Chapter 3

Selected Case Studies in Information Technology Research and Development

Contents

Page

Introduction 55

Case Study 1: Advanced Computer Architecture 55
Findings 55
Changing Computer Architecture 56
Computer Architecture R&D 57
Critical Areas of Research 63
Manpower 65
International Efforts.. 65

Case Study 2: Fiberoptic Communications 67
Findings 67
Advantages 67
Commercialization Trends 69
United States R&D..... 71
Directions of U.S. Research 73
Research in Japan 74

Case Study 3: Software Engineering 75
Findings 75
Introduction 75
Software R&D Environments 76
Content and Conductor Software Engineering R&D...... 79
International Efforts in Software Engineering R&D...... 85

Case Study 4: Artificial Intelligence 87
Findings 87
Introduction 87
Artificial Intelligence R&D Environments 89
Content and Conduct of Artificial Intelligence R&D 96
Foreign National Efforts in Artificial Intelligence R&D 105

Tables

Table No. Page

11. Major Federal Advanced Computer Architecture R&D Projects 57
12. Milestones in the History of Computer Architecture 5 7
13. 1982 Federal Spending for Computer Architecture R&D..... 58
14. Federal Open Access Supercomputer Facilities 58
15. Summary occurrent and Planned Government Open Access Advanced Computer Architecture Software Development Facilities. 59
16. NSF Plans for Computer Research .,.. 60
17. Summary of New Commercial Supercomputer Systems Under Development 61
18. 1983 DARPA, DOE, and NSF University Funding of Advanced Computer Architecture 62
19. Major Government Sources of Artificial Intelligence R&D Funding. 94
20. NSF Grant Awards in Artificial Intelligence 94
21. Strategic Computing Cost Summary . . 94
22. Representative Expert Systems 104

Figures

Figure No. *Page*

8. Computer Architecture Functional Elements 56
9. Optical Fibers Are Small, Lightweight and Versatile 68
10. The Software Life Cycle. 80
11. Artificial Intelligence 90
12. Artificial Intelligence Science and Engineering 92
13. Strategic Computing Program Structure and Goals 95
14. A Semantic Network.. 98
15. Markets for Artificial Intelligence 101

Chapter 3

Selected Case Studies in Information Technology Research and Development

Introduction

These case studies present a microcosm of the R&D process in information technology. The diversity of research and development efforts in information technology make it virtually impossible to examine in detail all of the many fields and disciplines. Therefore, four fields have been selected for detailed analysis. They are: Advanced Computer Architecture (ACA), Fiber Optics (FO), Software Engineering (SE), and Artificial Intelligence (AI).

These fields were selected for several reasons. They depict the wide range, diversity and inter-relatedness of the fields and applications of information technology. An analysis of them provides a broad overview of the scientific, technical and institutional issues in information technology R&D. These fields were chosen to include both hardware and software and both computer and communications technologies. They also illustrate the mix of long-term goals and near-term capabilities, thus reflecting the importance of these different perspectives in the R&D process. These four areas, moreover, are among those considered to be critical in determining the direction and pace of advance of information technology as a whole. The importance of advances in the four fields is exemplified by the development of government funded national R&D programs in Japan, Britain and the European Economic Community.

Case Study 1: Advanced Computer Architecture

Findings

- The technology of advanced computer design is critically important for the expansion of information technology in many fields. There is extensive R&D activity underway in universities, in industry, and in the National Laboratories aimed at producing and exploiting new computer designs; but there is considerable uncertainty over which new designs will be viable.
- Since their invention, electronic computers have been based on one architectural model, the von Neumann sequential processing architecture. The limits of computational speed achievable with this design are being reached; significant further increases in computer performance will require *parallel processing architectures,* which are inherently more complex to design and to use.
- VLSI (Very-Large-Scale Integrated Circuit) design facilities, based on powerful computers, are now being used to develop and test computer architectural designs, including parallel processors and special designs for certain dedicated operations, such as communications signal processing, image processing, and graphics.
- Software has been difficult to produce for computers of advanced, high-performance design. As the variety and complexity of architectural types increases, the difficulty of developing and integrating software will increase. Therefore, research in software development for novel computer designs will be critical.

38-802 0 - 85 - 5

- The Federal Government has been a major driver of advanced computer architecture because of its scientific and national security applications. The Government has influenced the evolution of computer design through the funding of R&D and the procurement of state-of-the-art systems. This leverage, though still important, is diminishing as commercial applications for advanced computers grow and as Federal requirements become a smaller fraction of sales.
- National programs in Japan, Great Britain, France, Germany, and the European Community have been established to pursue advanced computer R&D. The Japanese have recently demonstrated an ability to produce advanced architecture computers of competitive performance to American products.
- American companies and universities pursuing R&D in computer design face-difficulties:

  Universities cannot afford design and testing facilities for developing an architectural idea to the point where its performance can be assessed.

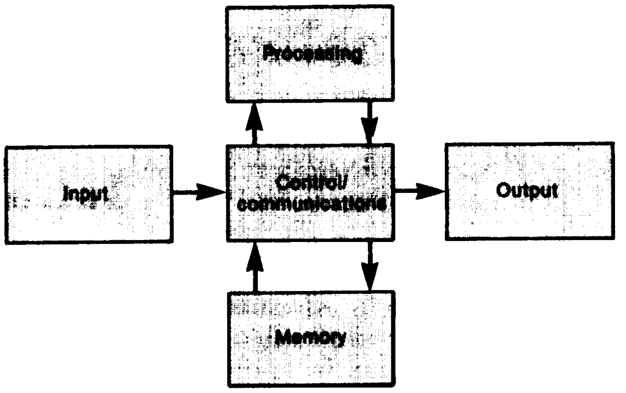
  Companies face large, risky investments in the design of new high-performance computer systems. Markets for novel machines are initially small and expand only slowly as new applications are exploited and software becomes available.

Changing Computer Architecture

Computer architecture is the internal structure of a computer, the arrangement of the functional elements that carry out calculations and information manipulations. (see fig. 8).

Since the early 1950s, all electronic computers (with a few exceptions) have been designed around one basic architectural model, the von Neumann machine, invented by mathematician John von Neumann. In this architecture, instructions *and data* are stored in memory, fetched one by one in sequential fashion, and acted on by the processor. Computer design is now changing, encouraged by two factors.

Figure 8.—Computer Architecture Functional Elements



SOURCE: Office of Technology Assessment.

First, the limits imposed by physical laws on the computational speed attainable with traditional computer design are being approached. Science and engineering demand continual advances in computational speed to increase the precision of calculations and to improve accuracy in models and simulations.[1] Sequential processing constitutes a severe restriction on the precision and completeness that these calculations and models can achieve in a reasonable amount of time. Therefore, computer designers are studying architectures that can make possible decomposition of large calculations into pieces for simultaneous processing by a number of computational units in *parallel.*

Second, as information technology is applied in more and more areas, special problems are encountered that impose unique demands on computer capabilities. Until recently, system designers have relied on software to apply the capability of von Neumann processors to problems. Now, it is possible to create special integrated circuits to address specific problems.

[1]These are the applications normally associated with so-called "supercomputers." Major current applications are in, for example, weather modeling and the simulation of nuclear weapons explosions. The reader is referred, for a discussion of the applications of and policy issues surrounding supercomputers, to *Supercomputers: Foreign Competition and Federal Funding* by Nancy Miller, Congressional Research Service, Issue Brief 83102, latest update July 12, 1984.

Computer architecture R&D is making possible the economical design of custom computer architectures for specialized applications including telecommunications and data acquisition signal *processors, image* and *graphics processors,* and *symbol processors* for the manipulation of nonnumerical information.

The impending changes in computer architecture promise cost-effective solutions to many problems, but they also challenge the designers, suppliers and buyers of computer systems. Designers will need to have more detailed appreciation of applications; computer vendors will be faced with more complexly segmented markets; and buyers will need to be more sophisticated in defining their needs and in choosing among a wider offering of products.

Computer Architecture R&D

Federal Government Involvement

The Government has had considerable involvement in advanced computer architecture R&D (see table 11), both as a funder and a performer of work. Major elements of the software development work for each generation of these systems have been done by the National Laboratories, especially Los Alamos (LANL) and Lawrence Livermore (LLNL).[2] Moreover, the impetus for the development of each successive generation of advanced architecture scientific computers has come predominantly from government demand for faster, higher capacity, and more sophisticated systems for weapons, intelligence, energy, and aerospace applications (see table 12). The National Labs still constitute the greatest concentration of users of supercomputers (see table 15).

The Federal Government has provided between $15 and $20 million in annual funding for advanced computer architecture R&D in recent years (see table 13). In addition to sponsoring research in universities and in industry, the Government has performed computer research at the National Labs.

[2]The first Cray-1 computer was placed in Los Alamos National Lab *without any software.*

Table 11 .—Major Federal Advanced Computer Architecture R&D Projects

Machine	Year delivered	Agency	Contractor	Major use
ENIAC	1945	Army	University of Pennsylvania	Ballistics calculations
NORC	1950	Navy	IBM	Ordinance research
—	1950	NSA	Sperry	Classified
CDC 1604	1959	NSA	Control Data	Classified
LARC	1961	AEC	Sperry	Nuclear weapons design
STRETCH	1961	AEC	IBM	Nuclear weapons design
CDC 6600	1964	LLNL	Control Data	Nuclear weapons
ILLIAC IV	1972	DARPA/NASA	Burroughs/University of Illinois	Aerodynamics
MPP	1983	NASA	Goodyear Aerospace	Image processing
s-1	–	LLN L/Navy	—	Signal processing

SOURCE Office of Technology Assessment

Table 12.—Milestones in the History of Computer Architecture

Class	Date	Typical machines	Major innovation
I	1953	IBM 701	Vaccuum tubes
II	1960	CDC 1604	Transistors
Ill	1964	IBM 360	1/0 processing
		CDC 6600	Freon cooling
Iv	1970	IBM 370	Integrated circuits
		CDC 7600	
v	1972	Illiac IV	Parallel processing
		TI ASC	Pipeline architecture
VI	1976	Cray-1	Vector processing

SOURCE Office of Technology Assessment

Table 13.—1982 Federal Spending for Computer Architecture R&D (millions of dollars)

Department of Defense	$12.0[a]
Department of Energy	2.9[a]
National Aeronautics and Space Administration	1.5[a]
National Science Foundation	1.2
Total	$17.6

[a]Estimates.

SOURCE: Department of Defense and Department of Energy numbers from the Office of Technology Assessment Workshop on Advanced Computer Architecture; NASA number from personal communication with Paul Schneck; NSF number from *Summary of A wards, FY 1982,* NSF Directorate for Mathematical and Physical Sciences, Computer Sciences Section,

DOE spent $645,000 at Los Alamos and $500,000 at Lawrence Livermore National Labs in fiscal year 1983 on two experimental research projects on advanced computer architecture hardware.[3] In addition, Los Alamos is leasing a Denelcor HEP-1 to experiment with parallel processing software concepts. The Navy, in conjunction with Lawrence Livermore, has been involved with the design and construction of an advanced architecture computer, termed the S-1 Project. The design is intended to handle signal processing tasks for Navy missions. Approximately $20 million has been spent over the last 4 years on the S-1 Project.[4]

Facilities and help in advanced computer applications development are provided to researchers in science and engineering fields to support the missions of several Government departments (DOD, DOE, NASA), and to further basic research (NSF). Seven Federal facilities provide limited open access to certain groups of researchers (see table 14).

The National Laboratories plan to add more supercomputers over the next few years, so it can be expected that Government scientists and engineers and contractors on mission agency work will have access to state-of-the-art large-scale computing facilities (see table 15). Academic researchers will have limited access to these facilities for work in fields related to agency missions (e.g., fusion energy, atmospheric and ocean sciences, and aerodynamics). These facilities also provide support for software development.

The Department of Defense, through the Defense Advanced Research Projects Agency (DARPA), has formulated ambitious plans for research and development in advanced computer-based systems. Included will be efforts to develop high speed signal processor architectures and to integrate numeric and symbolic processing in advanced computer architectures for use in intelligent weapons systems.[5]

In April 1983, the National Science Foundation organized a working group to study

[3]Edward Oliver at the OTA workshop on Advanced Computer Architecture, July 14, 1983.

[4]Personal communication from Sidney Fernbach, Consultant, Control Data Corp.

[5]This program, called "Strategic Computing, " is covered in some detail in the Artificial Intelligence Case Study later in this chapter.

Table 14.—Federal Open Access Supercomputer Facilities

Facility	Major system	Research users	Charges
NCAR	2 CRAY 1-As	Atmospheric and ocean sciences	No charge to NSF users. $2,200 per prime CPU hour for others
MFECC	2 CRAY 1 1 CDC 7600	Magnetic fusion energy community	No charge
LANL	Open 1 CRAY 3 CDC 7600s 1 CYBER 825	Government agencies, labs, and nonprofit institutions	$636 per prime CPU hour
NASA-Ames	CRAY 1-S CDC 7600	Computational fluid and aerodynamics	No charges to NASA grantees. $2,000 per CPU hour for CRAY
NASA-Goddard	CYBER 205	NASA-funded and NASA-project related	No charge to NASA grantees. $1,000 per CPU for others
NASA-Langley	CYBER 203	NASA and NASA-funded scientists	$1,300 per CPU hour
NASA-Lewis	CRAY 1-S	Principally aerodynamics related	No charge for NASA supported

SOURCE: *A National Computing Environment for Academic Research,* National Science Foundation, October 1983

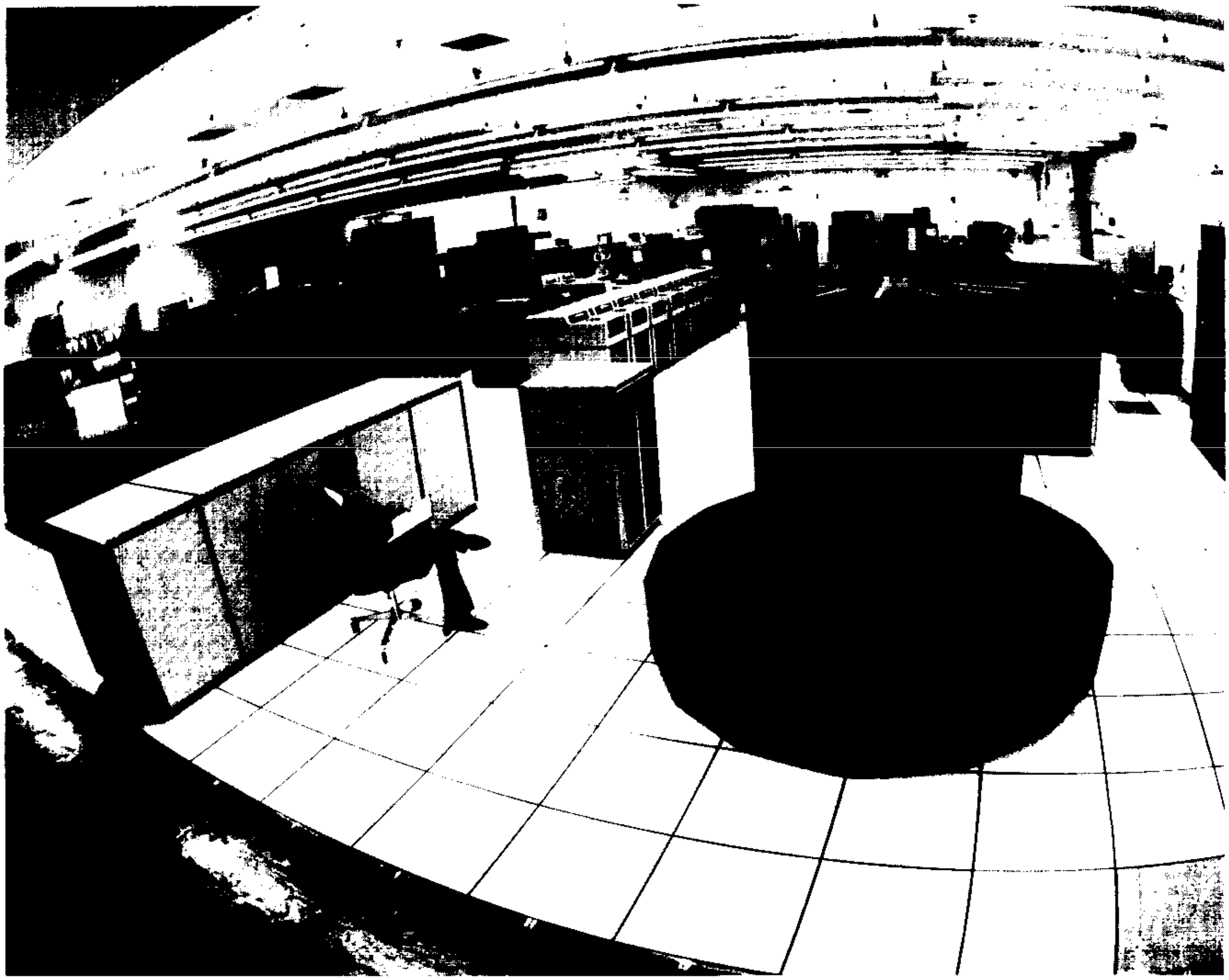


Photo credit: U.S. Department of Energy, Los Alamos National Laboratory

View of a part of the main computing facility at Los Alamos National Laboratory. CRAY 1 in foreground

Table 15.—Summary of Current and Planned Government Open Access Advanced Computer Architecture Software Development Facilities

Agency	Current systems	Number of users	Planned additions fiscal years 1984-88
National Science Foundation	2 Class VI	850	1 Class VII
Department of Energy	3 Class VI	6,400[a]	1 Class VI, 5 Class VII
National Aeronautics and Space Administration	3 Class VI, 1 special	3,000	1 VI, 1 VII, 1 special
Totals	8 Class VI, 1 special	10,250	2 VI, 7 VII, 1 special

[a]Includes researchers performing classified work.

SOURCE: *A National Computing Environment for Academic Research,* National Science Foundation, what was recognized as a critical scientific imperative. The report of that group stated:[6]

> Computing facilities have a decisive effect on the kind of research which is done by academic scientists and engineers. During the 1950s and 1960s the Government encouraged the growth of computing in research, research methods were transformed in discipline after discipline, and the United States enjoyed a large, ever-widening lead in quantitative research and modeling complex phenomena. In the 1970s Government support slackened and academic computing facilities no longer kept pace with advancing technology . . . Science has passed a watershed in using computers for research. Computers are no longer just tools for measurement and analysis but have become the means for making new discoveries . . . Academic research in computer architecture, computational mathematics, algorithms, and software for parallel computers should be encouraged to increase computing capability.

In response to this imperative, the NSF working group recommended an expansion in spending for academic research in advanced computers, and improved access to computing facilities including 10 new supercomputer facilities and special networks to make these systems widely available (see table 16).

The House Committee on Science and Technology considered R&D in advanced computers and access to powerful computer systems by scientists and engineers in many fields to be crucial elements in the advancement of science and technology. Accordingly, they approved a budget of $40 million in fiscal year 1985 for NSF's Advanced Scientific Computing initiative, thus doubling the administration's request for this program.[7]

Industry's Role

Three U.S. companies are developing next generation (Class VII) supercomputer systems. In addition, two Japanese companies (Fujitsu and Hitachi) have introduced new systems and a third (Nippon Electric-NEC) is developing a new supercomputer, planned for delivery in 1985 (see table 17).

Cray will introduce the Cray-2 in 1985. This will be a four processor vector machine.[8] The Cray-3 is scheduled for introduction in 1986. It will be an 8 to 16 processor vector machine with Galium Arsenide (GaAs) (see ch. 9) circuitry. Cray sees integrated circuit technology as critical. The Japanese are the major suppliers of state-of-the-art fast bi-polar memory chips, and one-half of the integrated circuits in current Cray machines are Japanese made.[9]

Control Data (CDC) spun off development work for its next generation advanced architecture machine to a new company, ETA Systems, which CDC capitalizes with $40 million for 40 percent ownership. This approach is being taken by CDC because small groups with dedication, entrepreneurial spirit, and a personal stake in the success of the project are considered important.[10] ETA Systems will spend $4 million to $6 million the first year on direct R&D costs. Plans are for the first demonstration machines to be available late

Table 16.—NSF Plans for Computer Research (million of dollars)

	Fiscal years 1984	1985	1986
Local facilities	$45.5	$ 90.9	$106.7
Supercomputers	14.4	70.0	110.0
Networks	2.1	7.3	11.5
Advanced computer systems and computational mathematics	8.0	20.0	33.0
Total	$69.9	$188.1	$261.2

SOURCE: *A National Computing Environment for Academic Research,* National Science Foundation.

[6] *A National Computing Environment for Academic Computing,* prepared under the direction of Marcel Bardon by the Working Group on Computers for Research, Kent K. Curtis, Chairman, July 1983, pp. 1-2.

[7] Authorizing Appropriations to the National Science Foundation, House Committee on Science and Technology, Report 98-642, Mar. 30, 1984, pp. 8-9.

[8] Vector computers have specialized architectures that achieve high speed calculation of mathematical formulas by treating entire arrays of data (vectors) as processable by single instructions, saving time on certain calculations that can be arranged as a series of vectors.

[9] L. T. Davis, "Advanced Computer Projects," presentation at the *Frontiers of Supercomputing Conference,* Los Alamos, Aug. 15, 1983.

[10] W. Norris, "A Conducive Environment for Supercomputers," banquet address at the *Frontiers of Supercomputing Conference,* Los Alamos, Aug. 18, 1983.

Table 17.—Summary of New Commercial Supercomputer Systems Under Development

Companv	Model	Maximum speed (M FLOPS')	Available
Cray Research	Cray-2	1,000	mid 1985
	Cray-3	NA	1985-86
ETA Systems	GF-10	10,000	1986-87b
	GF-30	30,000	NA
Denelcor	HEP-2	4,000	1985-86
NEC	SX-2	1,300	March 1985

NA—Not announced
aMFLOPS (million floating point operations per second) a measure of computer performance on high precision Calculations
bL M Thorndyke at the *Frontiers of Supercomputing Conference*, Los Alamos, August 1983.

SOURCE *The IEEE Committee on Super Scientific Computers.*

in 1986, and volume production of two machines per month is planned for 1987. The design will employ two to eight vector processors with the maximum eight processor version to sell in the range of $20 million.[11]

Denelcor, a former maker of analog computers, developed a parallel processing computer design (HEP-Heterogeneous Element Processor). Since 1982, this design has been available for sale or lease. Considered to be an experimental machine by users at facilities such as Los Alamos, this design is a step toward a new generation of computer architectures. Work is currently underway on the HEP-2, which should be competitive with Cray and CDC machines if component and software problems can be overcome. Moreover, the viability of Denelcor efforts will require higher sales than have so far occurred with the HE P-1 .[12]

In addition, other U.S. firms including computer companies (Digital Equipment, Hewlett-Packard, Honeywell, IBM, NCR and Sperry), telecommunications companies (Harris), semiconductor companies (Advanced Micro Devices, Intel, Monolithic Memories, Mostek, Motorola and National Semiconductor), electronics companies (Allied, Eaton, General Electric, RCA and Westinghouse) and aerospace companies (Martin-Marietta) are involved to some extent in research on parallel processing, data-flow or multiprocessor architectures.[13]

Industry representatives characterize the advanced computer architecture business as risky. The market is small: approximately 100 Class VI supercomputers have been installed worldwide as compared to tens of thousands of less powerful computers. Development costs are high: design tools include other advanced architecture machines for hardware simulation and software development.

The Role of Universities

Although as many as 50 U.S. universities are involved in advance computer architecture (ACA) research,[14] significant funding levels are available in only a few major schools. (See table 18.)

University research in advanced computer architecture is characterized by a series of stages of elaboration of a concept including: 1) theoretical paper and pencil work; 2) simulation of ideas on existing computer systems; 3) "breadboard" wiring of designs with off-the shelf components; and 4) full-scale engineering and construction of prototype machines in which state-of-the-art components, software and peripheral devices can be integrated to test the design on full-scale problems. OTA found that few if any projects currently under-

[11] L. M. Thorndyke, "The Cyber 2XX Design Process," presentation at the *Frontiers of Supercomputing Conference*, Los Alamos, Aug. 15, 1983.

[12] B. Smith, "Latency and HEP," presentation at the *Frontiers of Supercomputing Conference*, Los Alamos, Aug. 15, 1983.

[13] "Next-Generation Computing: Research in the United States," *IEEE Spectrum*, November 1983, pp. 62-63.

[14] The OTA workshop on Advanced Computer Architecture concluded that every major Computer Science and Electrical Engineering department has some interest.

Table 18.—1983 DARPA, DOE, and NSF University Funding of Advanced Computer Architecture

Institution	DARPA	DOE	NSF
California Institute of Technology	1,000[a]	550	
Massachusetts Institute of Technology	2,000	250	102
University of California, Berkeley	2,000[a]		
Stanford University	2,000[a]		
New York University		600	150
University of Illinois		155	200
University of Texas		107	120
University of Wisconsin		95	
Duke University			197
Other			570
Totals	7,000[a]	1,757	1,339

[a] Estimate

SOURCE DARPA numbers from a personal communication with Duane Adams; DOE numbers from Edward Oliver at the OTA Workshop on Advanced Computer Architecture; NSF numbers from Summary *of Awards: Fiscal Year 1983,* National Science Foundation, Division of Computer Research.

way in universities have funding to carry a design concept through to the final, systems engineering stage. Several projects will produce prototypes, but the elaboration of an idea into a system with software and supporting peripherals to demonstrate the performance and utility of the concept on real problems requires funding on the order of twice what the currently best funded projects receive.

The major distinction between efforts pursued in industry and in universities on ACA, aside from the commercial and product development orientation of industry work, is that more radical and advanced designs are being pursued in universities, whereas evolutionary designs are sought by industry. This is a result of the stake that industry has in the existing base of software and users and the need for upward compatibility of systems. University researchers have a greater ability to pursue revolutionary designs that could require completely new programming approaches and techniques.

Facilities Requirements

The increasing availability and capability of VLSI circuitry and computer-aided design tools are expected to have significant impact on computer design.[15] Prototype production time and cost will decrease. Both general purpose and custom application architectures are implementable in VLSI, opening opportunities for the testing and evaluation of many more computer architecture ideas. However, the initial investment required for VLSI design and fabrication equipment is very costly and will probably remain so. It is unlikely that most universities will be able to afford this equipment.

Other expenses associated with ACA research include computer-based simulation facilities. There may be a need for current generation supercomputers at universities to facilitate and test the design of new architectures. Bell Laboratories currently devotes most of its Cray computer to VLSI circuit design.

Supercomputers are also used by Cray and Control Data for software development, so that software is available when a new hardware design is completed. Universities would benefit from access to these software development tools, giving researchers the chance to test ideas experimentally. But here again, the costs associated with the procurement and operation of these design and computing resources are beyond the means of university project budgets. Some universities are forming consortia to spread the cost of microelectronics design and fabrication facilities across several institutions (see ch. 6). Shared supercomputer facilities are a key element to NSF plans for Advanced Scientific Computing.

[15] S. Trimberger, "Reaching for the Million-Transistor Chip, *IEEE Spectrum,* November 1983, p. 100.

Annual operating costs for government supercomputer facilities average more than $10 million." Only three U.S. universities currently operate such facilities and these are utilized at less than 50 percent of capacity. The reason for this low usage is the high cost of computer time on these systems, ranging from $2,000 to $3,000 per hour. University ACA research is therefore usually done on minicomputers which lack the capability of generating sophisticated, high-resolution graphics of importance to the design of integrated circuits.

Critical Areas of Research

Currently, the goal of most advanced computer architecture research and development is parallel computation.

The two major U.S. industrial developers of supercomputers, ETA Systems and Cray Research, and the Japanese manufacturers, Hitachi and Fujitsu, are pursuing parallel architectures in a conservative incremental fashion, contemplating the production of machines with up to 16 parallel processors by the late 1980s. Vector architecture will remain the dominant method for achieving fast numerical processing in these systems.

Universities, by contrast, are pursuing a number of methods of achieving "massively parallel" computation with upwards of 1,000 processors working in concert. One of the basic problems of computing in parallel is the requirement for communication and coordination among the individual processing elements when they are working on pieces of a single problem. Often the results of one process are required for another process to go forward. Several architectural solutions to these difficulties are under study, and extensive evaluation of different approaches must be done before their viability in real-world problems can be assessed. Detailed simulation of concepts and testing of prototypes is required, and present university funding is inadequate to support such work.

There are currently more than 50 concepts for parallel processing architectures under consideration in academic and industrial institutions. However, there are no standard metrics for comparing the performance of different architectural designs or the software to be used on parallel machines. Nor is it likely that any *one* metric could fully measure differences in performance, since different applications place different demands on systems. The development of such metrics and the establishment of suitable test facilities for implementing standard design evaluations are critical issues in advanced computer architecture. The Federal Government may have a role in this area by setting voluntary standards for computer performance measurement, and by providing facilities for testing.

Thus far little attention has been devoted to the problems of *symbolic,* as opposed to numeric processing architectures. In the past, the von Neumann architecture has been used for both kinds of computations; the focus of advanced computer architecture R&D has been on computers for "number crunching" applications, or high precision calculation, simulation and modeling for science and engineering. The increasing importance of artificial intelligence is encouraging the design of special machine architectures, both to ease the programming of artificial intelligence applications, and to speed the processing of symbolic computations. Several companies are now producing machines for artificial intelligence, and the list is expected to grow. '7

Three other areas of technology are critical to the development of advanced computer architecture systems: integrated circuits, circuit packaging, and algorithm and software design.

Integrated Circuits (IC)

An order of magnitude (1 OX) increase in computer speed is expected from improvements in IC materials and manufacturing *techniques.* Silicon will remain the dominant IC substrate material through 1990 because the technology

[16] *A National Computing Environment for Academic Research,* op. cit., p. 22.

"See the case study on Artificial Intelligence in this chapter.

is well understood and the practical limits of device density, speed and power consumption have yet to be reached. Silicon will be the basis of a growing set of special purpose VLSI architectures. Gallium Arsenide (GaAs) digital logic and memory circuits are growing in importance, and will be used in the Cray-3.[18] University based research in chemistry, physics and microelectronics are expected to make significant contributions to increased integrated circuit capability.

Packaging

Interconnections among the logic elements on some complex chips occupy over half of the useable chip area, and affect the performance speed of chip functions. Currently, three sandwiched layers of interconnection within a chip are typical, and it is expected that as many as 12 layers will be common in a decade.[19]

As chips have become more complex, containing greater numbers of logic elements, the number of "pins," or inputs and outputs, required for communication with them has grown. The connection of sets of chips has thus become more complex, and the difficulty of simultaneously housing and powering chips, and dissipating the waste heat from chip sets is forcing advanced computer designers to find more sophisticated methods of packaging them.[20]

Cryogenic liquid cooling equipment is required for most existing and planned supercomputers. Facilities must be provided for the refrigeration and storage of the coolant. And the size, weight, and reliability of the cooling equipment must be considered in the purchase and use of these systems.

Packaging is also a critical factor in supercomputer manufacturing costs. Cray machines are currently hand wired. In an effort to reduce costs, the Japanese are developing designs that lend themselves to automated manufacturing procedures.

Software and Algorithms

The lack of applications software for supercomputers has been a significant barrier to their adoption and use.

One half of recent Cray Research R&D funds have reportedly been devoted to software development,[21] and the vectorizing FORTRAN compiler, a software program that helps prepare standard FORTRAN code for execution on the Cray vector architecture, has been a major factor in the commercial success of the Cray-1 line." Users of vector computers, including the Cray, are quite pleased to obtain 20 percent of the maximum rated speed of these machines on typical problems.[23] Work is continuing in industry and universities to develop software to make vector machines more effective and easier to use, and this work will be of critical importance through this decade.

The introduction of parallel processing designs and the proliferation of special purpose computer architectures will make software production and design more complicated.[24] The creation of new high-level languages that are more easily understood by users, and other tools and programming support environments that facilitate the expression of logical, symbolic, mathematical, scientific, and engineering concepts in computable form, could greatly

[18] L. T. Davis, op. cit.

[19] J. A. Armstrong, "High Performance Technology: Directions and Issues," presentation at the *Frontiers of Supercomputing Conference*, Los Alamos, Aug. 15, 1983.

[20] MCC is devoting some of their initial efforts to packaging technology; and ETA Systems sources estimate that 60 percent of the R&D effort for the GF-10 will be in packaging. ETA is planning to use liquid nitrogen cooling to obtain a doubling in speed from CMOS silicon integrated circuits (L.M. Thorndyke at the Los Alamos Conference).

[21] Rollwagen, op. cit.

*'Nippon Telephone and Telegraph, the Japanese state telecommunications monopoly, has chosen a Cray-XMP over recently introduced Japanese supercomputers of comparable or superior speed, reportedly because of the software, and the experienced team of field representatives, available from Cray. "NTT Picks Cray Super CPU," *Electronic News*, Oct. 10, 1983, p. 87.

[23] David Kuck, at the OTA workshop on Advanced Computer Architecture, July 14, 1983.

*'Paul Schneck, at the OTA workshop on Software Engineering, Nov. 17, 1983.

expand the utility and lower the costs of operating advanced computer systems.[25]

In order for the speed potential of advanced computers to be used, problems must either be programmed to take advantage of the computer design, or the architecture must be designed to handle the unique characteristics of the problem. The mediating factor between the problem and the architecture is the algorithm, or the structured procedure for solving the problem. In the future, computer designers will have to be more cognizant of computational algorithms and the effects of computer architecture on programming and problem solution, and thus they will need greater knowledge of applications. Similarly, designers of complex programs, especially scientists, mathematicians, and engineers, will need to have a greater appreciation of the inherent capabilities and limitations of particular computer architectures as they become more dependent on advanced computers in their work. Collaboration between the *users and designers* of future computer systems is critical to both the utility and the commercial success of advanced architecture computers.[26]

Manpower

There is a shortage of people capable of designing software for advanced architecture computer systems. In particular, people skilled in the design of software and software tools for use in sophisticated scientific and mathematical applications are scarce.[27] There is a need for people who understand scientific problems in a range of disciplines, and who can design and implement computer systems to solve those problems.

Attracting talented faculty to train the next generation of computer researchers is a problem. The difficulty results, in large measure, from the uncompetitive salaries and low job mobility offered by universities. The problem is expected to become acute as demand for computer architectures employing symbolic processing and artificial intelligence capabilities increase. (See Artificial Intelligence case study.)

International Efforts

Japan

Two Japanese firms, Fujitsu and Hitachi, have introduced advanced architecture computers whose performance is competitive with the fastest available American supercomputers. Early copies of these machines have been installed in three Japanese universities. A third company, Nippon Electric, has announced plans to introduce a supercomputer in 1985.

The Japanese Government is funding two national efforts in advanced architecture research and development: "High Speed Computing Systems for Science and Technology' and the "Fifth Generation Computer System" program.

The Electrotechnical Laboratory of the Agency of Industrial Science and Technology (AIST), an arm of the Ministry of International Trade and Industry (MITI), is managing a 10 year project (January 1981—March 1990) called "High Speed Computing Systems For Science and Technology. " It is focused on microelectronics research and development (GaAs, Josephson Junctions and High Electron Mobility Transistor devices), parallel processing systems with 100 to 1,000 computing elements, and systems components, including mass storage and data transfer devices, to support high-speed scientific and engineering calculations. Total government funding will be on the order of $100 million. A consortium of six major Japanese computer companies (the Technology Research Association) has been formed to conduct much of this work in industrial laboratories on "consignment" from MITI.[28]

[25] M. B. Wells, "General Purpose Languages of the Nineties, " presentation at the *Frontiers of Supercomputing Conference*, Los Alamos, Aug. 17, 1983.

[26] OTA workshop on Advanced Computer Architecture, July 14, 1983.

[27] This point was emphasized in one applications area in Particular, telecommunications (Paul Ritt at the OTA workshop on Advanced Computer Architecture, July 14, 1983).

[28] "Super Computer-High Speed Computing Systems for Science and Technology, " *Science & Technology in Japan*, October-November, 1982, p. 16.

The "Fifth Generation Computer System" program is managed by the Institute for New Generation Computer Technology (ICOT). The program is described in more detail on page 105, as part of the artificial intelligence case study.

Great Britain

Great Britain has reacted to the Japanese efforts (in particular the Fifth Generation project) by establishing a national plan for research in advanced computer systems, known as the Alvey Programme for Advanced Information Technology. It is discussed below in the Software Engineering and in Artificial Intelligence Case Studies. Researchers in British universities have made significant contributions to advanced computer architecture research. For example, the University of Manchester has an advanced prototype of a "dataflow" architecture machine.[29] Work is also being pursued in industry.[30] Inmos, Ltd. has developed the Transputer, a device which combines processing and communications functions on a single chip. This device has been specifically designed for the connection of a number of units to achieve concurrent processing.[31]

France

France has considerable interest in advanced computer architecture research and development. Two industrial companies, Cii-Bull and CGE, as well as seven government funded institutions including five universities are pursuing work in this area.[32] Three "supercomputer" projects are expected to produce machines by 1985-88, but these systems are not likely to be speed rivals of American and Japanese systems of similar vintage.[33]

West Germany

West Germany has four universities and six industrial companies working on advanced computer architecture R&D.[34] The government is providing about $4 million per year for research at the universities on parallel processing. A project at the Technical University in West Berlin has received a grant from the Ministry of Research and Technology to develop a full-scale prototype.[35]

The European Community

The Commission of the European Communities (EC) has initiated ESPRIT (European Strategic Program for Research and Development in Information Technology) to pursue information technology R&D on a cooperative basis with industry, universities, and the governments of the EC countries pooling their efforts. Four institutions (three Belgian and one French) have thus far announced plans to study parallel processing with ESPRIT funding.[36] The proposed ESPRIT plan calls for 'he development and use of computerized facilities to study new computer designs.[37] (For more information on ESPRIT, see ch. 7.)

---

[29] A. L. Davis, "Computer Architecture, " *IEEE Spectrum*, November 1983, pp. 98-99.

[30] Five have been identified, see "Next-Generation Computing: Research in Europe, " *IEEE Spectrum*, November 1983, pp. 65-66.

[31] "Transputer Does Five or More MIPS Even When Not Used in Parallel, " Iann Barron, Peter Cavill, David May and Pete Wilson, *Electronics*, Nov. 17, 1983, p. 109.

[32] "Next–Generation Computing: Research in Europe, " *IEEE Spectrum*, November 1983, pp. 64-65.

[33] *Report of the IEEE Super Scientific Computer Committee*, Oct. 11, 1983.

[34] *IEEE Spectrum*, November 1983, op. cit., pp. 67-68.

[35] "Western Europe Looks to Parallel Processing for Future Computers, " *Electronics*, June 16, 1983. p. 111.

[36] A. L. Davis, op. cit.

[37] "Proposal for a Council Decision adopting the first European Strategic Programme for Research and Development in Information Technologies (ESPRIT), Commission of the European Communities, June 2, 1983, p. 24.

Case Study 2: Fiber Optic Communications

Findings

- Fiber optic communications technology is an important export (e.g., a $3 billion world market projected for 1989), and is a key element in improved productivity and reduced costs in telecommunications systems.
- Fiber optic communications technology is developing rapidly and the potential benefits from continued research are extensive.
- The most significant research is concentrated in only a few large firms and universities, in part because of the expense and the required long term commitment.
- Increased Government funding in this technology-now at a low level—would enlarge university participation and accelerate technological advances.
- The scarcity of trained research and development personnel is a continuing handicap. A large proportion of university researchers in fiber optics are foreign nationals and many return to their native land after completing graduate studies.
- Research conducted in Japan and Europe is among the world's most advanced and exchange of research information internationally among colleagues is critical to scientific advancement.
- Japan is the world's leader in several key aspects of the technology, and is a strong competitor to American firms.

The first successful transmission of voice signals using energy from the Sun was accomplished in 1880 by Alexander Graham Bell and Sumner Tainter using a device called the photophone, but was abandoned because weather made the system unreliable.[38] Since 1970, interest has resumed in using light energy for telecommunications, in the form of fiber optic communications because of two technological advances: the laser and light transmission through low loss silica glass fibers.

Advantages

Fiber optic communication is the transmission of light signals through transparent glass or plastic fibers, where the signals are generated by lasers or light emitting diodes (LEDs) and received by photodetectors, which convert light signals to electrical signals. The major components of fiber optics technology are the fiber cables and connectors, transmitters and receivers, and repeaters, or regenerators,[39] which amplify and reconstitute the signals periodically along the fiber.

There are several properties of fiber optic communications that make them attractive for telecommunications applications:

- large bandwidth, meaning that large amounts of information (voice conversations, computer data, graphics) can be transmitted rapidly. For example, a quarter-inch diameter optical cable with two fibers carries as much data as a 3-inch copper cable with 20,000 wires.[40] (See fig. 9.) Conservatively, the capacity of a single pair of fibers currently available commercially is about 4,000 voice grade circuits in field applications and about 400,000 voice grade circuits under controlled laboratory conditions;
- less susceptibility than copper wire to radio frequency interference, providing less cross-talk, higher quality signal transmission, and immunity from electromagnetic pulse (EMP) effects-characteristics of value for both civilian and military uses;
- lower loss of signal strength, meaning that fewer repeaters are needed;
- resistance to "noninvasive" or covert wiretaps;

[38] Report on Research at the University of Arizona, vol. 1, No. 1, fall 1983, p. 11, published by the Research Office, University of Arizona.

[39] In standard copper telephone wires, Signal regeneration is required at about one-mile intervals. Repeaters add significantly to the installation and maintenance costs of transmission systems.

[40] *High Technology*, "Fiber Optics: Light at The End of the Tunnel," March 1983, p. 63.

Figure 9.—Optical Fibers Are Small, Lightweight, and Versatile

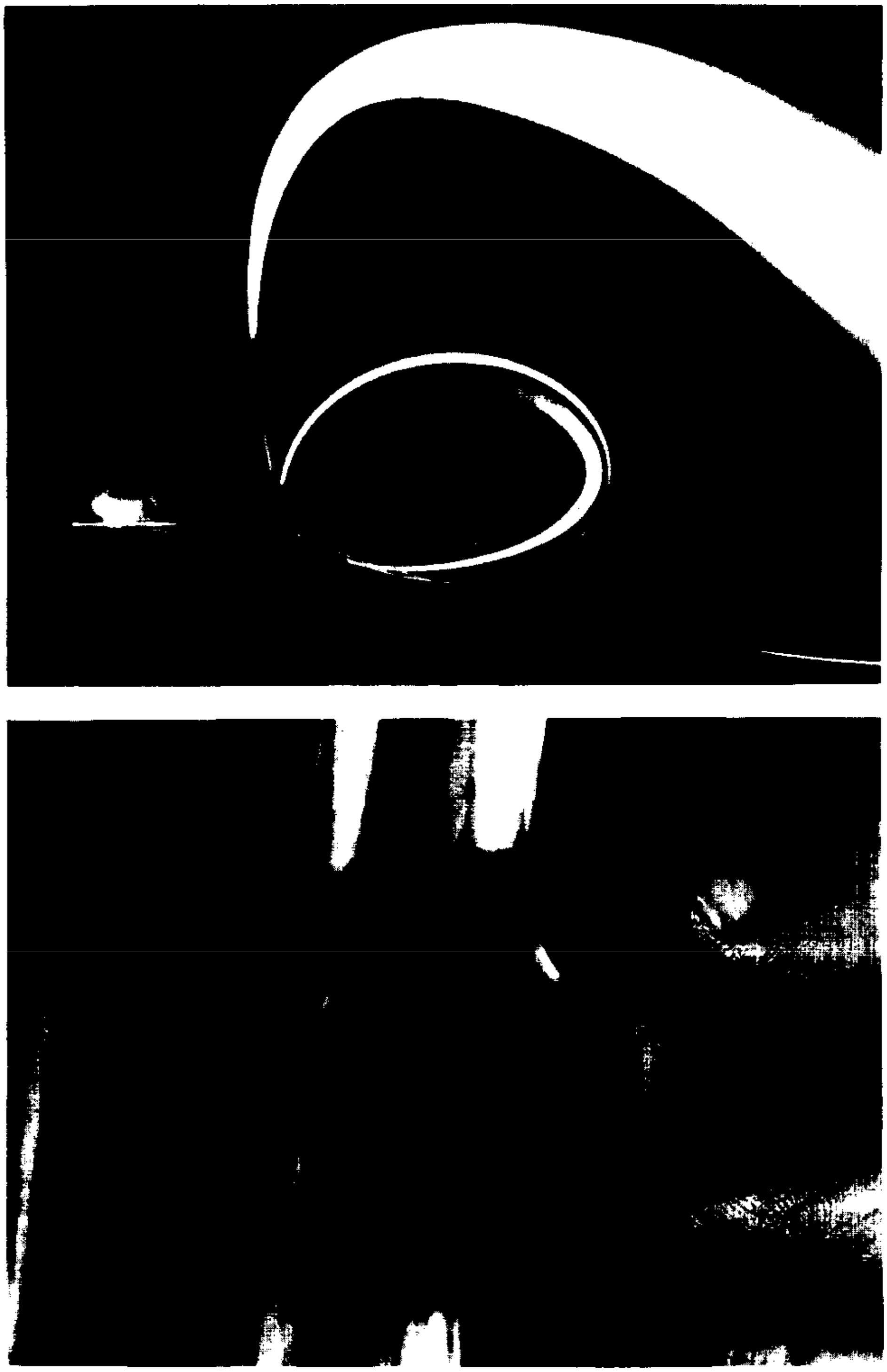


Photo credits: AT& T Sell Laboratories

- small size and low weight, factors that contribute to ease of transport and less need for underground and building duct space;
- declining cost compared to other terrestrial telecommunications technologies.[41]

Commercialization Trends

The technology has progressed rapidly into the commercial marketplace since 1966, when researchers[42] first proposed the possibility of purifying the glass used in optical fibers to reduce losses in signal strength. In 1970, the first low loss fiber was produced by Corning Glass Works, and by 1977 prototype systems were being installed by AT&T and General Telephone and Electronics Corp. in the United States. Today, many developed countries have fiber optic communication systems in operation or plan to install them.

The United States, Canada, Western Europe, and Japan accounted for an estimated 96 percent of a $550 million world market in fiber optics communications equipment in 1983. The world market was approaching $1 billion for 1984,[43] and is projected to expand to $3 billion by 1989,[44] as countries satisfy their telecommunications system expansion and replacement needs with the increasingly cost-competitive fiber optic communication systems rather than with microwave radio, copper twisted wire pair, and coaxial cable.

Applications in the United States

Telecommunications, the major market (85 percent)[45] for fiber optics, can be described in four segments: long distance; interoffice trunks that connect telephone central offices; local feeder lines; and local area networks. Long distance applications are currently the most cost effective for fiber optic communications, and comprise the vast majority of current use in the United States.

Interoffice trunking provides links between intracity telephone facilities. Feeder lines include intracity transmission links between carrier facilities and subscriber distribution points, while local feeder lines extend to subscriber locations. Local area networks (LANs) serve limited communities, for example, within a building or a building complex. They are just emerging and will become an important market for fiber optics.

There are a growing number of installations of fiber optic systems in long distance telecommunications. Among these are AT&T's Northeast Corridor route. In 1983 a line connecting Washington, DC and New York City was inaugurated. In 1984 this line was extended to Cambridge, MA, and Richmond, VA. The total Northeast Corridor line will use 45,000 miles of fiber over the 750 mile route. The first phase of a west coast route, which will eventually extend from Los Angeles to Oakland and Sacramento, has been completed. A transatlantic cable (TAT 8), engineered to carry 40,000 voice circuits at a cost of less than half that of its predecessor, is scheduled for installation in 1988. By March 1983, AT&T had already installed over 100,000 miles of fiber and projections are for another 300,000 miles by 1990.[46]

Other firms planning major systems include United Telecommunications, Inc., with its 23,000 mile, $2 billion lightwave network to be completed by 1987.[47] Southern New England Telephone Co., in a joint venture, will route its system through 20 States along railroad rights of way; MCI, with a 4,000 mile system intended to serve the east coast;[48] and Cable and Wireless, a British company, with

[41] This comparison is based on the relative cost per channel-mile, which is the number of voice circuit equivalents (channels) multiplied by the distance of the transmission link.

[42] Kao and Hockman, ITT Standard Telecommunications Laboratories, England.

[43] D. G. Thomas, "Optical Communications," Research and Development, June 1984, p. 203.

[44] Signal, September 1983.

[45] The remaining 15 percent of the noncommon carrier applications are said to be in vehicular, industrial control systems, and in CATV. *High Technology*, op. cit.

[46] Optoelectronics Supplement to *Electronic News*, "Fiber Optics Market Still Baffles Suppliers," p. 5, Apr. 4, 1983.

[47] "23,000 Mile Fiber Network Planned by United Telecom," *The Journal of Fiber Optics*, June 1984.

[48] *Fiber Optics Industry Service: Competitive Environment*, Gnostic Concepts, Inc., 1983, pp. 1-2.

a 560 mile network that will link major cities in Texas.

Fiber optics is already an attractive replacement for copper cable in some *local* telecommunications applications: 1) in trunk lines between telephone central (switching) offices where the average 10-mile distance can be spanned by optical cables without repeaters, thus installation and maintenance costs are lower, and 2) in local feeder lines in large cities where crowding in underground utility ducts is a growing problem.

Local area network applications are expected to grow rapidly according to some industry projections, particularly as user requirements for bandwidth increase to the 10 to 100 million bits per second (Mbs) range. The past slow growth of fiber optic applications in LAN systems is due to the variety of technical needs among different customers, and a lack of uniform technical standards. In contrast with long distance communications systems, which have traditionally paid extensive attention to technical standards development, a Federal telecommunications standards committee only recently (July 1984) held its first meeting to develop Federal guidelines for LANs. Once standards issues are settled, growth in the use of fiber optics within buildings and building complexes is expected to be rapid.[49]

Other factors limit the adoption of fiber optic communications technology. The large base of installed copper wire in the AT&T plant (some 827 million miles) is likely to be replaced slowly. The large capacity (up to 100 television channels) of some CATV systems and the fact that this expensive investment in coaxial cable has been made quite recently in many cities, suggests that fiber optics will not be widely used for cable television for some time. Advances in nonoptical transmission techniques are enabling considerable increases in the information carrying capacity of copper wire pairs.

[49]*Aviation Week and Space Technology*, **"Promising Future Seen for Optical Fibers," pt 1, Oct. 12, 1983, pp. 44-77,**

In applications other than communications carrier uses, such as aerospace and military systems, fiber optic technology is already being exploited. These are primarily in command, control, and communications applications including guidance and control systems for aircraft, spacecraft, and missiles; optically multiplexed data bus transmission systems; electronic warfare and sonar applications; and advanced instrumentation systems. The imperviousness of fiber optics to electromagnetic interference, along with light weight, small size, and high information rates, make it of special value in aerospace and military applications.

Foreign Applications

Installation of fiber optic communications systems have been completed, or are planned in a number of countries. A small sample of these includes:

- *West Germany*. By late 1984, 10 broadband integrated fiber optic local area networks will be built in Berlin, Hamburg, Hanover, Dusseldorf, Nuremburg, and Munich, called the BIGFON (Broadband Integrated Fiber Optic Local Network) network. In addition to having access to the public switched telephone network, subscribers will also be able to access the integrated telex and data network and receive radio, television, telephone, and full motion picturephone. The long-term goal is to include all telephone subscribers in the nation.
- *France*. Several fiber optic systems are being installed by the French. The government has decided to upgrade the nation's antiquated telecommunications network by leapfrogging toward the most advanced technology available-especially fiber optics. They also plan to install a submarine cable between France and Corsica in 1985, providing over 7,600 channels operating at 280 Mbs.
- *Japan*. One of several projects being undertaken is an 80 km fiber optic route in the suburbs of Tokyo. The system operates at 400 Mbs, providing video confer-

encing and facsimile services. In addition, the Japanese have installed a broadband network for data communications in Tsukuba to facilitate scientific and technical communications.

An Intelligent Network System (INS) is being developed by Nippon Telephone and Telegraph's Yokosuka Laboratory near Tokyo. Services include voice, data, CATV, still and motion pictures, facsimile, TV conferencing and high resolution TV. The INS is expected to make extensive use of fiber optic technology. Another project involves a 45 km undersea fiber optic cable south of Tokyo, operating at 400 Mbs. The system will be extended over a 1,000 km route between islands.

- *United Kingdom.* British Telecom, the nation's telecommunications authority, has committed itself to using only fiber optics in the trunk network from 1984 on. By 1990, half the trunk network will be fiber optic systems.

  Mercury Communications, a communications carrier in limited competition with the government telecommunications authority, has begun installing an intercity fiber optic network using the British rail rights of way.

  The British plan to construct a nationwide cable telecommunications system for the delivery of television programming, FM radio, pay television, and text. An experiment with a small number of homes is being conducted using fiber optics technology. This is considered to be the prototype for the national system.

United States R&D

Much fiber optics R&D in the United States is being conducted by a few large companies: AT&T Bell Laboratories; Corning Glass Works; ITT; and, to a lesser extent, GTE; a few universities; and some smaller companies. Many of the commercial products now available are a result of R&D performed by Bell Labs. AT&T began funding optic communications research in 1960 and related laser research in 1958.

There is concern that many U.S. companies have been inclined to undertake research only where the prospective payoff is likely to occur within a very few years. This attitude has the effect of shifting investigation away from promising areas such as research on infrared systems, where another 5 to 10 years of work may be required before commercial products become available. Thus, the importance of stable Federal funds for basic research is underscored by short-term planning within industry.

The expense of research in fiber optics makes it difficult for small firms to play a role in R&D, except in some areas of product commercialization. In addition, there is a tendency for equipment purchasers to prefer vendors of complete lines of components, which works to the detriment of small firms. As a result, small specialty firms often must rely on DOD for research funds, on takeovers by larger firms, or on venture capital in order to remain competitive. Regardless, these firms play an important role in the technology by providing innovative ideas and products, by serving as conduits for the commercialization of university-based research, and by filling niches that might not be attractive to larger firms.

Because of the considerable expense associated with research in fiber optics technology, and the low level of available funding, few universities have major research programs. The importance of cost is illustrated by the $500,000 or higher cost of Molecular Beam Epitaxy equipment (needed for growing alternating epitaxial layers on semiconductor light sources and detectors) and $200,000 for fiber drawing equipment (needed to produce fiber and to experiment with different fiber designs) required to perform research. Very little university research is focused on glass fibers, but instead is directed toward light sources and detectors. The principal universities with major research programs are the California Institute of Technology, the University of 11-

38–802 O - 85 - 6

linois, Stanford, Cornell, Princeton, and the Massachusetts Institute of Technology.

Government Funding

University research has been supported for decades through NSF funding for the support of research and, more recently, training and laboratory equipment. NSF supports several universities, such as the University of Arizona in Tucson, Northeastern University, and Cornell University's Submicron Facility. Although NSF funding levels in this technology are not large $1.75 million in 1983–they represent a consistent source of funds, often supporting fundamental research with very long term potential payoff. Additional levels of funding would likely accelerate the rate of technological advance.

Between 18 and 20 projects are being funded by NSF in about a dozen universities. Many of these are concerned with advancing theoretical knowledge in areas such as laser technology, the development of pioneering optic and optoelectronic integrated systems and bistable optical switching devices, research into infrared lasers and detectors, and the application of integrated optical interface circuits in local area networks at gigabit (billions of bits) per second data rates. Another $300,000 of NSF funding is available for upgrading university laboratory equipment.

The DOD funds fiber optics research through mission-oriented procurements. Approximately 95 percent of the research is carried out by industry. DOD has some $32.4 million committed to the development of cables and connectors, light sources and detectors, radiation effects exploration, and to sensor and communications applications. An estimated $12.6 million of this is committed to research, principally applied research. In addition, the military departments allocated about $22 million for fiscal year 1984 among seven procurement programs for applications ranging from surveillance, shipboard and long distance communications, and helicopter flight control systems.

Cooperative Research

NSF has several activities directed at encouraging cooperative research between universities and industry, and transferring technology into industry. One of these noted in chapter 2, the Industry/University Cooperative Research Centers program, provides planning grants to aid universities in establishing industry affiliations and support for specific scientific or engineering technologies.

NSF has awarded a $75,000 grant for planning purposes to the University of Arizona at Tucson, Optical Sciences Center. The University held its first meeting with industry in early 1984 to begin determining mutual interest in specific areas of cooperation and industry support. Most of this research is expected to be directed toward long term projects in physics and materials science with potential applications in optical logic circuitry and optical computers, with limited attention to fiber optics.

Another NSF activity funds specific projects where research is undertaken cooperative ly by universities and company investigators. Funding is at levels of about $100,000 per year over a 2 to 3 year period, on the average. One of the funded projects is being undertaken jointly by Bell Laboratories and the University of Arizona, Optical Sciences Center. This project's long-term research is in high speed optical, bistable switching devices operating at picosecond (1 trillionth of a second) rates.

The current level of cooperation in research is not extensive, but holds promise for broadening the base of research. Among the problems and issues to be worked out are finding ways for competing firms to share research data and establishing a balance between university investigators' interest in long-term research and companies' desire for short-term payoffs.

Manpower and Industry Support

Fiber optics research requires training in both physics and in electrical engineering. Because universities, in general, do not provide this cross disciplinary training at the bachelor and masters degree levels, companies hiring recent graduates provide supplementary in-house training. Some companies also establish an affiliation with universities. An example is the affiliation between Corning Glass Works and the University of Rochester Institute of Optics, in which the company provides some faculty and funding. Another example is the support from United Technology Research Center to the Rensselaer Polytechnic Institute which is providing real estate for new facilities and adjunct faculty to teach specialized engineering courses.

The University of Rochester is the only school in the United States to offer an undergraduate degree in optics. Only three schools in the nation offer graduate degrees in optics: The University of Arizona in Tucson, the University of Rochester, and Northeastern University.

A factor aggravating the availability of trained Ph.D. graduates in this field is that a large proportion of the students are foreign nationals. Industry argues that immigration laws make it difficult for the student-graduate to remain in the United States after graduation, although many would prefer to stay and perform research.

Directions of US. Research

Among the areas of R&D focus identified during the course of this study are:

Fibers

The early fiber optic cables put in use were of the *multimode* type, in which lightrays enter the fiber at a variety of angles and travel through the core of the fiber reflecting from its inner refractive surfaces. However, *single mode* fiber technology, where lightrays follow a single direct path along the fiber core, has important advantages. Single mode fibers have greater information carrying capacity and allow a tenfold increase in the distance between repeaters for regenerating signals. Today's single mode fiber systems are able to transmit, without repeaters, up to 200 Mbs (million bits per second) over 80 to 100 km. (This information rate is sufficient to carry simultaneously a video channel, high fidelity audio, data, and many telephone calls.) In laboratory tests, this performance has been exceeded by about 10 times, suggesting far greater potential gains from research. Improvements from research are expected to continue in both multimode and single mode fibers.

Improvements have been made in lowering attenuation (losses in signal strength) in both types of fibers by a factor of 100 since 1970, primarily due to development of methods to reduce impurities in the fibers.

Activities are being directed toward further improvements in optical fibers. These include research into different cross sections for fiber cores, such as circular, triangular, and oval, as investigation into new materials such as plastics, and improvements in fiber splicing techniques. Research into plastic materials is about at the stage of 1975 era research in glass fibers, and promises even lower cost, more durable fiber materials for some applications.

Longer wavelength (1.7 and 4.0 micron) material for fibers is also receiving attention, as these show promise of decreased attenuation by a factor of 10 to 100 over that of currently available fibers, with long-term prospects for transcontinental or transoceanic transmission without the need for repeaters.

Light Generators and Detectors

Light generator and detector technologies are important areas of research. Research is continuing to improve the lifetimes of these devices, their spectral stability, the narrowness of spectral emissions, switching speeds, current threshholds, and receiver sensitivity. The most recent advance is the cleave-coupled cavity laser—a device notable for its wavelength stability and capability of changing frequencies rapidly, making it attractive as a multisource generator. It has been demonstrated at 274 Mbs over 100 km of fiber optic cable by Bell Laboratories, and at 1.6 Gbs (billion bits per second) over 40 km by the Japanese.

Coherent detection, a technique to improve receiver sensitivity and to increase the information carrying capacity of fibers, is being pursued in many research laboratories, and may become important if a variety of obstacles can be overcome.

Research is also continuing into ways of integrating light sources, detectors, and the associated circuitry into single chips thus lowering cost and increasing reliability.

Optical Multiplexer

Optical multiplexer (and demultiplexers) are devices that combine (or separate) different signals so they can be sent through the same optical fiber. Wavelength multiplexing is already being used in AT&T's east coast and west coast systems, and a few experimental systems in Japan, Canada, and Europe. Research in wavelength multiplexing techniques should lead to important cost savings for wideband systems.

Connectors and Splicing Techniques

Research is continuing to simplify techniques for splicing together separate segments of optical fiber and to achieve lower losses due to the splice.

Bell Labs recently announced the development of an ultraviolet splicing system that contributes only 0.03 decibels to signal loss, using an optical test signal to assist in the alignment of fibers.

Switches

Switching permits a signal to be routed through specific paths to subscribers. Switching is a bottleneck in optical communications systems because the conversion of signals from light to electronic (current switch technology is electrical) causes delays in moving the signals through the system. Improvements in switching capabilities hold promise for reducing the number of conversions required from optical to electronic and vice versa. Current switches are expensive, limited in applications, and of unpredicted reliability.

Optical switching research is being conducted at Bell Laboratories and the University of Arizona, where experimental, room temperature switching rates, for "turn on, " of 50 picosecond (50 trillionths of a second) have been measured.

Storage

Research is continuing into methods for storing optical signals on fixed and volatile memory devices. Improvements in storage will make possible store and forward and electronic mail features for optical networks.

Amplifiers and Repeaters

Regenerative repeaters detect a signal, then amplify, reshape, and retime it into a replica of the original signal. The regenerated signal then modulates a laser or light emitting diode for transmission along the next span of optical fiber. Decreasing the number of repeaters required along a line depends in part on amplification capabilities.

Integrated Circuits

Research is continuing to improve capabilities for putting optical and optoelectronic light generators and detectors onto single integrated circuits, and to increase the operational bit rates. Recent breakthroughs hold promise for reducing the number of discrete components required in fiber optic systems and expanding bit rate capabilities.

Research in Japan

While research in fiber optics is being actively pursued in the United Kingdom, France, and West Germany, the most advanced foreign research has been undertaken in Japan.

Japan's research is being conducted, at least in part, to support the development of a new nationwide broadband telecommunications network that will make extensive use of fiber optic technology. Research is also being supported for future commercialization and international markets. Fiber optic research supporting this network is performed mainly in private companies, and is supported by the Ministry of International Trade and Industry (MITI) through the Optical Measurement and Control System (OE) project. OE plans to develop optoelectronic integrated circuits, transistors, and GaAs/GaAIAs lasers and detectors. The budget is approximately $100 million for the 1979-87 time period. Half of this amount is to be devoted to a coordinated research facility, information exchange among researchers, and the remainder on projects in six or seven companies, including Hitachi, Nippon Electric Corp., Toshiba, Mitsubishi, Fujitsu, and Matsushita. Japanese investment in infrared laser research is estimated at between $3 million and $4 million.

Case Study 3: Software Engineering

Findings

- Software is an important factor in information technology, exceeding four times the cost of hardware in large systems. The relative decline in hardware costs is shifting the focus of R&D to software. The complexity of new applications and information systems is also forcing the focus of information technology R&D toward software issues.
- R&D in software engineering has produced prototypes of software design tools and programming environments (integrated sets of tools) that promise significant productivity increases. But the cost of retooling, including retraining software manpower, and the uncertainty associated with innovation in software development are retarding adoption of innovative tools and techniques.
- In order to speed the acceptance of software engineering innovations, an applied research base needs to be established to scientifically test and validate new software development techniques, and to disseminate information on their performance in specific applications environments. Such an applied research base would link basic research in universities to applied research and development efforts in industry and government.
- Foreign efforts, particularly those in Japan but also national targeted efforts in Europe, show signs of movement toward such an applied software engineering research base.
- The Federal Government, through the Department of Defense, is making some efforts to create a software engineering applied research base for national defense purposes, but the applicability of this research base to the general problem of software productivity in the American economy is uncertain.
- It is difficult to differentiate software *production* activities from R&D, especially development. Much software production is a creative design effort. There are some aspects and types of programming that are clearly not R&D, but the dividing line is difficult to define.

Introduction

The term software refers both to the instructions that direct the operation of computer systems, and the information content, or data, that computer systems manipulate. Software is thus a logical rather than a physical product. An adequate organizing formalism or calculus for software creation has not yet been discovered.[50] Therefore, the development of large, complex software systems depends heavily on the insight and creativity of sys-

[50] "We are in a business that is 35 years old . . . and I invite you to think where civil engineering was when it was 35 years old . . . they had not discovered the right angle yet." Harlan Mills at the OTA workshop, Nov. 17, 1983.

terns designers and programmers. The methods presently employed to develop and test software are ad hoc, without a strong scientific basis. Thus, software systems are expensive to build and maintain, and can be unreliable in operation. The problems associated with inadequate software methods are growing as information technology uses spread and are relied on for larger, more complex, and more critical applications.

Research and development efforts have produced new methods that show promise for improvement in the productivity and reliability of software creation, testing and maintenance. These include *Advanced program editors* and *debuggers*, new *languages, design methodologies* and *integrated programming environments*. However, the introduction of innovative techniques into software production is proving difficult. The adoption of new methods often requires the conversion of large existing program inventories. This is expensive and risky because evidence that one methodology is better then another is not systematically collected. As well, most software development is oriented toward the single project at hand and often relies on antiquated programming habits and attitudes. The high job mobility of programmers and software systems designers perpetuates individualism and fragmentation in software methodology.

The establishment of software engineering then, involves not only the development of superior methodologies, but a transformation of the programming process—from an art to a *science* and from a *labor intensive* to a *capital intensive* effort. The pressures on software production for larger, more complex, more cost effective, and more reliable systems, make the present situation untenable." A concerted effort among researchers, educators, data processing managers, systems designers, and programmers, and support from corporate and Government management is required for this transformation to occur.

Software R&D Environments

Varieties of Software and Characteristics of Software Development

Software is classified as being of two general types: *applications software* that is designed to apply computer power to a specific task or tasks, such as computer-aided design of automobiles, or payroll or inventory management in a department store; and *systems software* that is used to manage the components of an information system itself, such as computer operating systems that control input and output operations.

In general, systems software is an integral component of the hardware because its job is to control the hardware, including the peripheral equipment–e.g., disk, printer, and memory usage, and to schedule and accommodate the creation and execution of applications programs. A recent trend, encouraged by the spread of personal computers, has been toward the use of standard systems software so that a large number of applications programs can be made comparable with hardware from different suppliers. The manufacturers of hardware or specialized software vendors write most systems programs, and more of the systems programs are being embedded in hardware-in programmable integrated circuit memory called ROM (Read Only Memory). Thus end-users now generally do not create or alter systems programs.

Much applications programming is done by users. For personal computers in homes and for supercomputers at the National Weather Service, programs must be written to tell the machines how to solve problems and organize

'l" According to a projection made at an early 1981 data processing managers conference, Department of Defense software costs would increase nearly three times as fast as the department's budget, and nearly 20 percent faster than expenditures for computers during the 1980s, " B. M. Elson, "Software Update Aids Defense Program, " *Aviation Week and Space Technology*, Mar. 14, 1983, p. 209. Just as the explosion in numbers of telephone operators in the 1930s and of bank clerks for check processing in the 1960s forced a move to automated systems, the sheer demand for manpower in computer programming seems to be increasing the pressure for the introduction of less labor intensive software development techniques; see T. C. Jones, "Demographic and Technical Trends in the Computing Industry, " *DSSD User's Conference.*

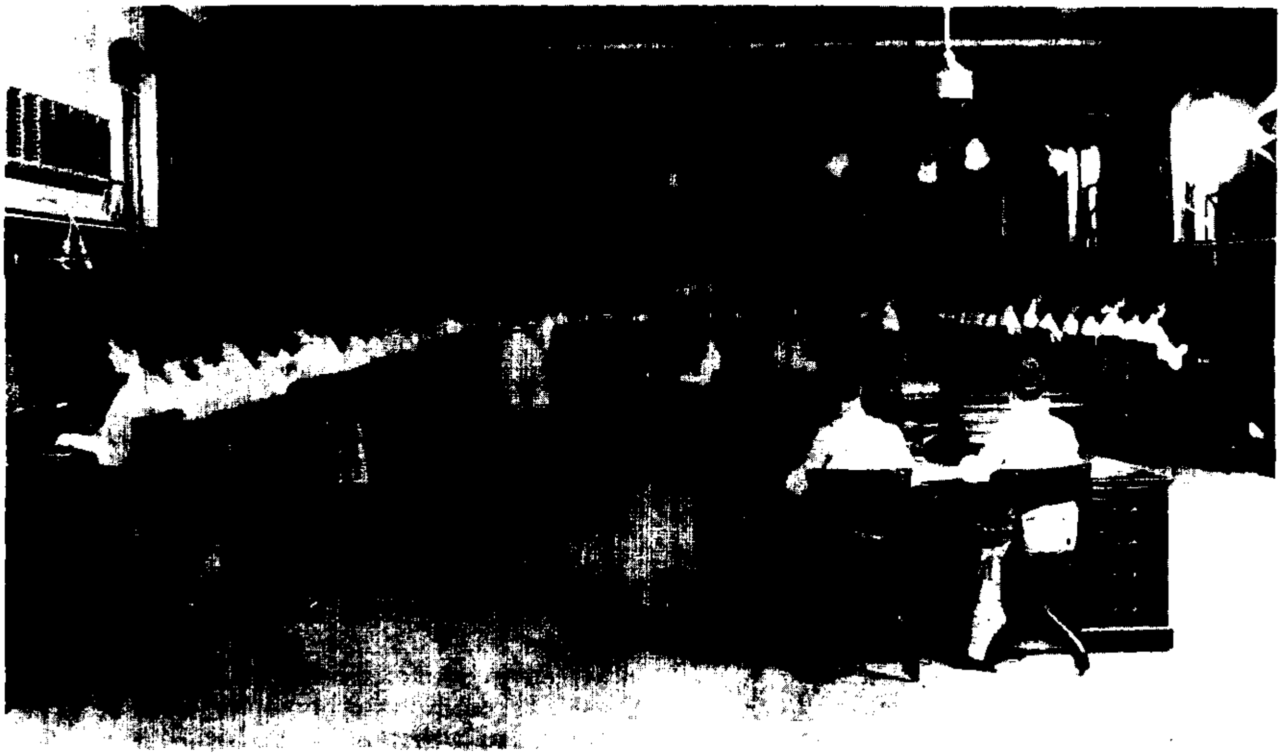


Photo credit: Smithsonian Institution

Like software development is today, telephone switching was once a labor-intensive effort

information. Much of this work is done by highly trained professionals knowledgeable about computer systems and the problems to be solved.

Not all applications programming is software R&D. For example, an economist writing a short program to calculate a unique set of statistics that may never be used again or by others is not engaged in software R&D. Conversely, a physicist developing a program for a supercomputer to calculate formulas used in nuclear reactor design certainly may be involved in R&D. There is a large area in between these examples that is ambiguous, and no hard data is available concerning the time spent creating different categories of programs. The Internal Revenue Service has faced this difficulty in defining the types of software work that are eligible for the R&D tax credit. Their proposed solution has been to consider the costs of developing computer software as not eligible for tax credits, unless the software is "new or significantly improved," or "if the programming itself involves a significant risk that it cannot be written."[52]

Software R&D in Industry

The information processing industry is devoting a large and growing amount of resources to software development. Purchases alone, currently some 12 percent of total spending for software, are expected to exceed $10 billion in 1984. Approximately 10,000 companies of various sizes, from one man operations to divisions of major corporations, are developing software for sale. Estimates of the total number of programmers in industry range from 500,000 to nearly a million,[53] and

[52] "Credit for increasing Research Activity," *Federal Register*, Jan. 21, 1983, pp. 2799-2800. There has been considerable controversy over this issue. See W. Schatz "A Taxing Issue," *Datamation*, June 1983, pp. 58-60.

[53] ADAPSO estimates, and T. C. Jones, op. cit. p. 83.

the distinction between hardware and software developers is becoming increasingly blurred.[54]

Every major computer and telecommunications equipment manufacturer and service provider develops programs and software tools to make their products suited to the needs of users. Some firms such as Hewlett-Packard, reportedly spend nearly tw0thirds of their R&D budgets on software.[66] More than 40 percent of the technical people at Bell Labs are involved with software development.[56]

Interest in the improvement of software productivity has led some large corporations to establish formal programs for the development and evaluation of software practices—to introduce scientific and engineering techniques into the evaluation of software development and use. AT&T currently employs approximately 300 Ph.D.s in software research, while IBM has approximately 150 and ITT has about 20 of these researchers developing and using formal experimental methods of evaluating software engineering techniques. Other large companies, including Control Data, Xerox, and Honeywell, are also beginning to use experimental R&D studies as the basis for improvements in software production.[57]

The Federal Role

The National Science Foundation funds university and some corporate research in software engineering. The NSF Software Engineering Program within the Division of Computer Research awarded $2.2 million in grants in fiscal year 1983.[56] Additional research related to software engineering is funded by other programs in the Computer Research Division (e.g., the Software Systems Science, Computer Systems Design, Theoretical Computer Science and Special Projects Programs) bringing the total funding to $5 million to $10 million per year.[59]

The Federal Government is the world's largest user of data processing resources. A recent GAO study found that 95 to 98 percent of the Government's applications software is custom developed.[60] The cost of software for the Department of Defense is estimated to be $4 billion to $8 billion per year.[61] DOD operates a patchwork of incomparable systems and computer languages.[62] The incomparability of software contributes to increased software development costs, through schedule slippages, lengthy testing programs, and problems in contracting for hardware and software services.

DOD has moved toward the development and use of a single standard computer language. The rationale for this is the potential for saving several hundred million dollars a year through lower personnel training costs, increased programmer productivity, and substantial reuse of standard code modules. After competitive development of four separate languages and extensive design evaluation, the Pentagon chose a language developed by the French company Cii Honeywell Bull. The name of the language, *Ada*, is trademarked, and compilers using the Ada name are strictly controlled and validated to assure that the language remains standard. Ada is expected to be the primary DOD computer language by 1987.

There is some resistance to use of Ada. Several years ago, the Air Force developed its own quasi-standard language, Jovial. Comparison tests between Ada and Jovial will continue for

---

[54] See, for example, S. B. Newell, A. J. De Geus, and R. A. Rohrer, "Design for Integrated Circuits," *Science*, Apr. 29, 1983, pp. 465-471. See also "The Changing Face of Engineering," *Electronics*, May 31, 1983, pp. 125-148.

[55] W. P. Patterson, "Software Sparks a Gold Rush," *Industry Week*, Oct. 17, 1983, pp. 67, 69-71.

[56] *AT&T: 1982 Annual Report*, p. 19.

[57] OTA workshop on Software Engineering, Nov. 17, 1983.

[58] *Summary of Awards, Fiscal Year 1983*, National Science Foundation, Division of Computer Research.

[59] Estimates by OTA.

[60] "Federal Agencies Could Save Time and Money With Better Computer Software Alternatives," General Accounting Office, GAO/AFMD-83-29, May 20, 1983, p. 1. Even standard applications such as payroll are largely custom designed. This GAO report found that there are at least 78 different Federal civilian payroll systems.

[61] Elson, op. cit., p. 209.

[62] I. Peterson, "Superweapon Software Woes," *Science News*, May 14, 1983, pp. 312-313. Testing of software alone is estimated to account for nearly half of this cost.

the rest of the decade.[63] The Army is more enthusiastic about Ada. The Navy, which has more software already written than the Air Force and Army combined, is interested, but the task of switching the Navy to Ada will be enormous. The Navy has over 450 different systems and subsystems with embedded computers, and the number of Navy computers has been doubling every 2 years.[64]

A limitation of Ada is that a new programming language only addresses about 20 percent of the total software problem. As detailed below, coding of computer programs is 20 percent or less of the effort of developing and maintaining software. Government computer systems in particular require an enormous amount of documentation, and Ada has no facilities for automated documentation.[65]

To deal with these problems of software design, production and maintenance, DOD has proposed a new initiative called Advanced Software Technology.[66] This is envisioned as a 10 year program costing $250 million.[67] The administration requested a funding level of $19.3 million in fiscal year 1984, but the Senate Armed Services Committee cut the program to $10.5 million because, "The Committee is not convinced that the necessary planning has been done to justify a budget of nearly $20 million in the first year."[68] Included in the plan, as it has thus far been developed, is a provision for a Software Engineering Institute to help formulate and standardize software engineering techniques and practices.[69]

[63] J. Fawcette, "Ada Tackles Software Bottleneck," *High Technology,* February 1983, pp. 49-54.

[64] Peterson, op. cit., p. 313. It has been estimated that there are some 50 million unique lines of Navy software code in a variety of languages currently in use, It would cost, it has been reported, some $85 billion and take several years to rewrite this mass of code.

[65] The OTA workshop on Software Engineering, Nov. 17, 1983.

[66] This initiative is known within DOD as STARS, Software Technology for Adaptability, Reliability and Serviceability.

[67] Peterson, Op. cit.

[68] *Omnibus Defense Authorization, 1984, Report to Accompany S.675,* Committee on Armed Services, U.S. Senate, p. 131.

[69] OTA workshop on Software Engineering, Nov. 17, 1983.

Content and Conduct of Software Engineering R&D

Software engineering ideally is a set of concepts and tools for transforming descriptions of tasks to be performed by computer systems into digital code that machines can understand. Software engineering research involves the study of methods to understand, improve, implement, and evaluate these concepts and tools, and to embody them in *software development systems* or *programming environments* to facilitate the entire software *lifecycle.*

The Software Lifecycle

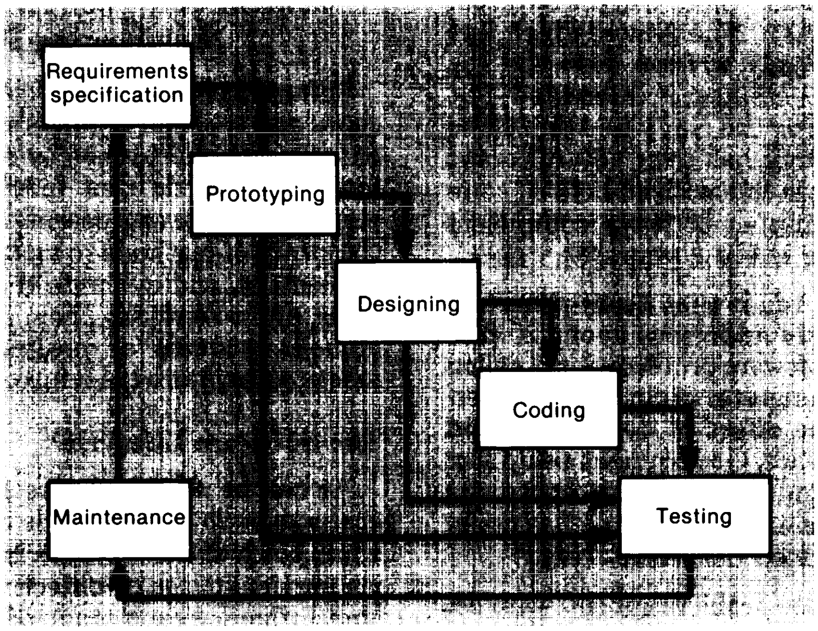
As can be seen in figure 10, there are several stages in the life-cycle of a piece of software. Each of these stages is characterized by its particular set of objectives and methods that influence each stage of research and expected improvements. Testing occurs continuously throughout the life of a software and is an integral part of software production and maintenance. Documentation, an activity not depicted in figure 9, is of preeminent importance in every phase of the software lifecycle. Comprehensive records of every activity and software characteristic must be created to aid designers, programmers, maintainers, and users in understanding the structure and operation of the software system.

Requirements Specification

This initial phase of software development involves the description of the system objectives and the tasks that the end users want performed. This description requires a thorough knowledge of the application by the software design project team. It is accomplished by rigorous and continuing communication between the project team and the end users.

OTA advisors and published sources emphasize that requirements specification is the most critical phase of project development, because all of the later stages must build on the foundation laid in this activity. At the same time, it is the most difficult activity in the software lifecycle to develop a rigorous method-

Figure 10.—The Software Life Cycle



SOURCE: Data Communications.

ology for. Requirements specification involves the understanding, explicit organization, and integration of what are often idiosyncratic practices of information usage. Often it is not possible to completely specify requirements at the beginning of a project; also, requirements can be expected to change over the course of a long software development project.

Computerized tools now available or under development may offer assistance in the specification process.[70] They are designed to help users describe and specify system properties, functions and performance requirements. Some of these tools allow the specifications to be checked by computer for consistency and correctness; some of them can generate simulations to help the user analyze the operation of a specified system. Thus far these tools have worked well only on a limited range of applications.

A difficult problem is specifying how the knowledge, experience and habits of workers in particular environments can be described and how these abstract specifications can be converted into machine code. Research is proceeding in *knowledge representation* and *knowledge engineering* (*see* Artificial Intelligence Case Study). Near term prospects are uncertain for identifying core concepts for organizing the different kinds of knowledge found in the variety of environments in which computers are applied, or designing broadly applicable *very-high-level languages* which can automatically render abstract specifications in machine-executable procedures.[71]

The experience now being gained in research and development of "expert systems," and the more fundamental research efforts in knowledge representation, machine learning and human cognition should eventually contribute to the process of computer systems requirements specification. But for the near term, this phase of software development will remain labor intensive and will require special skills (including human relations) and specific knowledge of applications among its practitioners.

[70] OTA workshop on Software Engineering, Nov. 17, 1983.

[71] R. Yeh, "Software Engineering," *IEEE Spectrum*, Nov. 1983, p. 92.

Prototyping

Once the project team has a working knowledge of the requirements of the software, a preliminary design of the final system is made so that the team can obtain feedback from the users. Thus, fine adjustments can be made at an early stage.

A *software prototype can* be as simple as a paper and pencil sketch that runs through the workings of the system, or as complex as a large and intricate computer simulation. The choice of prototype complexity is based on the nature of the application, the level of training and the sophistication of the end users, the size of the final software system, and the criticality of the application.[72]

Research and development of the prototyping phase involves the design, development and testing of methods and tools to make prototyping more understandable to systems designers and users, and more automated so that it is a logical and accepted part of software development. Research efforts in universities and industry are focusing on languages that make the quick production of prototypes possible, and on the design of man-machine interfaces to enhance and verify the effectiveness of communication between the system and its users.

The major issue in prototyping involves the acceptance of this activity and the awareness among system designers of its importance. Prototyping is commonly ignored in current software development practice because of its expense and difficulty and the lack of automated support. But experts contend that, in the long run, prototyping can save time and effort because changes to requirements specifications become more costly as each phase of software development proceeds.

Design

The design stage of software development involves the reduction of the software specifications to a set of *procedures* that can be programmed for the computer. The software design team analyzes the application and segments the design problem into subproblems or *modules* that can be programmed by individuals or small groups.

The segmentation or modularization of the software design breaks the problem into manageable pieces to maximize the ease of programming and testing the segments, and to facilitate the interchange or replacement of the modules to simplify maintenance. Proper segmentation is of particular importance because it has been demonstrated that as the size of the programming project team grows, more and more time is spent communicating among team members and coordinating communication among program segments and less in actually writing code. Thus segmentation has a crucial impact on the overall productivity of the software development team.

The design phase of software production relys heavily on the experience and intuition of the design team and is considered to be best learned by apprenticeship. Understanding of fundamental mathematical principles is also important.[73]

Research on the design phase of software engineering is embedded within the larger framework of R&Don "programming environmerits. " The objective of software engineering is to provide as much support as possible for the creative designer. Facilities should be made available to ease the rendering of abstract principles and problems into workable, testable, reliable, and cost-effective solutions. Software design can be made more productive in much the same way that design of complex integrated circuits has improved: by the application of computer-aided-design (CAD) tools based on sophisticated graphics workstations. A number of companies are making significant productivity gains using CAD tools for software design.

[72]For example, air traffic control software requires extensive simulation, whereas an inventory system might need only a simple sketch and some text to check against user needs.

[73]*Data Communications,* op. cit. p. 81, and *Mosaic,* op. cit. pp. 4-5.

In the longer term, software design productivity could improve significantly through the expanded development and use of *software Libraries*, or systems to store and access standard, reusable software modules that perform functions required in many programs. As now practiced, software design generally segments each program problem in a unique fashion, both to accommodate the unique features of new applications and to make the most efficient use of hardware. Program modules are usually not designed for reuse, and there is no strong incentive to do so because facilities are not provided for storing separate modules or for classifying and finding appropriate segments for new programs, and because the project development team is generally unable to spread the cost of creating reusable modules across a large organization. The result is that designers and programmers are constantly "reinventing the wheel" in software development.

Researchers in computer and information science have identified basic techniques and have designed tools to provide generic programming modules, data structures and algorithms, and to classify and organize them for storage and access. But the software community has been slow in adopting the facilities necessary to make software libraries work because of high initial cost, poor management understanding of such systems, and reluctance among programmers to modify long-established design practices.

coding

Coding is the stage of software production that makes a procedural description of the system executable by computer. More emphasis has traditionally been placed on this activity than on any other, even though it is estimated to comprise only 15 to 20 percent of the total software development effort.[74] The reason for the traditional emphasis on coding is that it represents the basic mechanism by which programmers control the operation of computer systems.

Many tools are already available to programmers to help produce efficient and reliable code. The most familiar of these are the high-level languages such as FORTRAN, COBOL, and Basic that were designed to aid programmers in coding certain broad types of applications. *Compilers* translate high-level language into machine executable binary digital code. *Editors* and *debuggers* facilitate the entry, testing, and correction of code lines.

Research and development in the improvement of computer code writing continues. Innovations over the last decade include *structured programming* which greatly simplifies the tasks of testing and deciphering code during maintenance, and *new high-level languages* designed to be easy to learn and use or technically sophisticated (no languages yet introduced appear to be both simple and highly flexible).

A recent trend, that has proceeded hand in hand with the introduction of personal computers, has been the development of coding systems or languages that can give end users of computer systems more control—and remove the need for the help of a professional programmer for each application. An example is Visicalc.

Large companies, including manufacturers of mainframe computers, are interested in decentralizing the control of computer programming and use by providing software systems that are responsive to the needs of noncomputer professionals. A proposed solution to the software bottleneck dilemma is to farm out some of the work now done by professional programmers to end users. IBM, for example, asserts that up to one-half of computer applications development can and should be handled by the end-users with personal computers.[75]

[74] *Data Communications*, op. cit. p. 58, and E. B. Altman, "Software Engineering," *Mini-Micro Systems*, December 1982, p.184.

[75] *Application Development in Practice*, Tech Tran User Survey, Xephon Technology Transfer, Ltd., 1983, pp. 53-54. Some experts consider this solution to be dubious, and believe that it may in fact merely spread software problems.

The proliferation of personal computers in the business world has some other possible implications, both for the process of software development and for employment patterns of professional software development personnel. End-users programming on personal computers could allow central data processing departments to concentrate their efforts on the larger, more complex, and more critical applications. A reduction in the backlog of small projects could also encourage the development of new large-scale applications that heretofor languished because of long waiting times. It is conceivable that the demand for professional programmers will slacken in such industries as financial services which have traditionally employed a large part of the data processing workforce in coding and maintaining applications programs. This could result from trends toward: 1) small-scale applications development by end-users with personal computers using commercially available software packages, and 2) increased use in data processing shops of standardized and automated software development tools, also commercially obtained, for large-scale projects. But increasing demand for innovative, off-the-shelf applications programs and software development tools suggests that demand for the *most talented* software designers will *increase* as competition intensifies in these expanding markets.

There is a secondary impact of the introduction of personal computers on software development. Upper management is becoming familiar with the problems and possibilities of software and computer use through experience with personal computers. Management aware ness of and commitment to change in the software development process may be a key factor in introducing and accepting software engineering concepts and innovations.

Large and complex applications will continue to depend on the efforts of professional software designers and programmers. Thus, the introduction of automated tools and techniques such as CAD are imperative. Pressure will continue for more productive and reliable software engineering as the complexity of computer-based systems increases and as more of society's functions are entrusted to computerized systems.

Testing and Validation

An important activity in software production and maintenance is the testing of the products of the various phases of the software lifecycle. Until recently software development budgets included sufficient funds for adequate testing in only a few highly critical projects.

It is impossible, no matter how extensive the testing activities, to *guarantee large* software systems to be error free. A risk assessment based on the cost associated with software defects or breakdown in a given application must be made to determine the level of effort in testing that is justified. A rigorous testing regime can double the cost of software development.

There are many concepts, tools and techniques available and undergoing research or development for testing and validating software. They range from traditional methods such as manual "desk checking" or "walk throughs, " to statistical methods to determine the probability of errors in a set of code, to some as yet highly experimental automated *program provers* that verify that properly structured programs perform as specified. The broad applicability of program provers will de penal on some fundamental breakthroughs in program theory.[76] Until the requirements specifications and design stages of software development are better understood and supported, manual techniques will be the key testing methods available. Although some automated tools are in use, few exist as integrated packages. They must generally be pieced together by individual projects; and

[76] Some experts contend that the pursuit of formal verification methods such as program provers may prove fruitless because of the increasing complexity of software and the rapid changes in software development practices. See R. A. DeMillo, R. J. Lipton and A. J. Perlis, "Social Processes and Proofs of Theorems and Programs, " *Communications of the ACM*, May 1979, pp. 271-280.

they themselves can become sources of high cost and errors.[77]

Maintenance

This is the phase of the software engineering lifecycle that is concerned with the revision of software that is in use. Software maintenance consists of two kinds of activities: correction of errors that went undetected in the course of software development and testing but that crop up in the use of the software; and changes in programs resulting from altered or additional requirements specifications. It has been estimated that 50 to 90 percent of current software costs involve maintenance. Some software systems may cost 25 times as much to maintain as to develop.[78] Because of poorly designed, badly written, and inadequately documented code, much of the maintenance programmers' time is spent trying to understand programs rather than changing them.

A solution to the maintenance problem adopted in many cases is to discard old code. This is because experience indicates that the "patches" made in programs when they are fixed or revised often increase the complexity and difficulty of maintaining code and make it inefficient to run. Therefore an average of 10 months worth of programming development effort is thrown away each year by mainframe based data processing organizations. Unfortunately, this old code is often replaced by poorly designed new code which again is difficult and expensive to maintain.

Poor requirements specification and design practices, idiosyncratic coding styles, and poorly organized documentation increase maintenance costs. Because requirements were vaguely specified or have changed, the design proceeded on assumptions that were false or later rendered invalid. Bad design assumptions lead to poor segmentation and structuring of programs which exacerbate the problems of making changes and testing. A change in one part of a poorly designed program will require changes in other parts of the program, multiplying the cost and time for maintenance. Novel coding of standard functions increases the difficulty of predicting the effects of changes. Large and cumbersome, poorly organized and incomplete documentation make it difficult for maintenance programmers to know what the system is supposed to do.[79]

Junior programmers with little experience are generally assigned to maintenance tasks, thus increasing the delays in bringing software back into use, and also perpetuating the knowledge and acceptance of poor programming styles and habits.

The Scientific Basis for Software Engineering

There are some well understood scientific methods that can contribute significantly to the improvement of software engineering tools and practices. The application of scientifically designed, experimentally based statistical methods for the evaluation of the effectiveness of software tools may produce the concepts and methods for at least an order of magnitude increase in software productivity.

In particular, in the critical area of human interaction with computers, a body of research methods have been developed in universities that, if widely and systematically applied, could identify, quantitatively assess, and produce improvements in programs for making computers more responsive and easier for people to program and use. Research in behavioral psychology and psychophysics is being applied to the design of computer systems and software engineering tools on a limited basis in a number of university programs.

The limited scale of commitment to *applied research in* software engineering maybe a major reason for lack of dissemination and use of these well understood techniques. Acad-

[77] W.R. Adrion, M. A. Branstad, and J. C. Cherniavsky, "Validation, Verification, and Testing for Computer Software," *ACM Computing Survey,* June 1982, p. 183.
[78] Olsen, op. cit. p. 58.

[79] OTA advisors related that program documentation captures, on average, only 60 percent of the information needed to fully understand a program.

emics are generally involved in small-scale, *fundamental research* oriented projects that cannot test large size programs or a large number of possible software development aids. Some researchers suggest that applied research on a massive scale following the model of medical clinical trials is necessary to disseminate and make use of the knowledge embodied in available techniques.

Industry participation is an important ingredient in increasing the scope of experimentation because of its large and diverse software efforts. As a user of innovative tools and systems, industry must also be involved in the research so that essential feedback is provided to tool designers and evaluators. Inadequate appreciation of the gains that can be achieved in software productivity and a concentration on short-term results have made industry management reluctant, until recently, to make a commitment to large scale or long-term experimental software development studies.

As mentioned earlier, some large companies are becoming interested in applied software research and are hiring computer and human factors scientists to develop research programs. However, competitive and proprietary considerations and the uncertainty of software protection by copyright or patent appear to have inhibited an integrated applied research effort *among* companies. Antitrust laws may also have a chilling effect on industry based joint applied research.[80]

An alternative approach is for the Federal Government to use its software development efforts as a test-bed for software engineering research. Some moves in this direction are being discussed within the defense community, and the National Bureau of Standards has a research effort in software engineering in its Institute for Computer Science and Technology.

[80] Eleven major defense contractor companies recently announced plans to form a consortium for the pooling of efforts in software engineering R&D. The Justice Department has reportedly given the go-ahead for the development of a formal business plan patterned after the Microelectronics and Computer Technologies Corp. (MCC). M. Schrage, "Software Research Group Set," Washington *Post*, Oct. 10, 1984, pp. C1-C2.

International Efforts in Software Engineering R&D

Japan

The Japanese have had a software engineering effort since 1970.[81] The Information Technology Promotion Association (IPA), a consortium of private companies and the government, together with long-term credit banks, provided $25 million in 1983 for software R&D and technology transfer. Current projects include: purchase or development of software packages for rent to users; a Software Maintenance Technology Project, to be completed in 1985, which is developing work stations for program analysis, documentation, testing, and production management; and a cooperative effort among users, suppliers, and researchers to develop computer-aided design, computer-aided instruction, and software engineering innovations.

A noteworthy aspect of Japanese software engineering efforts is the establishment of software *factories*. These are consortia, staffed by people from participating companies, which are creating integrated environments for software production, testing, and maintenance. Impressive programmer productivity gains are reported to have resulted from the average reuse of 30 percent of computer code in new applications.[82] Japanese software factories appear to owe their success in raising productivity levels, in part, to the fact that only restricted kinds of applications with fixed and well understood requirements are attempted. It is uncertain how effective these software factories would be on more difficult applications. Nevertheless, these concerted, cooperative efforts provide a test-bed for new software engineering concepts and tools in a well capitalized development and production environment. As well, they train people in the use of advanced software development techniques, serve as a dissemination mechanism

[81] *Summary of Major Projects in Japan for R&D of Information Processing Technology*, prepared by Arthur D. Little (Japan), Inc., under contract to OTA, July 1983.

[82] Raymond Yeh at the OTA workshop on Software Engineering, Nov. 17, 1983.

when those people return to their companies, and raise management awareness of the importance and efficacy of software engineering. Thus software factories have potentially significant implications for future Japanese software development efforts on a broad scale, regardless of their impact on individual software systems.

Britain

As a part of the Alvey Programme for R&D in information technologies (see ch. 7), Britain plans to spend approximately $100 million over 5 years (1984-89) on software engineering R&D, which is approximately 20 percent of the total Alvey Programme effort.[83] Britain's traditional strength in software will be built upon, expanded and modernized to minimize dependence on software imports.[84] An initial focus of the Alvey effort will be the measurement of the cost effectiveness of software engineering through analysis of software imports and exports. Also, the program will track the capitalization of software development and analyze the relationship of capital intensity to programmer productivity. Efforts will be made to establish formal output measures for software cost and quality to help evaluate new tools and methods and to establish a system for software warranties.[85]

The software engineering effort will consist of three main thrusts:[86]

1. *Exploitation* of existing tools and methods, the transfer of technology from universities to industry, and increased investment in software production capital and in training.
2. *Integration* of tools and methods for the improvement of hardware and software development which will be focused in an Information Systems Factory.
3. *Innovation* software engineering through increased levels of R&D aimed at developing new tools and evaluating their effectiveness.

The Information Systems Factory, to be established by 1989, is based on the premise that information systems' functional requirements may be written independently of whether a given function is to be implemented in hardware or software. Therefore, it is reasoned, the trend should be toward the design of modules whose uses are known in relation to other modules, such that the decision of whether to implement a function in hardware or software may be made on the basis of economic, timescale and other cost-benefit criteria. The Information Systems Factory concept will evolve into the 1990s, by which time advances in fundamental research are expected to make possible a truly intelligent software engineering production environment with fully automated and integrated tools and a large library of standard function modules.[87]

The British consider the understanding of software engineering concepts to be too primitive at this point to make a final determination of a single direction to pursue to produce a fully integrated software engineering environment. Like those in the United States, current British research projects are too small in scale to adequately test software in diverse applications to determine the efficacy of new tools and methods. The Alvey Programme will fund several approaches for full-scale development, both to select useful techniques and to build technology transfer bridges between research, development, and production. In this context, the software engineering effort is expected to contribute to, and to benefit from, the parallel Alvey Programme efforts in computer-aided design for VLSI circuits and knowledge-based systems.[88]

[83] *A Programme for Advanced Information Technology: The Report of The Alvey Committee*, Department of Industry, Her Majesty's Stationery Office, p. 47.

[84] *Alvey Software Engineering—A Strategy Overview*, prepared by Alvey Software Engineering, Director D. E. Talbot, Department of Trade and Industry, p. 1.

[85] *Software Reliability and Metrics Programme: Overview*, prepared for the Software Engineering Directorate by the Center for Software Reliability, Alvey Directorate, Department of Trade and Industry, April 1984, p. 4.

[86] *Alvey Software Engineering—A Strategy Overview*, op. cit., p. 3.

[87] Ibid., pp. 5-7.

[88] Ibid., pp. 7-8.

France

Like the British, the French have been noted in the past for the quality of their software.[89] The major French R&D centers for software engineering are run by L'Institute National de Recherche en Informatique et en Automatique (INRIA) and the Centre National d'Etudes Telecommunications (CNET). Together, these laboratories employ approximately 100 researchers in software engineering and related studies.[90] In addition, 11 other French Government and industry labs are pursuing R&D in software engineering.[91]

[89] Three significant computer languages, Algol, Prolog, and Ada, were originated in France.

[90] *Research and Development in Electronics USA-France 1982/1983*, French Telecommunications and Electronics Council, March 1984.

[91] *IEEE* Spectrum, November 1983, pp. 64-65.

Case Study 4: Artificial Intelligence

Findings

- Artificial intelligence (AI) research seeks to make computers perform in ways that demonstrate human-like cognitive abilities: perception and action in complex situations and environments; interaction with humans in natural language; and common sense and the ability to learn from experience.
- The capabilities of artificial intelligence are often subject to exaggeration. AI has gone through several waves of optimism and disappointment as new fundamental concepts have been discovered by research, as new computational techniques and equipment have allowed these concepts to be tested, and as the limits of these concepts have been realized when applied to real-world or large-scale problems.
- In the last several years, the first real commercial products of about 25 years of AI research have become available. These systems are of three types:
  - *expert systems* that aid human experts in analyzing complex situations and in making decisions,
  - *natural language processing programs* that serve to make the interaction of humans and computers more natural, and
  - *image or vision processing systems* that can make robots and other automated processes more flexible in operation.
- Presently available AI technology is of value in only a limited range of applications. But the promise and potential have led to several national efforts to push AI technology forward.
- In the United States, AI research has been supported principally by the Department of Defense. Plans for putting artificial intelligence concepts to work in defense applications have recently been announced. These applications are far beyond the capabilities of any present systems.
- Basic AI research is conducted at a relatively small number of the nation's top universities by a small number of researchers. These researchers are under conflicting pressures from companies wishing to capitalize on the production of highly valuable systems, from students demanding training in AI, and from personal desire to pursue their own research interests.
- Only a few universities have the facilities and equipment to compete with industrial AI labs. Therefore some highly motivated researchers are drawn out of academia where they would be available to perform needed fundamental research and to train future generations of researchers.

Introduction

The goal of artificial intelligence (AI) research is to create systems that demonstrate some of the following characteristics: the ability to assimilate unstructured information and to act independently in complex situations; the capability of natural language interaction (e.g., in English) with humans; common sense, and the ability to learn from experience.[92] The possibility that machines may be made capable of such activities has burst into public consciousness as a result of the increasing power of microelectronics-based computers and their application in many new environments, the establishment of national efforts with the aim of creating intelligent computer systems, and the sudden emergence of some commercial products that embody characteristics of artificial intelligence.

Some experts contend that the ultimate aim of computer science is to produce intelligent machines. Just as the industrial revolution was driven by the urge to amplify human and animal muscle power with more efficient mechanical devices, the computer revolution is driven by the desire to amplify human intellectual power with electronic information machines.[93]

An expanding user population and the increasing diversity and complexity of the applications of current and planned information systems, are compelling designers to find ways to build intelligence into computers. Computers are increasingly being required to exchange information among people with diverse interests, backgrounds, and levels of computer sophistication, so they are being designed with software that permits people to interact with them in forms more closely resembling natural language. New varieties of sensors and increasingly diverse sources of information are providing input, thus computer systems must be more flexible to make use of information with varying levels of importance and confidence attached to it. Processing and output facilities are needed that present information in forms that are tailored to the unique needs of individuals, or that can control operations in complex and perhaps unpredictable circumstances.

The complexity of the environment in which the system is to operate is perhaps the most important dimension to be considered in the design of AI computer systems. As the number of environmental variables increases, and the number of possible responses multiplies, automated systems must exhibit increasing degrees of intelligence. Currently, computer systems are being designed for environments in which they will be required to resist confusion and error from ambiguous or contradictory input signals, automatically coordinate diverse activities under changing conditions, and provide logical and understandable explanations of their actions to operators.

The highest degree of machine intelligence might be the ability to automatically adapt to conditions that were not specifically anticipated by the designers. Such ability would require a machine to have "common sense, " broad world knowledge from which to infer reasonable courses of action, and also the ability to assimilate new knowledge by learning, through self-organization of experience. Machines with this degree of intelligence are speculative, and attempts to push present AI concepts toward such capabilities illustrate both the difficulty of the problems, and the inadequacy of present concepts as a foundation for machine intelligence on a scale approaching the cognitive abilities of humans.[94]

There have been some notable recent successes in applying AI concepts to real-world problems. The introduction of a number of AI systems into commercial use has elicited a demand in industry for AI software that can make computer systems more responsive and productive. Three types of artificial intelligence products in particular are generating in-

[92] D. Waltz, "Artificial Intelligence: An Assessment of The State-of-the Art and Recommendation for Future Directions, " *AI Magazine*, fall 1983, pp. 55-66.

[93] For example, the Japanese 5th Generation project is motivated by a belief that intelligent computer systems will be one of the cornerstones of a healthy economy and society in the later years of the 20th century and beyond. See T. Moto-oka, "Challenge of Knowledge Information Processing Systems (Preliminary Report on Fifth Generation Computer Systems), " pp. 3-89, and H. Karatsu, "What is Required of The 5th Generation Computer–Social Needs and its Impact," pp. 93-106 in *Fifth Generation Computer Systems*, T. Moto-oka (cd.), JIPDEC-North Holland, 1982, 287 pages.

[94] "Some tasks requiring "intelligence" are already performed better by machines than humans, for example long division; other tasks may never be performed by machines. See Waltz, p. 55. See also M. Michael Waldrop, "The Necessity of Knowledge, " *Science*, Mar. 23, 1984, pp. 1279-1282.

tense interest among industrial companies. These are: *natural language data base interfaces* that allow users to access information stored in a computer data-bank with English language queries; *expert systems* that aid people in complex tasks that require experience and detailed knowledge to perform; and *robot vision systems* that promise increased flexibility in manufacturing and other robot applications. Industry sees these systems as enhancements or potential replacements for rare and/or expensive human skill.

Similarly, the Department of Defense is interested in artificial intelligence to enhance the power, efficiency and reliability of increasingly automated computer-based weapons and command and control systems. DOD's new "Strategic Computing" program seeks to push the frontiers of AI science and technology for use in battlefield management systems, autonomous (unmanned) tanks and submarines, and automated expert assistants for airplane pilots.

There is concern, among both AI researchers and observers of information technology R&D, that enthusiasm about recent successes and fear of foreign competition are encouraging irrational expectations for artificial intelligence. These forces are drawing limited AI R&D resources toward risky, short-term development work, and away from some of the tough questions that AI research should and could answer. As Arno Penzias, Nobel laureate and Vice President for Research at Bell Laboratories has written about artificial intelligence R&D:

> A crash effort in any one area would inevitably pull talented people away from other areas . . . whose payoffs to society might be even greater in the long run. Our challenge is to improve computers and extend their expertise into all the areas where people will need them and want them. We can best do that with a balanced program of research and development. . . using knowledge to help create a society that provides a meaningful life for all the people in it.[95]

The sudden discovery of promise in artificial intelligence by society at large raises several questions. Of immediate concern are the questions of whether AI, as it is now understood, can meet industry and military expectations for performance, or whether there will be expensive failures in applying this technology, and by implication a casting of doubt on the entire AI enterprise, because present concepts are immature and limited; and whether the present educational structure, or some reasonable extension of it, can meet the growing demand for trained AI R&D talent. For the longer term, there is a question whether the progress of AI research can be sustained in the face of feverish commercial development, or whether the limited number of researchers will be drawn away from the study of fundamental and difficult problems and the teaching of new AI people. There are also profound questions concerning how artificial intelligence technology might be used, and how these uses might affect society in general, especially in terms of employment and the potential the technology offers to enhance or compete with human intellect.

Artificial Intelligence R&D Environments

The Roots of AI

The idea of automated intelligence has intrigued mankind for more than a century.[96] In the 1950s, two important figures in the early history of electronic computers, John von Neumann and Alan Turing, both expressed confidence that within a short time computing machines would equal or surpass human intellectual capabilities.[97]

[95] A. A. Penzias, "Let's Not Outsmart Ourselves in Thinking-Computer Rush," *Wall Street Journal*, Sept. 13, 1983.

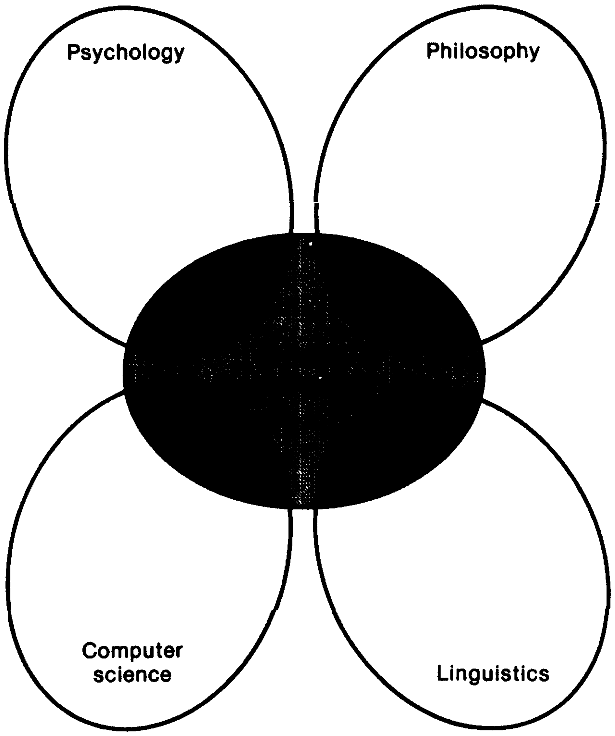
[96] E. Charniak, "Artificial Intelligence-An Introduction," presentation at the 1983 Conference of The American Association for Artificial Intelligence, Washington, DC. One can conceive of artificial intelligence as the culmination of the *mechanistic* paradigm of scientific thought that has been dominant since the time of Newton. See *Science and Change, 1500 to 1700* by Hugh Kearney, McGraw Hill, New York, 1971, for a discussion of the emergence of the mechanistic paradigm. Some historians suggest that Man has always sought to represent and embody life and intellect in his artistic and useful creations. See J. David Boulton, *Turing's Man: Western Culture in the Computer Age*, (Chapel Hill) University of North Carolina Press, 1984,

[97] See J. von Neumann, "The General and Logical Theory of Automata," pp. 99 and 109, and A. M. Turing, "Computing

Improvements in the understanding of computation and knowledge organization coupled with advances in computer speed and memory capacity made possible through improvements in electronics, have encouraged intermittent waves of optimism that super electronic "brains" were just a few years in the future. Time and time again, these waves of optimism have ebbed in the face of a common problem. As the systems that researchers contemplate and build are required to deal with more information, and as the situations in which they are required to operate become more realistic (more perceptually complex and unpredictable), the fundamental computational principles and methods they have to work with become less efficient and impossibly slow.[98]

The term "artificial intelligence" was coined by John McCarthy in a grant application in 1956 to describe the subject of a conference that he was organizing.[99] This meeting, held at Dartmouth College, brought together researchers in different fields whose common concern was the study of human and machine cognition. The conference established AI as a distinct discipline, and also served to define the major AI research goals: 1) to design machines that think, and 2) to understand and model the thought processes of humans. Thus AI research is grounded in computer science and electrical engineering and also has its roots in cognitive psychology, linguistics, and philosophy. There is an appreciation among many researchers that AI is an interdisciplinary study born of the interaction of these two goals, and that the intersection of the separate traditions constitute what has come to be a core of concerns that distinguish artificial intelligence from other fields of science and engineering.[100] (See fig. 11.)

Figure 11.—Artificial Intelligence



SOURCE: Randy Davis and Chuck Rich.

As an esoteric study melding some aspects of computer science, linguistics, psychology, and philosophy, artificial intelligence research has grown up, for the most part, in an exclusive set of universities. Some industry-based research efforts at AT&T Bell Laboratories, Xerox Palo Alto Research Center, Bolt Beranek and Newman, and SRI International span more than a decade, but most of the work has been concentrated at Massachusetts Institute of Technology, Stanford University, and Carnegie-Mellon University, the leading research and academic centers offering a full range of AI subdiscipline. The concentration of effort and expertise in these top tier universities has tended to restrict entry into the field, and to foster a very tightly knit community of researchers.

Some years ago, AI research and training programs began to spread to a wider set of universities, and this expansion of academic programs continues. These schools have programs built around one or two graduates of

---

Machines and Intelligence," p. 245, in *Perspectives on the Computer Revolution,* Z. Pylyshyn, (ed.), Prentice-Hall, Englewood Cliffs, NJ, 1970.

[98] J. T. Schwartz, "Research in Computer Science: Influences, Accomplishments, Goals," Report *of The Information Technology Workshop,* R. Cotellessa, Chairman, National Science Foundation, Oct. 10, 1983, p. 18.

[99] Charniak, op. cit., p. 5.

[100] Nils Nilsson, "Artificial Intelligence Prepares for 2001," Presidential Address given at the Annual Meeting of the American Association for Artificial Intelligence, Aug. 11, 1983.

the established programs, and thus they offer only restricted AI curricula. Also, a number of new industry-based R&D programs are being formed. The newer industrial programs in AI R&D have also tended to center on the ideas of one or two researchers from the top schools. The lifespan of these industrial efforts may be limited if they are unable to draw in new talent from the small available pool and are unable to obtain infusions of new ideas.[101]

The difficulty of establishing and maintaining new AI R&D programs results from the fact that there is only a small base of existing researchers. Second, the nature of AI work is such that an interdisciplinary team is required in most cases to produce programs of useful size and complexity. This "critical mass" of workers knowledgeable in AI concepts exists only in a few universities and companies and requires a number of years to pull together before fruitful work can be expected.[102]

OTA found that a major transformation is occurring in AI, resulting in two major directions for the R&D community. On one hand, fundamental research will advance theory. The other direction is taking existing AI concepts and developing their applications. As in other fields, the two directions reflect the divergence of interest between basic scientists and applied scientists or engineers (see fig. 12). However, AI is distinct from other fields that have undergone this transformation in at least four significant ways: 1) AI is a very young (scientifically immature) field, 2) there are few trained AI practitioners, 3) the expectations for the technology are high, and 4) the pace of transformation to an applied study is fast. The divergence of interests is expected to eventually result in the concentration of the traditional "top-tier" university centers of AI research on the fundamental scientific questions, an expansion of industrial R&D efforts to deal with the development of applications, and the establishment of new academic programs for the training of applied AI scientists and engineers.

The growing commercial and military importance of AI is forcing researchers to make difficult choices in how they spend their time. They must balance a number of commitments: to the study of research issues that have long-term significance; to the teaching and supervision of students;[103] to consulting for industry and government; and to startup companies in which they develop personal interests. To be sure, not all of these commitments are mutually exclusive. Long-term research and student supervision can reinforce one another. Consulting and entreprenuerialism can be complimentary. But the wearing of many hats by AI faculty may pose conflicts apart from the overcommitment of time.

First, there is a danger that students may be judged, consciously or unconsciously, on their contribution to the business interests of faculty, rather than on their progress in learning the science of AI.[104] Second, the progress of AI, as well as that of other sciences, is dependent on the free exchange and public dissemination of research results. Yet proprietary interests may inhibit that flow of information.[105] Third, many researchers face a conflict between the objectives of the sponsors of research and their own conception of the best direction for the research. (See ch. 6 of this report on New Roles for Universities.)

This latter conflict may be particularly acute among those in AI research because of the fact that the social implications of artificial intel-

[101] The OTA workshop on Artificial Intelligence, held Oct. 31, 1983. There is some controversy on this point, but it is generally agreed that some "critical mass" of AI trained people, working in close proximity where they can exchange ideas, is a critical factor in performing advanced AI work.

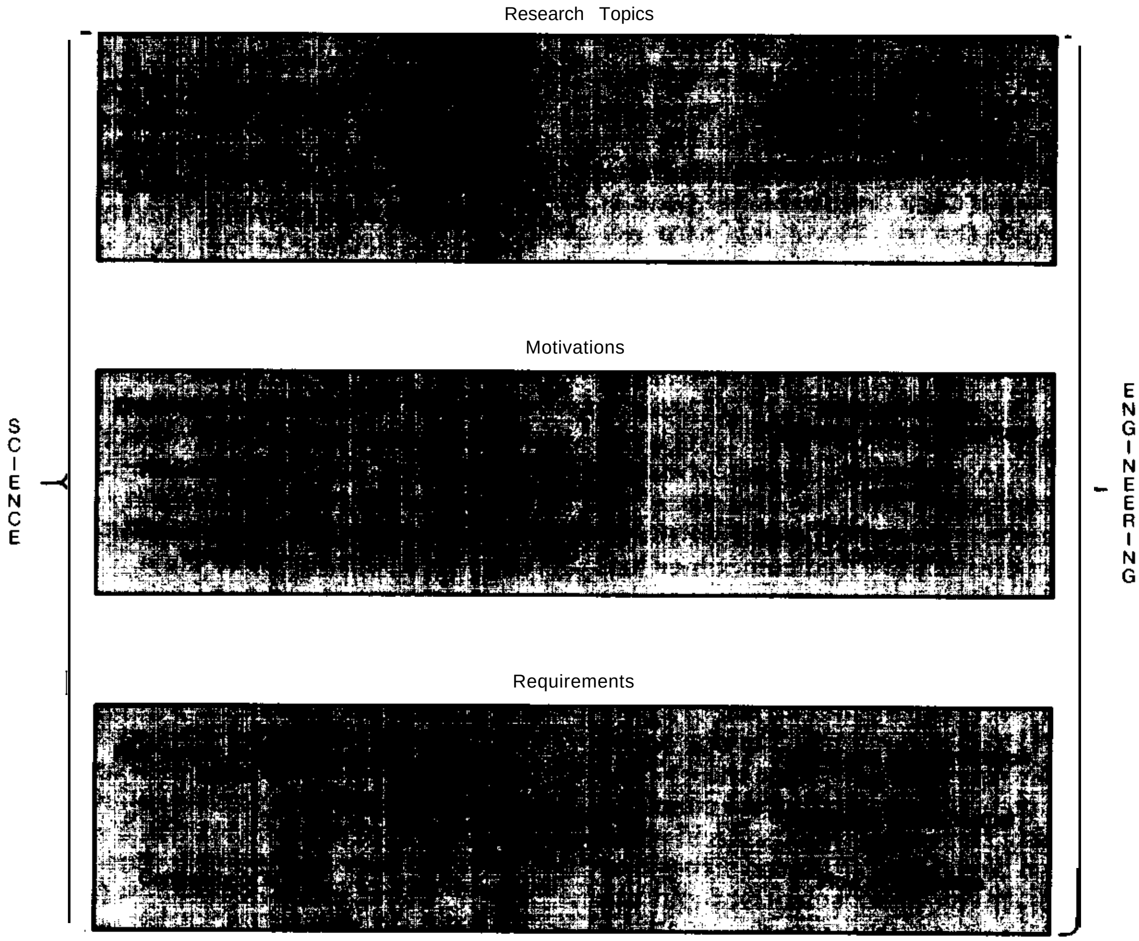
[102] Waltz, op. cit., pp. 64-65.

[103] One OTA advisor said that at his university, one quarter of the students now entering computer science and electrical engineering want to study AI. AI faculty across all schools are obliged to supervise, on average, some ten graduate students; in other fields the average ranges from less than one to about four. As well as teaching and advising, faculty must spend time writing proposals and seeking funding and equipment for graduate thesis work.

[104] MIT has developed a policy that prohibits faculty from supervising students that work for companies in which the faculty member has substantial interest. Patrick Winston at the OTA Workshop on Artificial Intelligence, October 31, 1984.

[105] This point was made by advisors at the OTA workshop on Artificial Intelligence, and is discussed by Jordan J. Baruch in an editorial in *Science*, Apr. 6, 1984.

Figure 12.—Artificiai Intelligence Science and Engineering



The evolution of the field of artificial intelligence is producing new relationships and new distinctions within the Al science and engineering community.

SOURCE. Office of Technology Assessment.

ligence are so profound. They include the replacement of humans with machines in skilled work, the increased reliance on automated systems and the risk of system failure, and affects on man's perception of himself, and his institutions. These conflicts raise ethical questions such as: how much responsibility do the creators of powerful technologies have for the social impacts of their creations? or, should scientists consider the public interest in deciding what research should be done?[106]

[106]The reader is referred to Relations of Science, Government and Industry: The Case of Recombinant DNA, " by Charles Weiner, ch. 4 in *Science, Technology, and the Issues of The Eighties: Policy Outlook*, A. H. Teich and R. Thornton (eds.), Westview Press, Boulder CO, 1980.

Until recently, artificial intelligence research was supported by four Federal agencies, primarily the Defense Advanced Research Projects Agency (DARPA) plus the National Science Foundation, the Air Force Office of Scientific Research, the Office of Naval Research, and a handful of companies. Now, significant levels of work are beginning to be funded in several corporations as the profit potential of AI grows. The Department of Defense plans to expand AI research in universities and defense contractor companies.

Industry Efforts

An increasing number of industrial firms have begun AI R&D efforts in the past few years. Market research sources report that industry spent a total of $66 million to $75 million in 1983 on AI products.[107] As many as 15 or more of the largest U.S. corporations are on the threshold of expanding the development of expert systems.[108] The Microelectronics and Computer Technologies Corp. (MCC) is devoting a fair amount of its early R&D efforts to AI related questions. [109] Three U.S. companies, Xerox, Symbolics, and Lisp Machine, Inc., currently offer computer systems especially designed for AI program development.[110] Four other computer companies, Sperry, Apollo Computer, Data General, Hewlett-Packard and Digital Equipment, are also developing AI computers.[111] Of most significance is the number of startup companies developing AI products, particularly expert systems.[112] At least a dozen firms have been founded in the past few years and they are obtaining people, techniques, and seed ideas from the top universities and from nonuniversity centers such as SRI International.

[107]T. Manuel, and S. Evanczuk, "Commercial Products Begin to Emerge From Decades of Research, " *Electronics, Nov.* 3, 1983, pp. 127-129. It should be noted that these numbers reflect a very broad definition of AI with which many experts would disagree.

[108]J. Johnson, "Expert Systems: For You?, " *Datamation, February 1984,* pp. 82, 84, 88.

[109]See ch. 6, in this report.

""These so-called "Lisp Machines' (because they use the AI programming language Lisp) represent an estimated market value of $50 million.

"'These machines are useful in many types of complex program development; nearly one-half of the sales of these machines, by at least one of the current vendors, are to users that are not necessarily developing AI applications. J. M. Verity, "LISP Markets Grow, " *Datamation,* October 1983, pp. 92-94, 98,100.

[112]See Verity, op. cit. p. 94, and Kinnucan, P., "Computers That Think Like Experts, " *High Technology,* January 1984, p. 42. One OTA advisor believes that venture capital will be a significant source of funding for AI R&D. At least $16 million has thus far been invested in startup companies developing expert systems, J. W. Verity, "Endowing Computers With Expertise, " *Venture,* November 1983, p. 49.

Government Funding

The Department of Defense is the lead agency funding AI research. The Defense Advanced Research Projects Agency (DARPA) established the three university "centers of excellence" at MIT, Carnegie-Mellon, and Stanford, and has awarded contracts since the early 1960s. The Office of Naval Research (ONR) and the Air Force Office of Scientific Research (AFOSR) have been funding AI work since the 1970s. The Air Force in particular is interested in AI techniques for image understanding. DARPA has proposed a significant expansion in its AI efforts (see "Strategic Computing" below). ONR and AFOSR funding is expected to remain stable in the near term. [113] (See table 19.)

The National Science Foundation has awarded grants for university and some industrial AI projects since the late 1950s. In recent years the NSF AI research budget has remained stable, between $5 million to $6 million annually. (See table 20.) Two NSF directorates, Computer Science and Information Science and Technology, provide the bulk of these funds, and a number of the awards (13 of 80 in 1983) are made jointly between these two divisions, or with the NSF Office of Interdisciplinary Research, or with other programs within NSF.

"STRATEGIC COMPUTING"

DARPA has embarked on a major effort to push the frontiers of computer technology and artificial intelligence for application in future military systems. The "Strategic Computing" program plans to spend $600 million over the next 5 years (see table 21), over and above past levels of funding, and spending will accelerate to unspecified levels as prototype development gets underway in the late 1980s. The project calls for the demonstration of significant mil-

1[1] 'Personal communications from Dr. David Fox, Director of Mathematics and Information Science, Air Force Office of Scientific Research, and Paul Schneck, Head, Information Sciences Division, Office of Naval Research.

Table 19.—Major Government Sources of Artificial Intelligence R&D Funding[a]

	Estimates	
	Low	High
National Science Foundation	5.0	6.0
National Institutes of Health	3.5	4.2
Office of Naval Research	1.2	1.4[b]
Air Force Office of Scientific Research	2.5	3.0
Defense Advanced Research Projects Agency[c]	15.0	20.0
Total	27.2	34.6

[a]Sources indicate that these levels have obtained for the past several years, and with the exception of DARPA, are expected to remain stable or increase modestly in the near term.
[b]This *does not* include $900,000 in robotics research, some of which is in artificial intelligence.
[c]This represents the base level of DARPA funding *before* the initiation of the "strategic computing" program (began in fiscal year 1984).

SOURCES: NSF numbers from *Summary of Awards* publications of the Division of Computer Research, FY 1982-83, and the Division of Information Science and Technology, FY 1981-82 and 1983, National Science Foundation. NIH numbers from Susan Stimler, Director, Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health. ONR numbers from Paul Schneck, Leader, Information Sciences Division, Office of Naval Research. AFOSR numbers from David Fox, Director, Division of Mathematics and Information Science, Air Force Office of Scientific Research, DARPA numbers from Ronald Ohlander, Program Manager, Defense Advanced Research Projects Agency.

Table 20.—NSF Grant Awards in Artificial Intelligence

	Fiscal year 1982	Fiscal year 1983–
Number–of Institutions	41	38
Number of awards		
Total awards granted ($OOO)[a]	$6,077	$5,150
Top 12 institutions:	(Dollars in thousands)	
MIT	946	240
Stanford University	485	587
Rutgers University	245	442
University of Maryland	351	317
University of Illinois	478	141
University of Pennsylvania	302	288
SRI International	302	202
New York University	360	130
Yale University	170	301
Carnegie-Mellon University	185	227
University of Massachusetts	87	229
University of Michigan	40	337
Total	3,951	3,441

[a]These numbers reflect the awards in each fiscal year and not I he amount of funding during those Years sixteen of the 67 awards made in 1982 were multi-year grants (most were 2 year grants), 5 of the 80 awards made in 1983 were multi-year grants Therefore, total funding probably Increased somewhat between 1982 and 1983

SOURCE Office of Technology Assessment, and the National Science Foundation, Divisions of Information Science and Tech. nology and of Computer Research

Table 21.—Strategic Computing Cost Summary

	Fiscal year				
	1984	1985	1986	1987[a]	1988a
	(Dollars in millions)				
Total military applications	$ 6	$15	$27	TBD	TBD
Total technology base	26	50	83	TBD	TBD
Total infrastructure	16	27	36	TBD	TBD
Total program support	2	3	4	TBD	TBD
Total	$50	$95	$150	TBD	TBD

[a]Out-year funding levels to be determined by program progress.

SOURCE: Defense Advanced Research Projects Agency.

itary capabilities within 10 years (see fig. 13).[114] The major areas of focus are:

- autonomous (unmanned) vehicles,
- pilot's associate systems to aid combat pilots in coping with the complexity of current and planned aircraft, and
- battle management systems to help commanders make decisions under conditions of uncertainty.

To lay the groundwork for the development of such systems, DARPA plans to fund the building of Galium Arsenide integrated circuit pilot production lines for the manufacture of low power and radiation resistant microelectronics. In addition, DARPA is focusing efforts on research and development of advanced computer architectures to meet the computational speed, and physical size and weight requirements of mobile systems (see the case study of Advanced Computer Architecture in this chapter), as well as artificial intelligence R&D in vision, speech, natural language, and expert systems.[115]

Some of the goals of the "Strategic Computing" program, particularly the vision system requirements for autonomous vehicles, have been described as "extremely ambitious."[116] Some describe this program as unprecedented in the history of U.S. Government funding of science and technology.[117] Unlike the Manhat-

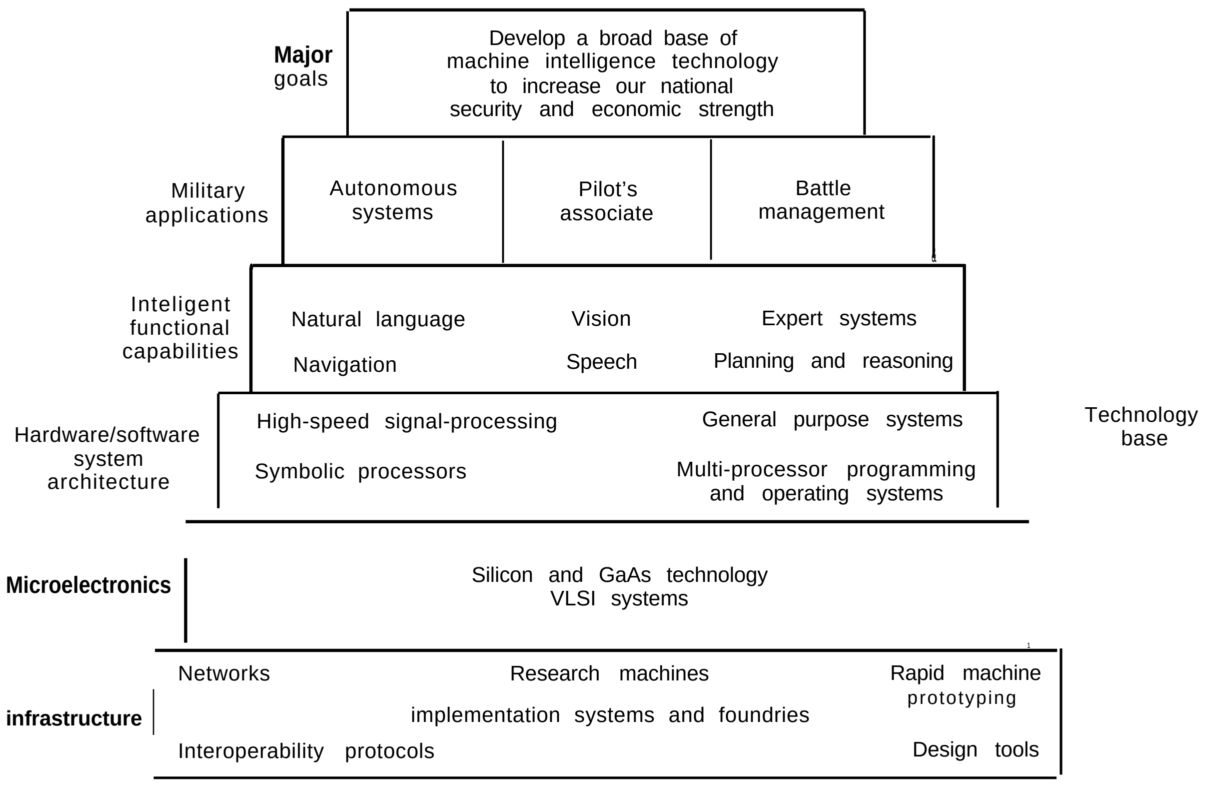
[114] Strategic Computing, "New-Generation Computing Technology: A Strategic Plan for its Development and Application to Critical Problems in Defense," Defense Advanced Research Projects Agency, Oct. 28, 1983, p. vii.

[115] Ibid., app. IV, p. 2.

[116] Ronald Ohlander, at the OTA workshop, Oct. 31, 1983; see also *Strategic Computing*, pp. 22-23.

[117] Mitch Marcus at the OTA workshop, Oct. 31, 1983.

Figure 13.—Strategic Computing Program Structure and Goals



SOURCE: DARPA.

tan Project or the Manned Moon Landing Mission, which were principally engineering problems, the success of the DARPA program requires basic scientific breakthroughs, neither the timing nor nature of which can be predicted.

> The kind of work that is being done now [in AI research] that would support, for instance, an autonomous vehicle system, is primitive in relation to the problems that are to be addressed [by the DARPA project]. Operation in a complex combat environment that may have multiple targets with camouflage, different kinds of obstacles with varying degrees of threat and impedance associated with them, and the integration of various kinds of sensors, for example touch and vision, and the appropriate knowledge representation to deal with them, is an enormous problem that will be solved only by significant strides in basic AI research, and not just development in narrow vehicle applications.[118]

It is expected that the Strategic Computing Program will approximately quadruple the annual Federal funding of R&Din AI and related hardware.[119] This major increase in R&D funding will have significant and far reaching effects on the artificial intelligence community and research environment. Undoubtedly, the increased flow of money will have some positive effects on AI research. In particular the availability of funds for modern equipment should make university laboratory facilities comparable with increasingly well equipped industry labs.[120] The increased availability of graduate student financial aid in the form of research assistantships should draw in and help hold more qualified post-graduate trainees, thus expanding the potential faculty base, and relieving some of the pressure on current faculty to obtain funding for graduate research. However, as industry scales up to meet military requirements, additional pressures on the limited AI manpower resources are expected.[121] Although the expansion of training programs should eventually increase the supply of manpower, the rate of that expansion is limited by the small existing faculty base.[122] The current imbalance in the supply and demand of AI manpower will continue, and likely increase, and may intensify competition among commercial, military and academic AI R&D agendas. Thus some applied research, such as work toward "intelligent library systems" and computer assisted education, and fundamental research that is unlikely to have high immediate commercial or military value, may be neglected.[123]

Content and Conduct of Artificial Intelligence R&D

Artificial intelligence as a field is a set of somewhat loosely related R&D activities that range from the study and implementation of machine sensing (e.g., vision and speech) and pattern recognition algorithms and systems, to theoretical and practical work on automatic problem solving, inferencing and reasoning strategies and programs. Of central concern is the concept of knowledge, a set of information (facts, procedures, patterns) that is integrated and processable as a whole, and thus constitutes a useful representation of a part or domain of the world. AI deals with how knowledge is built up and used in computer-based systems: how it is collected, stored, accessed, manipulated, and transferred. AI R&D seeks methods of formalizing and representing knowledge in consistent and unambiguous, yet flexible ways so that these tasks can be performed by machines.

[118] Marvin Denicoff, at the OTA workshop, Oct. 31, 1983.

[119] From about $30 million to approximately $120 million (averaging the $600 million program over its 5 year projection). One DARPA source estimated that one-fourth of the program funds would be spent in universities. (Duane Adams at the OTA Workshop on Advanced Computer Architecture, held July 14, 1983.)

[120] This may be an inducement for researchers to accept research appointments.

[121] Ronald Ohlander, at the OTA workshop, Oct. 31, 1983.

[122] In the three AI subfields from which most commercial and military applications are expected, expert systems, natural language understanding, and vision, there are approximately 60 faculty at the top schools turning out some 30 Ph.D. graduates per year, half of whom take industry jobs. D. Waltz, "Artificial Intelligence: An Assessment of the State-of-the-Art and Recommendations for Future Directions," prepared for the NSF Information Technology workshop, Jan. 5-7, 1983; and a personal communication from Azriel Rosenfeld, Apr. 25, 1984.

[123] David Waltz at the OTA workshop, Oct. 31, 1983. A private funding organization, the Systems Development Foundation, is sponsoring research directed at "intelligent libraries."

Although investigators in a number of fields perform work that contributes to and overlaps with the subject matter of AI, a number of areas are considered the core of AI research.

Symbolic Computation

Since AI is concerned with the processing of knowledge, intelligent machines are required to manipulate symbols that represent objects, concepts, and qualities, as well as symbols that represent numbers or quantities, which are the focus of traditional computation. These qualitative symbols may be represented within a computer as logical relationships or as lists of symbols with pointers linking various objects, concepts, and qualities. The manipulation of symbols provides a mechanical means of achieving inference, in particular deductive inference, in which a problem solution or a conclusion is tested against facts in a *knowledge base* to determine its truth or validity.

A branch of AI research and development, which forms a link with R&D in advanced computer architectures, is the design of computing machines that optimize efficient manipulation of symbols. Some in computer architecture research maintain that this R&D effort has not received sufficient attention.[124]

[124] Sidney Fernbach at the OTA workshop on Advanced Computer Architecture, July 17, 1983.

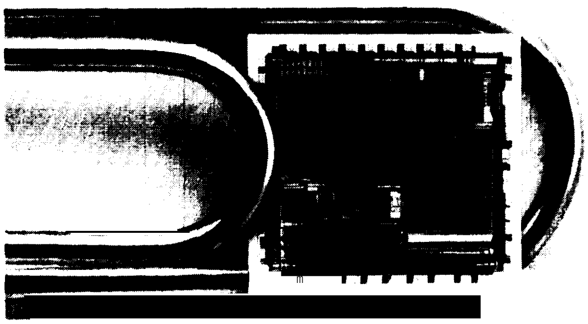


Photo credit, AT&T Be// Laboratories

Specialized microprocessors promise increasing levels of "intelligence" in computer systems

Pattern Recognition

Of particular importance to the branches of AI concerned with sense perception, such as machine vision and speech understanding, is the idea of matching input from the environment with symbolic representations of patterns (e.g., visual objects or speech utterances) stored in the system. Raw sensory input, in the form of electrical signals from a television camera or a microphone, must proceed through several levels of processing to produce patterns that are comparable with stored symbolic representations and recognizable as having unique meaning.

In general, these processing levels include:

- **Formation:** The sound or light signals are digitized and stored as a set of simple physical parameters such as frequency and amplitude.
- Analysis: Patterns of variations in the parameters, for example areas of light and dark in pictures or variations in pitch and intensity in an utterance, are detected and measured to produce a detailed physical description of the input.
- Interpretation: The patterns, now represented by sets of measurements, can be either directly compared with *templates,* stored descriptions of entire patterns, or further analyzed to extract *features,* which are parts of patterns that are of particular value in defining and identifying the patterns.

Knowledge Representation

This area of research deals with ways of expressing knowledge in computable form and making and exploiting connections among the facts or propositions in a knowledge base. A basic concept in knowledge representation is the *propositional formalism,* which is simply the structured way that facts are expressed so that ambiguities are avoided and processing operations are as efficient as possible.

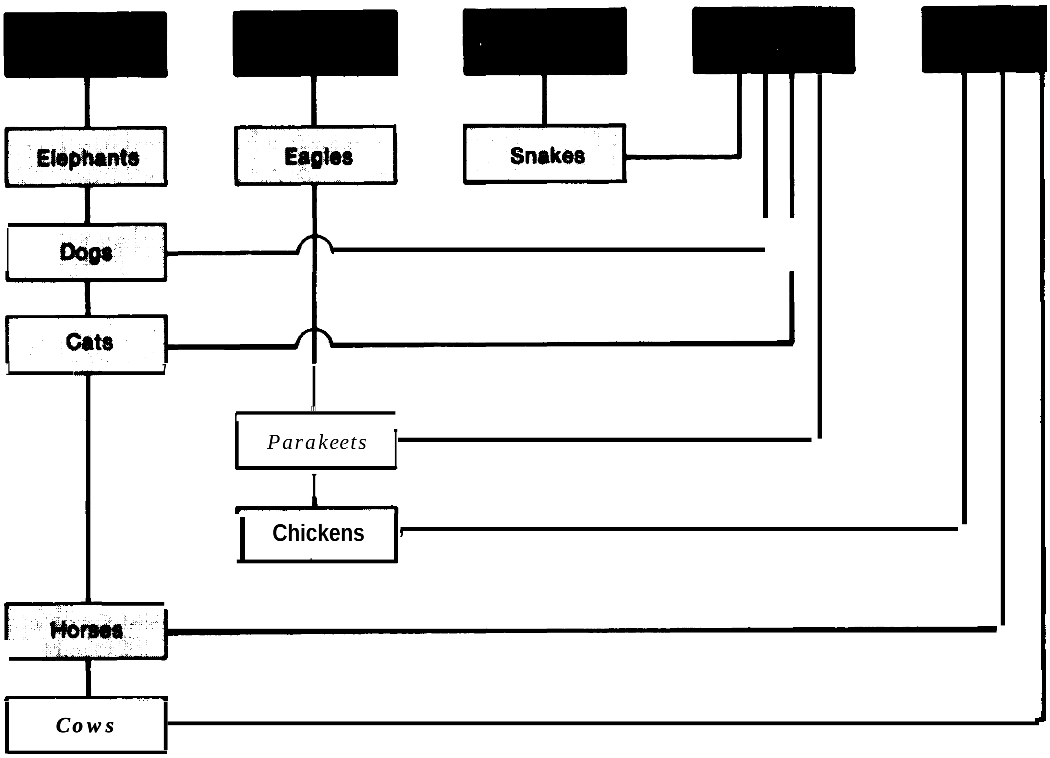
As well as a formal structure or syntax, knowledge representation requires a method of capturing meaning, or *semantics*, in a knowledge base. Meaning in a qualitative sense is expressed by the relationships that are stated to exist among concepts. One method of representing relationships is the *Semantic Network (see* fig. 14).

A knowledge representation scheme that is similar to semantic networks has been developed in which the *attributes* of concepts are expressed and related. *Frames* provide slots to fill to structure knowledge about things, thus they represent expectations about what attributes members of a given frame will possess. These "expectations" can be exploited to imply procedures. For example, a computer can be programmed to inquire about the attributes of a newly introduced object.

Similarly, knowledge representations called *scripts can* represent expectations about a knowledge domain. But instead of representing concepts, scripts describe *situations* and *actions* that are expected in those situations. A classic example is a restaurant script in which one expects people to order from a waiter (or waitress), to eat, and to pay the bill. Thus, invoking the restaraunt script would invoke the entire set of expectations surrounding restaurants that have been programmed.

Another knowledge representation scheme, one that was developed in the mid-1960s, is the *production system*. This essentially consists of a set of *if-then rules* that specify a pattern, for example, "if the temperature is less than 65 degrees" and an action to be taken, "then turn the furnace on. " The importance of this scheme is its ability to express *procedural knowledge* and to initiate operations depending on prevailing conditions. Its major weakness is that a fixed rule must be stated for every condition that is to be encountered.

Figure 14.–A Semantic Network



SOURCE: Randy Davis and Chuck Rich.

All of the knowledge representation schemes thus far adopted by AI researchers use considerable computer memory if they contain a sizable or complex domain of knowledge. Traditionally, researchers and builders of programs have had to "shoe-horn" knowledge into computer systems with limited and expensive memory capacity. The precipitous decline in the cost of computer memory is alleviating some traditional AI problems, making possible for the first time the production of cost-effective systems with enough knowledge to tackle "real-world" problems. Advances in computer architecture design, especially the ability to make customized VLSI processors dedicated to symbolic computation and capable of parallel processing, should have a great impact on the kinds of problems that may be addressed by knowledge-based systems.

Other difficulties inherent in the application of existing knowledge representation schemes will not necessarily be alleviated by lower memory costs and advances in hardware capability because they involve the limitations of the schemes themselves. Existing knowledge representations deal only with *surface knowledge,* which is either explicitly stated in relationships or deductively inferred by the chaining of propositions. *Deep knowledge,* knowledge gained from inductive inference (reasoning from facts to general principles), and *commonsense knowledge* that is routinely learned by children through experience, are not expressable using current knowledge representation systems. Other problems involve the inflexibility of procedural knowledge and the difficulties of programming machines to reason about unforseen occurances.

Given the large collection of facts, relations and patterns in a useful size knowledge base, even the fastest computers can bog down. Therefore, an AI program must concentrate the search of its knowledge base on those portions that are most likely to hold the facts, patterns, or solutions that are needed in a given situation. The problems associated with the search of knowledge bases are ubiquitous in AI. They result from what is termed the "combinatorics explosion": as the size of a knowledge base increases, the number of possible paths to a sought-after piece of information increases exponentially. AI interest was originally in developing general *knowledge* AI systems, but the combinatorics explosion sets a limit on how broad or useful such systems can be. More recent work has focused on specific problem domains where a detailed representation of the world can be built and exploited, and where the search is controlled by *knowledge about the problem domain,* or heuristics.

Machine Learning

The study of *heuristics* ("rules of thumb" or knowledge acquired from experience), how they may be used to process knowledge, and how they may be generated automatically by computer systems are major pursuits of AI research. Ideally, a computer could modify its repertoire of procedures as new facts are added to its knowledge base. Likewise, new procedures should suggest new relations among facts, but current knowledge representation and programming techniques are incapable of supporting changing knowledge and rule bases.

Commonsense Knowledge

Human knowledge is based on a wide range of experience acquired through sense organs. This experience is fed into a cognitive system that can integrate that knowledge in order to cope with a complex, dynamically changing world. Much of what people know involves relationships that are obvious but that depend on explanations beyond the comprehension of most. Representing commonsense knowledge in computer systems is a major challenge to researchers, and is essential for certain kinds of intelligent machine behavior, such as reasoning about the chronology of events in the course of a disease or reasoning about the movement and characteristics of robots.

Pragmatic Knowledge

Current artificial intelligence systems employ syntax (structure) and semantics (meaning) to represent knowledge about the domains in which they operate. Systems of higher intelligence, ones capable of choosing among a range of responses in complex situations, of adapting to changing (and unanticipated) conditions, and of interacting with humans and perhaps other intelligent systems on a high level of understanding, must employ ways of representing and reasoning about the needs *of users* and about *knowledge itself*. Truly intelligent machines would respond to people of different backgrounds and needs in ways that are appropriately tailored to individuals. They would also possess an understanding of the uses and limits of knowledge-a mechanism for judging the appropriateness of information to a given situation. For example, a robot tank would need to be able to distinguish between trees it could run over and those that it would have to avoid.

Current syntax- and semantics-based artificial intelligence systems employ knowledge representations that are inadequate to support *pragmatic* knowledge. Frame- and script-based representations can, in a rudimentary way, deal with expectations about situations, but those situations must be tightly circumscribed and clearly defined at the time of program design. Research has yet to identify knowledge representation schemes that can support the processing of pragmatic knowledge in complex and dynamically changing domains.

Commercially Available Artificial Intelligence

Notwithstanding the fundamental limitations of present artificial intelligence concepts, programs are being developed that are proving useful in a widening array of applications. Two market forecasts[125] describe the commercial potential of AI systems as exploding over the next decade. From the small 1983 sales base of $66 million to $75 million, the market for AI products is expected to rise to $2.5 billion to $8.5 billion by the early 1990s.[126] The changing mix of products and sectors in which they will be used is depicted in figure 15, which shows AI emerging from the laboratory into the home, the factory, and into schools. Three types of AI-based systems in particular are expected to experience rapid growth-vision systems, natural language understanding systems, and expert systems.

COMPUTER VISION SYSTEMS

Vision systems are being designed as input subsystems to enhance the utility and flexibility of computer-based automation. Vision systems are increasingly used in industry for quality control inspection, for product identification, and for robot guidance and control.[127] Machine vision will be a crucial function in many future automated systems, including aspects of the DARPA "Strategic Computing" initiative. For example, an advanced vision system will be an integral component of autonomous vehicles.

Some sources estimate a current annual market for machine vision systems of $35 million. This market is projected to double each year for the next 5 years, and should reach about $1 billion by the end of the decade.[128] More than 250 companies, most of them in the United States, are designing and selling machine vision systems.[129]

Vision systems consist of cameras and computer processors to analyze and interpret the information collected by the cameras. The cameras provide the processor with frames, typically 60 per second, collected by scanning the camera's visual field. A frame forms an image consisting of a matrix of dots, or pixels (picture elements), numbering in current systems 64 X 64 or 256 X 256 pixels. Higher resolution cameras are being developed, but as the resolution (number of pixels per frame) increases, higher speed processors are needed to handle the increased information within rea-

[125] International Resource Development, Inc., and DM Data, Inc.

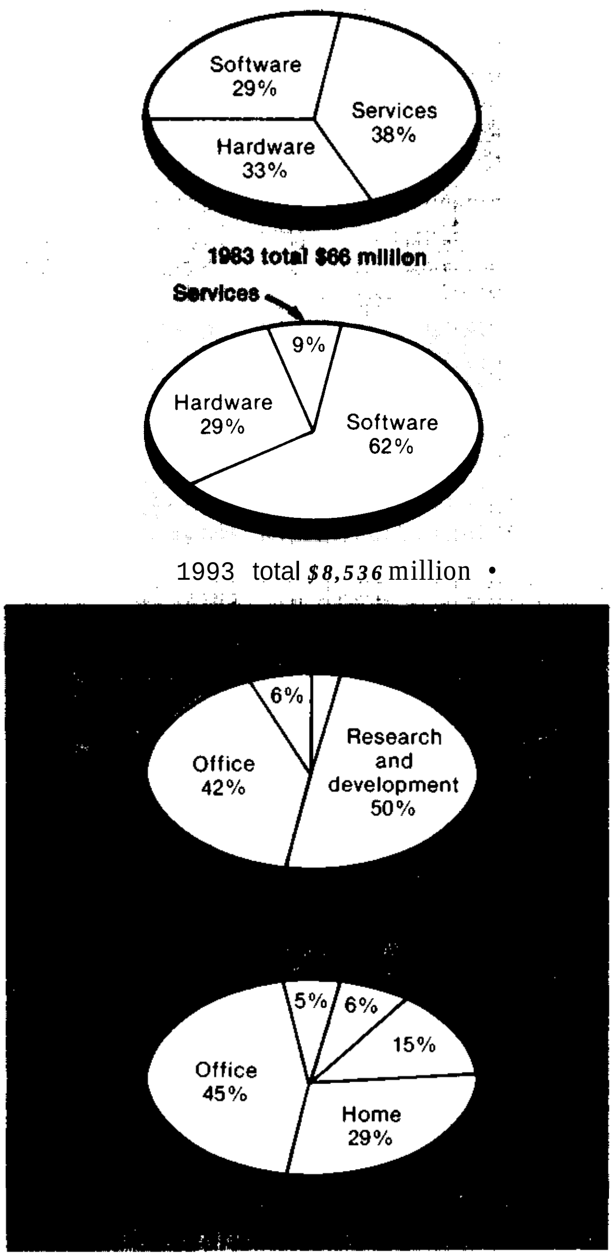
[126] Manuel and Evanczuk, op. cit., pp. 127 and 129.

[127] The reader is referred to the OTA report, *Computerized Manufacturing Automation: Employment, Education, and the Workplace*, April 1984, pp. 89-92, for a discussion of current industrial applications of machine vision systems.

[128] "Machines That Can See: Here Comes a New Generation," *Business Week*, Jan. 9, 1984, pp. 118-120.

[129] Ibid., p. 118.

Figure 15.— Markets for Artificial Intelligence



SOURCE International Resource Development Inc.

sonable time limits.[130] Each pixel represents an area of light or dark in the image. Most systems allow a pixel to represent only black or white, though systems are being developed and are in use in which pixels may be one of as many as 64 shades of gray, and color vision systems will soon be available. Gray scale and color systems will be capable of higher acuity, but such systems require much higher processor speed and more sophisticated algorithms to interpret the increased information flow.

The processor analyses the image to extract patterns that may represent edges or textures that can be used to characterize and recognize objects in the visual field. Simpler systems have stored "templates," or representations of entire objects, in memory to which the extracted patterns are compared. More sophisticated systems store "features," or characteristic parts of objects. In these systems, sets of features are extracted from the image and are combined to represent constraints on the list of possible objects that may be present; recognition involves the identification of an object (or objects) that could satisfy those constraints.

A major problem in machine recognition is inferring three-dimensional information from two-dimensional information in an image of the scene.[131] One method that has been adopted to solve this problem is to project "light stripes" on objects in the scene; depth cues and contours can be inferred from distortions in the stripes.

Current vision systems are primitive in comparison to human visual recognition capability. The human retina can perform at least 10 billion operations per second, and the brain is undoubtedly capable of much higher processing speeds.[132] As important, human experience with the visual world provides a store of knowledge with which the eye and brain may make inferences and interpretations to resolve

[130] A Stanford University researcher has developed a robot cart that can navigate in a simple environment at a speed of 3 to 5 meters per hour while analyzing one image of its surroundings for each meter of progress using a computer capable of processing a million instructions per second. For such a system to naviagate at walking speed would require a computer capable of from 10^9 to 10^{10} instructions per second. T. Kanade, and R. Reddy, "Computer Vision: The Challage of Imperfect Inputs," *IEEE Spectrum*, November 1983, pp. 88-91. Computers of this speed are being developed, but may require facilities as large as a room just to provide cooling for the hardware.

[131] Azriel Rosenfeld, personal communication, May 25, 1984.

[132] Kanade and Reddy, op. cit., p. 88.

ambiguities caused by shadows and variations in lighting, the orientation of objects, and the presence and often overlap of multiple objects in a scene. Also, subtle variations and movements can be detected and used by the human eye and brain to distinguish superficially similar objects (e.g., the faces of identical twins).

In order to make vision systems based on current processor architectures and algorithms work, the number of environmental variables that may cause ambiguities must be reduced. For example, objects on an assembly line may be arranged so that only one face is presented to the camera or only one object at a time is in the visual field. Lighting may be controlled so that perceived edges can be interpreted as facets of an object and will not be the result of shadows or variations in the reflectance of surfaces of an object; or structured lighting may be used to infer distance and contour.

If vision systems are to operate in less controllable circumstances, such as out-of-doors or in the home, much higher speed processors and more sophisticated algorithms for analyzing and interpreting visual scenes will be needed. Custom designed, massively parallel VLSI-based processors dedicated to visual analysis are being looked to to provide more powerful hardware. Research in vision algorithms is proceeding to provide more powerful concepts and more sophisticated software. The parallel advance of hardware and software solutions, and fundamental advances in AI research, will be required to produce systems capable of the interpretation of complex and unstructured visual scenes.

NATURAL LANGUAGE UNDERSTANDING SYSTEMS

These systems are intended to allow communication with computers in English or other human languages, freeing users from the need to learn a computer language. In the words of one researcher:[133]

> The ultimate goal of creating machines that can interact in a facile manner with humans remains far off, awaiting breakthroughs in basic research, improved information processing algorithms, and perhaps alternative computer architectures. However, the significant progress experienced in the last decade demonstrates the feasibility of dealing with natural language in restricted contexts, employing today's computers.

Automatic natural language understanding work began with the goal of producing systems that could translate one natural language into another. Such systems are in use now, but are proving to be limited to rough translation, requiring a human translator to produce final fluent text.

The goal of another class of natural language systems is to understand documents: to produce an abstract of a piece, perhaps alert people who might be interested, and answer questions on its content. Such systems might be paired with document generators that could produce instruction manuals.[134] Document understanding and generation systems are still largely experimental, but they have the potential to augment human knowledge through the production of intelligent library systems, aiding people in searching and evaluating large bodies of information.

More immediate-term applications are in using natural language systems as "front-ends" or interfaces to computer systems. Natural language systems are in operation that serve as interfaces to data bases, for example, to help store and find personnel or inventory records. One can expect that model systems will soon be available that demonstrate the feasibility of controlling complex systems such as power generation facilities and weapons systems with natural language interfaces. In conjunction with expert systems, natural language systems are being developed for computerized medical advisor, trouble-shooting and repair, mineral exploration, and investment analysis applications. Future applications, requiring advanced work to produce, might include use in the creation of graphical displays and in computer-aided education.[135]

[133] Waltz, *AI Magazine*, op. cit., p. *56.*

"'Ibid., p. 57.

"'Ibid.

There is also current work in producing *speech understanding* systems, which would replace keyboards with voice input devices. Primitive systems are available, but true *speaker independent* systems with large vocabularies that can handle *continuous speech,* as opposed to isolated words separated by pauses, are as yet unattainable with present technology. Advanced integrated circuit signal processors should produce some progress over the next decade.[136]

In part, the capabilities of natural language systems have been limited by the cost of computer memory. Natural language systems must have large vocabularies to be "natural' to users. This problem is diminishing because of the fall in computer memory prices.

Two other problems are replacing memory cost as an upper limit to system size and functionality. First is the "combinatorics explosion" problem: as the size of the system vocabulary increases, the time required for presently available computer architectures to process a sentence increases exponentially. Novel machine architectures based on parallel processing promise some relief for this problem. Second, as noted in the previous discussions of commonsense reasoning and pragmatic representation, researchers are finding difficulties in designing systems with broad and deep knowledge of the world, and with pragmatic understanding of situations. In language, this kind of knowledge is crucial to the solution of ambiguities.

EXPERT SYSTEMS

These AI-based programs are designed to serve as consultants in decisionmaking and problem solving tasks that require the application of experience and judgment. '37 Expert systems consist of a set of "if-then" rules which express the knowledge and experience of an expert, and the actions one would take when faced with a set of conditions in the domain of his expertise. They also generally have a separate knowledge base which states facts about the domain, to which the program can refer to make inferences and deductions about given situations and conditions.

Expert system programs are developed from extensive interviews with recognized experts in the field of application. Such interviews often reveal many unwritten (and even unconscious) rules and criteria for judgment that the expert applies in solving problems.[138] The interviewer, called a *knowledge engineer,* then codes the rules and facts in a way that is efficiently processable by computer. Extensive testing and validation must be conducted before a system is considered complete enough to use in actual practice.

This area of AI application is receiving considerable industry and military interest because expert systems may be capable of relieving some of the demand for high-priced experts in the fields in which they are applied. Ironically, because they require man-years of effort from scarce knowledge engineers to develop, expert systems are quite expensive. Therefore they are currently being attempted commercially only in applications that promise a particularly high payoff.[139]

Research in this field has demonstrated that expert systems can rival human expert judgment *if the domain of application is* properly

[136]For discussions of the state-of-the-art in voice input computer systems, the reader is referred to: A. Pollack, "Computers Mastering Speech Recognition," *New York Times,* Sept. 6, 1983, pp. Cl and C7; R. D. Preuss, and D. J. Jurenko, "Digital Voice Processing," *Astronautics and Aeronautics,* January 1983, pp. 44-46; and R. J. Godin, "Voice Input Output, " *Electronics,* Apr. 21, 1983, pp. 126-143.

"'Expert systems illustrate an interesting point about human intelligence, "Paradoxically, it has proven much easier to emulate the problem-solving methods of some kinds of specialists than to write programs that approach a child's ability to perceive, to understand language, or to make 'commonsense' deductions, " R. O. Duda, and E. H. Shortliffe, "Expert Systems Research, " Science, Apr. 15, 1983, pp. 261-268.

[138]This fleshing out of unwritten rules is a possible side benefit of the development of an expert system; it can lead to progress in the field of application itself. Conversely, some fear that this codification of knowledge in automated systems could "ossify" and constrain progress in certain fields by foreclosing the possibility of the unexpected discovery of new solutions through serendipity.

[139]Some of the enthusiasm recently derponstrated by companies for expert systems may be unfounded; OTA advisors were wary of overstated claims for the usefulness and applicability of current generation expert systems.

chosen.[140] This means that the application must have certain characteristics for an expert system to be successfully applied. These include: 1[4]1

- There must be recognized experts in the field;
- The task normally takes the expert a few minutes to a few hours to accomplish;
- The task is primarily cognitive (as opposed to manual);
- The skill is teachable to neophytes;
- The task requires no commonsense.

More generally, the application domain must be restricted enough to be expressable in a finite set of rules, numbering less than 2,000 at present, so that available computers can solve problems in a reasonable amount of time. Even so, most current expert systems take longer to perform a task than a competent human expert.[142]

Table 22 shows a representative sample of current expert systems and their uses. Several companies are forming or expanding efforts in developing expert systems, and venture capital interest in startup companies is reportedly brisk.[143]

[140] In fact, one OTA advisor said that the success of current applications depends more on the choice of domain than it does on how well the program is written. Patrick Winston at the OTA workshop, Oct. 31, 1983.

[141] R. Davis, and C. Rich, *Expert Systems: Fundamentals*, Tutorial at the AAAI Conference, Aug. 22, 1983, p. 31.

"'Waltz, *AI Magazine*, op. cit., p. 61.

[143] See *Electronics*, Nov. 3, 1983, pp.127-131, and *Venture*, November 1983, pp. 48-53.

Current systems, because of the limitations of the concepts on which they are based, all suffer from several serious weaknesses:[144]

- They are highly customized to specific applications and are useless in other, even closely related fields;
- Since the knowledge on which they are based is collected over a period of time, often from a number of experts, there can be inconsistencies in the programs that are difficult to detect and repair;
- The systems are necessarily based on narrow sets of rules and facts, therefore their judgments and recommendations can be myopic and naive;
- Since all of the current systems contain only surface knowledge from which to make inferences, if knowledge that is critical for a given judgment is missing, their performance is poor;
- Few current systems have natural language interfaces, therefore they can be difficult for the uninitiated to use.

Although many systems can provide explanation for the chain of reasoning that led to a given conclusion, these explanations are often unsatisfactory because they are not tailored to the needs or understanding of individual users. Neither do the explanations usually refer to underlying principles, such as physiology or geology, so human experts find them unconvincing.[145]

"'Waltz, op. cit.

[145] Duda and Shortliffe, op. cit., p. 266. See also G. O. Barnett, "The Computer and Clinical Judgement, " editorial in *The New England Journal of Medicine*, Aug. 19, 1982, pp. 493-494.

Table 22.—Representative Expert Systems

Expert system	Domain	Type of evaluation	Routine use
DENDRAL	Mass spectroscopy interpretation	Case studies	Yes
MYCIN	Antimicrobial therapy	Randomized trials	No
INTERNIST-1	Internal medicine diagnosis	Case studies	No
CASNET	Glaucoma assessment and therapy	Case studies	No
PROSPECTOR	Geological exploration	Case studies	No
RI	Computer layout and configuration	Case studies	Yes
Digitalis Advisor	Digitalis dosing advice	Randomized trials	No
PUFF	Pulmonary function test interpretation	Randomized trials	Yes
Microprocessor EXPERT	Protein electrophoresis interpretation	Case studies	Yes
HASP and SIAP	Ocean surveillance (signal processing)	Case studies	No

SOURCE' R. O. Duda and E. H. Shortliffe, "Expert Systems Research," *Science,* Apr. 15, 19s3.

Research is underway to help alleviate some of these shortcomings, including the study of improved knowledge representation schemes and inference methods. Novel computer architectures should push back some of the limitations now imposed on system size by computer speed. The most immediate and probably far-reaching problem is the difficulty and expense of building and testing expert systems, including the acquiring and validating of rules and facts. Applied research work is progressing in the development of programming aids (tools, languages, environments) to lower the time and cost of building expert systems. Some current basic research is concerned with the automatic acquisition and structuring of knowledge. Incremental advances can be expected over the next decade, but fundamental breakthroughs will be required to expand the usefulness of expert systems significantly beyond current limits.

Foreign National Efforts in Artificial Intelligence R&D

Artificial intelligence is a key pursuit of recently initiated national, government financed R&D programs in Japan and Great Britain. Britain's effort is in fact a reaction to the Japanese Fifth Generation Project [146] as it was detailed in a conference held in October 1981. '47 The achievements of the objectives of these targeted national programs will require unprecedented fundamental advances in basic research. Their ultimate success is much less certain than were previous national technical efforts, such as the Manhattan Projector the Manned Moon Landing. However, these national AI research commitments are likely to produce advances in AI science, technology, and commercial application.[148]

The Fifth Generation

The Japanese plan to build, from a base of research that is almost exclusively borrowed from the United States and Western Europe, significant artificial intelligence systems with the stated purpose of enhancing productivity in areas of the Japanese economy that thus far have proven resistant to automation.[149] The total funding for the life of the project is not publicly stated, but some sources estimate that up to $1 billion to $1.5 billion may be spent over the next 8 to 10 years.[150]

The Fifth Generation Program is centrally managed by the Institute for New Generation Computer Technology (ICOT), formed by the Ministry of International Trade and Industry (MITI) in 1982. '61 The program will be funded, for the most part, by eight industrial companies. The government has provided seed money and staffing; 42 of the total staff of 52 have come from MITI's Electrotechnical Laboratory. [152] Spending thus far has been $2 million in 1982, $12 million in 1983, and $27 million in 1984.[153]

Overall, the functions that the Japanese see their Fifth Generation system performing include: [154]

1. problem-solving and inference,
2. knowledge-base management, and
3. intelligent interface (with users),

These goals correspond roughly to previously discussed research in:

1. expert systems,
2. knowledge representation, and

[146]A British delegation reported that the extent and cohesiveness of the Japanese plan, and the reaction to it that could be expected from American industry, were a "major competitive threat. " *A Programme for Advanced Information Technology, The Report of The Alvey Committee,* Department of Industry, Her Majesty's Stationery Office, London, October 1982, see p. 5.

[147]*The International Conference on Fifth Generation Computer Systems,* Tokyo, Japan, Oct. 19-22, 1981.

[148]OTA workshop, Oct. 31, 1983.

[149]These include agriculture and fisheries, the office, and the service industries; applications are expected, as well, in areas that are already highly automated such as electronics and automoble design and manufacturing. *Fifth Generation Computer Systems,* T. Moto-oka (cd.), op. cit., see p. 3.

[150]P. Marsh, "The Race for the Thinking Machine," *New Scientist,* July 8, 1982, p. 85-87. Note, this is the *highest* estimate encountered.

[151]" 'Fifth Generation Computer System Under Development, " *Science & Technology in Japan,* January/March 1983, p. 24-26.

[152]" ICOT: Japan Mobilizes for the New Generation, " *IEEE Spectrum,* November 1983, p. 48.

[153]Personal communication from John Riganati, National Bureau of Standards, June 1984.

[154]*Science & Technology in Japan,* op. cit., p. 24.

3. natural language understanding.[155]

The Fifth Generation R&D plan calls for three phases of 3 to 4 years each:[156]

- **Phase I:** Development of "basic technologies" consisting of computer architectures and related software.
- Phase II: Development of prototypes of the inferencing and knowledgebase functional components, to correspond to the architecture and software developed in phase I.
- Phase III: A "total system" prototype is to be developed in the final phase:

The reactions of experts in this country to the Fifth Generation Program goals and methods have been mixed, ranging from enthusiastic to skeptical. However, there is agreement that the goals the Japanese have established are quite ambitious for the time frame that is set for achieving them, and the program is unlikely to succeed in all of its objectives.

The Alvey Programme

Named for John Alvey, chairman of the committee that developed the plan for a British response to the Japanese Fifth Generation Program, this academia-government-industry cooperative effort will concentrate on four segments of information technology, one of which is artificial intelligence, termed by the British, "Intelligent Knowledge Based" Systems.[157]

The British see a manpower shortfall in AI R&D similar to that in the United States, so their program will initially have a strong educational component, focusing on the expansion of both academic and industrial training facilities in the early part of the proposed 10 year effort. As the program progresses, emphasis will shift first to a broad effort in many aspects of basic AI research, and will later focus on specific demonstration systems to be developed in collaboration between the strengthened university programs and industries interested in particular applications.[158] Expected milestones include:[159]

- The research community (numbered in 1982 at some 150 people) is to grow by 50 percent in 2 to 3 years and double in about 5 years as a result of increases in graduate student funding, and numbers of faculty and research positions;
- Computer equipment and software support are to be expanded into an adequate base for research in 2 to 3 years;
- An increased understanding will be gained of knowledge representation and intelligent computer interface concepts, and knowledge based tasks and domains within 2 to 3 years;
- Within 5 years, there will be substantial progress in understanding the application of knowledge based concepts using logic programming languages and parallel processors, expert systems and natural language understanding systems;
- Over the 10 year time course of the program, progressively more sophisticated demonstration systems will be developed, some early target applications being teaching assistant programs, software production aids, and improved robot systems; later applications might include medical advisor systems, tactical decision aids, full 3-D vision systems, and office management and document production systems.

The government has pledged a total of $310 million for the entire Alvey program, with private industry expected to contribute an additional $230 million.[160] The plan calls for expenditures of about $39 million on Intelligent Knowledge Based Systems in the first 5 years, some $30 million of that going to universi-

[155] L. R. Harris, "Fifth Generation Foundations," *Datamation*, July 1983, pp. 148-150, 154, 156.

[156] Ibid., p. 25.

[157] The other areas are Software Engineering, Man/machine Interfaces, and VLSI. *A Programme for Advanced Information Technology*, op. cit., p. 21.

[158] Ibid., p. 35.

[159] Ibid., pp. 38-40.

[160] M. Peltu, "U.K. Eyes 5th Gen.," *Datamation*, July 1983, pp. 67-68, 72.

ties.[161] The Science and Engineering Research Council (SERC), roughly the equivalent of the American NSF, will develop plans and disburse most of these funds.[162]

The United Kingdom has a strong tradition in AI, particularly in the University of Edinburgh and Imperial College, London. But in the years immediately preceding the Alvey initiative, efforts had been scaled back because of an unfavorable government report on the prospects for artificial intelligence, and many researchers emigrated to the United States to continue work in American universities.[163] Alvey Programme funding may reverse these trends, but there is disagreement among experts concerning the ability of European universities, in general, to respond flexibly to the growth of a field such as AI, where interdisciplinary research across traditional academic departmental lines is so crucial.

[161] *A Program For Advanced Information Technology*, op. cit., pp. 47 and 49.
[162] Ibid., p. 48.

[163] The OTA workshop, Oct. 31, 1983.