

AD-A012 589

COMPUTER APPLICATIONS IN THE BIOLOGICAL SCIENCES

FEDERATION OF AMERICAN SOCIETIES FOR  
EXPERIMENTAL BIOLOGY

PREPARED FOR  
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH  
ADVANCED RESEARCH PROJECTS AGENCY

FEBRUARY 1975

DISTRIBUTED BY:

**NTIS**

**National Technical Information Service  
U. S. DEPARTMENT OF COMMERCE**

AFOSR - TR - 75 - 095 1

211155

COMPUTER APPLICATIONS IN THE BIOLOGICAL SCIENCES

February 1975

Prepared for

Human Resources Research Office  
Defense Advanced Research Projects Agency  
Arlington, Virginia 22209

by

Kenneth D. Fisher, Ph.D.  
C. Jelleff Carr, Ph.D.  
John M. Talbot, M.D.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)  
NOTICE OF TRANSMITTAL TO DDC

This technical report has been reviewed and is  
approved for public release in accordance with FAR 190-12 (7b).  
Distribution is unlimited.

D. W. TAYLOR  
Technical Information Officer

This research was supported by the Advanced Research Projects  
Agency of the Department of Defense and was monitored by the  
Air Force Office of Scientific Research under Contract No.  
F44620-74-C-0077 (ARPA Order No. 2808; Program Code 4D20).

Life Sciences Research Office  
Federation of American Societies  
for Experimental Biology  
9650 Rockville Pike  
Bethesda, Maryland 20014

PRICES SUBJECT TO CHANGE

DDC  
REPRODUCTION  
JUL 25 1975  
REGISTRATION  
B

(Approved for public release; distribution unlimited)

Reproduced by  
NATIONAL TECHNICAL  
INFORMATION SERVICE  
US Department of Commerce  
Springfield, VA. 22151

AD A 0 1 2 5 8 9

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <b>AFOSR - TR - 75 - 0951</b>	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER <b>AD-A012 589</b>
4. TITLE (and Subtitle) <b>Computer Applications in the Biological Sciences</b>		5. TYPE OF REPORT & PERIOD COVERED <b>Technical</b>
7. AUTHOR(s) <b>Kenneth D. Fisher, C. Jelleff Carr and John M. Talbot</b>		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS <b>Federation of American Societies for Experimental Biology* 9650 Rockville Pike, Bethesda, Maryland</b>		8. CONTRACT OR GRANT NUMBER(s) <b>F44620-74-C-0077</b>
11. CONTROLLING OFFICE NAME AND ADDRESS <b>Defense Advanced Research Projects Agency Arlington, Virginia 22209**</b>		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>61101E ARPA order # 2808 68/3/2</b>
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) <b>DEPARTMENT OF THE AIR FORCE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (NL) 1400 WILSON BLVD ARLINGTON, VIRGINIA 22209</b>		12. REPORT DATE <b>February, 1975</b>
16. DISTRIBUTION STATEMENT (of this Report)  <b>Approved for public release; distribution unlimited.</b>		13. NUMBER OF PAGES <b>27</b>
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		15. SECURITY CLASS. (of this report) <b>Unclassified</b>
18. SUPPLEMENTARY NOTES  <b>*Life Sciences Research Office      **Human Resources Research Office</b>		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
<b>PRICES SUBJECT TO CHANGE</b>		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) <b>The computer sciences have had a major impact on research and development in the physical sciences and engineering over the past several decades. Their influence on biological research has developed less rapidly. Computer sciences have assumed a more important role in biological and medical research as hardware has become more sophisticated, more widely available, and as its capabilities have been appreciated. The applicability of computers to many facets of biological research is on the threshold</b>		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

-1- SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

of further advances because these applications appear to be maximally efficient and useful at minimized costs. However, monetary constraints over the next several years may delay the extent to which further applications are pursued. There is a need for improved mechanisms of communication between life scientists with limited computer knowledge and computer scientists with the expertise required to address topics in biological research. More emphasis in biomathematics as a bridge between the life sciences and the computer sciences could lead to significant progress in the next several decades. However, this will require emphasis on and support of interdisciplinary efforts in these two research fields.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

## FOREWORD

The Life Sciences Research Office (LSRO), Federation of American Societies for Experimental Biology (FASEB), provides scientific assessments of topics in the biomedical sciences. Reports are based upon comprehensive literature reviews and the scientific opinions of knowledgeable investigators engaged in research in specific areas of biology and medicine.

This technical report was prepared for the Human Resources Research Office, Defense Advanced Research Projects Agency (DARPA), Department of Defense, under contract number F44620-74-C-0077 monitored by the Air Force Office of Scientific Research.

The Life Sciences Research Office acknowledges the contributions of the investigators and consultants who have assisted with this study. The report reflects the opinions expressed by the consultants and a judicious attempt has been made to incorporate their different views and opinions; however, the authors accept responsibility for the contents. The listing of the consultants' names in Section VII does not imply that they endorse the conclusions of the study.

The report has been reviewed and approved by the LSRO Advisory Committee (which consists of representatives of each constituent society of FASEB) under authority delegated by the Executive Committee of the Federation Board. Upon completion of these review procedures the report has been approved and transmitted to DARPA by the Executive Director, FASEB.

While this is a report of the Federation of American Societies for Experimental Biology, it does not necessarily reflect the opinion of all of the individual members of its constituent societies.

C. Jelleff Carr, Ph. D.  
Director  
Life Sciences Research Office  
FASEB

## SUMMARY

The computer sciences have had a major impact on research and development in the physical sciences and engineering over the past several decades. Their influence on biological research has developed less rapidly. Computer sciences have assumed a more important role in biological and medical research as hardware has become more sophisticated, more widely available, and as its capabilities have been appreciated. The applicability of computers to many facets of biological research is on the threshold of further advances because these applications appear to be maximally efficient and useful at minimized costs. However, monetary constraints over the next several years may delay the extent to which further applications are pursued. There is a need for improved mechanisms of communication between life scientists with limited computer knowledge and computer scientists with the expertise required to address topics in biological research. More emphasis on biomathematics as a bridge between the life sciences and the computer sciences could lead to significant progress in the next several decades. However, this will require emphasis on and support of interdisciplinary efforts in these two research fields.

TABLE OF CONTENTS

	Page
Foreword . . . . .	3
Summary . . . . .	5
I. Introduction . . . . .	9
II. Overview of Current Applications . . . . .	11
III. Some Opportunities for Applications of Computers in the Biological Sciences . . . . .	15
A. Pattern Recognition . . . . .	15
B. Cell Image Analysis . . . . .	18
C. Artificial Intelligence . . . . .	18
D. Environmental Systems Analysis . . . . .	19
E. Computer Networks and Component Miniaturization . . . . .	19
F. Technology Transfer . . . . .	22
IV. Concluding Statement . . . . .	23
V. Bibliography . . . . .	25
VI. Key Investigators . . . . .	29
VII. Scientific Consultants . . . . .	31
DD Form 1473 . . . . .	32

## I. INTRODUCTION

Computer science has expanded at a phenomenal rate over the past thirty years and has made significant impact upon virtually all facets of modern civilization. The evolution of analog and digital computers and computer systems with extensive storage and retrieval capacities has influenced substantially research and development in the physical sciences, engineering, and technology. However, their importance in biological research has been relatively slow in developing despite the observation of Ledley (1965) that perhaps the greatest utilization of computers would be in the biomedical sciences.

Computers and computer systems essentially amplify and extend man's capacity to process information. They perform operations that are not otherwise feasible because of time limitations; and once programmed they can operate in a self-directed manner. Most computers and computer systems are conceptually similar in that:

- they manipulate data in accordance with programmed instructions involving given steps and procedures;
- the data processed by the hardware have evolved from extensive collection processes and frequently are subjected to prior processing;
- they perform logic operations at high speeds;
- they can store and retrieve data and the capacity to do so is a function of their size;
- they are capable of communication with their operators as well as with other computers;
- they are accurate and the error rate per total number of operations is essentially zero.

An excellent example of how computer technology assisted in approaching a biological problem is the analysis of amino acid content of proteins. Dayhoff (1972) has collected data on 57,000 amino acid residues of proteins from 176 different taxonomic groupings of plants and animals. Computer analyses of these data on similarities and differences in amino acid sequences have been used in developing patterns of possible evolutionary relationships among organisms. A computer model of the point mutation process has been developed from this data base. Analysis of the

similarities and differences in the structure of cytochrome C and C<sub>2</sub> by these computer based techniques has led to modifications in modern concepts of evolution and phylogeny of plants and animals.

It is obvious that developments in computer science and technology have resulted from continued massive federal support by numerous agencies (Moraff, 1974). A significant percentage of this support evolved from the early recognition by the military and engineering communities that computers increase productivity when skilled manpower is limited and that they have the capacity to handle otherwise unmanageable large-scale programs or data. For example, development of minicomputers and microprocessors has been spurred by the needs for portable and on-board facilities in aerospace and military equipment. More recently, these types of hardware have been incorporated into larger systems as modular subunits (e.g., the PLURIBUS system being developed by Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts). The advantages of such systems include the relatively low cost and high reliability of minicomputers as well as the ability of the subunits to function independently, if necessary.

On the other end of the spectrum, computer networks and time-sharing central facilities were first proposed, supported and developed by the military. The Defense Advanced Research Projects Agency alone invested approximately 20 million dollars to set up one of the first network facilities (ARPANET) which is still a model for computer networks. As of 1974, ARPANET is limited to a number of ARPA-supported research groups and several Department of Defense agencies but the interest in the potential of this system in life sciences research has been recognized (Heart, 1974). Other federal agencies are also heavily committed to use of computer systems and networks. A compilation of all data bases and related software available from 60 federal agencies has been published recently (National Technical Information Service, 1974).

Recognition of the potential utility of computers in biological research is demonstrated by the number of conferences, reports and symposia on this subject in recent years. Expanding use as research tools in planning, implementing, and analyzing biological studies suggests that even more new vistas are emerging. This report is a review of certain current applications of computer science and large-scale data handling technology to research in the life sciences; several areas of research that may be consistent with DARPA research goals are suggested. It focuses upon the use of computers in the basic areas of the biological science; however, specific applications in medical research and development are outside the scope of this report.

## II. OVERVIEW OF CURRENT APPLICATIONS

In computer science, as in other disciplines of science and technology, three conceptual levels of activity are evident: invention, innovation, and diffusion. Invention and discovery are related to the fundamentals of computer science and development of conceptually new types of hardware that form the basis of the ability to rapidly manipulate, store and retrieve symbols. Innovation refers to the first application of hardware, a system or specific techniques; diffusion involves continued application and extension of innovation to a broader base of issues and problems. It seems logical to suggest that much of the potential for computers in the life sciences has involved and will continue to include innovation and the subsequent application, or diffusion of knowledge, into more aspects of biology and medicine at many levels.

Computers and related systems have made significant contributions to experimental and theoretical aspects of scientific research. Theoretical calculations derived from a model or example of some biological phenomenon provide a set of values against which experimentally derived data can be compared. Based on data analysis, either the theoretical model or experimental design can be manipulated to provide a better understanding of the bewildering array of dependent and independent variables inherent in biological systems. Within the biological sciences, computers are being used in model formulation, analysis, and statistical compilation and projection.

In experimental protocols, computers are used widely: 1) real time automatic control, 2) sequencing of experimental procedures, 3) monitoring critical steps, and 4) the introduction of feed-back control. For example, with these techniques, simultaneous analyses of multiple enzyme systems, numerous physiologic parameters, and complex behavioral patterns are now possible. Such areas of research heretofore impossible because of technologic limitations now are feasible. In addition, both analog and digital computers are useful in data computation, reduction, analysis, and comparison. The speed, memory capacity, and ability of these instruments to manipulate artificial symbols such as numbers in "real time" have markedly enhanced the ability of investigators to perform accurate, reproducible, and complex experiments. Indeed, the concept of the Laboratory Instrument Computer (LINC), one of the first systems developed 15 years ago by biologists for use in biological research, led to more advanced and less expensive computer systems.

A substantial portion of current and accomplished research efforts have been reviewed in the multivolume series, Computers in Biomedical Research (Stacy and Waxman, 1965, 1969, 1974). Certain applications and potential uses of computers from the life scientists' viewpoint were reviewed in Biology and the Future of Man, a comprehensive treatise of the National Academy of Sciences (Handler, 1970). Finally, the Federation of American Societies for Experimental Biology, together with the National Science Foundation and the National Institutes of Health, recently sponsored a comprehensive conference on many aspects of the use of computers as research tools in the life sciences.\* In his summary of this conference, Siler (1974) concluded that a number of significant contributions to research in the life sciences had been made through use of computers. These include:

- understanding the structure and function of biologically active macromolecules by means of computer graphics, information storage, retrieval and analysis;
- development of minicomputers with specific functions in controlling and sequencing experimental procedures;
- availability of ramified and sophisticated statistical programs for data computation, reduction, analysis and comparison;
- development of computer systems with random access memory and error correction capabilities for study of causality and intentional activity in biological systems; and,
- production of systems and computer networks with human speech understanding capabilities for a wide range of biological studies.

The most important contributions may come from the use of computer hardware in evolving new approaches to the understanding of the multifactorial components of cell life, complex interactions at the organismal and ecological levels, and complex aspects of human behavioral patterns. Considerable experimentation and theoretical modeling can be done most effectively with computers. However, the costs and benefits of these approaches must be justified. The cost of access to computer networks or

---

\*The Conference on the Computer as a Research Tool in the Life Sciences (Lindberg, 1974) was published in the December, 1974 Federation Proceedings. Copies were provided to the sponsoring agency with the report. Additional copies are available from Federation Proceedings, Subscription Department, 9650 Rockville Pike, Bethesda, Maryland 20014 at \$5.00 each.

acquisition of minicomputers must be considered in relation to their utility in enhancing research. Many scientists in the biological disciplines have limited experience in this aspect of research planning. Another serious handicap to the widespread effective use of computers lies in the lack of adequately trained biomathematicians who can bring the capabilities of computer science to bear on issues confronting the life scientist.

### III. OPPORTUNITIES FOR APPLICATIONS OF COMPUTERS IN THE BIOLOGICAL SCIENCES

The following paragraphs outline opportunities for research applications of available computer technology and suggest possible research by computer techniques in the biological sciences. The order of presentation does not indicate any priority, nor is this section of the report meant to be all inclusive. It is an overview of some opportunities that have been identified in the course of this study. Additional research areas have been reviewed by Stacy and Waxman (1965, 1969, 1974), were presented at the recent Conference on the Computer as a Research Tool in the Life Sciences (Lindberg, 1974), and appear frequently in current scientific publications.

#### A. PATTERN RECOGNITION

In general, the larger the organic molecule, the greater the number of possible methods of synthesis. For macromolecules of biological significance the number of potential routes of synthesis is enormous. Selection of the most feasible or practical routes of synthesis is often a tedious and time consuming task. A computer program, Logic and Heuristics Applied to Synthetic Analysis (LHASA), can process up to 64 carbon, nitrogen or other skeletal atoms in the molecular structure excluding attached hydrogen atoms (Anonymous, 1974a). Graphic display and access to the computer's memory bank provides the investigator with the capability to construct synthetic pathways in a stepwise manner based on programs developed from sequences of single step transformations.

The LHASA program remains small although it has been improved repeatedly in the past 15 years. To expand the capability of the system to handle larger molecules requires additional data storage capacity. To make LHASA or a similar more sophisticated system available to a larger group of geographically separated users would require investment in network facilities. The hardware and software technology exist, the potential applications are evident, yet fiscal constraints appear to be a limiting factor in further development.

A related area of interest involves molecular model building by computer. Interactive graphics allowing operator-computer communication, provide visual display of not only static forms of known structures but also dynamic conformational changes resulting from interactions with other molecules. Interactive three-dimensional computer graphics have become an indispensable tool in research in determining actual amino acid sequences

in specific proteins (Dayhoff, 1974) and structure of nucleic acids and their derivatives, e.g., the calcium salt of guanosine 3', 5'-cytidine monophosphate (Langridge, 1974). Drug concentrations in tissues and plasma can be predicted from single or multiple measurements by computerized solution of systems of differential equations. For example, McInnis *et al.* (1973) have developed a multicompartment model by computer simulation to establish optimal dose regimens for the drug, digoxin. The interfacing of computers with gas chromatography and mass spectrometry provides another opportunity to collect meaningful data from minute biological specimens. Accurate qualitative and quantitative assessments of substances at the subpicogram level are now feasible with such instruments. These techniques could be used to monitor or predict the metabolic fate of either drugs that affect human performance or naturally occurring substances that reflect changes in body functions, e.g., hormone and phospholipid levels in the blood during stress.

Sacco *et al.* (1972) have used pattern analyses on body subsystems to develop time-sequence projections or forecasts of patient conditions. By projecting data on blood pressure, hematocrit, fibrinogen, as well as blood serum potassium, creatinine and osmolarity, they were able to develop a set of projective criteria from physiological and biochemical parameters that characterized the overall deviation from normal for severely injured patients.

While the authors indicated applications of their techniques to certain medical situations such as triage, it is evident that similar approaches to mathematical analyses of changes in specific physiological parameters may be useful in studying human performance and response to stressful situations. For example, changes in catecholamine excretion and serum phospholipid levels are known to follow abrupt alterations in work-rest cycles or critical tactical operations. The techniques of pattern recognition developed by Sacco *et al.* (1972) may provide an opportunity to address current military needs for a fresh approach to prediction of human performance from biochemical and physiological data.

A number of computer-assisted studies of electrophysiological phenomena have provided techniques and approaches that were heretofore impossible. For example, it is well known that signals from the electrocardiogram (ECG), electroencephalogram (EEG) and electrooculogram (EOG) provide quantitative and qualitative data on basic biological processes as well as responses to external stimuli. A recent report (Anonymous, 1974b) noted that Mason has perfected a procedure for estimation of the eye fixation point and tracking vigilance over time from the EOG of sonar operators. Computerized data processing of EOG records provides a reproducible method of calculating eye fixation time, visual field sampling patterns, target fixation data, and acquisition speed of such operational skills.

Using both in-house computer facilities and ARPANET, Vidal (1974) investigated the possibility of using electroencephalographic data in operant conditioning of various aspects of decision-making in man-machine communication situations. Recent studies have focused on development of a model for the phenomena of evoked responses in the EEG based on sequential events of short duration in the bioelectric potentials. Computer analyses of EEG records have identified candidate EEG patterns that appear to reflect the retinal processing of red, green, and blue receptors. These observations indicate that the trichromatic receptive processes of the foveal cone cells are being reflected in distinct patterns of electrical potential in the scalp which are identifiable in topological dimensionality analyses. Vidal (1974) believes that a combination of short event detection with observations of amplitude changes at selected times will provide encoding of the evoked response in a particularly economical manner. If this model is accurate, then components of the EEG could be reinforced naturally in a man-machine communication scheme analogous to operant conditioning. Reinforcement could be useful in certain aspects of training for critical performance in tasks such as operation of sonar equipment.

Orr *et al.* (1974) have identified patterns of ultradian rhythms in heart rate and performance measurements of normal subjects required to perform a standardized vigilance visual-motor response task for periods as long as 48 hours. Using complex demodulation analyses of ECG and motor responses developed by Orr and Hoffman (1974) they observed changes in individual ultradian patterns over time that were indicative of cumulative effects from stress of sleep deprivation and continuous vigilance. The complex demodulation technique is essentially computer-assisted analog band-pass filtering. This method of analysis provides both phase and amplitude data throughout any time series of choice. Modulation of a frequency band can be studied in the time domain of interest for the duration of the measurement period.

Hegge referred\* to subsequent studies in which he and his colleagues monitored the EOG, EEG, ECG, and the electrogastrigram of normal subjects and drug addicts during withdrawal. Preliminary data analyses by related complex demodulation techniques suggest that there are a number of normal ultradian phenomena with a periodicity of 50 to 120 minutes. These investigators believe that ultradian rhythms may be labile, and may be compressed in time with stress. If this hypothesis is correct, then ultradian rhythmicity may be an important aspect underlying human performance, vigilance, and sustained activity. Alterations in periodicity of ultradian rhythms, if measurable, might serve as an indication of the impact of stress on individuals in critical operations or situations.

---

\*Meeting January 15, 1975 held at Life Sciences Research Office, FASEB.

## B. CELL IMAGE ANALYSIS

Automated analyses of cell populations on the basis of size, shape, surface morphology and density has been a goal of cytologists for several decades. One of the earliest systems developed utilized optical density frequency distribution patterns, but was limited by system capacity to adequately classify and identify the cells (Prewitt and Mendelsohn, 1966). Recently, Lemkin *et al.* (1974) and Carman *et al.* (1974) have reported the development of a real time picture processor for use in cytologic identification. The major component of this system is the processing unit that provides rapid serial digital processing of digitized gray scale images by special purpose hardware containing image buffer memories as well as fast addressing programs. Because this system handles gray scale data, rather than binary black and white images, a larger set of image patterns can be generated, stored, analyzed, and identified. Such systems will be extremely useful in research programs requiring extensive analyses of cell populations.

## C. ARTIFICIAL INTELLIGENCE

Newell (1973) has recently reviewed the subject of "artificial intelligence" with respect to the development of this area of computer science in the past several decades. He suggests that the concept of artificial intelligence is not yet firm and cannot be clearly defined. He indicated that there are three conceptual approaches to artificial intelligence that prevade contemporary research. These include:

- the use of artificial intelligence to model, as well as explore, intellectual functioning;
- the use of artificial intelligence to test and generate approaches to systems of cause and effect or relationships; and,
- the theoretical, psychological approach, which involves theoretical modeling of contemporary theories of human behavior.

One practical result of research in this aspect of computer technology is the development of speech understanding systems. According to Reddy (1974), by the early 1980's, systems with speech-understanding capabilities such as the Hearsay system being developed at Carnegie-Mellon University will be available for laboratory and field use. On the

other hand, Birkhoff (1975) has pointed out the over-optimistic predictions for developments in "artificial intelligence" made 15 years ago have yet to be fulfilled. He indicated that computer programs simulating the pattern recognition capability of the human eye have been written, but also noted that current voice processing computer systems require about 60 seconds to process one second of human speech.

#### D. ENVIRONMENTAL SYSTEMS ANALYSIS

Lee *et al.* (1974) have developed a program for storage and retrieval of environmental impact information using the Systems 2000, a generalized data base management system (MRI-Systems Corporation). The program was developed initially to evaluate the environmental impact of certain United States Army military programs; however, computer systems that can analyze and potentially predict the effects of large-scale human activities on the physical and biological aspects of the environment have broad applications in many areas of the environmental sciences.

In the field of systems ecology, of intense current interest, computers are playing a decisive role, with a relatively up-to-date survey provided by Patten (1971, 1972). Compartment analysis, originally developed for transport of metabolites among organs of humans and animals, has been extensively applied in ecosystem analysis. The DYNAMD language, developed originally for industrial management use, has achieved important applications in ecological modeling, notably the Forrester-Meadows World Model; a special "Energy Language" has been developed for ecosystem simulation. The field is currently so active that the National Science Foundation has set up a special Systems Ecology Program which is primarily concerned with the use of computers in ecosystem simulation and analysis. The importance of environmental data management was a topic of discussion at the recent Stockholm Conference on the Environment (Oppenheimer and Brogden, 1975).

#### E. COMPUTER NETWORKS AND COMPONENT MINIATURIZATION

The development of systems and networks with larger and larger memory capacity, greater operational capability, and interconnected terminals is but one of the rapidly expanding areas in the computer sciences. At the other end of the spectrum, incorporation of microelectronic circuitry into computers has provided an impetus for production and innovation with minicomputers and microcomputers. The handheld calculators designed with multiple functions are an example of such applications.

Miniaturization of components and expansion of microprocessor capabilities have already made significant impact in the computer sciences (Ackley and Nardello, 1973; Bogle *et al.*, 1973). Initially based on circuitry allowing simultaneous processing of four bits, recent advances in microprocessor technology have resulted in production of circuits capable of handling 8 and 16 bit word lengths. Even more advanced circuitry is becoming available. These developments as well as cost reductions in fabrication suggest that a wide range of applications for microcomputers will become more feasible as initial investment costs are further reduced.

Miniaturized computer systems have already been incorporated into sophisticated laboratory instruments with considerable success. For example, high speed general-purpose oscilloscopes, gas chromatographs, and distance measuring devices and microprocessing capabilities are commercially available. The development and wider availability of such microcomputerized research instrumentation will have considerable impact on research in the biological sciences (Robinson, 1974).

However, microprocessor systems are limited currently by the requirement that they be programmed in either machine or assembly language. There is a need for higher level symbolic languages such as FORTRAN, BASIC, and specialized application languages such as SKED, that can be used with cross compilers and cross assemblers. Because the majority of research investigators are not skilled in currently available microprocessor languages, the potential benefits from use of these systems is, at present, closely related to availability of programs and programmers.

It is recognized that development of higher level symbolic languages will involve considerable cost with low profitable return on the investment. The same issue was evident during the development of minicomputer language systems. The experience gleaned from that phase of development in computer science should be applicable to the current situation in regard to microprocessor language systems. Experience suggests that additional federal or industrial support will be required.

The early experiments with computer networking clearly established that scientists can use program systems over great distances. The SUMEX machine on TYMSHARE and ARPANET is an example of one effort to get specific programming systems ready for network use. Heart (1974) reported that there is considerable interest in using the ARPANET by life sciences research groups both within the government as well as among other user organizations. He concluded that access to the ARPANET would be beneficial to investigators in the biological sciences, particularly in relation to their objectives in gaining experience in sharing a resource such as ARPANET. The capacity for communication within this system is currently unavailable to many research programs in the biological sciences.

Heart (1974) called attention to the possible use of network facilities such as ARPANET in ameliorating the duplicative needs for expensive and specialized resources that are becoming essential to large-scale research efforts in many biological disciplines. A network and its attendant technology provides a convenient cost-effective method of sharing large data bases, storage, retrieval and analyses of data from sources isolated in time or by geography, and, specialized software packages to support research in particular fields. Programs that provide information exchange among systems within networks are also required. Current work on a graphics protocol for ARPANET suggests that the difficulties may be due to hardware incompatibility. Limitations in software compatibility can be overcome as semantic protocols are developed for interaccessibility of data bases from different disciplines utilizing network facilities.

It is recognized that a number of networks and shared facilities with impact on the life sciences are already operational. MEDLINE, a network facility of the National Library of Medicine, provides a powerful bibliographic system for the medical sciences. There is no such extensive network system in the other biological disciplines that is readily accessible to so large a number of separate terminals and users at minimal cost. Computer systems such as PROPHET, MIRACLE, CHEMCON, and CRYNET are examples of advanced technology in specific areas of current research (Heart, 1974; Raub, 1974). There are opportunities for use of networks and shared facilities in the biological sciences and the financial aspects of such technology favor widespread development of these types of computer systems.

In summarizing a recent symposium on this topic, Wilson (1974) concluded that additional efforts are needed in:

- "making networking available to smaller, less sophisticated users, which will require handholding; now, too often, the rich get richer; those who are already well endowed with computer access get more through networking;
- overcoming of political, organizational, and economic impediments; encouragement and subsidy of protocol development and standardization of access, even across different networks;
- bringing specialized and discipline-oriented computing centers into sophisticated networking, for example the Health Sciences Computing Facility, the National Center for Atmospheric Research, Lawrence Berkeley Laboratory, Brookhaven National Laboratory, and so on; in addition, serious consideration should be given to implementation of new discipline-oriented centers now made practical by networking."

Ideally, synthesis of the advantages of networks with recent developments in microcomputer capabilities could lead to a handheld personal computer for the working nonprogrammer life scientist. Such a terminal could act as an interface with multiple laboratory instruments for data collection; run routine data analysis; function as a network terminal providing access to data bases and other information. This concept of a personalized computer with network access could also function as a point of origination and reception for network dissemination of scientific publications.

#### F. TECHNOLOGY TRANSFER

The National Aeronautics and Space Administration (NASA) and the former Atomic Energy Commission (now a part of the Energy Research and Development Administration) have established a Technology Utilization Program for dissemination of information on technological developments which have potential utility outside the aerospace and nuclear sciences. The Technology Utilization Program monitors a number of areas of research including the computer sciences. Because NASA has long played a leading role in supporting the computer sciences, numerous potential applications are frequently identified in the publications of the Technology Utilization Program (National Aeronautics and Space Administration, 1974).

#### IV. CONCLUDING STATEMENT

Siler (1974) indicated that there is a unanimity of opinion that the utility of computers in many facets of biological and biomedical research was now at the threshold of further advances because such uses are efficient and fiscally within reasonable costs based on the benefits derived. However, fiscal constraints over the next several years may significantly affect the extent to which further applications are pursued. As investigators become more aware of the potential of computers and computer systems in the biological sciences, research administrators should recognize the importance of support for these efforts (Davis, 1974).

The scope of computer science is so broad and complex that it is not possible for any one individual to grasp the large number of current research efforts in this field. Knowledge concerning computers and their capabilities is dispersed nonuniformly among a limited but growing number of experts. The discipline needs to be better organized in ways that will permit more effective education of life scientists capable of effectively utilizing computer science and technology.

Experts in the fields of computer science and artificial intelligence may be constrained by lack of knowledge in specific disciplines of biology. Biologists frequently are limited by lack of knowledge or facilities to manipulate concepts and data. There are a number of instances where careful collaboration between computer scientists and bioscientists have produced results which neither could attain individually. This points to the need for individuals who see as their scientific function the bringing together of computer techniques and biological problems. The constraints of network use and the personal computer lead to not only special problems but also unique benefits.

The individual, functioning as a scientific interfacier, should be capable of drawing on the resources of the network to synthesize for the bioscientist an environment which is a combination of capability in the personal computer of the bioscientist and capability in the network. In its full development, this type of synthesis would yield specially adapted programs in the personal computer for data collecting, purifying, and modeling; and, protocols for accessing network data searches, bibliographic searching, data validation and dissemination, and scientific result publication. In addition, such a program could produce documentation or publication mechanisms specifically generated to reflect the needs of bioscientists.

Saltzberg (1974) noted that much of the current work with computers in the life sciences was being done by groups of investigators skilled in

computer science. He concluded that a need exists for mechanisms and procedures whereby life scientists with limited computer knowledge could communicate with computer scientists who might offer solutions to particular problems in biological research. It has been pointed out repeatedly that even with knowledge of computer programming, hardware, and interfacing of equipment, choosing a system to meet specific research needs is difficult and requires interdisciplinary efforts and support.

Successful merging of the common interests of these two groups remains a challenge to both research communities (Saltzberg, 1974). This challenge includes both diversion from productive research activities and the funding of initial costs. There appears to be a need for more emphasis in biomathematics *per se* as a bridge between the biologists with little computer expertise and computer scientists who can apply their background in mathematical modeling to the life sciences.

Siler (1974) reiterated this thought in his concluding remarks at the symposium, stating that if the use of computers in the life sciences is to be successful as a science rather than an art form, then training of life scientists should include training in biomathematics and biostatistics as well as in computer techniques. Further, he noted that there is a critical need for training of biomathematicians knowledgeable in the basic principles of the life sciences, deterministic and stochastic mathematical techniques, and in computer methods because these scientists would be competent to train life scientists in these quantitative techniques.

There are a number of obstacles to such extensive and expensive training programs and to supporting financially further applications of computers to problems in biology. However, many experts have pointed out that it is not possible to assess what this lack of opportunity for such training or absence of fiscal support are costing the nation in terms of missed discoveries, inadequately interpreted data, and time lost in tedious and perhaps unnecessary experimentation.

## V. BIBLIOGRAPHY

- Ackley, P.R. and R.A. Nardello. 1973. Cross tabulations applicable to minicomputers. *J. Amer. Soc. Info. Sci.* 24: 437-444.
- Anonymous. 1974a. Choosing chemical routes. [Reviews work of Dr. E.J. Corey] *Mosaic* (Fall): 21-26.
- Anonymous. 1974b. Eye sampling study. [Reviews work of Dr. R.L. Mason] *Naval Res. Rev.* 28(8): 22-23.
- Birkhoff, G. 1975. Mathematics and computer science. *Amer. Sci.* 63: 83-91.
- Bogle, R., J. Zarembo and A.D. Bender. 1973. Factors to consider in the financial evaluation of minicomputers for R&D. *Lab. Manage.* 11: 58-62.
- Carman, G., P. Lemkin, B. Shapiro, M. Schultz and P. Kaiser. 1974. A real time picture processor for use in biologic cell identification. II. Hardware implementation. *J. Histochem. Cytochem.* 22: 732-740.
- Davis, R.M. 1974. Meeting the real costs of the computer as a research tool in the life sciences. *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* 33: 2409-2411.
- Dayhoff, M.E., editor. 1972. Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Washington, D.C.
- Dayhoff, M.E. 1974. Computer analysis of protein sequences. *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* 33: 2314-2316.
- Handler, P., editor. 1970. Digital computers and the life sciences. Pages 531-584 *in* *Biology and the future of man*. Oxford University Press, New York, N.Y.
- Heart, F.E. 1974. Networks and the life sciences: the ARPA network and telenet. *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* 33: 2399-2402.
- Langridge, R. 1974. Interactive three-dimensional computer graphics in molecular biology. *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* 33: 2332-2335.
- Ledley, R.S. 1965. Use of computers in biology and medicine. McGraw Hill Book Company, New York, N.Y. 965 pp.

- Lee, E.Y.S., R.K. Jain, E.K.C. Lee and B. Goettel. 1974. Environmental impact computer systems. CERL-TR-E-37, Army Construction Engineering Research Laboratory, Champaign, Ill. 113 pp. AD-787 295/5WJ, National Technical Information Service, Springfield, Va.
- Lemkin, P., G. Carman, L. Lipkin, B. Shapiro, M. Schultz and P. Kaiser. 1974. A real time picture processor for use in biologic cell identification. II. Hardware implementation. J. Histochem. Cytochem. 22: 725-731.
- Lindberg, D.A.B., Conference Chairman. 1974. Conference on the computer as a research tool in the life sciences: keynote address. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2311-2321.
- McInnis, B.C., S. El-Asfour and T.E. Bynum. 1973. Modeling and computer simulation to determine digoxin dosage. Pages 69-70 in IFAC-APS Int. Symp. Physiol. Control, Rochester, N.Y.
- Moraff, H. 1974. Sponsorship of computing research in medicine and biology. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2413-2414.
- National Aeronautics and Space Administration, Technology Utilization Program. 1974. Computer programs: special applications; a compilation. NASA SP-5967(01), National Technical Information Service, Springfield, Va.
- National Technical Information Service. 1974. Directory of computerized data files and related software available from Federal agencies 1974. NTIS-SR-74-01, National Technical Information Service, Springfield, Va.
- Newell, A. 1973. Artificial intelligence and the concept of mind. Pages 1-60 in Computer models of thought and language, K.C. Shank and K.M. Kolbye, eds. W.H. Freeman Co., San Francisco, Calif.
- Oppenheimer, C.H. and W.B. Brogden. 1975. Meeting report: international conference on environmental data management. Bioscience 25: 113 and 115.
- Orr, W.C. and H.J. Hoffman. 1974. A 90-minute cardiac biorhythm: methodology and data analysis using modified periodograms and complex demodulation. IEEE Trans. Biomed. Eng. BME-21(2): 130-143.
- Orr, W.C., H.J. Hoffman and F.W. Hegge. 1974. Ultradian rhythms in extended performance. Aerospace Med. 45: 995-1000.
- Patten, B.C., editor. 1971. Systems analysis and simulation in ecology, vol. 1. Academic Press, New York, N.Y. 597 pp.

- Patten, B.C., editor. 1972. Systems analysis and simulation in ecology, vol. 2. Academic Press, New York, N.Y. 592 pp.
- Prewitt, J.M.S. and M.M. Mendelsohn. 1966. The analysis of cell images. Ann. N.Y. Acad. Sci. 128: 1035-1053.
- Raub, W.F. 1974. The PROPHET system and resource sharing. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2390-2392.
- Reddy, D.R. 1974. Computer as a research tool in speed-understanding systems research. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2347-2351.
- Robinson, A.L. 1974. Computers: first the maxi, then the mini, now it's the micro. Science 186: 1102-1104.
- Sacco, W.J., R.A. Cowley and W.S. Capes. 1972. Pattern recognition analysis of shock trauma data and its application to Army problems. Report prepared for Army Material Systems Analysis Agency, University of Maryland, Baltimore, Md. 12 pp. AD-784 129, National Technical Information Service, Springfield, Va.
- Saltzberg, B. 1974. Session II-computer applications II: rapporteur's summary. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2420-2421.
- Siler, W. 1974. Conference summary and conclusions. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2420-2421.
- Stacy, R. and B.D. Waxman, editors. 1965 (vol. 1 and 2), 1969 (vol. 3) and 1974 (vol. 4). Computers in biomedical research. Academic Press, Inc., New York, N.Y.
- Vidal, J.J. 1974. Biocybernetic control and man-machine interaction: technical reports number one. UCLS-ENG-7430. Prepared under contract no. DAHC-15-73-C-0303, School of Engineering and Applied Science, University of California, Los Angeles, Calif. 91 pp. AD 777 720, National Technical Information Service, Springfield, Va.
- Wilson, K.R. 1974. Session V-networks and shared facilities: rapporteur's summary. Fed. Proc. Fed. Amer. Soc. Exp. Biol. 33: 2407-2408.

## VI. KEY INVESTIGATORS

Margaret O. Dayhoff, Ph.D.  
National Biomedical Research Foundation  
Georgetown University Medical Center  
Washington, D.C. 20007

Frank E. Heart, M.S.E.E.  
Director  
Computer Systems Division  
Bolt, Beranek and Newman, Inc.  
Cambridge, Massachusetts 02138

Donald A.B. Lindberg, M.D.  
Director  
Information Science Group  
University of Missouri Medical Center  
Columbia, Missouri 65201

Peter G. Lykos, Ph.D.  
Director  
Computer Center  
Illinois Institute of Technology  
Chicago, Illinois 60616

Arnold W. Pratt, M.D.  
Director  
Computer Research and Technology  
National Institutes of Health  
Bethesda, Maryland 20014

John A. Starkweather, Ph.D.  
Director  
Computer Center  
University of California  
San Francisco, California 94143

William Siler, Ph.D.  
Chairman  
Department of Biomathematics  
University of Alabama Medical College  
Birmingham, Alabama 35294

## VII. SCIENTIFIC CONSULTANTS

Richard J. Feldman, Ph.D.  
Computer Research and Technology  
Computer Center Branch  
National Institutes of Health  
Bethesda, Maryland 20014

Tyron E. Huber, M.D.  
6002 Roosevelt Street  
Bethesda, Maryland 20014

Frederick W. Hegge, Ph.D.  
Chief  
Department of Experimental  
Psychophysiology  
Walter Reed Army Institute of Research  
Walter Reed Army Medical Center  
Washington, D. C. 20012

Herbert Pollack, M.D.  
Route 1, Box 70  
Upperville, Virginia 22176

William Siler, Ph. D.  
Chairman  
Department of Biomathematics  
University of Alabama  
Medical College  
Birmingham, Alabama 35294