AD-A021 274

REPORT OF THE ARPA STUDY GROUP ON ADVANCED MEMORY CONCEPTS

E. R. Berlekamp, et al

Science Applications, Incorporated

Prepared for:

Rome Air Development Center

February 1976

068094

ADA021274

RADC-TR-76-28
Final Technical Report
February 1976

REPORT OF THE ARPA STUDY GROUP ON ADVANCED MEMORY CONCEPTS

Science Applications, Incorporated

Sponsored by
Defense Advanced Research Projects Agency
ARPA Order No. 2886

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public including foreign nations.

This report has been reviewed and is approved for publication.

APPROVED: Edmund J. Kennedy

EDMUND J. KENNEDY
Project Engineer

Do not return this copy. Retain or destroy.

REPORT OF THE ARPA STUDY GROUP ON ADVANCED MEMORY CONCEPTS

Dr. E. R. Berlekamp
Mr. R. L. Garwin
Mr. D. E. Knuth
Mr. J. Lederberg
Mr. R. A. Leibler
Mr. C. Levinthal
Mr. J. Rajchman
Mr. R. Madden
Mr. V. A. Vyssotsky

Contractor: Science Applications, Incorporated
Contract Number: F30602-75-C-0098
Effective Date of Contract: 6 February 1975
Contract Expiration Date: February 1976
Amount of Contract: $400,000.00
Program Code Number: 5P10
Period of work covered: 1 Oct 74 - 30 Apr 75

Principal Investigator: Lee S. Baumann
Phone: 703 527-7517

Project Engineer: Edmund J. Kennedy
Phone: 315 330-3857

Approved for public release;
distribution unlimited.

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by Edmund J. Kennedy, RADC (ISIM), Griffiss AFB NY 13441.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
RADC-TR-76-28

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
REPORT OF THE ARPA STUDY GROUP ON ADVANCED MEMORY CONCEPTS

5. TYPE OF REPORT & PERIOD COVERED
Final Technical Report
1 Oct 74 - 30 Apr 75

6. PERFORMING ORG. REPORT NUMBER
SAI-75-631-WA

7. AUTHOR(s)
Dr. E. R. Berlekamp
Mr. R. L. Garwin
Mr. D. E. Knuth
Mr. J. Lederberg
Mr. R. A. Leibler
Mr. C. Levinthal (see reverse)

8. CONTRACT OR GRANT NUMBER(s)
F30602-75-C-0098

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Science Applications, Incorporation
1911 North Ft. Myer Drive, Suite 1200
Arlington VA 22209

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
62711E
B8860101

11. CONTROLLING OFFICE NAME AND ADDRESS
Defense Advanced Research Projects Agency
1400 Wilson Blvd
Arlington VA 22209

12. REPORT DATE
February 1976

13. NUMBER OF PAGES
54

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
Rome Air Development Center (ISIM)
Griffiss AFB, NY 13441

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
N/A

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)
Same

18. SUPPLEMENTARY NOTES
RADC Project Engineer:
Edmund J. Kennedy (ISIM)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Advanced Computer Memories
Innovative Technology
Architecture, Software and Theory
Neuro-Sciences

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Reports on a study conducted by experts in the field of advanced memory concepts convened by ARPA to review state-of-the-art and possible future technology in area of computer memories.

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

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

7. Mr. J. Rajchman, Mr. R. Madden, Mr. V. A. Vyssotsky

PREAMBLE

The computer industry has remained notably strong despite the current problems which plague the world. It is one industry where U.S. exports exceed imports, where expansion continues, and where, in many fields, the availability of jobs still exceeds the availability of highly skilled specialists. It is an area in which the graduate students involved in any university-based long-range research programs will be able to contribute to the national economic needs in the near future.

Many agencies of the federal government, including the Bureau of the Census, the Office of Management and Budget, and the Department of Health, Education and Welfare will rely heavily on computers for the foreseeable future. The Department of Defense requires computers for applications ranging from parts inventory to strategic weapons control. The National Security Agency, the nuclear energy research efforts at Livermore and Los Alamos, and the Worldwide Military Command and Control Systems or their renamed organizational descendants will all make heavy demands on the most advanced computer systems available in the year 2000.

For all these reasons, the federal government has an interest in ensuring adequate long-range research programs in computer science and technology. We were delighted to accept the invitation, which we received from Malcolm Currie at our initial meeting at the ARPA headquarters in Washington, D.C. on 21 October 1974, to outline the framework of a new program for the ARPA in the area of Advanced Computer Memory Concepts. Since then, we have met formally on three subsequent occasions for a total of four days. We have heard presentations from the eleven knowledgeable speakers listed in Appendix A. We have also discussed this matter with numerous additional sources in government, in industry, and in the universities.

One danger of a government-sponsored research program in the computer area is that it may become so low-risk, so short-term, or so oriented toward the development of some particular device that it might duplicate commercially-financed research and development activities already underway in industry. After surveying the industrial efforts now underway, we have formulated preliminary plans for a long-range basic research program which may yield significant long-term benefits to DoD and the nation. These plans stress the longer range requirements for innovative technology that are least likely to be mediated by already obvious lines of development.

Following a brief overview of short-term industrial trends, this report highlights four important research areas which should be prominently included in the ARPA program in Advanced Memory Concepts. Listed in the order of the immediacy of their applicability, these are: 1) innovative technology, 2) architecture, software and theory, 3) materials sciences, including solid state properties of organics, and 4) neurosciences. The final section of this report contains recommendations on how the Advanced Memory Concepts program should be managed.

A list of the technologies which will dominate computer systems for the next several years is shown in Figure 1. Because the faster memory technologies cost substantially more per bit of information stored than the slower ones, memories are typically organized into a hierarchy of the sort shown in Figure 2. The so-called "primary" or "main" memory is now usually constructed out of cores or large-scale integrated metal oxide semiconductros (LSI MOS). This is often back up by a larger, slower, "secondary" disc memory, which is substantially less expensive per bit. The secondary memory, in turn, is backed up by a much larger and slower tertiary memory which often consists of magnetic tapes. In the other direction, many computers also have an additional level of "cache" memory which is smaller and faster than the primary memory. Present-day cache memories are constructed of bipolar

	Basic Material	# Micro-miniature masking steps	Analog contents at physical address? (in any presently popular version)	Operating temperature	Time to decay after power loss	Estimated current National Research Budget/yr. (Times 0.5 to 2.0)
Josephson-effect	Pb Alloy	~ 5	No	$\sim 4^{\circ}K$	Hours	4M
Bipolar transistor integrated circuit	Pure Si	≥ 6	No	Room	2 ms	
Metal Oxide Semi-conductor, Field-Effect Transistor Large Scale Inte-grated circuit	Pure Si	≥ 4	See CCD	Room	2 ms	300M
Charge-Coupled Devices	Pure Si	4	Yes	Room	2 ms	5M
Magnetic bubbles	Fe garnet	1?	No	Room	∞?	10M
Magnetic core	Fe ferrite	0	No	Room	∞	20M
Magnetic discs	Fe film	0	No	Room	∞	20M
Magnetic tape	Fe film	0	--	Room	∞	20M
Punched cards or paper tape	Paper	0	No	Room	∞	?

Figure 1. Some Memory Technologies Being Developed in 1974, listed in approximate order of decreasing speed

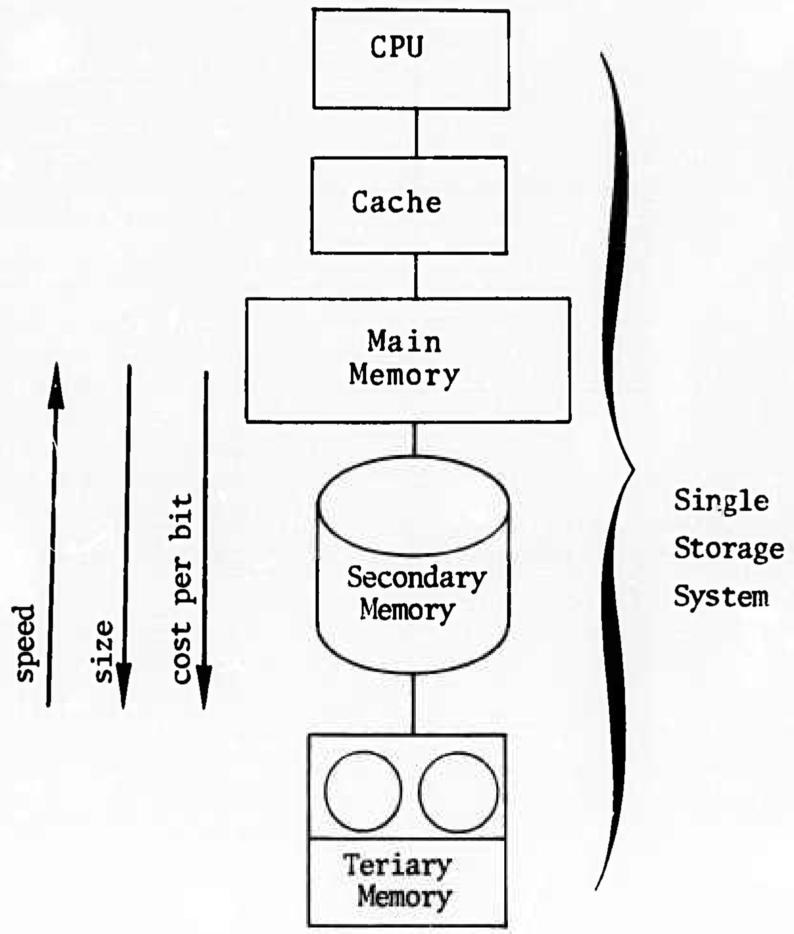

Figure 2. adopted from "Role of Optical Memories in Computer Storage", by R. L. Mattson, Applied Optics, Vol. 13, No. 4, p. 756, (April 1974)

integrated circuits. A typical 1974 installation might have a cache of about 10^3 words, a main memory of about 10^6 words, and on-line disc capacity of about 10^8 words, where each word has between 32 and 64 bits.

The large volume of production of semiconductors and magnetic recording devices now provides a significant level of funding for continued research in these two main branches of established computer technology. These technologies will continue to evolve and improve, and they will continue to dominate the cache, primary and secondary levels for the short-term future. The following technologies are now at a sufficiently advanced stage of research to enable us to foresee the possibility of a significant market impact in the not too distant future:

Semiconductor integrated circuits will dominate the fast end of the memory hierarchy for the foreseeable future. Semiconductors now play a fundamental role in all of electronics. For many years nearly all computer processors and buffer registers have been constructed from semiconductors, and in recent years semiconductors have been taking an increasing share of the main memory market away from magnetic cores. Many interesting variations of the semiconductor technology, including charge-coupled devices (CCDs), area are already here, and other technological variations including epitaxial growth on sapphire, silicon ribbon growth, ion implantation, and electron-beam formed patterns are being vigorously developed. Costs will continue to decrease and performance will continue to increase.

Magnetic recording now dominates mass storage (the secondary and tertiary levels of the hierarchy). It provides a form of information storage which can be permanently maintained with no power requirement. In the last three decades magnetic recording performance has improved by many orders of magnitude, thanks to great strides in the design of heads and bearings and to improved recording materials. Magnetic recording technology can rightly claim to be the technological basis of the computer age. It was already supreme before the advent of the transistor, and at most installations it still commands the lion's share of both the floor space and the hardware investment dollars. Despite its success and dominance, magnetic recording technology provides relatively bulky and expensive devices requiring considerable human attention. Furthermore, its reliance on mechanical motion has inherent limitations. It is far more efficient in accessing blocks of data than randomly located bits. While manageable trade-offs between economy and latency are possible with high bit rate transfers and quasi-random accessibility in discs, the architectural and operating situations for very large memories remain far from ideal.

Magnetic bubbles may soon move into the gap in the hierarchy between semiconductors and discs, possibly even replacing one or the other at some installations. Large efforts at BTL and IBM have obtained high bit densities, non-volatility, and some degree of bubble-to-bubble logic. There are hopes that a manufacturing process requiring only one micro-miniature mask may avoid the mask-alignment problems which will arise in the manufacture of semiconductor integrated circuits.

Josephson junction devices may find some applications at the very fast end of the hierarchy. These superconducting devices may be used for processors, cache memories, and even main memories. Despite the complication of needing a cryostat, the very fast switching times of these devices (tens of picoseconds, as opposed to nanoseconds for bipolar transistors commercially available today) make these devices potentially attractive. The fabrication processes are similar to those of bipolar transistors. The uniformity of the oxide layer is more crucial, but ingenious ways to insure it automatically have already been developed. Further efforts to develop this technology are underway at IBM, and a small, very fast prototype computer based on this technology may become operational within a few years.

Video discs will provide very inexpensive mass-produced read-only mass memory within a few years. These rotating devices, now being vigorously developed by RCA, Zenith, Philips, and others, will provide recorded television on easily replicable discs, at a mass-production factory cost of about 20¢/disc. The pressed disc has a spiral groove at the bottom of which are submicron-sized depressions. These depressions are read-out either through a capacity pick-up in a needle-in-the-groove (RCA), or through a light beam (Philips, Zenith, and CSF in France). There is an equivalent 10^{10} bits per disc. Access is serial but some pseudo-random access by groove selection is also possible.

It is not yet clear what utility these products may have for the computer field.

1. Innovative Technologies for Computer Memories

Despite the differences emphasized by their enthusiasts, all of the technologies above magnetic cores in Figure 1 (page 2) have a great deal in common. All are planar. All are based on micro-miniaturization. These technologies will continue to evolve and improve, and they will continue to dominate the faster levels of the memory hierarchy for the foreseeable future.

The large volume of production of magnetic recording devices also provides a significant level of funding for continued research in the evolutionary development of this technology. This technology will also continue to evolve and improve, although not as rapidly as semiconductors.

We believe that innovative technologies have better prospects of competing with conventional mass memory devices than with conventional fast memory devices. At the very large end of the memory spectrum, several devices based on technologies other than magnetic recording are already operational. The list of such devices includes the IBM photostore at Livermore and Precision Instruments' Unicon 609.

As we recommend in the next section, the systems aspects of nonerasable (i.e., write-once-only) memories merit further study. It is not yet clear how fast the value of such a memory declines with increasing minimum delay between the writing and the first reading, nor how an overall system might be configured to make best use of a very large, very cheap, fast nonerasable memory. But we foresee the possibility that many future tertiary memories might be nonerasable. Even today, magnetic tapes are de facto used this way; writing the entire reel is a much more common operation than scanning through it and overwriting selected parts. The overall system economics do not appear to be significantly affected by the fact that it is possible to rewrite an entire reel rather than throw it away and buy a new one.

New devices and techniques will depend on new materials and new ways of interacting with or modifying existing mediums. For example, visible light can be focused easily to a spot size of about 1μ in diameter. By using ultraviolet light and sophisticated optics one can obtain spot sizes that are 3 or 4 times smaller. Furthermore, photographic emulsions exist which have a small enough grain size so that pictures with better than 1μ resolution can be recorded. Thus, in principle, there are both materials and physical devices for reading and writing data with light with a planar packing density of about 10^{12} bits/m^2. However, photographic film requires a relatively slow and complicated developing process and other methods of modifying material with light are either slower, more complex, or have a lower resolution and therefore a lower planar packing density.

Furthermore, the depth of focus for light spots can easily be less than 5μ . Therefore, devices which operate in three dimensions, even for a very thin slab, would result in a considerable gain in the effective packing density. New devices will be required to move focused light spots in three dimensions, especially since this must be done rapidly and over a large area. But even more important is the fact that new materials will be needed to record the information contained in this moving light spot. This is particularly true if any use is to be made of the three-dimensional possibilities of such a system. Clearly, one way of increasing the signal to noise ratio in such an optical memory is by spreading the information for each bit using holographic recording. Although various aspects of this approach are currently being pursued, many innovations both in devices and materials are still needed before this type of technology is ready for a major developmental effort.

Even more spectacular packing densities could in principle be used if the resolution currently available in commercial electron-optics instruments could be employed in a memory system. Scanning transmission electron microscopes are now functioning in research laboratories with a resolution and spot size of a few angstrom units. If there were any way of making use of this high resolution, it would be possible to record a bit of information in a square 100 angstroms or so on a side. Even considering the need for paging marks and other alignment indicators, one could easily imagine memories with a packing density of the order of 10^{15} bits/m^2. The basic problem is that there are currently no materials which can be modified in a detectable way if an electron beam strikes them with these very small spots. However, there are several possibilities using molecular crystals and various organic materials which could be explored. (c.f. Appendix B). Here also, new devices, as well as new materials, would be needed in order to develop a useful system in which an electron beam can be focused and deflected at high speeds and which could be coupled with a less precise system for mechanically moving a target platter so that a sufficiently large region could be accessed by the beam.

We mentioned the particular examples of light and electron beam-accessed memories only as examples of possible new developments which could lead to major improvements in mass memory technology. Some other approaches which also merit further study are discussed in Appendix C. General notions which could guide would-be memory inventors are given in Appendix D.

The possibilities for additional ideas on how to build memories are numerous, and we think that clever inventors should be encouraged. We believe that the initial, conceptual stages of the inventive process can be most effectively encouraged by a program of many small grants.

Once an initial idea has jelled, it is often possible to identify some of the technical obstacles which must be surmounted before the device can be built. Sometimes these obstacles look sufficiently small that an enlarged development program appears justified, but more often the obstacles are serious research problems in other disciplines. For example, a new memory invention might be only marginally feasible using commonly available materials, but it might offer great advantages if one could find new materials with slightly better properties. Or the new invention might offer a very effective way for building memories organized in some highly nonconventional way of unproven value. In either such case, the idea should be kept alive pending further progress on the related problems. However, when the new problems are at the research frontiers in specialized disciplines other than the inventor's, there is no need to support a massive development effort at the inventor's laboratory. The next contributions are most likely to be made by different individuals and groups. If and when the needed progress in other disciplines occurs, a major development effort may again become warranted.

To guide a new invention all the way to production may require several shifts of role and emphasis in the effort needed from any particular development group. The management of such an effort requires broad technical sophistication and the flexibility to move quickly. This is precisely the forte of some major industrial laboratories, and we do not think that ARPA can or should sponsor competitive programs leading toward the development of a production-line for some type of memory device. ARPA should support the initial stages of the inventive process, but it must be very wary of the dangers of premature development efforts. Such efforts not only consume vast sums of money; they may actually retard progress by distracting talented people from the more fundamental research problems.

2. Architecture, Software, and Theory

Memory organization has had an interesting history. The conceptual machines studied by Turing and others in the 1930's used tapes with no addressing facilities other than the ability to step forward and backward. When von Neumann started writing machine-language programs for his paper machines, he initially assumed a sequentially accessed (delay-line) memory hardware. He assumed that memory addressing problems would be handled by special subroutines that would idly count time until the proper data arrived on the delay lines. Rajchman persuaded him that random access addressing could be handled more readily in the hardware than in the software, and the art of programming for machines as if they had random-access memories began to develop rapidly even in the days when most of the machines had drum memories. Since then, the rise and reign of magnetic core technology has obviated software efforts outside the random-access mainstream. With much of the addressing problems solved by the core wiring, the economic advantages of delay-line memories disappeared. With memory cores wired so as to be unable to perform additional logic, the distinction between memory and processing was clear.

Most thinking about computer software and applications today is preoccupied with the conventional random access memories to which programmers have become accustomed. Clever programming tricks, such as inverted indexing, including hash coding, help surmount the problems of searching for the contents of these conventional memories. These programming techniques, which have evolved gradually over a period of many years, are now part of our cultural heritage. Hardware architects evaluate new machines in terms of the speed with which the new machine runs a stereotyped set of test programs. Only rarely are the test programs readjusted to take greater advantage of the new machines, and only rarely are the new machines evaluated by their capability of attacking new problems.

While this conventional modus operandi leads to continuing evolutionary improvement in the computing art, it is unlikely to yield any radical organizational changes. Researchers who have access only to conventional hardware are not likely to pursue truly radical ideas. Industry will not put much effort into developing software for peculiar machines that do not exist, nor is it likely to put effort into building them if the programming techniques to capitalize on their advantages are not yet known.

With the approaching demise of magnetic cores, the technological reasons for conventional memory organization are no longer so convincing. In many current machines the memory, as well as the processing, is done with semiconductors. Hence it is no longer possible to incorporate much of the addressing function into the core wiring. Between 30% and 50% of the active area on current memory chips is used for addressing rather than memory in the strict sense. Microcomputers and minicomputers now have processors in precisely the same MOS technology as the memory, and there is nothing to prevent memory and processing from being implemented on the same chip. <u>Schemes for organizing and using fast memories are now limited by our lack of imagination and insight more than by the semiconductor technology, which is extremely versatile</u>.

A million word semiconductor memory is a single circuit, consisting of several million gates and flip-flops having the property that, at most, one millionth of its flip-flops can change at any given clock time. This seemingly inefficient use of logic yields a very large ratio of stored-bits to pins on the chip, a ratio that has a great influence on the cost of a conventionally-organized memory. For the same number of components in the same area, one could accomplish thousands or even a million times as much signal processing, with only a moderate increase in the ratio of pins to cells. If these components were organized in a sufficiently regular manner for applications requiring sufficient masses of computation, it might be possible to achieve this big jump in computing capability with little or no increase in fabrication cost.

A number of ideas on how to achieve some of this potential improvement have already been proposed. All involve some form of parallelism. In its most extreme form, this might involve some limited processing at every location in the memory, as in an associative memory. Both human memories and those of animals do simultaneous "associative" searching to retrieve partially-specified information. The single-minded location-oriented memory of today's computer is incapable of achieving such performance, and this is a major bottleneck in large data-base applications. But today we are so accustomed to conventional memories that we can't really imagine how associative memories should best be exploited. It will be wise to have a reasonably large number of people thinking seriously about how to exploit large associative memories if they were inexpensive and readily available. Experience with nature suggests that such memories are not infeasible and that many present limitations of computers could be transcended. One of the goals will be to develop new applications.

Much can be learned by devising paper machines with very large memories, and programming them by simulation on existing machines. Such studies need to be undertaken before the memories are actually built, since so many design decisions need to be made regarding the operation of the machine; merely saying "associative memory" by no means uniquely specifies the concept, and indeed the large literature on this subject (approximately 1000 citations in Computing Reviews, October 1971, and a more recent review in Computer Journal, November 1974) suggests that the term "associative memory" already conveys an improperly fixed connotation to too many people. There are many conceivable ways to organize a distributed-control memory; for example, we might look for the data that most clearly matches a specified pattern, or we might have a drastically different sort of memory in which the data forms itself into clusters of related items and the program examines these clusters and specifies new rules of combining them. Programming experience with paper machines will provide important feedback to the design of "mega-parallel memories", and will suggest new characteristics that large memories should have. Important new algorithms will probably be discovered.

As a less extreme version of parallelism, one might partition the memory into a number of submemories and assign a separate processor to each. Rudimentary versions of this scheme include not only semiconductor memories, but disc controllers having limited processing capabilities dedicated to each head. Slightly more powerful parallel machines now in existence include the Illiac IV and the CDC Star, in which tens of processors march along together under the command of a common control sergeant. More flexible schemes involving thousands of processors with independent programs have been considered, but the difficulties of resolving potential conflicts and locking problems have frightened all but the most ardent enthusiasts. But there is no a priori reason for such problems to be insurmountable, and the potential advantages of intelligently-organized parallelism are so great that further studies are urgently needed.

The programmer's conceptual notion of a single processor is already out of touch with the details of reality of most current machines, in which various system functions are parcelled out to peripherals and minicomputers of various kinds. Some specialized "blister" units with specialized processing skills have already appeared. As more and more standard arithmetic subroutines become implemented in special purpose hardware rather than software, future computers may look more like a tightly linked organization of minis than like the monolithic structure shown in Figure 2 (page 4). The total amount of memory dedicated to the various subunits of this organization might conceivably be much greater than that aggregated in any single "primary" organizational unit. If there is still any single chief executive unit in such a machine, it will probably spend most of its time issuing high-level commands (e.g., invert this matrix; solve this partial differential equation; find all references to these kinds of accounts; ...) rather than the current low-level commands (e.g., find the quotient of these two numbers; store this number at that location; compare this number with that; ...). Like the human executive, it will probably issue many additional orders to many other units and make many other decisions before reading any progress reports about the execution of the first command.

While studying how to utilize nonconventionally organized memories, appropriate symbolic languages to specify the algorithms will be invented. These may well have a payoff in other scientific fields too, since it will lead to ways of specifying and analyzing parallel processes for which neither mathematical models nor conventional computer models are appropriate.

Besides studies of associative memories and other forms of parallelism, there is a continuing need for analysis of more conventional memory structures. Some possibilities for important studies of this kind are:

1. Analysis of the performance of typical classes of computer programs when running on machines with hierarchical memories of given types and speeds.
2. A customer study of presently understood needs for large memories.
3. Development of new applications for large memories.
4. Study of how to use a large nonerasable memory.

Several theoretical studies have already identified problems for which parallelism is inherently limited, while in other cases parallelism is known to be highly advantageous. It can be expected that important results will be learned by further studies of this kind, as computational complexity is analyzed under new ground rules corresponding to new memory organization techniques. Proposals for such continued studies should be solicited since they can be expected to provide considerable insight into finding the best mixture of different kinds of computer memories.

To conclude this section, we again emphasize that ARPA should also be receptive to truly imaginative unsolicited proposals for theoretical studies in other areas not among those listed above. For example, someone might propose a truly novel way to design reliable memories containing unreliable elements. We have deliberately avoided giving this topic greater emphasis because there is already a 15-year gap between the theory of error-correcting codes and current practice. Cheap and effective ways of correcting quite high error rates at the system level are already known, but little used! This area now needs education and advertising more than additional research. On the other hand, methods of designing and manufacturing large semiconductor chips (perhaps 3-inch square or even larger) which function perfectly despite numerous random local defects might prove to be of great interest and commercial value.

As another example, someone might propose studies of fundamental limits. Fundamental limits of algorithms and related questions of computational complexity are more concerned with processing than with memory, but these topics are a very important branch of current computer science and some good small proposals in this area might be considered within this program. Expositions of specific technologies which spell out their eventual practical limits and the obstacles to reaching them should also be encouraged. But the only physical limits on memory known to us which are fundamental in a truly technology independent way (e.g., those based on Heisenberg's uncertainty principle) are so far beyond our wildest current dreams that we think it premature for ARPA to solicit proposals to study "fundamental" physical limits of memory.

3. and 4. Basic Sciences

The discovery of a new physical phenomenon, such as the transistor in 1948 or the Josephson effect in 1962, often provides the initial impetus for a major new thrust in research and development. It is evident that any truly long-range program must not only seek inventive and novel ways of utilizing known properties of materials, it must give ever more encouragement to the discovery of additional phenomena, and to the recognition of their potential applications to computer technology.

Among the many sciences which might conceivably contribute to memory technology, those which have the most evident prospects of yielding economic returns within the next decade are already well-financed. The prospects of ever getting economically valuable insights about memory from the remaining sciences are somewhat speculative, but we have identified two in which we believe the very long-term prospects to be sufficient to warrant ARPA support. These are as follows:

Solid-state physics of new materials, including organics.

While organic chemists have long been able to synthesize many new compounds, it is only recently that the electrical and magnetic properties of organic compounds have attracted much interest. Several organic crystals are now known which have large variations in conductivity associated with phase transitions occurring at temperatures much higher than any known super-conducting metal, and this area appears intellectually ripe for further research. While we do not wish to restrict this program to organics, we think it is important to mention organics in order to increase the likelihood that organic chemists will perceive "new materials" and "memory" to be relevant to their favorite compounds as well as to more conventional materials such as silicon and metallics.

While there is now a large amount of research going on in organic chemistry, relatively little of it is motivated by interest in properties potentially relevant to memory technology. We therefore believe that research funds spent in this area have a potentially large leverage for attracting increased interest from other organic chemists.

We recommend an increased ARPA concern with the properties of materials which are likely to be useful to memory technology. Whether this should be done through the materials science office, or through the information processing techniques office, or through some combined effort is an administrative decision appropriate for the Director of ARPA.

Neuro-sciences

While it is obvious that human memories are organized much differently from computer memories, no one knows enough about human memories to build their more advantageous features into computer memories. Just as the study of birds once challenged those whose conception of man-made flying machines was limited to balloons, so now the study of human memories challenges us whose conception of memory is limited to present-day computers. Even if the structure of human memory is understood a half-century from now, people might then still choose not to design all of the same mechanisms into computers. It might be like trying to build airplanes with flapping wings, or even with feathers. But there might also be other features of the structure of human memory, analogous to the aerodynamics of how birds glide, which might yield simple new principles that will provide important new insights valuable in the design of computer systems. For this reason we believe that the basic sciences which study biological memories merit support.

We must stress that we expect the benefits under this program to be very long-term. Biological memories are currently understood too little to justify any immediate effort to build or simulate neutral nets for technological rather than scientific objectives. But tangible progress is now being made on questions concerning the pre-processing in the visual and acoustic input channels, the connectivity of these channels to the brain, and other fundamental structural questions. The neutral nets which have so far been analyzed in any detail either physiologically or anatomically have been very small, usually involving a few interacting cells. Although it is obviously impossible to attempt the experimental analysis of any significant fraction of the billions of cells in a complex brain, there are limited groups of cells which seem to carry out well-defined processing tasks. With the application of modern techniques it may be possible to analyze the anatomical connections and precise physiological functions of nerve networks having as many as several hundred interacting cells. This would yield increased understanding of the nature of some of the elementary sensory preprocessing functions.

The general support of neurobiological research is primarily the responsibility of NIH and NSF. Hence, the support of basic work in these sciences should be a relatively minor part of the total ARPA memory program. Nevertheless, this support should function as a means of ensuring a continuing flow of communication about non-ARPA funded work in these sciences to other persons working on more immediately applicable aspects of the ARPA memory concepts program. However, we are divided on the best ways to encourage this communication. Some think that these sciences should be supported as elements of interdisciplinary projects which embrace engineers and computer scientists at the same institution as the neuro scientists; others think that carefully planned conferences and publications, as well as consulting and cooperative projects, might achieve the same objective at less expense.

MANAGEMENT OF THE ARPA ADVANCED MEMORY PROGRAM

At our first meeting, we devoted part of our mental efforts to an attempt to invent large new computer memories. It soon became clear that this attempt to do years of research in real time exceeded our combined mental capacities. In three months we could at most hope to plan a memory program; there was clearly no way in which we could carry the program out. Nevertheless, we have subsequently been tempted to subtler versions of this same fundamental error.

After abandoning attempts to do the research ourselves, there was then a temptation to identify a few new major technologies which ARPA might emphasize. Several nonconventional technologies already exist. Bubbles is now being developed at a significant rate, and we do not think progress in bubbles would be substantially enhanced by additional funding from ARPA at this time. The IBM Josephson junction program is progressing with a budget comparable to that of a good-sized ARPA project; and it is also adequately funded by other sources. We now realize that any technological approach which has progressed sufficiently far to warrant a development effort of this size is very likely to attract adequate industrial financing. There may be a few relatively small high-risk development projects which have sufficient technical merit to justify consideration within this program, but the major emphasis of the program must be on the generation of new memory inventions and organizations rather than on the evolution of those which already exist.

Even after we abandoned the goal of specifying the technologies to be developed, there was still a temptation to specify the desired parameters of the inventions we wish to encourage. It is trivial to wish for sizes and speeds orders of magnitudes better than today, but less easy to specify the tradeoff between speed and size, or the extent to which noisy, nonerasable, or delay-line memories might be acceptable, or the extent to which incentives should be offered for memories having bonus properties (e.g., associative memories). It then became clear that problems of memory organization and use are themselves major topics for research.

In the previous sections, we did identify several types of areas in which proposals should be solicited. The committee is unanimous in its recommendation that a significant portion of the funds spent on the Advanced Memory Concepts Program should go to small grants (i.e., $\leq$ \$200,000/year). We also recommend that these grants be incrementally funded with a three-year phaseout period to enable the investigators to devote their energies to scientific efforts rather than to survival politics. We also recommend that the program in Advanced Memory Concepts should be widely advertised in an effort to generate a large number of unsolicited proposals. In order to advise ARPA in selecting appropriate referees for these proposals, we recommend the prompt formation of advisory panels in the following disciplines:

1. Inventive memory technology (5-8 members)
2. Architecture, software, and theory (5-6 members)
3. Electromagnetic properties of new materials (4-5 members)
4. Neuro-anatomy, biology, psychology (4 members).

We also recommend the formation of a coordinating council consisting of people such as the chairmen of each of the above four panels, representatives of appropriately selected DoD computer consumers, and possibly one or two additional members.

In order to minimize conflicts of interest, we think it is important that a majority of the panel and council members be persons who will <u>not</u> receive ARPA grants.

In view of the great interest in computers at numerous government laboratories, at numerous industrial laboratories (some of which are not interested in ARPA contracts), and at numerous academic institutions (some of whose members already have adequate funding from non-ARPA sources), we do not foresee any serious difficulties <u>finding</u> first-rate people whose institutional biases average out to near zero. Recruiting them to serve will not be trivial; they must be convinced that ARPA has a durable commitment to the program, that the other panelists are of a calibre they admire, that the advice of the panels will be taken seriously, and that reimbursement and other staff work will be efficient and prompt.

We urge ARPA to sponsor a week-long conference on Advanced Memory Concepts in September, 1975. We think the organization of this conference should be one of the first goals of the coordinating council. This council should then evaluate the conference and devise further steps to ensure that the results of the research in Advanced Memory Concepts are communicated to all interested parties both inside and outside of the ARPA community. This might include subsequent conferences, as well as ARPA support of relevant conferences co-sponsored by appropriate technical societies.

We also recommend that the council assist ARPA in seeking out appropriate authors and initiating a series of special monographs. A book with a theme such as "Fundamentals of Computer Memory Concepts" would be an extremely valuable aid in focusing the attention of experts in electrochemistry, organic materials, neuro-sciences or any other spcialized field of science on the problems of memory. The current literature tends to be so specialized, fragmented and technical that it is difficult for an outsider to know where to begin, while most insiders are preoccupied with research in some narrow subspecialty. Yet the problem of memory has a broad generality and presents a fascinating intellectual challenge for imaginative minds.

In addition to the small grants to support novel and inventive work on Advanced Memory Concepts, we can identify a few types of projects which might merit a major grant (i.e., budget level $\geq$ \$500,000/year). Like the smaller grants, these would be incrementally funded with a phaseout period of three to five years.

1. <u>An interdisciplinary institute encompassing all of the areas we suggest emphasizing in this program</u>.

The main argument in favor of such an institute is the need for significant interaction among the people working in the four different areas which we think should be supported under the ARPA memory program. New possibilities of computer organization, construction, internal housekeeping, and usage may call for a different emphasis im memory characteristics. Conversely, the conception of a potentially powerful memory that does not fit conventional usage may call for ingenious computer organization. The interplay between physically-oriented and computer-systems-oriented innovations is clearly very important in the construction of any usable system.

2. <u>A laboratory to develop a particular memory technology</u>.

This is another obvious example. As mentioned earlier, this type of project would either duplicate or compete with the efforts of strong industrial research laboratories. We must caution ARPA to be wary of such projects. To the extent that they are to be supported at all, they should probably be carried out at an existing national not-for-profit laboratory or at an industrial laboratory which is willing to supply matching funds. The goals of such a project are most likely to be more immediate than the goals of this program, and it may be more appropriately sponsored by some other branch of DoD.

3. <u>An institute to study novel machine architecture and use</u>.

As suggested earlier in this report, much of the effort would be devoted to paper studies and simulation of nonconventional machines. Some software and operating system questions require team effort, which might more readily forthcoming from an institute than from a collection of geographically separated individuals. The institute might also be more able to provide adequate facilities for simulating unconventional machines. These simulations might entail some special-purpose hardware as well as software, but the major goal would definitely <u>not</u> be the construction of any particular nonconventional machine.

On the other hand, some committee members feel that a firm expression of ARPA's intent to proceed with contracts for the construction of prototype nonconventional machines at some future time, on the basis of careful prior study of diverse proposals and simulations on conventional machines, would be a powerful incentive for more imaginative research leading to significant innovations.

While we are unanimous in our recommendation that a significant portion of the available funds by used to finance small grants, we have differing opinions about which of the three recommended major grants should receive the most emphasis. Much depends on the management of the program, both at ARPA and at institutes or laboratories receiving the major grants. The difficulties of finding appropriate principal investigators for large grants should not be underestimated. It may easily happen that a superstar academic researcher may not be a very good manager. There will also be additional difficulties in obtaining significant effort from people in disciplines whose value structures rate other problems as substantially more important than the problem of finding new concepts for computer memories.

At our first meeting, Dr. Lukasik asked this group to consider the relative merits of implementing the Advanced Memory Program at universities, at industrial laboratories, or via new institutional entities which we were invited to devise. Except for the existing not-for-profits and the obvious (and perhaps insufficiently used) possibility of enlisting otherwise unavailable academic talent through off-campus consulting arrangements, we have no additional types of institutional entities to suggest. These remarks are not intended to prevent the establishment of new institutions of the existing types. A small institute or laboratory might function better if it is totally autonomous than if it is a wholly-owned subsidiary of a large, established bureaucracy. The larger and older institutions not only have higher overhead rates: the "services" which this overhead supports are sometimes more of a hindrance than a help. These organizational questions merit close scrutiny for every grant.

The great and obvious advantage of research programs based at universities is that they will attract graduate students. This is an indispensable mechanism for exerting a significant influence over the field for many years to come. For this reason, we are unanimously agreed that a major portion of the funds available to the ARPA program in Advanced Memory Concepts should be spent at universities.

Industrial laboratories and national not-for-profit laboratories often have more experienced management and greater flexibility than universities. For these reasons, and others detailed in Appendix E, we are in unanimous agreement that the dominant portion of the computer research and development funds of the Defense Department (of which ARPA is a relatively small part) should be spent in industry. There is no question that projects which emphasize development more than research are unsuitable for universities.

Despite our broad agreement on these general principles, we have differing opinions over the precise balance we would like to see between universities and industrial laboratories within the Advanced Memory Concepts Program. Some feel that all proposals should be considered solely on their merits; others feel that university-based work should receive special recognition for institutional side-effects. In the opinion of the chairman, this disagreement may not be too important. Whatever grading system is used, I think it likely that much of the work on innovative technologies will be based in industrial laboratories, that the bulk of the work in the remaining areas will be based in universities, and that the not-for-the profit laboratories should get some share of each.

APPENDIX A

SPEAKERS AT MEETINGS OF ADVANCED MEMORY CONCEPTS GROUP

November 7, 1974 - San Francisco, California

- Robert N. Noyce, President of Intel Corporation, Santa Clara, California.
- John McCarthy, Director of Artificial Intelligence Laboratory, Stanford University.

November 18, 1974 - Newark, New Jersey

- Wilhelm Anacker, IBM Watson Research Laboratory, Yorktown Heights, New York.
- Sidney Fernbach, Director of Computing, Lawrence Livermore Laboratory.
- Richard Karp, Associate Chairman for Computer Science, University of California-Berkeley.
- Joel Moses, Deputy Director for Project MAC, MIT.
- Allen Newell, University Professor, Carnegie-Mellon University.
- Michael L. Dertouzos, Director of Project MAC, MIT.
- H. Chang, IBM Watson Research Laboratory, Yorktown Heights, New York.

December 3, 1974 - San Francisco, California

- Forest Baskett, Electrical Engineering and Computer Science Department, Stanford University.
- Dr. Mueller-Westerhoff, IBM Research, San Jose, California

APPENDIX B

A HIGH DENSITY MOLECULAR MEMORY

by

C. Levinthal

During the last fifteen years there have been three developments in the electron microscopy of biological macro-molecules which suggest a new method of constructing a computer memory which can, in principle, store a very high density of information.

The first development is that several protein molecules, small viruses and other macro-molecular aggregates have been shown to form very regular three-dimensional crystals whose surface periodicity can be visualized in the electron microscope. The number of particles regularly spaced in such a surface lattice is of the order of 10^{11} to $10^{12}/cm^2$. If such particles could be modified in a detectable way one would have a memory system with 10^{15} bits stored in a square meter. (THE MOLECULAR OTULINE OF HUMAN GI IMMUNOGLOBULIN FROM AN EM STUDY OF CRYSTALS by L. W. Labaw and D. R. Davies, Ultrastructure Research 40, 349-365 (1972) by Academic Press, Inc.)

The second development was the observation that any material held together with co-valent chemical bounds, and particularly organic macro-molecules, are disrupted when irradiated by an electron beam. The theory of radiation damage is not fully understood. However, the experimental observations on the mass-loss of organic material when it is subject to ionizing radiation are well established. Total charge of approximately 10^{-3} to 10^{-2} coul/cm^2 cause rapid and large scale loss of organic matter. These currents correspond to .6 to 6 primary electrons passing through each square angstrom of the target material. Furthermore, the approximate cross section for inelastic scattering can be calculated to show that the observed mass-loss occurs when secondary electrons are ejected from a large fraction of the target atoms.

(BEAM-INDUCED LOSS OF ORGANIC MASS UNDER ELECTRON MICROPROBE CONDITIONS by T. A. Hall and B. L. Gupta, Journal of Microscopy, Vol. 100 Part 2, March (1974), pp. 177-188.)

Scanning electron microscopes (SEM) with field emission electron sources and methods for focusing electron beams which produce very high current densities in very small spots represent the third relevant development. Current densities of 10^4 to 10^5 amps/cm^2 can be achieved with a spot having a diameter of less than 50 angstroms. Furthermore, in currently available SEMs the stability and precision of the deflecting system is sufficiently great so that points in the target plane can be addressed with an accuracy of one in 10^5 to 10^6 in each dimension. Thus, each of some 10^{10} points can be addressed in a two dimensional target by electronic deflection of the beam. In addition, systems have been developed for controlling beam position with a computer and for detecting secondary electron emission, primary electron scattering, and electromagnetic radiation induced by the electron beam. (For example, Model 106 STEM, Coates and Welter, Inc., Sunnyvale, CA.)

In order to visualize organic material in an SEM, particularly in the reflection mode, the material is ordinarily coated with a thin metal film. This film, deposited by sputtering or evaporating in a vacuum, increases the contrast of surface contours of the organic material by a very large factor. In addition, the shape of the surface is stabilized by the metal fi after it is deposited, there is no longer any change in the shape or scattering power of the surface when it is further exposed to the electrom beam.

These properties of organic material in an SEM mean that it is possible to construct a write-once, read-only computer memory in which each bit corresponds to one biological macro-molecular particle with dimensions of the order of one or two hundred Angstra in a two-dimensional lattice of memory molecules. Data could be written in the memory at the rate of about 10 Megabits/sec and after a block of data in an area of approximately 0.1mm^2 had been written, it could be stabilized by evaporating a metal film on the small surface. Subsequently, it could be read by using the scanning electron microscope in a conventional manner and processing the output video data to yield a bit stream at a rate of approximately 10 Megabits/sec.

Since the uncoated memory molecules are unstable under electron bombardment they cannot be examined before writing. Therefore, the scan corrections for the rotation and translation of the memory lattice must be determined by examining the selected region around its periphery. This can be done by evaporating a grid of metal-coated strips over the entire lattice and making use of the fact that within such strips the location of the memory molecules could be determined without radiation damage. Such grid strips would be prepared by conventional lithographic techniques with widths of approximately 0.5 micron and the arrangement of the lattice memory molecules within a grid square could be determined by scanning with the electron beam around its periphery.

The basic principles described here are an attractive basis for further study: we still need empirical verification to know how long would take to write one block and with what reliability or error rate could the system operate. These molecular crystal are known to contain defects at which crystal planes slip with respect to one another. The process of scanning the periphery of a grid square could help in detecting such defects and, if one were present, the square could be rejected. Obviously, the immediate problem in developing a molecular memory of this kind is connected with the target material. There are many different macromolecular arrangements which could, in principle, be used and their ease of preparation, stability and ability to be modified would have to be studied with currently available SEMs before any real design could be made for a practical system.

In the practical operation of a molecular memory of the type descirbed here, one might have a target a approximately one square meter composed of many small molecular lattices embedded in a plastic matrix. Each of the separate lattices would be mounted on a wire mesh and a particular opening in the mesh could be moved into position by a mechanical translation of the target. The electron beam would be used to locate the edges of the wire and therefore determine the reference position for a square, and within the square, the orientation of the molecular lattice could be determined by examining the molecules in the strips precoated with metal. The time for the mechanical motion would be of the order to 50 to 500 milliseconds and once the orientation of the lattice is determined, the time to position the beam to a defined point within the square would be of the order of 2 to 20 microseconds. Then a faster scan could be initiated during which the transfer rate would be of the order of 1 to 10 Megabits/sec. Provision would have to be made for coating a small region of the target with metal, without disturbing the rest of it. If one arranged the target so that the molecules are on a plastic backing away from the electron source, then low voltage electrons could be used to find the position of the wire while higher voltage electrons, of perhaps 25 to 50 kilovolts could be used to address the molecular memory èlement which would then be on the bottom of the plastic.

APPENDIX C

INNOVATIVE TECHNOLOGIES FOR COMPUTER MEMORIES

by

Jan Rajchman

The following is a list of means to construct a memory ranging from established techniques to possible approaches. Commentaries about items 1-5 appear on pages 5-6 of this report; commentaries about items 6-14 are sketched out below.

The possibility of other approaches is, of course, the main hope of the ARPA enterprise. Indeed we believe that the probability of success is good. There has been a stupendous advance in relevant technologies with many more options to consider by many more competent individuals to consider them than was the case when the present techniques came into being. What may help most is a fresh look.

List of Memory Techniques

A. Established Techniques

1. Semiconductors
2. Magnetic recording

B. Systems or techniques recently developed or the subject of recent research

3. Magnetic bubbles (mostly BTL and IBM)
4. Memories and processors based on the Josephson effect (IBM)
5. Read-only rotating discs (RCA, Zenith, Philips, and others)
6. Read-only, write-once, mass storage system (Unicon 609)
7. Optically addressed write-read rotating discs (IBM, Honeywell, Overseas labs)
8. Holographic Memories (RCA, BTL, RADIATION, CSF, Siemens, and others)
9. Electron beam accessed memories (SRI, Microbit, and others)
10. Electrochemical Memories (University of Michigan)

C. Possible topics of interest

11. Delay line memories
12. Colloidal, photochromic, and macromolecular materials
13. Organic materials
14. Molecular biology - materials contribution

6. Read-only, write-once, Mass storage system (Unicon 609)

This system stores 2 x 10^9 bits on a thin rhodium-plated data strip approximately 4-3/4" wide and 31" long. Reading is by detection of the difference in reflectivity between a "burnt" and "non-burnt" hole. The strips are mounted on drums and tracks are selected by galvonometer deflection of the laser beam. Total storage capability with 500 strips is 10^{12} bits.

The system is manufactured by Precision Instruments Company and is presently undergoing operating tests at NASA Ames Research Laboratory at Moffett Field, California. It is an example of very high capacity mass storage with typically best access time possible with high density mechanically transferable medium. It also provides an experimental unit to determine the utility of a read-only write-once mass memory with a low cost disposable storing medium.

7. Optically addressed write-read rotating discs (IBM, Honeywell, overseas labs)

The disc is coated with a thin film of magnetic material with strong magneto-optic properties. Writing is achieved by heating micron sized bit spots by lasers to temperatures over the Curie point and by letting them cool in the presence of a magnetic field whose direction determines the stored bit. Non-destructive weaker light is used for reading, and is influenced by the direction of the stored magnetic field.

A significant project at San Jose IBM with cryogenically cooled discs using films of europium oxide and arrays of gallium arsenide lasers proved not to be competitive with magnetic recording. Considerable work at room temperatures with films of manganese bismuth is continuing, notably at Honeywell and many laboratories abroad.

No additional support is recommended in this area unless a new radical concept appears.

8. Holographic Memories

A mass memory with the storage capacity of mechanically accessed magnetic discs and yet with random accessibility at electronic speeds became a possible concept with the advent of holography. The laser beam is a long pointer at the end of which interference phenomena (due to another conjugate beam) creates a fine structure--that makes possible redundant area storage of bit density 10^6 to 10^7 bits/cm^2. Powerful selection addressing schemes are possible due to the independent freedom in page location and composition.

All aspects of the concepts were experimentally demonstrated in a scaled-down working system (RCA) and detailed analysis showed the desired mass storage goal to be attainable (Siemens, CSF, RCA, Radiation, and others). The main difficulty resides, not in the information handling and system aspects, but in the energy aspect. There simply are not sensitive enough storing light responding materials or practically usable lasers that are powerful enough.

We believe that imaginative proposals for very sensitive in situ recording materials would be a legitimate area of support, as all industry has essentially ceased further development on this approach. Also novel approaches should be considered that would radically increase light utilization by an order of magnitude (only 1% is used at best in present designs) or simplify optics (very high quality lenses are now required).

Incidentally, no other concept has yet been proposed that combines in one device permanent recording capability with the high speed random accessing (and even content addressable accessing).

Besides write-read work, a number of read-only holographic memories were developed. Capacities up to 10^8 were attained (in Japan) and larger capacities are possible. So far none has found any relevant uses. (BTL, Siemens, several Japanese companies.)

9. <u>Electron beam addressed memories (See also p. 10 and Appendix B)</u>

The electron beam is the electronic "pointer" par excellence and ever since the 1940's it has been the basis for "storage" and "memory" tubes. It was soon found in the early days that random deflection of the electron beam, unaided by feed-back markers on the target, proved difficult and surprisingly low capacities of at most a thousand bits per tube were realized at first. Feed-back systems were found both cumbersome and slow. There was the conception and development of the purely digital (but expensive) selectron. Over the years advances in electron optics and in A to D converters made far more ambitious beam addressed densities reasonable. The ultimate in precision of deflection has been reached with the scanning electron microscope where perhaps as many as a million positions can be reached at random. This is obtained with extreme care in electron optic lens design, stability of power supplies, shielding against stray fields, etc. Relatively small electron beam currents are obtained. There have also been important concepts in compounded selection that demand in each of two consecutive deflection steps only the square root of the ultimate resolving power.

For storage, most systems have been using surface electrostatic charges obtained by primary bombardment and secondary emission from an insulating surface or floating metallic islands. (Among recent efforts are: extended work at SRI, also at CE, in Japan and various individual efforts, such as Ph.D. theses).

A significant recent development (at Microbit), as in the use of charge stored below the surface of semiconductive silicon at a depth penetrable by the high speed electrons from the beam. The mechanism of storage provides for inherent amplification of charge in the silicon that increases by two orders of magnitude the stored charge and read-out signal. It also tends to minimize detrimental interactions between adjacent bits. It appears at present that densities of 10^7 bit/cm^2 are expected but have not yet been demonstrated. Also, estimates indicate possiblities of one or two orders of magnitude greater densities. Hence, with a reasonable number of tubes, memory system capacities of 10^8 to 10^{11} are considered by Microbit. It appears reasonable to consider this record and is likely to yield smaller storage capacities than estimated by its proponents, to be a candidate for the "gap" between discs and high speed transistor memories, rather than as a mass memory.

Despite the fact that the electron beam (or more generally electron optics) is among the oldest of all technologies used for memories, it nevertheless should be considered as a prime contender for new ones. Indeed electron optics permits to "see" at the smallest scale yet attained, the electron beam carries some of the highest energy densities attainable, and very low inertia permits high speed of communication. However, in this case, particular care must be exercised in deciding what is pertinent for ARPA to support.

10. Electrochemical Memories

It is possible to store a bit by electroplating a chemically active metal on one or the other of two chemically inert electrodes. This idea is not new, but a recent study at the University of Michigan just published, has demonstrated the potentialities for a voltage-coincident arrayed memory based on it. According to the study, one could hope for capacities of 10^8 bits, access time in hundreds of microseconds, and relatively easily made structures because packing densities need not be excessively high. Considerable difficulties remain mostly because of non-ideal behavior of the electrolyte. This non-conventional approach may have very significant potentialities. The critical issue is the behavior of electrolytes at high speeds and high current densities. These are areas in electrochemistry and science and art to which very little attention was given so far.

11. <u>Delay line memories</u>

Disturbances propagating naturally in a medium, such as sound waves, can be the basis of a memory by simply restoring the weakened and distorted outputs to their original strength and shape and feed them to the input. This is an old idea that was the basis of the early computers. The greater versatility of random access memory has largely displaced delay memories.

Today, it may be significant to have another look at this approach, which is inherently of low cost since it employs a continous medium and requires no construct-per-bit. In the first place, another view of computer organization may make it more compatible with serial access than the conventional one, as is considered in some detail in this report.

In the second place, recent technological developments may in fact make the serial nature of the delay compatible with conventional organization. There has been considerable recent development of wrap-around delay lines in which the information is stored in surface acoustic waves travelling over crystals or non-crystalline solids. Recent work of I. Mason and others at University College, London, employs piezoelectric films on lense-shaped quartz wafers. Shiren of IBM Research has described a mechanism occuring in piezoelectric semiconductors which traps signals in the form of acoustic bulk waves at frequencies up to 100 GHz.

A more speculative idea would be to use optical delay lines. Very low loss lines were obtained by multiple reflections between spherical mirrors about ten years ago (BTL). It may be possible to use such mirrors for many beams simultaneously. Furthermore, a beam could carry a wide frequency band of information that is multiplexed in and out by optical beams. The large bandwidth capability of light barriers could be used without the concomitant necessity of electronic switches of that bandwidth. The electronic circuits operating at their highest bandwidth, which is still much lower, could be tricked to effectively yield a random access memory. Such an approach would try to exploit some of the recent advances in optics, such as non-linearity, the whole area of diffraction technology revitalized by holography, laser technology, etc. This is given here as pure speculation, perhaps as an example of ideas that should be proposed to ARPA for study.

12. Colloidal, photochromics and macromolecular materials

The electrolytic cell operates by transporting material from electrode to electrode, through ionic sized particles. More material per unit charge can be transported the larger the transporting corpuscle.

An electrophoretic display device developed at Masishuta (Ota et al) depends on transporting micron-size colored particles suspended in a liquid of another color between front transparent electrodes where their color is visible and a back electrode where it is masked by that of the suspending liquid. Since the particles stay where last driven, this form of display is a memory.

Similarly, there is considerable work with so-called photochromic materials which produce large changes of color upon application of an electric field due to the change of generally large organic macromolecules. Here again relatively small electronic charges produce large effects.

13. Organic Materials (See also pp. 18-19)

Thus far electronic devices did not exploit any properties of organic materials. Yet the extraordinary richness of properties that seems possible with the modern understanding and material synthesis techniques would suggest that this is an area which could be exploited if the imagination and skill of the disciplines of the organic chemist, the electronics and computer experts could be combined.

A suggestive example is the work at Penn which synthesized materials conducting one to two orders of magnitude more in one direction than the two others. Very non-linear dielectrics would be another possible example.

While this is a possible area that should be supported, there is the danger of directing ARPA funds to basic scientific work--however meritorious it may be--if there is not some notion of the effort's applicability to memory. We do not believe that the ARPA funds are sufficient to "seed" all basic work. Fruits will be born more likely by the association of invention and science, than unguided science alone. Clearly, some phenomena are so intriguing by the strangeness or magnitude that it may be taken for granted that their greater understanding would lead to useful notions for memories. It is a question of judgment for each case.

14. <u>Molecular biology - materials contribution (See also pp. 19-20)</u>

The human memories organization and biological memories system may or may not be of great significance to the development of artificial "stores of information" that we are in the habit of calling memories.

There are likely to be contributions from the biosciences other than on the systems level. It is possible for example that some of the special materials that are bio-synthesized may be very useful to construct memories.

APPENDIX D

GENERAL NOTIONS ABOUT MEMORIES

by

Jan Rajchman

There are general notions about memories that could guide would-be inventors. These are mostly notions of former researchers who have led them to success or else to some unsurmountable obstacles. Often these notions are so vague, so intuitive, so controversial, or so trivial that they are not clearly articulated and seldom if every published. There should become encouragement to bring these ideas to light.

Here are some examples:

1. One general characteristic of the thirty year history of modern computers is that technologies for storing information have lagged behind those for logic processing. This is so much so that usually the characteristics of memory and storage devices is taken as the main description of the whole system.

Perhaps a basic reason for this is simply the following. The unit of storage--be it a discrete cell, a spot on a continuous medium or a travelling disturbance--has associated with it a means of identification or "addressing mechanism" that is much simpler than that required in a unit for manipulative logic. This mechanism is at most an elementary logic "AND" in core or other matrix-like memories, a controlled displacement of an electron beam or laser, the timing of the motion of a disc or tape, or of the natural propagation of a disturbance. Despite this much greater simplicity of the storing cell as compared to the information-generating-logic cell, still in general, the required number of storing cells is so much greater that the overall storage system is more onerous than the overall logic system. Fortunately, many fewer bits need to be manipulated with each other than simply stored, for otherwise these simpler storage cells would make no sense.

It seems therefore, that the technologies for storage and logic that are different, have good reason to remain different.

The recent success of transistor technology in memories tends to run counter to this historical and seemingly inherent trend. Even here, LSI is most successful in high speed memories when the logic unit cell of storage deviates most from a true logic cell and is simplified to the ultimate by a "capacity sotrage with dynamic refresh" approach. Then its per-bit-cost is low enough to provide the high-speed needs of ROM. Cheaper yet are CCD memories with lesser cell complexity, but they provide only serial rather than random access.

The normal evolution of industry will show where the LSI technology will lead to. It may in fact be sufficiently promising to warrant a fresh look at distributed high-speed memory-logic and content-addressable memories that thus far simply were not economically attractive in comparison to alternatives for the same functions.

However, outside these areas, the logically potent semiconductor technology is much too expensive for the simpler tasks of mass storage of large amounts of information, where specially tailored other technologies are needed.

2. The physical nature of the bit storing unit and the mechanism of identifying or addressing the stored bit are two essential and usually inseparable aspects of any memory concept.

The state of magnetization, polatization, electronic energy level occupancy, surface charges, elastic deformation, persistent or tunneling currents are all well-known examples of physical states used in memory cells. The identifying mechanism generally depends on the physical location of the storing cell, hence the general use of the term "addressing". Addressing is by three principal means (i) by coincidence of excitation of conductors in a 2, 3 or n-dimensional array, (ii) through a "pointer" such as an electron beam, a laser beam, a magnetic head or a recording needle. Either the pointer moves on a fixed medium or vice versa.

It is interesting to observe that in the whole art of electronics there are only these four main pointers. (iii) through the natural propagation of a disturbance such as an elastic wave in a delay line.

Holographic recording spreads any bit over an area (or a line) and bears the same relation to spot bit recording in the space domain as does frequency modulation to pulse modulation in the time domain. Here the identification of the stored bits is not strictly related to their physical location.

One could envisage a volume of inhomogeneous material such that if it were irradiated as a whole by some radiation bearing an appropriate code (such as sharp spectral line or lines, or time sequences of specific frequencies), certain parts of it would be transformed and in effect store the event of such irradiation. Subsequent irradiation would cause the storing material to emit some other radiation revealing their state. Such a highly speculative example (or fantasy) is given here to illustrate the possible idea of a "lock-and-key" type of memory in which information is broadcast and finds its own specific responders. Here identification would be fully disassociated from explicit physical addressing.

3. While memory is essentially an information-centered device, it necessarily requires energy for its operation. In general the energy expanded to store a bit (in write in-situ memories) is relatively large because of the relative insensitivity of most storing materials and of the necessity of obtaining a read-out signal sufficiently large to dominate over inherent and man-generated competing signals. Energy delivery is in fact an important aspect of addressing. It is very efficient in magnetic recording. In memories with no mechanical motion, energy delivery is most efficient in electron beam and laser pointers that have high evergy densities and carry energy and address information simultaneously. In matrix addressing much energy is wasted in non-selected locations. This would be even more so in a volume broadcast-type identification memory system.

4. Bit densities have constantly risen both on magnetic tapes and discs, in cores and in semiconductor memories. Higher bit densities in a moving medium provide higher bit rates and economies in required storing surface. Higher bit densities in semiconductors provide greater economy in silicon--a significant factor in determining cost--and tend to increase speed. (The same arguments were valid for core.)

It is only common sense to demand miniaturization in memory. When dealing with information, matter and energy are things to put up with, hence the less of them the better. We generally agree with that notion.

It may turn out, however, that there is some optimum size, not necessarily the smallest. When making a mass memory the controlling factor is cost. The convenience of fabrication is highly sensitive to size, very large and very small things are more expensive than those for which most shop techniques are applicable. Furthermore, elements that have been miniaturized to the extreme--far below the wave length of light for example--can be used only with very sophisticated detecting techniques such as high precision electron beams. The apparatus to produce and control such beams requires a volume that is ridiculously larger than the volume occupied by the storing medium. Hence, no reduction in overall volume of the hardware results. Cells with larger areas occupying the same or a smaller volume may, in fact, be less costly. The matter is clearly subject to scrutiny.

Most memories today can be thought of as two-dimensional; surface of the recording magnetic medium, or surface of the silicon chips or surfaces in most other types. The question as to whether three-dimensional storage is more desireable has often been raised. In general, it turns out that it is easier to increase one-dimensional density by the square root of its attained value (from n to $n^{3/2}$ leading to n^3) than to shift to a three-dimensional array with the same one-dimensional density ($n \rightarrow n^3$) because in the three dimensional implementation the freedom for inventing various art effects is quite limited. For example, it is easier to make memory with a plane 1000 x 1000 than with a cube 100 x 100 x 100. Matrix type memories with n leads passing through each cell can be thought of as n-dimensional. For 3 leads, they can be thought of as a three-dimensional memory. In the case of 3-way addressed cores, one finds it more convenient to wire the cores by planes.

These remarks on miniaturization and dimensionality are made here not as having absolute validity, but as cautions against sole reliance on loosely defined common beliefs.

APPENDIX E

ENCOURAGING INDUSTRIAL RESEARCH IN ADVANCED MEMORY SYSTEMS

by

V. A. Vyssotsky

November 12, 1974

In this note, I shall briefly consider how ARPA funding of advanced memory systems research may fit with the industrial R&D picture. Much of the note is background material which is not deep or novel but which I include because I'm not familiar with any place where it's written down in just this form.

1. A Hypothetical Example

To set the stage, let's consider a mythical example of an advanced memory system project as contemplated by some major company. Our memory will be based on new device technology, will have access time in the range of microseconds, will have capacity 10^{15} bits, and will sell for $10M ($10^{-6}$ cents/bit), including 40% gross (pretax) margin. Research leading to this memory will start in 1975, and production will begin in 1995. R&D cost will be:

$0.6M/year in 1975-1984
$1.0M/year in 1985-1989
$2.0M/year in 1990-1994
$1.0M/year in 1995-1999

Incidently, although the time scale is plausible, these R&D costs are quite low. But let's forge ahead.

How many will we sell? At $10M a piece, the overall market is only likely to be a few hundred, and our company won't get the whole market, so let's assume we sell 100, for total sales volume of $1000M, with unit sales by year as follows:

1995 - 2
1996 - 5
1997 - 10
1998 - 10
1999 - 15
2000 - 20
2001 - 15
2002 - 10
2003 - 8
2004 - 5

The market for such memories will, of course, continue past 2004, but by then we will need improved versions, requiring more R&D, so we must amortize initial development costs over these 100 units. We'll need a production facility. Suppose we provide this in two phases, one in 1994, the other in 1996. Each phase takes a $40M capital investment, depreciated on a straight line over eight years. Each phase also taken $5M in expensed start-up costs in the first year of production, and another $1M the following year.

Putting all of this together, and with a 50% effective tax rate, the earnings impact by year is shown in Table E1. This looks good. We net $141M, with an economic ratio of 11.75.

But it isn't good. For one thing, we've ignored the time value of money, which is maybe 10% per annum. For another, we've ignored the various risks. Then, too, it may not be convenient to supply $80M of capital when it's needed. And, finally, we've ignored the opportunity cost of other things we could have done instead. No company evaluates projects quite the way I'm about to do, but it's not too far off, though oversimplified, to lump all of these additional considerations together in an assumption that the time value of earnings impact in such a venture is 20% per annum. This gives us the result shown in Table E2. We now have a net loss (discounted) of $0.337M, and an economic ratio of -0.18.

Without going into details, I'll just observe that although changing the assumptions, as for example by doubling sales, will change the result, even selling 200 units doesn't make the picture compellingly attractive. This, in a nutshell, is why most companies don't do much long-range R&D. The CEO has better uses for his money.

2. Why Companies Do Research

The example above makes it seem a little puzzling that organizations like IBM, Kodak, DuPont, and Bell System do any long-range R&D at all. They do so for a variety of reasons, but my observation leads me to believe that one key train of thought dominates, though it is seldom stated explicitly in entirety. As Peter Drucker has pointed out, a large company is not run to achieve maximum profits, but rather to assure the long-term health of the enterprise. Adequate earnings are an essential ingredient of this, but are only part of it. There are several other factors. One of the major ones is preservation of markets, and in some cases expansion into new markets. But even if no new market is envisioned, the existing market isn't static, and its changes are unforeseeable. So preserving the health of the enterprise requires having a continuous flow of options available to the CEO to introduce new products and services.

Now a company with a small share (say less than 10%) of the total market of interest to it can depend on having new product opportunities made for it, by its competitors or by its suppliers or customers or by the general marketplace. The company with a small market share, then, can thrive if it's alert enough to seize opportunities which are created external to it. But the company with a large share of its market is vulnerable to having major segments of its market redefine themselves away from the company's line of products and services. Recent classic examples of this are the Penn Central and the Cunard Line. It happened to Ford in the late 1920's. The company may recover from such a traumatic shift, as Ford did, or it may go down, as the Penn Central did, but no alert management will risk such an episode if it can help it.

So the company with a large market share (and these are often large companies, but not always) will seek to anticipate market needs and encourage the overall market by "getting there fastest with the mostest". Research is an essential component of being able to do that consistently, even though research may be a net drag on earnings. So far as I can see, when you get rid of all the other, less important factors, this is why companies do long-range research.

3. Relationship to the ARPA Program

In view of all this, how can ARPA get the most effective participation by industry in research on advanced memory systems? Well, first off, there's not a big multiplier to be gained in having a company do a research project unless the company perceives the possible results as potentially applicable to its field of business. Lots of companies will happily spend government money on R&D just to get the cash flow, and do an honest job of R&D for a buck, but that's an end to it unless the company foresees a possible product or service in its own market.

Second, the closer the time when the company foresees some possible sales to recover its own R&D costs, the more likely it is to put in its own money. Thus, for example, if in pursuit of "molecular memories" ARPA wants to encourage research in protein biochemistry, a pharmaceutical company is more likely to take off and run than is a computer company, because the pharmaceutical house can envision a market for tailored proteins per se, which the computer company cannot; the computer company would have further intervening steps to take before it could bring a computer product to its market. Nor would the computer company normally try to enter the marketplace for pharmaceuticals or chemicals; it lacks experience in that market, and can't reasonably expect to build a self-supporting product line.

Next, let's observe in Table 2 the disproportionate impact of the first few years of R&D expense. The discounted earnings impact of the $3M R&D expense in 1975-1979 is about enough to offset the discounted earnings impact of an added $500M in sales in 1995-2004. Thus, if ARPA wishes to support industrial R&D, the maximum multiplier is likely to accrue from supporting the earliest stages of R&D; such support can make a project which otherwise falls just off the bottom of the company's "will do" list seem attractive enough to undertake.

4. The Industry/University Balance

When universities are sorely in need of research funding, why not place all the ARPA money with universities, and none with industry? This may turn out to be the appropriate result, but it would be unwise to adopt it as an a priori constraint for three reasons. First, certain types of research cannot reasonably be conducted in universities, because of prerequisite concentrations of skills or physical facilities which exist only in industry. Much process research falls in this category.

More importantly, where a project ARPA wants done is also in the self-interest of a company to carry out, but just falls off the bottom of the company's "will do" list, ARPA funding assistance can result in a commitment of company money much larger than the ARPA grant. This multiplier may be far greater in an industrial setting than is typically the case in a cash-starved university. So spending some portion of the ARPA money in industry can be a way of stretching the limited supply of ARPA dollars.

Finally, and also important, if a company is putting its own money, as well as ARPA's, into a research project, corporate management will be asking itself regularly how the research results can best be incorporated into saleable products. In this case ARPA gets, free of charge, a lot of product planning effort which may not get done at all if the same research is done in a university.

5. Pitfalls for ARPA

ARPA is presumably well aware of the pitfalls of dealing with industry, so I'll just mention one which seems to me especially dnagerous. Most industrial R&D organizations, like most universities, are short of money. This being so, it may be attractive to some companies to try to get a free ride on ARPA money, in either of two senses. First, the company may see an ARPA-funded project as a way of keeping together a research group which would otherwise have to be disbanded. This is not of itself a bad motivation, but a project done on this basis is unlikely to lead to a large corporate commitment of follow-on dollars.

Second, a company may find it attractive to use ARPA dollars to pay for research which the company would have done anyway, thereby releasing its own research dollars to be used on other projects which may be of no interest at all to ARPA. If this were to happen, the ARPA expenditure would have zero impact in the areas of interest to ARPA.

6. Pitfalls for Industry

Some industrial R&D labs are very wary of government funding. There are several reasons for this, of which four are perhaps the most significant. These may not be so apparent to ARPA as the other side of the coin, above, so I'll take a little more space on this.

First, many companies feel that government funding is unreliable; it seems to increase or decrease for reasons which may be apparent in retrospect, but which are very hard for corporate management to foresee and base plans upon. This has a destabilizing effect on the company's own R&D budget, since most companies strive to maintain a stable work force and a stable work program in R&D.

Second, some companies feel that government funding of R&D complicates the management of work. In most industrial research efforts the technical management finds it necessary from time to time to speed up, slow down, change emphasis, detour temporarily, or even alter direction completely; the one sure thing about research is that you will be surprised, and not all of the surprises are pleasant. Some companies have had bad past experience in this respect with government-funded work; managers have tales of frustration about urgently needing to change course, but being delayed by the need for agency approval of the change. To be sure, there are lots of other projects where this has been no problem.

Third, reporting requirements on government contract work often seem inordinate to industrial contractors, and in some cases this biases a company against doing any work at all on contract for the government.

Finally, every company is concerned about the possibility of having valuable information created by the company at its own expense flow into the public domain, through a government contract pipeline. Such a loss can be avoided, of course, but it takes considerable management attention and effort to preclude it.

7. Summary

ARPA will surely wish to deploy in a university setting much of its money for advanced memory systems research. It should also consider funding research by industry. For certain types of research, universities may not be equipped to do that research as well as industry (or, indeed, to do it at all).

In addition, ARPA may be able to get a large multiplier effect for its dollars by using the dollars for industrial R&D. To achieve this multiplier effect on any particular project, ARPA must seek satisfactory answers to the following questions:

- Does this research fit the company's own business thrust?
- Will the company commit its own money, as well as ARPA money, to this line of research?
- Would the company do the project on its own money if ARPA didn't provide any help?
- If good research results are achieved, is the company likely to carry the project forward through development of its own?
- Does the proposed arrangement give the contractor enough assurance of funding, protection of rights and freedom of maneuver to make the contractor's management feel comfortable with the job?

(signed)

V. A. Vyssotsky

VAV:clf

Att.
Tables E1 and E2

Copy (with att.) to
Members of Advanced Memory Concepts Group

TABLE E1

AFTER-TAX EARNINGS IMPACT ($M)

Year	R&D	Depreciation	Start-up	Sales Margin	Total	
1975	-0.3				-0.3	
1976	-0.3				-0.3	
1977	-0.3				-0.3	
1978	-0.3				-0.3	
1979	-0.3				-0.3	
1980	-0.3				-0.3	
1981	-0.3				-0.3	
1982	-0.3				-0.3	
1983	-0.3				-0.3	
1984	-0.3				-0.3	
1985	-0.5				-0.5	
1986	-0.5				-0.5	
1987	-0.5				-0.5	
1988	-0.5				-0.5	
1989	-0.5				-0.5	
1990	-1.0				-1.0	
1991	-1.0				-1.0	
1992	-1.0				-1.0	
1993	-1.0				-1.0	
1994	-1.0				-1.0	
1995	-0.5	-2.5	-2.5	4.0	-1.5	-12.0
1996	-0.5	-2.5	- .5	10.0	6.5	
1997	-0.5	-5.0	-2.5	20.0	12.0	
1998	-0.5	-5.0	- .5	20.0	14.0	
1999	-0.5	-5.0		30.0	24.5	
2000		-5.0		40.0	35.0	
2001		-5.0		30.0	25.0	
2002		-5.0		20.0	15.0	
2003		-2.5		16.0	13.5	
2004		-2.5		10.0	7.5	153.0
					TOTAL NET	141.0

TABLE E2

AFTER-TAX EARNINGS IMPACT ($M)

Discounted

Year	Impact From Table 1	Discount Factor	Discounted Earnings Impact	
1975	-0.3	1.0000	-.300	
1976	-0.3	.8333	-.250	
1977	-0.3	.6944	-.208	
1978	-0.3	.5787	-.174	
1979	-0.3	.4823	-.144	
1980	-0.3	.4019	-.121	
1981	-0.3	.3349	-.100	
1982	-0.3	.2791	-.084	
1983	-0.3	.2326	-.070	
1984	-0.3	.1938	-.058	
1985	-0.5	.1615	-.081	
1986	-0.5	.1346	-.067	
1987	-0.5	.1122	-.056	
1988	-0.5	.0935	-.047	
1989	-0.5	.0779	-.039	
1990	-1.0	.0649	-.065	
1991	-1.0	.0541	-.054	
1992	-1.0	.0451	-.045	
1993	-1.0	.0376	-.038	
1994	-1.0	.0313	-.031	
1995	-1.5	.0261	-.039	-2.071
1996	6.5	.0217	.141	
1997	12.0	.0181	.217	
1998	14.0	.0151	.211	
1999	24.5	.0126	.309	
2000	35.0	.0105	.368	
2001	25.0	.0087	.218	
2002	15.0	.0073	.110	
2003	13.5	.0061	.082	
2004	7.5	.0051	.038	1.694

Net Discounted Impact -.377

MISSION
of
Rome Air Development Center

RADC is the principal AFSC organization charged with planning and executing the USAF exploratory and advanced development programs for information sciences, intelligence, command, control and communications technology, products and services oriented to the needs of the USAF. Primary RADC mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, and electronic reliability, maintainability and compatibility. RADC has mission responsibility as assigned by AFSC for demonstration and acquisition of selected subsystems and systems in the intelligence, mapping, charting, command, control and communications areas.