I   Narrative Description

This is an annual report for the Stanford University Medical
EXperimental computer resource for applications of Artificial Intelligence
in Medicine (SUMEX-AIM). It covers the period between May 1, 1981 and
April 30, 1982.

This is the first year of a 5-year renewal of the SUMEX resource
grant. It begins an important and exciting new phase for SUMEX-AIM
community research. Successes in developing expert systems, many of them
stemming from projects in the SUMEX-AIM community, have stimulated very
strong and growing interest in AI research from many popular and industrial
fronts. At the same time, the on-going revolution in computational tools,
made possible by larger and larger scale microelectronic integration, is
making routine applications of AI systems more practical and effective and
is providing tools for new avenues of research. Our approved project goals
focus principally on a merging of state-of-the-art community research in
biomedical AI applications with these new computing tools and on the
challenges they will bring to the SUMEX-AIM community and resource. We
expect that the integration and exploitation of these emerging computer
technologies that will have a profound effect on the development and
dissemination of practical biomedical AI systems.

The earlier phases of the SUMEX-AIM resource were characterized by
the building of a national community of biomedical AI collaborators around
a central resource located at Stanford University. Beginning with 5
projects in 1973, the AIM community grew to 11 major projects at our
renewal in 1978 and currently numbers 15 fully authorized projects plus a
group of 4 pilot efforts. Many of the computer programs under development
by these groups are maturing into tools increasingly useful to the
respective research communities. The demand for production-level use of
these programs has surpassed the capacity of the present SUMEX facility and
has raised important issues of how such software systems can be optimized
for production environments, exported, and maintained.

We continue to seek interesting new AI applications in our community
of biomedical and computer scientists interacting through electronic media.
However, we expect the SUMEX-AIM community to develop a somewhat different
character in the coming years. It will become more decentralized in terms
of computing resources, more diverse in scope, and even more heavily
dependent on network communication facilities for interactions,
collaborations, and sharing.

The following sections report on the activities of the SUMEX-AIM
resource this past year including brief summaries of the objectives of
SUMEX-AIM, a characterization of biomedical AI research, resource
organization and operating procedures, recent core progress in system
development and basic AI research, and progress in the collaborative
projects.

I.A   Summary of Research Progress

I.A.1   Overview of Objectives and Rationale

SUMEX-AIM ("SUMEX") is a national computer resource with a dual mission: 1) promoting applications of computer science research in artificial intelligence (AI) to biological and medical problems and 2) demonstrating computer resource sharing within a national community of health research projects. The central SUMEX-AIM facility 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.

I.A.1.1   What is Artificial Intelligence

Artificial Intelligence research is that part of Computer Science concerned with symbol manipulation processes that produce intelligent action [1 - 7]. By "intelligent action" is meant 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 ways in which people use computers to accomplish tasks can be "one-dimensionalized" into a spectrum representing the nature of the instructions that must be given the computer to do its job; call it the What-to-How spectrum. At the How extreme of the spectrum, the user supplies his intelligence to instruct the machine precisely how to do his job, step-by-step. 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.

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. Still, he demands 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 shall be done by a real computer. The user wants to provide this specification of What to do in a

E. A. Feigenbaum   2

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).

The research activity aimed at creating computer programs that act as "intelligent agents" near the What end of the What-to-How spectrum can be viewed as a long-range goal of AI research.

Expert Systems and Applications

The national SUMEX-AIM resource is an outgrowth of a long, interdisciplinary line of artificial intelligence research at Stanford concerned with the development of concepts and techniques for building "expert systems" [1]. 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.

Currently authorized projects in the SUMEX community are concerned in some way with the application of AI to biomedical research(*). The 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 formal problems in organic chemistry.

(*) Brief abstracts of the various projects can be found in Appendix C on page 277 and more detailed progress summaries in Section II on page 81.

I.A.1.2   Resource Sharing

Besides the biomedical AI research theme of SUMEX-AIM, another central goal is an exploration of the use of computer-based communications as a means for interactions and sharing between geographically remote research groups engaged in biomedical computer science research. This facet of scientific interaction is gaining importance with the explosion of complex information sources and the regional specialization of groups and facilities that might be shared by remote researchers [8]. We expect an even greater decentralization of computing resources in the coming years with the emerging VLSI(*) technology in microelectronics and a correspondingly greater role for digital communications.

Our community building effort is based upon the current state of computer communications technology. While far from perfected, these developing 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. The network experiment also enables diverse projects to interact more directly and to facilitate selective demonstrations of available programs to physicians, scientists, and students.

We have actively encouraged the development of additional affiliated computing resources within the AIM community and expect such decentralization to become the "way of the 80's". Since 1977, the facility at Rutgers University has allocated a portion of its capacity for national AIM projects and our network connections to Rutgers and common facilities for user terminals have been indispensable for effective interchanges between community members, workshop coordinations, and software sharing. The "Caduceus" project (page 180) at the University of Pittsburgh has installed a VAX computer and is converting their software to run initial clinical tests at Pittsburgh. Since last year, the "Simulation of Cognitive Processes" project has been doing most of their work on their own VAX computer, and several more projects have proposed machines dedicated to their own use.

The proliferation of distributed machines will serve to increase the importance of electronic communications to facilitate interactions and sharing. Even in their current developing state, communication facilities enable effective access to the SUMEX community resources from a great many areas of the United States and to a more limited extent from Canada, Europe, Japan, Australia, and other international locations.

     (*) Very Large Scale Integration

I.A.1.3   Impact of AI in Biomedicine

Artificial Intelligence is the computer science of symbolic representations of knowledge and symbolic inference. 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 of their 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 cumulation of knowledge, for example:

- codification of theoretical and heuristic knowledge

- effective use of the wealth of information implicitly available in textbooks, journal articles and from practitioners

- dissemination of 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

E. A. Feigenbaum

We believe that computers are the most hopeful technology to help 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 information and using it 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 individual practitioners without reducing the significance of the individuals.

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; 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 therefrom 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.

The researchers of the SUMEX-AIM community currently constitute a large fraction of all the computer scientists whose work is aimed at the development of symbolic computational methods and tools. SUMEX-AIM is laying the scientific base so that medicine will be able to take advantage of these technological opportunities for inexpensive computer power. Medical diagnostic aids and tools for the medical scientist that operate in an environment of a network of professional workstation computers have the practical possibility of large-scale and low-cost use because of anticipated near-term developments in the computing industry.

I.A.2   Synopsis of Recent Progress

       We can report substantial further progress during year 09 of our
grant toward the overall goals of the SUMEX-AIM resource. We have
continued the refinement of an effective set of hardware and software tools
to support the development of large, complex AI programs for medical
research and to facilitate communications and interactions between user
groups. We have worked to maintain high scientific standards and AI
relevance for projects using the SUMEX-AIM resource and have actively
sought new applications areas and projects for the community. Many
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. As discussed in the sections describing the
individual projects, a number of the computer programs under development by
these groups have matured into tools increasingly useful to the respective
research communities. The demand for production-level use of these
programs has surpassed the capacity of the present SUMEX facility and we
have been actively investigating the general issues of how such software
systems can be moved from SUMEX and supported in production environments.

       A number of significant events and accomplishments affecting the
SUMEX-AIM resource occurred during the past year:

   1) SUMEX has acquired and integrated 5 Xerox Dolphin InterLisp
       workstations into the computing environment at Stanford to begin
       research on the important software issues involved in using personal
       computing environments for AI systems. Substantial systems work was
       done to develop Ethernet services to make the Dolphins work smoothly
       and effectively in our existing facility. This has been shared
       widely with other research groups outside of Stanford in order to
       broaden and assist the growing user community. Research has begun
       on moving the ONCOCIN, MOLGEN, AGE, and GUIDON systems to run on the
       Dolphins with promising early results. Shortly, one of the SUMEX
       Dolphins will be moved to Rutgers to enable researchers there to
       begin exploring its potential for biomedical AI systems.

   2) We took on development and support of a VAX 11/780 UNIX system
       purchased with ARPA funds for shared access by the Stanford
       Heuristic Programming and Network Graphics projects and members of
       the SUMEX-AIM community. We have carried out necessary systems work
       to integrate this machine well into the Stanford network system,
       install FranzLisp and the USC-ISI implementation of InterLisp-VAX
       for AI research, and expand the stable of available user software.
       Various Lisp performance measurements have been made this past year
       with disappointing results.

   3) The AI Handbook core research project completed publication of a
       three-volume, 1500-page compendium of knowledge about the field of
       Artificial Intelligence. It has been edited by Avron Barr, Paul
       Cohen, and Edward Feigenbaum at Stanford University, with textual
       contributions from students and investigators at several other

research institutions across the nation. The work contains hundreds
of articles covering most of the important ideas, techniques, and
systems developed during 26 years of research in AI.

4) The SUMEX-AIM collaborator project community has continued vigorous
    development of their respective programs. Details are reported by
    the individual investigators in Section II. The VM and ONCOCIN
    projects have continued preliminary clinical testing/evaluation this
    past year using SUMEX network and computing resources. The CADUCEUS
    project has begun setting up its own local VAX computing resources
    which should help reduce the load on SUMEX for newer pilot efforts.
    We have continued to work hard to meet the needs of collaborating
    projects and are grateful for their expressed appreciation.

5) We continued to support the highly successful GENET experiment for
    the dissemination of the MOLGEN programs into the molecular biology
    community. "Advertised" through presentations and demonstrations by
    MOLGEN investigators at several professional conferences, several
    hundred molecular biologists from over 60 institutions have used the
    system and most have found it easy to learn and highly effective as
    a research tool for their investigations. A new GENET Executive
    Committee has been set up to manage this sub-community and to direct
    its efforts from production analysis of DNA sequences to new
    developments of computer science applications in molecular biology.

6) We have continued support of the existing SUMEX facility hardware,
    software, and network systems to enhance throughput and to assist
    user access to existing and planned resources. Over the past year,
    we have improved internetwork software and developed Ethernet
    services to provide a gateway to other campus networks and to expand
    general terminal access. We are also completing a more efficient
    Ethernet interface for the KI-10 system and have supported CSD
    efforts to develop a mass bus Ethernet interface.

7) We have continued to investigate long term options for SUMEX-AIM
    equipment plans including professional workstations and central
    machine development. For the next 5-10 years, it is clear that the
    "ideal" configuration must comprise a heterogeneous environment
    including Lisp workstations and a substantial central resource for
    time-shared use by groups unable to afford enough workstations to
    support their planned staff. Recent work has indicated that the VAX
    is not an advantageous machine for LISP applications. We are
    therefore developing plans to upgrade the aging KI-10 system to a
    DEC 2060 TOPS-20 system.

I.A.3   Details of Technical Progress

The following material covers SUMEX-AIM resource activities over the
past year in greater detail. These sections outline accomplishments in the
context of the resource staff and the resource management. Details of the
progress and plans for our external collaborator projects are presented in
Section II beginning on page 81.

I.A.3.1   Facility Hardware

Over the past year, the SUMEX facility hardware configuration,
including the main KI-10 machine (Figure 1), the 2020 satellite machine
(Figure 2), the new shared ARPA VAX/UNIX system (Figure 3), and system
network interconnections (Figure 4), have continued to develop according
to plan and to operate effectively within capacity limitations.

The primary facility hardware development efforts this year have been
directed at:

1) Installation and integration of 5 Xerox Dolphin InterLisp
   workstations.

2) Development of the shared VAX 11/780 configuration.

3) Upgrade and expansion of Ethernet cable coverage, completion of
   Ethernet interface equipment for the KI-10 and development of other
   network server facilities.

4) Investigation and planning of hardware configuration alternatives for
   facility development.

5) Planning and purchase of initial file server configuration.

6) Follow-up on AMPEX memory expansion late in year 08.

7) Support of local project hardware needs.

Phasing of Hardware Acquisitions

In the SUMEX renewal, both an augmentation of the central resource in
terms of address space and capacity and exploratory work with Lisp
workstations were planned in the first few years. The Initial Review Group
recognized in their special study section report the importance of
optimizing the timing of our planned hardware acquisitions to coincide with
the availability of desired technological developments and community needs.
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 VAX during year 09 and our
first workstations during year 10. However, we judged that the order of
these purchases should be reversed for several reasons:

1) The VAX implementation of the InterLisp language, the basis for much
   SUMEX-AIM community AI research, was behind schedule and a
   preliminary working system was not expected until 1982. A widely
   usable production version would take somewhat longer.

2) An evaluation of the prospects of InterLisp-VAX performance was being
   prepared by Dr. Masinter (see Appendix A) and would not be available
   until late summer. This and the actual measured performance of
   InterLisp-VAX would strongly affect the suitability of the VAX for AI
   research.

3) In order not to delay non-LISP SUMEX-AIM work involving VAX, we were
   able to arrange that SUMEX-AIM would have shared access during year
   09 to a VAX 11/780 funded by ARPA to support Heuristic Programming
   Project research.

4) InterLisp Dolphin workstations were available for delivery in the
   summer of 1981 on which research toward adapting expert AI systems
   for the interactive workstation environment could begin.

Dolphin Workstation Installation

We ordered 5 Xerox Dolphin InterLisp machines for delivery early in
grant year 09. The machines were delivered on time, 2 in August and the
remaining 3 in October. The installation of these machines proceeded very
smoothly, given the previously established Ethernet setup at Stanford.
Experience with the physical Ethernet installation was important but the
software support that had been derived from Xerox PARC under a university
grant license and the extensions made to this software by our own personnel
were more important still. Facilities to monitor network activities, move
files around, and connect terminals between systems have proven especially
useful.

The machines received were preproduction models and were fully
configured but quite noisy. The hardware has in general proven reliable
except for the Shugart model 4008 disk drives that are part of the package.
We have had several failures of memory, processor, and other component
boards as might be expected with new machines. We have had 7 disk drive
failures, however, in the last 9 months -- far more than should occur.
These failures have concentrated in two of the machines, both of which are
housed in machine room environments. This suggests an interaction between
the machine and the disk in some undiscovered way. In all cases, Xerox
personnel have been responsive in repairing problems.

Performance of the Dolphin workstations has improved substantially
since February 1982. Dolphin performance is based largely on microcoded
implementation 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, it became clear that additional improvements were necessary. Xerox responded promptly to this new information and produced the Con Brio version of InterLisp-D which 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. The Con Brio release made the Dolphins much more comfortably usable as programming workstations. Still further improvements are needed and planned, especially in numeric computations and in the speed of compilation of Lisp functions.

We have agreed to share one of the 5 Dolphins purchased by SUMEX with the Rutgers resource, subject to two conditions. First, we must complete development work to integrate the machines in the Stanford Ethernet environment and second, Rutgers must allocate funds to purchase a second Dolphin of their own and establish a local Ethernet to connect them with other host machines. The first condition is now met with our work during the past year. The second is in process now and we expect to ship a Dolphin to Rutgers this summer.

Other aspects of the integration of the Dolphin workstations into the SUMEX computing environment are described in a following section on local network developments (see page 26).

Shared ARPA VAX 11/780

During year 09, as part of the plan to accelerate purchase of the Dolphin workstations, we gained shared access to a VAX 11/780 which was 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. This past year we augmented the machine by adding 2 Mbytes of memory and expanding the file system with a newly-announced DEC RP07 disk drive (512 Mbytes). Both of these expansions were funded by ARPA. Approximately 30% of the machine is allocated for HPP and SUMEX use.

Over the past year, much experience has been gained with the VAX as an environment for AI work using various Lisp implementations. This experience has been disappointing for the most part because of either shortcomings in the Lisp system itself (FranzLisp) or in the performance of the system on the VAX (InterLisp-VAX). This is discussed in more detail until the rationale for future hardware developments (see page 14)

Local Network Systems

Over the past 2.5 years, we have been developing a local, high-speed Ethernet 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. Since the early 3 Mbit/sec Ethernet

was given to Stanford and several other universities by Xerox, an agreement
has been reached between DEC, INTEL, and Xerox on the standards for an even
higher performance network [13]. The new network runs at 10 Mbits/sec and
supports a significantly larger packet address space. Xerox has started to
market products for the new network but debugged interfaces, software, etc.
for general use are just now becoming routinely available. Furthermore,
even though three companies have agreed on a set of low level protocols and
interface conventions, the rest of the world may not go along. There is
already an alternative (but closely related) IEEE specification in
preparation and alternative wideband network systems being delivered. Even
among the three parties in the Ethernet specification, there is no
agreement on higher level protocols.

All of this suggests that it is not time to jump to the newer and
faster networks yet. We feel the 3 Mbit/sec network is adequate for our
bandwidth needs in the near future and there is already a significant
hardware and software investment in 3 Mbit/sec network equipment at
Stanford related to SUMEX community interests. In the longer term, we will
want to upgrade to whatever hardware and protocol standard is broadly
adopted.

In the meantime we are continuing to develop our 3 Mbit/sec PUP
network services. This places a heavier burden on us to develop and
maintain our own equipment for Ethernet support. We have tried to minimize
the "home-brew" nature of this work by sharing common hardware and software
designs with other groups in the same situation

The current Stanford network configuration related to SUMEX-AIM work
is sketched in Figure 4. Over the past year, we have made considerable
progress in a number of areas of network development:

1) KI-10 DMA interface complete -- The initial KI-10 Ethernet interface
   was made via a PDP-11 connected to the I/O bus which is inefficient
   under heavy traffic. The extra demand made on the KI-10's for
   performing file server functions for the Dolphin workstations, high-
   speed terminals, printer traffic, and other file transfers underscore
   the need for a more efficient direct memory access interface. This
   interface is now debugged and uses a phase decoder (design borrowed
   from the SUN terminal project at Stanford) to detect the incoming
   serial Ethernet signal, an internal packet buffer to prevent overruns
   to and from the TENEX time-sharing system, and a memory bus interface
   to transfer data. The interface is not in operation yet because of
   problems with the phase decoder in handling low level signals which
   we receive from the very old cable being used to connect our gateway
   to the Computer Science building across campus. This cable will be
   replaced shortly and the new interface will go into full operation.

2) UNIBUS interface complete -- In our initial Ethernet connections of
   the KI-10, 2020, and VAX hosts, we used a UNIBUS interface board
   designed by E. Markowski at Xerox. Because of the limited
   availability of these boards for our future work, we designed a PDP-
   11 interface board that uses the serial phase decoder network front

end mentioned above. This design provides several features not
available on the Xerox board including more explicit error
information and a more sophisticated filter on source addresses for
incoming packets. A printed circuit version of this board has also
been produced and is being used in our PDP-11 TIP described below.

3) Network cabling upgraded and extended -- We have extended the
   coverage of our Ethernet this past year to cover the SUMEX and MYCIN
   office areas. This allows connection of Dolphin workstations in the
   offices to other network resources and also facilitates developmental
   work on Ethernet servers such as the gateway and TIP. Our gateway
   connection to other campus Ethernets has been made via an old cable
   dating back to the ACME facility. This cable is an RG-58 video cable
   and is very far from the standard Ethernet specification. Still,
   since it was the only cable available at the time, we installed
   special signal conditioning and detection electronics to make it
   work, however marginally. We are in the process of replacing this
   cable with true Ethernet cable. In the meantime, we have found a
   number of weaknesses in the current net interface designs. Because
   the old cable has a high loss (more than 100:1), problems with DC
   offsets, encoding skew, and bandwidth limitations have been been
   common. The phase decoder serial detector has proven particularly
   sensitive to these problems. Were we to make the decision again, we
   would not choose this method to decode the Manchester coded network
   signals. Rather we would use a threshold trigger/delay detector such
   as was used in the original Xerox design. This approach can also be
   extended to run at 10 Mbits/sec for the newer network protocols
   whereas the phase decoder cannot.

4) Ethernet gateway -- We chose to isolate the SUMEX Ethernet from other
   networks on the Stanford campus in order to avoid electrical
   interactions during development and to facilitate different
   administrative conventions for the use of the various networks. We
   therefore developed a gateway to couple the networks together as
   shown in Figure 4. This gateway consists of a PDP-11/05 with
   Ethernet interfaces on the SUMEX and Computer Science networks. It
   is located in Pine Hall which is a convenient intersection point for
   the networks. The major gateway effort was not in configuring the
   hardware but rather in developing the necessary software described on
   page 27.

5) Ethernet TIP -- Both because of a shortage in terminal ports to some
   of the host systems and the desire to allow terminals to connect
   freely to various hosts instead of being directly connected to one or
   another, we developed an Ethernet Terminal Interface Processor (TIP).
   This server is similar in function to the familiar ARPANET TIP. 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. We have built two versions of the TIP, one based
   on a PDP-11/05 processor and one based on the SUN MC-68000 processor.
   The PDP-11/05 can handle up to 8 lines without memory management

because of buffer space limitations in the 32K address space. The
MC-68000 can handle many more (probably up to 32), limited by
processor speed. The hardware configurations are straightforward and
include a processor, network interface, and groups of line
interfaces. Again the major effort has been in developing software
as described later.

Future Hardware Configuration Planning

Over the past year we have continued to evaluate strategies and
alternatives for planned system configuration development. In particular,
we have had a chance to evaluate the Dolphin InterLisp machines and shared
VAX, reassess the role of the dual KI-TENEX system, and reach a local
consensus about what the long term configuration of the SUMEX-AIM facility
should be. In summary, we feel that the best resource configuration for
the coming decade is 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 now available and
scheduled for announcement.

1) Lisp Workstations -- Since delivery of the Dolphin InterLisp machines
    about 6 months ago, considerable experience has been gained in their
    use as Stanford AI projects have begun adapting their systems to the
    workstations. The concept of an individual workstation, especially
    with the high-bandwidth graphics interface, has proven ideal. Both
    program development tools and facilities for expert system user
    interactions are substantially improved over what is possible with a
    central time-shared system. The main shortcomings of these systems at
    the present time are processing speed and cost. Benchmarks of Dolphin
    performance have varied widely depending on what program is being run.
    These have ranged from roughly half the speed of a KL-10 to an order of
    magnitude or more slower. The "average" experience with performance at
    Stanford has been comparable to KI-10 speed under a load average of 4-
    5. This is marginally acceptable in that human resources are still
    badly constrained by machine speed but the systems are usable in
    comparison to the heavy overload of the SUMEX resource. We expect
    other workstations to be available soon (Xerox, Symbolics, LMI, HP,
    etc.) that should improve performance and hopefully lower costs.
    Still, for machines costing in the range of $50,000, it will be some
    time before we can afford to equip most of the SUMEX-AIM community with
    individual workstations.

2) VAX -- Several groups have experimented with using the shared ARPA
    VAX/UNIX machine for AI system development. One group has used the
    FranzLisp derivative of MacLisp and the other has used the USC
    InterLisp implementation for the VAX. Both of these groups have
    reported poor performance of the VAX environment for various reasons.
    Both FranzLisp and InterLisp have proven very expensive on the VAX with
    InterLisp turning the VAX essentially into a single user machine.
    FranzLisp is a fairly minimal system compared to other modern Lisp
    systems and the development group at Berkeley is not providing

energetic support in extending or maintaining the system. The
InterLisp-VAX system is based on the "Byte Lisp" system used for the
Dolphin InterLisp system at Xerox and so is nicely compatible with
other InterLisp's. We undertook a separate study of the InterLisp-VAX
prospects under contract with Dr. Larry Masinter of Xerox, one of the
key developers of InterLisp 10/D (see Appendix A). This study pointed
out the difficulties of adapting InterLisp to the VAX and predicted
that because of the internal data structures of InterLisp, a VAX
implementation would be quite slow. It is a tribute to the USC-ISI
group under Dave Dyer that the VAX implementation has reached a usable
state. However, even with the further improvements predicted by USC,
it does not appear that the VAX will be a cost-effective Lisp
development machine. Preliminary results of another study of Lisp
systems running on a wide variety of machine environments being done by
R. P. Gabriel of the Stanford AI Laboratory, confirm this conclusion.

3) <u>Dual KI-TENEX Role</u> -- The SUMEX KI-10 system has served the AIM
   community well but is becoming a growing liability in terms of
   maintenance, staff commitment, and divergence from the mainstream of
   outside developments. For years we contributed to and derived from the
   TENEX community a wide variety of system software. That community is
   dwindling as more and more TENEX sites (e.g., USC, BBN, and SRI) move
   toward increasingly TENEX-incompatible TOPS-20 systems. We are thus
   left with a growing set of problems. New software being developed by
   the TOPS-20 community must undergo major changes to run on TENEX
   because of JSYS differences and the use of user-mode extended
   addressing in TOPS-20 with the release 5 monitor this past year. Our
   own software contributions have become of minimal use to other sites
   because they are incompatible. We are dependent on specialized and
   isolated in-house systems expertise to keep these machines running.
   Should we lose this in the competitive market for systems programmers,
   we would have a very difficult time finding or more likely training a
   replacement since there is no incentive for a programmer to learn a
   dying system. Even within DEC, it is becoming increasingly hard to
   find knowledgeable engineers for the KI-10 hardware and we are taking
   on more and more of the system diagnostic and troubleshooting tasks.
   We are coming up against a major software change caused by ARPA moving
   to the new IP/TCP protocol. DEC is taking on the monitor changes for
   the TOPS-20 community and others in the community are adapting user
   level programs to the change. No similar upgrade effort is underway
   for TENEX and we will likely be faced with developing our own software
   changes or lose our ARPANET connection. It is time to retire the KI-
   TENEX system and upgrade it to a more modern system that will provide
   the needed larger address space and free our systems staff to
   concentrate on more productive development efforts for the community
   such as related to professional workstations and compatible Lisp
   support.

4) <u>Planned Configuration</u> -- The cost and performance of individual Lisp
   workstations projected for the next few years means that much useful
   research can be done in experimenting with their use for expert AI
   systems but that we will not be able to equip a substantial part of the

AIM community with such systems. Thus, in order to continue support
for the SUMEX-AIM community at large and to provide facilities for new
AI efforts, the best facility configuration will continue to be a
central shared system surrounded with growing numbers of workstations
coupled by network. It does not appear feasible to continue support
for the KI-TENEX system for that period and the VAX does not appear to
be a cost-effective alternative. An upgrade of the KI-10 system to a
DEC 2060 TOPS-20 system is the best solution. The 2060 is in the
mainstream of on-going systems work; it has an extended address space
for which some Lisp systems are being adapted (perhaps even an
InterLisp); it has a processing capacity of 2-3 times the KI-10 badly
needed for our community; and it is more compact, reliable, and
maintainable. We are therefore preparing a plan for review by the AIM
Executive Committee, NIH, and council to begin implementation of such
an upgrade.

Initial File Server Development

The initial development of an Ethernet file server has been an
integral part of our council-approved year 09 equipment plan with further
expansions approved for later years. We have recently been able to take
advantage of an exceptional offer by Digital Equipment Corporation, through
their corporate external research sponsorship program, to purchase a VAX
11/750 as the processor part of the file server. In exchange for a
substantially reduced price for the machine, we (among several groups at
Stanford) have agreed to share information with DEC on the use of such
machines in our "personal computing" environment.

In the initial file server configuration, we are also planning to
purchase two 600 Mbyte disk drives for main file storage, one 800/1600 BPI
tape unit for long term archives, and one 300 Mbyte removable pack drive
for cyclic backups. We will most likely buy non-DEC peripheral equipment
as the most cost-effective approach. We will then complete the planned
initial file server configuration in year 10.

For a file server with such large storage capacity, the old approach
of periodic tape dumps for backup becomes impractical, especially in terms
of operations time required. We are therefore implementing a system
whereby specific user requests to archive files will cause the files to be
copied to tape. Other backups for system reliability purposes will be
written on a removable pack disk drive and a circular queue of packs
maintained to give backup recovery for the most recent 1 month period.
Such a system has been working successfully at Xerox and elsewhere on large
file server systems.

AMPEX Memory Expansion

At the end of year 08, we reported the expansion of our KI-10 AMPEX
memory from 256K to 512K, funded after council approval of our 5-year
renewal. This increase has improved system efficiency as predicted. We
observed an average reduction in page trap handling overhead from 10.7% to
6.7% and a reduction in system overhead from 36.1% to 28.3% thereby
delivering more processing to user jobs.

E. A. Feigenbaum   16

In the past years, we have observed infrequent parity errors in the AMPEX memory. By examining systematic records of these errors, we found they typically involved a single bit and all occurred in locations that are loaded once at system reload time and then not referenced until a system dump, monitor DDT operation, or other rare event. These errors could not be correlated with memory module and would change location when the monitor was reassembled. No information about the problem was available from AMPEX Corp. Since the problem was so highly intermittent and suspected to involve some backplane interaction that induced loss of contents of a static memory cell bit, we tried writing a program that every so often scans all memory thereby rewriting all cells during the read/refresh cycle. No further occurrences of this problem have been seen and we have not expended further resources on it.

Other Hardware Development

We have undertaken other hardware efforts as appropriate during the past year. These have included:

1) Installation of a machine room temperature and power monitor coupled to the Stanford Medical Center security office.

2) Purchase and installation of a Canon laser printer with an interface from Imagen, Inc. This system is just now being integrated into the SUMEX facility and will be reported in more detail next year.

3) Installation of a Vadic 1200/1200 dial-up rotary for the system using previously surplused hardware from the NIH.

4) Assistance in the debugging of the Mass bus Ethernet Interface System being developed by Computer Science.

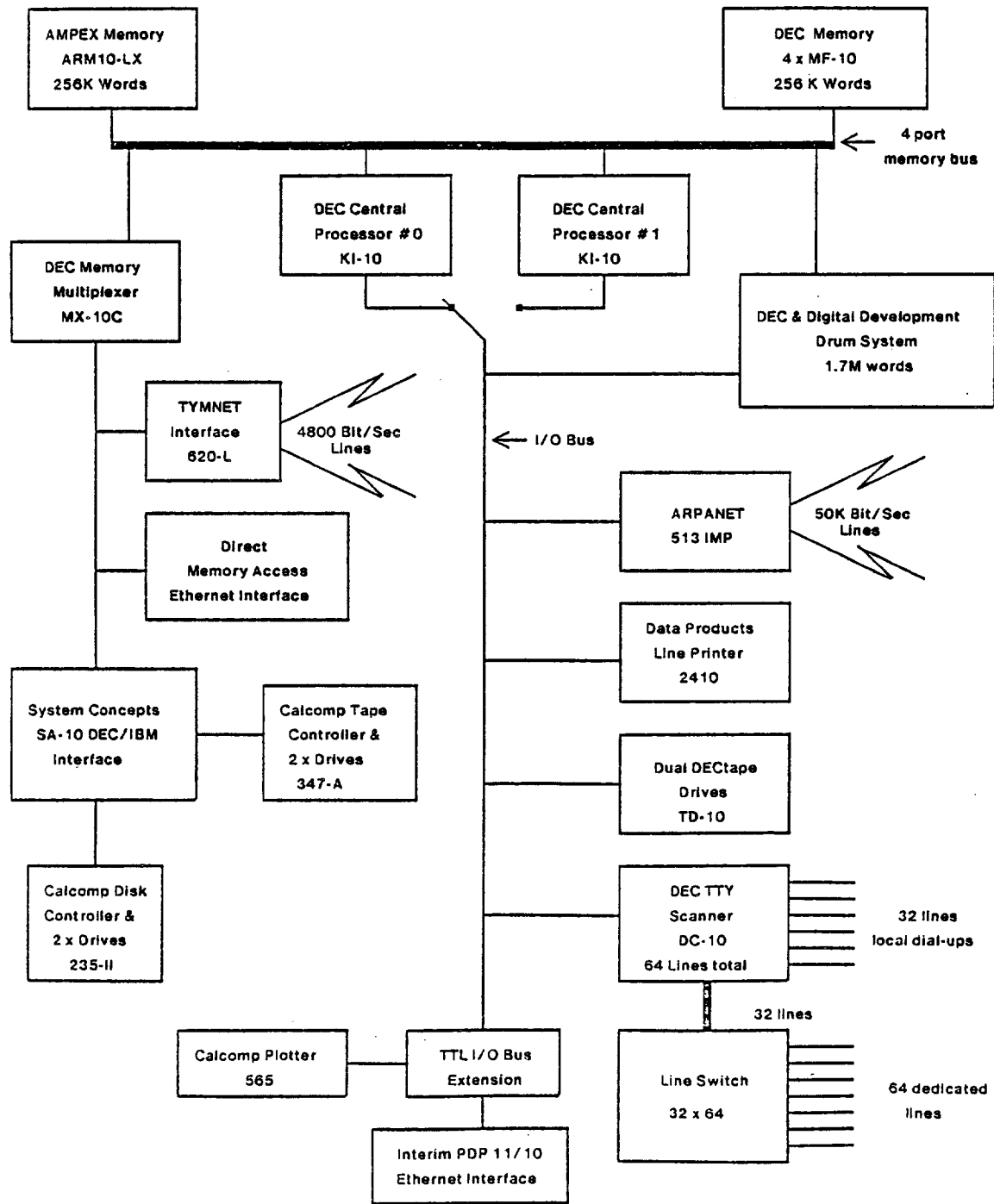In addition we have provided broad support to users for terminal and communications connections and repairs.

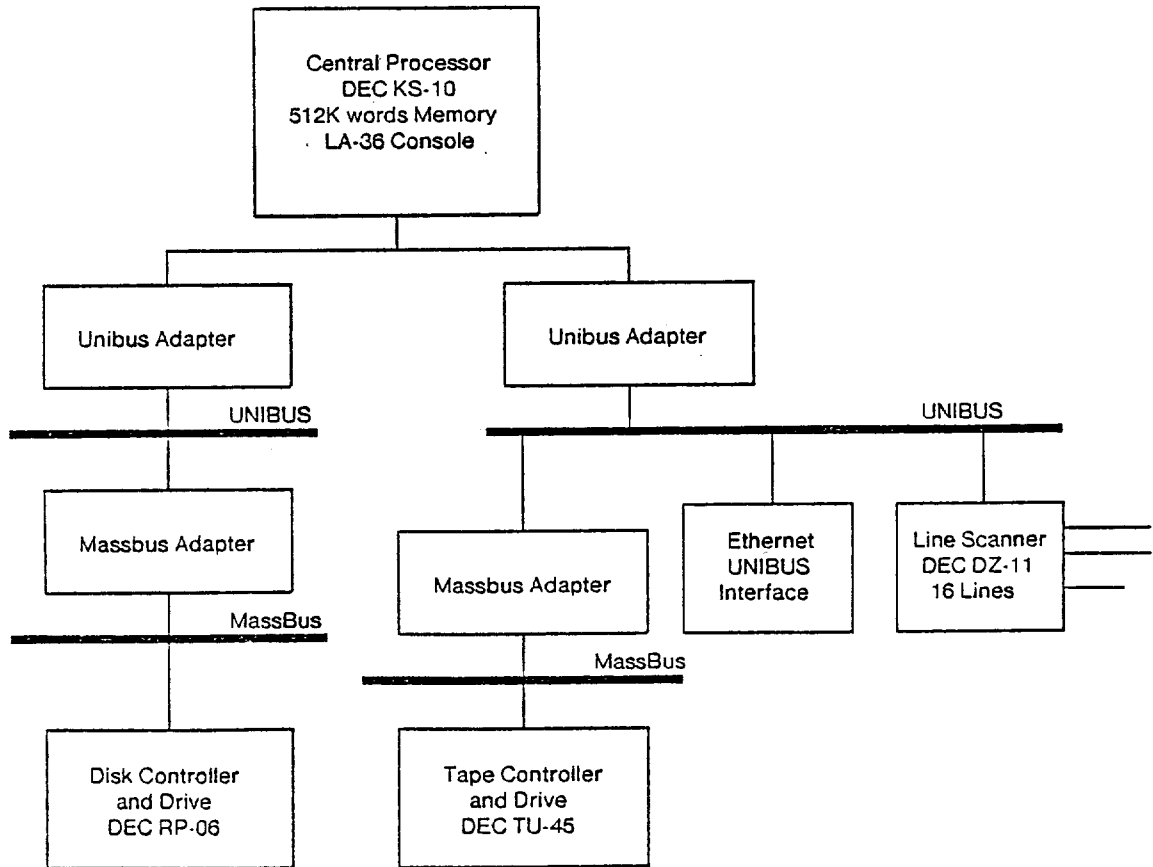Figure 1. Current SUMEX-AIM KI-10 Computer Configuration

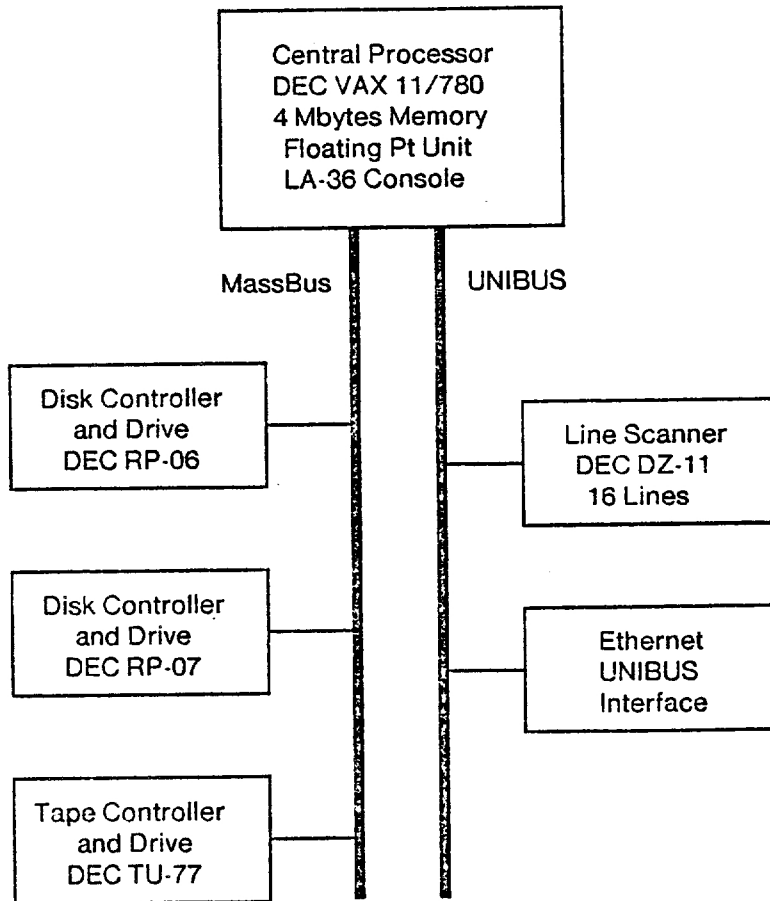Figure 2. Current SUMEX-AIM 2020 Computer Configuration

Central Processor
DEC VAX 11/780
4 Mbytes Memory
Floating Pt Unit
LA-36 Console

MassBus   UNIBUS

Disk Controller
and Drive
DEC RP-06

Line Scanner
DEC DZ-11
16 Lines

Disk Controller
and Drive
DEC RP-07

Ethernet
UNIBUS
Interface

Tape Controller
and Drive
DEC TU-77

Figure 3. Current Shared VAX Computer Configuration

Electrical
Engineering

Margaret Jacks
Heuristic Prog Proj

Pine Hall
Center for Information
Technology

CIT

SCORE 2060
3 Dolphins
Shared VAX
Ether TIP
Dover Printer
Other CSD Equip

G

Psychology

R

G

MYCIN Offices

Dolphin

R

Medical Center
SUMEX Mach Rm

SUMEX Offices

R Repeater

G Gateway

Dual KI-10
2020
File Server
Xerox Alto
SUN - Devel
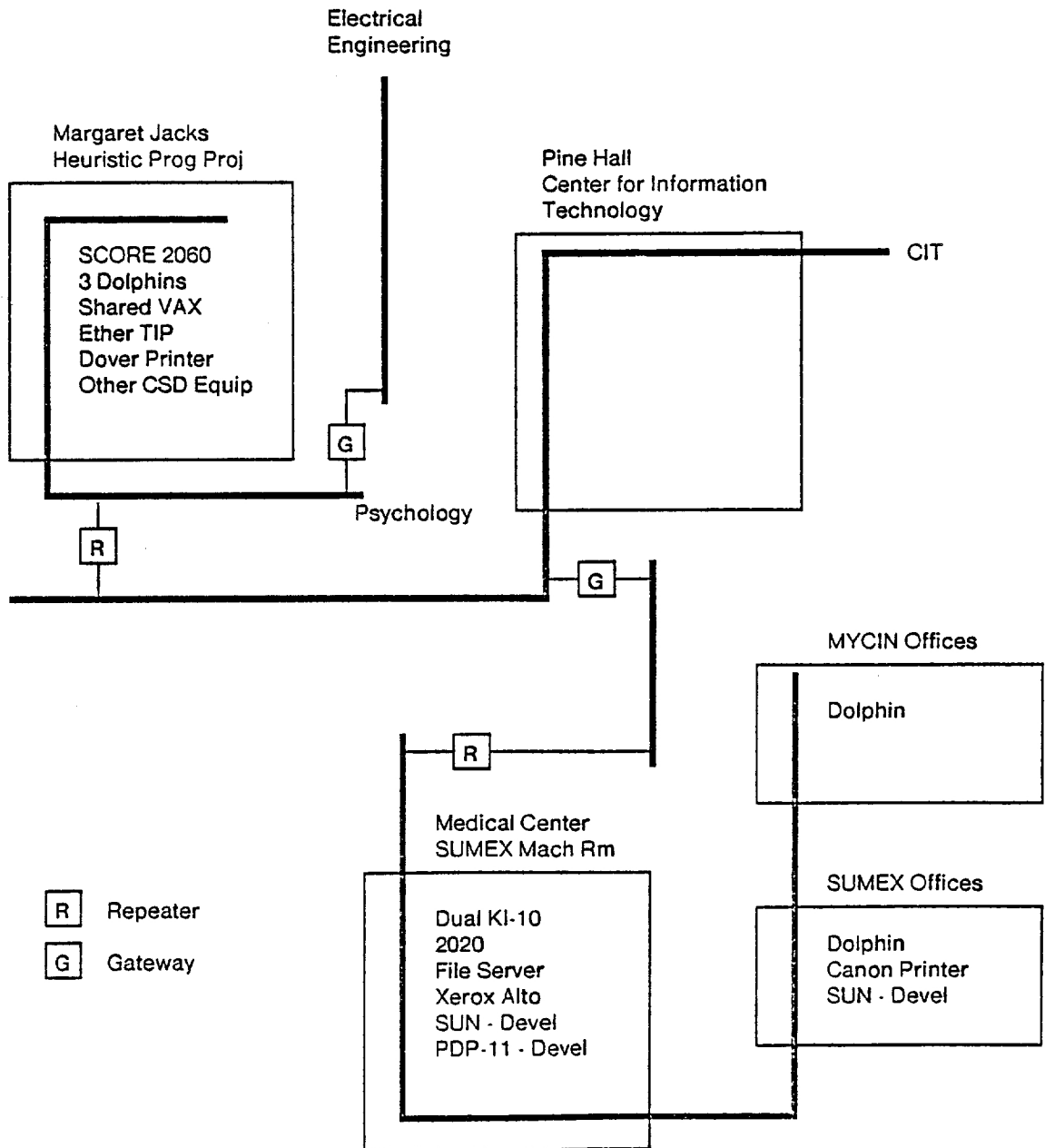PDP-11 - Devel

Dolphin
Canon Printer
SUN - Devel

Figure 4. Intermachine Connections via ETHERNET

I.A.3.2   Host System Software

Our system software work this past year has concentrated on several
areas including changes to support hardware development projects, upgrading
and enhancing network interface service, correcting encountered system
bugs, and implementing new features for better user community support. In
addition we have invested substantial effort in becoming familiar with the
VAX/UNIX system which has played a key role in our investigations this past
year.

Dual KI/TENEX System

DMA Ethernet Interface -- In parallel with our hardware efforts this
past year to extend and improve the KI-10 Ethernet connection, the
necessary monitor changes were made to support the new hardware. Initial
software work was begun earlier but the major part was completed this past
year including a new interrupt driver module for the monitor and an
extensive set of user-mode diagnostics. The dual processor system has
proven extremely flexible in order to carry out hardware debugging while
user time-sharing is in progress. Since most of the system hardware
devices are on one CPU, developmental hardware can be connected to the
other for debugging. That processor can be taken off-line temporarily for
hardware changes without bringing down the rest of the system. Also,
special diagnostics involving interrupt handling and other low level
monitor functions such as mapping special memory pages used by the hardware
under test can be performed on that machine without affecting other users.
The interface and software are now checked out and are awaiting
installation of the new gateway Ethernet cable so that reliable
communications with the phase decoder packet interface will be possible.

File System Expansion -- In past years, we have reserved one disk
drive as a backup in case of failure so that we could keep the system
operational. Because of increasing pressures for more file space, we made
the necessary monitor changes to bring this drive on-line, thereby
expanding the file system by 76,760 pages. This has met user community
needs at the expense of hardware backup.

ARPANET Protocols -- Effective January 1, 1983, the ARPANET is
scheduled to change from the current NCP (Network Control Protocol) to
IP/TCP (Internet Protocol/Transmission Control Protocol). This will
require a major change in the monitor service that supports our ARPANET
connection. In the past, changes of this type were funded for the
community by ARPA or the hardware vendor so that individual sites would not
have to redevelop such software. This is happening for the TOPS-20
implementation of IP/TCP with coordinated efforts between BBN and DEC. An
old experimental system runs under TENEX but it does not integrate the
network support into the standard file system calls as is the convention
for TENEX/TOPS-20. These changes will have to be done by the local staff
unless we can upgrade the KI-10's to a 2060 to take advantage of the
community effort. At least 6 man-months of effort will likely be involved
in such an effort.

System Loading Controls -- We previously reported on the system load
controls we have implemented on the KI-10 system to allocate available
system capacity effectively among projects and users according to Executive
Committee guidelines. These continue to operate effectively and we have
not made any substantial changes in this area.

Executive Program -- In order to better control the GENET community
(see page 73) who share a single directory, we designed a subdirectory
login facility in the EXEC that provides a two-step process for login.
First the user must know the global account name and password for GENET.
The subdirectory facility now requires an additional private user name and
password for entry. We instituted this control in order to preserve the
form of the GENET guest account but to screen out inappropriate users such
as industrial people. Together with the limit of 2 simultaneous jobs, this
mechanism has worked well.

Monitor Bug Fixes and Improvements -- We have continued to repair
important bugs in our TENEX monitor. In general the system runs extremely
reliably with most problems coming from explicit hardware malfunctions or
periods of instability following significant monitor changes. We found
several more subtle bugs in the system this past year that had been causing
various problems. By now, all of the "obvious" bugs have been located and
so those remaining are much more elusive, occurring infrequently or only
after a long chain of rare events that is difficult to reconstruct.

One of these involved a problem with mapping temporary pages in the
monitor in which the order of two instructions was inverted and the page
was marked as free before the necessary reference to the page was complete.
Another showed up during Ethernet interface debugging in the handling of
processes forced to run on one processor or another. And finally we
implemented an automatic recovery system for a bug that causes the ARPANET
service to run out of buffers infrequently.

2020/TOPS-20 System

Monitor Upgrade -- Our 2020 system has continued to run very reliably
this past year. We have upgraded the monitor to release 4 which required
substantial work to remerge the Ethernet service changes in the new monitor
code. At the same time we brought up the release 5 Executive program being
run at SCORE which is a beta test site for monitor release 5. We also
upgraded the COMND JSYS emulator package for TENEX to be compatible with
release 4. There will likely be few further monitor releases for the 2020
since it does not support extended addressing and there are no plans to add
this feature.

Demo Controls -- We previously instituted a mechanism for reserving
the entire 2020 for demonstrations and developmental testing of various
expert systems (e.g., DENDRAL, ONCOCIN, CADUCEUS, etc.). Because of the
unpredictability of usage during these reserved times and the resulting
waste of 2020 capacity, we changed this reservation policy by taking
advantage of the pie-slice" scheduler in the release 4 monitor. We now
guarantee dedicated users a large fraction of the machine but also allow

others to do useful work when the demo demand is low. This system has
nicely met the needs of both groups.

VAX/UNIX System

This past year we took on systems support for the ARPA VAX/UNIX
system shared by the SUMEX-AIM community. Substantial effort was spent in
becoming familiar with the UNIX monitor and in bringing up various user
subsystems such as FranzLisp and InterLisp. The new Berkeley release 4.1
monitor was installed after merging local Ethernet software and the
interprocess communication software from Carnegie-Mellon was imported and
brought up. When the file system was expanded with the RP07 disk, the new
device service and swapping off of both the RP06 and RP07 disks was added
and appropriate file system partitioning instituted between the major user
groups. A group protection facility was added to UNIX to facilitate more
file privacy between disparate groups of users. Finally, a substantial
effort was made to track down a serious bug in the multiplexed file code
used extensively by FranzLisp programmers. This bug caused the file system
to become corrupted when an "i-node" was erroneously written out twice.
After isolating the cause of the problem, we were able to import a fix from
CMU and the system has restablized nicely.

Dolphin System

Our system work on the Dolphins has centered mostly on the network
services provided by the various host machines. Still a significant effort
has gone into effective liaison with Xerox PARC and Electro-Optical Systems
to coordinate bug and benchmark information, install new software releases
and upgrades, and assure proper hardware maintenance. We actively
supported the distribution of TENEX, TOPS-20, and UNIX file server software
needed to interface the Dolphin to these host systems.

I.A.3.3   Network Systems

A highly important aspect of the SUMEX system is effective
communication with remote users and between the growing number of machines
available within the SUMEX resource. In addition to the economic arguments
for terminal access, networking offers other advantages for shared
computing including improved inter-user communications and more effective
software sharing. We continue to base our local communications on the 3
Mbit/sec Ethernet and remote communications on TYMNET and ARPANET for
reasons detailed in previous annual reports.

Users asked to accept a remote computer as if it were next door will
use a local telephone call or hardline connection to the computer as a
standard of comparison. Local networks stand up well in this comparison
but remote network facilities do not. Data loss is not a problem in most
network communications - in fact with the more extensive error checking
schemes, data integrity is higher than for a long distance phone link. On

the other hand, remote networking relies upon shared use of communication
lines for widespread geographical coverage at substantially reduced cost.
However, unless enough total line capacity is provided to meet peak loads,
substantial queueing and traffic jams result in the loss of terminal
responsiveness. Limited responsiveness for character-oriented TENEX/TOPS-
20 interactions continues to be a problem for network users.

TYMNET

TYMNET provides broad geographic coverage for terminal access to
SUMEX, spanning the country and also increasingly accessible from foreign
countries. Technical aspects of our connection to TYMNET have remained
unchanged this past year and have continued to operate reasonably reliably.
As noted earlier, however, users have complained periodically about having
their connections dropped and we have implemented a data collection
facility in the EXEC program to help document and classify these failures.
There are definitely episodes in which all connections are lost and the
jobs are detached. These occur about once every few days and we continue
to analyze these data to try to separate out local from network causes.

TYMNET has made few technical changes to their network that affect us
other than to broaden geographical coverage. The previous network delay
problems are still apparent although better cross-country trunks into New
York and New England are available improving service there. TYMNET is
still primarily a terminal network designed to route users to an
appropriate host and more general services such as outbound connections
originated from a host or interhost connections are only done on an
experimental basis. This presumably reflects the lack of current economic
justification for these services among the predominantly commercial users
of the network. Whereas TYMNET has interfaces meeting X.25 protocol
standards, the internal workings of the network will likely remain the
same, namely, constructing fixed logical circuits for the duration of a
connection and multiplexing characters in packets over each link between
network nodes from any users sharing that link as part of their logical
circuit.

We have continued to purchase TYMNET services through the NLM
contract with TYMNET, Inc. Because of current tariff provisions, there is
no longer an economic advantage to this based on usage volume. SUMEX
charges are computed on its usage volume alone and not the aggregate volume
with NLM's contribution to achieve a lower rate. We have implemented the
"dedicated port" charging system for SUMEX use and have realized a
substantial reduction in monthly usage costs. We will continue to work
closely with NIH-BRP and NLM to achieve the most cost-effective purchase of
these services.

ARPANET

We retain our advantageous connection to the Department of Defense's
ARPANET, now managed by the Defense Communications Agency (DCA). This
connection has facilitated close collaboration with the Rutgers-AIM
facility and many other computer science groups that are also on the net.