VOLUME 2 □ NUMBER 1 JANUARY 1965

IEEE spectrum

features

Departments: *please turn to the next page*

the cover

The illustration on the front cover of this issue is of a portion of a "Twystron" hybrid amplifier manufactured by Varian Associates. It combines the RF driver input section of a multicavity broadband klystron with a traveling-wave-tube output circuit. For the future of the electron tube, see page 50

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

VOLUME 2 □ NUMBER 1 IEEE spectrum JANUARY 1965

Spectral lines

Always a Student! Continuing education may be a newly recognized need in the engineering field, but neither the name nor the method is new to many of our Land Grant-derived engineering colleges. The philosophy of such schools, and of some others as well, has always been to take education to the student, if the student could not come to the campus.

Perhaps much of our past effort in continuing education has been directed to improving the subprofessional, and we have somewhat neglected professional needs. But this emphasis can be excused if we remember that in the past we tacitly assumed that technology would change little during one man's lifetime, and many engineering designs were often based on past experience stored in the handbooks. Perhaps assumptions such as these were valid for a stable and bounded technical world. For a changing technology, for an unbounded world, for a world in which the results of research continually flow into production, it has become apparent that continued education for the professional is paramount. The machinery for such education already exists in many colleges; only a little oil and direction are needed.

It is obvious that the practicing engineer needs to know of new materials, new energy sources, new devices, but it is more important that he adopt new philosophies and concepts as well. He must accept the importance of reliability over mere functioning, of the transient over the steady state, of the system over the component, and of the mathematical method over cut and try. Therein lies difficulty—a fact may be accepted upon proof, but a concept must often be lived with to be adopted.

It might be possible to teach of new materials, energy sources, or devices through seminars, through tutorial talks, through symposia, but teaching the new concepts cannot be so easily handled. Attempts at such education disclose the considerable failure of past technical education. There was not sufficient breadth or depth in past education to provide a foundation for new concepts. One must start almost at the beginning and this will not be done through a few odd-hours courses.

It has been pointed out that most engineers with B.S. degrees granted prior to World War II, or even prior to 1950, now face problems unless they have actively continued with technical education. Since 1950 our engineering schools have granted more than 400 000 B.S. degrees, and if we place the current size of the profession in the United States at 900 000, it appears that perhaps one-half million persons need a basically new education. The problem is numerically staggering.

Added to this group are many who are in engineering work, but who are not holders of engineering degrees. Some Midwest data indicate that "engineering staffs" may have thirty to forty per cent of their manpower in this category. They too need help, but the starting point for this help will have to be adjusted to the needs.

Graduate education and degrees are not a solution for such numbers. Graduate schools usually assume that only the graduates in the top half of their classes have laid a sufficiently strong foundation to profit from advanced study. Obviously, the half million in need of continued education did not all have such high standing.

Education at night after eight hours of work will not solve the problem either. Such classes cannot attract large numbers of students because they must compete with the wife, the family, and community interests also challenging for those hours. Industry is going to have to accept responsibility, and carry on this work as a company or community contribution to the future. It must use teachers from company laboratories, from area universities, from community colleges, or recruit teaching talent where it exists. It must then make classes available on company time, and indicate that attendance is expected as for any company-sponsored training program.

I hope we are now educating to provide the proper base for continued learning for the classes of 1965 and beyond; we cannot forever undertake the job of rebuilding the base. Proposals to turn the clock back—to again educate "nuts and bolts" graduates—seem to overlook all the indications that pre-1950 engineering education has been shown insufficient for a changing world.

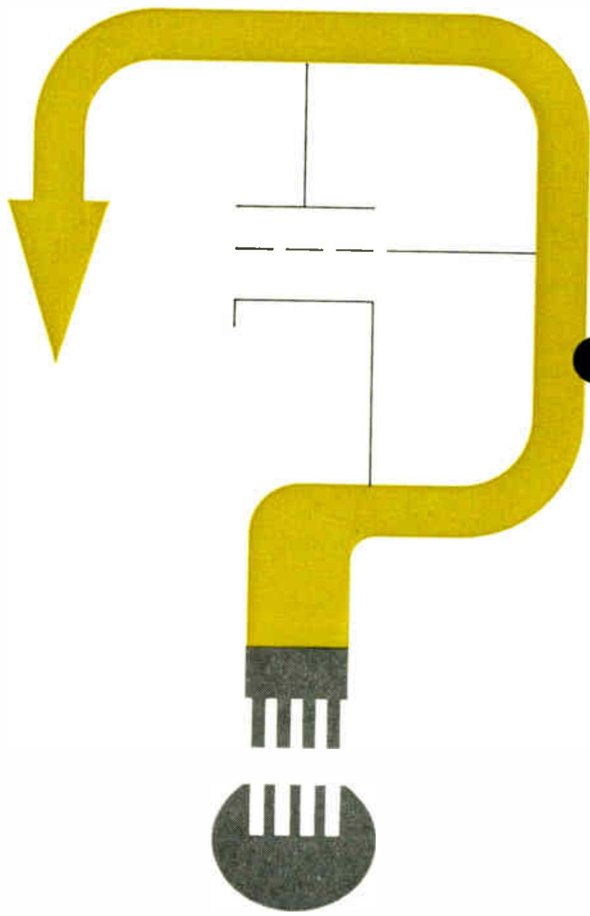
The problem actually has deep roots, and we do too little research on education. For a century the colleges of engineering have attempted to supply that education (or training) which they believed industry wanted. This objective must be seriously questioned for several reasons, among which are: (1) the needs of our diverse industry can no longer be described by a simple or single formula; (2) much industry, with its necessity for profit, cannot afford to take the long view that should be inherent in the educational process; and (3) education for industry may not be education for the man.

A new philosophy for engineering education must recognize the engineer as a member and leader of the community, must provide for sympathetic understanding of other fields, and must make it possible for the engineer to grow in mental capacity as the technical world grows in scope. It is time that our colleges direct their philosophies to the training of minds in technical learning, and to providing the mental drive and desire that will lead to continued learning throughout life. *J. D. Ryder*

The future of the electron tube

The future of the electron tube has been profoundly affected by the emergence of solid-state devices—but it is far from defunct. Tube needs are growing in such applications as power conversion, display, and optical sensing. But tubes of the future may have to be manufactured for service rather than for sale

E. W. Herold *Varian Associates*



It is a truism that almost every engineer's livelihood has been profoundly affected by the electron tube. Born with the invention of grid control by DeForest about 1906, it has been the mainstay of electronics for the past five decades. Radio, television, computers, sound motion pictures, and much of the field of energy conversion and automation, were either made possible or profoundly modified by outgrowths of the DeForest device. In addition, its influence has been felt in many other areas. The chemical industry, medicine and biology, transportation, education, finance—all have advanced and been altered by its use.

With the advent of the transistor in the 1950s, the character of the electron tube's influence began to change. The costs of many semiconductor devices have now become competitive with those of tubes, and the new microelectronic integrated circuits promise another major cost reduction, to a small fraction of the separate component costs. These developments have prompted predictions that solid-state devices will replace *all* tubes in the not too distant future.

An examination of present trends, however, does not bear out such a conclusion, particularly if the term "electron tube" is used in a broad sense to include new devices, tubelike in nature, but more closely related to present tubes in regard to technology rather than to principles of operation. Paralleling the growth of solid-state devices during the last 15 years has been a substantial change in the "product mix" of electron tubes. Television picture tubes, new microwave tubes, and pickup and storage tubes—all of which bear little resemblance to the DeForest invention—have now relatively high dollar value. Consequently, in 1964, the total dollar value of electron tubes produced in the United States remained much the same as it had been for the past ten years, and appeared to be less affected by the rising solid-state market than might have been expected. In the present article, we shall anticipate changes in the next decade or two to determine the product mix of the future electron-tube market. If the analysis is correct, it is important to both the tube manufacturer and the tube user that appropriate action be initiated today to take maximum advantage of the changing conditions.

Two interrelated factors can help in our estimate of future trends in the use of the electron tube. One factor is the anticipated needs and requirements of the electronics industry; the other is current research and development. These factors are said to be interrelated because research and development are so often stimulated and directed by anticipated needs and requirements. Both factors are considered in the following discussion. For the broad outlines of product mix, chief reliance will be placed on extensions of sales data into the future. For more specific illustrations, current research and development in materials and techniques, and in new applications of tubelike principles, will be used as an indication of what is to come. What is discovered in the research laboratory today inspires the development of tomorrow and the product of the day after that.

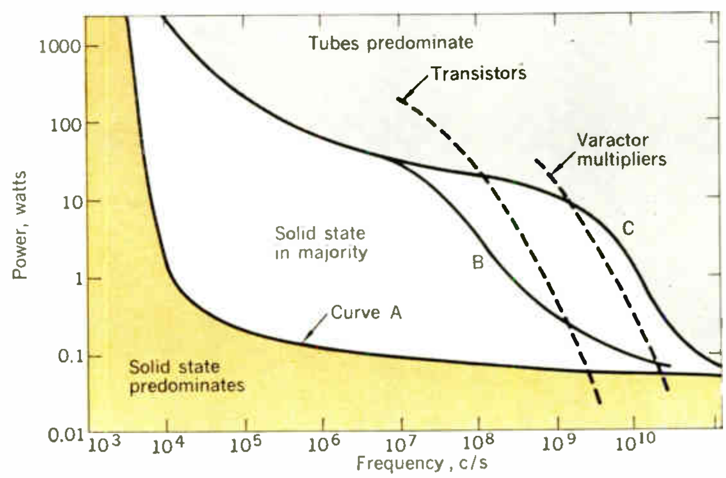
Tubes vs. solid-state devices for amplification and generation

A major inroad of the solid-state device, particularly the transistor, has been on the oldest electron-tube type—that using grid control for small-signal amplification or low-power oscillation. No effect of the transistor on high-power generation at high frequencies has yet been observed. Although some may believe that this condition is only a matter of time, it can be argued that generation in the latter region of the power–frequency plane is best initiated by control of currents in vacuum or in ionized gas; i.e., by the electron tube. The reason is simple. In the solid-state device, the interaction of the controlled currents with the crystal lattice takes place in the active amplifying region. This interaction produces heat. As the frequencies become higher, the dimensions become smaller, and there is difficulty in dissipating this heat when the power level is high. In the electron tube, on the other hand, the active region is a vacuum (or, at worst, a tenuous gas), incapable of interacting with the electron stream to produce heat. Instead, the electrons can be eventually directed to a metallic collector whose heat-dissipating ability is high, whose location can be remote from the active region, and whose size need not be directly related to the wavelength.

In an endeavor to show the anticipated relationships on the power–frequency plane, curves were prepared by Salvano and Garoff for an unpublished paper presented at the 1959 AIEE Winter General Meeting held in New York. Figure 1, which is a chart of average power vs. frequency for generator–amplifier devices, shows these relationships in somewhat modified form. As of 1964, electron tubes can be built for almost any spot on the plane. Transistors and other solid-state devices have not yet attained the highest powers and frequencies, although they may yet do so. For these reasons it would seem that, above curve *C*, electron tubes exhibit enough advantages to insure their predominance in the market, at least for the next two decades. During this period, only a parallel connection of lower-power individual solid-state devices is likely to compete above the 100-watt 100-Mc/s level. Since a single electron tube will fill almost any spot in this region with ease, the advantages of solid-state devices (other than their novelty and glamour) are believed to be insufficient to compete.

Below curve *A*, however, there appears to be little future need for the electron tube; the advantages are all in favor of solid-state devices, and they are eventually expected to operate at any point in this region. Between curve *A* and curve *B*, the electron tube may be fighting a losing battle although, as of 1964, the tube is far from superseded here. The greatest uncertainty lies between curves *B* and *C*. As indicated by the dotted lines, which show the 1964 state of the art, both the transistor and the varactor multiplier have penetrated this area; the

Fig. 1. Chart of average power vs. frequency for generator–amplifier devices. Tubes are expected to be predominant above curve C, while solid-state devices will establish market predominance below curve A.



eventual outcome is uncertain, but is not significant to the main point of the present study.

A conclusion drawn from Fig. 1, already widely accepted, is that the electron tube will be almost completely eliminated as the active element in all small-signal information-handling applications. These include small-signal amplification, generation, digital processing, and logic circuitry. In other words, the transistor and its allies will soon completely supplant the type of tube commonly called the "receiving tube." This conclusion will be utilized in predicting the product mix of the future. A second conclusion, if Fig. 1 is accepted as realistic, is that the electron tube will continue to predominate at the highest powers and frequencies.

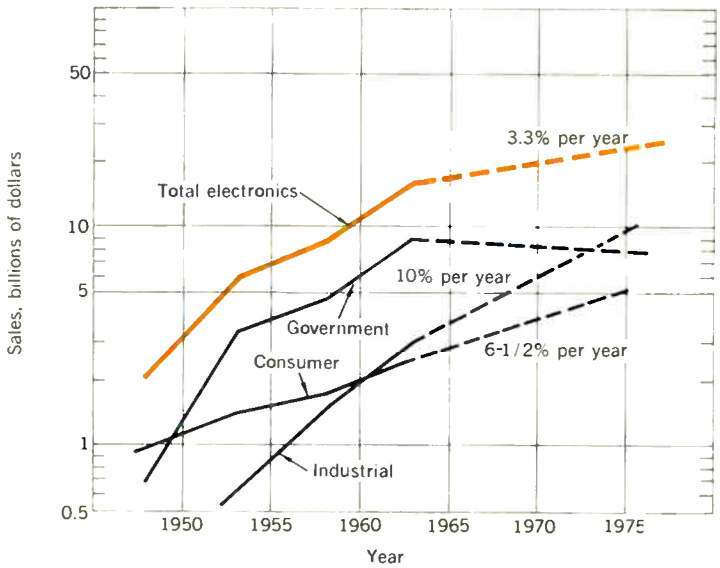
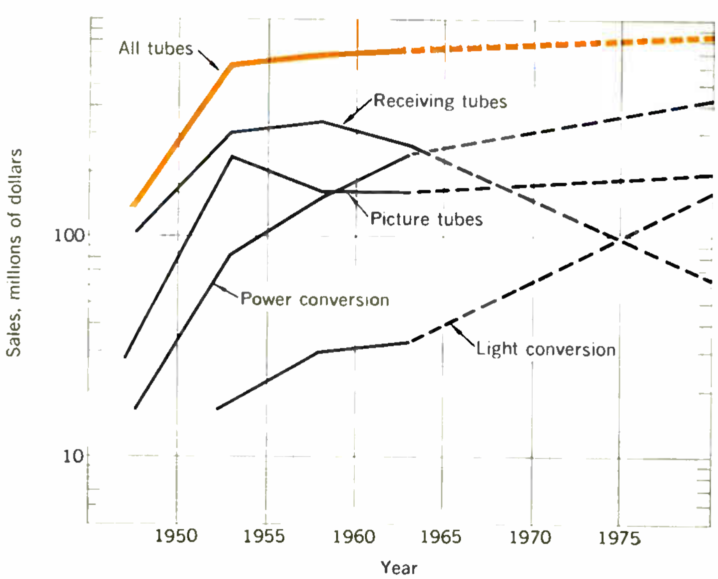


Fig. 2. Current electronic sales as reported by EIA (solid curves), and as predicted for the future (dotted curves). The major prediction is the rapid rise of industrial electronics in the years ahead.

Fig. 3. Current electron-tube sales as reported by EIA (solid curves), and as predicted (dotted curves). Note the ultimate dominance of power-conversion types.



Anticipated future needs

The future need for the electron tube will be based on anticipated markets for electronics; figures for the United States will be used as indicative of world-wide trends. In Fig. 2, the solid lines show approximate past electronic sales figures in dollars, smoothed between five-year points; the data are from a report by the Electronic Industries Association (EIA).[1] The top curve is the total for the three portions shown below it.

For the future, shown by the dotted lines, it was assumed that government sales, chiefly military and space, would drop by one per cent per year. If consumer items continue at a 6½ per cent per year rise, and industrial products taper off slightly to a 10 per cent per year increase, then the total electronic sales figure approximates the upper dotted line, with a 3.3 per cent per year rise. Others[2] have predicted a 7 per cent per year rise for this total, but this includes a slight rise in government sales. For comparison, the United States gross national product has been increasing at a rate of about 5.5 per cent, and the United States population at a rate of 2 per cent.

If one accepts the predictions of Fig. 2, or similar ones, the striking feature is the rise of the industrial market, which, by 1973, becomes the largest outlet for electronic products. This is perhaps the most significant point for the present discussion. The greatest need for future electron devices can be expected in the industrial field. If this conclusion is combined with that drawn from Fig. 1 (i.e., that all small-signal information-handling processes will use solid-state devices), we must accept the fact that the major role of the electron tube lies in power conversion, display, and sensors, particularly optical sensors.

An interpretation of the EIA past sales data for electron tubes[1] is shown by the solid lines of Fig. 3. The dotted curves are extended into the future on the following basis. Receiving-tube sales, according to the first-mentioned conclusion, are reduced at a 9 per cent per year rate; eventually the sales will be largely of replacement tubes. Picture-tube sales extend into the future at a slight increase, which reflects the higher dollar volume in color tubes and ignores the possibility of a solid-state equivalent in the next decade. The applications expected to contribute most to industrial markets are power conversion and light conversion (pickup tubes, displays, and related special types); they are expected to increase demand sufficiently to keep the approximate total tube market at a high level, and perhaps even to raise that level a little.

As stated, our first objective is to find a product mix for the future—and as power conversion is expected to predominate in dollar value, this area should be examined. Fig. 4 shows power-conversion-tube sales by interpretation of EIA data.[1] The military and space programs are the dominant users of high-power microwave tubes for communications (including radar and its sister arts). Thus, with the government sales trend of Fig. 2 in mind, it is seen that the need decreases, and the projection of microwave sales into the future is shown at a decline of one or two per cent per year. Other power tubes used for communications, although less affected by government sales, have limited prospects because of limited broadcast and other channel space, and are shown with no future increase. If the industrial market develops as in Fig. 2, one can readily justify a 12 per cent per year rise in noncommunication power conversion; i.e., high-power tube types that convert low frequencies to high and the converse, or that convert heat to electricity, or that have other related applications. Some of these applications will be discussed later in more detail.

From a summation of the conclusions in Figs. 1 to 4, a table can be prepared to illustrate tube product mix, as it is in 1964 and as it appears it will be two decades hence, in 1984. Most of the assumed trends are extensions of those in the figures, with an important exception. In the display-device and picture-tube application, it has been assumed that, between 1974 and 1984, a combination of reduced color tube costs and the emergence of practical solid-state displays will decrease the picture-tube dollar volume substantially. With some minor rearrangement, then, Table I can be set up.

The data of Table I are striking, if one accepts the predictions. The information-handling application, once the mainstay of the receiving tube, has virtually disappeared, having yielded to solid-state devices. Power generation for communications, in all the low-power and lower-frequency applications, will probably be accomplished by solid-state devices. The higher-power communication applications remain and, as shown, are estimated at the 20 per cent level in 1984. Solid-state devices and lower costs have a reducing effect on display applications. For control and rectifier uses, a major reduction is expected.

In only two instances does the prediction forecast an upward trend in the product mix—the use of light-sensing pickup tubes and of noncommunication power-conversion tubes. The last item includes cooking, industrial heating, chemical processing, thermionic energy conversion, plasma and magnetohydrodynamic power generation, ion propulsion, particle and electron accelerators, microwave power transmission, and controlled fusion plasma apparatus; all of these, in the broad context of this article, are considered to be electron tubes and to use electron-tube technology. These comprise 40 per cent of future markets, and they offer the greatest promise for long-range utilization of tube-making capacity. The writer sees no competition from solid-state devices in these applications.

A word of explanation is in order concerning item 4 of Table I. One may wonder why this application for electron tubes is believed to be relatively immune to replacement by solid-state devices. The key lies in the word "scanning." Sensing and pickup tubes that use optical imaging, plus electron-beam scanning, will be hard to replace. Practical optics often lead to small images. The smallest microelectronic tool available today is the electron beam, and these scanning tubes use it. One could say that such electron tubes are already solid-state devices. Also included in item 4 are other special tubes for coding, image storage, and related applications in which the scanned beam represents the closest approach to a microelectronic-size variable connector. With regard to the increase in product mix for this application (from 6 to 20 per cent), the writer believes that this type of device will become increasingly important in industrial markets, without decreasing in importance in consumer and government markets.

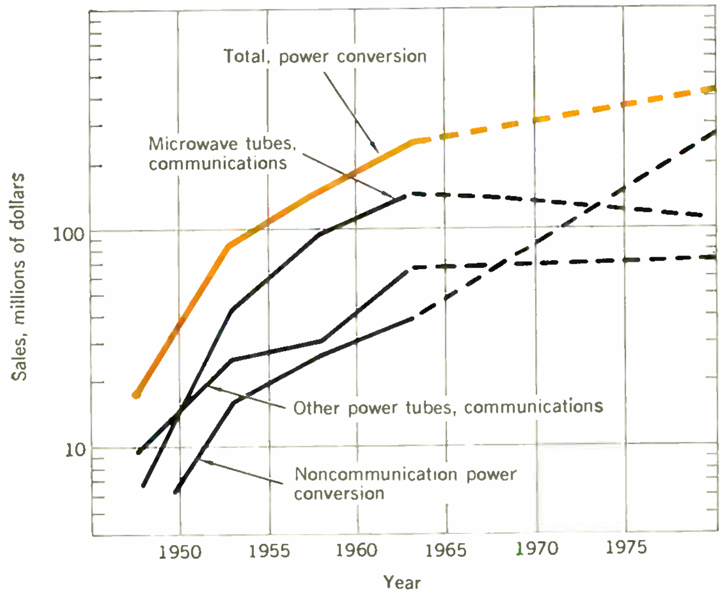


Fig. 4. Power-conversion-tube sales showing ultimate dominance of tubes for noncommunication power-conversion applications. Solid curves are as interpreted from EIA data; dotted curves are predictions.

I. Electron-tube applications

	(per cent of dollar sales) 1964	1984
1. Information handling	36	4
2. Power generation for communications	26	20
3. Display (picture tubes)	22	15
4. Light-sensing and pickup and other beam scanning	6	20
5. Control and rectifier	5	1
6. Noncommunication power conversion	5	40
	100	100

Electron-tube marketing in the future

Based on the changes in applications predicted in Table I, major changes are anticipated in methods of marketing electron-tube products.[3] Although many of these changes are only of indirect concern to the engineer, one is so basic to electron-tube design and performance that it deserves special mention. Our affluent society has found it easy to manufacture products for sale. Much greater difficulty is encountered in selling—in creating the need and promoting obsolescence so that manufacturing can continue at a profitable rate. A principle we have employed in the electron-tube industry is to design and make a salable product; i.e., to manufacture for sale. Our greatest sales and profits then come from a product with fast obsolescence, high volume, and low initial cost. Reliability and low maintenance cost have a smaller effect on salability, perhaps about as much as an attractive appearance or a glamorous name.

However, our society is becoming increasingly aware that its need is for a service rather than for a product.[3] Ownership of a product is useless if the product does not serve its function. Conversely, a function may be provided as a service without ownership by the user. A leading example is the leasing of automobiles, which has grown at a much higher percentage rate than has the sale of automobiles to user-owners. It is possible that the electron tubes of the future may have to be designed, manufactured, and sold to provide a service, particularly in the industrial markets. If they are manufactured for service, rather than for sale, the greatest profits are promoted by a completely different set of criteria. Long life, high reliability, and low operating cost and maintenance are the important factors. The engineer who designs such a product uses entirely different standards than when he designs something for sale.

Fortunately, the government has purchased electron tubes with service in mind, and many parts of the electron-tube industry are strongly oriented toward a highly engineered, extremely reliable, long-lived product. Conversion of such industry to civilian industrial applications can be easily made, if marketing is geared toward not selling the product, but selling the service. Leasing and extended warranty are two older methods that accomplish this end; and many more ingenious marketing developments are anticipated in the future.

For the engineer, such changes in marketing provide him with the opportunity to incorporate his finest ideas in the product. He is relieved in part of the continual frustration of a situation wherein an improvement often cannot be used if it raises the initial cost. Above all, his designs are for optimum use, instead of for optimum salability.

Materials and techniques of the future

As with solid-state devices, the future of the electron tube is strongly, if not entirely, dependent on new or improved materials and techniques. Many of the advances in the industry will result from general materials work, independent of the electron tube, and others will come from semiconductor technology, or from electronically active materials originally prepared for another purpose.

Among the materials under intensive research study are those that produce light—luminescent materials or phosphors. The nature of some of these is poorly understood today; both cathodo-luminescence under electron bombardment and electroluminescence via passage of current in powders are still highly empirical phenomena in most materials. However, injection luminescence using carrier recombination in semiconductors is now a science, and it should not be long before a major advance may be made in application of this knowledge to the electron tube. The result may be better display and picture tubes, with possibly enhanced light output combining an electroluminescent screen, triggered by a scanned electron beam. New techniques may also play a role, and deposition of oriented crystallites may eliminate some of the disadvantages of single-crystal phosphors and randomly oriented microcrystal phosphors.

Similarly, the converse aspect of light conversion, using photosensitive materials, is under continuous study. Although one may not expect as much of a breakthrough as with the phosphor, the improved understanding will be of value in achieving uniformity and reliability.

Single-crystal materials of high perfection should be valuable in electron tubes, particularly for special designs where maximum strength and minimum cross section are desired. However, it does not appear that single-crystal technology will ever play the role in electron-tube development that it did in perfecting the transistor, because so much of the electronic activity in the tube takes place outside the solid material. At the same time, pseudo-single-crystal behavior via grain-oriented polycrystals may be as revolutionary for the electron tube as the single crystal was for the transistor.

In the writer's opinion, the most significant materials technique for future electron tubes lies in the use of chemical vapor deposition (or gas plating). With the use of one or more gaseous compounds of solid materials, it is possible to produce a chemical reaction at a surface, so as to deposit an element or compound on the surface at temperatures well below the melting point of the deposited material.[4,5] Most remarkably, some deposited materials can be deposited either amorphously (with no crystal structure), or epitaxially (following the grain structure of the substrate), or with a polycrystalline random grain orientation, or with polycrystalline grains all oriented in one direction[6,7] (normal to the surface, independent of substrate). Such orientation produces a surface of arbitrary contour, yet with one crystal direction pointed outward everywhere. Hundreds of elements and compounds are available for such deposition. Those that can be controllably grain-oriented can be used to enhance or suppress emission, or to enhance or suppress almost any other surface property; the orientation produces a substantially uniform surface that is otherwise possible only with a planar cut on a single crystal.

Two useful properties of materials for future electron tubes are anisotropy and controlled inhomogeneity. Anisotropy suggests different behaviors in different directions, while controlled inhomogeneity suggests different behaviors in different places. Anisotropic behavior is possible in single crystals and, above all, in the grain-oriented crystals previously discussed. For example, pyrolytic graphite can be so oriented; it has electric and thermal conductivities hundreds of times greater in one direction than in another.[6] Even normally isotropic crystals, when deposited in grain-oriented fashion, are anisotropic for diffusion, since the grain boundaries tend to all run in a parallel direction.[8] Controlled inhomogeneity is the result of a purposeful material design in which the composition and properties are varied from place to place in order to produce better results than obtained with a uniform material.[9] A trivial example is the copper-plating of an Invar microwave cavity to reduce surface resistance while retaining low expansion. Neither anisotropy nor controlled inhomogeneity have yet been used extensively in electron tubes.

Other techniques include new methods of obtaining emission, such as by a laser beam, or by heating the electrons inside a solid, while keeping the solid itself cold. Long life and high reliability are commonly achieved in expensive tubes by a small appendage vacuum pump of the getter-ion type. In the future, such pumping action may well be designed as a basic part of the working structure, thereby simplifying operation and power supplies. Widespread use of certain gas discharges in tubes will require a better approach to elimination of gas cleanup; this is an area in which research is badly needed and the writer does not know of a solution now in the research stage. It may be expected that solidly embedded heaters inside cathodes, which use conduction rather than radiation heating, will supplant present types and improve ruggedness, reliability, and life. It is likely that ferrites, ferroelectrics, piezoelectrics, and semiconductors all may be found useful some day for the inside of an electron tube.

Noncommunication power conversion

The anticipated future importance of the electron tube in the field of power conversion is enough to justify a more detailed discussion. The type of power conversion that is not useful for communication is often "low grade" in comparison. Frequency and amplitude stability, low noise, and phase linearity, which are part of communications power generation, are rarely needed in other fields. Electron tubes for industrial power conversion, therefore, may be quite unusual. As an example, if someone were to devise an efficient plasma microwave generator with the simplicity of a fluorescent lamp, it would probably have little use in communications, but might revolutionize microwave cooking and heating.

There is no *a priori* reason why the low operating cost of thermal ovens for food preparation cannot be equalled by some form of microwave device. However, as already learned by those in the field, one must sell the service, not the tube; we are still a long way from either the innovation in the electron tube or the innovation in marketing that will make this form of cooking universal. Perhaps the same thing is true of microwave heating in industrial processes. However, in the latter instance there are applications in which no other process will do, and even present-day equipment will serve.

Thermionic energy conversion[10]—direct conversion of heat to electricity—is an intriguing concept, and there is nothing on the solid-state horizon to equal its potential. The reason is simple. The thermionic converter is like the thermoelectric couple, except that (1) the hot side is well insulated from the cold side, and (2) the hot side may be run at temperatures well above the melting points of the good thermoelectrics. Both differences make the thermionic device capable of much higher efficiency than is possible with thermoelectrics. This type of conversion is particularly promising for nuclear reactor use, in which core temperatures of the future may exceed any now common in power plants. Only a modest advance in today's conversion results would suffice to make the thermionic converter adequately efficient.[11] However, greater advances in materials and tube technology are needed to achieve the reliability and low operating cost that would be necessary for central station use.

For conversion of heat to electricity, power generation by means of magnetohydrodynamics, or other methods using high-speed plasma ejection, are definitely a part of the future.[10] Whether or not this technique should be included in a discussion of electron tubes is a moot point. The problems of electrodes, electrode erosion, ion production, and cooling of parts certainly resemble problems that also arise in gas discharge tubes. At the present time, solutions of some of these problems, on a scale needed for megawatt powers and under low operating cost conditions, seem far off indeed. One may hope, however, that electron-tube engineers can offer solutions based on tube experience that may improve on those proposed by the plasma physicist.

Particle accelerators of all kinds are now important in research, but some forms soon may be a major part of industrial electronic equipment.[12] Electron linear accelerators are in common use for X-ray radiography and cancer therapy, and will continue to find experimental use in food irradiation, plastics manufacture, etc.[13] This type of accelerator uses the microwave electron tube as a power source and is, itself, a microwave tube of a special kind. If these relatively experimental uses develop into practical embodiments, the electron-tube engineer will find himself with much to do, and reduction of overall usage cost (which is now very high) will be a major requirement.

Another form of particle accelerator that has some potential for future use is the ion propulsion device.[14] Devised by rocket physicists, ion propulsion made best progress only after it attracted the attention of tube engineers. The ion engine is really not a different problem for the tube industry. Unfortunately, its use awaits development of a suitable low-weight power supply for space travel, and this is much more difficult.

The gas laser, although currently under consideration chiefly for communications, may some day have important conversion uses at higher levels. It is, after all, an electron tube, and in practical form will make greater use of tube-making facilities and techniques. It should certainly be on the list of potential future applications.

Finally, power transmission in space, via microwaves,[15] and controlled-fusion apparatus[16] are suggested; both are possible applications for electron-tube technology, but both seem decades away, so far as extensive use is concerned.

This article is based in part on papers presented at the 7th Annual Tube Techniques Conference, New York, N.Y., September 28, 1964 and at the IEEE Electron Devices Group Annual Meeting, Washington, D.C., October 19, 1964.

REFERENCES

1. Electronic Industries Association Yearbook (1964 Edition), Washington, D. C.

2. Dulberger, L. H., Strasser, J. A., and Carroll, J. M., "Electronic Markets," *Electronics*, vol. 37, no. 1, Jan. 3, 1964, pp. 53–56.

3. Garretson, R. C., and Mauser, F. F., "The Future Challenges Marketing," *Harvard Business Rev.*, vol. 41, Nov.–Dec. 1963, pp. 168–188.

4. Powell, C. F., Campbell, I. E., and Gonser, B. W., *Vapor Plating*. New York: John Wiley & Sons, Inc., 1955.

5. Bakish, R., "Vapor-Phase Metallurgy," *Intern. Sci. and Tech.*, vol. 1, no. 18, June 1963, pp. 54–60.

6. Pappis, J., and Blum, S. L., "Properties of Pyrolytic Graphite," *J. American Ceram. Soc.*, vol. 44, Dec. 1961, pp. 592–597.

7. Weissman, I., and Kinter, M. L., "Improved Thermionic Emitter Using Uniaxially Oriented Tungsten," *J. Appl. Phys.*, vol. 34, Nov. 1963, pp. 3187–3194.

8. Weissman, I., "An Improved Thermionic Emitter: Thorium on Grain-Oriented Polycrystalline Tungsten," *Ibid.*, Jan. 1965.

9. Herold, E. W., "Future Circuit Aspects of Solid-State Phenomena," *Proc. IRE*, vol. 45, Nov. 1957, pp. 1463–1474.

10. Kaye, J., and Welsh, J. A., eds., *Direct Conversion of Heat to Electricity*. New York: John Wiley and Sons, Inc., 1960.

11. Wilson, V. C., "Thermionic Power Generation," *IEEE Spectrum*, vol. 1, May 1964, pp. 74–83.

12. Cooper, D. I., "Radiation Processing," *Intern. Sci. Tech.*, vol. 1, no. 27, March 1964, pp. 22–34.

13. Miller, C. W., "The Linear Accelerator as a Power Source of Irradiation Processing," in *Radiation Sources*, Charlesby, A., ed. New York: Pergamon/Macmillan Co., 1964, chap. 6.

14. Langmuir, D. B., Stuhlinger, E., and Sellen, J. M., eds., *Electrostatic Propulsion*. New York: Academic Press, 1961.

15. Okress, E. C., Brown, W. C., Moreno, T., Goubau, G., George, R. H., and Heenan, N., "Microwave Power Engineering," *IEEE Spectrum*, vol. 1, no. 10, Oct. 1964, pp. 76–100.

16. Rose, D. J., and Clark, M., *Plasmas and Controlled Fusion*. Cambridge, Mass.: M.I.T. Press, 1961.

The MAC system: the computer utility approach

The broad system goal of Project MAC may be regarded as the development and operation of a community utility that is capable of supplying computer power to each "customer" where, when, and in the amount needed

R. M. Fano *Massachusetts Institute of Technology*

A major program devoted to research efforts on advanced computer systems and the exploitation of these systems is currently under way at the Massachusetts Institute of Technology. This work is being carried out under the project name "MAC," an acronym derived from two different, but appropriate, terms: "machine-aided cognition," and "multiple-access computer." The first of these terms could be considered as expressing the broad objective of the project, and the second as describing its principal tool.

The notion of machine-aided cognition implies an intimate collaboration between a human user and a computer in a real-time dialogue on the solution of a problem, in which the two parties contribute their best capabilities. In order for this intimate collaboration to be possible, a computer system is needed that can serve simultaneously a large number of people and that is easily accessible to them, both physically and intellectually. The present MAC system is a first step toward this goal.

The case for machine-aided cognition with the aid of a multiple-access computer system was very eloquently argued by Professor John McCarthy in a 1961 lecture.[1] The views he presented were largely the consensus of a committee that had just completed a comprehensive study of the future computational needs of M.I.T. These concepts were also embodied in the Compatible Time-Sharing System (CTSS),[2] developed by the M.I.T. Computation Center under the leadership of Professor F. J. Corbató; an early version of this system was first demonstrated in November 1961.[3] The MAC computer system is a direct descendant of CTSS.

The last section of McCarthy's lecture introduced the notion of a community utility capable of supplying computer power to each "customer" where, when, and in the amount needed. Such a utility would be in some way analogous to an electrical distribution system; that is, it could provide each individual with logical tools to aid him in his intellectual work, just as electric tools today aid him in his physical work. In this regard, it might be said that the present state of the computer as a source of logical power is similar to that of the early steam engine as a source of mechanical power. The steam engine could generate much more power than could any man or animal, and therefore could perform tasks that were previously impossible. However, the power generated could not be supplied on an individual basis to aid men in their daily work until after the advent of electric power distribution.

The analogy between electric power and computer power illustrates only one of the aspects of a computer utility—namely, its ability to provide the equivalent of a

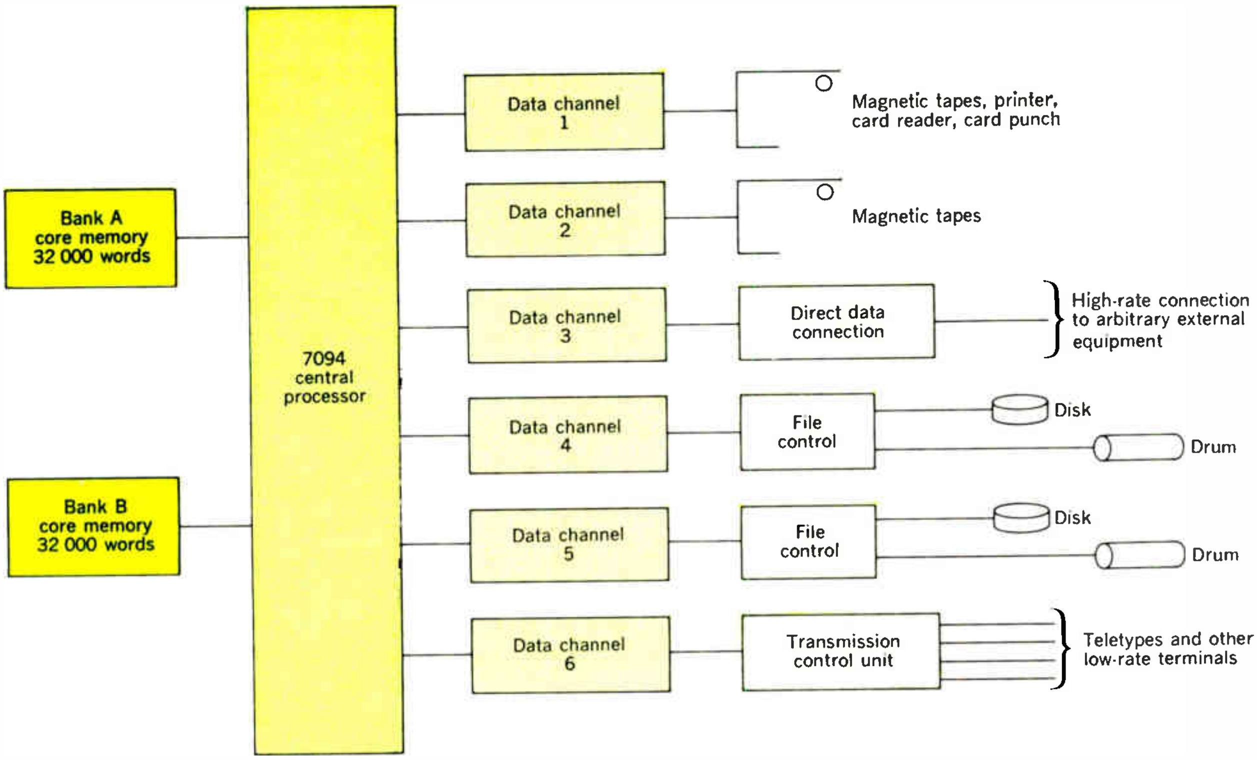


Fig. 1. Block diagram showing equipment configuration of MAC computer installation.

private computer whose capacity is adjustable to individual needs. Of much greater importance to the individual customer would be the benefits that such a utility could make available to him, by placing at his fingertips a great variety of services in the form of public procedures, data, and programming aids, and by allowing him to store and retrieve his own private files of data and programs. Furthermore, a computer utility could provide customers having common interests with convenient means for collaboration. For instance, designers who are working together on a complex system would be in a position to check continually the status of the complete, overall design as each of them develops and modifies his own contribution.

The MAC system is an experimental computer utility which, since November 1963, has been serving a small but varied segment of the M.I.T. community. It is the most extensive of several time-sharing systems in operation.[4–6] In spite of its experimental character and its limited capabilities, it has gained quick acceptance as a daily working tool.

The purpose of this article is to present a somewhat brief description of the current system, to report on the experience gained from its operation, and to indicate directions in which future developments are likely to proceed. The scope is limited to the general organization of the system and to the basic services it can provide. In particular, no mention will be made of the problem-oriented languages and other special programming facilities that are being added to the system by the system users themselves, thereby increasing its intellectual accessibility. Thus, it should be evident that the work reported here represents only a relatively small portion of the overall research effort that is encompassed by Project MAC.

Equipment configuration

The primary terminals of the MAC system are, at present, 52 model 35 Teletypes and 56 IBM 1050 Selectric teletypewriters (adaptations of the "golf-ball" office typewriter). These are located mostly, but not exclusively, within the M.I.T. campus. Each of these terminals can dial, through the M.I.T. private branch exchange, either the IBM 7094 installation of Project MAC or the similar installation of the M.I.T. Computation Center. The supervisory programs of the two computer installations may, independently, accept or reject a call, depending on the identity of the caller. Access to the MAC system can also be gained from any station of the Telex or TWX telegraph networks. Some tests and demonstrations already have been conducted from European locations, and experiments are now being planned in collaboration with a number of universities for the purpose of providing further experience with long-distance operation of the system.

Although Teletypes and other typewriterlike terminals are adequate for many purposes, some applications demand a much more flexible form of graphical communication. The MAC system includes for this purpose the initial model of a multiple-display system developed by the M.I.T. Electronic Systems Laboratory for computer-aided design.[7] The system includes two oscilloscope displays with character generator and light pen, as well as some special-purpose digital equipment that performs the light-pen tracking and simplifies the task of the computer in maintaining the display and in performing common operations, such as translating and rotating the display. The two oscilloscopes can be operated independently of each other; communication with the computer can be achieved by means of the light pen, and also through a variety of other devices, such as knobs, push buttons, toggle switches, and a typewriter. The meaning of a signal from one of these input devices is entirely under program control.

The whole display system communicates with the MAC installation's IBM 7094 equipment through the direct-data channel, and the display data are stored in the central memory of the 7094. Thus, the display must be located in a room adjacent to the computer installation. Remote operation would require the addition of a memory and some processing capacity for local maintenance of the display.

A separate, very flexible display terminal is provided by a DEC PDP-1 computer, which can communicate from a remote location with the MAC computer installation through a 1200-bit-per-second telephone connection. The PDP-1 can also be used as a buffer between the MAC computer and the display system just described, thereby permitting simulation and study of remote operation of the latter.

All of these terminals can, in principle, operate simultaneously and independently of the MAC computer installation by time sharing its central processor. However, in order to insure prompt response, the number of terminals active at a given time is limited by the supervisory program to 24. This number has already grown to 24 from 10, and is expected to grow further and possibly to double in the next few months, although maximization of this number is not considered to be a primary objective at this time.

The equipment configuration of the MAC computer installation is illustrated in Fig. 1. The IBM 7094 central processor has been modified to operate with two 32 000-word banks of core memory and to provide facilities for memory protection and relocation. These features, together with an interrupt clock and a special operating mode (in which input–output operations and certain other instructions result in traps), were necessary to assure successful operation of independent programs coexisting in core memory. One of the memory banks is available to the users' programs; the other is reserved for the time-sharing system supervisory program. The second bank was added to avoid imposing severe memory restrictions on users because of the large supervisory program, as well as to permit the use of existing utility programs (compilers, etc.), many of which require all or most of a memory bank.

The central processor is equipped with six data channels, two of which are used as interfaces to conventional peripheral equipment, such as magnetic tapes, printers, card readers, and card punches. A third data channel provides direct data connection to terminals that require high-rate transfer of data, such as the special display system.

Each of the next two data channels provides communication with a disk file and a drum. The storage capacity of each disk file is 9 million computer words, and the capacity of each drum is 185 000 words. The time required to transfer 32 000 words in or out of core memory is approximately two seconds for the disk file and one second for the drum. The two disk files, with a total capacity of 18 million words, are used to store the users' private files of data and programs, as well as public programs, compilers, etc. The two drums are used for temporary storage of active programs that have to be moved out of the core memory to make room for other programs. Thus, in this respect they act as an extension of the core memory. Drums with transfer rates that are four times faster will be substituted for the present ones in the near future.

The transmission control unit (IBM 7750) consists of a stored-program computer, which serves as an interface between the sixth data channel, and up to 112 communication terminals capable of telegraph-rate operation (approximately 100 bits per second). An appropriate number of these terminals are connected by trunk lines to the M.I.T. private branch exchange and to the TWX and Telex networks. Higher-rate terminals can be readily substituted for groups of these low-rate terminals; for instance, on the present MAC system, three 1200-bit-per-second terminals are installed, one of which provides the communication channel to the PDP-1 computer. All of these terminals are compatible with Bell System Data-Phone data sets. Part of the core memory of the transmission control unit is used as an output buffer, because the supervisory program and its necessary buffer space have grown in size to the point of occupying the whole of the *A* bank of core memory. The design intent was and still is to provide sufficient input–output buffer space in the main memory to eliminate unnecessary core-to-core transfers; the present mode of operation is actually a makeshift that has been necessitated by equipment limitations.

The Digital Equipment Corporation's PDP-1 computer is not an integral part of the MAC time-sharing system, except when connected to it as indicated above. It was acquired to permit early experimentation with light-pen interaction with a display, and for other very-high-speed interaction work. It includes a 16 000-word core memory, microtapes, a high-speed channel, and a scope display with character generator and light pen. It is anticipated that this equipment will be replaced by a PDP-6 computer in the near future.

Operating program

The operating program of the MAC computer system is a direct descendant of CTSS. Many parts of the operating program have since been rewritten to facilitate system maintenance, and various facilities have been added that had been described in the CTSS manual[2] but had not then been implemented. Furthermore, it now includes various user-developed features such as the input–output adapter for the display system described in the preceding section, compilers for various new languages, and other programming aids.

The heart of the MAC system is the supervisory program, which resides at all times in the *A* bank of core memory. This program handles the communication with all the terminals, schedules the time sharing of the central processor on the part of active programs, moves these programs in and out of core memory, and performs a variety of bookkeeping functions necessary to protect users' files and to maintain detailed accounting of the system usage.

The services that the system is capable of providing are organized in the form of commands—that is, instructions that system users can give to the system. The programs corresponding to the commands are permanently stored in the disk files; when a command is issued, a copy of the corresponding program is made, loaded, and executed just as if it were a user's program. The language facilities available in the system include FAP, MAD, MADTRAN (a translator of FORTRAN into MAD), COMIT, LISP, SNOBOL, a limited version of ALGOL, on-line, extended versions of SLIP and GPSS, and two problem-oriented languages named COGO and STRESS, developed for civil engineering applications.

The system is rapidly evolving through the addition of new language facilities and other utility programs and programming aids. The operating program itself is now being modified by system programmers working on line, and modifications are occasionally introduced without even interrupting the operation of the system. The entire system, exclusive of users' files, includes approximately half a million words of code, of which a little more than 50 000 were specifically written for the system and the rest consist of adaptations and modifications of compilers and other programming aids previously developed at M.I.T. and elsewhere.

Some of the basic services available from the system are illustrated in the six parts of Fig. 2, which represent a complete record of a demonstration session held on June 9, 1964, between 10:36 and 11:39 a.m. The total computer time used was 1.3 minutes. The author's typing appears in lower-case letters, and the computer replies are in capital letters. All digits represent computer replies except those intermixed with lower-case text, and the two lines following "type range n1. to n2. . ." in Fig. 2(E). Each command is immediately acknowledged by the computer with a character W, followed by the time of the day in which the first two digits are hours, the next two are minutes, and the one following the period indicates tenths of a minute. The end of an interaction is indicated by the letter R, followed by the sum of the two numbers indicating the total number of seconds of computer time used during the whole interaction. The first of the two numbers indicates the processing time, and the second represents the time that is wasted in the swapping of programs back and forth between the core memory and the drum.

The session was started by issuing the *login* command, followed by the author's problem number and name. The computer then asked for his password, which is not recorded because the printing mechanism of the typewriter is automatically disconnected while the password is typed. All the lines assigned to the author's user group were already in use; therefore, a stand-by line was assigned because the total number of lines in use was less than 24. This condition introduced the possibility of an automatic logout of the author's problem. The rest of Fig. 2(C) contains various information, including the allotment of computer time in minutes, which had been made that very morning and of which none had yet been used.

Figures 2(A) and (F) illustrate some of the facilities for writing, editing, filing, and compiling programs, and for retrieving and printing private files. The *input* command causes the computer to type out successive line numbers as each statement of the program is written. The program in Fig. 2(A), a corrected version of which is printed out in Fig. 2(F), can be used to compute prime numbers as illustrated in Fig. 2(E). The program is written in the MAD language. Some of the typing errors in Fig. 2(A) were intentional, others accidental. The quotation mark erases the preceding character and itself; thus, two successive quotation marks erase the preceding two characters. The question mark erases the entire line up to that point. Each line is terminated by giving a carriage return. A carriage return (i.e., entering a null line) is also used to enter the manual mode of input, as illustrated in line 230. In the manual mode, preceding lines can be deleted or corrected by issuing appropriate commands. After the deletion of two superfluous lines, the program was filed under the name PRIME MAD as part of the author's private file.

Next, the *mad* command was issued to cause the program to be compiled by the MAD translator. The first attempt at compilation failed because of an error for which a diagnostic was printed. The error consisted of the omission of the word *print* in line 130. The *edit* command was then used to reopen the program file, the error was corrected in the manual mode, and the program was refiled under the same name. The second attempt to compile the program was successful, as indicated in Fig. 2(F), and the corrected program was then printed out through use of the *printf* command.

The binary version of the PRIME program was then loaded (together with the necessary library subroutines) by use of the *load* command, and the resulting core image and machine state were recorded by use of the *save* command. The latter command created a new file named PRIME SAVED, as indicated by the system reply to the command *listf prime saved*. The program was then started by issuing the command *resume prime*. It could also have been started by issuing the command *star prime* instead of *save prime*, or by loading and starting the program in one operation by means of the command *loadgo prime*.

The PRIME program asked for the numbers n1 and n2 defining the desired range of prime numbers; the numbers 1 000 000 and 1 001 000 were given. The typing out of the prime numbers was interrupted by pressing twice the break button on the Teletype, which resulted in the system's replying with the word QUIT.

The command *listf* without arguments causes the system to list the contents of the user's own private file directory as illustrated in Fig. 2(B). There are four items with PRIME as first name: PRIME MAD, the original program shown in Fig. 2(F); PRIME BSS, its translation in machine language; PRIME MADTAB, a table (storage map) useful for debugging purposes; and PRIME SAVED, the complete machine state after loading the program. The item having the name PRIMES MAD is a slightly different version of PRIME MAD.

Items are deleted from the file by issuing the *delete* command followed by the names of the items, as shown also in Fig. 2(B). The command *delete*, with an argument consisting of an asterisk followed by a space followed by a name, causes all items with the stated name as a second name to be deleted. The result of issuing the two *delete* commands shown in Fig. 2(B) was the elimination from the file of all of the items except for PRIME MAD and FILTER SAVED, as indicated in the following printout of the file directory.

```
input
W 1038.3
00010              print format text,$ $
00020              print format text, $type range n1. to n2. on 2 lines $
00030              reat"d format b,fn1
00040              read format b, fn2
00050              n1=fn1
00060              n2=fn2
00070              print format text,$primes are$
00080              through loopb, for n=n1,1,n.g.n2
00090              whenebe n.l.3?whenev?          whenever n.l.3, transfer to print
00100              throught loopa,for 1=2,1,l.g.(n/l)
00110              through loopa,for l=2,1,l.g.(n/l)
00120              "loopa    whenever (n-(n/l)*l).e.0,transfer to loopb
00130              print format a,n
00140 loopb        continue
00150              print format text,$range finished$
00160              execute exit
00170              execute dormant""""nt.
00180              vector values a=$(i8)$
00190              vector values b=$(f20.8)$
00200              integer n,l,?          vector values text=$(12a6)$
00210              integer n,l,n1,n2
00220              end of program
00230
 MAN. delete,100
 MAN. delete,160
 MAN. file  prime mad
W 1057.6
R .800+3.616

mad prime
W 1058.4

 THE FOLLOWING NAMES HAVE OCCURRED ONLY ONCE IN THIS PROGRAM AND ARE ALL
 ASSIGNED TO THE SAME LOCATION. COMPILATION WILL CONTINUE.

PRINT

ERROR 02051 IN PART3. CARD NO.   00090
INVALID MODE FOR SOME OPERAND IN PRECEDING STATEMENT
 NO TRANSLATION.
R .600+.833

edit prime mad
W 1059.2
00230
 MAN. 130 print     print format a h""""a,n
 MAN. file prime mad
W 1101.1
R 1.066+1.200
```

A

B

```
listf
W 1118.5
   10 FILES    61 TRACKS USED
DATE             NAME          MODE   NO. TRACKS
 6/09/64     PRIME   SAVED      P          17
 6/09/64     PRIME     BSS      P           1
 6/09/64     PRIME  MADTAB      P           1
 6/09/64     PRIME     MAD      P           1
 6/08/64     MON04   SAVED      P          12
 6/08/64    PRIMES     MAD      P           1
 6/05/64    FILTER   SAVED      P          19
 5/12/64     GSUBA     BSS      P           3
 5/12/64     SCANA     BSS      P           3
 5/08/64     SCANT     BSS      P           2
R .616+.800


delete prime saved prime bss prime madtab primes mad mon04 saved
W 1122.9
R 3.200+.400

delete *bss"""" bss
W 1123.7
R 1.666+.200

listf
W 1123.9
    2 FILES    21 TRACKS USED
DATE             NAME          MODE   NO. TRACKS
 6/09/64     PRIME     MAD      P           1
 6/05/64    FILTER   SAVED      P          19
R .616+.616
```

C

```
login t193 fano
W 1036.4
PASSWORD
 PARTY LINE BUSY, STANDBY LINE HAS BEEN ASSIGNED
 T0193  2859 LOGGED IN  06/09 /64    1036.7
 CTSS BEING USED IS       MAC05K
SHIFT           MINUTES
      ALLOTTED      USED SINCE 06/09/64 1036.7
 1       100          0.0
 2       100          0.0
 3       100          0.0
 4       100          0.0
LAST LOGOUT WAS 06/09/64   1036.7
TRACK QUOTA=          P,       500 Q. 0041 TRACKS USED.
R 5.550+1.000
```

Fig. 2. Six printouts from a demonstration session held at M.I.T. on June 9, 1964.

D

```
copy p mon04 saved
W 1125.9
*R 2.616+.400

resume mon04 1
W 1126.1

 CTSS UP AT  902.8 06/09/64.

  NUSERS= 22    TIME= 1126.2
  29.7   29.7   BACKGROUND,
  46.6   46.6   FOREGROUND,
  17.6   17.6   SWAP TIME,
   6.2    6.2   LOAD TIME,
  15.5   15.5   USER WAIT,
   6.6    6.6   SWAP WAIT,
   3.9    3.9   LOAD WAIT.

  NUSERS= 23    TIME= 1127.2
  19.8   29.6   BACKGROUND,
  43.6   46.5   FOREGROUND,
  27.3   17.7   SWAP TIME,
   9.3    6.2   LOAD TIME,
  25.5   15.6   USER WAIT,
  10.3    6.7   SWAP WAIT,
   5.5    3.9   LOAD WAIT.
 QUIT,
R .000+2.616

delete mon04 saved
W 1127.9
R 1.000+.400


logout
W 1132.9
 T0193  2859 LOGGED OUT 06/09 /64    1139.1
 TOTAL TIME USED=   01.3 MIN.
```

A program for monitoring the system operation, by the name of MON04 SAVED, is available in the public file, accessible to all system users. This program must first be copied into one's own private file by means of the *copy p* command, as shown in Fig. 2(D). Since this program is stored as a record of a previous machine state, it is started by issuing the *resume* command. A second argument in the command (the digit 1) causes the program to be run at one-minute intervals. At time 1126.2 there were 22 active users; one minute later there were 23. The other figures are percentages of the computer time devoted to various operations.

The word FOREGROUND refers to the computer time devoted to running programs requested by on-line users. BACKGROUND refers to the computer time available for running normal batch processing, which takes place automatically when no service is requested by on-line users. SWAP TIME is the processor time wasted in moving a program from core memory to drums, and vice versa. LOAD TIME is the processor time devoted to loading programs from the disk files into core memory.

The first four figures in each of the four columns of percentages in Fig. 2(D) add to 100 per cent (or approximately so). The last three figures refer to the times in which the processor is totally idle while input–output operations are taking place. The USER WAIT is part of the foreground, and it is already included in the foreground figure. Similarly, SWAP WAIT and LOAD WAIT are already included in swap time and load time. The figures in the second column of each table are percentages evaluated from the time at which the system was last placed in operation; for example, 902.8 in Fig. 2(D). The figures in the first column are percentages over the last one-minute interval; in the first table they are identical to those in the second column.

After the monitor program was deleted from the file, the session was terminated by issuing the *logout* command. The session lasted approximately one hour, and was interrupted by two telephone calls lasting for a total of approximately 15 minutes. The total computer time used—with both swap and load time included—was 1.3 minutes.

The record of the demonstration session illustrates a few of the most basic commands of the MAC system. The total number of commands available to all users is, at present, 68. This number is continually increasing as new facilities, developed by users as well as by the system programmers, are added to the system. In addition, a variety of programs of lesser general interest are available in the public file, from which they can be copied into private files. This same public file is also used to facilitate the transfer of programs and data between system users. Among the special facilities available for operation of the display system are provisions for displaying text, for drawing on the screen, and for rotating three-dimensional objects. The public part of the system consists at present of approximately half a million words of code. The users' private files barely fit into the two disk storage units, the total capacity of which is approximately 18 million words.

Operating experience

The MAC system has been operating in roughly its present form since the middle of November 1963. It is now in operation 24 hours a day, 7 days a week. Maintenance, disk dumping and loading, and occasional nontime-sharing operation require approximately four hours per day. The on-line use of the system steadily increased from November to April; the total number of computer hours charged to on-line users (the sum of the numbers printed out by the system on completion of each command) was 311 in April and 297 in May. In other words, the computer time devoted to serving on-line users amounted to approximately 42 per cent of total clock hours. The background use is not included in these figures. The total number of user-hours between logins and logouts turns out to be approximately 17 times the number of computer hours used.

E

```
load prime
W 1111.1
R 3.100+1.200

save prime
W 1111.4
R .866+.600

listf prime saved
W 1111.7
 6/09/64     PRIME   SAVED    P          17
R .650+1.416

resume prime
W 1112.0
#EXECUTION.

TYPE RANGE N1. TO N2. ON 2 LINES
1000000.
1001000.
PRIMES ARE
 1000003
 1000033
 1000037
 1000039
 1000081
 1000099
 1000117
 1000121
 1000133
 1000151
 1000159
 1000171
 1000183
 1000187
 1000193
 1000[illegible]
 1000[illegible]
 1000- QUIT,
R [illegible].400+2.216
```

F

```
mad prime
W 1105.4
LENGTH 00205, T.V. SIZE 00005, ENTRY 00043
R 2.400+1.000

printf prime mad
W 1106.0
00010              PRINT FORMAT TEXT,$ $
00020              PRINT FORMAT TEXT, $TYPE RANGE N1. TO N2. ON 2 LINES $
00030              READ FORMAT B,FN1
00040              READ FORMAT B, FN2
00050              N1=FN1
00050              N2=FN2
00070              PRINT FORMAT TEXT,$PRIMES ARE$
00080              THROUGH LOOPB, FOR N=N1,1,N.G.N2
00090              WHENEVER N.L.3, TRANSFER TO PRINT
00110              THROUGH LOOPA,FOR I=2,1,I.G.(N/I)
00120 LOOPA        WHENEVER (N-(N/I)*I).E.0,TRANSFER TO LOOPB
00130 PRINT        PRINT FORMAT A,N
00140 LOOPB        CONTINUE
00150              PRINT FORMAT TEXT,$RANGE FINISHED$
00170              EXECUTE DORMNT.
00180              VECTOR VALUES A=$(I8)$
00190              VECTOR VALUES B=$(F20.8)$
00200              VECTOR VALUES TEXT=$(12A6)$
00210              INTEGER N,I,N1,N2
00220              END OF PROGRAM
R .416+1.016
```

The system is usually fully loaded (24 on-line users) during the day, and almost fully loaded in the evening until midnight, and sometimes later. The system is very seldom idle, even in the early morning hours.

Facilities for detailed monitoring of system operation have become available only very recently, and therefore no dependable data can be presented at this time. It must be stressed in this regard that it is far from obvious what measurements would provide a useful characterization of the system performance, in view of the many variables involved, and of the complexity of their interactions. Furthermore, the frequency and character of user requests vary with time; in addition, they are highly unpredictable in nature, and of course cannot and should not be restricted or controlled in any realistic measurement of system performance.

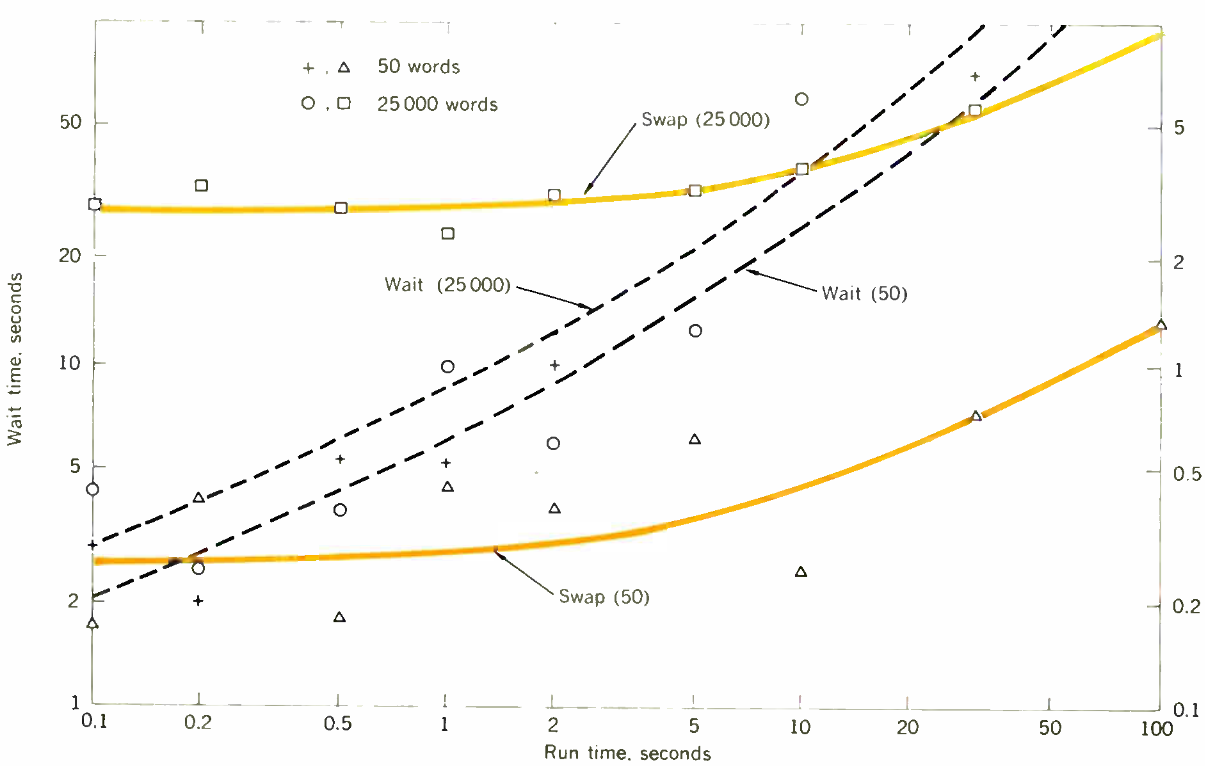
The performance figure of greatest interest to the user is the response time (the time interval between the issue of a request and the completion on the part of the computer of the requested task) as a function of the bare running time of the corresponding program. The response time depends on the scheduling algorithm employed by the system, as well as on the number and character of the requests issued by other users.

The scheduling algorithm operates roughly as follows. Each user request is assigned an initial priority, which depends only on the size of the program that must be run. The highest priority is assigned to the smallest programs. The highest-priority programs are allowed to run for a maximum of 4 seconds before being interrupted, whereas lower-priority programs are allowed to run for longer intervals, which are multiples of 4 seconds. The lower the priority, the longer is the allowed interval. If a program run is not completed within the allowed interval, the program is transferred from core memory to drum (the state of the machine being automatically preserved), and its priority is reduced by one unit.

The curves of Fig. 3 illustrate the behavior of the wait time and swap time as a function of program run time for programs 50 words long and 25 000 words long. The swap time is defined for this purpose as the time spent in transferring the program between disk file and drum on the one hand, and core memory on the other; wait time is the time during which the computer is performing tasks that are not pertinent to that program. Obviously, the total response time is the sum of the bare run time (the abscissa in Fig. 3), the swap time, and the wait time. The points in Fig. 3 were obtained from measurements performed by M. Jones, and the curves are rough interpolations between the points. Each point represents the average value of 30 consecutive measurements. The number of users during the time that the measurements were being made varied from a low of 13 to a high of 21; it was between 17 and 20 most of the time. The scattering of the points in Fig. 3, together with the variability of the number of users while the measurements were being performed, should make clear the kind of difficulties one faces in obtaining precise and accurate measurements of system performance.

It is clear from Fig. 3 that the swap time constitutes a large overhead for short runs. In this figure, the swap time is defined to include also the initial transfer of the program from disk file to core memory and the final transfer back into the disk file; the two-way transfer of 32 000 words between disk file and core memory takes approximately 4 seconds, and the same transfer between drum and core memory takes approximately 2 seconds. The wait time increases less than linearly with run time,

Fig. 3. Behavior of wait time and swap time as functions of run time.



and, as expected, is not greatly affected by the size of the program.

The MAC computer installation has experienced a normal number of hardware failures. It should be realized however, that hardware failures are much harder to diagnose in a multiple-access system because of the impossibility of reproducing the machine conditions at the time of failure. Moreover, the probability distribution of machine states in a multiple-access operation is quite different from that in a batch-processing operation; therefore, hardware troubles that become apparent in the former mode of operation often go unnoticed in the latter. Above all, it is often extremely difficult to determine whether a particular malfunctioning is the result of a hardware failure, a system-programming error, or even an error in the logical design of the machine. There is no doubt that multiple-access systems present substantially more difficult maintenance problems than do conventional systems.

The most delicate aspect of the operation of a multiple-access system of the MAC type is the responsibility assumed by the system managers with respect to the users' programs and data that are permanently stored in the disk files. Elaborate precautions must be taken to protect the contents of the disk files against malfunctioning of the system, as well as against actions of individual users. The supervisory program restricts the access of each user to his own private file and to public files, which he cannot modify. A full copy of the contents of the disk file is made twice a day to minimize the loss in case of malfunctioning. While losses of users' programs and data have occurred, their frequency and seriousness have not discouraged users from entrusting their work to the system.

The system users number a little more than 200, with ten academic departments represented among them. Although most users have had previous experience in programming, there is a growing group of users who are working entirely with programs developed by others or through high-level problem-oriented languages that enable them to communicate with the system in an English-like language so set up that it is appropriate to their professional field. As stated previously, detailed descriptions of these special language facilities designed for use in specialized fields are considered to be outside the scope of this article; however, readers interested in pursuing the subject further may find additional information in other published material.[8–11]

Enthusiasm mixed with a great deal of frustration is the most common reaction to the system on the part of its users. The system was very quickly accepted as a daily working tool, particularly by computer specialists. This quick acceptance, however, was accompanied by the kind of impatience with failures and shortcomings that is characteristic of customers of a public utility. The capacity of the system is limited, and therefore users are often unable to log in because the system is already fully loaded. Furthermore, the system may not be in operation because of equipment or programming failures just when one is planning to use it. In other words, the system is far from being as reliable and dependable as one might expect a utility to be.

Experience since last November, however, has shown that it is perfectly feasible for a computer system to be the object of research and development for some people and, simultaneously, an effective working tool for others. System experimentation and use are not only compatible but mutually beneficial.

It is becoming increasingly evident that the system's ability to provide the equivalent accessibility of a private computer is only a secondary, although necessary, characteristic. What users find most helpful is the fact that the system places literally at their fingertips a great variety of services for writing, debugging, and compiling programs, and facilities for working on problems in their own particular fields through the use of appropriate problem-oriented languages.

The system users themselves are beginning to contribute to the system in a substantial way by "publishing" their work in the form of new commands. As a matter of fact, an editorial board is being established to review such work and formally approve its inclusion in the system. Thus, the system is beginning to become the repository of the procedural and data knowledge of the community that it serves. A corollary of this trend is, of course, the increasing difficulty that users find in ascertaining just which facilities of interest to them are included in the system. In other words, it can be said that the system begins to take on the undesirable as well as the desirable characteristics of a library.

System trends

The organization of the MAC system appears to be at the threshold between two basically different points of view on computer systems and their utilization. The traditional view is that of a processor serving one user at a time and executing programs in succession, with a negligible amount of interaction during execution with the user himself or any other part of the outside world. A corollary of this view is that the processor, the memory, and the peripheral equipment must be designed to fit the requirements of the "typical user" rather than the average requirements of users, considered as a group. Therefore, the system as a whole can be used efficiently only if programs are tailored specifically to its peculiarities.

The present MAC system is still organized in a traditional manner in the sense that programs, whether public or private, are executed in succession as independent and indivisible entities. The fact that one program may be interrupted by a higher-priority program is irrelevant for the purposes of the present discussion. For instance, the execution of a system command generates a copy of the corresponding program (stored in the disk file), which is then run just as if it were a user program. Thus, if several users are simultaneously compiling programs written in the same language, several copies of the same compiler are simultaneously and independently shuttling back and forth between core memory and drum. Another consequence is that any interactive program must be present in main memory in its entirety when data or instructions are needed from the user, even though the presence of one or two of its subroutines may be sufficient to accomplish the interaction.

A further and very serious aspect of the present mode of operation is that only one complete, executable user program can reside in core memory at any one time. The implication is that the central processor must remain totally idle while a new program is being transferred into core memory or while necessary input–output operations are taking place. The idle time is very substantial in the present MAC system, as indicated by the WAIT percentages in Fig. 2(D).

These system inadequacies are clearly the result of an organization keyed to the traditional point of view of a central processor executing independent programs one at a time. Moreover, the characteristics of the present equipment would preclude in practice, if not in principle, any substantially different system organization. In order to overcome these limitations, one must approach the system design problem from a point of view considerably different from the traditional one.[12,13] We observe in this regard that even the term "time sharing" is inappropriate and somewhat misleading because it emphasizes the necessary but secondary goal of providing the equivalent of a private computer. The term "multiple access" could also be considered misleading when it is applied to the central processor.

The emphasis instead should be on the system ability to provide convenient and flexible multiple access to an ever-changing structure of procedures and data capable of interacting as parts of distinct processes. In other words, it is the software structure that is important, and the hardware assumes the secondary role of providing convenient means for access to it.

If the system goal is to provide convenience of access to such a software structure, one is naturally led to view the system itself as memory centered rather than processor centered. Furthermore, the memory that forms the heart of the system would be not the main core memory, but the bulk memory, consisting of disk files or similar devices in which the procedures and data are normally stored. The main memory would play, instead, the role of a giant buffer matching the relatively low transfer rate of the bulk memory to the fast processing rate of processors and input-output channels. When the system is looked at from this point of view, it assumes the appearance more of a message-store-and-forward communication system than of a traditional computer system. Indeed, the smooth flow of messages is of paramount importance to efficient operation; thus, the major function of the supervisory program is to organize the transfer of messages (procedure and data, as well as input-output messages) in such a way as to avoid the problems associated with unnecessarily long queues and to insure efficient utilization of the available equipment.

It is also clear that a computer system intended to serve a large number of people simultaneously must be organized so that it will avoid any unnecessary duplication of information in either main or bulk memory and unnecessary transfers between the two. This statement implies that procedures and data must be executable as common parts of processes that have been simultaneously and independently initiated by different users. It also implies the possibility of executing processes involving several subroutines in such a manner that only the necessary subroutines are in core memory at any given time. The program segmentation scheme advocated by J. B. Dennis is keyed to these objectives.[13]

A corollary to this view of a computer system is the functional subdivision of the hardware into pools of equipment serving the same function, with each piece of equipment being duplicated to meet the average system demand. The point here is that if the objective of the system is to provide convenient access to the procedures and data stored in the bulk memory, enough equipment must be provided to perform the necessary functions that a bottleneck in one part of the system will not prevent the full utilization of other parts. The equipment itself can then be designed to match the average demand of users as a group, rather than the requirements of the "typical users." Furthermore, if each piece of equipment is duplicated within the system, one can envision the achievement, through on-line maintenance, of a level of reliability and continuity of operation that is unthinkable for the present MAC system.

In conclusion, the experience with the present MAC system suggests a trend toward memory-centered (as opposed to processor-centered) systems—including pools of bulk memories, core memories, central processors, and input-output channels, all communicating with one another, with the core memories acting as buffers. On the software side, the trend seems to be in the direction of executing processes that consist of many subroutines and data structures, which are never assembled into a single program and some of which may be common to other independent processes simultaneously in execution. This view of computer systems is indeed very different from the traditional one. Its implications are far from clear. Their exploration is a major objective in the development of the next MAC system.

This article is based on a paper presented at the Symposium on Computer Augmentation of Human Reasoning, Washington, D.C., June 16, 1964, and published in the Proceedings of this Symposium. Project MAC is sponsored by the Advanced Research Projects Agency, Department of Defense, under Office of Naval Research Contract Nonr-4102(01).

The author acknowledges the accomplishments and ideas of many people associated with Project MAC, particularly F. J. Corbató of the M.I.T. Computation Center and his staff, who are responsible for the development of the MAC system. He also expresses gratitude to the sponsoring agencies—the Advanced Research Projects Agency and the Office of Naval Research—and to the International Business Machines Corporation and the New England Telephone and Telegraph Company.

REFERENCES

1. McCarthy, J., "Time-Sharing Computer Systems," in *Management and the Computer of the Future*, M. Greenberger, ed. Cambridge, Mass.: The M.I.T. Press, 1962, pp. 221–236.

2. M.I.T. Computation Center, *The Compatible Time-Sharing System*. Cambridge, Mass.: The M.I.T. Press, 1963.

3. Corbató, F. J., Daggett, M. M., and Daley, R. C., "An Experimental Time-Sharing System," *Proc. AFIPS Conf.*, vol. 21, 1962, pp. 335–344.

4. Boilen, S., Fredkin, E., Licklider, J. C. R., and McCarthy, J., "A Time-Sharing Debugging System for a Small Computer," *Ibid.*, vol. 23, 1963, pp. 51–57.

5. Schwartz, J., "A General-Purpose Time-Sharing System," *Ibid.*, vol. 25, 1964, pp. 397–411.

6. Dennis, J. B., "A Multiuser Computation Facility for Education and Research," *Commun. ACM*, vol. 7, Sept. 1964, pp. 521–529.

7. Stotz, R., "Man-Machine Console Facilities for Computer-Aided Design," *Proc. AFIPS Conf.*, vol. 23, 1963, pp. 323–328.

8. Ross, D. T., and Feldman, C. G., "Verbal and Graphical Language for the AED System: A Progress Report," Tech. Rept. MAC-TR-4, M.I.T., Cambridge, Mass., 1964.

9. Biggs, J. M., and Logcher, R. D., "Stress: A Problem-Oriented Language for Structural Engineering," Tech. Rept. MAC-TR-6, M.I.T., 1964.

10. Weizenbaum, J., "OPL-1: An Open-Ended Programming System Within CTSS," Tech. Rept. MAC-TR-7, M.I.T., 1964.

11. Greenberger, M., "The OPS-1 Manual," Tech. Rept. MAC-TR-8, M.I.T., 1964.

12. Corbató, F. J., "System Requirements for Time-Shared Computers," Tech. Rept. MAC-TR-3, M.I.T., 1964.

13. Dennis, J. B., "Program Structure in a Multi-Access Computer," Tech. Rept. MAC-TR-11, M.I.T., 1964.