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 14, our training activities are summarized starting on page 53, and the progress of our collaborating projects is detailed starting on page 93.

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 93.

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 56.

III. Narrative Description

III.A. Summary of Research Progress

III.A.1. Overview

It is now thirteen years since the SUMEX-AIM resource was established in 1973 and before discussing the details of our progress this past year, we take this opportunity to reflect on the broad progress of SUMEX as a resource. Computing and communications technologies and biomedical artificial intelligence (AI) research have achieved remarkable results. The SUMEX-AIM resource has both profoundly influenced and responded to those changing technologies. It is widely recognized that our resource has fostered highly influential work in biomedical AI -- work from which much of the field of expert systems has emerged -- and that it has simultaneously helped define the technological base of applied AI research.

SUMEX has been 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.

SUMEX 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. 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. Interest in AI research and application continues to grow. 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. Nevertheless, 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.

Even as we reflect on this substantial progress, however, at the deepest research level 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 needs for more and more powerful 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. This has been the major focus of our core system research this past year.

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, 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, 6, 7, 8]. 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, users supply their intelligence to instruct the machine precisely *how* to do the job, step-by-step.

At the other extreme of the spectrum, users describe *what* they wish the computer to do for them to solve problems. They want 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 they are addressing an intelligent agent that is using knowledge of their world to understand their intent, complain or fill in their vagueness, make specific their abstractions, correct their errors, discover appropriate subgoals, and ultimately translate what they want done into detailed processing steps that define how it should be done by a real computer. Users want to provide this specification of what to do in a language that is comfortable to them and the problem domain (perhaps English) and via communication modes that are convenient (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 enabled a long, interdisciplinary line of artificial intelligence research at Stanford concerned with the development of concepts and techniques for building expert systems [3]. 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 B and more detailed progress summaries in Section IV. 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 experimental programs, 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 [5, 2]. And, as projected, 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 powerful tools 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. This central role has already evolved significantly and continues to change 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 9 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 87). 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.

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.

E. H. Shortliffe

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. mainframe 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.

- 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

E. H. Shortliffe   12

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.

E. H. Shortliffe

III.A.3. Details of Technical Progress

This progress summary covers the nucleus of the SUMEX-AIM resource. 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 year (May 1985 - April 1986), focusing on the resource nucleus.

- We have made additional significant improvements to the SUMEX-AIM computing environment in order to optimize computing support for the community. These include the addition of 18 Xerox 1186 workstations, 20 Texas Instruments Explorer workstations, and 4 Symbolics workstations. The purchase of these Lisp machines was funded jointly by NIH, DARPA, and machine vendor gifts. We continue to operate the mainframe computers (DEC 2060, 2020, VAX 11/780, and VAX 11/750's) for the community. Because of the broad mix of research in the SUMEX-AIM community, no single computer 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 core development of the SUMEX software tools 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 have been enhanced with new protocol and internet routing capabilities. We have developed many new software packages to enhance the computing environments of the Lisp workstations and to link them to other hosts and servers on our networks.

- We have continued the dissemination of SUMEX-AIM technology through various media. We have distributed 73 copies of our AI software tools (EMYCIN, AGE, MRS, SACON, and BB1) to academic, industrial, and federal research laboratories. We have also continued to distribute the video tapes of some of our research projects including ONCOCIN, and an overview tape of Knowledge Systems Laboratory work to outside groups. The KSL overview tape won a CINE Eagle award for excellence and a recommendation to represent U.S. scientific efforts at events abroad.

- 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 89).

- The Medical Information Sciences program, begun at Stanford in 1983 under Professor Shortliffe as Director, has continued to grow over the past year to include about 25 outstanding PhD and MS students, and a search for an additional faculty member has been authorized and is underway. The specialized curriculum offered by the MIS program focuses on the

development of a new generation of researchers able to support the development of improved 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 equipment gifts from Xerox and 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, further 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, work continues on 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.

- 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.

- At the end of the current reporting period, we are actively planning the move of the SUMEX and Medical Computer Science offices into newly constructed Stanford Medical School office space, funded by the university. This space, in the Stanford Medical Center complex, provides us with almost twice the area we previously occupied and it is laid out so as to promote better interactions between out groups and among our students and research staff. The move will take place in June.

- 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).

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 1981, when, using DARPA funding, we added a VAX 11/780 system running the UNIX operating system (see Figure 3). By 1982, the KI-10's were becoming difficult to maintain, both in terms of hardware and software, and so were upgraded to a DEC-supported 2060 with about twice the capacity (see Figure 1). The interactive computing environment of this facility, with its AI program development tools and its network and interpersonal communication media, was 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 late 1970's and early 1980's, computer system research on early microprocessors and compact minicomputers suggested that large mainframe computers would not be essential or even the dominant source of computing power for AI research and AI program dissemination. Thus, we began to implement a strategy for computing resources marked by the integration of heterogeneous systems -- mainframes, Lisp workstations, and service systems (e.g., for file storage and printing) all linked together by local area networks. Since the purchase of the first Xerox InterLisp Dolphin workstations in the summer of 1981, many more workstation products have come on the market with significant improvements both in performance and lower cost. Thus, over the years, we have configured the optimal resource computing environment around shared central machines coupled through a high-performance network to growing clusters of personal workstations.

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 early workstation systems were their limited processing speed and high cost. But in the few years since our first experimental systems, processing power has increased by a factor of 10 (e.g., in Texas Instruments Explorers, Symbolics 3600's, and SUN workstations) and cost has decreased by a factor of 3-4 (e.g., in Xerox 1186's). As a concrete overall system example, SUMEX was among the first sites to receive the new Xerox 1186 (DayBreak) last fall. Each one costs less (70% less) than a Dolphin purchased in 1981. Each is roughly 4 times as fast as a Dolphin, has a larger display (20% more pixels), larger disk (38% more space), more memory (245% more), and is small enough and cool enough to fit in a private office or student carrel, eliminating the need for an umbilical cord to a machine-room or floor area therein.

Lisp Workstations

Work in the SUMEX-AIM community and the KSL draws heavily on both of the major dialects of Lisp, Interlisp and the derivatives of MIT's MacLisp. Thus, our workstation purchases have included machines for both environments. We have added new

workstations paced carefully with the developments of higher performing, more compact, and lower cost systems (see Figure 5). By early 1985, we had acquired with NIH funds, DARPA funds, and industrial gifts, the following workstations: 5 Xerox 1100's (Dolphins), 21 Xerox 1108's (Dandelions), 3 Xerox 1109's (DandeTigers), 1 Xerox 1132 (Dorado), 1 Symbolics LM-2, 4 Symbolics 3600's, and 2 Symbolics 3670's.

In late 1985, after long evaluation and vendor negotiations, we acquired a large number of additional workstations including 20 Xerox 1186's (DayBreaks), 20 Texas Instrument Explorers, 1 Symbolics 3600, 1 Symbolics 3640, and 2 Symbolics 3645's. These were purchased mostly with DARPA funding and vendor gifts, and also include two additional drives for the Xerox file server donated last year special software environments for the workstations. These workstations have been broadly integrated into our faculty, staff, and student offices and into public work areas. They are used to support all of our research projects and the gift machines also support various courses offered at Stanford in AI.

We continue to evaluate Lisp workstations as the technology is changing rapidly. Systems based on the SUN Microsystems workstation and the IBM PC-RT have benchmark data rivaling the performance of other specially microcoded Lisp machines (e.g., those from Xerox, Symbolics, and TI), but the software environments are not nearly so extensively developed yet.

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. We now support primarily 10 Mbit/sec Ethernets (see Figure 6) 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.

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 using Motorola MC-68000 systems.

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

Because our Lisp workstations have only limited local file space, the development of effective shared file servers is essential to our resource operation. We had previously implemented two file servers, based on DEC VAX 11/750 machines purchased through a special price arrangement with DARPA (see Figure 4). 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. Because of problems with the SI controller, we replaced it this past year with an Emulex so that the two systems are now identical. We also purchased another 450 MByte disk drive to provide needed capacity expansion. Finally, we are in the process of purchasing a SUN 3 file server with two 450 MByte disks using DARPA funds in order to increase file server performance and capacity.

Other Network Hardware

Over this past year, Stanford University has undertaken a major effort to install cabling throughout campus as part of a Stanford University Network (SUNet) development and installation of a new telephone system. These installations have helped improve the connectivity and performance of our network, including redundant links between the new SUMEX and MCS space in the Medical School Office Building and the SUMEX machine room and other parts of campus.
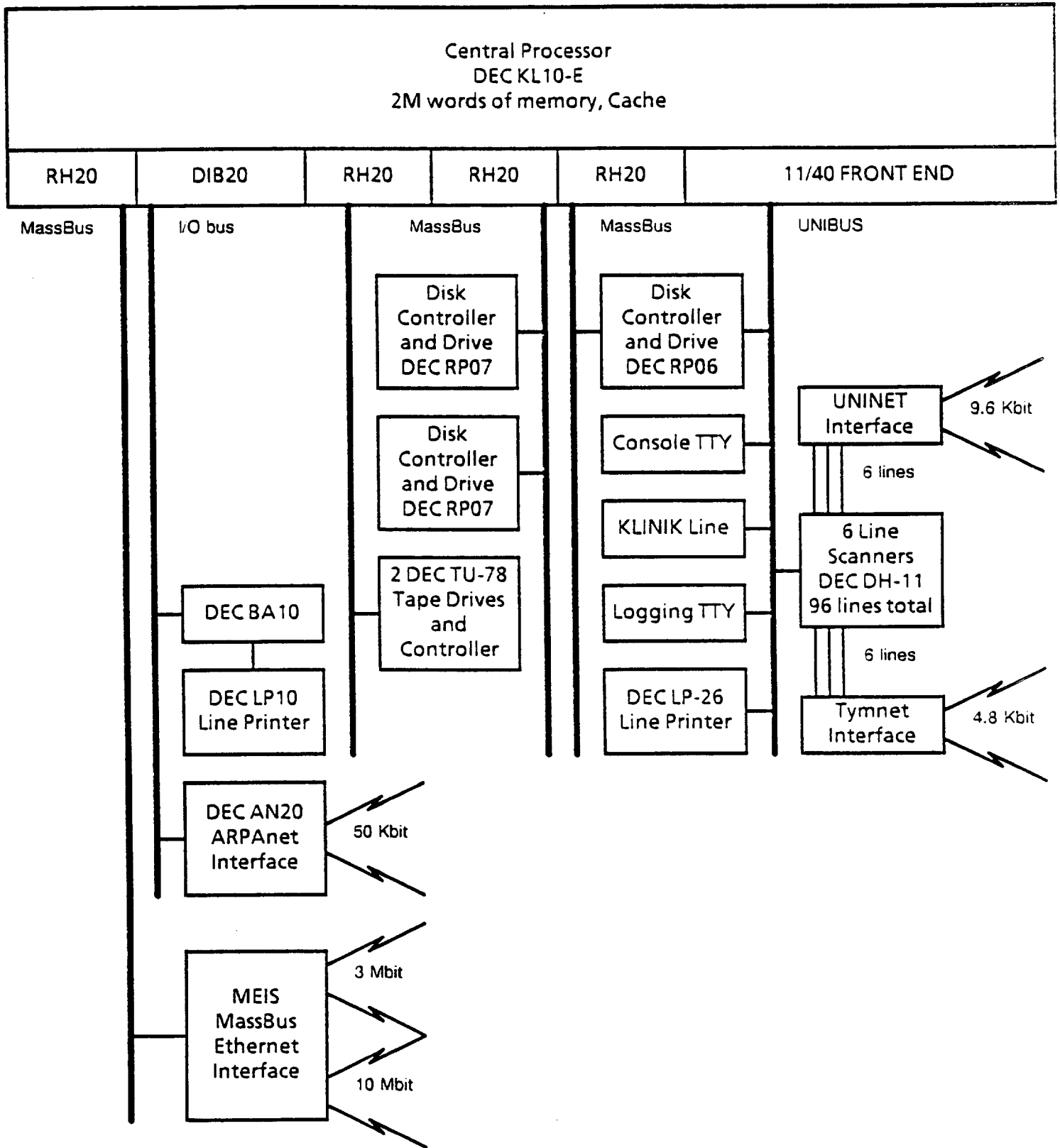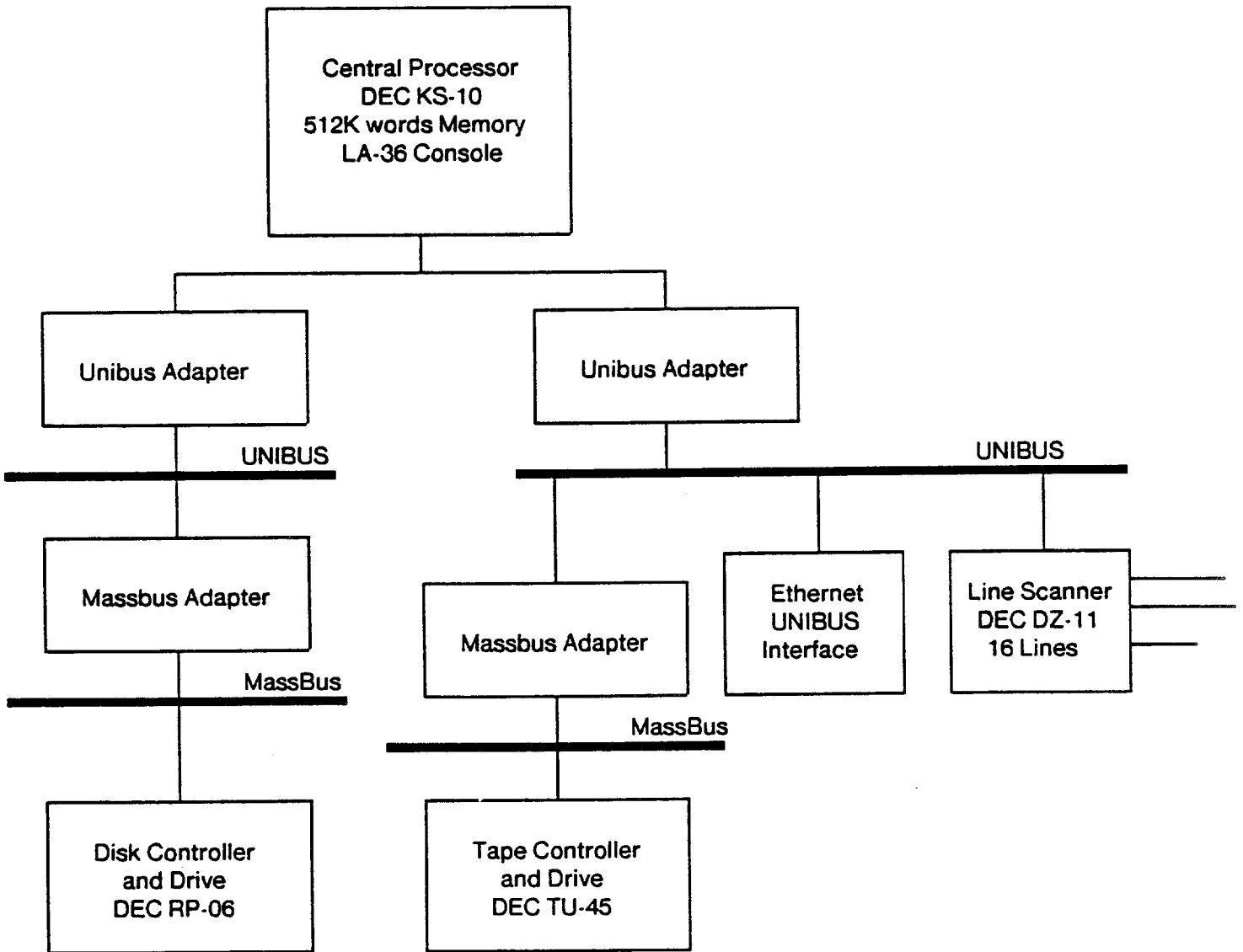
Figure 1:   SUMEX-AIM DEC 2060 Configuration

E. H. Shortliffe

Figure 2: SUMEX-AIM DEC 2020 Configuration
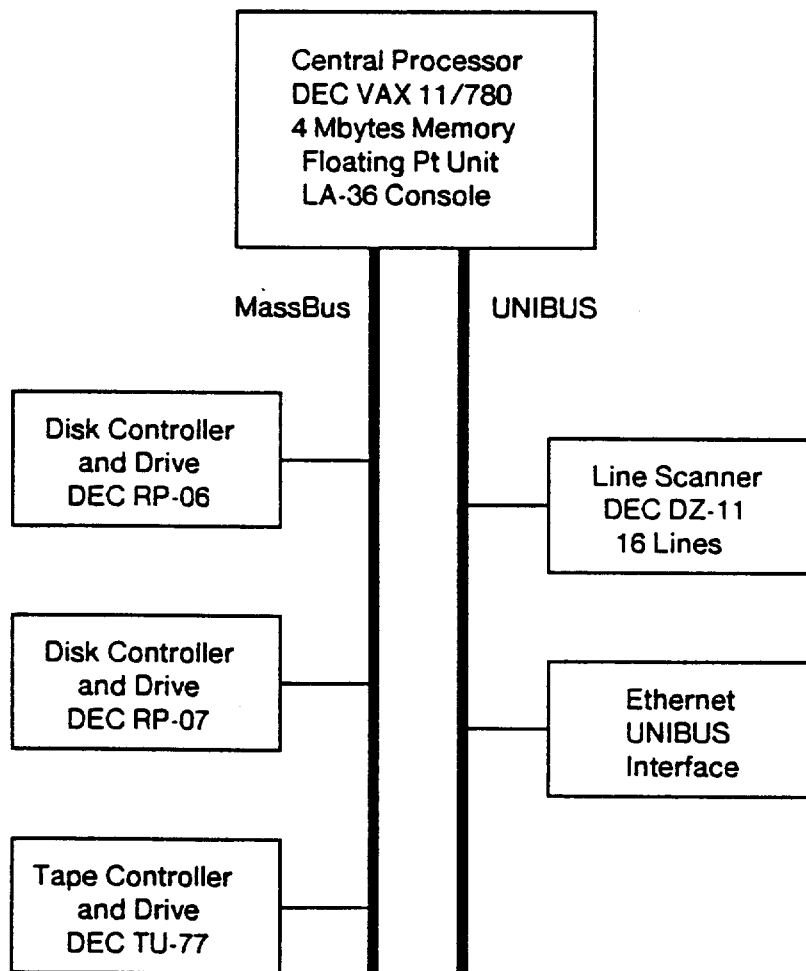
Figure 3: SUMEX-AIM Shared DEC VAX 11/780 Configuration

E. H. Shortliffe
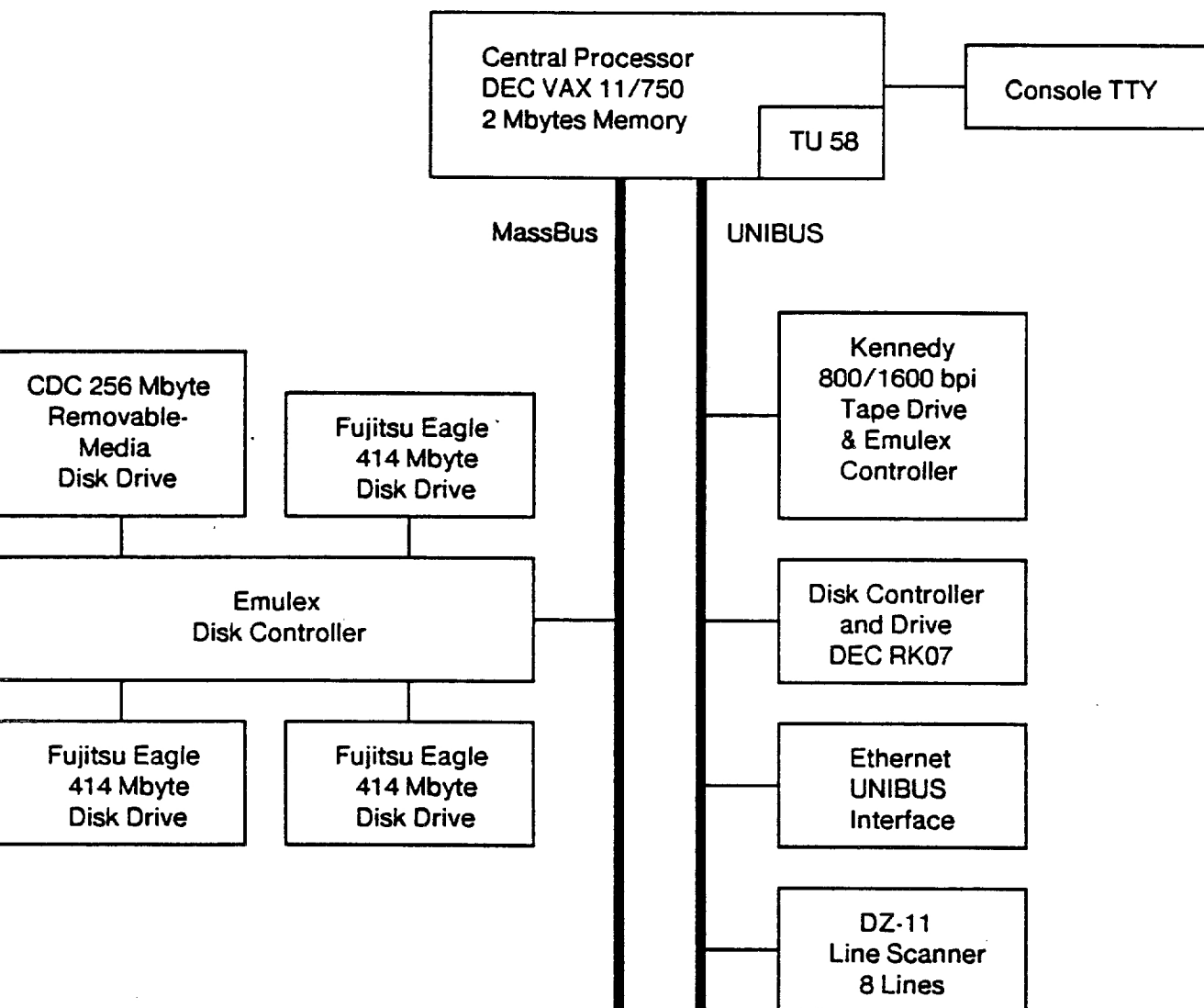
Figure 4:   SUMEX-AIM File Server Configuration
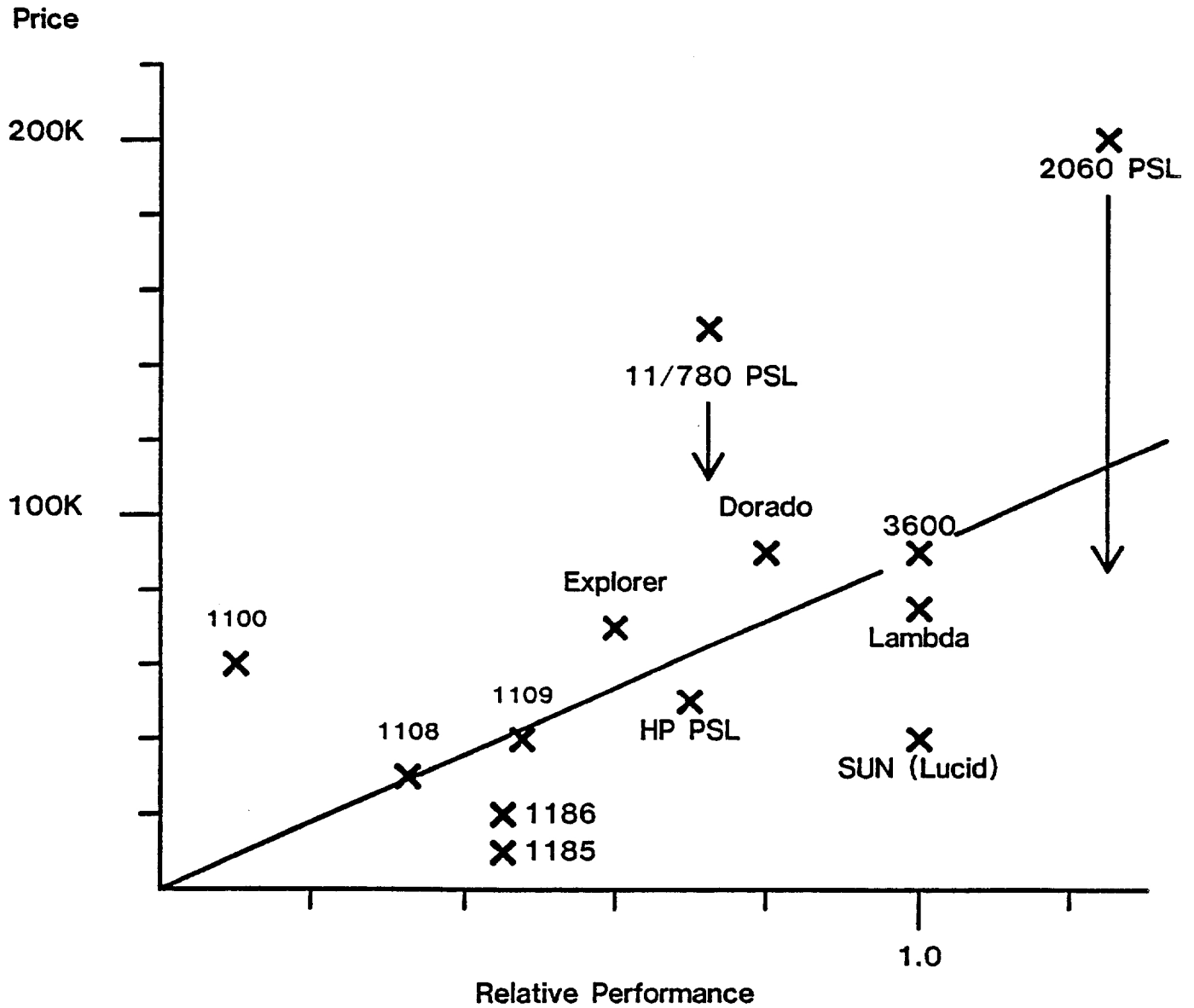
Figure 5: Price/Performance Comparison of Lisp Workstations

Margaret Jacks Hall
Logic Group

SCORE 2060
Symbolics LispM's
Xerox D-machines
Ether TIP
Xerox laser printers
Other CSD Equipment

Electrical
Engineering

G

R

Pine Hall

CSLI

G

G

SUMEX VAX-780

Medical School Office Building
Symbolic Systems Resources Group
Medical Computer Science Group

Xerox D-machines   Applenet
H-P 9836's
TI Explorer
Imagen laser printers   G
Xerox laser printer
SUN - Devel
Ether TIPs

R

G

Medical Center
SUMEX Machine Room

Med. A   Med. B

R

Med. C   Nozaki   Fairchild

R

Unk.2nd   Radiology Unk.3rd

R

SUMEX 2060
SUMEX 2020
Xerox D-machine
VAX-750 file server
Xerox NS file server
Imagen laser printer
Ether TIP

Xerox D-machines

R

G

R   Repeater

G   Gateway

1.25 Mbit
phone line

Whelan Building (Welch Road)
HPP and HELIX

10 Mbit ethernet

Xerox D-machines
TI Explorers
Symbolics LispM's
Silicon Graphics Iris
VAX-750 file server
Ether TIPs
Imagen laser printers
Apple laser printer

G

Xerox Alto
Xerox D-machines
Xerox laser printer
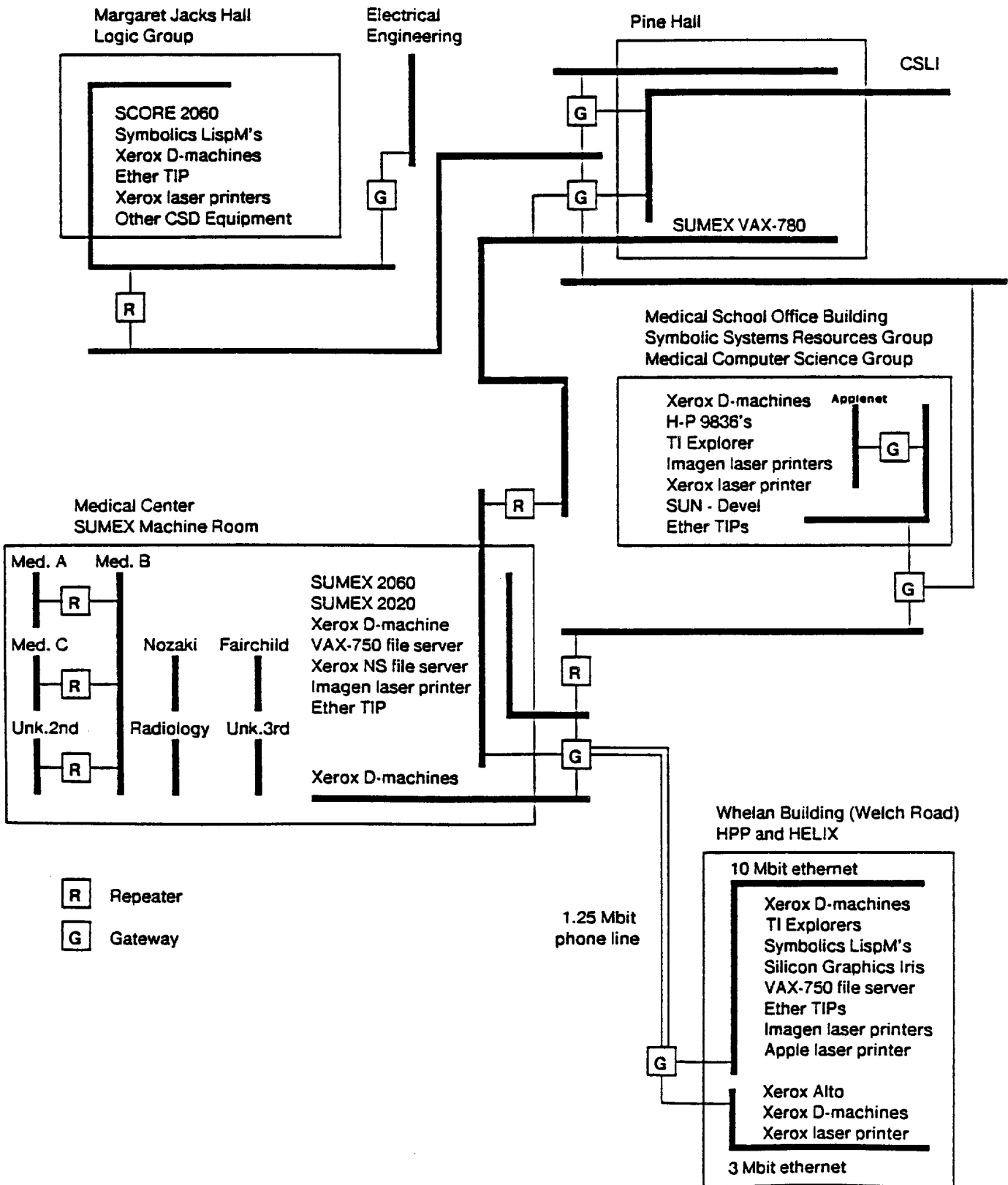
3 Mbit ethernet

Figure 6:   SUMEX-AIM EtherNet Configuration

III.A.3.3. Core System Development

Operating System Software

The various parts of the SUMEX-AIM computing environment require development and support of the operating systems that provide the interface between user software and the raw computing capacity. In addition to performance and relevance to AI research, much of our strategy for hardware selection has been based on being able to share development of the operating systems among a large computer science community. This includes the mainframe systems (TOPS-20 and UNIX) and the workstation systems. Following are some highlights of recent system software developments.

TOPS-20 Development and Support

With our long term plan to phase out the 2060 mainframe system, our development efforts in that area are beginning to wind down. Nevertheless, over the past year, considerable work was required to keep the TOPS-20 systems running effectively for the community. This has included the periodic updating, checkout, and installation of new versions of system software. An important upgrade involved moving from the previous 5.3 TOPS-20 monitor release to 6.1. This required a large effort to incorporate all the local SUMEX changes, not only in the monitor itself but in other service software like the EXEC, Galaxy and its spoolers, CHECKD, and ACJ.

Other activities included installation and checkout of the new MCA25 Cache which we purchased the previous grant year to enhance performance by doubling the size of cache memory, the addition of a two-way associative page table, and the addition of a "keep" bit to retain frequently accessed executive page table entries. Several problems occurred in the installation requiring detailed analysis. SUMEX staff pointed out a bug in DEC's MCA25 diagnostic resulting in a detailed reexamination of the installation which uncovered the omission of one of the backplane wire additions. After correcting the installation, the MCA25 cache worked well as verified by timing tests.

We continued to develop and improve our laser printer spooler, IMPSPL. This included adding many new features such as variable typesizes, page reversing, and manual paper feed for printing on high quality paper. This also required changes to the EXEC and Galaxy to support the new features. IMPSPL appears to be quite solid now, having run for almost a year without any problems.

A significant effort was required "tending the system". In the past year, we analyzed approximately 75 memory dumps of the TOPS-20 operating system in order to pinpoint hardware and software problems and bugs. To facilitate this, we continued the development of QANAL, a Quick automatic dump ANALyze program. Briefly, this program reads the dump and a copy of the monitor file that was running when the failure occurred, and produces a report detailing the state of the system at the time of the failure. Included among these reports is a SYSTAT type output, giving detail to the individual fork, or process, level of all jobs logged in. During the past year, we upgraded QANAL to analyze TOPS-20 release 6.1 data structures, as well as adding several other refinements.

We have continued to track network protocol and service (e.g., file transfer and electronic mail) developments. Many monitor bugs related to TOPS-20 implementation problems relating to the DOD IP/TCP protocols. This complex software required significant effort on our part because SUMEX-AIM has become a major communications crossroads and so exercises the network code very heavily. This has raised many bugs and performance problems that we have worked to improve. We have played an active role in network discussion groups related to areas such as electronic