge that gives power is not merely the knowledge of the textbook, the lecture and the journal, but the knowledge of *good practice*--the experiential knowledge of *good judgment* and *good guessing*, the knowledge of the practitioner's art that is often used in lieu of facts and rigor. This heuristic knowledge is mostly private, even in the very public practice of science. It is almost never taught explicitly, is almost never discussed and critiqued among peers, and most often is not even in the moment-by-moment awareness of the practitioner.

Perhaps the the most expansive view of the significance of the work of the SUMEX-AIM community is that a methodology is emerging for the systematic explication, testing, dissemination, and teaching of the heuristic knowledge of medical practice and scientific performance. Perhaps it is less important that computer programs can be organized to use this knowledge than that the knowledge itself can be organized for the use of the human practitioners of today and tomorrow.

Evidence of the impact of SUMEX-AIM in promoting ideas such as these, and developing the pertinent specific techniques, has been the explosion of interest in medical artificial intelligence and the specific research efforts of the SUMEX community. In SUMEX's second decade, we have found that the small community of researchers that characterized the AIM field in the early 1970's has now grown to a large, accomplished, and respected research community. The American Association for Artificial Intelligence (AAAI), the principal scientific membership organization for the AI field, has 7000 members, over 1000 of whom are members of the medical special interest group known as the AAAI-M. This subgroup was founded by members of the SUMEX-AIM community who were active in AAAI and is the only active subgroup in the Association. The organization distributes semiannual newsletters on medical AI and provides a focus for cosponsoring relevant medical computing meetings with other societies (such as the American Association for Medical Systems and Informatics -- AAMSI). Medical AI papers are prominently featured at both medical computing and artificial intelligence meetings, and artificial intelligence is now routinely featured as a specific subtopic for specialized sessions at medical computing and other medical professional meetings. For example, members of the AIM community have represented the field to physicians at the American College of Pathology and American College of Physicians meetings for the last several years. A mere decade ago, the words "artificial intelligence" were never uttered at such conferences. The growing interest and recognition are largely due to the activities of the SUMEX-AIM community.

Another indication of the growing impact of the SUMEX-AIM community is its effect on medical education. For reasons such as those outlined above, there is an increasing recognition of the need for a revolution in the way medicine is taught and medical students organize and access information. Computing technology is routinely cited as part of this revolution, and artificial intelligence (and SUMEX-AIM research) generally figures prominently in such discussions. Such diverse organizations as the National Library of Medicine, the American College of Physicians, the Association of American Medical Colleges, and the Medical Library Association have all called for sweeping changes in medical education, increased educational use of computing technology, enhanced research in medical computer science, and career development for people working at the interface between medicine and computing; reports of all four organizations have specifically cited the role of artificial intelligence techniques in future medical practice and have used SUMEX-AIM programs as examples of where the technology is gradually heading.

In summary, the logic which mandates that artificial intelligence play a key role in enhancing knowledge management and access for biomedicine -- a logic in which we have long believed -- has gradually become evident to much of the biomedical community. We are encouraged by this increased recognition, but humbled by the realization of the significant research challenges that remain. Our goals are accordingly both scientific and educational. We continue to pursue the research objectives that have always guided SUMEX-AIM, but must also undertake educational efforts designed to inform the biomedical community of our results while cautioning it about the challenges remaining.

III.A.2.4. Summary of Current Goals

The following summarizes SUMEX-AIM resource objectives as stated in the proposal for the on-going five-year grant, begun on August 1, 1981, and provides the backdrop against which specific progress is reported. These project goals are presented in the three categories used in the previous proposal: 1) resource operations, 2) training and education, and 3) core research.

1) Resource Operations

- Maintain the vitality of the AIM community by continuing to encourage and explore new applications of AI to biomedical research and improving mechanisms for inter- and intra-group collaborations and communications. User projects will fund their own manpower and local needs; will actively contribute their special expertise to the SUMEX-AIM community; and will receive an allocation of computing resources under the control of the AIM management committees. There will be no "fee for service" charges for community members.

- Provide effective computational support for AIM community goals, including efforts to improve the support for artificial intelligence research and new applications work; to develop new computational tools to support more mature projects; and to facilitate testing and research dissemination of nearly operational programs. We will continue to operate and develop the existing KI-10/2020 facility as the nucleus of the resource. We will acquire additional equipment to meet developing community needs for more capacity, larger program address spaces, and improved interactive facilities. New computing hardware technologies becoming available now and in the next few years will play a key role in these developments and we expect to take the lead in this community for adapting these new tools to biomedical AI needs. We planned the phased purchase of two VAX computers to provide increased computing capacity and to support large address space LISP development, a 2 GByte file server to meet file storage needs, and a number of single-user "professional workstations" to experiment with improved human interfaces and AI program dissemination.

- Provide effective and geographically accessible communication facilities to the SUMEX-AIM community for remote collaborations, communications among distributed computing nodes, and experimental testing of AI programs. We will retain the current ARPANET and TYMNET connections for at least the near term and will actively explore other advantageous connections to new communications networks and to dedicated links.

2) Training and Education

- Provide community-wide support and work to make resource goals and AI programs known and available to appropriate medical scientists. Collaborating projects are responsible for the development and dissemination of their own AI programs.

- Provide documentation and assistance to interface users to resource facilities and programs and continue to exploit particular areas of expertise within the community for developing pilot efforts in new application areas.

- Allocate "collaborative linkage" funds to qualifying new and pilot projects to provide for communications and terminal support pending formal approval and funding of their projects. These funds are allocated in cooperation with the AIM Executive Committee reviews of prospective user projects.

- Support workshop activities, including collaboration with the Rutgers Computers in Biomedicine resource on the AIM community workshop and with individual projects for more specialized workshops covering specific application areas or program dissemination.

3) Core Research

- Explore basic artificial intelligence research issues and techniques, including knowledge acquisition, representation, and utilization; reasoning in the presence of uncertainty; strategy planning; and explanations of reasoning pathways, with particular emphasis on biomedical applications.
- Support community efforts to organize and generalize AI tools that have been developed in the context of individual application projects. This will include work to organize the present state-of-the-art in AI techniques through the AI Handbook effort and the development of practical software packages (e.g., AGE, EMYCIN, UNITS, and EXPERT) for the acquisition, representation, and utilization of knowledge in AI programs.

III.A.3. Details of Technical Progress

This progress summary covers only the resource nucleus. Objectives and progress for individual collaborating projects are discussed in their respective reports in Section IV. These collaborative projects collectively provide much of the scientific basis for SUMEX as a resource and our role in assisting them has been a continuation of that evolved in the past. Collaborating projects are autonomous in their management and provide their own manpower and expertise for the development and dissemination of their AI programs.

III.A.3.1. Progress Highlights

In this section we summarize highlights of SUMEX-AIM resource activities over the past 4 years, focusing on the resource nucleus.

- We have continued to recruit new user projects and collaborators to explore further biomedical areas for applying AI. A number of these projects are built around the communications network facilities we have assembled, bringing together medical and computer science collaborators from remote institutions and making their research programs available to still other remote users. At the same time we have encouraged older mature projects to build their own computing environments thereby freeing up SUMEX resources for newer projects. Nine projects now operate on their own facilities, including three that have become BRTP resources in their own right. Nine projects in the community have completed their research goals and their staffs have moved on to new areas.

- SUMEX user projects have made good progress in developing and disseminating effective consultative computer programs for biomedical research. These performance programs provide expertise in analytical biochemical analyses and syntheses, clinical diagnosis and decision-making, molecular biology, and various kinds of cognitive and affective psychological modeling. We have worked hard to meet their needs and are grateful for their expressed appreciation (see Section IV).

- We have made significant strategic improvements to the SUMEX-AIM computing environment in order to optimize computing support for the community. These developed in ways somewhat different from the initially projected plan. The DEC VAX computer did not prove to be an effective machine for running Lisp [23], while Lisp workstations have in fact become available from a number of vendors as tentatively expected at the time of our proposal (first Xerox, then Symbolics and LMI, and more recently Hewlett-Packard and Texas Instruments). Thus, rather than augmenting our mainframe resources with the purchase of large address space VAX's, we upgraded the KI-TENEX system to a DEC 2060 and at the same time, began moving aggressively toward a Lisp workstation-based research environment, with the approval of an ad hoc site visit group. We did secure VAX capabilities for our community by means of access to an 11/780 purchased under DARPA funding. We made an initial purchase of Xerox Dolphins with NIH funding and subsequently added more Xerox and Symbolics machines with NIH and DARPA funding and with industrial gifts. Because of the broad mix of research in the SUMEX-AIM community, no single workstation vendor can meet our needs so we have undertaken long-term support of a heterogeneous computing environment, incorporating many types of machines linked through multiprotocol Ethernet facilities.

- We have continued the dissemination of SUMEX-AIM technology through various media. We have distributed various AI software tools to many research laboratories, including over 200 combined copies of the GENET, EMYCIN, AGE, MRS, SACON, GLISP, and BB-1 systems. Several of our software systems have been adapted as commercial AI tools such as the Teknowledge S.1 and M.1 systems derived from EMYCIN, the Texas Instruments Personal Consultant system derived from EMYCIN, and the IntelliCorp KEE system derived from UNITS. We have also prepared video tapes of some of our research projects including ONCOCIN and an overview tape of Knowledge Systems Laboratory work.

- Our group has continued to publish actively on the results of our research including more than 45 research papers per year in the AI literature and a dozen books in the past 5 years on various aspects of SUMEX-AIM AI research (see page 81). These books have included the three-volume set of the *Handbook of Artificial Intelligence*, edited by Barr, Cohen, and Feigenbaum; a book on *Readings in Medical Artificial Intelligence: The First Decade* by Clancey and Shortliffe; and a book on *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* by Buchanan and Shortliffe.

- We completed the GENET project, begun in 1980 as a collaboration between the MOLGEN investigators and SUMEX, to make a set of DNA sequence analysis computing tools available to a national community of molecular biologists. This was an experiment in using a SUMEX-like resource to disseminate sophisticated software tools to a computer-naive community and proved extremely successful. GENET served over 300 molecular biologists before being phased out in early 1983. Subsequently, a new resource called BIONET has been funded by NIH at IntelliCorp to provide routine service of the type pioneered by SUMEX/GENET.

- A program in Medical Information Sciences was begun at Stanford in 1983 under Professor Shortliffe as Director. A group of faculty from the Medical School and the Computer Science Department argued that research in medical computing has historically been constrained by a lack of talented individuals who have a solid footing in both the medical and computer science fields. The specialized curriculum offered by the new program is intended to overcome the limitations of previous training options. It focusses on the development of a new generation of researchers with a commitment to developing new knowledge about optimal methods for developing practical computer-based solutions to biomedical needs. The feasibility of this program resulted in large part from the prior work and research computing environment provided by the SUMEX-AIM resource. Over 20 PhD and MS trainees will be enrolled in the fall of 1985. It has been awarded post-doctoral training support from the National Library of Medicine, received an equipment gift from Hewlett-Packard, and has received additional industrial and foundation grants for student support.

- We made significant progress in core AI research. In the area of knowledge representation, work was done on the representation of explicit strategy knowledge, temporal knowledge, causal knowledge, and knowledge in logic-based systems. In the area of architectures and control, we worked on a new implementation of a blackboard architecture with explicit control knowledge. Under knowledge acquisition studies, three PhD theses were completed covering experiments in learning by induction, by analogy, and learning from partial theories. In the area of knowledge utilization, results include work on reasoning with uncertainty and using counterfactual conditionals. We continued work on a number of existing tools for expert systems and on building new ones such as the BB1 system. And finally, significant work was done on the inference of user models, skeletal planning, defining a taxonomy of diagnostic methods, and reasoning with causal models.

- We have continued the core development of the SUMEX facility hardware, software, and networking systems to enhance the facilities available to researchers. Much of this work has centered on the effective integration of distributed computing resources in the form of mainframes, workstations, and servers. Network gateways and terminal interface machines based on MC-68000 microprocessors were developed to link our environment together and are now the standard system used in the campus-wide Stanford University network. We developed a gateway interface between Apple equipment (e.g., the Macintosh and Lisa) and EtherNet hosts that is now in wide use at universities around the country. We have developed many other software packages to enhance the computing environments of the Lisp workstations and to link them to other hosts and servers on our networks.

III.A.3.2. Resource Equipment Details

The SUMEX-AIM core facility, started in March 1974, was built around a Digital Equipment Corporation (DEC) KI-10 computer and the TENEX operating system which was extended locally to support a dual processor configuration. Because of the operational load on the KI-10's, in the late 1970's, we had added a small DEC 2020 system (see Figure 2) to support more dedicated testing of systems like ONCOCIN and Caduceus and for community demos. This facility provided a superb base for the AI mission of SUMEX-AIM through 1982. Its interactive computing environment, its AI program development tools, and its network and interpersonal communication media were unsurpassed in other machine environments. Biomedical scientists found SUMEX easy to use in exploring applications of developing artificial intelligence programs for their own work and in stimulating more effective scientific exchanges with colleagues across the country. Coupled through wide-reaching network facilities, these tools also give us access to a large computer science research community, including active artificial intelligence and system development research groups.

The Heterogeneous Computing Environment

In the renewal for the current grant period, both an augmentation of the central resource in terms of address space and capacity and exploratory work with Lisp workstations were planned. The Initial Review Group recognized in their special study section report the importance of optimizing the timing of our planned hardware acquisitions to coordinate community needs with the availability of important technological developments in vendor-supported systems. They recommended in their report that we be allowed considerable flexibility as to phasing of equipment purchases within the 5-year renewal period.

We had initially planned to purchase a large VAX in 1981 and later, our first Lisp workstations. However, we speeded our push toward workstations for several reasons. The state of VAX Lisp implementations and projections of their performance were very discouraging (a study of the VAX InterLisp implementation was done at the time as documented in [23]). And the first Xerox InterLisp Dolphin workstations were available for delivery after the summer of 1981. These machines were the prototypes on which research toward adapting expert AI systems for the interactive workstation environment could begin. So, we purchased 5 Dolphins for the fall of 1981 and, in order not to delay non-Lisp SUMEX-AIM work involving VAX machines, we were able to arrange shared access to a VAX 11/780 funded by ARPA to support Heuristic Programming Project research. One of the Dolphins we purchased was loaned for several years to the Rutgers Computers in Biomedicine resource for experimental work.

We continued to evaluate strategies and alternatives for planned system configuration development. In particular, we had a chance to gain experience with the Dolphin InterLisp machines and the shared VAX, reassess the role of the dual KI-TENEX system, and reach a consensus about what the long term configuration of the SUMEX-AIM facility should be. This was validated by an ad hoc study section review in 1982. In summary, it was decided that the best resource configuration for the coming decade would be a shared central machine coupled through a high-performance network to growing clusters of personal workstations. The central machine should be an extended addressing TOPS-20 machine and the workstations will be chosen from the viable products available and scheduled for announcement.

The concept of the individual workstation, especially with the high-bandwidth graphics interface, proved ideal. Both program development tools and facilities for expert system user interactions were substantially improved over what is possible with a central time-shared system. The main shortcomings of these systems were their processing speed and cost, but the prospect of other workstations to be available from Xerox, Symbolics, LMI, HP, and others reassured us that these were the right choices for AI system in the long term. Still, at the time, it was not possible to equip very much of the SUMEX-AIM community with individual workstations.

Upgrade of the KI-10's to a 2060

Meanwhile, on the mainframe front, given the continued need for a central machine, the poor Lisp performance of the VAX, and the increasingly untenable difficulties in maintaining the KI-TENEX system, we decided it is time to retire the KI-10's and upgrade them to the then (1982) more modern DEC 2060 TOPS-20 system. This would free our systems staff to concentrate on more productive development efforts for the community such as work related to professional workstations and compatible Lisp support. The 2060 had a processing capacity of 2-3 times that of the dual KI-TENEX system, badly needed for our community, and it was more compact, reliable, and maintainable. Pending the arrival of more cost-effective and generally-available Lisp workstations, this would allow us to continue support for the SUMEX-AIM community at large and to provide facilities for new AI efforts.

In late 1982, we implemented the upgrade. The purchase price of the DECsystem 2060 reflected a substantial price reduction based on an external research grant from Digital Equipment Corporation to the Heuristic Programming Project in exchange for access by DEC to the AI software systems and knowledge-based systems expertise developed by the HPP. The remainder of the system was funded jointly by NIH and DARPA. The system configuration is shown in Figure 1. Of course, the transfer of service required a substantial investment of hardware engineering effort as all of the local line and network connections had to be changed over. This was all effected invisibly to the user community by running the old KI-TENEX and the new 2060 systems in parallel for more than a month.

Using DARPA funding, we also made some upgrades to the shared VAX 11/780 which was initially purchased by ARPA for HPP research as well as work in network graphics and VLSI design. The configuration of this machine is shown in Figure 3. In 1983, we augmented the machine by adding 2 Mbytes of memory and expanding the file system with a DEC RP07 disk drive (512 Mbytes). Approximately 60% of the machine is allocated for HPP and SUMEX use.

The overall facility model then became the central shared 2060, 2020, and VAX 11/780 systems surrounded with growing numbers of workstations and intercoupled by a local area network.

Additional Workstations

After the purchase of the 5 experimental Dolphin workstations, much work went into their development by Xerox, based on feedback and interactions with groups such as ours using them for AI applications. Performance of the Dolphins improved substantially based largely on improved microcoding of frequently used primitives and facilities. The initial optimizations of the Dolphin microcode were based on work at Xerox observing their own programs running. When the Dolphin was exposed to other AI systems such as ours, it became clear that additional improvements were necessary and were implemented, including enhanced performance for CONS operations, function calls, disk management, garbage collection, and other areas. Improvements in individual areas of performance ranged from factors of 2 to 10.

By 1983, other contenders were entering the Lisp workstation market in addition to Xerox. Because work in the HPP and the SUMEX-AIM community draws heavily on both Interlisp and the derivatives of MIT's MacLisp, we broadened our workstation experiments into both areas.

With NIH funding in 1983, we purchased 6 Xerox 1108 workstations (Dandelions) and in 1984, 3 Xerox 1109's (DandeTigers). With DARPA funding we purchased 2 Xerox 1108's and 1 1132 (high-performance Dorado) in 1984. In early 1985, the ONCOCIN group received a grant from Xerox of 13 1108's and additional printing and file server equipment. These machines represent the second generation of Xerox Lisp workstations and include significantly higher performance and functionality.

With DARPA funding in 1983 we bought a Symbolics LM-2 running the ZetaLisp system. In 1984, we added 3 Symbolics 3600's and a 3670 and in early 1985, another 3670 -- all with DARPA funding. We are also planning the purchase of additional workstations in the near term with DARPA funding.

Local Area Network Server Hardware

Since the late 1970's, we have been developing a local, high-speed Ethernet environment to provide a flexible basis for planned facility developments and the interconnection of a heterogeneous hardware environment. Our development of Ethernet facilities has been guided by the goals of providing the most effective range of services for SUMEX community needs while remaining compatible with and able to contribute to and draw upon network developments by other groups, dating back to the early 3 Mbit/sec Ethernet given to Stanford and several other universities by Xerox. We now support both 3 and 10 Mbit/sec Ethernets (see Figure 5) running numerous protocols and extended geographically throughout the SUMEX-AIM and related Stanford research groups. This network is the "glue" that holds the rest of the computing environment together and consists of numerous servers such as gateways and servers for terminal access, file storage and retrieval, and laser printing.

In the early phases, a substantial amount of special hardware was developed by our group for network interfaces including a high-performance direct memory access interface for the dual KI-TENEX system and a serial phase decoded UNIBUS interface that are used on our DEC 2020, VAX's, and early PDP-11 gateways and TIP's. The KI Ethernet interface served well for a period until we upgraded the system to a 2060, at which time we installed the 2060 mass bus EtherNet interface designed and built by the Stanford Computer Science Department. Our KI-10 interface is still seeing service in connecting another KI-10 system (Institute for Mathematical Studies in the Social Sciences) to the net.

Hardware for Gateways and TIP's

As we evolved a more complex network topology and decided to compartmentalize the overall Stanford internet to avoid electrical interactions during development and to facilitate different administrative conventions for the use of the various networks, we developed gateways to couple subnetworks together. These first used PDP-11/05 hardware and then Motorola MC-68000 systems as they became available.

Similarly, we designed gateway between Apple equipment such as the new Macintosh terminal, that may play a role in our future virtual graphics work, and EtherNet using a MC-68000 gateway and a locally-designed Apple Bus to Multibus EtherNet interface. This system incorporates an 8530 Zilog chip to communicate with the Apple Net and software to manage the protocol packaging.

We also developed a MC-68000 terminal interface processor (TIP) to provide terminal access to network hosts and facilities. It is basically a machine that has a number of terminal lines and a network interface and software to manage the establishment of connections for each line and the flow of characters between the terminal and host. It can handle up to 32 lines. Both of these systems are now widely used throughout the Stanford network.

File Server Hardware

The development of an EtherNet file server was an integral part of our council-approved equipment plan with further expansions approved for later years. With joint NIH and DARPA funding, we were able to take advantage of an exceptional offer by Digital Equipment Corporation, through their corporate external research sponsorship program to DARPA contractors (the HPP), to purchase two VAX 11/750 machines as the processor part of our file servers. In the initial file server configurations, we also bought Fujitsu Eagle 450 MByte disks and controllers (one each from Systems Industries and Emulex) with one 800/1600 BPI tape unit for long term archives, and one 300 Mbyte removable pack drive for cyclic backups.

Other Network Hardware

We have developed numerous local network connection systems that have taken advantage of existing cabling rather than invest in expensive trenching and recabling. For example, in The Heuristic Programing Project (HPP) move to 701 Welch road, a high-performance network link to other SUMEX and campus network facilities was essential. Several communication schemes for establishing a reliable and relatively fast link were considered, including microwave, infrared laser, direct ethernet (by trenching and placing a direct ethernet cable), telephone company T1 service and others. All of these would have involved high cost and so we developed a communication link using bare copper telephone pair already in place. The wire distance between the HPP Welch Road location and the SUMEX machine room in the Medical Center is approximately 2000 ft. Utilizing high capacity differential drivers and ultra high speed, high sensitivity receivers, a half-duplex transceiver was developed for plain copper twisted pair that achieved error-free transmission at 1.25 Mbits/sec in each direction, utilizing Manchester data encoding. This communication link has been in operation for well over a year now without any appreciable down time or noticeable error rate or data delays.

In addition to the normal continuous flow of maintenance problems, we have reconnected the very reliable line printer from the old KI-TENEX system to the 2060. This required substantial modification of the printer controller to adapt to the different 2060 bus signal standards. We have also installed lots of communications equipment, including dial-in and -out modems and laser printer connections.

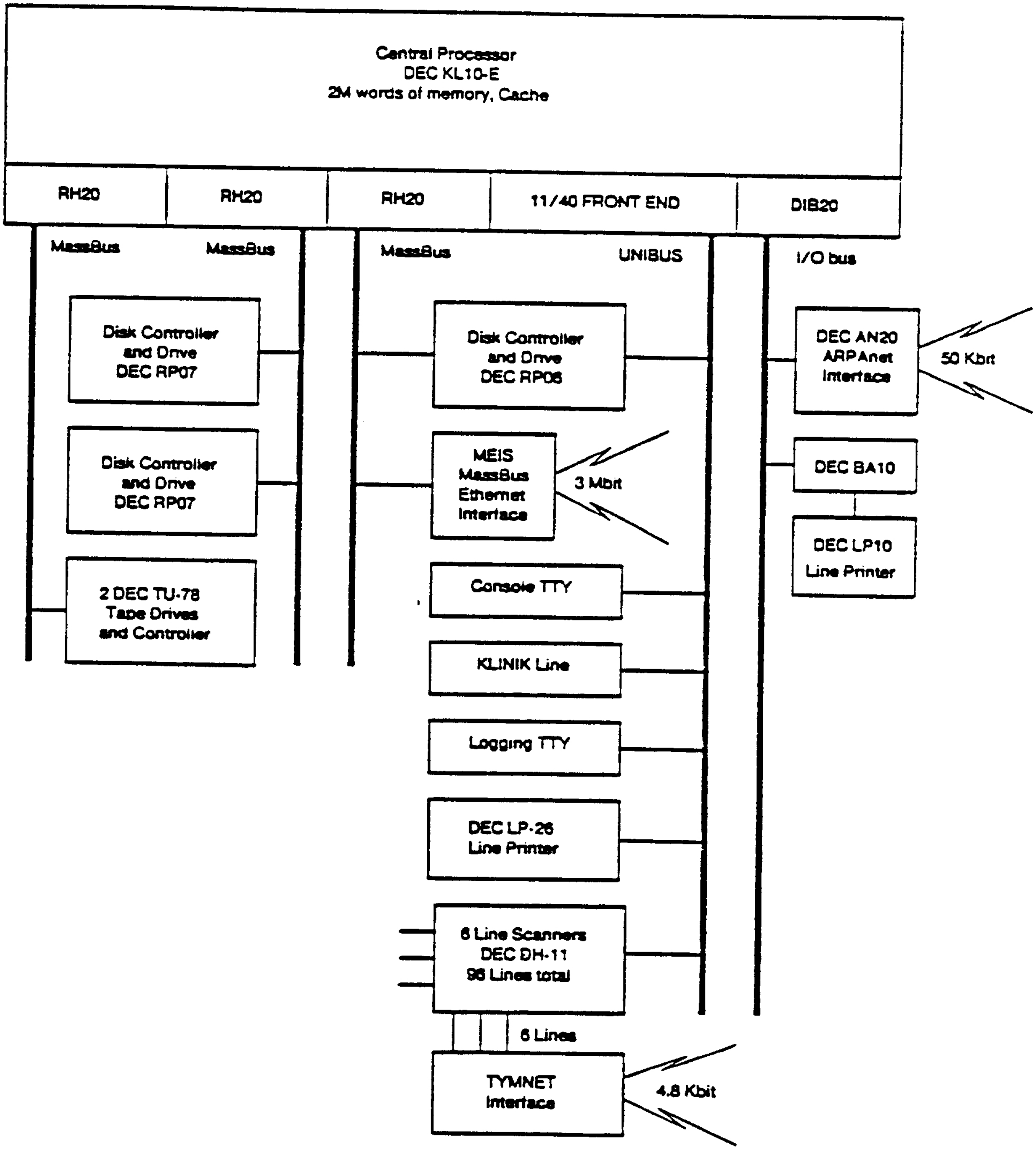

Figure 1: SUMEX-AIM DEC 2060 Configuration