STANFORD UNIVERSITY MEDICAL CENTER
STANFORD, CALIFORNIA 94305 • (415) 497-5801

STANFORD UNIVERSITY SCHOOL OF MEDICINE
SUMEX COMPUTER PROJECT
Department of Genetics
Professor J. Lederberg
Principal Investigator

18 October 1974

TO : Dr. S. Amarel
Dr. W. R. Baker, Jr.
Dr. C. Brewer
~~Dr.~~ J. Lederberg
Dr. E. Levinthal

FROM: T. Rindfleisch

The attached are belated minutes to the July 30 SUMEX-AIM Executive Committee meeting, including updates reflecting events and activities since that meeting. Please review the minutes and forward any corrections or additions to me as soon as possible. (please do not let my example of promptness guide you).

Tom Rindfleisch

THE SUMEX AIM PROJECT IS A SHARED RESOURCE FOR RESEARCH IN ARTIFICIAL INTELLIGENCE IN MEDICINE SUPPORTED BY THE BIOTECHNOLOGY RESOURCES BRANCH, NATIONAL INSTITUTES OF HEALTH

TO: SUMEX-AIM EXECUTIVE COMMITTEE

FROM: T. Rindfleisch

SUBJECT: MINUTES FOR SUMEX-AIM EXECUTIVE COMMITTEE MEETING (7/30/74) AND CURRENT STATUS REPORT

The following are minutes for the meeting on July 30 of the SUMEX-AIM Executive Committee at Stanford University and also including updates reflecting developments since that meeting. Attendees included Dr. Brewer (NIH), Dr. Amarel (Rutgers), Dr. Lederberg (Stanford), Dr. Levinthal (Stanford), Mr. Rindfleisch (Stanford), and Ms. Carpenter (Stanford). The agenda for the meeting is reproduced in Attachment 1. The discussions of agenda items are summarized below in the order in which they were considered.

1) DEDICATION CEREMONY AND ANNOUNCEMENTS

Dedication:

Drs. Lederberg and Brewer summarized plans for the SUMEX-AIM dedication ceremony to be held in mid-November. The one-day session would include morning technical presentations by the five initial collaborator projects and an afternoon session covering broader policy issues. The tentative main speaker for the afternoon session was to be Dr. Herb Simon from the Carnegie-Mellon Psychology Department.

Dr. Lederberg would also invite local Congressional leaders with an interest in attending.

Subsequent to this meeting a number of aspects of the dedication planning have changed. Because of changes in the federal administration, the presentation of government policy views and goals related to resource sharing will be made by Dr. Thomas Bowery, Director of the Division of Research Resources. Dr. Simon is unable to make the keynote address because of schedule problems. In his place, Dr. Licklider, current Head of ARPA-IPTO, has agreed to give an overview of computer sharing through networks and its importance to scientific interactions.

The dedication date is fixed to be November 14.

Announcements:

We are in the process of announcing the existence of SUMEX-AIM in a variety of ways. The prospectus, available for distribution to interested people, is shown as Attachment 2. An excerpted version of the prospectus was published in the August SIGART (see Attachment 3). Dr. Levinthal is also organizing a session at the November SIGBIO meeting in San Diego to present a project description and summaries of on-going work by collaborating Principal Investigators.

2) SYSTEM STATUS AND PLANS

Communications:

A summary of network connection plans was presented. Connection to the TYMNET was scheduled to begin in early August and to be complete by late August. A potential problem was forecast in installing the 2400 baud telephone line linking the SUMEX node to the main network because of a possible telephone worker strike.

The telephone delay did not materialize and the equipment was installed and, for the most part, debugged during August. The TYMNET link has been provisionally operational since early September but a series of intermittent hardware problems, software debugging, and attempts to verify specified error rate performance have delayed fully operational status. These should be solved shortly.

The administrative question of paying for TYMNET usage through the NLM contract to achieve their favorable rates has been solved by Dr. Baker of BRB. SUMEX network usage will fall under the NLM contract with funding through an intra-agency transfer.

Approval to hook up to the ARPANET as a Very Distant Host was approved in early July by ARPA-IPTO. The approved plan included purchasing an interface from BB&N between the SUMEX KI-10 and a 50 kbit/sec line into the Stanford AI Lab IMP. The schedule was projected to be 6-9 months for an operational network connection.

The user bandwidth impact of a VDH connection as opposed to a local host connection has been difficult to quantitate On the one hand, facilities operating as local hosts seem to experience average effective network bandwidths of 10-15 kbits/sec - well below the 50 kbit/sec VDH link. On the other hand, no TENEX system has been connected as a VDH before and possible delays may come from the speed with which the time-sharing monitor can service network interrupts. Such response lapses could cause message retransmissions between the IMP and the SUMEX VDH thereby reducing the effective bandwidth. BB&N has agreed to work closely with us to minimize these problems and the interface has been slightly redesigned to include more buffering to reduce such overruns. Such delays, if a problem, can be eliminated by putting a PDP-11 network interface between the PDP-10 and the network as has been discussed in our previous planning. This would serve both to eliminate real time interrupt response delays and to off-load network service (NCP) overhead from the PDP-10. In heavy usage, this latter component can use 20-30% of the machine. We will keep the Executive Committee informed of this situation.

We have proceeded with the VDH approach and have orders in with BB&N for the interface and with the Range Measurement Lab of Patrick AFB for the telephone lines.

The initial plan to link to the Stanford AI Lab IMP has been altered. Taking into account the time required to install 50 kbit/sec lines and projected ARPA uses of local IMP ports at SRI, Stanford, XEROX, etc., ARPA has proposed that we connect to a port on a TIP operated by TYMSHARE. We have agreed to this change as no obvious disadvantages seem to result (TYMSHARE is highly oriented to reliable service, we are dealing with them relative to TYMNET services, and we may have redundant ARPANET access through their TIP).

The current schedule for line and interface installation is end of December 1974. We therefore should be able to begin experimentation with network operations by January 1975.

Dr Amarel asked if additional IN-WATS lines could be installed in the interim until network services were operational. After discussing the various factors involved (added cost, expected operational date for the TYMNET, current user load, and system capacity without a swapping device) it was decided not to increase the IN-WATS lines in August.

Swapping Storage:

The status of the swapping device was summarized. Digital Development Corporation has encountered serious technical problems in producing the model A7312 disk system we had ordered. Expectations as of the July 30 Executive Committee meeting were that the device would be installed in August. Additional problems with head crashes prevented this and we have reached an agreement with DDC to provide alternate swapping storage, based on an older, more expensive technology, in place of the A7312. This alternative will provide us equal or greater capability compared with the A7312 at no additional cost. This new device will be available in mid November and in the mean time, an interim, lower capacity, slower device will be installed. The interim device has been installed and is operational as of this week (10/5/74). Even though it is less capable than the final configuration will be, a substantial improvement in performance was observed. The fabrication of the final configuration is currently on schedule.

System Augmentations

Because the VDH ARPANET connection reduces the associated hardware costs over those previously budgeted for the TIP, we looked at the effects on system capacity for new AIM users of augmenting various components of the system configuration. The details of that study are described in Attachment 4. In summary, we proposed augmenting the file system (users are already approaching the available capacity), the core memory, and swapping storage. The latter two additions serve to reduce the overhead in swapping users in and outfrom about 30% in a simulated high paging load to about 10%.

A discussion ensued supporting the desirability of ensuring as much capacity as possible to new AIM users. Dr. Amarel expressed the desire that proposals such as this be distributed before the meetings so that more considered discussion could be held. We agreed with that viewpoint in general but explained that the timing of ARPA's decision on the VDH (early July) and the effort required to do the technical studies on augmentation rationale precluded such prior distribution for this meeting.

Dr. Brewer emphasized that BRB wanted to make sure that SUMEX-AIM allocated adequate resources, other than computing, to remote users and particularly to "have not" research projects which are not local to large computing facilities. These resources include terminals, communications, and other incidental help for enabling such users to get started. He also stated that the timing of system augmentations should match projected needs. We pointed out that such was the case at present since very rapid growth of the user community could be expected with the opening of network access.

The matter was taken under advisement by BRB, noting an approval by the Executive Committee to proceed with the augmentation. We have subsequently received budget authorization from BRB for augmenting the disk file storage and core memory. Funds proposed for added swapping storage are held in reserve pending future approval. In addition, a budget of $9,600 has been allocated for "have not" terminal rentals during the next grant year. It is the SUMEX-AIM policy that terminals are NOT generally provided to user groups - only in circumstances where abnormal fiscal or administrative conditions require it to allow new users to get started.

Arrangements for Backup and Reliability:

Dr. Lederberg discussed the desirability of making arrangements with other computr facilities for backup purposes or other mutual advantage. This may entail quid pro quo agreements in the form of mutual backup support, software exchange, or other "barter" arrangements. The Committee acknowledged the desirability of such agreements within the charter of the SUMEX project and authorized Dr. Lederberg to enter into appropriate relationships for the benefit of the facility.

Language Support:

Interest was expressed in two additional languages for the SUMEX machine - the Stevens Institute version of SNOBOL (SITBOL) and PL-1. Dr. Amarel's group has used SITBOL extensively in developing the programs at Rutgers and would like to use it on SUMEX. The lease cost of the software is $400 per year.

Dr. Lederberg indicated that the PL-1 language may be available through a development project in West Germany. A group at the German Cancer Research Center at Heidelberg may convert the PL-ACME language to run on the PDP-10. Dr. Amarel expressed interest in PL-1 as well.

A discussion ensued on two issues: 1) that SUMEX has limited resources and cannot properly support a long list of languages, and 2) that the needless proliferation of languages impedes software exchange. It was pointed out that there were virtues in using several languages in that each had strengths which could be exploited and that having various language processors available facilitated importing software using those languages. There was an acknowledgement that SUMEX resources are limited and a suggestion that Rutgers take on some of the language maintenance. Dr. Amarel indicated his resources were limited as well.

It was agreed that SUMEX would attempt to bring up SITBOL as possible with available resources. The general question of support for other languages and software systems within the SUMEX community will be considered at later meetings of the Executive and Advisory Committees.

We have subsequently obtained a provisional copy of SITBOL and brought it up nder TENEX. A number of file access problems were encountered which require sources programs to understand. We are in the process of obtaining these.

Jerry Feldman's group at the University of Rochester has suggested that BB&N's version of BCPL is a good system programming language and would let us bring up a number of programs that BB&N is developing. We are looking into the possibility of obtaining a copy of BCPL.

3) RESOURCE ALLOCATION POLICIES

It was pointed out that the need for allocation policies for the various consumable resources of SUMEX would be evident shortly, particularly as more and more new user groups come on. Specifically it was noted that the file system capacity was being consumed by on-line files and that even with the proposed augmentations, allocations to the various projects would have to be made and enforced. This is true of other resources as well (e.g., CPU cycles) although disk storage is most noticeable at present.

A tentative policy statement (Attachment 5) was distributed as a plan to control the use of file space within allocation boundaries. Little discussion followed because of meeting time limitations other than to note that other facilities (e.g., USC-ISI) were imposing various forms of access limitation (general users, message-only users, etc.). A fuller discussion of resource allocation policies was deferred until a later meeting.

4) NEW USER PROJECTS

A number of projects with potential AI relevance have emerged from the Health Manpower Act (769-A) proposals. Dr. Levinthal presented synopses of the nature of the various projects with discussion following on their relevance to SUMEX-AIM. A summary of the disposition of the 5 projects considered is given below:

a) Dr. Orthner, Geo. Washington University

Project Objectives -

Develop a mini network computer system and apply it to developin a general physiological model.

Committee Disposition -

It was agreed that the project as currently constituted does not fall within the working definition of AI. BRB will send Dr. Orthner a copy of the SUMEX-AIM prospectus noting that the proposal must be recast to be eligible for SUMEX.

b) Dr. Safir, Mt. Sinai Hospital (New York)

Project Objectives -

Collaboration with Dr. Amarel's glaucoma and eye disease diagnosis project.

Committee Disposition -

Dr. Safir will be a user of the SUMEX facility under Dr. Amarel's auspices.

c) Dr. McCormick, University of Illinois at Chicago Circle

Project Objectives -

Implement an effective computing support system for the University of Illinois Eye Clinic including patient data record automation and developing diagnostic aids. The system will include mini-computer record management using efficient switching theory cover algorithms to implement pattern matching procedures.

Committee Disposition -

This project has AI components which will likely grow with time. BRB will encourage Dr. McCormick's people to interact with Dr. Amarel's group's work preparatory to a later formal review as an independent project.

d) Dr. Pople, University of Pittsburgh

Project Objectives -

Expand and validate the data base in the DIALOG system, improving its human interface and testing its utility more extensively. DIALOG simulates the inferential processes of diagnosis within a "disease tree" data structure including a taxonomy of diseases and the likelihood of symptomatic relationship to underlying disease.

Committee Disposition -

BRB will invite Dr. Pople to get in touch with SUMEX personnel to plan details of his use of the facility.

e) Drs. Gorry and Schwartz, MIT

Project Objectives -

Develop the theory of clinical cognition including a simulation of expert history taking (data collection, diagnostic inference, use of uncertain information, etc.), representation of clinical knowledge, model-based decision making, and the application of common sense. The work would use goal-directed programming languages under development at MIT.

Committee Disposition -

BRB will encourage Dr. Gorry to stay informed of SUMEX and to cooperate with AIM goals. It is unclear at present if Gorry will get his own machine resources as proposed or may become a direct SUMEX user.

At present Dr. McCormick's group and Dr. Pople are actively using the SUMEX system.

5) AIM ADVISORY GROUP

Dr. Amarel reported that he had contacted the previously discussed list of candidates for membership on the AIM Advisory Group and that he has received a high rate of acceptance. The tentative membership includes:

Dr. Amarel	Rutgers University	AI Research
Dr. Feigenbaum	Stanford University	AI Research
Dr. Feldman	University of Rochester	AI Research
Dr. Bobrow	Xerox PARC	AI/Systems Research
Dr. Reddy	Carnegie-Mellon University	AI Research
Dr. Lindberg	University of Missouri	Medical Computing
Dr. Abelson	Yale University	Social Psychology
Dr. Safir	Mt. Sinai Hospital	Ophthalmology

A number of other candidates were discussed to broaden the scientific contact of the Advisory Group. Specifically, memberships for Dr. Minsky (AI) at MIT and Dr. Corey (Chemistry) at Harvard were proposed. These contacts will be pursued by Dr. Amarel

Executive Committee members unanimously approved the proposed Advisory Group membership and authorized Dr. Lederberg to send out formal letters of invitation to the nominees. These letters have been sent.

Dr. Amarel agreed to serve as temporary chairman of the Advisory Group for its early organizational phases but requested that a permanent chairman be elected by the group as soon as possible. That person would represent the Advisory Group on the Executive Committee.

It was agreed that Advisory Group members should be asked for nominal 3-4 year commitments with 1 or 2 rotations each year to provide group continuity. Dr. Minsky's commitment could be limited to 1 year if that would make him more willing to participate.

[Added note: Dr. Lederberg's formal letter of invitation specified just one year as the term of appointment as we had not yet specified the protocol for rotation. It is expected that Advisory Group members will be told informally about the rotation scheme and that this will be implemented through specified annual reappointments.]]]

The first meeting of the Advisory Group will be scheduled around the time of the SUMEX dedication ceremony.

6) EXECUTIVE COMMITTEE ORGANIZATION

Dr. Amarel raised the general question of the SUMEX-AIM committee organization and specifically the relative roles of the Executive Committee and the Advisory Group. It was agreed that the Executive Committee was directly responsible for making decisions on the allocation of SUMEX-AIM resources, user project authorization, and establishing policy guidance for the development of the SUMEX-AIM half of the resource.

The Advisory Group advises the Executive Committee on SUMEX-AIM priorities including the active promotion and recruitment of new projects and the screening of proposed new applications.

Dr. Lederberg agreed to serve as chairman of theExecutive Committee and to be responsible for informing users of the results of Executive Committee actions.

7) AIM WORKSHOP PLANNING

Dr. Lederberg presented a slight revision of the AIM Workshop Management Plan proposed by BRB. The revisions were minor to make the Workshop Plan consistent with the SUMEX-AIM management plan. Dr. Amarel questioned the need for a workshop management plan, given its allocated $25,000 budget. It was agreed that the proposed plan was not very elaborate and served to make clear the relative responsibilities of the various groups involved. The Executive Committee accepted the plan as modified by Dr. Lederberg.

Dr. Amarel reported on the very early workshop planning under way. He is accumulating information from earlier workshops held at Carnegie-Mellon and Washington University on other subjects to better take advantage of lessons learned there. A tentative plan includes an agenda for about 30 invited people over 7-10 days. These people will combine AI people, non-AI computer scientists, medical people, etc. Separate sessions may be necessary in order to tailor subprograms to the interests of a diverse group of attendees (biochemists, ophthalmologists, etc.). The theme of the Workshop will be "Knowledge-based Systems in Biomedicine" and it will be held at Rutgers.

8) FUTURE MEETINGS

The meeting was adjourned with the next session to be arranged by the chairman, Dr. Lederberg. The next meeting will be scheduled around the time of the dedication ceremony after the Advisory Group meets.

Attachment 1

AGENDA

SUMEX-AIM Executive Committee Meeting

July 30, 1974

1) System Status and Plans

   Swapping storage (DDC)
   TYMNET
   ARPANET
   Proposed augmentations

2) Allocation Policies

   File storage
   CPU
   Accounting

3) Health Manpower (769-A) AI-relevant Projects

4) Executive Committee

   Select AIM community member

5) AIM Advisory Group

   Define membership
   Plan organizing meeting

6) SUMEX Dedication and Announcement

   Oct/Nov dedication plans
   Prospectus
   SIGART announcement

7) AIM Workshop

   Review management plan
   Plan first workshop
      Time
      Theme: "Applications and skepticism of AI in Medicine"

8) General Discussion and Meeting Schedule

Attachment 2

SUMEX-AIM PROSPECTUS

INTRODUCTION AND OBJECTIVES

In partnership with and with financial support from the Biotechnology Resources Branch (BRB) of the Division of Research Resources, National Institutes of Health (NIH), Stanford University is developing and operating a National Shared Computing Resource to explore advanced applications of computer science in health research. The SUMEX (Stanford University Medical Experimental Computer) facility, under Professor Joshua Lederberg as Principal Investigator, is national in scope in that a major part of its computing capacity will be made available to authorized research groups throughout the country by means of a communications network. There are two main objectives of the facility: 1) the specific encouragement of applications of artificial intelligence in medicine (AIM) and 2) the managerial, administrative, and technical demonstration of a national shared technological resource for health research.

The emergence of more economical technologies for data communications allows a liberation from geographically organized facilities in favor of the concentration of functionally specialized capabilities at different sites. In addition to the economic advantages of resource sharing, a closer interaction between diverse research efforts is expected to promote a more systematic exchange of research products and ideas. This is particularly true in computer science where technical complexities have tended to encourage the development of relatively isolated groups, each pursuing a line of research and program development and making limited use of the working programs available from others. The SUMEX-AIM project seeks to lower these barriers in the specific area of artificial intelligence applied to health research. Indeed, multilateral community-building rather than unilateral service is the project's essential mandate.

"ARTIFICIAL INTELLIGENCE" RESEARCH

The term "artificial intelligence" (AI) is applied to research efforts aimed at studying and mechanizing information processing tasks that have generally been considered to require human intelligence. The current emphasis in the field is on efficient acquisition and utilization of material knowledge, and representation of conceptual abstractions in problem solving processes. AI systems are characterized by complex information processes that are, to a large extent, non-numeric, e.g., graph searching and symbolic pattern analysis. They involve procedures whose execution is controlled by different types and forms of knowledge about a given task domain, such as models, and fragments of "advice" in the form of systems of constraints or heuristic rules. Unlike conventional algorithms commonly based on a well-tailored method for a given task, AI procedures typically use a multiplicity of methods in a highly conditional manner - depending on the specific data in the task and on a variety of sources of relevant information.

For example, an AI system for data interpretation would assist a user at a higher conceptual level than the application of a numerical algorithm for curve fitting by suggesting and evaluating analytical expressions to match the data. Thus it may include deciding what model best describes the data source and what actions to take as a result. Such a model would be based on the data itself and on other pieces of information which may be relevant to constraining a solution. Inherent in these processes are a store of general symbolic knowledge about the problem domain describing what is fact, what is reasonable and unreasonable, and what solutions have worked in the past. Using an ability to communicate effectively with the problem data source and a human user, the programs might examine the knowledge base to construct and test plausible explanations of a particular set of data and to project and decide among subsequent alternative courses of action.

This type of "intelligent" assistance by computer program is perhaps best illustrated by a number of examples taken from on-going research efforts identified as initial users of the facility. The DENDRAL project at Stanford is aimed at assisting the biochemist in interpreting molecular structures from mass spectral and other chemical information. In cases where the characteristic spectrum of a compound is not catalogued in a library, these programs carry out the rather laborious processes a chemist must go through to interpret the spectrum from "first principles." By symbolically generating "reasonable" candidate structures from hints within the spectrum and a knowledge of organic chemistry and mass spectrometry, the program infers the unknown structure to be the one which best explains the observed spectrum. There is no direct algorithmic path available to determine such a molecular structure from the spectral data - only the inferential process of hypothesis generation and testing within the domain of reasonable solutions defined by a knowledge of organic and physical chemistry.

This process, as implemented in the computer, is a simplified example of the cycle of inductive hypothesis - deductive verification that is often taught as a model of the scientific method. (Whether this is a faithful description of contemporary science is arguable; and how it may be implemented in the human brain is unknown. In any case, these are useful leads rather than absolute preconditions for the pragmatic improvement of mechanized intelligence for more efficient problem solving.) The elaboration of these approaches as deeply as we can with existing hardware and software technologies is the most promising approach to enhancing the application of the computer to the vaguely structured problems that dominate our task domains.

A project related to DENDRAL, carried out in collaboration with the University of California at San Diego, seeks to infer the structures of complicated proteins from x-ray crystallographic data through a similar paradigm of hypothesizing reasonable structures and testing their ability to explain the observed data.

Other projects at Stanford and Rutgers University seek to assist experimentally in diagnosing disease and suggesting treatment. In these cases the input data are clinical symptoms and physical and biochemical measurements and the knowledge domain is the physiology and pathology of various organ systems as well as the effects and interactions of courses of treatment. Current applications of this work are in the areas of infectious diseases and diseases of the eye. Long term applications of these types of computer programs might be to consolidate and reconcile the knowledge from a diverse group of experts or to enable more effective treatment of disease in locations that lack access to specialized expertise.

Other examples, in areas of psychology, are aimed at building and testing complex models of human cognitive and affective processes. Programs are being designed at Stanford to simulate paranoid or other behavior patterns in response to natural language discourse. Embedded in such programs are a general symbolic model for the behavior type and an ability for natural language communication. In response to a human user, the program seeks to understand input discourse in terms of the behavioral model and to produce appropriate English language sentences in response. Applications of this work may be to improve our understanding of particular behavior patterns by systematizing the characteristics of models which emulate them or to assist in training medical students in psychiatry.

A system is being developed at Rutgers which takes as an input a social episode, i.e., an account of a sequence of actions involving the interactions of several persons within some social context, and it generates an interpretation of the episode in terms of intentions and reasons that might have motivated the persons in the episode to perform their actions. This system is based on a model consisting of a body of rules of belief (about specific people, actions, and motivations) and a strategy of interpretation. Applications of the system may be in improving communication processes such as the psychiatric interview.

These examples are given for the purpose of concrete illustration of the types of problem areas we seek to explore. They are not intended to bound the domain but rather to stimulate new ideas along the lines of "intelligent" programs with medical applications. These terms are not precisely defined at this time and our objectives might better be phrased in terms of "advanced computer science concerned with mechanized theory formation and problem solving in medical research and practice." We will emphasize the past achievements of AI-oriented and AI-labeled research in providing facilities for further advances such as symbolic knowledge representation and manipulation, concept formation, problem solving, learning, and natural human communication (e.g. language, speech, vision, etc.).

MANAGEMENT AND USER QUALIFICATIONS

The SUMEX-AIM facility is community-oriented and its organizational structure is being established to emphasize user support and interactions among user groups. The following summarizes the overall structure as it is tentatively constituted. Additional detail will be provided as appropriate, giving specific information about how prospective users may gain access to the facility.

The user community is divided for administrative purposes into two groups: 1) those at the Stanford Medical School (local to the facility) and 2) those elsewhere in the country and at Stanford. The facility resources (computing capacity and manpower) will be allocated in equal portions to these two groups. Stanford Medical School users will be qualified for access to the facility by Dr. Lederberg in his capacity as Principal Investigator for the SUMEX grant. The national user group will gain access to and be represented in the design, development, and allocation of the facility resources through an advisory group for a national program in Artificial Intelligence in Medicine (AIM). The AIM Advisory Group will consist of members at large of the AI and medical communities, facility users, and the Principal Investigator of SUMEX as an ex-officio member. A representative of the NIH-BRB will serve as Executive Secretary. It will advise the AIM Executive Committee, whose responsibility is to give overall direction to the national AIM program. The AIM Executive Committee consists of the Principal Investigator of the SUMEX Project, the Principal Investigator of a series of workshops on AI applications in medicine, a representative of the NIH-BRB, and a representative of the AIM Advisory Group.

Besides its charge to review potential uses of the SUMEX-AIM facility, the AIM Advisory Group will advise the AIM Executive Committee on the allocation of funds needed to assist in interfacing recommended new users. The SUMEX-AIM computing resource will be made available to qualified users without any charge, which of course entails a careful review of the merits and priorities of proposed applications. At the discretion of the advisory group, attendant communication and transportation costs to allow specific users to gain access to the facility may be covered as well.

Qualifications for new users and the details of AIM assistance to users will be more specific at a later time. In general terms, however, potential users will be judged on the basis of:

1. The scientific interest and merit of the proposed research.

2. The relevance of the work to the AI approach of SUMEX-AIM as may be indicated in part by the need for the specialized SUMEX facilities as opposed to other computing alternatives.

3. The prospective contributions and role of a user in the community, e.g., developing and sharing new systems or applications programs, sharing use of special hardware, etc.

4. The user's capability and intentions of operating in a community-effective style for mutual advantage. Besides the programming innovations that some users may be able to contribute, all are expected to furnish expert knowledge and advice about the existing art in the fields relevant to their special interests.

The overall objective will be to promote AI applications of high scientific merit among an extensive group of competent users, consistent with maintaining a responsive and productive computing environment. The initially approved loading is estimated to be about 30% of the total SUMEX-AIM capacity, leaving a substantial capacity for new research groups to enter at this time.

TECHNICAL CAPABILITY

The computer facility, consisting of a Digital Equipment Corporation model KI-10 central processor operating under the TENEX time-sharing monitor, is scheduled to be available in a limited fashion in June 1974, and to be fully operational in September. The system will have initially 197K words (36 bit) of high speed memory; 1.3M words of swapping storage; 40M words of disk storage; 2 9-track, 800 bpi industry-compatible tape units; 1 dual DEC tape unit; a line printer; and communications network interfaces providing user terminal access. At the present time the choice of a communications network approach is being finalized. However, specialized users will in any case find that the communications costs are small compared to the value of access to the system. This hardware complement may be expanded in the future based on available funding and justified user community needs.

Software support will evolve from the basic system as dictated by user research goals and requirements. Initially available programs will include a variety of TENEX user, utility, and text editor programs. Major user languages will include INTERLISP, SNOBOL, SAIL, FORTRAN-10, BLISS-10, BASIC, and MACRO-10.

INFORMATION CONTACT

For further information about the SUMEX-AIM project or to discuss potential research applications, please contact:

Dr. Joshua Lederberg
SUMEX-AIM Computer Project
Stanford University Medical Center
Stanford, California 94305

(415) 497-5801

Atta men 3 →

STANFORD IS SITE FOR MEDICAL AI COMPUTER — SUMEX

A five-year NIH grant of $2.75 million has been awarded to Stanford University Medical School to establish the first shared national AI computer facility for medical research. Funded by the Biotechnology Resources Branch of the Division of Research Resouces at NIH, the facility will be known as SUMEX - Stanford University Medical Experimental Computer - and directed by Dr. Joshua Lederberg, Professor and Chairman of the Department of Genetics.

Initially, SUMEX will consist of a TENEX PDP-10, scheduled to be fully operational in the early Fall and throughout the country on a time-shared basis over a computer-communication network. The two main objectives of the SUMEX program are (1) the specific encouragement of applications of artificial intelligence in medicine (AIM) and (2) the managerial, administrative, and technical demonstration of a national shared technological resource for health research.

The emergence of more economical technologies and data communications allows a liberation from geographically organized facilities in favor of the concentration of functionally specialized capabilities at different sites. In addition to the economic advantages of resource sharing, a closer interaction between diverse research efforts is expected to promote a more systematic exchange of research products and ideas. This is particularly true in computer science where technical complexities have tended to encourage the development of relatively isolated groups, each pursuing a line of research and program development and making limited use of the working programs available from others. The SUMEX-AIM project seeks to lower these barriers in the specific area of artificial intelligence applied to health research. Indeed, multilateral community-building rather than unilateral service is the project's essential mandate.

More particularly, SUMEX will be concerned with mechanized theory formation and problem solving as it occurs in both medical research and practice. It hopes to advance traditional AI subject areas such as symbolic knowledge representation and manipulation, concept formation, learning, and natural human communication as it relates to medicine.

Scientists who have been identified as initial users of SUMEX include Dr. Saul Amarel of Rutgers University, who is directing several projects in AI and biomedicine such as developing diagnostic aids for diseases of the eye and studying human communication processes in psychiatry, Drs. Stanley Cohen and Stanton Axline of Stanford Medical School planning strategies for the diagnosis and treatment of infectious diseases, Dr. Kenneth Colby of the Stanford AI project with PARRY, Drs. Edward Feigenbaum and Carl Djerassi continuing on DENDRAL, and Dr. Stephen Freer of the University of California at San Diego seeking to infer protein structure from x-ray crystallographic data.

Management

The user community is tentatively divided for administrative purposes into two groups: (1) those at the Stanford Medical School (local to the facility) and (2) those elsewhere in the country. The facility resources (computing capacity and manpower) will be allocated in equal portions to these two groups. Stanford Medical School users will be qualified for access to the facility by Dr. Lederberg in his capacity as Principal Investigator for the SUMEX grant. The national user group will gain access to and be represented in the design, development, and allocation of the facility resources through an advisory group for a national program in Artificial Intelligence in Medicine (AIM). The AIM Advisory Group will consist of members at large of the AI and medical communities, facility users, and the Principle Investigator of SUMEX as an ex-officio member. A representative of the NIH-BRB will serve as Executive Secretary. It will advise the AIM Executive Committee, whose responsibility is to give overall direction to the national AIM program. The AIM Executive Committee consists of the Principal Investigator of the SUMEX Project, the Principal Investigator of a series of workshops on AI applications in medicine, a representative of the NIH-BRB, and a representative of the AIM Advisory Group.

Besides its charge to review potential uses of the SUMEX-AIM facility, the AIM Advisory Group will advise the AIM Executive Committee on the allocation of funds needed to assist in interfacing recommended new users. The SUMEX-AIM computing resource will be made available to qualified users without any charge, which of course entails a careful review of the merits and priorities of proposed applications. At the discretion of the advisory group, attendant communication and transportation costs to allow specific users to gain access to the facility may be covered as well.

User Qualifications

Qualifications for new users and the details of AIM assistance to users will be more specific at a later time. In general terms, however, potential users will be judged on the basis of:

1. The scientific interest and merit of the proposed research.
2. The relevance of the work to the AI approach of SUMEX-AIM as may be indicated in part by the need for the specialized SUMEX facilities as opposed to other computing alternatives.
3. The prospective contributions and role of a user in the community, e.g., developing and sharing new systems or applications programs, sharing use of special hardware, etc.
4. The user's capability and intentions of operating in a community-effective style for mutual advantage. Besides the programming innovations that some users may be able to contribute, all are expected to furnish expert knowledge and advice about the existing art in the fields relevant to their special interests.

The computer facility, consisting of a DEC model KI-10 CPU operating under TENEX, will have 197K words (36 bit) of high-speed memory, 1.3M words of swapping storage, 40M words of disk storage, two 9-track, 800 bpi industry-compatible tape units, a dual DEC-tape unit, a line printer, and communications-network interfaces providing user terminal access. The initial communications network will be TYMNET, although SUMEX is also expected to be available as a remote host over the ARPANET at a later time.

Software support will evolve from the basic system as dictated by user research goals and requirements. Initially available programs will include a variety of TENEX user, utility, and text editor programs. Major user-languages will include INTERLISP, SNOBOL, SAIL, FORTRAN-10, BLISS-10, BASIC, and MACRO-10.

Information Contact

For further information about the SUMEX-AIM project or to discuss potential research applications, please contact:

Dr. Joshua Lederberg
SUMEX-AIM Computer Project
Stanford University Medical Center
Stanford, Ca 94305
(415) 497-5801

RECOMMENDED SYSTEM AUGMENTATIONS
WITHIN FIRST YEAR BUDGET

The initial SUMEX computer configuration plan, approved by the AIM Executive Committee in November 1973, was a compromise between the technical demands of establishing an effective community AI computing facility and the budget constraints imposed by Council. Within the projected budget at that time, we attempted to balance the configuration in terms of available file space, core size, and swapping storage.

As discussed at earlier Executive Committee meetings, ARPA has found it necessary over the past 6 months to reconsider its policies as they relate to ARPANET expansion and use by non-DOD agencies. The result of these deliberations has been a decision in early July by ARPA that SUMEX can become a Very Distant Host (VDH) on the ARPANET rather than a new TIP node as initially planned. We have revised the earlier network plan to implement a VDH interface and to augment the interim line scanner capacity to handle local terminals (previously to be handled by the TIP). We are also in the process of interfacing to the TYMNET in order to provide low bandwidth terminal support on a broader geographical and administrative basis than is afforded by the ARPANET at the present time.

Some reductions in first year costs have resulted from the reconfiguration and delays in implementing the ARPANET connection. These include delayed project staffing, delayed operational status, and reduced communications fees as well as the inherently lower cost of the VDH connection. The overall reductions amount to approximately $148,000 and afford the opportunity to reconsider other aspects of the machine configuration to give a larger capacity to better meet the needs of the AIM community.

Whereas the SUMEX facility is just coming to a fully operational state, we can project a number of areas where augmentation would be of benefit to system performance. These projections are based on observations of current SUMEX utilization as well as experiments on a KA-TENEX system at the Institute for Mathematical Studies in the Social Sciences (IMSSS). The IMSSS machine allows a more parametric measurement of performance sensitivity to hardware changes because it has a larger configuration from which the effects of reducing various component capacities can be observed. The following summarizes these recommendations.

FILE SPACE

Even in these early stages of SUMEX operation, it has become clear that the file system capacity will be a limiting factor to AIM community expansion. This derives from the interactive nature of the TENEX system making on-line files essential, the large files involved in AI program images, and the large data files currently in use and expected increasingly as data base-oriented AIM projects are identified. The capacity of the current file system is not yet fully utilized and we have issued only verbal requests to economize on file space. However, the trend toward early consumption of the file capacity is clear as summarized by recent file utilization statistics.

Out of a total of 81,200 available pages (4 RP-03 disk drives), the following are averages of the space in use including all system and user directories:

Mid-June	47,500 pages
Late June/early July	53,000 pages
Mid-July	52,000 pages
Late July	52,000 pages

We have developed a policy statement on file space allocation and control which is attached. In this policy, current data on disk requirements for various aspects of the system and user projects are integrated to allocate the overall available space (81,200 pages):

I. TENEX/AIM SYSTEM (common to both SUMEX-SUMC and -AIM)

Operating Monitor	5,000 pages		
Supporting Direct. (lang., lib., etc.)	10,000 pages		
AIM management and SUMEX staff	10,000 pages		
File system reserve for temporary overflows	6,200 pages		
		TOTAL	31,200 pages

II. SUMEX-SUMC Users

TOTAL	25,000 pages

III. SUMEX-AIM Users

TOTAL	25,000 pages

	81,200 pages

Among the initial SUMEX-SUMC projects (DENDRAL, Protein Structure Modelling, MYCIN, and various pilot efforts) approximately 17,500 pages are in use. On the SUMEX-AIM side only 8,000 pages are allocated because delays in network connections have precluded Dr. Amarel's and Dr. Colby's groups from actively using the system.

Based on these data, we recommend adding 4 more drives (81,200 pages - this is also the limit of the number of drives which can be put on the existing controller) to augment the SUMEX-AIM component of the file system. This would provide room for an additional 8-16 projects at 5,000-10,000 pages per project. At $13,000 (plus tax) per drive, the total cost for this augmentation would be $55,120.

MEMORY AND SWAPPING STORAGE

The operational status of the SUMEX KI-TENEX system has been approaching "routine" since May for the local community primarily. Over this period we have begun to collect statistics on the performance of the system but note that swapping is implemented on a provisional, inherently inefficient basis on the moving head file system disks. A sample of these data is shown in Figure 1. During the prime time shown, the system load was 10-14 jobs including 2 or 3 LISP users and miscellaneous EXEC, editor, and private program jobs. Plots are shown in Figure 1 of the percent time allocated to running user programs and the percent time consumed in system overhead (waiting for pages to be swapped in and out to make a program runnable, managing core allocations, and handling page fault traps). It is significant to note that the overhead consumes on the average about 35% of the machine under this load and in excess of 60% at times. This is predominantly a result of I/O waits on the relatively slow disks used for swapping. During this period, the maximum demand for swapping storage was 1750 pages.

A dramatic improvement in efficiency is expected when our permanent fixed head swapping device is installed in August, but these data raise obvious questions about the system capacity which will be allocatable to additional user projects. In conjunction with Mr. Rainer Schulz of the Stanford IMSSS facility, we have collected a preliminary set of data illustrating the relationship between system overhead and hardware configuration. The IMSSS KA-TENEX facility was used because they have a total configuration of 256 K words of memory and a large swapping drum in operation so that by limiting each of these parameters, we could evaluate the overhead under a "standard" load. The results of this experiment are shown in Table 1.

At present, the SUMEX machine is operating in a configuration similar to box 5 in Table 1 and with the installation of the swapping device will operate somewhere between boxes 1 and 3. (Note that the amount of virtual address space overflowing the "drum" determines the relationship of box 3 between boxes 1 and 5). The interaction between overhead, configuration, and job mix is complex. Witness for example, some data not shown in the Table. By adding 2 100 page jobs to the 4 200 page jobs in boxes 3 and 4, the overhead in box 3 is lowered while that in box 4 is raised. Nevertheless, several general relative trends can be noted. Increasing the speed of swapping storage reduces system overhead by reducing the I/O wait time for moving pages in and out of memory. Increasing memory size also reduces the overhead by allowing more working sets to be resident simultaneously thereby giving more candidate jobs to be run while waiting for pages to be swapped for other jobs.

It must be noted that the jobs run in this test simulate the effects of simultaneous very large jobs. In general there will be a spectrum of job sizes which will tend to reduce the overhead in all configurations (more working sets resident). On the other hand, the overhead estimates for swapping off of moving head disks are low because no data files were in use during the test thereby necessitating fewer time-consuming head seeks than would be encountered normally. Also the test programs addressed their arrays sequentially so that large blocks of pages would tend to be sequentially resident on disk. Thus in swapping programs in and out, less seeking would be required than normal.

From these estimates of relative system overhead as a function of configuration, it is clear that substantial gains can be made by adding memory to the system and by guaranteeing enough capacity so that paging occurs off a fast, fixed head device. This relative overhead can be reduced from something in excess of 30% (box 3) to something in excess of 11% (box 4) by adding memory and from greater than 11% to about 8% by adding more swapping storage. The improvement in efficiency by adding swapping storage would in fact be more than is apparent from the above data, taking into account the additional inefficiencies involved in more randomized disk seeks. Note that on the day data were taken for Figure 1, the maximum swapping space in use was 1750 pages. The fixed head swapping device we are getting will have a capacity of 2600 pages. Thus, in normal operating circumstances the probability that swapping storage will overflow to the slower moving head disk is real.

Even for a 100% efficient system, the number of users which can effectively be accommodated is limited by the response time for each user given roughly by a subdivision of the CPU capacity between the total number of users. It is very hard to pin down this number at present because it will depend on the nature of the jobs in execution. In the grossest terms, we might expect one limiting complement of users to be on the order of 5-10 LISP jobs (300-400 pages each) and 20 smaller jobs (50-100 pages each) for a total of something over 4000 pages of address space in use. This would clearly overflow the 2600 page swapping device.

For the above reasons, even though the firm limits of the current machine configuration have not been reached by existing user community demands, augmentations of the system memory and swapping storage would be beneficial to the AIM mission in allowing a larger community of projects to participate. Within the first year budget allocation, 64 K words of fast memory can be added ($50,000 plus tax) and the swapping storage doubled ($37,600 plus tax). Based on the relative data in Table 1, these additions, while costing about 10% of the overall facility, may free up approximately 20% of the machine capacity from overhead. This extra capacity is significant in terms of added AIM user support, we therefore recommend these augmentations in addition to the file system expansion discussed previously.

The total augmentations can be accommodated within the expected first year budget underrun:

File storage	$55,120
Memory	$53,000
Swapping storage	$39,850

TOTAL	$147,970

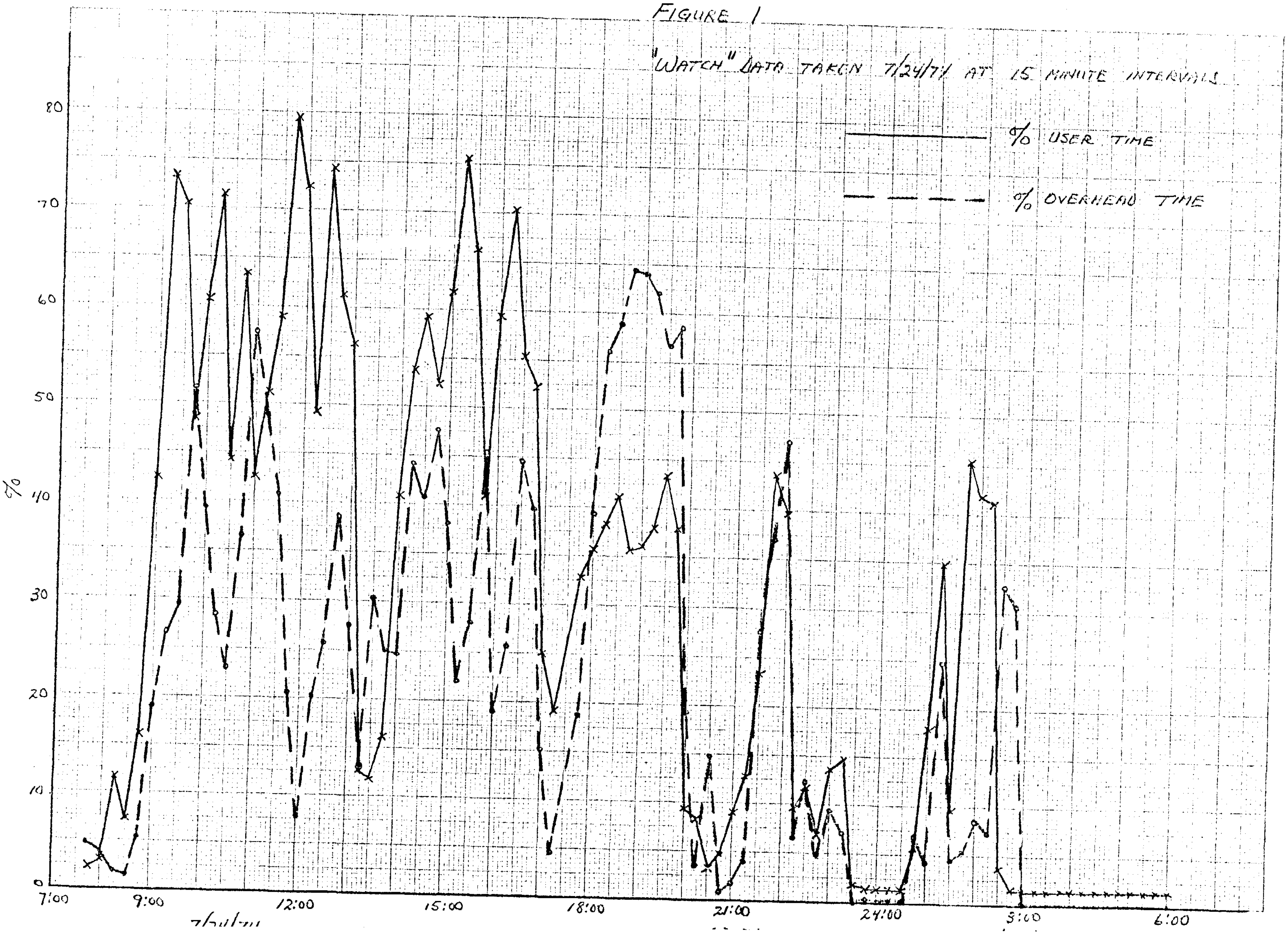
FIGURE 1
"WATCH" DATA TAKEN 7/24/71 AT 15 MINUTE INTERVALS
% USER TIME
% OVERHEAD TIME
%
80
70
60
50
40
30
20
10
0
7:00
9:00
12:00
15:00
18:00
21:00
24:00
3:00
6:00

Table 1

System Overhead as a Function of Configuration

	Memory 196 K	256 K
All "drum" (fixed head)	1 9 x 100 pages: 6% 4 x 200 pages: 16%	2 9 x 100 pages: 4% 4 x 200 pages: 8% 4 x 300 pages: 24%
Part "drum" / Part "disk" **	3 9 x 100 pages: (11%)* 4 x 200 pages: (33%)*	4 9 x 100 pages: (5%)* 4 x 200 pages: 11%
All "disk" (moving head)	5 9 x 100 pages: 17% 4 X 200 pages: 51%	6 9 x 100 pages: 6% 4 x 200 pages: 15%

* Estimated by interpolation because actual measurements were not available

** The drum space was limited to 450 pages with any overflow moving to disk

SUMEX FILE SPACE ALLOCATION AND CONTROL POLICY

The following summarizes the procedures to be followed within the SUMEX project for allocating on-line disk file storage to system and user directories as well as for enforcing these allocations. Up-dates to these policies will be maintained on <DOC>FILE-SPACE.POLICIES.

OVERALL DISK CAPACITY SUBDIVISION

The SUMEX file system hardware currently consists of 4 DEC RP-03 disk drives, each holding 20,300 pages (1 page = 512 36-bit words) of storage. Of this total of 81,200 pages, 31,200 are allocated to cover TENEX system-related functions, AIM (Artificial Intelligence in Medicine) management and SUMEX staff, and a file system reserve as follows. 5,000 pages are allocated for operating monitor functions including swapping storage, resident and swappable monitors, system crash dump area, and monitor bootstrap images. 10,000 pages are allocated for supporting monitor directories such as subsystems, libraries, documentation, bulletins, accounting, operations, and system program source files. An additional 10,000 pages are allocated to directories for AIM management and SUMEX staff. Finally 6,200 pages are held in reserve for the overall file system to accommodate temporary directory overflows.

This leaves 50,000 pages for user directories; subdivided into 25,000 pages for SUMEX-SUMC user projects and 25,000 pages for SUMEX-AIM user projects. SUMEX-SUMC is the half of the SUMEX project allocated to user projects from within the Stanford University Medical Center and SUMEX-AIM the half allocated to projects from around the country and elsewhere at Stanford.

USER PROJECT ALLOCATIONS

Within the SUMEX-SUMC and SUMEX-AIM moieties respectively, file space will be allocated to user projects consistent with established policy guidelines. As new projects are authorized, their file storage needs will be estimated and negotiated with the Principal Investigator of the new project. Based on these decisions, the new project will be allocated a portion of the SUMC or the AIM file capacity as appropriate. As such new projects are added, revisions to the other user project allocations may be necessary or the need for hardware augmentation established.

After an allocation of file space has been made to the user project Principal Investigator, he may subdivide his space between individual user directories authorized by him in any way he sees fit. System accounting and file maintenance programs will assist each of the Principal Investigators in keeping track of the file space utilization within his project.

Requests for changes in file space allocation from projects already using the system are made directly to the SUMEX Principal Investigator for SUMC projects and to the Executive Committee through the SUMEX Principal Investigator for AIM projects.

All communications regarding user directories and file space allocations should be routed by SNDMSG or telephone to Ms. K. Carpenter (415-497-5141 or Username = CARPENTER).

FILE SPACE ALLOCATION ENFORCEMENT

Users and projects are free to use file space as they wish within their allocations. As a directory exceeds its allocation, the system will inform anyone logged in to that directory of the amount the directory is over allocation. This notification stands as a warning that the user should delete obsolete files and use the archive system as appropriate to bring his directory within allocation. Any directory over allocation is subject to auto-purge clean-up by the system without further notice, should this be necessary to ensure continued system operation. Tools are currently available (e.g., DSKSTAT, DIRECTORY, DELETE, DELVER, EXPUNGE, ARCHIVE, INTERROGATE, etc.) to maintain directories within allocations. These will be extended as appropriate to assist in directory management.

At present files are not automatically expunged upon deletion and remain on the system in a deleted state (see pp 48-51 and 59-61 of the TENEX EXEC Manual), even when a job is logged out. This facilitates retrieving files which have been erroneously deleted but requires an explicit EXPUNGE to be executed to completely eradicate a file. If a system purge does not necessitate it sooner, all deleted files will be expunged once a week prior to the full system dump. A message announcing the up-coming system expunge will be issued in advance of the dump but no subsequent indication will be sent of the names of the deleted files which were expunged. This policy is based on the premise that files marked as deleted or as temporary (;T or ;S) are intended to be discarded after a brief interval of retention for security.

When the file system-reserve (6,200 pages) begins to be consumed by the aggregate of directories over allocation, decisions will be made as appropriate to purge the file system such that all directories are brought within allocation. If possible, all users will be notified when a purge is imminent with an estimated time and date. Users who are over allocation must clean-up their directories immediately or the system will clean them up at the time of the purge. The purge may occur at any time after the notification (if necessary, before the estimated time) to permit continued system operation.

At the time of a purge, previously deleted files will be expunged system-wide. In addition, for directories which remain over allocation, active files will be archived, deleted, and expunged according to a "longest time since last used" algorithm until each directory is brought within its allocated space. Each affected user will be notified by SNDMSG of the files archived during the purge. We will consider providing additional algorithms for the system selection of files to be removed. Each user may opt for the algorithm within those available which is most convenient to his style.

In addition, when the aggregate file space in use encroaches beyond half of the reserve space, restrictions will be placed on moving files into directories over allocation. In particular, it will not be possible to de-archive files, copy tape files, or transfer files over the ARPANET (FTP) into such a directory until it is brought under allocation so as to accommodate the additional file.

Attachment 6

; <RINDFLEISCH>MESSAGE.TXT;1   FRI 26-JUL-74 9:41AM

text:

ARTIFICIAL INTELLIGENCE IN MEDICINE (AIM)
Workshop Management

The AIM workshops are designed to promote better insight into and wider understanding of existing and potential contributions of artificial intelligence (*) research to biomedical research and the improvement of health.
These workshops,part of the overall AIM activity, are based at Rutgers University,Computer-in-Biomedicine Resource, They are intended to help develop and promulgate the appreciation and use of AI techniques to their full potential in biomedicine

As an augmentation of the SUMEX-AIM activity, the AIM-Workshops are logically administered within the framework of the SUMEX-AIM resource management plan.

POLICY-MAKING AND ADVISORY FUNCTIONS FOR AIM WORKSHOPS

Group	Composition	Function
AIM Executive Comm	SUMEX P.I.; BRB Repres've Rutgers P.I.; Action agent on AIM Workshop matters	Establish and coordinate advi- Allocate AIM Workshop resources. Coordinate workshop activities with SUMEX-AIM specify policies for AIM Workshops. Evaluate effectiveness of each workshop in meeting objectives.
AIM Advisory Group	SRAG Repres've,Chairperson BRB Repres've,Executive selected others	Recommend policies for the AIM Workshops--nature and scope Recommend presenters,participants for each session.
BRB Shared Resource Advisory Group (SRAG)	National membership selected to cover technology and science of all BRB Shared Resources. BRB Repres've, Executive Secretary.	Advise BRB on policies and actions to meet the needs of biomedical research through sharing of research resources.
