VOLUME 3 ☐ NUMBER 4 **APRIL 1966**

IEEE spectrum

features

Departments: *Please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

departments

the cover

Graphical language techniques in time sharing, described in article starting on page 62, are opening up a "new world" for scientific investigators and designers. Cover shows four views of a complex 3-D form taken from computer-controlled display tied into Project MAC.

VOLUME 3 ☐ NUMBER 4 IEEE spectrum APRIL 1966

Spectral lines

Spectrum. At its March meeting, the Board of Directors of the IEEE made a significant change in the management of SPECTRUM by appointing C. Chapin Cutler of the Bell Telephone Laboratories as its editor. This change, which was proposed by the IEEE Publications Board, should open the way for the development of SPECTRUM into a periodical of increased value to the membership. To assist Mr. Cutler, a new editorial board has also been appointed.

Since its first issue in January 1964, SPECTRUM has become a publication of considerable stature—primarily through the efforts of the Headquarters' staff headed by E. K. Gannett and Ronald K. Jurgen, under the general guidance of the IEEE Publications Board and a board of consultants. However, it has been the feeling of the Publications Board that SPECTRUM needs the direct and continued attention of an editor and editorial board who are broadly representative of the Institute's membership. SPECTRUM's mission as the core publication of the Institute is a most important one, and its potentialities for significant service are great.

It should not be inferred from these statements that SPECTRUM's development has been unsatisfactory to date. Its growth has been strongly assisted by John D. Ryder, Walter K. MacAdam, and Thomas F. Jones, who, as members of the Publications Board, assumed special responsibility during its early stages. At first, there was considerable misunderstanding as to SPECTRUM's role. It was not intended to be a direct replacement for either of the primary publications of the IEEE's predecessor societies; it was meant to be a new publication that would represent the broader fields of scientific and technical interest of the new Institute.

It is interesting to see how SPECTRUM has carried out its mission.

The 28 issues published through March 1966 have presented 175 feature articles on a variety of technical subjects. It is not possible to categorize these articles in any simple way, but based on a division of interest similar to the present Group structure, it is fair to say that they fall into about 25 technical areas. The number of articles per issue has been increased from between five and six to about seven. The average length of an article has been decreased from ten pages to seven. These changes have been designed to make SPECTRUM more readable; it is hoped that the result will be that more members will read more articles.

A SPECTRUM feature that has been developing quite satisfactorily is the Special Conference Reports. It was felt that a valuable service could be rendered to the member by publishing summaries of the highlights of the various significant technical meetings held throughout the world. So far 15 such reports have appeared in SPECTRUM, and it is anticipated that the conference coverage will be even greater in the future.

The Headquarters' staff responsible for SPECTRUM includes several writers who have been assigned to develop articles on subjects of special interest. Such articles have supplemented the contributions of members and specialists writing on topics in their own fields. About ten percent of the articles published have been staff-written. In many cases they have been very timely and have provoked much favorable reader response.

It is evident from this description of SPECTRUM that many decisions must be made each time an issue is released for publication. Is the selection of feature articles appropriate in regard to subject matter and presentation? Are the subjects for staff-written articles well chosen? Should more or fewer articles be staff-written? Are the regular departments, such as "News of the IEEE," "Transients and Trends," "Focal Points," and "Book Reviews," useful and significant? Should new departments be instituted?

It is hoped that SPECTRUM's new editor and editorial board will help to answer these questions. It will be their immediate responsibility to direct the work of the Headquarters' staff assigned to SPECTRUM. To assist them in their work, it is planned to make more comprehensive readership surveys than those presently available. Every effort will be expended to produce a publication that will be even more interesting, more widely read, and more useful to the membership.

The Institute has a tremendous asset in having available to it the voluntary services of men of the caliber of Mr. Cutler and his associates on the newly appointed editorial board. I am sure they have the best wishes of the members. Their success will be our gain.

F. Karl Willenbrock

Authors

Clear air turbulence detection (page 56)

R. T. H. Collis was born in London in 1920. He was a pilot in the Royal Air Force from 1940 to 1947, during which time he acquired extensive flying experience ranging from instructing to fighter operations. Among a number of unusual assignments, he spent some time flying captured German aircraft; he also commanded one of the first units of jet aircraft in the British Air Force. Later he attended Oxford University, from which he received the M.A. degree. A graduate of the Royal Naval School of Meteorology, from 1950 to 1955 he served in the British Royal Navy as a meteorologist specializing in aviation forecasting. During this period he was involved with some of the clear air turbulence phenomena experienced by military jet aircraft. He is at present head of the Radar Aerophysics Group at Stanford Research Institute in California, which he joined in 1958. He is currently working on the development of lidar techniques in atmospheric research. His interests also extend widely in radar and radio meteorology, particularly in connection with operational applications. He has developed original techniques for measuring rainfall by radar and has carried out a number of studies relating to the remote observation of the atmosphere from space satellites. Before joining SRI he was with Decca Radar Ltd. in England, where in addition to serving as a meteorological specialist he also worked on marine and aviation applications of radar. Mr. Collis is the author of a number of technical and operational articles dealing with radar and lidar meteorology, and has been granted patents in the fields of aircraft armament, radar reflector design, and weather radar data processing. He is a Fellow of the Royal Meteorological Society and a professional member of the American Meteorological Society. He became a citizen of the United States in 1964.

Detection of coherent optical radiation (page 73)

Marvin E. Lasser received the B.A. degree in physics from Brooklyn College in 1949 and the M.S. and Ph.D. degrees from Syracuse University in 1951 and 1954, respectively. At Syracuse he also held a teaching assistantship from 1949 to 1951 and a research assistantship from 1951 to 1954. He joined the Research Division of the Philco Corporation in 1954 as a project engineer. In addition to investigating surface properties, he was responsible for initiating the Philco infrared detector program. In 1956 he was promoted to research section manager in charge of a group investigating energy conversion phenomena. He was named manager—applied physics research in 1958, with responsibility for the general applied physics program, including lasers, material preparation, dielectric properties, photovoltaic devices, and electrochemical studies. In 1965 he was appointed director of the Applied Research Laboratory. Dr. Lasser is a Fellow of the American Physical Society.

High-speed information channels (page 79)

Richard T. James received the E.E. degree from Rensselaer Polytechnic Institute in 1935. After graduation he worked on technical assignments for the General Electric Company in Pittsfield, Mass., and for radio station WOR in New York City. He joined the Long Lines Department of the American Telephone and Telegraph Company in 1942, where he was concerned with the transmission aspects of broadband carrier systems, both cable and radio. His duties included work on the video program and telephotographic networks, as well as various special services. He was named district plant engineer in 1949 and later became a staff assistant in plant operation and maintenance. In 1952 he was transferred to the General Engineering Department, where he worked on special services. At present he is in charge of a system engineering group involved in providing wide-band systems for special services of all types. These include high-speed data and graphic systems, as well as telemetry systems.

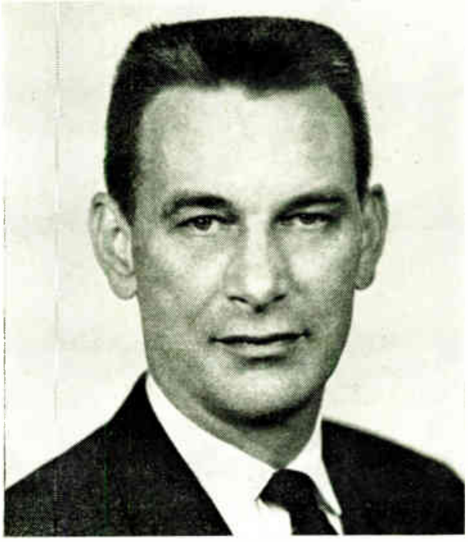
Engineering–economic systems: a new profession (page 96)

W. K. Linvill (M) is concerned with problems of systems analysis and decision making. The areas of application in which he has worked range from computer-coordinated systems to long-range planning. He received the joint B.S.E.E. and M.S.E.E. degree in 1945 and the doctoral degree in 1949, both from the Massachusetts Institute of Technology. He was an assistant professor at M.I.T. from 1949 to 1953 and associate professor from 1953 to 1958. His research interests were in computer control systems, and a considerable portion of his time was devoted to computer control systems for air traffic control and air defense. In 1954 he became more specifically interested in the systems area, and in 1956 he took a two-year leave of absence from M.I.T. to lead a project on NATO air defenses, sponsored by the Institute of Defense Analysis. In 1958 he became a senior staff member of the Rand Corporation. Since 1960 Dr. Linvill has been a professor of electrical engineering at Stanford University. In 1963 he became chairman of Stanford's Institute in Engineering–Economic Systems, where he established a graduate systems training and research program, in which field internship plays a very important role.

Interplanetary spacecraft telecommunication systems (page 103)

Glenn A. Reiff (M) was recently named NASA headquarters manager of the Mariner 67 program, which includes a planetary exploration mission to Venus and resumption of communications with Mariner IV during 1967. Previously he was in charge of the Mariner 64 and Pioneer programs. The former culminated in the successful completion of the first mission to Mars and the return of close-up pictures of the planet. The latter resulted in the flight of Pioneer VI and the return of valuable data on the interplanetary environment.

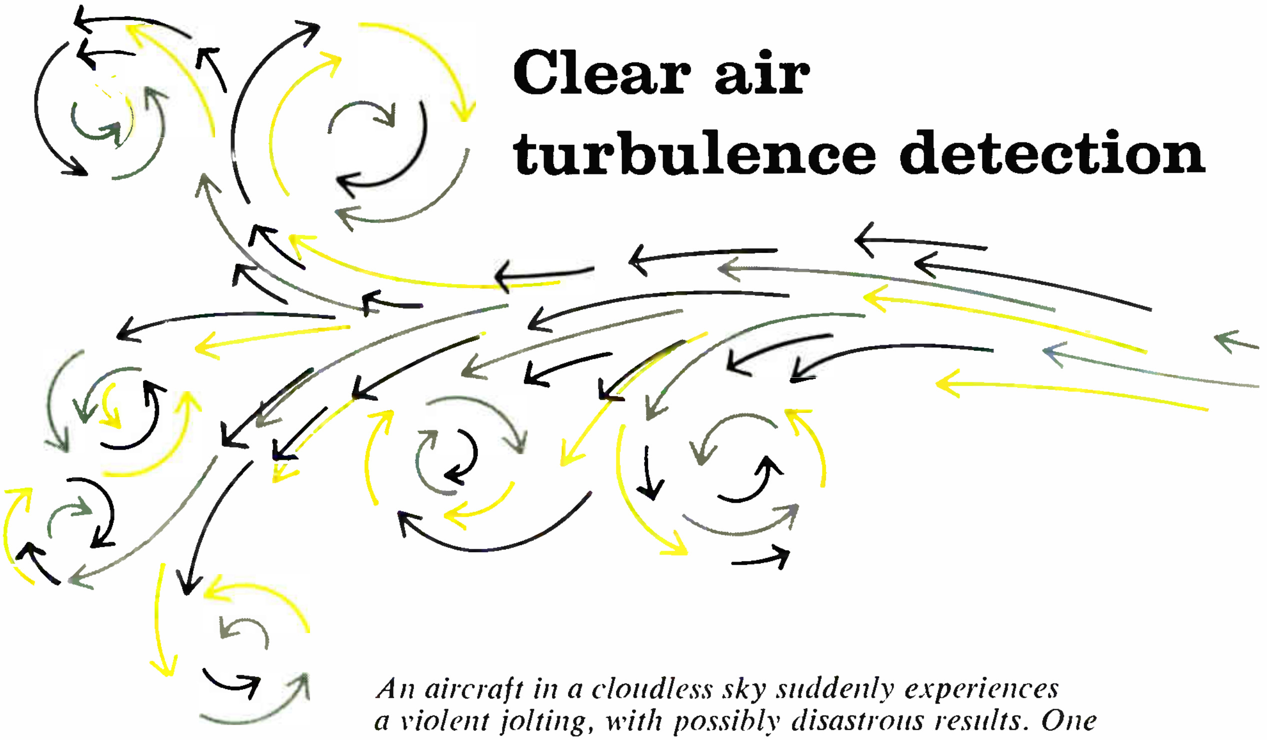
Mr. Reiff attended Texas A & M College during 1941–1942 and was graduated from the U.S. Naval Academy in 1945. His assignments at sea included duty aboard aircraft carriers, in all-weather reconnaissance and antisubmarine squadrons, and in fleet air staffs. Upon completion of the engineering electronics course at the Naval Postgraduate School, he was awarded an additional B.S. degree and an M.S. degree, and was assigned to aeronautical engineering duties. Before joining NASA in 1962, he worked on oceanographic and undersea warfare systems for Sanders Associates, Nashua, N.H.



On the mechanization of creative processes (page 112)

Saul Amarel (SM) received the B.S. degree in 1948 and the degree of Ingenieur E.E. in 1949 from the Israel Institute of Technology, Haifa. For six years he was associated with the Scientific Department of the Israeli Ministry of Defense, for whom he led the development of control and communication systems and also conducted research on computer simulation methods related to operations research problems. He was with the Electronics Research Laboratory at Columbia University from 1953 to 1955, where he developed operational methods for the analytical study of dynamic systems. He received the M.S. degree in 1953 and D.Eng.Sc. degree in 1955, both from Columbia.

Since 1957 Dr. Amarel has been a member of the technical staff of RCA's David Sarnoff Research Center at Princeton, N.J. At the present time he heads the Computer Theory Group of the Computer Research Laboratory. His work encompasses the fields of artificial intelligence, computational linguistics, and switching theory. He has written several papers and has presented talks and conducted courses on these subjects. He is a member of the American Mathematical Society, the Association for Computing Machinery, and Sigma Xi.



Clear air turbulence detection

An aircraft in a cloudless sky suddenly experiences a violent jolting, with possibly disastrous results. One way in which we can help the pilot to cope with this problem of clear air turbulence is to develop an effective warning system

R. T. H. Collis *Stanford Research Institute*

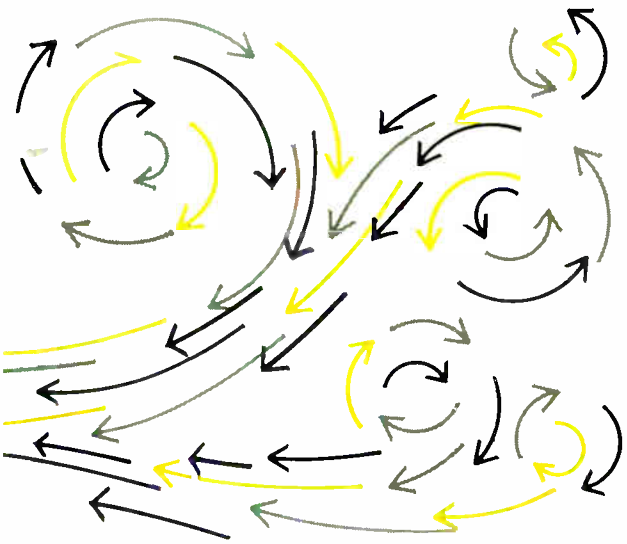
Although it is difficult to describe the nature of clear air turbulence (or CAT), we know its effect—the sudden vibration experienced by an aircraft in flight in a stormfree, cloudfree sky. It is evident that this turbulence is an eddy motion of the air that disrupts its uniform flow, a motion such as that associated with jet streams; however, we do not know exactly how it is generated. Because the effects of CAT can range from mere passenger discomfort to actual loss of an aircraft, it is important to devise some method of warning the pilot of turbulent areas to be avoided. Various indirect and direct detection systems, employing both active and passive techniques, are currently being investigated, but thus far all have exhibited serious limitations on effective performance. One reason for this is the lack of data concerning the phenomenon they are designed to detect.

Clear air turbulence (CAT) is an aviation problem. It concerns most directly those who fly in aircraft, those who operate them for profit, and those who make them. Others who are interested include meteorologists, air traffic controllers, insurance carriers, and developers and manufacturers of instrumentation. The last mentioned have recognized a potentially rich market for airborne CAT detection and warning systems and are addressing themselves to the problem in increasing numbers. This article is an attempt to bring the picture into focus from their viewpoint, a task which would be easier if the subject could be seen clearly from any point of view.

What is clear air turbulence?

The difficulty of answering the initial question as to what clear air turbulence is immediately underlines the lack of understanding and precision that characterizes the topic. (See the bibliography for a selected list of publications related to this general subject.) As generally understood in the present context, it is a phenomenon encountered by aircraft in flight, usually at high altitudes (about 20 000 feet) in areas free from clouds and not in the vicinity of obvious convective storm activity. It is manifest by its effect on the aircraft, which ranges from a mild vibration to a violent jolting, or worse.

The aircraft's behavior in such turbulence involves both the aerodynamic characteristics of the aircraft and the control maneuvers of the pilot. This fact makes it difficult to distinguish the intensity or nature of CAT on an absolute basis, and, in practice, it is generally reported in a rather subjective, qualitative manner. More objective data can be acquired by accelerometers and at least one system of direct measurement of a critical parameter related to turbulence independent of the aircraft has been developed.[1] Thus, although in special research investigations the incidence of CAT is comprehensively and objectively reported, for the most part our knowledge is based on the accumulated experience of pilots and their personal accounts of CAT encounters. Since the criteria for distinguishing the critical conditions involved are by no means clear-cut, considerable uncertainty is inevitable.



For example, it is often very difficult to ascertain what is meant by "clear" conditions. The presence of tenuous cloud is hard to detect in flight, particularly at night. It is often equally difficult to appreciate whether there is significant cloud in the general vicinity of the aircraft or whether convective activity is involved. Even among the meteorologists there is a lack of complete agreement as to the nature of the phenomenon. This stems in part from the shortage of good quantitative observational data and in part from the complexity of the causative mechanisms.

Atmospheric turbulence as experienced by aircraft is an eddy motion or gustiness of the air that disrupts its uniform flow. Although the mean airflow can be evaluated, the irregular turbulent motion can only be described statistically. The intensity and size of the eddies, and their frequencies (which are interrelated), extend over a very wide range. The larger eddies in general involve the higher air velocities and contain the most energy; they occur randomly in the three dimensions. Another form of "turbulence" experienced by aircraft is actually a more regular wave motion, in the atmosphere, that can extend over a wide range of amplitudes and lengths. Although the effect of such motion at critical wavelengths can obviously affect aerodynamic flight, most meteorologists believe that this mechanism is only important as a source of CAT when such wave motion contributes to the breakdown of regular flow.

What causes CAT?

Given that the atmosphere is in motion due to large-scale meteorological phenomena, turbulence occurs when the smooth airflow is disrupted as a result of a failure to transfer energy by continuous motion. This occurs when velocity gradients cause shear stresses that cannot be relieved quickly enough by smooth flow. Apart from the variations in wind speed and direction (wind shear), the breakdown of smooth flow also involves in a very complex manner such properties of the atmosphere as its thermal and density stratification and buoyancy.

The velocity gradients leading to turbulence usually result from the mesoscale features of the meteorological situation but they can be induced by mechanical effects. Thus, surface friction and topographical features can give rise to turbulence at low levels. In mountainous terrain the disturbance of the airflow may extend to much greater heights. In particular, where a mountain range lies across a strong airstream, wave motion is generated in the lee of the range and can be propagated to great heights and to a considerable distance downwind. These lee waves can have large amplitude and can give rise to strong up and down currents. It is believed, however, that the severe turbulence often associated with them can be attributed to the breakdown of smooth flow induced by the large shears resulting from the wave motion. The wave motion appears to be particularly prone to breakdown when thermally stable layers are encountered, and it has been noted that mountain wave turbulence tends to be worse at levels near and below the tropopause.

Of the purely meteorological phenomena that produce CAT, undoubtedly the most important is the jet stream. This is a factor of the general circulation of the atmosphere which is found at the upper levels (above 20 000 feet). It takes the form of a belt or ribbon of high-speed winds, some hundred or more miles wide and a few miles deep, in which velocities as great as 200 knots are encountered. In both hemispheres such a jet occurs with some regularity in the subtropics (at around 30° latitude) circling the earth. In middle and high latitudes less regular jet stream systems occur; the position and intensity of these are variable and are functions of the interplay of air masses of different thermal characteristics. The jet streams are much influenced by seasonal and geographical factors. The presence of strong wind shears around them is an obvious predisposing factor to the incidence of CAT, and a good deal is known, both theoretically and empirically, about the association of CAT with jet streams. Other dynamic features of upper atmospheric airflow can be associated with CAT, but again, although good correlations can be established statistically, the way in which CAT is generated in these cases is by no means fully understood. Recent work has shown that certain distributions of wind velocities and temperature can lead, intrinsically, to a situation in which strong wind shears develop as a kinematical property of the flow.[2]

Where does CAT occur?

If we could say, in detail and with confidence, where CAT occurs, there would be no CAT problem. In practice, although the broad features of the incidence of CAT are fairly well known with respect to the overall meteorological situation, it is not possible at the present time to pinpoint hazardous areas. One reason, of course, is that the present input of data regarding conditions aloft is inadequate for specifying areas of CAT hazard in terms of the time and space scales operationally necessary.

Meteorological soundings of the upper atmosphere are made at 12-hour intervals from a network of radiosonde stations around the world. The density of this network of course varies widely; even over the continental United States, where there are some 75 stations, it is only adequate for determining the large-scale situation.

Most turbulence is associated with certain areas of middle- and high-latitude jet streams, generally at the height of maximum winds. Clear air turbulence is also found in association with certain other airflow patterns aloft, particularly where strong wind shear is present.

Clear air turbulence can occur at all heights and is frequently present in the stratosphere at heights up to 70 000 feet or more. It often occurs in fairly shallow layers. The flow of air over mountains can cause or accentuate turbulence for hundreds of miles downwind and to heights well into the stratosphere (above 40 000 feet). Over the United States in particular the geographical and seasonal habits of jet streams are fairly well known, as are the areas of mountain wave effects. Thus, although the incidence of CAT is widespread, certain "preferred" areas are emerging but it is noteworthy that existing domestic air routes are laid out without regard to such areas. There is a strong possibility that as flying extends to parts of the world now considered remote or to higher altitudes, i.e., the middle stratosphere, CAT may be found to be even more widespread than is currently apparent.

Although increasingly good work is being done by meteorologists in providing general warnings of CAT, the most effective method under normal operational conditions would be one by which hazardous areas could be detected from the aircraft in flight. (At present, probably the best defense against the danger is the practice among pilots of exchanging information by radiotelephone of the CAT situation along the air routes.)

The why and how of CAT detection

The immediate effects of encountering CAT in flight are obvious—it disturbs the comfort and threatens the safety of the passengers and air crew. Physical damage can range from clothing spoiled by spilled beverages through physical injury as occupants fall or are flung about the aircraft to (in extreme cases) structural damage. Indirectly, CAT is believed to be responsible for contributing to structural fatigue. This possibility, together with the need to provide adequate margins of structural strength to withstand the effects of routine CAT encounters, imposes construction practices that are initially costly and operationally uneconomical. It is clearly advantageous, therefore, to detect CAT so that it may be avoided in flight. Indeed, at the present time it would be valuable to provide prior warning of CAT encounters even if they could not be avoided.

If these factors are important in modern commercial and military jet aircraft operation, it is certain that they will be even more important in the case of the supersonic transports now being designed. With such aircraft, the penalties of carrying unnecessary structural weight or the fuel to make unnecessary detours will be even more onerous. There is also the possibility that the effects of CAT on such aircraft at both subsonic and supersonic speeds will be particularly hazardous because of their size, weight, and radically different configuration.

Another quite important factor to consider is that even if a practical solution cannot be found for detecting CAT on a routine basis, any system that could be used to increase our present knowledge of the incidence and nature of this phenomenon in detail would be extremely valuable.

Currently, as noted, it is only possible to provide general warnings as to CAT-prone areas on the basis of general meteorological information. Such warnings are used to a certain extent to plan flight routes and select flight altitudes. Although considerable progress has been made in this respect, the approach is clearly limited. Between the recognition of predisposing meteorological conditions and the actual encounter with CAT, there is a complete hiatus at the present time. A number of possibilities exist that might fill this gap. Although dividing lines are not well marked, two distinct steps are indicated. First, it would be of enormous value to be able to recognize or distinguish CAT-prone areas with more precision and on a more detailed scale in terms of space and time. The general prior warning of CAT hazard used in planning a flight could then be complemented by a particular warning of a CAT-hazardous area should one be encountered in flight, thus enabling suitable avoidance procedures to be followed. The second step would be to provide specific warning of actual CAT conditions lying in the path of the aircraft.

The primary purpose of providing warning of CAT on either basis is to enable pilots to avoid such areas. If evasive action is impractical, then a second purpose is to give the aircraft's occupants an opportunity to prepare for the onset of turbulence. Even a minute or so of warning would enable the air crew to make the appropriate adjustments to the aircraft's controls for optimum penetration, and would also give the cabin crew and passengers a chance to ready themselves for the encounter.

The most important requirement, then, in providing warning of CAT is to give the earliest possible indication of the imminence of the hazard. It is almost equally important, unless the rather negative approach of accepting the inevitable is adopted, to know what evasive action may be taken to avoid an encounter. Since modern passenger aircraft fly at a speed of some ten miles per minute, and speeds of twice or three times that will be routine with supersonic transports, CAT hazards must be recognized at a range of many miles if any useful forewarning is to be achieved. Aircraft flying at such speeds can only change direction relatively slowly if accelerations unacceptable to structure and occupants are to be avoided. In addition, their freedom of maneuver in the vertical is limited by virtue of the fairly narrow margin that exists between cruising speed and critical high and low speeds at high altitude. A fully effective CAT warning system must be capable not only of providing warning of the presence of a CAT hazard many miles ahead in the aircraft's path, but of providing information on its spatial extent at sufficient ranges to enable other paths to be selected.

In practice this means that even for subsonic aircraft the CAT hazard must be appreciated at ranges of the order of at least 30 miles ahead, and that a detection system should cover the volume of air space some 5000 feet above and below the flight path for as many miles as possible on either side (with useful directional information). Even so, the effectiveness of such a system for evasive purposes would depend upon the spatial extent of any CAT areas lying in the aircraft's path. (The requirements for supersonic aircraft are of course considerably more exacting.) The problems of devising such systems are formidable, and it is already apparent that only a major breakthrough would offer a really useful solution. For these reasons it seems evident that the most promising approach lies not in attempting to detect CAT as such, but rather in recognizing, in flight, the particular areas in which CAT can be expected by reference to the characteristics of the upper atmosphere in the locality of the aircraft. The alternative, that of specifically detecting CAT as such, involves a difficult question: What exactly is to be detected?

Many concepts relating to the development of airborne CAT warning systems are currently being investigated. They cover a wide range, but so far have met with little success. Rather than attempt an exhaustive review, it is proposed to describe the main lines of approach.

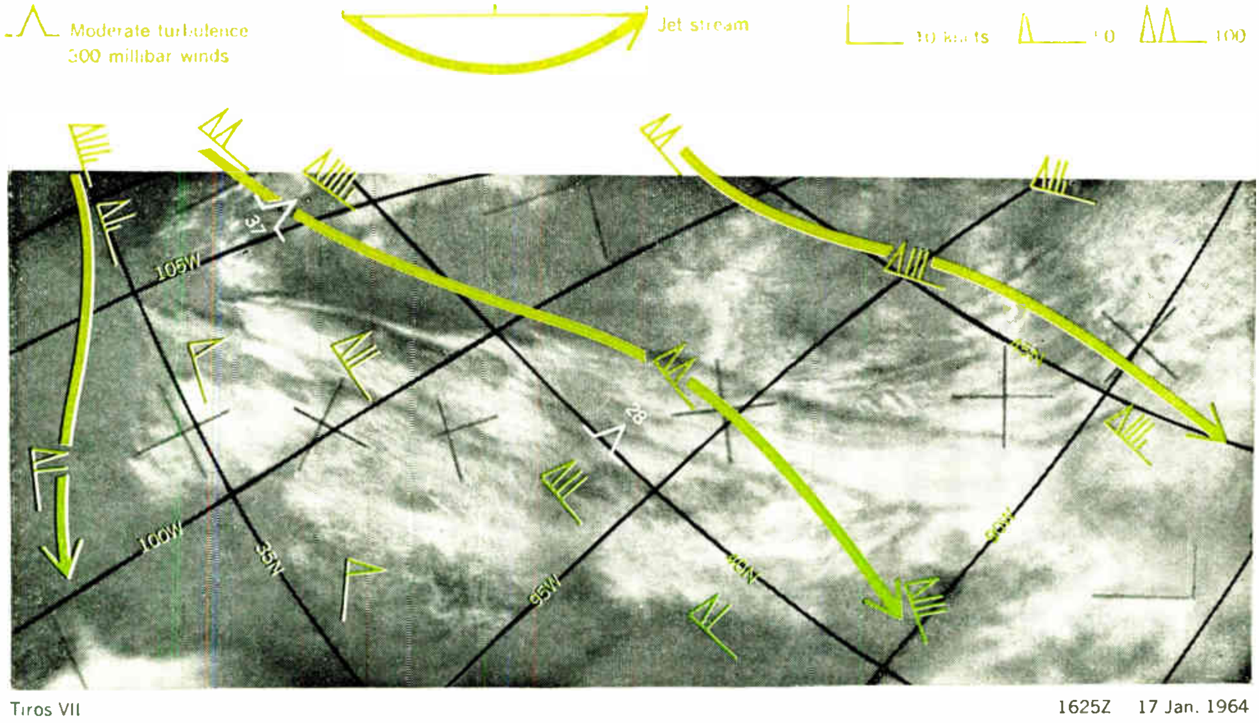
Direct detection of CAT

Active techniques. The direct method is largely concerned with detecting dielectric variations or discontinuities in the atmosphere caused by turbulence. It is believed by some that such variations may be detected remotely by active probing techniques using appropriate parts of the electromagnetic spectrum. The detection of dielectric inhomogeneities in the atmosphere by ground-based radar has prompted the application of such radars in the present context. For instance, after having conducted promising ground-based experiments,[3] the Boeing Company is now experimenting with a 150-kW 220-MHz radar mounted in a 727 trijet. (The antenna is a 40-foot-long half-Yagi array employing a series of monopole elements mounted along the top of the fuselage.) The dielectric gradients that can be expected from turbulence in a homogenous air mass would yield very low signal intensities. It appears that the stronger echoes observed by the Boeing ground-based radar are due to the dielectric gradients produced by the mixing of air of different characteristics in areas of strong wind shear, such as are associated with jet streams.

At optical frequencies the dielectric properties of the air also depend upon temperature and density. It thus has been proposed to use laser radar (lidar) to detect turbulent inhomogeneities in the atmosphere.[4] Although this is theoretically possible, it has been shown that the backscattering to be expected from dielectric discontinuities is much too small to consider seriously.[5] Also, differences in backscattering from the gaseous atmosphere resulting from variations in static pressure caused by turbulence are virtually negligible, and would in any case be masked by backscattering due to particulate matter.

The backscattering of light energy from particulate matter, even in what passes for clear air, offers quite a different approach to probing the atmosphere by remote means. Lidar observations from the ground and from aircraft have shown that it is possible to distinguish discontinuities in what appears to be quite clear air.[6] This has prompted the hope that it may be possible to distinguish a characteristic structure in turbulent air as observed by this method. Both Prof. P. Franken of the University of Michigan[7] and NASA (Langley) are carrying out airborne experiments in connection with this hypothesis, using lidars developed by Lear Siegler, Inc.* Another interesting concept regarding the use of lidar for CAT detection involves the detection of Doppler shifts in backscattered returns from the gaseous atmosphere or particulate matter therein, by exploiting the coherent properties of laser energy.[8] Studies—by North American Aviation, Inc., among others—have shown that this approach faces very formidable difficulties, which involve both engineering problems and the uncertainties regarding the distribution of particulate matter in the upper atmosphere.

Characteristic cloud patterns associated with strong jet stream flow as viewed by TIROS VII weather satellite, January 17, 1964. Superimposed wind data for 30 000-foot level based on 1200Z observations. Locations of two incidents of moderate turbulence that occurred within one hour of the time of photograph are shown with the heights indicated in thousands of feet. (Central location is over Omaha, Nebr.; that in the W just south of Denver, Colo.) Airline reports show that CAT was unusually prevalent throughout the region on that day.



---

* See *Focal points*, March SPECTRUM, p. 212.

Passive techniques. A novel, passive, electromagnetic technique is under investigation at Stanford Research Institute in collaboration with United Air Lines. This research seeks to establish a relationship between the atmospheric electrical field and CAT.[9] In an entirely different earlier experiment, a connection was noted between corona discharge currents measured in the aircraft and turbulence in clear air. The present experimental program is investigating this phenomenon further in the hope that such charging is attributable to the aircraft's passage through potential gradients associated with CAT. If this is so, sensitive instrumentation might make it possible to detect such field changes at a distance, and thus provide warning of CAT. Indications to date, however, suggest that the charging is the result of triboelectric effects as the aircraft passes through subvisible clouds of particulate matter, but anomalies in the magnitude of the effects observed still must be resolved.

The passive detection of turbulence-caused dielectric inhomogeneities using stars as distant sources of light energy is being considered; but although stellar scintillation and astronomical "seeing" conditions have been related to atmospheric turbulence and air motion, there are obvious practical difficulties in applying this principle to an airborne CAT detection system. To monitor the aircraft's path it would be necessary to employ a star on the forward horizon and the method could only readily be employed at night or twilight.

It has also been suggested that the thermal discontinuities present in a turbulently mixing atmosphere might be distinguished remotely by radiometric techniques.[4,10] Because of the problems involved in resolving such discontinuities on an appropriately small scale, it seems that such techniques would be more useful in detecting the broader-scale variations in temperature that are associated indirectly with CAT areas. This concept is discussed in the following.

Indirect detection of CAT

The basis of the indirect approach to providing warning of CAT is the detection of characteristic conditions from which the presence, location, and intensity of CAT can be inferred with high probability. This approach has taken various forms, most of which rely on passive techniques. One of the first positive efforts of this type was the visual observation of cloud patterns that could reveal the presence of jet streams.[11] Attempts to relate cloud observations to CAT have been partially successful,[12] but the difficulty of making and interpreting such observations has precluded a fully useful outcome of this approach. Another early technique also concerned with the detection of jet streams and significant dynamic features of the upper atmosphere was the monitoring of ambient temperature. Although this technique is valuable to aircraft navigating in the vicinity of certain jet streams,[13] it was early suspected that it had serious limitations in providing warning of turbulent areas.[14] The reasons for this inefficiency became apparent in later analyses of the potential of this method for CAT warning. Flight trials carried out by Eastern Airlines,[15] using instrumentation developed by Litton Industries, have revealed a correlation between thermal gradients and CAT incidence in certain circumstances. However, it has been shown that thermal gradients can occur relative to the jet streams and CAT-prone areas in such a way that in many cases an aircraft encounters CAT before passing through the significant temperature gradients.[16] This follows from the basic anatomy of jet streams. Again, thermal gradients are frequently detected aloft without any CAT being encountered.[14,15]

Investigations are continuing, however, to establish more closely the relationship between thermal gradients and CAT as some hopes are entertained for detecting such thermal gradients remotely by radiometric methods. Two regions of the spectrum are being studied for this purpose—the infrared and millimetric microwaves. Both techniques rely upon the rapid change of absorption that occurs on the skirts of atmospheric absorption lines [e.g., the CO_2 line at 15 μm (micrometers) in the infrared or the 5-mm O_2 line in the millimetric band]. The approach is well illustrated by an instrument, developed by Barnes Engineering Company, which is essentially a spectrometer scanning in a band from 13.5 to 14.5 μm.[17] If the instrument is pointed into an atmosphere having a constant temperature, its output throughout the spectral scan will correspond to that produced by a blackbody radiator at the air temperature. Where a thermal gradient is present, along the pointing direction, the received energy (which is a function of both the emission integrated along the path and the absorption taking place along the path) will show variations in intensity with wavelength. The evaluation of these in terms of thermal gradients or discontinuities is very complicated. It is anticipated that any practical system relying on this technique will depend upon a pragmatic approach in which characteristic readings are related to the incidence of CAT on the basis of experience.

Other indirect methods of discerning the structure of the atmosphere, and, in particular, of locating jet streams and wind shear regions with which CAT may be associated, include those that measure changes in the airflow in the locality of the aircraft. Northwest Airlines, for example, has investigated the relation between the rate and magnitude of drift oscillations, as detected by the aircraft's Doppler navigation system, and CAT with limited, but positive, results. Basic Devices, Inc., has developed a wind-gradient measuring system employing probes that operate on the principle of the hot-wire anemometer. These probes, mounted one on each wing tip or one above the other, could be used to determine differential air motion over horizontal or vertical intervals.

The most substantial indicators of air motion aloft, however, are the water-substance particles of clouds, and also the particulate matter present in "clear" air. Apart from the more obvious forms of high-level water-substance cloud, there is increasing evidence that subvisible stages of such clouds exist and that haze or dust concentrations are very variable. It remains to map such clouds or variations in particulate matter concentration so that airflow patterns may be interpreted. If this were possible the features of jet streams and other dynamic phenomena could be distinguished and areas in which CAT is likely could then be inferred, probably on largely empirical grounds. It is hoped that active radar techniques, specifically lidar, will enable such delineations to be made.[18] It is important in this context to recognize that the nature of air motion can readily be inferred in certain cases by reference to visible indicators at a considerable distance from the area in question. Thus, in the appropriate conditions, a lee wave motion can confidently be expected in clear air at heights several thousand feet above the bands of alto cumulus cloud that show characteristic wave configurations.

Conclusion

Current experience and developments in aviation indicate that CAT is a problem of increasing seriousness. While alarmism is to be deplored, it would be greater folly to think that the problem is a minor one that can be solved with relatively little effort. The search for effective detection and warning systems must be pursued with increased urgency.

The rewards to the developer of a successful airborne system would be very great, for, as with weather radar which has been so valuable in the case of convective storm hazards, every large aircraft would be fitted with one.

The form such a system would take is still unknown as no really promising approach to an early solution has yet emerged. Direct detection systems face formidable problems in achieving useful performance. Indirect systems face similar problems, and, in addition, suffer even more acutely from our lack of adequate knowledge regarding clear air turbulence.

Any system carried in an aircraft must obviously be of reasonable size, weight, and cost. To be fully effective, in addition to warning of the imminence of CAT in the aircraft's path, it must provide information to the pilot, in an acceptable form, of how he can best avoid the hazard. This involves the presentation of data in at least two dimensions. Because CAT often occurs in relatively shallow layers, and since the height of an aircraft can quickly be changed with relatively mild maneuvers, there are grounds for suggesting that such a presentation could well be of the vertical plane through the aircraft's path.

It is quite possible that the best solution to the problem lies outside the aircraft. In particular, the ability of meteorological satellites to monitor large areas and provide virtually instantaneous information on the nature and distribution of clouds that may be used to infer the dynamic features of the atmosphere is an exciting possibility.[19, 20]

One thing is quite certain. More detailed knowledge and a better understanding of the nature, cause, and incidence of CAT are urgently needed. Relevant research is being carried on, sponsored mainly by government agencies. Data on CAT are being collected routinely by the U.S. Weather Bureau and in special flight programs by NASA and the USAF. There is clearly a long way to go, however, before CAT is adequately understood, and much more work is necessary. In the meantime, attempts to find instrumental solutions to the warning problem should be pursued with cautious realism.

In addition to the sources mentioned in the text, the writer owes many thanks to his colleagues at Stanford Research Institute and in particular to R. M. Endlich and S. M. Serebreny for their advice and comments.

REFERENCES

1. MacCready, P. M., "Standardization of gustiness values from aircraft," *J. Appl. Meteorol.*, vol. 3, no. 4, p. 439, 1964.

2. Endlich, R. M., and Mancuso, R. L., "Objective analysis and forecasting of clear-air turbulence," SRI Final Rept., U.S. Weather Bureau Contract Cwb-10871, June 1965.

3. Buehler, W. E., and Lunden, C. D., "A note on VHF backscatter from turbulence in the upper troposphere," *J. Appl. Meteorol.*, vol. 4, no. 1, p. 151, 1965.

4. Weihe, V. I., "Airborne sensing of clear-air turbulence," *Navigation, J. Inst. Nav.*, vol. 10, no. 3, p. 242, 1963.

5. Munick, P., "Backscatter of light by turbulent air," New York Meeting Rept.—1964, Optical Society of America, p. 28.

6. Collis, R. T. H., and Ligda, M. G. H., "Laser radar echoes from the clear atmosphere," *Nature*, vol. 203, p. 508, 1964.

7. Franken, P. A., Jenney, J. A., and Rank, D. M., "Airborne investigations of clear air turbulence with optical radar," University of Michigan, Progress Rept., Office Naval Research Contract NONr 1224(51), 1965.

8. Goyer, G. G., and Watson, R., "The laser and its application to meteorology," *Bull. Am. Meteorol. Soc.*, vol. 44, no. 9, p. 564, 1963.

9. Nanevicz, J. E., Vance, E. F., and Serebreny, S. M., "Correlation between clear-air turbulence and aircraft electrical activity," SRI Final Rept., Air Force Cambridge Research Laboratories, Contract AF 19(628)-3308, 1965.

10. Merritt, E. A., and Wexler, R., "Radiometric detection of clear-air turbulence," presented at 5th Conference on Applied Meteorology, American Meteorological Society, Atlantic City, N.J., 1964.

11. McLean, G. S., "Cloud distribution in the vicinity of jet-streams," *Bull. Am. Meteorol. Soc.*, vol. 38, p. 579, Dec. 1957.

12. Kadlec, P. W., "A study of flight conditions associated with jet stream cirrus, atmospheric temperature change and windshear turbulence," Eastern Airlines Final Rept., U.S. Weather Bureau Contract Cwb 10674, 1964.

13. Serebreny, S. M., and Wiegman, E. J., "Characteristic properties of the jet stream over the Pacific," Pan American World Airways, Tech. Rept. No. 1, Navy Contract N 189-S 90981, 1953.

14. Serebreny, S. M., "Some major considerations of clear-air turbulence," presented at Flight Safety Foundation International Air Safety Seminar, Chandler, Ariz., 1960.

15. Kadlec, P. W., "Flight data analysis of the relationship between atmospheric temperature change and clear-air turbulence," Eastern Airlines Final Rept., U.S. Weather Bureau Contract Cwb 10888, 1965.

16. MacLean, G. S., "Air investigation into the use of temperature gradient as an in-flight warning of impending clear-air turbulence," Air Force Cambridge Research Laboratories AFCRL 65117, Environmental Research Paper No. 85, 1965.

17. Norman, S. M., and Macoy, N. H., "Remote detection of clear-air turbulence by means of an airborne infrared system," presented at AIAA 2nd Annual Meeting and Technical Demonstration, San Francisco, Calif., July 1965.

18. Collis, R. T. H., "Lidar detection of CAT," *Astron. and Aeron.*, p. 52, Dec. 1964.

19. Oliver, V. J., Anderson, R. K., and Ferguson, E. W., "Some examples of the detection of jet-streams from Tiros photographs," *Monthly Weather Rev.*, vol. 22, no. 10, p. 441, 1964.

20. Wiegman, E. J., "The distribution of clear-air turbulence reports and cloud patterns as seen in satellite photographs," SRI Final Rept., U.S. Weather Bureau Contract Cwb 10791, 1965.

BIBLIOGRAPHY

Proc. 1966 ION/SAE National Air Meeting on Clear Air Turbulence, Society of Automotive Engineers, New York, N.Y.

Colson, DeVer, "Summary of high level turbulence over United States," *Monthly Weather Rev.*, vol. 91, Oct.–Dec. 1963.

Endlich, R. M., "The mesoscale structure of some regions of clear-air turbulence," *J. Appl. Meteorol.*, vol. 3, June 1964.

Fusca, James A., "Clear air turbulence," *Space Aeron.*, p. 6, Aug. 1964.

Harrison, Henry T., "The mountain wave," NASA Contractor Rept. CR 315, Oct. 1965.

Hildreth, W. W., *et al.*, "High altitude clear air turbulence," Lockheed-California Rept., Air Force Systems Command Contract ASPTDR 63-145, June 1963.

Reiter, E. R., *Jet-Stream Meteorology*. Chicago: Univ. of Chicago Press, 1963.

Reiter, E. R., "CAT and SCAT," *Astron. and Aeron.*, p. 60, May 1964.

Reiter, E. R., and Nania, A., "Jet-stream structure and clear-air turbulence (CAT)," *J. Appl. Meteorol.*, vol. 3, p. 247, June 1964.

Steiner, R., "A review of NASA high altitude clear-air turbulence sampling programs," AIAA Paper 65-13, Jan. 1965.

Young, Warren R., "Turbulence—hidden giant in the sky," *Life*, p. 86, Dec. 18, 1964.

Human factors in engineering

Part II—

Advanced man–machine systems and concepts

The advanced man–machine systems—on-line man–computer interaction, manned space systems, satellite communications, advanced displays—none of which existed 20 years ago, present challenging new problems to human factors investigators

Nilo Lindgren *Staff Writer*

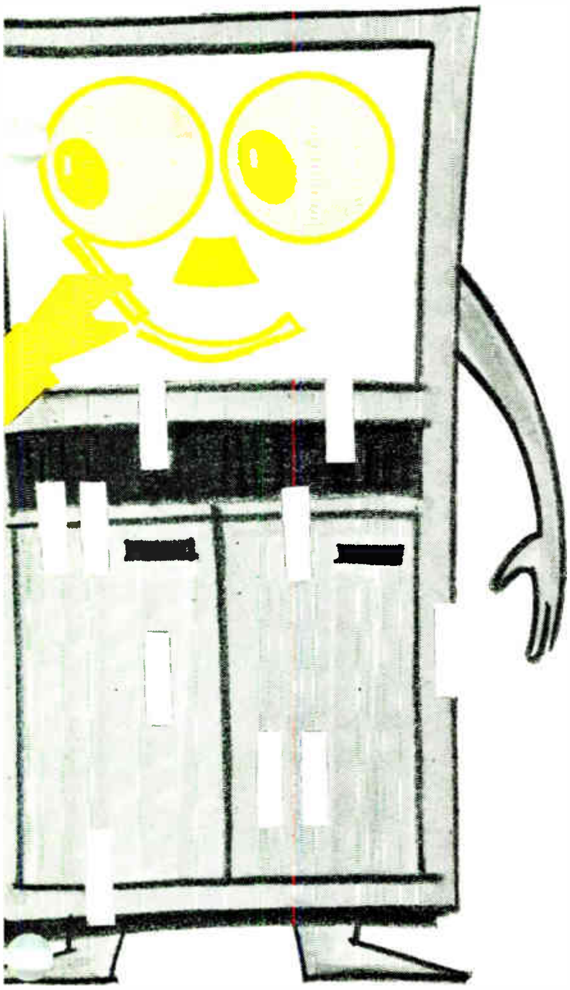
This second and concluding article scans over some of the new kinds of human factors problems that have emerged as a result of recent technological advances. The article discusses aspects of these new problems in the following major areas: man–computer interaction (psychological problems related to on-line usage, the need for more "natural" interaction languages, graphical communication); man–vehicle control and guidance systems (better models of human controller, novel systems); communications (computers in telephone information services, satellite relays); advanced space systems ("learning" and display problems related to incorporation of computers in prelaunch activities). Last, a recent study in human information processing is cited that points up a general problem—the need for rational optimal models of human functioning.

The folk wisdom, "Give a little boy a hammer and suddenly everything needs pounding," has its modern counterpart in "Give a little boy a computer and suddenly everything needs computing." It is the story of our age. In fact, many of the advanced man–machine systems with which human factors investigators must now deal involve computers in some form—as "workers" with whom communication is necessary, as adjuncts or components of huge complex systems, as tools for the simulation of devices and of aspects of individual and social behavior, and so on.

Man–computer interaction

There can be little dispute about the statement that the computer is the most fascinating machine man now interacts with. It is to the generation of this mid-century what the heavier power machines were to the mid-19th century. It has crashed in on the modern consciousness with the force of a silent atomic explosion, liberating terrific energy and setting many investigators off on a new intellectual adventure—namely, how to come to terms with the computer and make it a more natural ally; in other words, how to bring the functions of man and computer into a truly symbiotic association. As Dr. J. C. R. Licklider recently described that ally, to those who have been displaced by the computer, he "appears in the form of a superhumanly fast, accurate, methodical, tireless worker. To most of those who have prepared the procedures for computers to execute, [he] appears in the form of such a worker who is also illiterate, perversely literal, devoid of initiative, and so high salaried that one must strive continually to keep him from running out of work."[1]

For a normal human to get along with such a creature would appear to demand an inhuman patience and perseverance. But until very recently, computers were still so exorbitantly costly in relation to human labor, and computer designers were scrambling so fast to keep up with technological developments, pressed on by sharply competitive demands, that the inhuman demand had to be tolerated to a large extent, and was not attacked in any logical, organized fashion. Now, however, the picture has changed. In general, computers are becoming so economical, relatively speaking, that ways must be found of making better use of the people who use the computers. In short, the determination of what the human factors problems are in man–computer problem solving and



information processing has been drawn into the center of the target. Although it is far beyond the scope of this slight survey to go deeply into the fascinating psychological or human factors problems inherent in the man–computer interaction situation, we shall at least point out some of the present major directions.

Just six years ago, in the lead article of the very first issue of the IRE Transactions on Human Factors in Electronics, it was also Dr. Licklider who set forth the broad problems of "man–computer symbiosis," a term and a paper that have since been much cited.[2] The major problems he described then—the "several hurdles that stand between the nonsymbiotic present and the anticipated symbiotic future"—are still with us, although they do not loom as large because in the interim much intensive work has gone into whittling them down.

Licklider blocked out the problems into a number of areas: the speed–cost mismatch between man and computer; the physical interface, that is, the console with its displays and controls; the language mismatch between man and computer. In each of these areas, there are profound problems that excite the interest and concern of human factors investigators. Furthermore, the problems are of a radically new order, for they involve the analysis of man's mental processes to a far greater degree than has previously been necessary.

Time-sharing systems. The fact that "computers go very fast, men go very slow" is at the crux of the speed–cost mismatch. To allow a single man to monopolize the time of a large, expensive computer is simply not economical—this would be like giving over an entire electric power plant for the use of one individual—and the larger society is too tyrannical in its economic structure to tolerate such a situation, even for "a very creative man." And yet very creative individuals do feel that this is precisely the best way to interact with computers—to engage in prolonged and intimate "conversation" with them. By contrast, in most present man–computer exchanges, there exists a long interval between experimental steps because of the need to reduce all communications to typed "microstatements," a process that Dr. Ivan Sutherland compares to the operation of the postal system.[3] He says, "We have been writing letters to rather than conferring with our computers." This correspondence system has involved an arduous translation process between our natural language and the programming language. Dr. George A. Miller, the very well known authority in what might be called the field of "psycholinguistics of man–computer interaction,"[1] says of the binary language of computers: "It would be difficult to design a language that was more difficult or unnatural for human beings to learn to use."[4] This feature, he believes, poses a major obstacle to man–computer symbiosis. In effect, the language mismatch compounds the speed–cost mismatch problem.

One answer to this essentially multivalued problem of nonsymbiosis, and one that is interlocked with the "step-display-look" approach that is emerging, has been the implementation of the concept of time-sharing whereby many users at different locations are able to tap the same central computer "power" source. There are already in existence in the United States a number of large, experimental time-sharing systems; probably the best known is the Project MAC ("machine-aided cognition" and "multiple-access computer") system at the Massachusetts Institute of Technology.[5, 6]

In time-sharing systems, the speed mismatch between man and computer is turned to advantage. The computer can work on one "customer's" problem for a short time, switch to the next user, then to the next, and so on, and come back to work on the first user's problem so quickly that it appears as though the computer has been working continuously on his problem alone.[7]

Perhaps it should not be surprising that in such systems "software" and human factors problems of an entirely new sort are encountered and, typically, the discussers of the time-sharing systems seem disposed to prefer the elucidation of the software problems—i.e., these problems are more interesting and crucial in the successful development of these informational "public utilities."

Generally speaking, it is now both necessary and possible to determine just what a person does when he interacts with a computer. In former days, Dr. Licklider points out, fantastically little attention was given to how programming was done.[8] How a programmer worked—his mental processes, his methods for planning an attack on a problem—were hidden in obscurity. However, in the on-line situation, it is possible to find out exactly how he proceeds, step by step. What he does becomes overt. Thus, the habits, procedures, and needs of many on-line users are now the object of intensive investigation. At M.I.T., for instance, there is in progress a special Technical Information Project that centers around the MAC time-shared computer, and which among other objectives is investigating the role of human factors in literature search.[9] The MAC facility consists of a central computer with about 100 remote consoles that are available to about 400 people (only about 30 can use MAC at the same time). The interaction between user and the system proceeds by means of a language very close to natural English. The user "operations are observed and monitored by means of test procedures and feedback. The information thus obtained will guide the future evolution of the system by suggesting component and procedural innovations."[9]

That on-line users of time-shared computer systems are entering a "new world" quickly becomes evident from the amusing stories of what is happening to these users. As Clark Weissman points out in a discussion of the experiences of the System Development Corporation (SDC) Time-Sharing System (TSS),[10] there can be a great deal of helplessness and frustration in on-line programming efforts when there are communication failures owing to the user's ignorance of proper procedures, and he loses his programs. Some of the stories have positively Chaplinesque overtones: users drop in on each other's programs, get lost, wipe out another user's programs by accident, peer over each other's shoulder, and what not. It all sounds fascinating and frustrating—a world in which one is likely to run into wandering ghosts and "lost souls."

In one of Weissman's anecdotes, the computer operators received this cryptic message from a remote user: "I'm alone; please decipher the above error message." And, of course, the error message was only on the user's terminal. At another time, during an important demonstration of one system setup, the demonstration was punctuated by a message announcing the score of the second game of the World Series. The system users are vulnerable to messages from "jokers" or from users who are "careless," ignorant, or confused. "When computers were 'downstairs' and proper experts were 'around the corner,'" writes Weissman, such failures "might not have been as disconcerting to the off-line user. But, to the remote time-sharing user, possibly thousands of miles from the computer or the expert, communication in its broadest sense is of paramount importance."[10]

Of course, the message in such tales for those who look on computers with less than unbounded enthusiasm is to wait until the avant-garde investigators have perfected these potentially marvelous information assistants to the point where we can plug into them from a console by our office desks and converse with them in our native tongue. That day, says Licklider, though it may sound a bit fantastic now, is coming.

The language mismatch problem. In his 1960 paper, Licklider cited the "basic dissimilarity between human languages and computer languages" as being probably the biggest impediment to man–computer symbiosis. Today, despite the advances, the language mismatch remains a serious obstacle. The realization of computer systems that will understand natural speech, which stands as a kind of ultimate goal, seems far off.[11]

The dissimilarity between man and computer languages has been the source of new psychological problems, problems that were "scarcely dreamed of twenty years ago." What has happened is that there has been a gradual shift in emphasis from computer-oriented and procedure-oriented languages to problem-oriented languages more akin to natural language and human users. The psychological principles guiding this evolutionary shift are traced out by Dr. George A. Miller in one of his (as is usual with him) marvelously lucid papers.[4] He explicates the "psychology" behind the introduction of coding, relative addressing, and the hierarchical structuring of programs, in which the development of subroutines (by means of which programmers could borrow whole chunks of programs from one another) led to the concept of the compiler. With compilers it has been possible to construct languages that are much closer to the way a man thinks (FORTRAN, COBOL, ALGOL, IPL, COMIT, etc.). Essentially, the programmer selects from a library of subroutines the ones he needs, and the compiler translates it for him into the microsteps of machine language. Thus, the man is able to plan his program in larger hierarchical chunks that are more natural to him, and the machine does the dogwork translation into binary microsteps more natural to it.

The development of compilers has had a revolutionary impact for it brought the interface between machine and man a giant step closer to the man, and in the same step brought whole new classes of users[12] into interaction with computers—workers in a wide variety of disciplines for whom the "linguistic" barrier had previously been more or less insurmountable.

Although much of this evolution from machine-oriented to problem- and man-oriented languages proceeded empirically, without it occurring to the programmers to call in the psychologists and the human factors people, the problems they were coping with were essentially psychological or human factors problems. Nowadays, with the heightened sophistication of man computer interaction, the workers from these disciplines are very much in the act. They feel that the trend toward more natural languages of communication between man and machine will continue, although they are not so sure just how this will come about. Miller says, "We are today at a point where the subtler aspects of human language and human psychology are beginning to intrude, and we are not clear enough in our understanding of human languages and human language users to be confident of what step to take next in reducing the gap"[4] between man and machine. He cites some of the features of natural language that might be introduced into compilers yet to be developed—e.g., recursion and greater use of context—but in this direction, the investigators are brought to the brink of profound problems involving them in, among other things, the new theoretical studies in linguistics.[11]

Another direction being taken by some investigators is the study of existing programming languages to discover what is common in them, with the aim of developing a homogeneous language that will understand all the languages in the computer libraries, but this work is in a primitive state. Dr. Licklider speculates that there "is not going to be a *single* language"[8] in the man–computer interaction process.

Still another tack is taken by Allen Newell. He stresses the need for investigating the "planning process" in man[13]—how man goes about getting his ideas for solving problems—so that the computers can be made more "intelligent" in the interaction process. Newell points out that the problem is one of psychology. "First," he says, "we must understand the structure of human plans: What is it that the human has when he has the 'idea of solution'? Only when considerable is known about this can we understand what the computer must do to take a plan as input and carry it out. We need to devise a language of plans sufficiently similar to the 'natural language' of human planning so that the man is able to utter his plan to the computer. If this language is too difficult, then the human will face a translation task to put his plans into the language. Thus we need to know in considerable detail what information is available in a human plan and how these plans are reduced to action."[13] As a consequence, he says, "the work on artificial intelligence becomes directly related to the work on man–computer communication." And, he says, "if there is one final conclusion to be drawn . . . it is the importance of psychological investigation to the problem of man–computer communication. The task is not one for computer technology alone."[13]

Rather than having come to the end of a line, then, it seems that those who are in search of a more intimate man–computer symbiosis have broached some "beautiful" new problems. In general terms, one can conclude that the next order of business in man–computer interaction problems is to bring better organization into the languages, theories, methods, and devices of programming. Significantly, Dr. Ivan Sutherland, now director for information processes technology at the Advanced Research Projects Agency (and successor to Dr. Licklider in the post), says that his office is stressing the software area in its many contracts aimed at finding ways in which men can make better use of computers.[3]

The physical interface. In discussing the language mismatch problem, we have failed thus far to point out a most significant advance that may be viewed as a complementary hardware and language development—namely, man–machine graphical communication devices. About 1960, an interest began to build up in developing computer display systems whereby man and computer could "converse rapidly through the medium of line drawings."[14] The first really striking advance in this direction was "Sketchpad" (at the M.I.T. Lincoln Laboratory), the work of Ivan Sutherland.[14] With Sketchpad, it was possible to make two-dimensional sketches with a light pen directly on a computer CRT display, to modify and move parts of the drawing as one wished, while preserving the topology of the drawing, and more important, to carry out computations on the figures so drawn. These capabilities showed that such graphical language devices could serve as tools of extraordinary power in design and other applications. For instance, Sketchpad could be used to draw basic bridge shapes, which were then subjected to various loading conditions, so that the stresses and strains in the constituent members could be computed automatically. The Sketchpad system was subsequently extended by T. E. Johnson to permit the user to work with three-dimensional figures. These graphical communication systems have clearly just opened a Pandora's box, and now many organizations are working on such systems. Another very well known one, for instance, is the RAND tablet.[15] Dr. Licklider reports that IBM is now in the process of developing five different advanced display systems,[8] although details about them are still in the realm of trade secrets.

A most extraordinary and sophisticated system is the

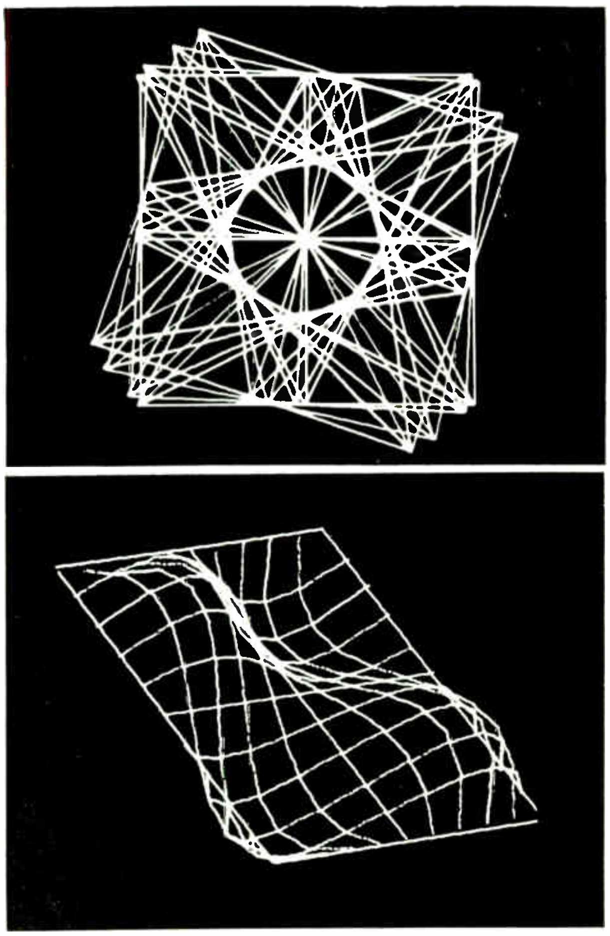
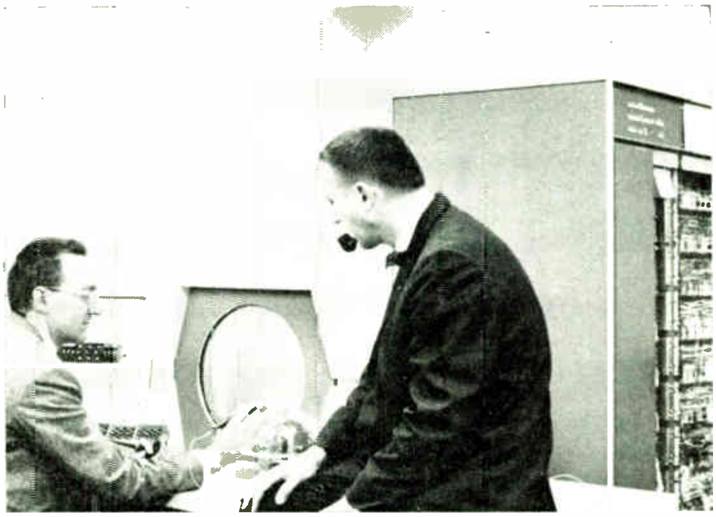


Fig. 1. Computer-controlled displays. Top figure is multiple exposure. Bottom figure is somewhat similar to that shown in different projections on cover.

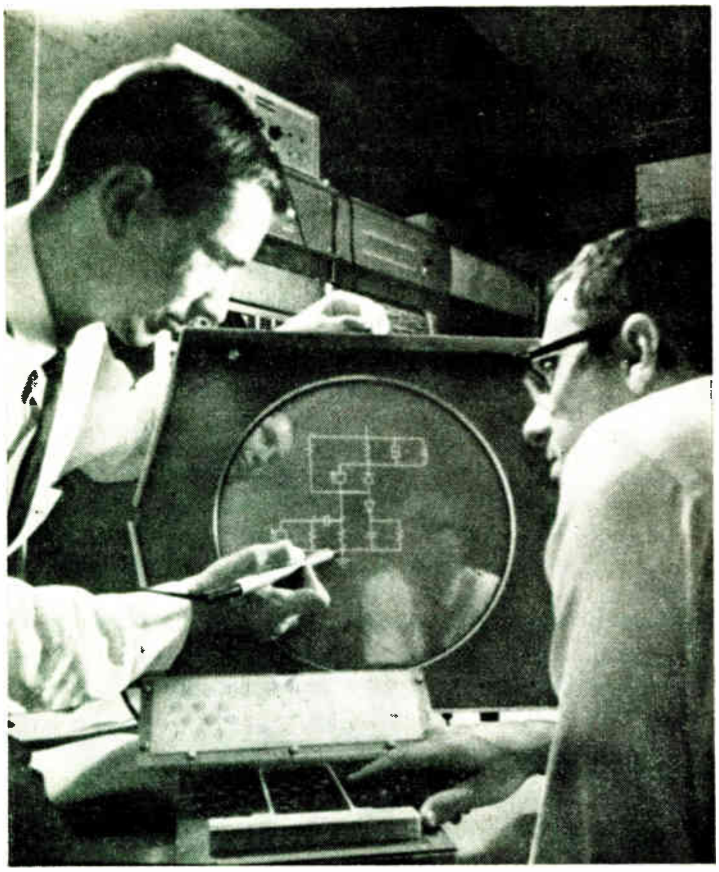
Fig. 2. The ESL display console, being developed at M.I.T., is a specialized computer that automatically converts three-dimensional drawing commands into arbitrary two-dimensional projections as in Fig. 1. The three-dimensional rate-control joy stick (crystal ball, near CRT) allows the user to rotate (in real time) the figure appearing on the scope face in almost any fashion. Light-pen tracking is fully automatic.



M.I.T. Electronic Systems Laboratory display console, which since January 1964 has been in operation in the Project MAC Time-Sharing System.[16] With this specialized computer console, it is possible to make directly three-dimensional drawings and convert them into arbitrary two-dimensional projections, to rotate them in real time, to translate them, to change their scale, and so on (see Fig. 1). Figure 2 shows the console. The three-axis globe in front of the scope is used to rotate and twist the drawing about as desired, so that one can get a feeling of the figure from all sides (the SPECTRUM cover shows different projections of a complex form). John Ward, assistant director of ESL, points out that this console, which was designed with special attention to the needs of computer-aided design under the restriction of a time-sharing system, is so flexible that it can be used for many applications. For instance, it is being used in the design and evaluation of ship hull forms.[17] In this setup, three-dimensional hull surfaces displayed on the CRT can be altered in a few seconds by typed-in changes in parameters, and can be rotated to any desired viewing angle for closer study.

Computer-aided circuit design. Another sophisticated graphical-type research program that uses the Project MAC computer involves the on-line simulation of electrical networks.[18, 19] In this program, being carried forward by Prof. M. L. Dertouzos and his colleagues at M.I.T., an electronic circuit designer interacts directly with the computer through a typewriter and CRT graphical input–output equipment. The designer communicates his circuit configuration an element at a time to the computer by placing the light pen on the CRT, as shown in Fig. 3, and by pressing the appropriate element button. Through this process, he can compose on the screen any circuit—including resistors, inductors, capacitors, diodes, and tunnel diodes—without knowledge of computer programming; he can then ask the computer to analyze the circuit by a typed command. Prof. Dertouzos points out that a crucial point in this approach is that it involves simulation rather than analysis. The circuit simulation is achieved by representing each electrical element by an appropriate block of storage registers and by embedding the laws of circuit theory within the computer. Then, for any given excitation, the computer "walks the circuit through" its behavior much as a real circuit operates. A most significant consequence of this man–machine interaction process is the short turnaround time, of the order of seconds, between a modification of the circuit by the designer and the computer response, whereas in the past this time was of the order of days.[20] The moral of this advance should wallop design engineers on the head—i.e., the 99.99 percent of engineers who are designing circuits without on-line graphical-language facilities are, in one sense, already "living in the past."

Fig. 3. Prof. M. L. Dertouzos and C. W. Therrien show how with a light pen and their new computer program tied in to Project MAC, they are able to invent, analyze, and display immediately electrical networks. Computer-aided circuit design should have profound repercussions in the electronics art.



As Prof. Dertouzos notes, "this reduction of delay in the design loop has obvious economical and technical repercussions."[20] He is careful to point out, however, that breadboarding and testing will not be eliminated by this design process, but all the same he is tempted into speculating on the future of man–machine interaction along these lines: "It is not difficult to imagine a not so distant time when a designer seated before a console attains, after some man–machine dialog, a circuit design he considers satisfactory. He then asks the computer to search component-manufacturers' tapes for components that will meet circuit and designer requirements. Having done so, he then commands numerically controlled tools to fabricate a prototype of the circuit, while all related financial transactions are completed automatically."[20]

And the exciting part about Dertouzos' projection is that it represents but *one* meaning of that Anglo-Saxon–Latin–Greek mouthful, "man–computer symbiosis." For he goes on: "What is harder to imagine and perhaps more challenging to speculate about is the contribution of the man–machine dialog to the *creative* or *conceptual* stages of circuit development. Certain mechanisms of creative design are observable and to some extent comprehensible. For example, a designer confronted with the design of a circuit to meet certain specifications resorts to his own memory for 'past experience' in designs of 'this sort.' He may then postulate a first-order circuit configuration that should accomplish the desired task. In evaluating his 'experience' or ideas suggested to him by the problem, the designer often uses some measure of goodness to reject or accept the proposed configuration. In decomposing a large system he may start from the output specifications and advance toward the input, or vice versa. We see that from a qualitative point of view, the notion of 'experience' is mechanizable by an organized memory. Evaluations of conceived configurations or 'good moves' are mechanizable as heuristic weightings in the decision process; while the conception of heretofore unknown configurations may be accomplished by a somewhat constrained random search."[20]

The possibilities and experiments with on-line facilities, with such a degree of man–machine symbiosis, have already become so numerous that no adequate account can be given. A glimpse of the breathtaking spectrum of such studies, embracing wide classes of users in both the physical and social sciences, can be obtained from a Project MAC Progress Report.[21] To scan but a few of the others: at Bolt Beranek and Newman, Dr. Jerry Elkind is heading up work on teaching through man–computer interaction[22] and is carrying out on-line control of human factors experiments;[22] at the MITRE Corporation, Dr. Edward Bennett and his colleagues have been developing an experimental on-line information control system called AESOP, which serves as a prototype of a class of management or command information systems capable of giving the members of the using organization (all the way up to the highest executives) immense on-line control over system performance;[23] the question of decision making in man–machine systems,[24] taken up in a current special issue of the IEEE TRANSACTIONS ON HUMAN FACTORS IN ELECTRONICS, is reviewed in this month's "Scanning the issues," page 120.

In discussing this exploding field, released by the genie of man–computer interaction, one must bring oneself to an artificial stop. So many concepts and possibilities are now emerging that one must ask: "Just how far can this thing go?" Dr. Sutherland gives us one view. He has speculated that the "ultimate displays" may allow a user to draw abstract systems that operate under any laws he wishes to impose, and then, through special television headsets, he might "feel" his way into these abstract environments. Such fantastic systems might bring about invaluable discoveries by investigators who fail now to understand certain physical phenomena simply because they cannot "see" the underlying dynamics of the phenomena. Although such ultimate systems (one can hardly describe them as mere man–machine "interfaces") are still far in the future, creative people are now thinking about them seriously for the simple reason that, owing to the developments of the past few years, they are no longer impossible.

Man–vehicle control and guidance systems

Unlike the field of man–computer relationships, the study of the human factors involved in the manual control, guidance, and stabilization of vehicles has been going on for a long time, and for all kinds of vehicles—motorbikes, automobiles, aircraft, and now spacecraft. Although formal study of the "controlling function" of the human operator began about 1900, well before the development of formal theories of servomechanisms,[25] and although every aerospace organization (as well as just about everybody else) can boast whole bibliographies of manual control studies, we can include here only a few points about recent investigations in this rich and "venerable" domain of human factors.

Traditionally, for each vehicle design, the designer has had to decide, on as *rational* a basis as possible, the degree of the guidance, control, and stabilization functions to be assigned to the man and to the machine. In the past, however, this goal has been easier to postulate than to realize. Much less has been known about the engineering characteristics of human operators than about the engineering characteristics of automatic guidance and control systems. As Prof. T. B. Sheridan of M.I.T. has so picturesquely put it: "It must be appreciated that in dealing with the human operator we are dealing with the 'blackest' of 'black boxes' and the most complicated of physical systems."[25] Because "man is a nonlinear, discrete, adaptive, time-varying, statistical controller with a 'mind of his own,' "[26] it has been difficult to develop adequate models for human operators, so that one of the major avenues to man–vehicle designs has been the simulation of vehicle control for varying degrees of human operator participation. There can hardly be an engineer alive who hasn't seen at least one form of such simulators. Thus, in the evolution of manually controlled vehicles, designers were very much dependent on subjective opinion ratings expressed by the human operators as to the vehicle's handling qualities. The aura of the test pilot did not grow through public relations gimmickry—an engineer could not know what the man–vehicle situation would be really like until the pilot took off, flew the vehicle, and came back to tell him. Thus, there have been assembled extensive catalogs of vehicle dynamic parameters given as functions of pilot ratings,[27] which no doubt will continue to be used. However, as Duane McRuer and Dunstan Graham (both of Systems Technology, Inc.), and Ezra Krendel and William Reisener, Jr. (of the Franklin Institute) point out in a recent report, "In a fundamental sense, these catalogs are only reports of specific test results—they fail to explain adequately the mutual interactions between the pilot and the vehicle, and they are difficult, if not impossible, to extrapolate to new situations and novel vehicle characteristics."[27] Accordingly, of late there have been increasingly extensive efforts to generate better theoretical models of human operator dynamic characteristics, especially in view of the increasing space work in which man–vehicle design failures would be more expensive and devastating than anything that has gone before.

The general situation is summed up by Prof. L. R. Young of M.I.T. as follows. He points out that psychological testing has been very fruitful in that it has led to a certain level in devising models for the human operator. Now, however, "for some of the adaptive and optimization control problems, which the human operator involves, we have little appropriate mathematics available to treat them. I think the time has come for the human factors people to look more intensively into mathematical possibilities. We must build up some theoretical backlog strictly for the manual control problems rather than continue to live off what has been ready-made for us in automatic control."[28]

Most analytical studies of human operators have viewed the man as a simple controller in a simple closed-loop operation—the man gets information on the vehicle motion through external sensors, such as gyroscopes, radar, etc., and through his observation of the error between his actual and reference orientation he stabilizes his craft by means of a control stick or whatever. For this type of closed-loop system, McRuer and Krendel developed in the late fifties a set of quasilinear transfer functions that quite accurately predicted the control characteristics of the human operator over a wide range of inputs and vehicle control dynamics.[29] This pilot dynamics model, in fact, found considerable application, and just recently these same investigators completed a comprehensive five-year program of developing better pilot models for use in handling qualities and manual control system analyses.[27] These new models, which are much more precise than the circa 1960 model, are expected to have far-reaching applications in the future. They embody, among other things, new knowledge that "should have significant impact both on the content and nature of the information displayed to the pilot and on the design of the manipulative devices with which the pilot imparts his desires to the vehicle."[27]

Other very extensive investigations of control models of the human operator have been going forward at the M.I.T. Man–Vehicle Control Laboratory, which is part of the Center for Space Research. They have been working on, among other matters, the extension of models of humans in orientation control tasks to include the description of the role of the nonvisual motion sensors (particularly the tactile and vestibular sensors).[26, 30] Figure 4, which represents a general block diagram of the man–vehicle control problem, shows the major areas now under research at M.I.T. It should be remarked in passing that a very good, readable, general discussion of the many questions involved in the assignment of roles for men and machines in space systems appears in a recent paper by C. S. Draper, H. P. Whitaker, and L. R. Young, all of the M.I.T. Department of Aeronautics and Astronautics.[31]

Among the more earthbound, but sophisticated, studies in progress at M.I.T. is that of a stabilized motorbike, which is being used as a test vehicle for dynamic orientation. In these studies, which have been going on under Prof. Yao Tzu Li's direction for the past few years, there are two aims: one is to find out more about how a human rides a bicycle (that is, to identify the rider's control function); and the second is to build an autopilot, an adaptive automatic stabilization device that will duplicate the human rider's performance.[28]

The bicycle project is but one of a broader program of studies of the adaptive functions of man in vehicle control systems. In many respects, adaptive control systems attempt to imitate the adaptability of humans. (M.I.T.'s Prof. T. B. Sheridan is credited with the first study of the adaptive characteristics of the human controller—in 1960.) But, Professor Li and his colleagues point up the irony that it was not the versatility of human adaptive control but rather its limitations that led to the great surge in the development of automatic adaptive control systems in the last few years. In general, the "human outshines the automatic system with his huge capacity of open loop or programmed control; but he lacks the capacity and speed for making on-line computation which is needed in the operation of active continuous adaptive systems."[32] Thus, "the objective of adaptive control systems is to reduce the range of adaptation required of the human operator."[32]

For space workers only. One of the most novel human-operator–vehicle systems now under investigation is that of a completely foot-controlled platform that free-floating space workers might eventually use.[33] The study into this use of the human balancing reflex for stabilization and control of vehicles in one-*g* and sub-*g* environments is going on in the Life Sciences Section of the Grumman Aircraft Research Department under the sponsorship of NASA. The research embraces triple aims: the use of the foot-balancing concept in one-*g* vehicles; its application to zero-*g* conditions; and basic studies of the human balancing mechanism. The appeal of a jet-propelled foot-controlled space scooter, says Thomas Keller, head of the Life Sciences Section, is that it makes use of a neuromuscular response that has already become a reflex. "We feel," he says, "that the foot-control vehicle would be far superior to the Buck Rogers type space-pack."[34] Theoretically, the big advantage is that the space worker would require no special control training, and

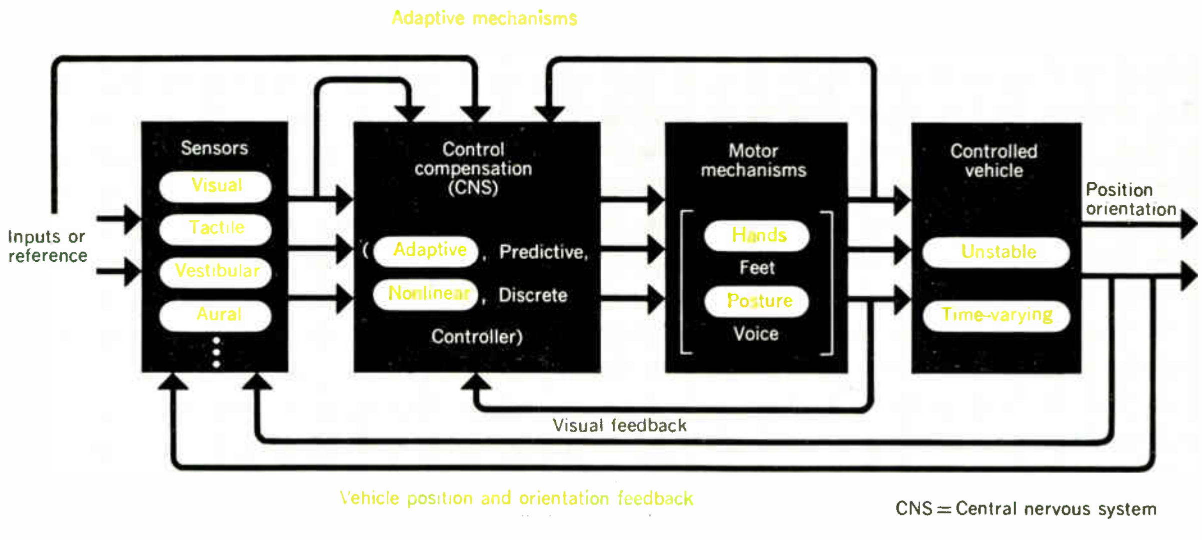


Fig. 4. New investigations going on at M.I.T.'s Man-Vehicle Control Laboratory aimed at extending control models of the human operator in dynamic space orientation are indicated here by circled items.

his control reflex would be better under conditions of stress. In addition, such a scooter would leave the space worker's hands free. Figure 5 shows a zero-g simulator that is being used in the research project.

New aircraft landing aid. Just to relieve any impression that the reader may be getting that all modern human factors problems are concerned with complex systems, we shall cite an at least deceptively simple development that has been under way at the U.S. Naval Research Laboratory.[35] It is an optical-geometrical system, called the Rainbow Optical Landing Aid, which its designers (H. P. Birmingham, A. W. Baldwin, and Miss Barbour Lee Perry of NRL) hope will increase the safety of landing high-performance aircraft aboard aircraft carriers in nighttime operations, a hazardous feat at best.

The new landing aid is designed to provide a pilot with information about the appropriateness of his descent during his landing approach, information which present landing aids do not adequately provide (i.e., accident rates have been very high). Altitude-rate or sink-rate error information is transmitted to the incoming pilot by a dynamic tricolor light beam (see Fig. 6), which by its color sequence tells him not only his angular error from glide path, but which is also coded to tell him continuously whether or not his angular rate of descent is too small, too large, or just right. This latter quality of the landing-control system—the inclusion of "lead" or error-rate information—is important, for it reduces the difficulty of the control task required of the pilot. (In NRL terminology, this represents the inclusion of one "quickening" term. Any modification to a closed-loop man-machine control system that reduces the need for the human operator to perform analog differentiations is defined as quickening. The concept has been applied in many human factors designs.)

Human factors in communication systems

We have seen something of the human factors and psychological problems involved in communications between man and machine. The communications between man and man through machines as intermediaries brings different types of human factors into the foreground of attention. In this respect, the problems of voice communication through telephone systems are especially interesting.

Some of the investigations carried out at Bell Laboratories, for instance, where a Human Factors Research Department was established in 1948, include: the measurement of human channel transmission characteristics; studies of telephone users' habits and desires; problems of echoes and time delay on telephone conversations; users' responses to push-button telephone keysets; verbal behavior in retrieving telephone information; how customers use number services; and many others.

More recently, with the experimental study of transmission through satellites, and with the prospect of computer services in the telephone system,[36] new human factors problems have come to the fore.

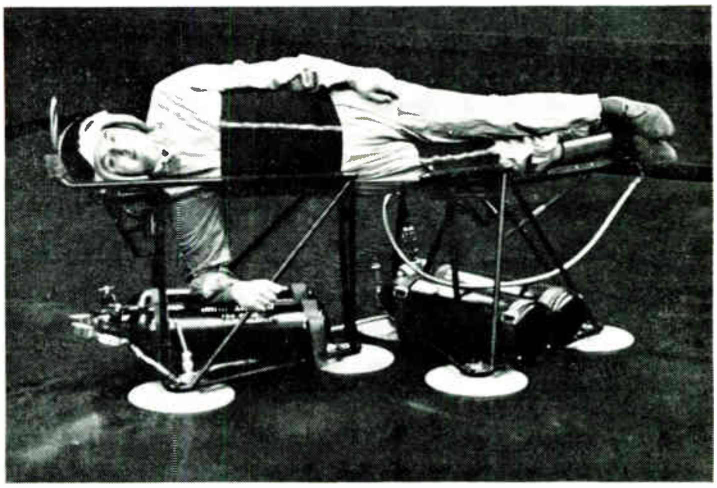
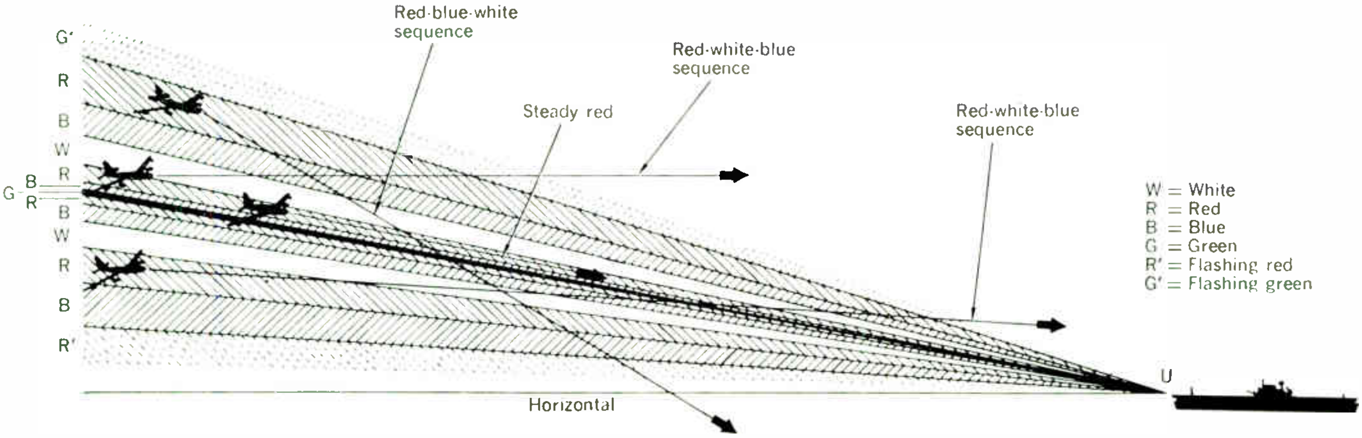


Fig. 5. (Right) Zero-g simulator at Grumman Aircraft Engineering Corporation's Research Department. The man is turned on his side so as to eliminate the gravity vector. This research is aimed at the development of a small space platform that a "space worker" might control by means of his foot-balancing reflex, thus obviating special control training.

Fig. 6. (Below) Rainbow Optical Landing Aid, being developed at U.S. Naval Research Laboratory, simplifies pilot's tasks in nighttime carrier landings. The dynamic array of color-coded light beams guides the pilot from any path of approach into the correct steady-green glide path. The Rainbow unit (U) would be space stabilized to compensate for ship pitch, roll, and heave movements.



One of the problems involves the use of computers to store telephone numbers. In the proposed system, when an information operator receives a customer request, she keys the request into a console by pushing buttons, and a set of numbers appears on a display before her from which she selects the number fitting the request. However, since computers are so expensive, one problem being investigated by the Bell human factors group is whether or not this method is enough faster to compete with telephone books on cost. The design of the console input and output—the various keys and display characteristics—constitutes one area of the problem, but it is, according to John Karlin, head of the Bell group, the least important. The greater share of work is that of looking into the best search strategies for the computer. A major problem arises because the computer cannot think as a human operator can. If given information is incomplete or inaccurate in any way, the computer will believe what the operator tells it. Accordingly, the really tough problem is to devise an appropriate search strategy and to determine how complete and accurate the information the operator gives the computer must be. The big cost savings would come from determining how much of a customer's request the operator should key into the computer. Karlin reports that if all goes well in the study, they hope to have the final results this summer.[37]

The Bell project also presents an interesting sidelight. Human factors engineers always stress the importance of being in on design studies as early as possible so that their investigations can properly influence the final system design. In the Bell case, the engineering department is not even going to begin to implement the computer lookup system until they have seen the complete results of the human factors study. . . . which amounts to the realization of a human factors dream.

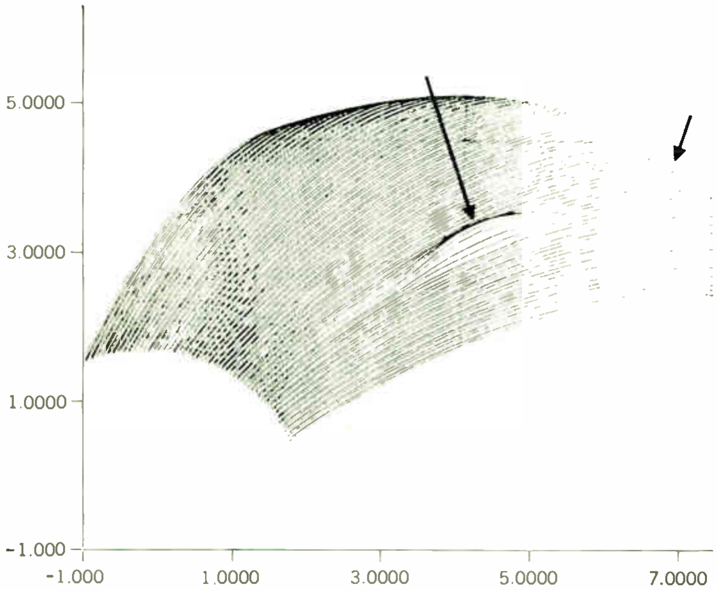


Fig. 7. On this kind of display surface (developed by means of computer programs), errors or abnormalities show up instantly as "bulges." Information displayed here is equivalent to that contained in 42 000 eight-digit numbers. If the same information were in "page" format, it would take a man hours of analysis to discover the errors.

The other big human factors study at Bell involves the Early Bird satellite, and looks into the question of how well people can talk over a satellite connection. Over a period of some months now, normal, commercial telephone traffic has been relayed over Early Bird as well as over cable. After each call, a subscriber is queried about his experience with the connection (he does not know which system handled his call). Results of these experiments will be announced within the next two to three months.[37]

Advanced space systems

More than any other man-machine system, the advanced manned spacecraft missions due to come off before the end of this decade represent the most crucial problems yet encountered. They have conscripted human factors efforts across the entire spectrum—man-vehicle control dynamics, advanced displays, computer backup, sophisticated real-time mission simulations, communications, life support systems, atypical environmental effects, allocation of tasks to several crew members, ground support equipment. There is almost no way of getting a human measure of the effort involved. They say 30 billion dollars worth, but what does that mean in human terms? Grumman Aircraft, which a few years ago had no human factors people around to speak of, now has a permanent staff of 50 people, 30 of them working on problems associated with LEM (Lunar Excursion Module) alone. And LEM itself, as big and fraught with problems as it is, sits like a pea atop the colossal Saturn vehicle that is to start it on its mission. But size alone does not convey very much meaning. It merely *symbolizes* the level of human busyness. Saturn and its associated problems are to our day and our culture what, perhaps, the Great Pyramid was to its day and the people of the Fourth Dynasty. Yet, such analogies don't provide much ground to stand on either. Finally, in discussing engineering efforts of such magnitude, what one must do is select (somewhat arbitrarily) a single aspect and let it stand for the rest.

Human factors in prelaunch activities. Now that the day of the "really big" space programs has arrived, many of the *ad hoc* procedures for prelaunch checkout developed during the earlier "smaller" programs, such as Mercury and Gemini, must now become much more formalized, while the highly automated ballistic missile checkout procedures (e.g., Minuteman) are inappropriate to manrated vehicles. In the earlier manned space programs, the use of ground computers was practically nil, but for the Apollo program, for instance, in which the launching alone will involve something like 10 000 men, and for which thousands of data points (upwards of 6000) will be monitored at central stations, the assistance of computers has become mandatory; and the problems of online man-computer interactions become crucial.

Although computers are being incorporated into missile and space programs, questions arise as to what level of automation will be most effective in prelaunch checkout. From a human factors point of view, this presents a new kind of problem for which there exist no adequate guidelines for the appropriate allocation of tasks between the men and the computers, or for the selection of effective control and display techniques. In addition, the launch checkout crew finds itself faced with less "direct contact" with the space vehicle, due to the interposition of checkout computers used primarily for sequencing and go/no-go determinations. In the face of an obvious need for, and recently demonstrated success with, automated space vehicle testing, checkout personnel nevertheless must retain ultimate responsibility for launch decisions. This responsibility, which weighs most heavily in the manned space program, can be carried out most effectively when the vagaries and special characteristics of individual vehicles are fully understood. Thus, for example, the period between delivery of a space vehicle to Cape Kennedy and the launch is a period not only of intensive activity but of continual learning. In a recent interview, Joseph G. Wohl, of Dunlap and Associates, Inc., a firm that has made many studies of prelaunch man–machine problems, stated the problem in a form very close to that of a human factors principle: "The full significance of missile and space vehicle preparation and prelaunch checkout is understood only when one realizes that these learning activities are indeed the primary means by which the residual unknowns, having escaped the filters of engineering analysis and factory testing, may be discovered, understood, and brought under control."[38] Checkout computers and display systems can aid this learning process. Thus, man–computer interaction in prelaunch activities represents in fact a general problem confronting all future advanced missile and space systems.

Displays. One of the important concerns of prelaunch activities is the detection of nonnormal (but within tolerance) conditions, a process that is very much dependent on the manner in which the test information is displayed. Says Wohl, "It is clear that detection of cyclic disturbances and second-order trends in sampled data can be enhanced by some display techniques and suppressed by others."[38] In their studies of existing display capabilities, Wohl and his colleagues point out potential inadequacies and suggest improvements based on human pattern-recognition capabilities. Present and planned test monitoring facilities differ from the Gemini and Mercury facilities (which used indicator lights, meters, and strip-chart recorders) in that they include CRT displays packed with information in alphanumeric form. Such so-called page- or list-formatted CRT's, although they allow the presentation of a great amount of information, submerge the test engineer's capabilities of observing data trends, oscillations, and other abnormal indications. In such displays, a value that goes out of limits is indicated only by a periodic interruption of the appropriate line of information. Nor does such a display give the human monitor information about instantaneous rate of change or leave behind a time history. Even at the relatively slow pace of manual checkout, this display makes it difficult to observe trends in test data. For fully automated tests run at "system speed," these displays are nearly useless in this respect, and would certainly not enhance the test engineer's understanding of what he was observing. Wohl states: "To help test engineers detect anomalies while attempting to monitor a large mass of CRT data at system speeds, we recommend the development of a real-time pattern recognition format for CRT's in which test point readings are 'mapped' onto a two-dimensional display in a coding scheme designed to enhance detection of impulses or other fleeting noncyclic disturbances."[38] One example of the type of thing Wohl is talking about is shown in Fig. 7, a display developed at the Computer Sciences Division of the Illinois Institute of Technology Research Institute. This plot displays information contained in approximately 42 000 eight-digit numbers. Through use of a computer program, a display surface is formed that instantly reveals errors or abnormalities by bulges on that surface.

Visual sampling behavior—a paradigm

In these two articles, we have referred fairly often to the stress that human factors investigators now place on the need for greater scientific generality in their models of aspects of human behavior. With better theoretical models, they say, it should become possible ultimately to solve analytically many human factors design problems that have had to be treated empirically through largely *ad hoc* methods. They want, in other words, to make a science out of this business.

Examples of this trend have been cited under the discussion of man–computer[24] and man–vehicle systems. An example of the weighty power of newer theoretical methods in modeling human pilots has also been cited in a recent SPECTRUM article.[6] Still another recent investigation, which can have an important bearing on the analysis of manned systems, is that undertaken by John W. Senders of Bolt Beranek and Newman, Inc.

One of the questions with which Senders has been concerned is how people absorb information; i.e., human information processing in relatively complex tasks. Senders says that he "spent a number of years dealing with more or less classical kinds of experiments in which you hold everything constant and vary one thing at a time and so on. These days people have discovered that you can be led into traps when you try to extrapolate the results of single-valued functions to multi-valued functions, for the interaction terms are very great indeed."[39] Thus, he has been led in his more recent efforts into a study of the nature of the effects of complication, and he has been trying to "generate mathematical models, which are rational models as opposed to purely empirical ones, in an effort to characterize and predict what will happen when [human operators and monitors] are confronted with very complicated information flows, complicated displays and complicated controls."[39, 40]

These studies lead to a rather comprehensive theory about human visual sampling behavior in particular and about the nature of *attention* in general. The extended theory incorporates ideas about what Senders calls "conditional sampling behavior, in which an observer inter-sample interval is a function of the value of the signal read on the previous sample." What this means is that an observer breaks up the environment in terms of attentional demand. In effect, he says, "This demands more attention than that." For instance, if he is observing dials and displays, he is most interested and attentive when their values indicate that a variable is approaching a boundary condition. The closer a dial needle, say, gets to a boundary, the more often the observer should look at it. This behavior is what Senders calls conditional sampling, and its mathematical formulation might be identified in textbooks one day as "Senders Sampling," a name that has more than mere alliteration to recommend it. Senders advances the idea "that as a consequence of the single-channelness of attention, queueing theory provides a general method of analysis of the switching of attention, of the attentional demand of a stimulus source, of the probability of simultaneous demand from two or more sources of stimuli, and of the notion of overload. The conditional sampling models provide the probability distributions which enter into the queueing model."[40]

Through the mathematical formulation of sampling behavior, Senders points out that it should be possible to predict how often an observer will need to look at the various control instruments and information displays in a particular man–machine situation. Thus, in a complex display situation, it would be possible to calculate the probability that one instrument has been left "screaming for attention" while another is being attended to. Furthermore, it may be possible to get an analytical measure of the reliability of the man in a complex system, to "see what the probability is that he will miss an event of interest, that he will clobber into something,"[39] a result that, as Senders points out, "strips the problem out of psychology."[39] Thus, one can measure an operator's actual performance, calculate the theoretical optimum, and then characterize the man's performance in terms of a percentage of the theoretical optimum. Through similar means, it also becomes possible to evaluate the human engineering design of an instrument or a pilot's cockpit. The significance of these results should not be underestimated. What they mean is that it is now possible to quantify things that have always been somewhat fuzzy in the past. As Senders says, "This is why human engineering and human factors, and so on, have always been rather fuzzy, too, because we have had no optimum, no 100 percent, no limiting condition against which to scale the behavior or to scale the quality of the design."[39]

At the present time, Senders and his colleagues are mechanizing the mathematical structure of this kind of sampling behavior (that is, simulating the sampling and scanning processes) and testing it against what people actually do.

Investigations like this illustrate the general direction of the human factors field—it is leaving behind the descriptions of *ad hoc* experiments that clutter the literature, and moving toward the unifying solidity of science.

The writer expresses his appreciation to the following people who generously gave assistance and information during the preparation of this survey: B. H. Manheimer, consultant; H. P. Birmingham, R. Chernikoff, J. H. Hill, B. L. Perry, NRL; R. L. Deininger, J. E. Karlin, Bell Labs.; J. C. R. Licklider, IBM; I. Sutherland, R. Taylor, ARPA; M. L. Dertouzos, W. R. Ferrell, T. B. Sheridan, J. E. Ward, L. R. Young, M.I.T.; E. Krendel, The Franklin Institute; J. I. Elkind, J. W. Senders, Bolt Beranek and Newman, Inc.; R. W. Pew, Univ. of Mich.; E. M. Bennett, The MITRE Corporation; J. Fox, T. Keller, Grumman Aircraft Engineering Corporation; M. A. Tolcott, J. G. Wohl, Dunlap and Associates, Inc.; B. Shackel, EMI Electronics Ltd.; K. S. Teel, Autonetics; and R. O. Besco, Hughes Aircraft Company.

REFERENCES

1. Licklider, J. C. R., "Man–computer communication—introduction," in *Communication Processes*, F. A. Geldard, ed. New York: Pergamon, 1965.

2. Licklider, J. C. R., "Man–computer symbiosis," *IRE Trans. on Human Factors in Electronics*, vol. HFE-1, Mar. 1960.

3. Sutherland, I., Private communication.

4. Miller, G. A., "Man–computer interaction," in *Communication Processes*, F. A. Geldard, ed., New York: Pergamon, 1965.

5. Fano, R. M., "The MAC system: the computer utility approach," *IEEE Spectrum*, vol. 2, pp. 56–64, Jan. 1965.

6. Senders, J. W., *et al.*, "Men and computers—Sixth annual symposium on human factors in electronics," *IEEE Spectrum*, vol. 2, pp. 104–108, Aug. 1965.

7. Licklider, J. C. R., "Problems in man–computer communications," in *Communication Processes*, F. A. Geldard, ed. New York: Pergamon, 1965.

8. Licklider, J. C. R., Private communication.

9. Kessler, M. M., "The MIT technical information project," *Phys. Today*, Mar. 1965.

10. Weissman, C., "Communication techniques for an international time-sharing users group," presented at IEEE 6th Annual Symp. on Human Factors in Electronics, May 6–8, 1965.

11. Lindgren, N., "Machine recognition of human language—Part I," *IEEE Spectrum*, vol. 2, pp. 114–136, Mar. 1965; Part II, pp. 44–59, Apr. 1965.

12. Rowe, W. D., "Why systems science and cybernetics?," *IEEE Trans. on Systems Science and Cybernetics*, vol. SSC-1, Nov. 1965.

13. Newell, A., "The possibility of planning languages in man–computer communication," in *Communication Processes*, F. A. Geldard, ed. New York: Pergamon, 1965.

14. Sutherland, I. E., "Sketchpad: A man–machine graphical communication system," Tech. Rept. No. 296, M.I.T. Lincoln Lab., reissued May 19, 1965.

15. Davis, M. R., and Ellis, T. O., "The RAND tablet: A man–machine graphical communication device," Memo. RM-4122-ARPA, The RAND Corp., Aug. 1964.

16. Stotz, R. H., and Ward, J. E., "Operating manual for the ESL display console," M.I.T. ESL Internal Memorandum 9442-M-129, Mar. 9, 1965.

17. Hamilton, M. L., and Weiss, A. D., "An approach to computer-aided preliminary ship design," M.I.T. ESL-TM-228, Jan. 1965.

18. Dertouzos, M. L., and Therrien, C. W., "CIRCAL: On-line analysis of electronic networks," M.I.T. Electronic Systems Lab. Rept. ESL-R-248, Oct. 1965.

19. Reintjes, J. F., and Dertouzos, M. L., "Computer-aided design of electronic circuits," presented at WINCON Conf., Los Angeles, Calif., Feb. 2–5, 1966.

20. Dertouzos, M. L., Private communication.

21. *Project MAC Progress to July 1964.* M.I.T., 171 pp. (Available from Defense Documentation Center, Document Service Center, Cameron Station, Alexandria, Va.)

22. Elkind, J. I., Private communication.

23. Bennett, E., *et al.*, "AESOP: A prototype for on-line user control of organizational data storage, retrieval and processing," MTP-23, The MITRE Corp., Nov. 1965: also *Proc. Fall Joint Computer Conference*, 1965.

24. Edwards, W., "Introduction: Revision of opinions by men and man-machine systems," *IEEE Trans. on Human Factors in Electronics*, vol. HFE-7, p. 1, Mar. 1966.

25. Sheridan, T. B., "The human operator in control instrumentation," in *Progress in Control Engineering—1.* London: Heywood & Co., Ltd., 1962, pp. 143–187.

26. Young, L. R., and Li, Y. T., "Studies of human dynamic space orientation using techniques of control theory," Man-Vehicle Control Lab Rept., Dec. 1964.

27. McRuer, D., *et al.*, "Human pilot dynamics in compensatory systems," Tech. Rept. AFFDL-TR-65-16, AF Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, July 1965.

28. Young, L. R., M.I.T., Private communication.

29. McRuer, D. T., and Krendel, E. S., "The human operator as a servo system element," *J. Franklin Inst.* vol. 267, pp. 1–49, May, June 1959.

30. Meiry, J. L., "The vestibular system and human dynamic space orientation," Sc.D. dissertation, T-65-1, Man-Vehicle Control Lab., M.I.T., June 1965.

31. Draper, C. S., *et al.*, "The roles of men and instruments in control and guidance systems for spacecraft," presented at the XVth Internat'l Astronautical Cong., Warsaw, Poland, Sept. 7–12, 1964.

32. Li, Y. T., *et al.*, "Adaptive functions of man in vehicle control systems," International Federation of Automatic Control (Teddington) Symp., Sept. 14–17, 1965.

33. Keller, T., *et al.*, "A study of the mechanics of human balancing for potential application to the control of vehicles," Grumman Research Dept. Memo. RM-299, Oct. 1965.

34. Keller, T., Private communication.

35. Perry, B. L., "The Rainbow Optical Landing Aid," NRL Rept. 6184, Feb. 1, 1965.

36. Romnes, H. I., "Satisfaction and success in the highly organized society," *IEEE Spectrum*, vol. 2, p. 67, Dec. 1965.

37. Karlin, J. E., Private communication.

38. Wohl, J. G., Private communication.

39. Senders, J. W., Private communication.

40. Senders, J. W., *et al.*, "An investigation of the visual sampling behavior of human observers," Rept. No. 1246, Bolt Beranek and Newman, Inc., May 10, 1965.

Detection of coherent optical radiation

Recent research work in laser detection includes superheterodyne and nonlinear parametric studies in the optical spectrum. Operation at these increased wavelengths can result in a significant improvement in sensitivity

Marvin E. Lasser *Philco Corporation*

Sensitive coherent detection techniques, identical in principle to the methods applicable in the RF or microwave regions, are being investigated at optical wavelengths. Since we already can reach the natural limit of being able to count individual photons in both the short-wavelength visible and near-ultraviolet regions, there is no possibility for sensitivity improvement in those spectra. However, at increased wavelengths, where the detectors are several orders of magnitude away from being able to count photons, much higher sensitivity can be achieved through the use of coherent detection methods. In optical superheterodyne detection the mixing in the superheterodyne receiver arises because optical detectors measure the intensity rather than the amplitude of the light waves, and thus square-law detection results. Another detection method is based on a truly nonlinear effect, in which the electric field associated with the incident radiation is so strong that it drives the polarization out of the linear region. Parametric amplification of the optical signal then becomes possible.

With the announcement of coherent optical radiation sources, a number of investigators began to consider the potential advantages and the practicality of utilizing the coherent nature of this light in the development of sensitive detection systems. Because much improvement in detection sensitivity previously had been achieved in both the RF and microwave regions through the use of coherent rather than incoherent detection methods, there was a great temptation to apply similar techniques to the optical spectrum. In fact, the coherent detection methods that are being studied at optical frequencies are identical in principle to those used in the RF or microwave regions. There now appears to be little question that coherent optical detection systems can be used to advantage.

There is a natural limit to the minimum amount of radiation that can be detected, even in an ideal system that is, in principle, free of external noise sources. It is clear that we can do no better than count individual photons. Since this operation can already be accomplished with cooled photomultipliers in the short-wavelength visible region and in the near-ultraviolet region, there appears to be no possibility for further improvement in sensitivity. The only application in which coherent systems might be of interest at the shorter wavelengths is in FM detection; however, there appears to be only limited active work in this area.

In the wavelength region from about 0.7 to 1.3 μm (7000 to 13 000 Å), the efficiency of photoemissive surfaces has dropped to about 1 percent, as compared with efficiencies of 20 percent in the blue and ultraviolet regions. Also, solid-state photodetectors come into their own at wavelengths beyond 0.7 μm, at least as far as detection efficiency is concerned. Until quite recently, however, it was not possible to demonstrate the sort of gain mechanism now so readily achieved with photomultipliers. Many laboratories are currently investigating solid-state diode structures having built-in gain mechanisms.

As we go to wavelengths beyond 1.3 μm there is no choice but to consider solid-state devices,* since photoemission simply does not take place at the longer wavelengths. In the wavelength region from 1 to 2 μm, the minimum detectable energy possible with available incoherent solid-state detectors is two or three orders of magnitude too high to make the full counting of photons

---

*We are explicitly excluding thermal-energy detectors from consideration here. They serve their purpose very well in any given wavelength region, and will probably be used widely until operating devices that are capable of detecting the electromagnetic radiation directly become available.

possible. Beyond 2 μm we are an additional two orders of magnitude away from the capability of recording individual photon events.

As we go to longer wavelengths it follows directly that we must use devices that will respond to low-energy excitation. As the minimum energy for excitation decreases, there are many competing excitation processes that interfere with the incoming optical signal. For example, thermal generation of excited charge carriers can create problems. Although cooling the device helps, it is not completely effective. Many other competing phenomena (i.e., other sources of noise) are present, such as optical excitation due to the background, which may occur in the case of a cooled detector whose surroundings are at room temperature. The room-temperature blackbody radiation from the surroundings generates carriers that contribute to the noise.

It is at these longer wavelengths, where it becomes more difficult to achieve high sensitivity with incoherent methods, that coherent systems offer the greatest potential improvement. It is hoped that a continuing study along these lines in the infrared region will make it possible to close the gap between the optical and microwave regions. Coherent sources in the region of a few hundred micrometers[1] are being continually improved. By the time these systems are developed, the requirement for sensitive detectors will be quite severe. It is difficult to see how sufficiently effective sensitivities can be obtained without the use of coherent systems.

Optical superheterodyne detection

The operation of a superheterodyne detector is based on the well-known empirical fact that all optical detectors measure the intensity, rather than the amplitude, of the incident electromagnetic wave. If we now assume that two electromagnetic light waves of different frequencies fall on the same area of a detector, the resultant amplitude of the waves is given by

$$E_t = E_1 \cos \omega_1 t + E_2 \cos \omega_2 t \tag{1}$$

where E_1 and E_2 are the amplitudes and ω_1 and ω_2 are the angular frequencies of the two incident waves. We are implicitly assuming that the two waves are parallel and normal to the surface of the detector. If these conditions are not met, we must realize that spatial phase cancellations might result.

To detect the presence of the signal we measure the photointensity, which is given by the square of Eq. (1):

$$E_t^2 = E_1^2 \cos^2 \omega_1 t + E_2^2 \cos^2 \omega_2 t + E_1 E_2 \cos (\omega_1 - \omega_2) t + E_1 E_2 \cos (\omega_1 + \omega_2) t \tag{2}$$

We can assume that the detector has a high-frequency cutoff above the difference frequency $(\omega_1 - \omega_2)$. We cannot, however, assume that the detector will respond to signals at the frequencies given in the first, second, and fourth terms of (2).

It has been shown that there is a minimum time required for the generation of electron–hole pairs.[2] Known as the electron–photon correlation time, it is of the order of 3×10^{-14} second. The detector, therefore, cannot respond to signals at optical frequencies. Under these conditions we can only detect the average power in these high-frequency terms. The average of $\cos^2 \omega t$ is one half, whereas the average of the linear cosine term, $\cos (\omega_1 + \omega_2) t$, is clearly zero. The result, then, is

$$E_t^2 = \frac{E_1^2}{2} + \frac{E_2^2}{2} + E_1 E_2 \cos (\omega_1 - \omega_2) t \tag{3}$$

A more physical picture can be given to explain the meaning of the effects described in Eqs. (1) through (3). If two coherent electromagnetic waves are superimposed on each other in space, the vector amplitudes of the two waves will add, giving rise to a single amplitude-modulated wave at the difference or beat frequency. An optical detector placed in the path of this composite wave will detect simply the envelope of the modulated light wave, thus yielding an electrical signal that is modulated at the beat frequency. There is no power at the microwave frequency in the input wave; it is only after detection that a microwave signal is generated.

Let us now consider the problem of sensitivity. The signal power is given by an expression

$$S = 2\eta^2 I_L I_S f(\omega) \tag{4}$$

where η is the quantum efficiency (photoelectrons per incident photon), and ηI_L and ηI_S are the currents generated by the local oscillator and signal source respectively. The $f(\omega)$ term is included to indicate how the frequency response of the device can be taken into account. The factor 2 comes about because of the coherent nature of the light.[3,4] The shot noise is given by an expression of the form

$$N = 2e\,[(\eta I_S + \eta I_L + I_D) f(\omega) + I_R]\,\Delta f \tag{5}$$

where I_D is the dark current of the device, I_R is the noise equivalent load of the following preamplifier, and Δf is the bandwidth. If a very strong local oscillator is used, the signal-to-noise ratio reduces to

$$\text{SNR} = \frac{\eta I_S}{e\,\Delta f} \tag{6}$$

If we let $I_S = en$, where n is the number of incident photons, then

$$\text{SNR} = \frac{\eta n}{\Delta f} \tag{7}$$

If I_L is sufficiently strong, it is possible to count photons in the incident beam.

If we are to take advantage of the fact that the carrier is of a very high frequency, we must consider information bandwidths that cause operation to be in the microwave region. Experience has shown that both photoemissive and solid-state devices are useful for this purpose.

Photoemissive superheterodyne detectors

With photoemissive devices, two modes of operation have been employed successfully: the photomultiplier type and the traveling-wave type.

The frequency response in the photomultiplier is limited by the time delays associated with the transit time of the carriers. It is not, however, the transit time itself, but rather its dispersion, that sets the upper frequency limits. If the carriers stay bunched together in their travel from dynode to dynode, the frequency information will not be degraded. This close bunching can be accomplished by the use of a magnetic field crossed with a high electric field.[5] If the fields are sufficiently strong, the dispersion resulting from initial velocity variation can be essentially eliminated.

In the traveling-wave tube (TWT) there is no transit time limitation, since the beam and circuit waves are synchronous. The bandwidth is limited only by the bandwidth of the helix and its couplers; see Fig. 1.

A photo TWT, designed by A. E. Siegman,[6] is ideal for superheterodyne detection, especially if the wavelength of interest is in the region in which the efficiency of the photoemitter is high. In a conventional TWT an ac signal is impressed on a large direct current, and ac amplification is brought about by increasing the depth of modulation of the combined currents. If there is no direct current, clearly there can be no current amplification; however, there will be an impedance transformation, which can be helpful. This is the case for a photocathode that has no appreciable dark current. If, on the other hand, a large direct current is impressed on the TWT, the shot noise of the system will be increased.

If superheterodyne detection is used, the only requirement is that the photocathode be illuminated by a strong local oscillator (LO) light beam. Heterodyning will take place at the photocathode. The microwave tube elements will then amplify the signal by increasing the depth of modulation into the LO-generated direct current. As shown in Eq. (6), however, the increase in shot noise caused by the LO-generated background does not degrade the signal-to-noise ratio.

Solid-state superheterodyne detectors

The high-frequency solid-state devices that have been the most successful are the photodiodes,[7,8] rather than the photoconductors. Here, again, the problem of transit time arises; and since the junction of a photodiode can always be made thinner than that of a photoconductor, the photodiode will have as high or higher frequency response.[9] Therefore, this section will be restricted to a discussion of the photodiode.

Let us consider some of the difficulties that arise when we try to build photodiodes that have frequency responses out to the gigacycle region. The frequency limitation can be considered either to be RC-limited or limited by the transport time of the carriers within the device. Let us take the latter problem first. Once light is incident on a solid-state material, free holes and electrons are generated. These generated carriers can move either by diffusion or under the influence of an electric field. The speed with which carriers move, or the distance they move in a given time, is determined by the diffusion constant of the material. In a given time T, the distance X that the carriers will move, in the presence of other carriers which set up a diffusion current, is given by

$$X^2 = DT \tag{8}$$

where D is the diffusion constant. With this in mind, we see that it is necessary to construct devices in which the distance carriers have to move by diffusion is extremely small (see Fig. 2). It is here implicity assumed that the optical absorption coefficient for the light is extremely high, so that the incident photons would, with high efficiency, generate free carriers in the distance X shown in Fig. 2; that is, we would like to satisfy the condition that $1/\alpha \approx X$, where α is the optical absorption coefficient.

These conditions can be fairly well satisfied in the III–V compounds, such as GaAs, InAs, and InSb. The type of structure employed is shown in Fig. 2. The thickness X is of the order of 1 μm, and therefore the time it takes for carriers to diffuse across the distance X (for typical diffusion constants in these materials) is of the order of 10^{-10} second.

Let us now consider what can be done if we use materials such as Ge or Si, in which the absorption process near the band edge requires the combined effect of photons and phonons and, as a result, has a lower value for α at wavelengths near the absorption edge. In this case we find that we will not have the proper conditions for both high collection efficiency and high frequency response. If the material is less than $1/\alpha$ thick, only a

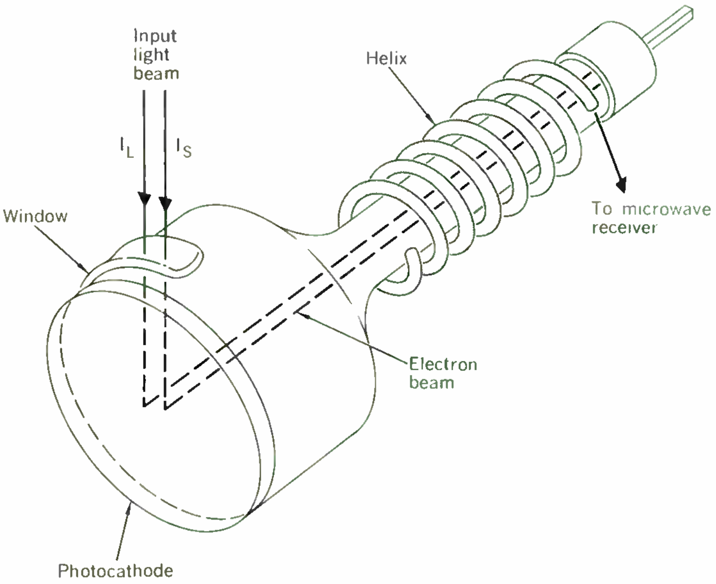
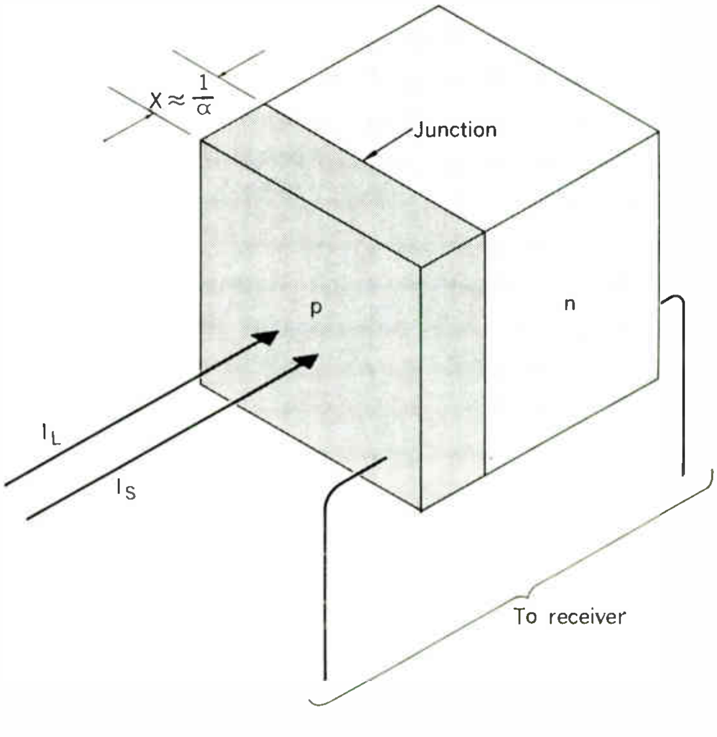


Fig. 1. A photo traveling-wave tube.

Fig. 2. Structure of a high-frequency photodiode in a material having a high optical absorption coefficient α.



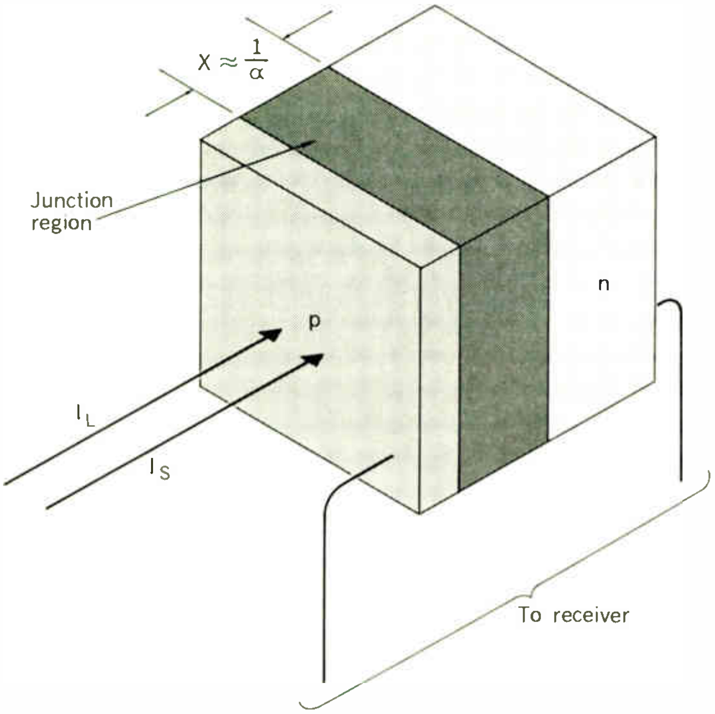


Fig. 3. Structure of a high-frequency photodiode in a material having relatively low values of α.

small fraction of the carriers will be absorbed; however, if the material is made thicker, the diffusion time will become too long. It is with these materials of low α that we can take advantage of the higher velocities attainable in the presence of a field. The structure that must be employed is one in which a field exists over a distance comparable to the distance $1/\alpha$, as shown in Fig. 3. The application of only a few volts across a device of this type will bring the velocities of the carriers moving in the field up to the saturated drift velocity—the highest velocity that can be attained by the carriers. The value V_s is of the order of 10^6 cm/s. If α is about 10^3, the time necessary for the carriers to drift across the field region is about 10^{-9} second. Again we find ourselves operating in the microwave region.

The limiting value of the RC time constant is determined by the series resistance R_s and the junction capacitance C_b of these devices; see Table I. The series resistance in all cases is made as low as practicable, and experience has shown that this value is typically less than 20 ohms. The capacitance of the junction is proportional to the area and inversely proportional to the junction width. The dependence on junction width makes the devices shown in Fig. 3 have an inherent advantage in that, for the same area, they would have a lower capacitance than those of the type shown in Fig. 2. In both cases, however, it is best to make the junction area as small as possible, consistent with the optical requirements of the system. Typically, capacitances in the diffusion devices with areas of 10^{-4} cm² are about 3 pF, and in the drift devices are about 1 pF. Table I shows high-frequency cutoff characteristics, of a number of solid-state devices, including the cutoff frequencies due to the RC limitation (f_{RC}) and due to the transit time (f_{TR}).

In connection with Eq. (6) we saw that in order to obtain the requisite high sensitivities a strong local oscillator is required. As lasers improve, this requirement becomes easier to meet; however, it would be advantageous to be able to work with a weaker, more readily available local oscillator. This can be accomplished if we take advantage of the avalanche or photomultiplier-type effect referred to earlier.

Breakdown diodes have been extensively studied in the past. Two mechanisms give rise to the breakdown process: tunneling and avalanching.[10] Tunneling breakdown takes place when the junction is extremely thin, so that the carriers can tunnel from one side of the junction to the other. It is, of course, this tunneling mechanism that is the basis of operation of the tunnel diode. If, however, the junction is made thicker, tunneling cannot take place, and avalanching will then give rise to the breakdown of the field across the junction. Avalanching occurs when free-charge carriers impinge on other bound carriers, providing enough impact energy to release them; these released carriers can, in turn, excite and release others. If the junction is biased so that it is not quite in a breakdown condition, the presence of an optically generated carrier can give rise to a controlled multiplication process. The most stringent requirement for avalanching is that the junction be uniform. If it were not, breakdown would be initiated at only limited regions of the diode and avalanching would be uncontrolled.

A word or two about the frequency response of the

I. High-frequency cutoff characteristics of several solid-state devices

Material	Wavelength Region, μm	Area, cm²	R_s, ohms	C_b, pF	f_{RC}, Gc/s	Transport Distance, μm	f_{TR}, Gc/s
Si, p-i-n	0.65–0.75	6×10^{-5}	5	0.5 (−6 V)	45	2.5	15
	0.65–0.75	2×10^{-4}	5	1.5 (−10 V)	12	2.5	15
	0.5–0.95	3.5×10^{-2}	65	5.0 (−100 V)	0.5	75.0	0.5
Ge, n-p	0.5–1.5	4×10^{-4}	6	5.6 (−6 V)	17	0.4	8
GaAs, p-n	0.4–0.8	1.3×10^{-4}	3.0	7.0 (−4.5 V)	7.5	1.0	8
		7×10^{-4}	2.5	32 (−4.5 V)	2	1.0	8
InAs, p-n	0.5–3.5	3.2×10^{-4}	12	3.0 (−5 V)	4.5	2.0	10
		2×10^{-3}	8	30 (−2 V)	0.65	2.0	10
InSb, p-n (77°K)	0.5–5.2	5×10^{-4}	18	7.1 (−0.2 V)	1.2	2.0	10

avalanche devices is in order.[11] If in the avalanche process we consider one carrier multiplication—for example, let us assume that the pairs are produced due to the impact of electrons but not holes—the frequency response would be determined by the time it takes for the electrons and an equal number of holes to be swept out of the junction region. In most semiconductors the mobility of electrons and holes are not greatly different from each other. The time it takes for the holes to be swept out is usually of the same order as that required for electron sweepout; therefore, avalanche gives rise to an increase in the gain–bandwidth product of the device. On the other hand, the avalanche generation of pairs by both avalanching holes and avalanching electrons, giving rise to further increase in gain, will also result in a decrease in response time. This can be seen by referring to Fig. 4. We see that a carrier—such as an electron—can travel across, and give rise to a second hole and electron at point A; this hole then travels back and can generate another hole–electron pair at point B. Therefore, by a process of this sort, the time for complete sweepout can be appreciably lengthened. In this case, we do reach an upper limit for the gain–bandwidth product. There is a concomitant decrease in signal-to-noise ratio as well because of the uncorrelated contribution of the noise from the electron multiplication and from the hole multiplication.[12] In the case of one-carrier multiplication the signal-to-noise ratio remains constant, although both the signal and noise do, of course, increase. In general, it is therefore considered advantageous to operate in the region of one-carrier multiplication.

It is also possible to use the nonlinear characteristics of the junction current–voltage curve, and thereby achieve parametric amplification of the input signal. As noted previously, the microwave signal is generated at the surface and the junction is required only to detect the presence of the signal. It is possible, therefore, to swing the bias so as to drive the diode into its nonlinear response and then have it interact with and parametrically amplify the microwave signal, which is generated by the amplitude-modulated light. This technique has been shown to work quite effectively.[13]

Although there is still a great deal of work being done on the devices described thus far, research in this field appears to be coming to an end. The efforts along these lines are now being directed toward device structure and ultimate noise figures. The mechanisms involved have been well studied and, with very few exceptions, well understood.

Nonlinear optical detection

Nonlinear optics, an entirely new area related to the detection of radiation, has only recently become the subject of intensive study; the first experiments were performed in 1961.[14] As a preliminary step to our discussion, it might be advisable to describe the related effect in "linear optics." (This latter term has arisen, only since the advent of nonlinear optics, as another way of referring to conventional optical techniques.) In general, a medium becomes polarized when it is struck by light. If the medium is transparent, and therefore lossless, the energy that induces the polarization is reradiated. The original radiation and the reradiated emission combine to give the standard refractive effects.

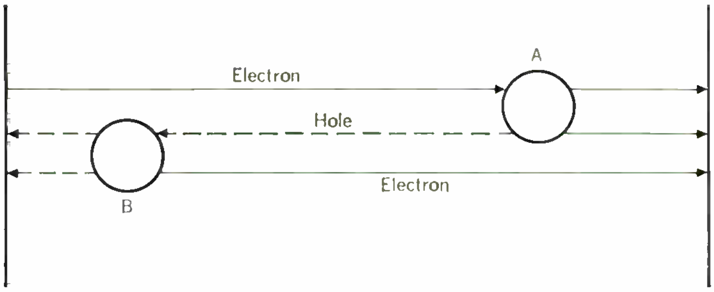


Fig. 4. Schematic representation showing the time delay that arises when both the electrons and holes cause the avalanche generation process to occur.

If now the electric field associated with the incident radiation is so strong that it drives the polarization of the optical material out of the linear region, "nonlinear optics" effects start to appear. The standard approach that is used to describe nonlinearities of this type mathematically is to expand the polarizability in a power series; that is, instead of the polarization being proportional to the applied field, as in the expression $P = \alpha E$, we have a series of terms of the form

$$P = \alpha_1 E + \alpha_2 E^2 + \alpha_3 E^3 + \ldots \qquad (9)$$

Let us consider, in particular, the second term, $\alpha_2 E^2$. We see that if there is an applied optical electric field $E = E_0 \sin \omega T$, then this quadratic term provides a contribution P_2 to the polarization of the crystal

$$P = \alpha_2 E_0^2 \sin^2 \omega T \qquad (10)$$

which reduces to

$$P_2 = \frac{\alpha_2 E_0^2 (1 - \cos 2\omega T)}{2} \qquad (11)$$

We note that there is a dc term, which arises from the quadratic linearity in much the same way as in square-law detectors operating in the RF or microwave region; in fact, this term has been observed in experiments. The cos $2\omega T$ term gives rise clearly to an optical signal at twice the frequency of the incident light signal. Therefore, an incident signal in the red can be doubled so that the light emerging from the crystal contains a detectable amount of ultraviolet radiation at just double the frequency (half the wavelength) of the incident strong optical radiation. Experiments of this type have shown, however, the occurrence of a phenomenon that must be taken into consideration. In a normally dispersive optical material, the index of refraction at the shorter wavelengths is higher than that in the longer-wavelength region. This means that the harmonically generated optical signal at the frequency 2ω will be traveling slower than the high-intensity signal at frequency ω. As a result, there will be a continuing difference in phase between the light at 2ω generated early in the process with that generated further into the crystal. It is clear that the continuously varying phase difference will give rise to phase cancellation, so that there is only a limited distance over which it pays to try to generate the harmonic signal. However, an ingenious method of index matching, independently developed by Giordmaine[15] and Maker *et al.*,[16] takes advantage of the fact that in certain anisotropic crystals it is possible to choose a direction of propagation

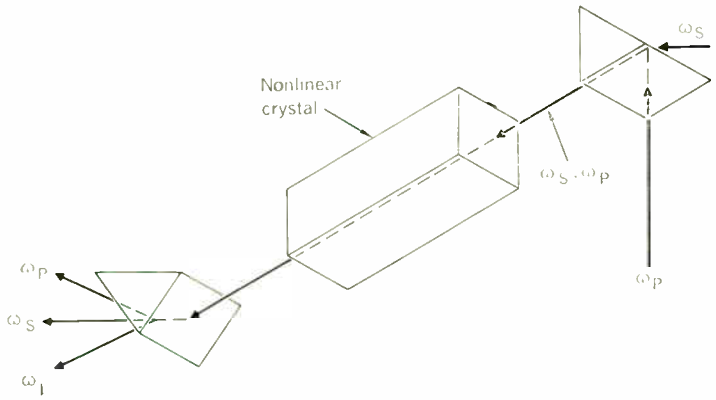


Fig. 5. Experimental setup used in observations of optical parametric amplification effects.

such that the ordinary refractive index at one frequency is equal to the extraordinary refractive index at the other frequency. If the crystal is properly oriented, the second-harmonic signal can then be made to continuously build up. The same concept may be stated in a slightly different way. Let us consider a traveling wave whose propagation is given by e^{ikz}, where $k = 2\pi n/\lambda$; we then have to satisfy the condition that $k_2 = k_1 + k_1$. In other words, we have two photons at ω_1 adding up to give us a single photon at ω_2. Again, the resulting condition is that $n_1 = n_2$.

The extension from harmonic generation to parametric amplification is now a very simple one. In the optical parametric amplifier we have a signal frequency ω_S, and in the presence of a strong light signal at the pump frequency ω_P, ω_S is amplified. In the process an idler frequency ω_I is also generated. The condition to be satisfied is that $\omega_P = \omega_S + \omega_I$. What happens here is that ω_P mixes with ω_S, generating a signal at ω_I; this in turn reacts back with the strong signal at ω_P, creating additional photons at ω_S. As this process continues through the crystal, energy is lost from the pump and converted into energy at ω_S and ω_I. At the same time, if this is to occur throughout the length of the crystal, we must again worry about the phase- or index-matching conditions. In this case the resultant condition is that $k_P = k_S + k_I$. The experimental setup that first gave rise to these results is shown in Fig. 5.[17]

Comparison of linear and nonlinear detection

Some attempt should probably be made here to compare linear and nonlinear detection methods. Nothing that has been said can, of course, contradict the statement that the detection process can do no better than count photons; therefore, in either case, the minimum detectable energy will be determined by the energy of a single photon.

The nonlinear amplifier has to be tuned as given by the index matching condition, but it is much less critical than the tuning required in the linear device where it is necessary that the signal frequency be no more than 10^9 or 10^{10} hertz away from the local-oscillator signal. In the optical parametric amplifier operating in a nonlinear mode, we could go two orders of magnitude beyond this and still consider the device to be tuned.

The difficulty with the nonlinear device is the need for a high-powered source. All of the experiments reported to date have utilized a pulsed ruby laser for the pump, and thus the amplification takes place for only a short period of time. There is much hope that both with improved lasers that operate in the CW mode at reasonably high power and with greatly improved nonlinear optical crystals this will no longer be a problem. It has been reported that it should be possible, through use of a one-milliwatt CW laser and a crystal of $LiNbO_3$, to achieve optical parametric oscillation of the nonlinear variety.[18]

There is also no reason to believe that the nonlinear and linear optics must work separately; in fact, it is reasonable to propose that a nonlinear amplifier that can give rise to a stronger optical signal amplified at the optical frequency can be used as a preamplifier in front of the linear superheterodyne detector.

The nonlinear effect can also be used to up-convert rather than to amplify. In other words, ω_I could be a higher frequency than ω_S, and thus we would have not only amplification but also a shift in frequency to a detector that has a higher sensitivity at ω_I than at ω_S.

REFERENCES

1. Patel, C. K. N., Faust, W. L., McFarlane, R. A., and Garrett, C. G. B., "CW optical maser action up to 133 μ (0.133 mm) in neon discharges," *Proc. IEEE (Correspondence)*, vol. 52, p. 713, June 1964.

2. Pershan, P. S., and Bloembergen, N., "Frequency response of the photomixing process," *Appl. Phys. Letters*, vol. 2, p. 117, 1963.

3. Oliver, B. M., "Signal-to-noise ratios in photoelectric mixing," *Proc. IRE (Correspondence)*, vol. 49, pp. 1960–1961, Dec. 1961.

4. Haus, H. A., Townes, C. H., and Oliver, B. M., "Comments on 'Noise in photoelectric mixing,' " *Proc. IRE (Correspondence)*, vol. 50, pp. 1544–1546, June 1962.

5. Gaddy, O. L., and Holshouser, D. F., "A microwave frequency dynamic crossed-field photomultiplier," *Proc. IEEE*, vol. 51, pp. 153–162, Jan. 1963.

6. McMurtry, B. J., and Siegman, A. E., "Photomixing experiments with ruby optical maser and traveling wave microwave phototube," *Appl. Opt.*, vol. 1, pp. 51–53, Jan. 1962.

7. Lucovsky, G., Lasser, M. E., and Emmons, R. B., "Coherent light detection in solid-state photodiodes," *Proc. IEEE*, vol. 51, pp. 166–172, Jan. 1963.

8. Riesz, R. P., "High-speed semiconductor photodiodes," *Rev. Sci. Instr.*, vol. 33, p. 994, Sept. 1962.

9. Lucovsky, G., Schwarz, R. F., and Emmons, R. B., "Photoelectric mixing of coherent light in bulk photoconductors," *Proc. IEEE (Correspondence)*, vol. 51, pp. 613–614, Apr. 1963.

10. Chynoweth, A. F., "Internal field emission," in *Progress in Semiconductors*, vol. 8. New York: Wiley, 1960. pp. 95–123.

11. Emmons, R. B., and Lucovsky, G., "The frequency response of avalanching photodiodes," *IEEE Trans. on Electron Devices*, vol. ED-13, pp. 297–305, Mar. 1966.

12. Tager, A. S., "Current fluctuations in a semiconductor dielectric under the conditions of impact ionization and avalanche breakdown," *Soviet Phys.-Solid State*, vol. 6, pp. 1919–1925, Feb. 1965.

13. Penfield, P., and Sawyer, D. E., "Photoparametric amplifier," *Proc. IEEE*, vol. 53, pp. 340–347, Apr. 1965.

14. Franken, P. A., Hill, H. E., Peters, C. W., and Weinreich, G., "The generation of optical harmonics," *Phys. Rev. Letters*, vol. 7, pp. 118–119, Aug. 15, 1961.

15. Giordmaine, J. A., "Mixing of light beams in crystal," *Phys. Rev. Letters*, vol. 8, pp. 19–20, Jan. 1, 1962.

16. Maker, P. D., Terhune, R. W., Nisenhoff, M., and Savage, C. M., "Effects of dispersion on the production of optical harmonics," *Phys. Rev. Letters*, vol. 8, p. 21, Jan. 2, 1962.

17. Wang, C. C., and Racette, G. W., "Measurement of parametric gain accompanying optical difference frequency generation," *Appl. Phys. Letters*, vol. 6, p. 169, April 15, 1965.

18. Askin, A., and Boyd, G. D., "Optical nonlinear parametric and harmonic interaction in $LiNbO_3$," presented at ONR Conf. on Physics of Quantum Electronics, June 28–30, 1965, San Juan, P.R.

High-speed information channels

Along with the tremendous increase, especially in the past few years, in the amount and kinds of information being transmitted, there has been an intensified need for high-speed channels and systems. At present the greatest growth potential appears to be in the transmission of graphic information

Richard T. James *American Telephone and Telegraph Company*

Following a brief review of the progress in the communication art that has led to today's high-speed transmission of information, including the evolution of the carrier spectrum, the reasons for the growth of high-speed communication are analyzed. The basic operation of parallel and series data systems is illustrated through the use of typical examples. The West Coast operation of The Wall Street Journal, described here, demonstrates one way in which high-speed transmission techniques can be profitably employed in the publishing field.

In the normal course of our existence in modern society we spend an appreciable percentage of our time handling information. Most frequently we transmit it by speaking or receive it by listening. This means of handling information, the "human channel," is universal; even the most primitive of people have a language. Almost as universal is the use of writing materials of one kind or another. With these, and a set of symbols (an alphabet), we can also transmit information by writing and receive it by reading. We do not know when or where this recording capability was added to the human channel, but it was a very great addition indeed. Among other things, it gave us a history.

The speed of the human channel, of course, varies from person to person. Using the English language, we are capable of speaking, and of comprehending the spoken word, at a rate of 180 to 300 words per minute. We read, and with special training can also write, at about the same rate. Studies have indicated that our speed limitation is not physiological (we can utter words faster than this) but mental, in that the main determinant is familiarity with the words being used. Therefore, it does not appear that the speed of the human channel has changed very much through the years. The ancient Assyrian, except for the handicap of having to thumb through his clay tablets, probably handled information nearly as fast as we do today.

For many centuries the human information channel, operating either in its elemental linguistic mode or the more advanced recording mode, has served mankind well. There was always a need to extend it over greater distances, and many difficulties occurred because this could not really be done; nevertheless, the channel capacity was adequate for the needs of the times.

With the advent of electrical communication systems in the early 19th century, this troublesome range limitation of the human channel began to disappear. To all intents and purposes it did disappear early in the present century as the capabilities for voice communication became international in scope. Most of the world's communication plant today is devoted to the needs of the human channel and the great bulk of it is not handling information any faster than this.

Quite recently, however, channels and systems that handle information at speeds so high that they dwarf the human channel have made their appearance. The environment of these systems, the reasons for their existence, relevant technical considerations, and a few typical applications will be the concern of this article.

Historical background

It is helpful at the outset to present a brief review of the progress in the communication art that has made today's high-speed information systems possible. This progress really began with the extension of the range of the human channel by means of the manual telegraph. In the United States, it is generally considered to have started in the year 1844 with the transmission of Morse's famous "What hath God wrought" message over the Government-sponsored Washington–Baltimore line. After numerous demonstrations, this line was opened for commercial business in 1845.

The telegraph was a very great addition to the communication capabilities of the nation, and the network grew rapidly. It had linked the Atlantic and Pacific coasts by 1861, and by 1877 the Western Union Telegraph Company was operating 7500 commercial offices throughout the country. In that year alone the network handled over 21 million messages for the general public. This was not by any means all the information transmitted by teleusually the more economical and practical; and in most cases it leads to a high-speed information-handling system of some kind.

Another governmental activity that involves high-speed information systems is the space effort, along with its multitudinous ancillary activities. The amount of information generated in this area is truly amazing. In most cases the information is processed at distant points, so transmission systems are required. The time element is of importance, since most of the activities in the space area operate in real time to some degree. This factor, coupled with the large amounts of information handled, leads to the need for high-speed transmission systems.

In the business community there is a continuous drive toward more efficient operations. For large corporations, increased efficiency generally means a more closely integrated operation, which in turn calls for close budgetary and inventory control, rapid order processing, and availability of information for comprehensive business status reports. With certain large complex commercial operations, high speeds soon become a necessity in certain parts of the overall information system.

The most popular reason for high-speed information systems is a sort of all-inclusive concept called automation. It mainly involves interconnecting computers, etc., into information-handling *complexes* (so called for the lack of a better name)—the object being that the complex, as a whole, can undertake tasks that any of its components could not accomplish alone, or at any rate could not accomplish soon enough. Although this may appear to be a rather narrow and specialized reason for the use of high-speed systems, it has nevertheless caused the construction of several large networks. One of these networks will be described later.

Finally, we come to the last reason for high speed, which is, quite simply, that new and better ways of doing things often require information transmission systems of this kind. One area where this is becoming a factor is the publishing business; here the interest seems to lie in the transmission of graphics—pictures alone or words and pictures composed on a page. The present systems used for this are not really facsimile systems nor are they quite the same as telephotographic systems. The application of one of these systems will also be discussed.

Technical considerations

Most of the high-speed systems presently operating are working in the 48- and 240-kc/s carrier bandwidths. Their use represents a real expansion in the total amount of information handled by the carrier systems. However, the price to be paid for this increased efficiency is that the high-speed system must operate in a very restricted technical environment.

The underlying cause of most of this restriction lies in the basic design concept of the carrier systems used in the national network. The great majority of these use a frequency multiplex based on 4-kc/s spacing of the VF channels. Furthermore, the long-haul high-capacity carrier systems all use a single-sideband suppressed-carrier modulation technique, which leads to a rather special distribution of signal energy on the carrier system when it is carrying a voice load.

Figure 2 illustrates this situation for a typical 48-kc/s channel group assigned to voice use. The frequency assignment is shown in Fig. 2(A). The individual channel carriers are placed as shown; the channel 1 carrier is at 108 kc/s and the other 11 carriers are spaced at 4-kc/s intervals down to 64 kc/s, the channel 12 carrier. The voice signals are modulated to the lower sideband positions shown. The upper sidebands and the carriers are suppressed. The output of this channel group modulator is then modulated as a group to its final position in the carrier spectrum. No further individual modulation of the channels takes place.

The energy spectrum of this channel group, as seen by the transmission line and the receiving channel group terminal, is shown in Fig. 2(B). The signal power in each channel will be a maximum at 500 to 600 c/s, and will be substantially extinct at 200 and 3200 c/s. The result is an asymmetrical distribution of energy. Experience has shown that not all channels will be active simultaneously. The activity factor will be about 25 percent, as indicated in Fig. 2(B) by the areas in color. Thus, the sideband energy coming out of the terminal is in a constantly changing number of pass bands, all about the same shape; but on the average no more than three of these are active simultaneously.

The transmitting terminal also suppresses the channel carriers as shown by the suppression bands. One carrier in particular, 104 kc/s, is suppressed very carefully because a pilot frequency *P* of 104.08 kc/s is added later in the terminal for regulation and maintenance purposes. Finally, any energy beyond the edges of the channel group—that is, outside of the range of 60 to 108 kc/s—is also suppressed.

Thus the energy distribution in a channel group used for its normal purpose, voice, is indeed rather special. A high-speed system using the same carrier spectrum will not usually have an energy spectrum that even remotely resembles the voice load. For example, the spectrum of a typical high-speed data system is shown in Fig. 2(C). The envelope extends from 62 to 103 kc/s, but it is not really continuous. It will consist of various combinations of active energy bands (shown dashed), depending on the data being sent at the time. In these high-speed systems, a voice channel (channel 1) is usually furnished and used for coordination purposes. This is also shown, as is the pilot frequency. It may also be noted that, in addition to a different spectral distribution, the high-speed system will have a different pattern of use. It is likely to have a much higher activity factor. Taken collectively, these considerations make a very close control of the high-speed system imperative; otherwise, it will contribute noise to the other VF channels in the carrier system.

The VF channel noise is of several types, but only two are of interest at this point. First there is the inherent noise of the radio or coaxial transmission path itself, which can be controlled by the addition of gain elements in the system. There are a great many of these elements in modern carrier systems. The fact that, as a practical matter, they cannot be made perfectly linear leads directly to the second source of noise that is of interest here. The situation is illustrated in a simplified fashion in Fig. 3 through the medium of an input–output (or transfer) characteristic, which can represent, for example, the input of a carrier system at New York and its output at Los Angeles. Ideally, the situation shown in Fig. 3(A) should prevail. Here the transfer characteristic is perfectly linear and therefore a single-frequency input is undistorted at the output, as shown. Moreover, two input frequencies,

(Continued on page 87)

X and *Y*, emerge as only two frequencies at the output.

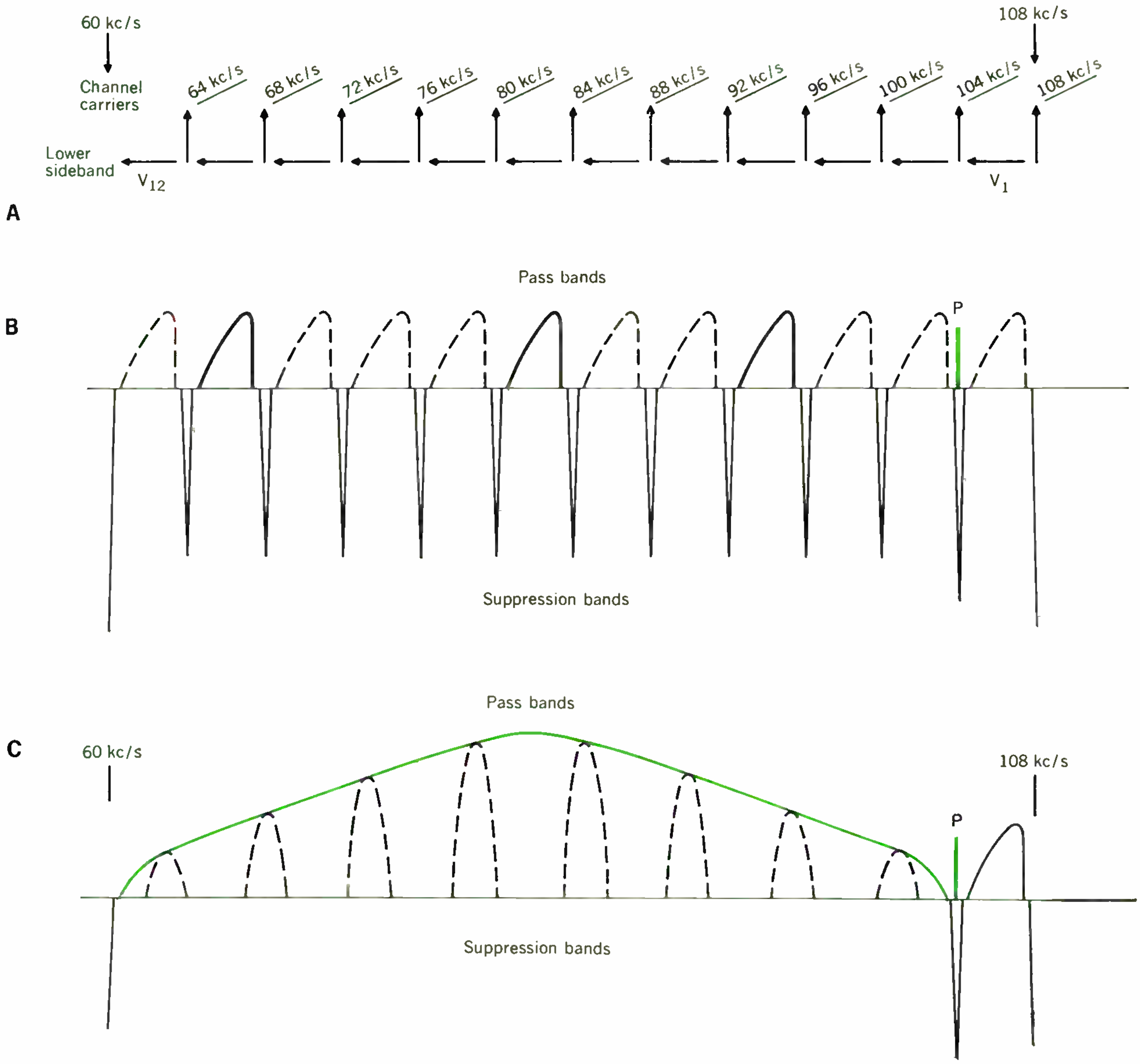
The actual state of affairs, however, is shown in Fig. 3(B), in which the transfer characteristic is curved rather than straight. A single-frequency input comes out distorted as shown, and a spectral analysis would show the existence of frequencies other than those in the original input. For *X* and *Y* these extraneous frequencies could possibly include all the modulation products shown. These unwanted products may ultimately show up as noise or crosstalk in the other voice channels of the carrier system.

The first means for controlling this noise is, of course, to make the transfer characteristic of the carrier system as linear as practicable. Second, by careful choice of modulation frequencies, the multiplex scheme is arranged so that as many of these extraneous products as possible fall between channels. Third, the load on the carrier system is carefully controlled. (The system linearity can be expected to change with the traffic load; it will usually be worse during the busy hour.) High-level single-frequency tones represent the most hazardous load of all.

Figure 4 illustrates what can happen in a carrier system if the load is not carefully controlled. Here the cross-modulation effect of a strong single frequency is shown. The translation of three voice-bandwidth channels (taken as 4 kc/s for convenience) are shown from baseband to transmission-line frequencies. At the extreme left, for example, the 4-kc/s baseband is first translated to 96–100 kc/s (channel 3 of the channel group), thence to 368–372 kc/s (group 2 of the basic supergroup), and finally to 248–252 kc/s for line transmission in supergroup 1. The direction of the arrows indicates whether the baseband spectrum is normal or reversed after the various modulation steps. The modulation frequencies are shown at the right. Three channels are shown; the other two end up in supergroups 3 and 7.

Finally, a strong 84-kc/s tone is shown, translated first to group 3 of the basic supergroup and finally to super-

Fig. 2. Channel group spectra. A—Frequency assignment. B—Voice spectral distribution. C—High-speed data spectral distribution.



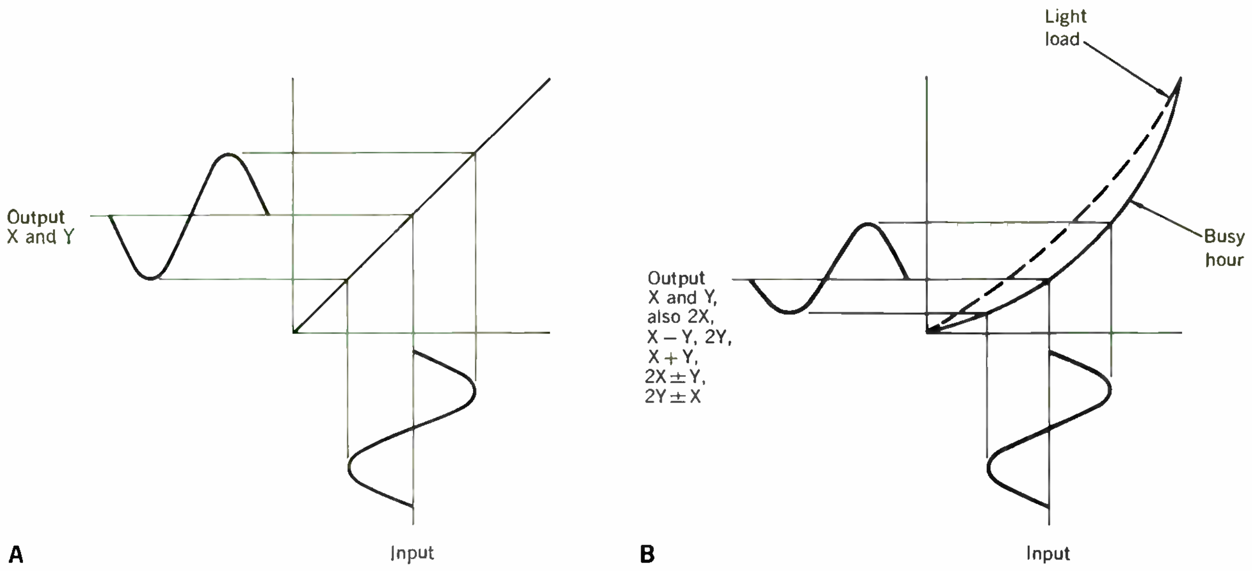


Fig. 3. Transfer characteristic under (A) ideal and (B) actual conditions.

group 4, where it ends up on the line at 932 kc/s. This is the villain of the piece; since it is at an excessive level it modulates with everything else on the line. As shown, one result is that the channel in supergroup 3 is modulated and appears not only in its normal position, but also (dashed arrows) inverted in supergroup 1 and correctly in supergroup 7. This modulation takes place in the transmission line; and once it happens it cannot be corrected. The result is that the channel in supergroup 1 is noisy and the channel in supergroup 7 now has intelligible crosstalk in it.

The consequences of strong single-frequency signals are much more complex than this. Since these signals actually interact with all channels on the system—not just two or three—their level has to be very carefully controlled. This also holds true for any other signal (as in data or telegraph transmission) that will be relatively constant on the line. In fact, it holds true for any signals that differ markedly from those generated by two human beings holding a telephone conversation.

Some of the consequences are shown in Fig. 5. The line designated "Message load" is the power load on the carrier system generated by the normal voice activity. A single-channel load, shown at 4 kc/s, will be about −16 dBm average power referred to the standard 0 dB reference level point. Increasing the number of channels to 60, which gives a carrier spectrum of 240 kc/s, will result in a total power load on the system of +2 dBm, as shown. Between these points, with the scale used, the load varies linearly.

The load will continue this linear increase up to the maximum capacity of the system: 600 channels or more. The system can handle this load quite adequately so long as it is a voice load with the spectral distribution shown in Fig. 2. If the load is anything other than this, some precautions have to be taken. For example, if single frequencies or data signals with strong single-frequency components are being carried, the permissible loading is sharply decreased. This is illustrated in Fig. 5 by the area in color. The lower limit (−14 dBm) applies to the single frequencies that are 4-kc/s multiples. Some liberties may be taken if the single frequencies are far enough away from 4-kc/s multiples, as represented by the upper limit shown. The actual limit applicable to a particular multiplex will usually be somewhere in the cross-hatched region. When several types of carrier systems are involved, it is prudent to choose the lower limit (−14 dBm) for all applications. The result is that a high-speed system with strong single-frequency components cannot exceed an average power level limit of −14 dBm. This limitation holds true for system bandwidths from 5 to 240 kc/s of carrier spectrum, as shown.

The noise that the system encounters, however, will change with the system bandwidth. It will increase roughly as ten times the log of the channel-bandwidth ratio. A 40-kc/s channel, for example, can be expected to have 10 dB more noise than will a 4-kc/s channel. This is shown in the system noise line drawn in the figure. The origin of this line on the left is the 4-kc/s channel noise for a typical 4000-mile (6400-km) radio carrier system (−44 dBm), extrapolated to wider bandwidths.

If we look at the 4-kc/s bandwidth, it becomes apparent that we have ended up with a projected signal-to-noise ratio of 28 dB or so. A 48-kc/s system, however, being limited to only a slightly higher power level, has a projected SNR of only 19 dB. For the 240-kc/s system the end result is an SNR of about 12 dB. Neither of these SNR values could be considered adequate in view of other system variables, which are not shown.

If the high-speed system can be arranged so that its signal spectrum has about the same effect as a pure voice load, the system power level can be raised back to the system load line. This technique results in a more adequate SNR. High-speed systems that cannot be so arranged must work in a very difficult technical environment.

There are, of course, other troublesome technical problems in the high-speed area—envelope delay distortion and regulation, to name just two. However, the loading and cross-modulation problems just outlined seem the most basic so far as analog carrier systems are concerned.

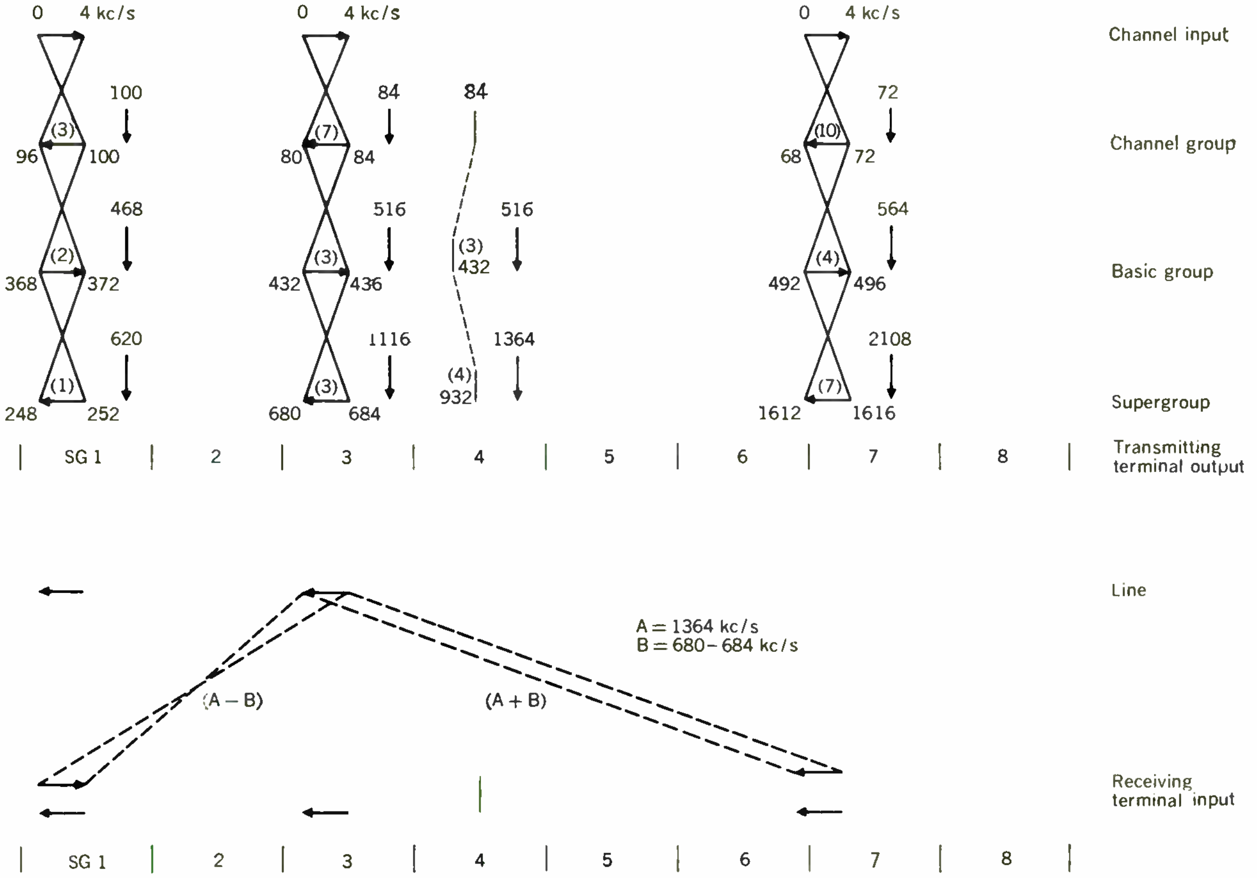
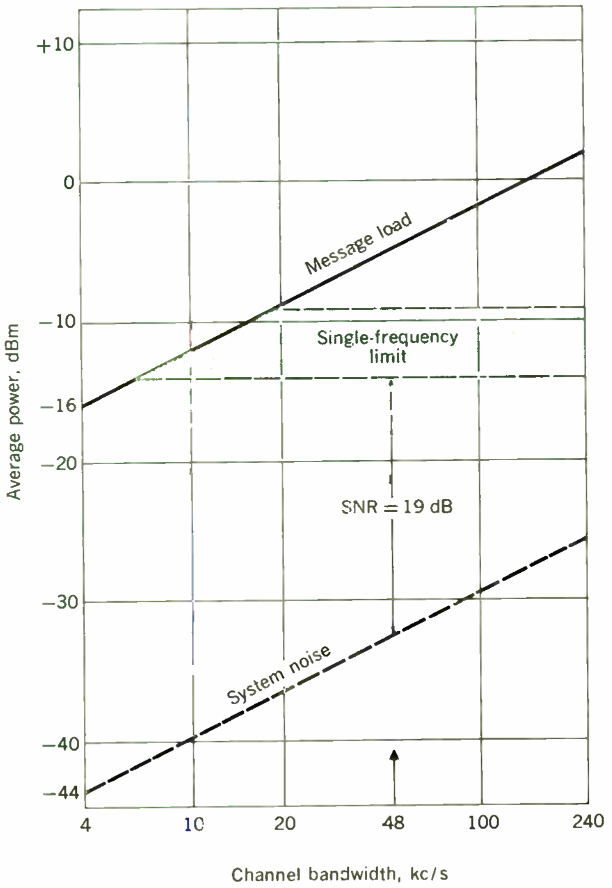


Fig. 4. Cross modulation. All frequencies are in kc/s. Numbers in parentheses indicate channel, group, or supergroup number.

Fig. 5. Noise margin.



Digital carrier systems using time multiplex schemes are not subject to the same limitations. Here the ground rules are quite different.

Up to this point we have discussed the evolution of high-speed systems, the reasons for them, and some of the relevant technical problems. It is now appropriate to discuss some high-speed applications.

High-speed parallel data systems

The parallel data systems considered here are designed for the transmission of data from seven-track magnetic tapes. They operate at speeds of 15, 41.6, or 62.5 kilocharacters/second. The most widely used speed at present is 15 kilocharacters/second (nominally 105 kilobits/second). Enough control functions are included to enable a computer to read information from a tape unit in a remote location in the same manner that it reads tapes from units in the computing center itself. These parallel systems are the oldest of the high-speed data systems; the first one was installed in 1958. Most of them also involve switching arrangements, many of which are quite elaborate. One of these will be described in some detail.

The signals on a perfect tape are shown schematically in section A of Fig. 6. Each character is an array of "ones" and "zeros" across the tape. In each track, successive 1's come out of the tape adaptor and control unit as pulses of opposite polarity (bipolar signals); 0's are a no-signal condition. The sketch shows 1's as the solid lines; 0's as the dashed lines. For a 15-kilocharacter/

second system the nominal bit rate on each track of the tape is 15 kilobits/second. In actual practice, slight irregularities in the relative positions of the reading heads in the tape unit may be expected, as may slight irregularities in the travel of the tape past them. One result is that the bit rate in each tape channel is not constant. It can exceed the nominal rate by as much as 30 percent.

A typical 15-kilocharacter/second system connecting two points, locations 1 and 2, is shown in the upper portion of Fig. 6. For clarity, only the circuits needed for transmission of tape information from 1 to 2 are shown. The remote tape unit is shown at the extreme left. Its seven data outputs, one for each track on the tape, are connected to the tape adaptor and control unit (heavy lines). One function of this unit is to shape the tape signals to the form shown in section B. The output of this unit is connected to the transmitting data terminal and transmitted over the line in parallel.

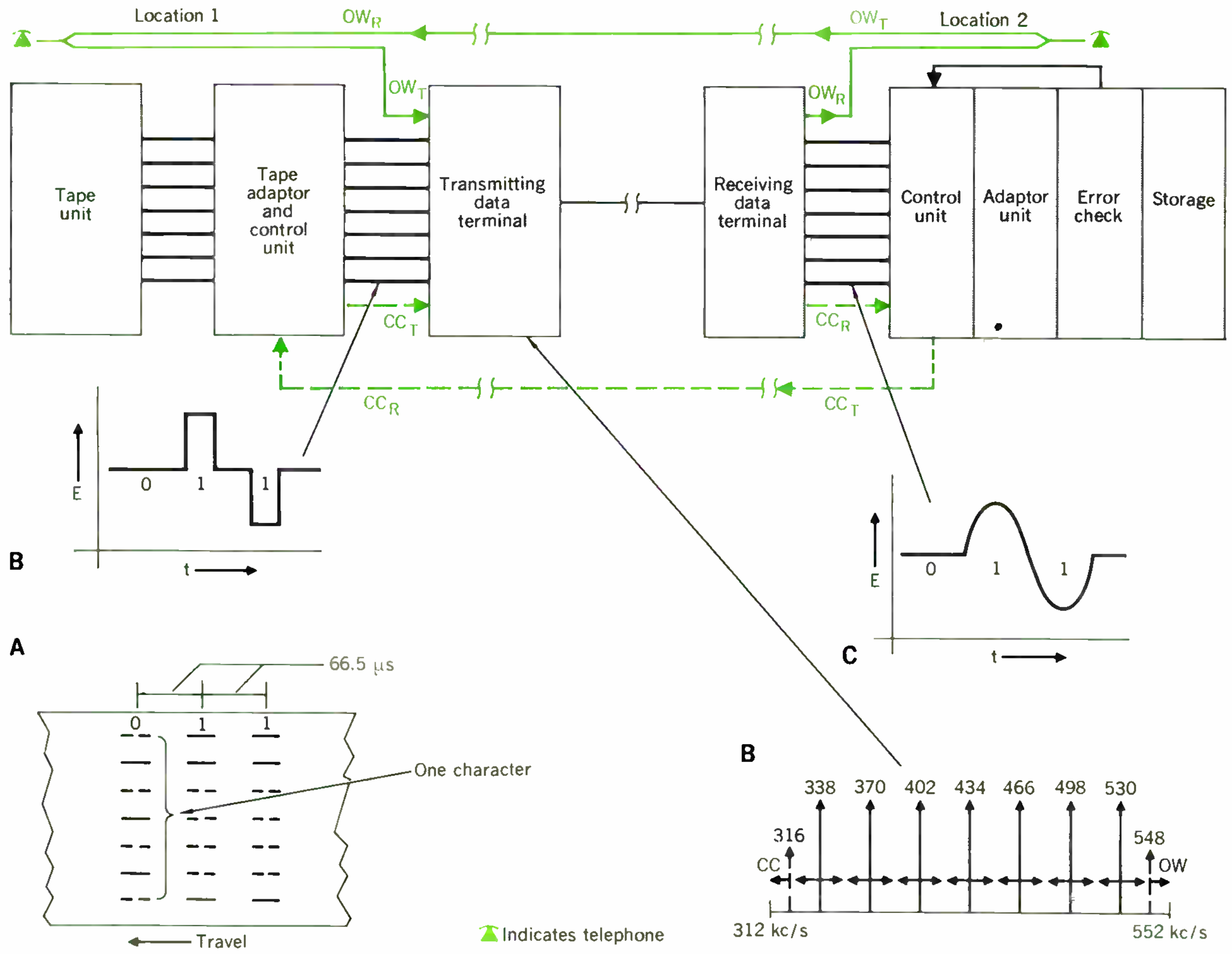
This system requires 240 kc/s of carrier spectrum bandwidth. One function of the transmitting data terminal is to translate all of the tape signals into the basic supergroup allocation, 312–552 kc/s. The frequency allocations within this range are shown in section D of Fig. 6. The seven tape signals are transmitted via transmitted-carrier double-sideband transmission systems with carefully controlled pass-band shaping. (These carriers are indicated by the solid arrows: 338 kc/s, 370 kc/s, etc.) In actual practice, this 312- to 552-kc/s output of the data terminal would be modulated to some supergroup position (Fig. 2) for line transmission. For reasons of simplification, these modulation steps have not been shown.

At location 2, the receiving data terminal demodulates the incoming signals (in the range of 312–552 kc/s) back to baseband. (The shaped tape signals at the output of this unit are shown in section C.) These signals pass through the control unit, are read in the adaptor unit, are checked for errors, and are finally either put in core storage for immediate computer use or read out of storage to another tape unit.

These systems include a voice-bandwidth circuit to transmit signals between the control units (CC in Fig. 6). This circuit, which is furnished on a four-wire basis, occupies the position shown in section D of Fig. 6. In some systems control signals go in both directions; in general, however, the control signals are only operative in the reverse direction from the data flow. A voice-bandwidth circuit for voice-coordination purposes (OW) is also furnished; see section D.

System operation can best be described as "start–stop." Data does not flow over the line continuously, but is transmitted in blocks. The longest block currently in use is approximately 1000 characters. Each block is coded in

Fig. 6. Typical parallel data system.



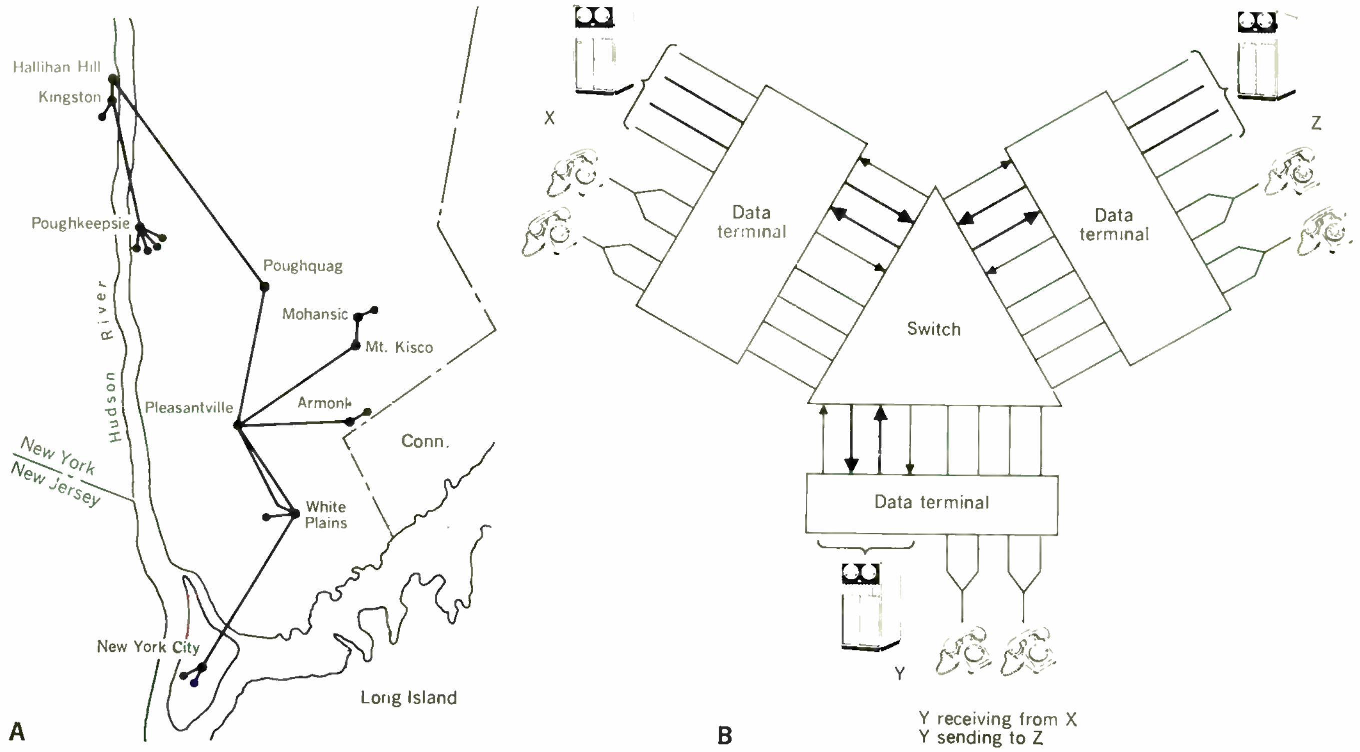


Fig. 7. Typical high-speed data network. A—Geographical layout. B—Switching scheme.

parity along the tape and each character is coded in parity across the tape. Failure to verify these checks at the receiving end causes retransmission of the block. Most systems are programmed to go through this cycle automatically a number of times (eight or nine times, for example) whenever an error occurs; if the block is still not received correctly, the system stops and actuates alarms. With this commonly used operational scheme, the "forward" speed of the system may be quite high. However, with present techniques its net speed or "throughput" cannot exceed half the forward speed, since transmission of the next block does not start until the received block is completely checked and read out.

In actual applications these parallel systems are not simple two-point layouts, but fairly complicated networks. One of these networks, shown in Fig. 7(A), was installed in October 1962 to interconnect nine computing centers of the International Business Machines Corporation in the lower Hudson Valley. The operational speed of this network was 15 kilocharacters/second. Switching capabilities were provided via a dial-controlled hub switching arrangement at White Plains, N.Y.

The switching scheme is shown schematically in the lower portion of Fig. 7(B). Any point on the network, Y for example, can send data to point Z and receive data from point X simultaneously. This is known as "split duplex" operation. The heavy lines connecting the tape control units and the data terminals are the seven data channels and the light lines directly adjacent are the control channels. For reasons that need not concern us here, this type of operation required the use of both VF circuits (CC and OW) in the data terminals to provide these control circuits. For voice coordination and switch control, each point also has two separate VF circuits to the switch. These circuits are also shown in the data terminal for simplicity and because they carry certain terminal alarms and maintenance controls, but they are not provided by the terminals; instead they are on a separate layout.

Note that in setting up any path, Y–X for example, the switch has to connect a four-wire VF circuit, a two-wire wide-band data circuit, and a two-wire control circuit. The VF circuit is switched at voice frequencies, whereas the other two circuits are switched at basic supergroup frequencies (312–552 kc/s) by eight-wire switching. A VF switching train is used to switch the VF circuits and extra switches are slaved on this train to switch the two high-frequency circuits. A two-digit dialing scheme is used.

The performance of such a complex network may be of interest. So far as error performance is concerned, detailed results, such as bit error rates and distributions, are not generally available for these high-speed systems. Their performance is generally measured in terms of the number of blocks in error; whether just one character or every character in the block is wrong is not generally known. On this basis the following tabulation, which covers a

III. Summary of troubles detected in New York–Poughkeepsie transmission path

Trouble Location	Percentage of Total
Data terminal apparatus	8
Switching apparatus	7
Wide-band facilities	
Interchange	44
Local loops	4
Voice-frequency facilities	20
No cause found	17

nine-month period, summarizes the performance of the longest transmission path (New York–Poughkeepsie). The blocks transmitted were quite short, generally 50 to 200 characters long.

Number of blocks transmitted	4 766 000
Number of blocks in error (all causes)	1157
Percentage of blocks in error	0.024

Although this tabulation indicates a very satisfactory performance record, it does not tell the whole story. Even though the number of blocks transmitted represents a vast amount of information, that leg of the network was not loaded to anything approaching a 100 percent duty cycle. When the duty cycle is light, trouble conditions can come in and clear out without being detected. Some idea of the extent of the troubles that were found is outlined in Table III, which summarizes the 59 troubles detected during the first 11 months of operation.

Note that although the largest number of troubles, 48 percent, occurred in the wide-band layout, the conventional VF layout also contributed 20 percent of the total. Thus, predictably, the facilities alone were responsible for

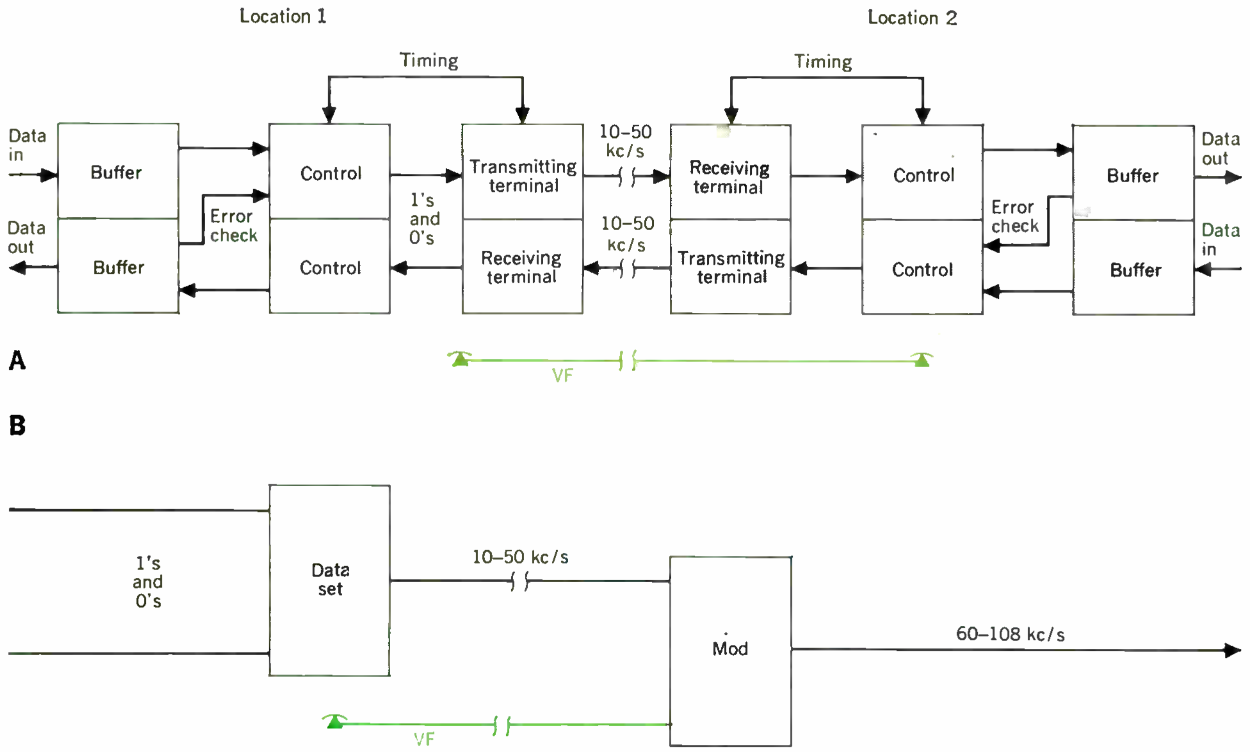
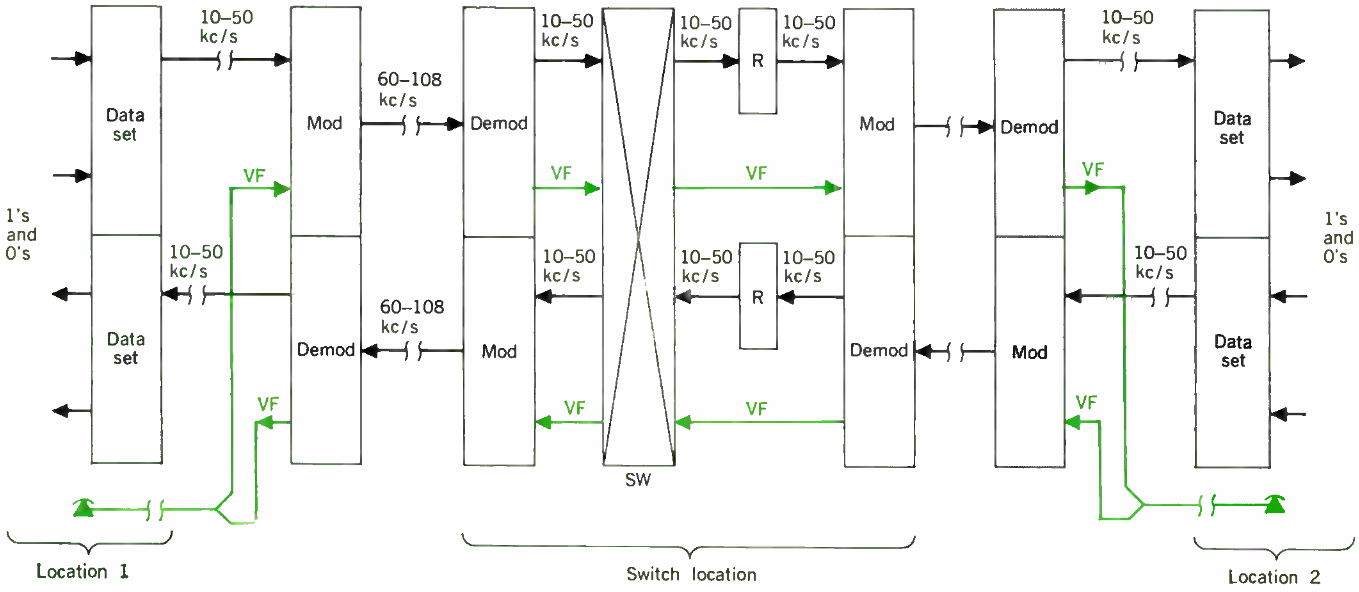


Fig. 8. Two-point serial data system. A—Overall system. B—Transmitting terminal shown in greater detail.

Fig. 9. Switched serial data system.



68 percent of the network troubles. Note also the high percentage (17 percent) of troubles not located; that is, the condition cleared before it could be pinpointed. This type of situation is to be expected in complex networks.

By late 1963 the expanding operational requirements of the IBM Corporation began to make this network inadequate, and thus it was replaced by a serial-type high-speed data network in December 1964. Since this new network may well be indicative of those to be expected in the future, it will be described in the next section.

A high-speed serial data system

The serial system to be described here works on a synchronous basis. The "forward" transmission speed is 40.8 kilobits/second, consisting of 5100 eight-bit characters/second. The input and output signals of the system at the interfaces with the connecting data apparatus are 1's and 0's. A logical 1 is a current less than 5 mA into 100 ohms; a logical 0 is represented by a current greater than 23 mA into the same impedance. In addition to the data, a variety of control functions are also available at the interface. All of these signals are transmitted over the line via a four-phase transmission scheme. The first step of modulation places them in the 10- to 50-kc/s range. The second modulation step—if the transmission is via one of the long-haul carrier systems—places them in the 60- to 104-kc/s portion of the channel group bandwidth. In this second step a VF circuit is also provided for coordination purposes; it occupies the portion of the channel from 104 to 108 kc/s. These allocations are shown in Fig. 2(C).

Figure 8(A) illustrates a typical system connecting locations 1 and 2. The transmitting and receiving terminals are shown in simplified form for clarity; they take and return logical 1's and 0's at the interface with the connecting data apparatus (shown as "control" and "buffer") in the sketch. Line transmission is in the range of 10–50 kc/s and only the first modulation step is shown. As noted, the transmitting and receiving terminals are always running in bit synchronism. If they are not transmitting real data, they transmit an "idle code" (10–00, 10–00) continuously to keep the transmitting and receiving clocks in step. This timing can be either taken from or fed to the connecting data apparatus. It should be understood that this connecting data apparatus is shown only in a functional fashion.

Transmission is on a block basis. The transmitting control unit (at location 1) first inserts a frame synch character in the bit stream followed by a "write" order. This is recognized by the receiving control unit at location 2, and if the buffer there is clear the control unit at 2 inserts a frame synch and "read" character in the bit stream toward 1. On receipt of this character, the control unit reads the block to be transmitted out of the buffer and transmits the block character by character. This information is passed to the receiving buffer and thence, usually, to storage. At the conclusion of the block, if the error check (EC) is verified, the receiving control unit generates another "read" order (again framed into the bit stream) and the transmitting apparatus goes on to the next block. This error check is based on a 4-out-of-8 code for each character. As shown, the system also includes a VF circuit for coordination purposes.

It may be noted here that the error and control signals are not fed over separate circuits, as in the parallel systems, but are interleaved in the bit stream, which is always continuous in both directions. This leads to "one way at a time" or half-duplex operation; that is, either of the locations sends data, but they do not send simultaneously.

The transmitting and receiving terminals are shown in simplified form in Fig. 8(A). The transmitting terminal is shown in somewhat more detail in Fig. 8(B). The 1's and 0's are actually fed to a data set, which is usually situated within about 100 feet of the connecting data apparatus. As noted, these input signals include control and status functions as well as the data. The data set modulates these signals in the 10–50-kc/s band. Present practice is to transmit this line signal for a maximum distance of possibly 20 miles. This distance is great enough to insure reaching the terminal of a carrier system. The carrier terminal shown (MOD) is that of a long-haul carrier. This terminal modulates the 10–50-kc/s signal, along with the paralleling VF circuit to the 60–108-kc/s channel group band. Succeeding steps of modulation that are not shown (supergroup and master group steps) place the channel group in the desired position for line transmission.

The switching arrangement used in the IBM network is shown in Fig. 9. Here locations 1 and 2, both remote from the switch location, are shown connected together through the switch (SW). From 1 and 2, only the first step of modulation, to the 60–108-kc/s channel group, is shown. The main interest at this point will be at the switch location. Both the data and the voice signals are translated back to the 10–50-kc/s and VF bands, respectively, and are so switched. This is "eight-wire" switching. The dial pulses that actuate the switch are carried by the VF circuit; here again two switch trains are used, the 10–50-kc/s switching matrix being slaved on the VF matrix. On long circuits (the circuit to location 2 is one of these), data regenerators (R) are added at the switch location. These regenerators are, in effect, modified data sets working "back to back," which effectively remove the accumulated jitter in the data signals.

The switch location for this network is at White Plains, N.Y. Table IV lists the locations connected to the switch at present, together with the approximate circuit mileages

IV. Service points and mileages for IBM network

Location	Circuit Mileage to Switch
Mohansic, N.Y.	16
Armonk, N.Y. (2)*	7
Poughkeepsie, N.Y. (5)	48
Fishkill, N.Y.	40
Endicott, N.Y.	140
White Plains, N.Y. (2)	1
New York, N.Y. (3)	25
Kingston, N.Y. (3)	63
Cambridge, Mass.	165
Washington, D.C.	228
Mechanicsburg, Pa.	180
Los Angeles, Calif.	2454
San Jose, Calif.	2556
Chicago, Ill.	720
Rochester, Minn.	973

*Figures in parentheses indicate the number of points connected in these locations.

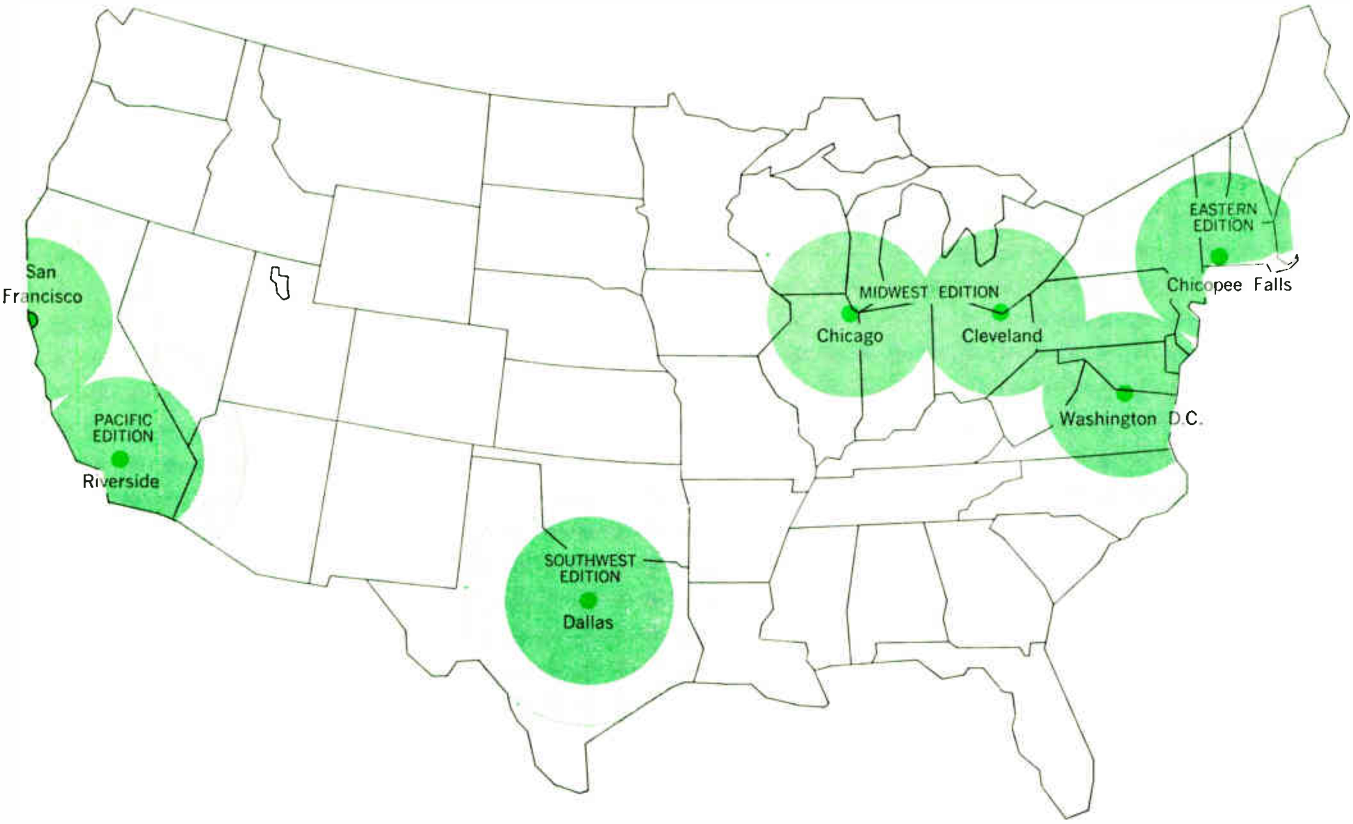


Fig. 10. Coverage areas for various editions of The Wall Street Journal.

for each. Unless otherwise noted (by figures in parentheses), only one point in each location is connected.

High-speed graphics—a press facsimile system

In concluding our discussion of high-speed information channels we will describe one that is an integral part of a modern publishing operation. The production and distribution of a daily newspaper is an intricate and involved process. The operation becomes even more complex when the newspaper serves a large geographical area. Such a newspaper is *The Wall Street Journal*. In terms of circulation, it is now the second largest in the United States.

The *Journal* is distributed to most of its readers via second-class mail and the objective is for it to arrive in the first business mail of the day. To achieve this, the papers have to start through the postal system at about 9 P.M. on Monday, for example, to reach the readers in the first business mail on Tuesday morning. The mechanics of the postal system place a limit on the geographical area that can be served in this manner. Roughly speaking, this area is within a 200-mile radius of the point at which the newspaper is printed. Considerably less-certain coverage may be attained within a 300-mile radius. Larger coverage areas can be attained only by the addition of more printing plants. Figure 10 shows the points at which the *Journal* is presently printed and the approximate coverage areas of each plant. Areas that are within a 200- and 300-mile radius are shown in a dark and light tint respectively.

Four major editions are published daily in seven printing plants. The Eastern Edition is printed simultaneously in Chicopee Falls, Mass., and Washington, D.C. The Midwest Edition is printed simultaneously in Chicago and Cleveland. The Southwest Edition is printed in Dallas; the Pacific Coast Edition is printed simultaneously in San Francisco and Riverside, Calif. All news and editorial copy is handled by the national editing staff in New York City. All editions are supplied with the same news material, so that the news content is identical in all editions, insofar as is practical. Advertising copy varies from one edition to another, depending on whether the space sold is for national or regional editions. The *Journal* concentrates its efforts on news of interest to the business and financial communities. Each printing plant has an editorial staff responsible for checking the composing-room work and layout of the news content.

In 1961 the *Journal* coverage area was not quite as shown in Fig. 10. The only printing plant on the West Coast, at San Francisco, served both the San Francisco and Los Angeles areas. The Los Angeles area was being served at a price, however. The papers destined for this area could not be mailed in San Francisco; they had to be flown to Los Angeles daily and mailed there. (Nine tons of newspapers daily!) Although this approach worked most of the time, the obvious long-term answer to the problem was to establish another printing plant in the Los Angeles area.

The Wall Street Journal could very easily have established another conventional printing facility in the Los Angeles area complete with a composing room, stereotype department, and a news production department. But for a long time its managers had been interested in the development of facsimile. It seemed obvious that a system that would permit transmission of a whole page at a time—skipping the laborious typesetting, correcting, and assembling process—would save time and eventually, perhaps, costs.

Accordingly, a new experimental production plant was built at Riverside. It is equipped with complete press facilities and can make the plates that the press requires, but it has no composing room or editorial staff. The existing San Francisco composing room and editorial staff serves both locations. As each page of the Pacific Edition is set up by the San Francisco composing room a "proof" is printed on high-quality paper. This is then transmitted to Riverside over a high-speed high-definition facsimile link.

The facsimile system employs a photographic process. Thus each proof page transmitted from San Francisco comes out as a full-size negative of the newspaper page at Riverside. The Riverside plant produces the plates for the press from these negatives via a photoengraving process. To produce the paper, an average of 44 pages are transmitted daily. Speed of transmission is approximately 4.7 minutes per page, and the total transmission time is 3 to 5 hours. The press run at Riverside starts no later than 1½ hours after the San Francisco press run begins.

To produce printing plates of adequate quality, very high definition is required in the negatives. The process used can give a definition of 1000 to 1500 lines per inch. It reproduces either black or white; as the system is presently adjusted, it does not reproduce a gray scale. The usual newspaper page, of course, does have photos, etc., that give the illusion of a gray scale by the halftone process. The facsimile system is capable of resolving these halftone dots and so transmits the graphical material with little observable degradation. The line-transmission scheme is via a transmitted carrier, double-sideband system with the carrier at 2.0 Mc/s. Approximately 2.0 Mc/s of line spectrum is required; if necessary, two of these systems can be fitted in the normal video bandwidth. The line signaling speed presently used is approximately 1.3 million elements per second.

The use of video circuits for this purpose represented an excursion into rather unfamiliar territory. It was known that there were occasional line "hits" or other transmission defects that could cause flashes, picture rolls, etc. These defects, though annoying, did not greatly harm video program material; but they would be recorded on the facsimile negatives, which had to be substantially perfect to produce good plates. Experience has indicated, however, that the transmission path is stable enough to provide adequate service. Service was started in May 1962, and for a short period the defective negatives averaged about one out of every four sent. During this period a succession of line troubles—minor in nature so far as video was concerned, but fatal to this service—were cleared up, and the performance improved considerably. During the first nine months of 1963, the failure rate dropped to 0.05 percent; out of a total of 5179 negatives, only 28 were defective. This level of performance continues at present.

Looking ahead

It may be safely concluded that high-speed systems will be with us for a long time. They will probably not ever be numerous compared with the number of voice or telegraph systems, but their number will continue to increase.

Predictions are risky in this field because the reasons for high-speed transmission itself are so complex and variable. In the conventional data areas—such as computer links—the data speeds for networks covering large geographical areas appear to be those that can be provided in the 48-kc/s channel group bandwidth. Economics seem to be against any higher speeds for the present. For networks covering limited geographical areas, however, the speed trend is upward, and systems using supergroup (240-kc/s) and wider bandwidths are either in operation or in the design phase.

With respect to numbers of systems, the graphics area probably has the greatest growth potential. The publishing business will account for some of this growth, but most of it will probably come in lower-definition dry-copy systems for the general business community. These systems were not described here, but several large networks are already either operative or in the engineering stages. One network of particular interest is the railroad communication system referenced in the bibliography.

The author wishes to thank those people who so kindly reviewed this article and pointed out his mistakes. The permission of the AT&T Company, the IBM Corporation, and *The Wall Street Journal* to use certain relevant material is also gratefully acknowledged.

BIBLIOGRAPHY

Abraham, L. G., "The complexity of the transmission network," *Bell Lab. Record*, vol. 38, pp. 43–48, Feb. 1960.

Affel, H. A., and Kendall, B. W., "A twelve-channel carrier telephone system for open-wire lines," *Bell System Tech. J.*, vol. 18, pp. 119–142, Jan. 1939.

Baldwin, G. L., and Snow, N. E., "Remote operation of a computer by high-speed data link," *Proc. 1962 Fall Joint Computer Conf.*

Bennett, W. R., "Noise characteristics and origins," Monograph 2567, Bell Telephone Laboratories, Inc., Murray Hill, N.J., 1956.

Bennett, W. R., and Davey, J. R., *Data Transmission.* New York: McGraw-Hill, 1965.

Caruthers, R. S., "The type N-1 carrier telephone system: objectives and transmission features," *Bell Telephone Tech. J.*, vol. 30, pp. 1–32, Jan. 1951.

Coe, R. H., Lawrence, G., and James, R. T., "A high-speed (105,000 bit per second) data transmission system," presented at the 1960 AIEE Winter General Meeting, New York, N.Y.

Crane, R. E., Dixon, J. T., and Huber, G. H., "Frequency division techniques for a coaxial cable network," *Trans. AIEE*, vol. 66, pp. 1451–1459, 1947.

Elmendorf, C. H., Ehbar, R. D., Klie, R. H., and Grossman, A. J., "The L3 coaxial system—system design," *Bell System Tech. J.*, vol. 32, pp. 781–832, July 1953.

Evans, W. E., and Pace, N. V., "Total railroad communication," presented at the annual meeting of the Communication and Signal Section, Association of American Railroads, Sept. 1965. See also "Southern's revolutionary room at the top," *Railway Age*, pp. 26–41, Sept. 13, 1965.

Green, C. W., and Green, E. I., "A carrier telephone system for toll cables," *Bell System Tech. J.*, vol. 17, pp. 80–105, Jan. 1938.

Hoth, D. F., "The T1 carrier system," *Bell Lab. Record*, vol. 40, pp. 358–363, Nov. 1962.

James, R. T., "Data transmission—the art of moving information," *IEEE Spectrum*, vol. 2, pp. 65–83, Jan. 1965.

Mertz, P., "Information theory impact on modern communications," *Trans. AIEE (Communication and Electronics)*, vol. 76, pp. 431–437, Sept. 1957.

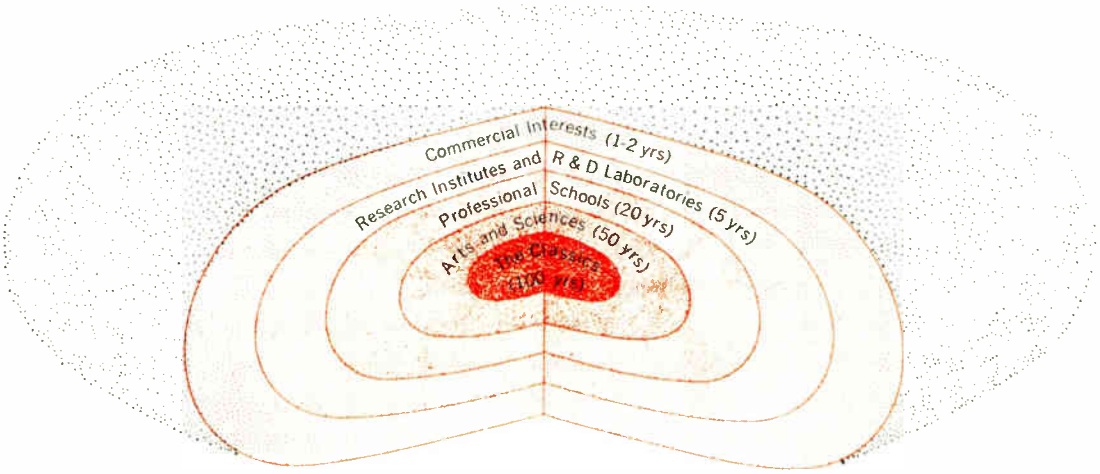
Nast, D. W., "Wideband data transmission, 40,800-bit-per-second service," presented at the 1964 Internat'l Symp. on Global Communications, University of Pennsylvania, Philadelphia.

Pierce, J. R., and Karlin, J. E., "Reading rates and the information rate of a human channel," *Bell System Tech. J.*, vol. 36, pp. 497–516, Mar. 1957.

Reid, J. D., *The Telegraph in America.* New York: Derby Brothers, 1879.

Shonnard, J. R., "High-speed communication of graphic intelligence with hard copy readout," *Trans. AIEE (Communication and Electronics)*, vol. 81, pp. 176–178, July 1962.

Travis, L. F., and Yaeger, R. E., "Wideband data on T1 carrier," *Bell System Tech. J.*, vol. 44, pp. 1567–1604, Oct. 1965.



Engineering–economic systems: a new profession

Planning in the upcoming age of automation will require the integrated efforts of academicians and practitioners, focused on real problems

W. K. Linvill *Stanford University*

Technological advances in computers and automation over the last 20 years have far outstripped our ability to use them effectively. Profound changes in industry, government, and business will emerge as the new technology unfolds. As an example: the whole concept of a "big" operation or company will have to be reshaped in terms of our new ability to organize and operate a system efficiently. To exploit the possibilities of this new technology rationally, broad interties between the academic and practical worlds are indispensable. The academic world contributes a philosophical structure providing vital breadth and flexibility, while the real world offers a simplifying practicality that insures workability and early realization. In the initial stages, planning for the use of the new technology must be done as a joint research and development effort between both worlds. As this joint effort grows, a more complete and integrated structure will undoubtedly evolve. A prototype form of the engineering–economic systems profession is being evolved in a program at Stanford University.

We are at the threshold of a new profession in engineering–economic systems that involves more management than the engineering professions, more technology than is usually involved in management, and more applied mathematics than either engineering or management usually involves. The problems of planning in this new era require a much broader view than has been taken previously, either by managers or by engineers.

We are in danger of missing the really significant advances now feasible if we make simple extrapolations of present practices instead of trying to capture the scope of the new technical potential at the outset.

The range and direction the field is taking will be described in this article although the details cannot and need not be spelled out since they are changing and evolving as needs of the new technology are being met.

The new technology and its needs

System problems. Consider the problem of developing an automated inventory and distribution system. While there are several strictly technical innovations to be made in automated distribution systems, the main problems are with the system and not simply with the devices it includes. To reap really significant benefits from computerized inventory control, one must combine it in many applications with a high-speed transportation system. Although such a system is technically feasible, it also requires a fast ground pickup and delivery system, better packaging methods, possibly a much broader base in air traffic control than is now provided, and perhaps even especially designed cargo aircraft. For such an automated system to be workable requires a mixture of private enterprise and publicly owned and operated support facilities.

Were an engineer or manager within a given company to view the development of his company's product distribution system, he would necessarily take as constraints to his planning the outside transportation facilities, both available and projected. At present, however, the types of transportation systems needed still await translation from feasibility to reality. A broader view of the problem—consideration of several regions, of the distribution needs of several companies, of several types of transportation and terminal delivery systems, and of the elastic relationship of demand for new products to their prices and breadth of availability—would not only lead to a better system but to one that would serve longer as a result of its comprehensiveness.

Several pertinent questions should be asked: Where should a common carrier take over? What part of the delivery system should be operated by the manufacturer of the goods to be delivered? What should be the balance between ground and air systems? Where distances are long enough and the premium on short delivery time high enough, air freight is preferable to ground freight. If the demand for air freight in a given region is great enough to warrant them, large-scale terminal facilities and common ground support transportation systems can be built economically. What are the alternatives to present practices in ground transportation? How do storage and delivery procedures and costs affect the balance between additional expenditures for warehousing and additional expenditures for shorter delivery times?

These are but a very few of the important problems to be faced in the design of an automated distribution system. These problems involve technical, economic, and political issues. Many of an established system's components will be used for many years under conditions sharply different from the initial ones. The demand pattern, for example, will undergo dynamic changes and will be subject to unavoidable uncertainties. A technical language is needed to describe, analyze, and forecast such situations, and to make the design decisions necessary to specify workable systems. The new professional discipline in engineering–economic systems is being developed to meet needs such as these.

Collaboration of academic and industrial worlds. The consequences of the new technology are both broad and revolutionary. Adequate planning for the exploitation of this emerging technology requires a combination of the comprehensive breadth of the academic world and the active practicality of industry, commerce, and government. The problem is to combine the strengths of the two components without seriously weakening their long-term basic functions.

Let us consider the relationships between the various groups in our society who are concerned with academic or technical matters. Imagine the various kinds of academic community activity as layers around a core. The classics provide the core of the intellectual activity that abstracts data and formulates basic principles. Workers in this area are sheltered from the outside world and they are characterized as carrying on extremely long-term activities. In engineering terms, we would say their time constants may be of the order of 100 years or so. They make changes to adapt to the world on a very deliberate basis. Another layer of intellectual activity is provided by the schools of arts and sciences in which the time constants are substantially shorter than those of the classics. Mathematics, physics, economics, and political science are examples of such areas. Perhaps time constants of the order of 50 years are appropriate to characterize this layer. A third layer is represented by the professional schools. These are devoted to professional training of students for the real world in ways that will be useful during their professional lifetimes. Time constants of the order of 20 years are appropriate to this sort of activity. Still another outer layer of the intellectual activity is represented by work in research institutes and the research and development laboratories of industry and government. They usually have time constants of the order of five years. Finally, companies with commercial interests and large work forces and government entities taking significant direct action must see responses from their efforts within only one or two years.

Let us picture the nature of the relationships between successive layers of the technical society. A given layer receives real-world stimuli *from* the outer layers and in turn provides philosophical structure *to* them for monitoring and extrapolatory purposes. During periods of high stresses, such as those imposed by rapid technological change, there is a very strong need for a given layer to be strongly coupled to both the next-inner layer, from which it receives philosophical direction, and the next-outer layer, from which it receives stimuli from the real world. Without the essentially conservative and contemplative influences of the inner layer, a given layer may be corrupted by short-term external stimuli, which, although strong, are not relevant to the long-term extrapolations. Without the strong signals and drive for action from the outer layer, a given layer may be isolated from the real world. Research—or theory—is generally involved with the process of relating to the inner core; practice is generally involved with relating to the outside world.

Professional schools have always served the extremely important function of coupling the real world to the academic. Without disciplined development of philosophical underpinnings—the abstractive, restructuring, and generalizing function of the academic community—the experience and data gathered in the practical, real world must remain disconnected and their usefulness limited. The exceedingly rigorous, general, and pure output of the departments of mathematics and sciences has unnecessary generality and too long a time constant for effective application to the rapidly changing technology of the real world. The professional school adapts the rigor of mathematics and science to the real world in a less general form with somewhat shorter time constants (more flexibility) so that the professional man can respond effectively to the practicalities of the real world. The professional schools have the somewhat specialized function of doing research to develop the philosophical aspects of practical fields and of training future practitioners of the new professions. In their development of the new field, these schools face the inevitable necessity to relate this new field to the established philosophical framework of the arts and sciences. Thus professional schools couple the arts and sciences to the real world and at the same time prevent the corruption of the arts and sciences that would occur if these layers dealt with the real world directly in enough breadth to do their data collecting alone.

The aforementioned relationship between theory and practice applies regardless of the professional area and it has been working effectively for many years. Uniquely in the system area, however, the relation of theory to practice must take place simultaneously across a broad range of areas in the academic community, involving mathematics, economics, physical science, and political science, and an equally broad range of areas in the real world, involving physical control, operational control, and public and private planning in many industries and regions. Because of this breadth of connections to both entities, special efforts to couple them are required. These special efforts must be expended because the needs and rewards of developing the new technology justify them.

Fast-changing technology imposes new demands and provides new opportunities for industry. Much as in the case of the university, industry is faced with adapting old patterns of operation to new situations. For example, automation provides remarkable advantages for industry but it also brings substantial financial risks as well as fast obsolescence of both human and physical resources. Because of the risks involved in getting into new areas, careful preliminary exploration is vital. The philosophical breadth furnished by the university is extremely valuable in such exploration. The technical capability of its engineering staff is a capital resource of a company. This resource is depreciated by obsolescence, which can be prevented only by continual training and research to keep the engineering staffs up to date. Thus, industry needs the university for both exploration and retraining.

As in the case of automated distribution systems, many of the most important system opportunities require joint action by several separate industries, and probably involve government agencies as well. Before any individual private industry, even a large one, could commit strong research effort or funding to such a joint venture, the system would need to be laid out in considerable detail, and various alternative structures and forms would need to be invented and analyzed by a responsible but neutral agency. At present, there is no entity set up to do such initial analysis and planning. A joint academic–industrial combination would be invaluable in this exploratory work.

Adequate planning for the exploitation of the emerging technology, therefore, requires a combination of the comprehensive breadth of the academic world and the active practicality of industry, commerce, and government. This university–real world interaction will exist in many different forms, depending upon the needs of the particular problem and the constraints of the companies and agencies involved. If the problems are primarily action-oriented in the real world, their solution will be supported and controlled mainly by industry or government. If a problem is primarily analysis-oriented or training-oriented, its solution will come primarily from the university. Whether a given system program is centered in the field or in the academic world, there must be strong components of both academic and real world participation for effective work.

Now let us consider what can be done in planning the development of the systems field as a profession—a system problem in itself.

The engineering–economic systems profession

The broad scope of the systems work to be done has been shown to require strong joint academic–real world effort regardless of whether the main emphasis is academic or practical. We turn now to the problem of establishing a new profession.

The development of the system profession involves a close interaction between three adjacent layers, as previously discussed. While we are concerned primarily with the center or professional school layer, we must also relate to the two adjoining layers. Foundation disciplines from the schools of arts and sciences will provide the coupling to long-term academic structures that is essential to build flexibility into the system man's capabilities. Also, substantial encounter with the real world through casework in the field promotes development of relevant tools and approaches to present-day system problems. These two outer layers are bound together by the system profession, which on one hand adapts the abstract language of mathematics and the concepts of physics and economics into forms more applicable by the professional, and on the other hand combines, compares, and systematizes the approaches to specific field problems into a broader and more general set of tools and system concepts. The interrelations between the three layers are represented in Fig. 1. We will present in detail the function of the three layers in the development of the system field. First, we will consider the foundation disciplines; second, the work in the field, since it already exists; and finally, the engineering–economic system core that links the two.

Flexibility through emphasis on foundation disciplines. Engineering training has been typical of many professional training programs. The strongest coupling to arts and sciences in the past was the requirement for engineers to have a good background in mathematics and physics. Once the students obtained this start, however, there was often little further interaction between engineering schools and mathematics and physics departments. Largely, the students were prepared for their real-world encounters by laboratory work, by practical engineering problems tailored for the classroom, and, once they got into industry, by a modest amount of on-the-job training. With the traditional specialization, this procedure worked well and was adequate, except for a few individuals who wanted special training for engineering research. These students would take graduate work in science or a mixture of science and engineering.

At the outset of World War II, when new technology was being created to meet the needs of the war, the engineers who were being called on to help in the new developments often found themselves badly outclassed by mathematicians and physicists whose broader academic approach gave them much greater flexibility to move from field to field. As a result, the physicists and mathematicians did much of the engineering research and development during the war, and after the war there was a swing toward "engineering science" in the engineering schools. While, on the whole, this swing to academic breadth was healthy, it still is not broad enough for present needs. As systems become larger and more significant, inevitably economic and political considerations become increasingly important. For the system man to be adequately founded in economics and political science is obviously necessary.

Generally, the stronger background a system man has in the foundation disciplines, the more easily he can shift from field to field in the real world. Experience in the academic world, in which the disciplines are devoted primarily to being philosophically consistent and complete, provides a man with a life-long set of concepts and values that can accommodate fairly radical changes among casework areas in the real world.

Coupling to the real world. Effective work in the real world results from structuring specific real world problems according to the set of basic principles derived from the academic core. Yet the principles developed solely in the academic environment do not fit the real world situation adequately. The problem of approximating the real world situation in terms of academic principles is very challenging indeed. Several examples will show the difficulties.

It is possible to coordinate the operations of an entire steel-making plant by means of a central data-processing system. Dynamics problems arise in connection with the production scheduling that do not fit the traditional problems of physical dynamics, and yet have important similarities to these patterns. A comparative study of various types of dynamical models would be extremely valuable to the system analyst but not particularly interesting to a mathematician.

Allocation problems often lead to linear or nonlinear programming formulations. The choice of retaining nonlinear cost functions and reducing the number of variables to one extreme or of linearizing cost functions and retaining many variables at the other is a difficult one. Which approximation tactic would be preferable would depend upon the practical issues to be resolved.

In many physical control problems—as well as in inventory control and marketing problems—the engineer or manager has a very interesting choice. He can either control the process—with insufficient or noisy data on its state—by elaborate but ineffective control processes, or he can invest substantial measurement effort to determine the state more adequately. Although a specialist in dynamics and probability would have the general background required for this problem, any solution he would formulate would require a very high order of imagination and clever analysis.

Since the fit of classical models to new situations is only very approximate, and since optimization involves more than strictly mathematical issues, training in foundation disciplines is, while necessary, clearly insufficient. Coupling to the real world is necessary for the relating of theory and practice in order to develop more useful tools for analysis and design.

Given that the professional school must relate both to arts and sciences and to the real world, our next question is one of *method*. As was mentioned before, the uniqueness of system problems arises from the fact that relation of theory to practice now must take place simultaneously across a broad range of areas in the academic community and an equally broad range of areas in the real world. For engineering professors in the conventional disciplines, practical testing of theory is fairly straightforward in a single technical area. One merely goes to the laboratory. In the system area, the need for comparative study of several systems in real world problems makes the world the laboratory. For practitioners in conventional disciplines the relation of theory to practice is also fairly easy. A practicing civil engineer can find out about structures or materials by simply consulting a specialist on structures or materials in the university. But now the planner of automatic inventories and high-speed distribution must rely on a team of specialists and he will find it very difficult to integrate their contributions. Integration is the function of the system specialist.

In this day of large, complex, and expensive systems, the problems of the real world simply are not portable. Whereas the older problems in engineering could be worked out on the campus, the newer ones can be met with only in the field. This task of utilizing the world as a laboratory is new to engineering schools but is similar to that encountered by medical schools for many years. For both teaching and research in medicine, the connection of university medical schools with practicing hospitals is well established, with the students as interns and the professors as practitioners. This pattern of cooperation is one that engineering schools should consider. Graduate students in the system area should spend approximately 20 to 40 percent of their doctoral program in the field where they accept and discharge project responsibility. Their project-directed casework in industry should be followed by concept-directed casework at the university that translates specific real-world practice into broader engineering principles through thesis research. Commonly accepted consulting activities of engineering professors can be extended so that as much as 20 to 30 percent of the professor's time can be applied to field problems. If a specific plan is set up so that the professor's industrial activity is related to the work of his graduate students, his industrial consulting can provide a continuing vital input to the development of his professional field. While the professor would assume responsibility for *projects* in industry, he would not assume responsibility for industrial *programs*, and thus be free to utilize his main energies as architect of the professional discipline.

From the standpoint of industry, the exploratory work done at the research laboratory in collaboration between industrial and academic personnel can be carefully evaluated by industry from the extensive analytical data provided. Those promising projects can be further developed, and followthrough in industry for profit is the natural outcome.

In consideration of the nature of the work to be done in the field, the real world provides a very extensive menu of practical areas. It is not possible that any one man have experience in all these areas. Indeed, the entire field program of one university cannot encompass them all. The object of a field program is to immerse each student in the environment of the real world where all general types of system problems are represented, even though all technical areas cannot be. These types of problems were discussed previously at some length.

Developing the engineering–economic system core concepts. Core concepts in systems are the bridge between the foundation disciplines and the work in the field. (See Fig. 1.) The system language is the language of mathematics.

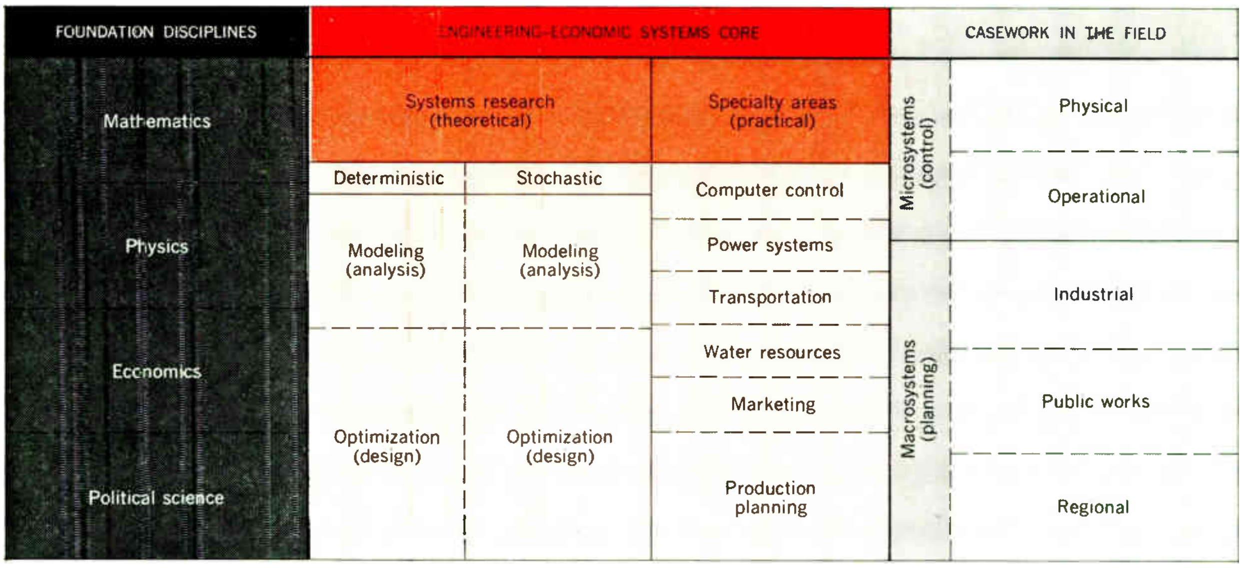


Fig. 1. Relationship of the engineering–economic systems core to both the academic and the real world.

Yet pure mathematical language and concepts are structured primarily to provide generality and rigor, often at the expense of introducing complication, while system concepts require flexibility and simplicity to be maximally useful. Accordingly, we are willing to sacrifice generality to gain simplicity and flexibility, and so we develop a set of system concepts that are closely related to, but still distinct from, mathematical concepts.

The two main functions of mathematics in systems are modeling and optimization. System analysis is generalized from each specific case by having a set of typical system models. Models of matrices and determinants, eigenvalue analysis and normal coordinates, iterative procedures for solving simultaneous equations, and analysis of signal flow graphs are typical examples. System models are either deterministic or stochastic; Markov processes are probably the most common example of stochastic dynamic models.

Optimization is to design what modeling is to analysis. Most optimization procedures can be related to the generalization of finding a maximum by finding a zero derivative. Constraints are handled by Lagrange multipliers. Linear and dynamic programming are particularly useful when computers are available.

Decision making under uncertainty is a combination of modeling and optimization. Here, the concept of utility is a cornerstone. The Bayesian models allow one to describe the consequences of new information. Optimization of a payoff function is essential in the decision-making process.

For both modeling and optimization, typical problems from physical and social science (the foundation disciplines) and from systems reliability, marketing, automatic control, resource allocation, maintenance and replacement policies, search procedures, inventory control, and scheduling (the field problems) must be related to the primary system concepts.

We would not claim the same life for system concepts that we expect for basic concepts of mathematics, economics, and physics. The value of system concepts is dependent on their relevancy to the real world. The 20-year life estimated earlier probably is appropriate to them. As one gets closer to the field problems, life becomes shorter.

Overlaying the whole set of system concepts related both to modeling and optimization is computer-aided system analysis. The speed, capacity, and flexibility of the digital computer have a profound effect on system analysis and design. Simulations of scheduling and dynamical systems as well as optimization in many variables are made feasible by the availability of computers.

Up to this point, we have been concerned with system theory, which relates analysis and design to the foundation disciplines. An equally significant facet of the engineering–economic system core is its connection to the real world. Practical areas such as physical control, power systems, transportation planning, and regional development are relevant. We would not expect the time constants of practical areas to be as long as those of the theoretical areas. System practice would represent a comparative study of specific casework projects. At the same time, system theory would be common to all system practical areas.

Numerous considerations have been presented in detail in the foregoing sections with the object of making clear (1) that a new professional entity is indeed involved here; and (2) that close interaction between the academic world and the practical world of technology is most essential for many reasons. Not only is this relationship vital in its implications for complementary activity between these two sectors, but also in its political and economic implications for society at large. With these assumptions fairly well established, then, let us consider a specific answer to the problem.

The engineering–economic systems program at Stanford

We will now describe a particular program in some detail to show how the field is developing at one large engineering school. Obviously, there is strong need to adapt the development of the program to the surrounding conditions a given university provides, and the reader should not assume that the plan described here is necessarily the best plan for other schools to follow.

Organization and objectives. The Institute in Engineering–Economic Systems at Stanford University is an interdisciplinary, interdepartmental organization developing a broad graduate research and coursework program with three specific objectives: (1) to develop interdisciplinary research activity in systems; (2) to establish industrial and governmental internships in the field to couple theory and practice; and (3) to present in courses system concepts derived from foundation courses in mathematics and physical and behavioral sciences as well as from practical casework. The interdisciplinary organization includes affiliated faculty members from the Departments of Civil, Electrical, and Industrial Engineering, Economics, Political Science, and Statistics, and from the Graduate School of Business.

The institute provides a center for research and project work on problems that require *integrated* contributions from several disciplines, especially from engineering, economics, and management. Attention is focused on areas in which planning and system considerations dominate. Particular stress is placed on study of physical or operational systems with complicated interaction between parts; on the situations in which decision making must take place under uncertainty; and on those situations in which characteristics or states evolve with time and in which control is a significant factor. In general, model making and computer simulation are emphasized, and various optimization and procedures receive strong attention. New professional courses being developed stress connections to the foundation disciplines and also to field work.

Since system problems are problems of the real world, environment for interaction with the real world for both training and research purposes is provided in two complementary ways. The first way is to immerse the student and the professor in difficult casework problems by means of internship programs with industry, research organizations, and governmental agencies at home and abroad. The second way is to bring practitioners from the field to the university, to free them from daily administrative chores and allow them to regenerate their approaches through study and interaction with students and faculty, thus contributing to the mutual growth of all three groups. Visiting practitioners are Senior Industrial and Government Fellows in Engineering–Economic Systems.

Graduate programs are offered with varying degrees of specialization in engineering–economic systems, ranging from supplemental training given as a part of a regular engineering program in some particular department to a doctorate in engineering–economic systems. The graduate program involves an intensive program of study at the university and extensive project-directed casework in the field, followed by research to unify and generalize the work done.

Two graduate degrees are being planned in engineering–economic systems. The first is the post-master's degree of System Engineer, which will require about a year and a half (academic year) of coursework plus a concept-directed casework thesis following one field project. The second is a Ph.D. degree in Engineering–Economic Systems. It involves heavy emphasis on mathematics, physics, and economics, plus two or more extensive casework projects in the field followed by a doctoral dissertation. For both degrees, about one third to one half of the courses are graduate courses given in mathematics, physics, and economics departments in the School of Humanities and Science. The remainder are professional courses.

Casework. The casework programs involve two areas of activity: macrosystems and microsystems. Macrosystems are characterized by broad scope and long-term activities. Usually, structure in relation to long-term function is more important than detail in their analysis and design. Uncertainty is a dominant consideration. Most planning problems are macrosystem problems. Casework programs in system planning include public works projects and regional development planning; the latter involves both public and private enterprises. In macrosystem casework, improvements in specific criteria and methodology for decision making are being made. Here, feasibility studies, market surveys, technological evaluations, models of growth, and simulation of operations are important.

Microsystems is more limited in scope and time than macrosystems. Detailed system structure and constraints are emphasized. Control of operations or physical systems is usually more important to microsystems work than is planning. The casework program in system control emphasizes the technico-economic problems of automation. In these studies the physical problems of systems analysis and design and the economic problems of evaluation specification and planning are both considered. In both macro- and microsystem projects, the field research is aimed at establishing specific results in specific areas as well as developing general theoretical results.

The IEES macrosystem projects have been concerned with the development of water resources, water quality control, urban and long-range transportation, airline planning, highway planning, hydroelectric power development, study of power system interties, regional economic development, and foreign aid for economic development. The projects in microsystems have been in automation in the electric power industry, including modeling and control of steam boiler and turbine systems; automatic load dispatching; simulation of multivariable dynamic systems; automation in the steel industry, including modeling and control of rolling mills; analysis of the steelmaking process and its computer control problems; dynamic scheduling of steel-making and steel-working operations; coordination of production and marketing in the steel industry; problems of inventory control and marketing; automated aids to instruction, including learning models, man–machine systems, optimal presentation of material and testing, and design studies of terminal equipment. Not all projects are active at any one time, but are under way when there is some particular interest in them. Though the inclusion of any one project is usually not vital, it is important that all general types of system problems always be represented. No exclusiveness is intended in the listing of practical areas. The aim is to undertake whatever projects are technically interesting, have practical significance at the moment, and involve theoretical considerations of general interest.

Internships. The casework projects described require substantial connections with companies and agencies outside the university. The institute has interns in private industries and consulting companies, in governmental agencies operating in the field, and with the Executive branch of the government in Washington, D.C. There is no attempt to make all internship relationships identical. The graduate students work in the field and are paid salaries for doing jobs that are considered technically challenging and practically significant. In all cases the student intern assumes project responsibility in the field. His project continues until he completes it, usually in from 6 to 15 months. These project-directed casework assignments are alternated with periods spent at the university taking graduate courses and doing concept-directed casework to follow up and generalize work done in the field.

Research in engineering–economic systems. A systems research program has been initiated covering both the macro- and microsystems areas in order to develop a theoretical base that combines foundation disciplines and casework. Mathematical and statistical principles have been applied to optimum allocation of effort, optimal control, system theory, analysis and simulation of multivariable systems, scheduling, and traffic control problems. Continued research in these areas will be encouraged and supported independently of immediate applications. The meaningfulness of theoretical work is enhanced, of course, by correlating it to the practical area. Some of our more significant research projects are described in the following.

Modeling of dynamic multivariable systems. Systems are designed and decisions are made more on the basis of concept than quantity. Accordingly, the structure of a mathematical argument is usually more important to the system man than the detailed methodology involved. Determination of mathematical structure is the essence of modeling. Most mathematics books give the most general logical structure, which is too impractical and overconcerned with improbable special cases to be most useful to system engineers. Most applied mathematics and engineering mathematics books are preoccupied with techniques for solution of equations and do not emphasize concepts or logical structure. The common threads among most system problems arise from similarity in structure and objectives. Mathematical structures that are not completely general usually involve abstraction to remove irrelevant detail; they are rich in useful system concepts. We have developed a set of principles that involves a geometric view of matrix theory, eigenvalue analysis, linear functionals, discrete system analysis, state variables, least squares, and constrained maximum problems, and a text is being prepared on this material. Much remains to be done in the area of modeling. In particular, models of multivariable systems of high dimensionality still present many unsolved problems. Analysis of systems that combine deterministic and stochastic components still need more work.

Principles of optimization. The past several years have seen extensive development of optimization techniques to solve problems in a number of physical system areas (e.g., production planning and inventory control). The expanding literature consists primarily of applications of various optimization techniques useful in solving classes of problems.

Little work has been devoted to unification of these optimization techniques into a set of basic principles. Some of the more familiar techniques have been treated abstractly and are now well understood. This treatment has resulted in fundamental concepts, such as the calculus of variations, the Kuhn–Tucker theorem, and least squares. In the case of least squares, the principle is a simple extension of the well-known geometric result that in a three-dimensional space the shortest vector from a given point to a given plane is the vector from the point that is perpendicular to the plane. This simple, intuitively appealing idea may be extended to infinite-dimensional spaces and, in this form, the principle may be applied to a wide variety of optimization problems. To apply the method, a particular problem is formulated in terms of minimizing a distance in an appropriate vector space. For example, the Wiener filter problem—the optimal estimation of a stochastic process—can be cast into this form by considering the vector space of random variables.

Decision making under uncertainty. The most important problems are those involving decisions—decisions that commit an organization to allocate its resources in a particular way. The mathematical theory of decision has arisen to provide a formal framework for the process of making decisions. To apply the theory we must (1) determine the possible states of nature, (2) specify the alternatives available to the decision maker, (3) assess the costs of following each alternative if each state of nature should occur, (4) assign probabilities to each state of nature, and (5) establish a criterion for selecting the best alternatives. Although the basic theory of this process is well developed, much work remains to be done before the formal representations of decision procedures provide the practical benefits they promise.

Summary. With regard to the Stanford program, we must observe that the pressures of establishing a new professional discipline while continuing the growth and development of existing fields poses important allocation problems for the administration. Dean Pettit's foresight in seeing the problem early and strongly supporting its solution allows the Institute in Engineering–Economic Systems enough flexibility, independence, and strength to develop in functionally meaningful but unconventional ways to meet the unconventional problems and opportunities of automation. A substantial and broadly based grant by the Ford Foundation provides the resources to initiate the program and test our plans for establishing the new profession.

Conclusion

Technology is at a crossroad. The revolutionary advances in ability to communicate and coordinate brought by computers and automation will change and can enrich the lives of us all. If the changes are to be orderly, productive, and dynamic, we must clarify objectives, functions, and priorities at a scope not heretofore attempted or necessary. Independent entities in industry must coordinate their activities. Broad plans for public facilities must be generated and government support for them obtained. The best analyses of the problems of the new age must be evolved by academicians and practitioners working together. Professional-school professors do not intend to stand idly by conjecturing whether the new offspring is more like an engineer or a manager. It is we who can best tie the academic and practical worlds together. Let's get on with the job!

Interplanetary spacecraft telecommunication systems

Telecommunication systems may soon reach a stage of sophistication where communication with spacecraft flying in the vicinity of the outer planets of the solar system will be possible

Glenn A. Reiff *National Aeronautics and Space Administration*

During the past six years, a class of telecommunication systems designed to provide tracking, telemetry, and command functions at interplanetary, multimillion-kilometer distances has been developed and proved out on flights along heliocentric orbits and into the vicinity of the planets Venus and Mars. These developments are contributing substantially to the exploration of the solar system and, in addition, are one measure of the state of the electronic arts. If the capability continues to increase over the next several years at the same rate, we shall soon be able to transfer large quantities of information to the earth from the surface of Mars.

Prior to 1960, the United States had launched only one spacecraft with sufficient energy to escape essentially the gravitational influence of the earth; however, plans had been made and work started to exploit the capability of scientifically exploring deep space with spacecraft and of maintaining radio communications over distances extending to many millions of kilometers.[1] Much of this work has now been completed. Earth-based sending and receiving stations exist at locations around the world that can provide 24-hour telecommunication coverage, and four spacecraft—Pioneer V, Mariner II, Mariner IV, and Pioneer VI—have been successfully launched on interplanetary missions extending to multimillion-kilometer distances from the earth. The radial distances from the sun of the orbits of these spacecraft are shown in Fig. 1. Such flights have made it possible to obtain direct measurements by which to evaluate the various theories of the nature of the interplanetary environment and of the means by which solar disturbances are propagated throughout these inner solar systems. These missions have been conducted during a phase of relatively low activity in the 11-year solar cycle (Fig. 1).

The name "Pioneer" is currently assigned to a class of spacecraft characterized by their small relative size and by the fact that their orientation in space is maintained by spinning the entire spacecraft. This class has been designed to carry instrumentation for measuring the particulate radiation, magnetic, and micrometeoroid environment of space between the planets Venus and Mars. Pioneer V, the first in the series, was launched on March 11, 1960, into a heliocentric orbit that carried it toward the sun. During its useful lifetime, it approached the sun to within approximately nine tenths of an astronomical unit, and confirmed the existence of interplanetary magnetic fields.

In addition to measuring the interplanetary space environment, the Mariner class of spacecraft was developed to make the initial reconnaissance of our neighboring planets. A comparison of the Mariner craft has already been published.[2] These spacecraft, which automatically orient their three axes by sensing the directions of two celestial references, are somewhat larger than the Pioneers. Mariner II, launched August 27, 1962, was the first vehicle from the earth to send back data from the

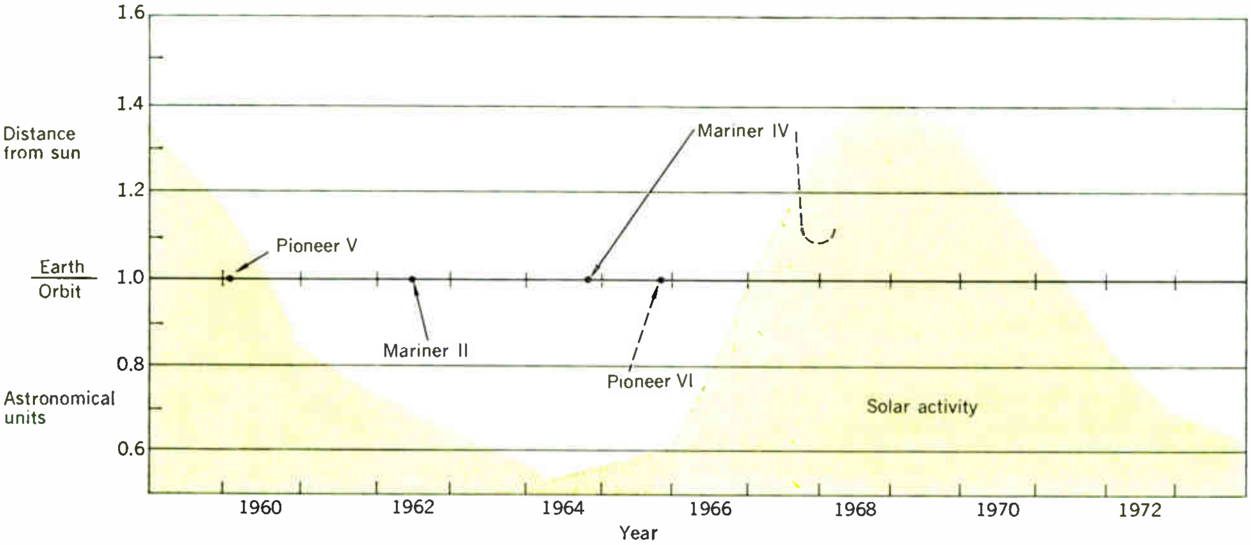


Fig. 1. Interplanetary missions of United States spacecraft.

vicinity of Venus, and it proved the existence of a continuous plasma streaming from the sun. Having been launched on November 28, 1964, Mariner IV completed its flight to Mars on July 15, 1965, and subsequently returned close-up photographs and other measurements of the red planet. During its journey away from the sun, Mariner IV detected numerous solar flares, provided the first unique identification of low-energy electrons in several of these flares, and indicated an increased concentration of cosmic dust between the orbits of the earth and Mars. On October 1, 1965, after ten months in space, an unprecedented threshold telecommunication distance of about 305 million km had been reached; reception of telemetry was terminated when a command from the earth shifted the spacecraft transmitter from the directional to the omnidirectional antenna. As indicated by the dashed line in Fig. 1, it may be possible to obtain telemetry from Mariner IV again in 1967 when the spacecraft approaches to within 50 million km of the earth.

To continue the studies of the interplanetary environment, Pioneer VI was launched on December 16, 1965, along a heliocentric orbit that is carrying it toward the sun. To the date of this writing, it has functioned well and has returned high-quality data. Additional Pioneer and Mariner launches are planned during the next several years.

With this introduction to interplanetary missions, let us proceed to an account of the more fundamental mission and spacecraft characteristics as they pertain to telecommunications, a review of the basic radio communication principles employed, a description of the elements of typical interplanetary communication system, and a recounting of telecommunication system performance to date as well as of possibilities for extending future telecommunication performance capabilities. Since Mariner IV and Pioneer VI are the most recently designed spacecraft in this series, emphasis will be placed on the telecommunication systems that they employ. Early papers have provided descriptions of the particular systems utilized on previous flights and have been used as a basis for the tables in the present article.[3-6]

Mission and spacecraft characteristics

The basic operational requirements that characterize unmanned interplanetary missions are that the spacecraft and its associated telecommunications system must be capable of reliably transferring information over multimillion-kilometer distances for months while the spaceborne equipment is unattended and also while large relative velocities are present between the transmitters and receivers.

The most important spatial references during these missions are the ecliptic—the plane of the earth's orbit about the sun—and the time-varying positions of the earth and sun relative to the spacecraft. The changing locations of these references and the method of spacecraft stabilization strongly influence the spacecraft and telecommunication system design and operational procedures.

The propulsive energy required essentially to remove a spacecraft from the gravitational influence of the earth and to place it in its own heliocentric orbit can be achieved only by obtaining high performance from the launch vehicle and by carefully balancing the total spacecraft weight with the booster capability. During the course of a flight, the spacecraft moves appreciably closer to or further from the sun than the orbit of the earth, and, as a result, there are wide variations in the radiant solar energy impinging upon the spacecraft. Solar cells have typically been used as the primary source of electric power, and their array must be sized to allow for this variation in the total energy available from the sun. Similarly, thermal control louvers and other means for keeping the spacecraft temperatures within specified bounds as the electric loads vary must be sized with regard to the changing distance from the sun.

The pointing directions of the telecommunications antennas and the view angles of the spacecraft orientation sensors and scientific instruments must be arranged in such a way as to provide the necessary responses, as the angular relationships between the earth, sun, and other celestial references at the spacecraft change throughout the entire mission.

I. Interplanetary spacecraft weight allocations

	Pioneer V	Mariner II	Mariner IV	Pioneer VI
Percentage of total weight allocated to:				
Scientific experiments	10	11	11	25
Communications, timing, and data handling	30	16	18	25
Electric power	35	24	26	20
Structure, pyro, cabling, and thermal control	25	30	26	24
Attitude control and propulsion	0	19	19	6
Total spacecraft weight, pounds	95	445	575	140

II. Summary of interplanetary missions

	Pioneer V	Mariner II	Mariner IV	Pioneer VI
Spacecraft weight, kilograms (pounds)	43 (95)	200 (445)	260 (575)	63 (140)
Date launched	3/11/60	8/27/62	11/28/64	12/16/65
Lifetime achieved, months	3.6	4.3	Still active	Still active
Maximum telemetry distance achieved, million kilometers (million miles)	36 (22.5)	87 (53.9)	305 (190)	
Total bits of information received	3×10^6	9×10^7	2.5×10^8	
Data storage, bits	None	None	5×10^6	1.5×10^4
Number of experiments	4	7	8	6
Primary power supply	Solar cells	Solar cells	Solar cells	Solar cells
Minimum power available, watts	25	200	300	60
Type of stabilization	Spin	Celestial reference	Celestial reference	Reorientable spin
Communication antennas	Quarter-wave stub	Turnstile, dipole, and paraboloid	Crossed slots and elliptic paraboloid	Slot omni and Franklin array
Signal polarization	Linear	RH circular	RH Circular	Linear
Frequency, Mc/s	378 Tx, 402 Rx	960 Tx, 890 Rx	2298 Tx, 2116 Rx	2292 Tx, 2111 Rx
Transmitter power, watts	5/150	3	10	8
Telemetry modulation	PCM/PSK/PM	PCM/PSK/PM	PCM/PSK/PM	PCM/PSK/PM
Telemetry data rate, b/sec	64, 8, 1	33.3, 8.33	33.3, 8.33	512, 256, 64, 16, 8
Launch vehicle	Thor–Able IV	Atlas–Agena B	Atlas–Agena D	Thrust-augmented improved Delta

Within the total spacecraft weight allowed by launch vehicle capability, the relative weights for telecommunications, guidance and control, structural, primary electric power, scientific experimental, and data-handling equipment must be carefully balanced in response to the objectives of the mission. Table I shows the weight allocation within the spacecraft under consideration. The final configuration of a spacecraft evolves from the nature of the mission and from an iterative process involving many trade-offs in which compromises have to be made among the requirements for the fulfillment of various detailed design objectives.

The balancing of these requirements in the design of interplanetary spacecraft has resulted in the allocation of some 15 to 25 percent of the total spacecraft weight to elements of the telecommunication, timing, and data-handling subsystems.

In each subsequent spacecraft design from Pioneer V to Pioneer VI and from Mariner II to Mariner IV, the capability and versatility of each class have been substantially increased. But the additional capability has been necessarily followed by an increase in size, weight, and complexity. Concurrently, interplanetary spacecraft reliability has been enhanced through the use of redundancy, comprehensive programs for screening parts, and more thorough environmental testing of the entire spacecraft. The approximate number of electronic parts per pound of spacecraft weight provides a rough indication of the additional capability as well as the increased complexity:

Spacecraft	Electronic Parts per Pound
Mariner II	60
Mariner IV	120
Pioneer VI	390

Of course, this also reflects the advancement in circuit fabrication techniques and the trends toward smaller discrete or solid-state electronic circuits. The electronic parts count used to obtain these ratios include some 10 700 solar cells on Mariner II, 28 224 aboard Mariner IV, and 10 368 on Pioneer VI. In addition, the Pioneer VI parts count includes 15 620 magnetic memory cores.

Although the telecommunication techniques employed with each of these spacecraft are similar in many ways, the requirements of each mission, the detailed spacecraft design, and the availability of facilities have caused important variations in the characteristics of both the earth-based and spaceborne equipment. Several of the more important features of each mission as they pertain to the communication systems are summarized in Table II.

III. Telecommunication system characteristics

	Spacecraft-to-Earth Link				Earth-to-Spacecraft Link			
	Pioneer V	Mariner II	Mariner IV	Pioneer VI	Pioneer V	Mariner II	Mariner IV	Pioneer VI
Frequency, Mc/s	378	960	2298	2292	402	890	2116	2111
Spacecraft antenna effective area, square feet	0.55	8.3	2.95	0.186	0.48	0.39	2.75	0.215
Ground antenna effective area, square feet	2.7×10^4	3.3×10^3	2.9×10^3	2.9×10^3	2.4×10^1	3.1×10^3	2.2×10^3	2.2×10^3
Transmitting power, watts	5/150	3.0	10	8	10 000	10 000	10–100 $\times 10^3$	10–100 $\times 10^3$
System noise temperature, °K	250	220/41	270/55	55	9200	9130	3190	3190
Receiving and transmitting circuit losses, dB	6	1	2	5	6	7	2	3

Contemporary techniques

An interplanetary spacecraft telecommunication system is comprised of the spacecraft-borne equipment and the ground-based terminals located at longitudes differing by about 120° around the world. The functions of this system are to track the position and velocity of the spacecraft, telemeter engineering and scientific data from the spacecraft, and transmit commands to the spacecraft. To perform these services reliably over great distances for many months, all current means of achieving extreme system efficiency must be employed. The unmodulated radio-frequency (RF) signal contains the tracking information and when modulated provides a carrier for command and telemetry data. Consequently, the general techniques of processing unmodulated radio frequencies will be summarized, followed by a brief description of the modulations utilized to provide data transmissions in the spacecraft-to-earth—or telemetry—link and the earth-to-spacecraft—or command—link. Some basic characteristics of these links are listed in Table III.

Radio-frequency channel

Range rates of this class of spacecraft relative to the earth at times have reached some 17 km/s. These rates cause Doppler shifts in the RF signal by as much as 300 kc/s at S-band during the course of a mission between our neighboring planets, or by as much as 12 kc/s during a 10-hour ground-based station view period. Figure 2 shows the two-way Doppler shift of an S-band signal at the range rates typically encountered during these flights. Phase-lock-loop receivers that track the frequency and phase of the received signal provide a means of detecting and following this varying radio frequency while maintaining narrow bandwidths for sensitivity. Tracking bandwidths of 12 c/s have been achieved in the ground-based receivers, and of about 20 c/s in the spacecraft receivers.

Fig. 2. Two-way Doppler shift of an S-band signal at range rates typically encountered during flights.

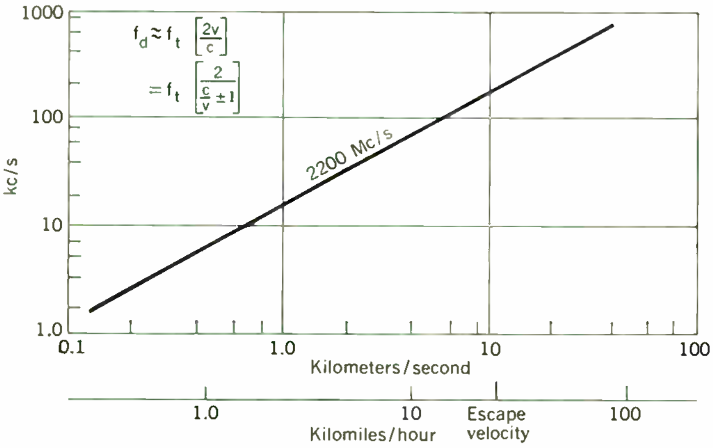


Figure 3 is a simplified schematic of the spacecraft portion of the system. The paths of the telemetry and command signals are shown by the heavy blocks. In the simplest and longest range mode of operation, no signal from the earth is present and the "out-of-lock" switch connects the noncoherent oscillator through the phase modulator and driver to the power amplifier. From the power amplifier a signal is radiated through one of the antennas to the earth at a frequency of about 2295 Mc/s. As will be noted later, this mode of operation was used to detect the presence of the Mariner IV signal during the time when the spacecraft was at its maximum distance from the earth.

The radio frequency of the signal transmitted is shifted by an amount dependent upon the earth–spacecraft range rate and is received by a ground-based station where it is tracked by means of the phase-lock loops that are located in the reference tracking receiver. The 85-foot-diameter paraboloidal reflector antennas installed at these stations are equatorially mounted (moving parallel in hour angle and perpendicular in declination to the earth's equatorial plane) and have a pointing accuracy of better than 0.02°. These antennas contain a tracking feed, a maser and a parametric preamplifier, and an electrohydraulic servo system. From a combination of the antenna feed signals processed by the phase coherent receivers, pointing error signals are obtained and used by the servo system to position the antenna automatically. The low noise temperature—about 15°K—obtained with the maser preamplifier is a key to obtaining an effective system noise temperature of about 55°K and maximum range performance.

Another mode of operation provides more accurate spacecraft range rate data. In this mode, the ground-based transmitter radiates to a spacecraft at a frequency of approximately 2115 Mc/s. Upon its receipt by the spacecraft, the signal is heterodyned, as indicated

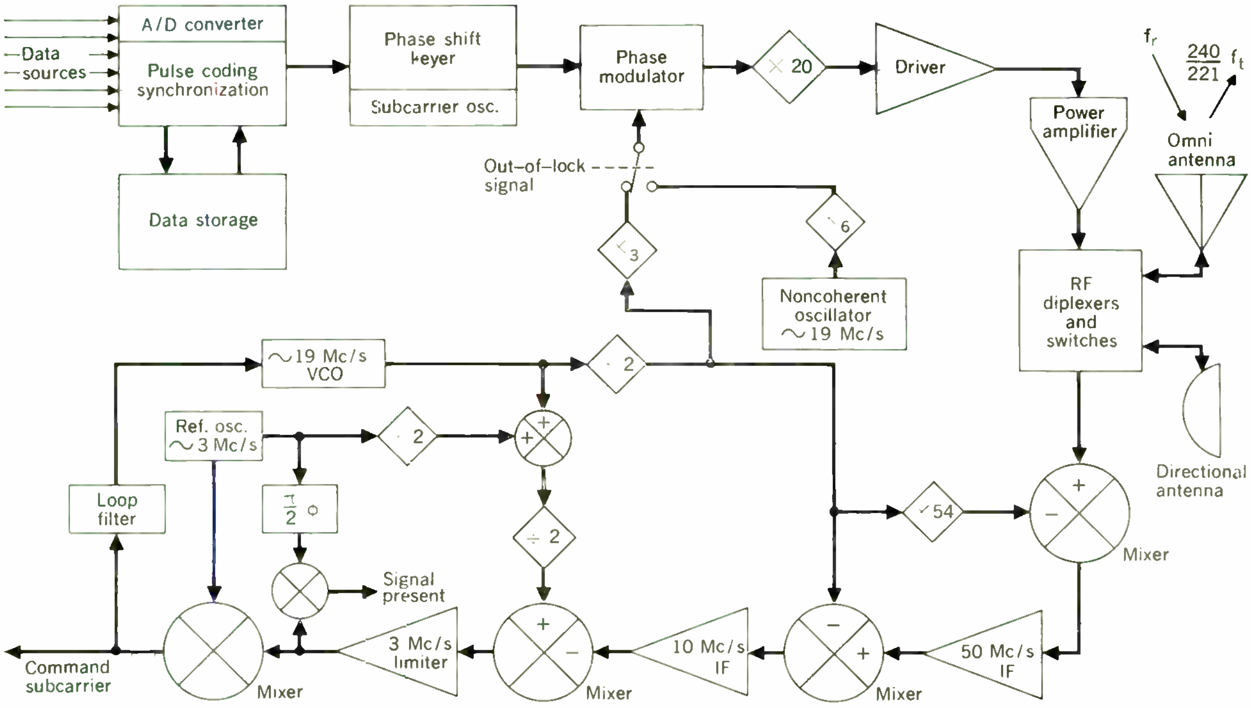


Fig. 3. Typical interplanetary spacecraft telecommunication subsystem.

in Fig. 3, by the phase-lock circuits to a series of intermediate frequencies. When the signal has been detected and lock obtained, a signal present indication is received. With this signal present, the "out-of-lock" switch replaces the noncoherent oscillator with a frequency multiplier chain driven by the incoming signal thereby closing the earth–spacecraft RF loop, which maintains phase and frequency coherence within the entire telecommunication system. Consequently, the radio frequency transmitted from the spacecraft is in an exact ratio to the frequency received.

The frequency of the ground-based klystron amplifier is precisely controlled by a rubidium frequency standard. This atomic standard has a short-term accuracy of about one part in 10^{11}. As a result, the range rate of the spacecraft can be obtained to an accuracy of a few millimeters per second by obtaining a Doppler count from the difference between the transmitted and received frequencies.

With the same ground station sending and receiving, the two-way Doppler mode is established. When one ground station transmits and another receives, the mode is known as three-way Doppler.

The only information contained in the functions thus far described are the facts that there is a spacecraft signal present in a known direction with respect to the ground station and that this signal is moving at a precisely measured rate relative to that station. To provide the necessary telemetry or command information, the RF carriers are phase- or frequency-modulated by subcarriers. Modulation indexes ranging from about 0.5 to 0.9 radian peak have been typically used, resulting in an RF carrier suppression of 50 to 60 percent. Thus, with 10 watts transmitted from the spacecraft, some 4 to 5 spaceborne RF watts are used to provide the range rate and position tracking functions.

Data links

Command and telemetry are provided in both the Mariner and Pioneer telecommunication systems by synchronous binary data-handling circuitry.

Inputs from the data sources indicated in the upper-left-hand corner of Fig. 3 may be producing data in either digital or analog form. Analog-to-digital (A/D) converters change the analog data into a digital representation, and all sources are serially sampled. The digital representations of the source data are pulse coded to obtain the baseband signals. These are the basic data waveforms to be transmitted. The Mariner IV command and telemetry links, as well as the Pioneer VI command link, utilize a nonreturn-to-zero-level (NRZ-LEVEL) baseband waveform in which the binary "1" is represented by one voltage level and "0" is represented by another level. The Pioneer telemetry link utilizes a nonreturn-to-zero mark (NRZ-MARK) baseband waveform in which the binary "1" is represented by a change in voltage level and "0" is represented by no change. With these waveforms the symbol interval varies inversely with the information transmission rate. Data are transmitted over the command links at 1 b/s (bit per second), so the symbol interval is one second. For telemetry at 8 b/s the symbol interval is 125 ms, and at 512 b/s the interval becomes 1.92 ms.

In all the data links, with the exception of the Pioneer command channel, the baseband signal is used to biphase-shift key to a subcarrier oscillator. In the Pioneer telemetry the frequency of the subcarrier square wave is fixed at 2048 c/s whereas in Mariner IV the frequency of the square-wave subcarrier can be either 150 or 600 c/s, depending upon the bit rate used. The subcarrier provides a source of modulation for phase modulating the carrier.

The method of modulating the Pioneer VI command link is substantially different from that used in the other links. This channel utilizes the baseband signal to frequency-shift key the transmitter with one of two frequencies. One frequency is transmitted for one second to represent a "1" or mark and the other is transmitted for the same duration to represent a "0" or space.

Synchronization of the digital signals is accomplished differently in the Pioneer and Mariner systems. The Pioneer VI telemetry and command channels are synchronized by inserting unique binary combinations in the baseband signals. A telemetry word consists of seven bits, with the seventh bit used for a parity check. A frame consists of a sequence of 32 words, and within a frame two words are used for synchronization.

The telemetry and command channels in the Mariner system are synchronized by means of a pseudorandom code used to modulate a second subcarrier, which is linearly added to the telemetry subcarrier to form a composite modulating signal. Consequently, the transmitted Mariner RF signal is phase modulated by the combined subcarriers. For Mariner IV telemetry a square-wave synchronizing subcarrier drives a redundant pair of code generators that provide the 63-bit pseudorandom code. A set of word gates, in turn, generates bit and word synchronization pulses that are used to synchronize (1) the stepping of the commutator, (2) the analog-to-digital converters, (3) the readout of the scientific measurements from the data automation system, (4) the readout of the event registers and timers, and (5) the playback of the stored data. The word-synchronizing pulses occur once per cycle of the code, while the data-bit-synchronizing pulses occur once every nine code bits, or seven times per code cycle. Thus, each Mariner IV engineering data word is also seven data bits long.

In the Mariner IV and Pioneer VI spacecraft-to-earth links, 11 to 15 percent of the total power radiated is used to provide synchronization.

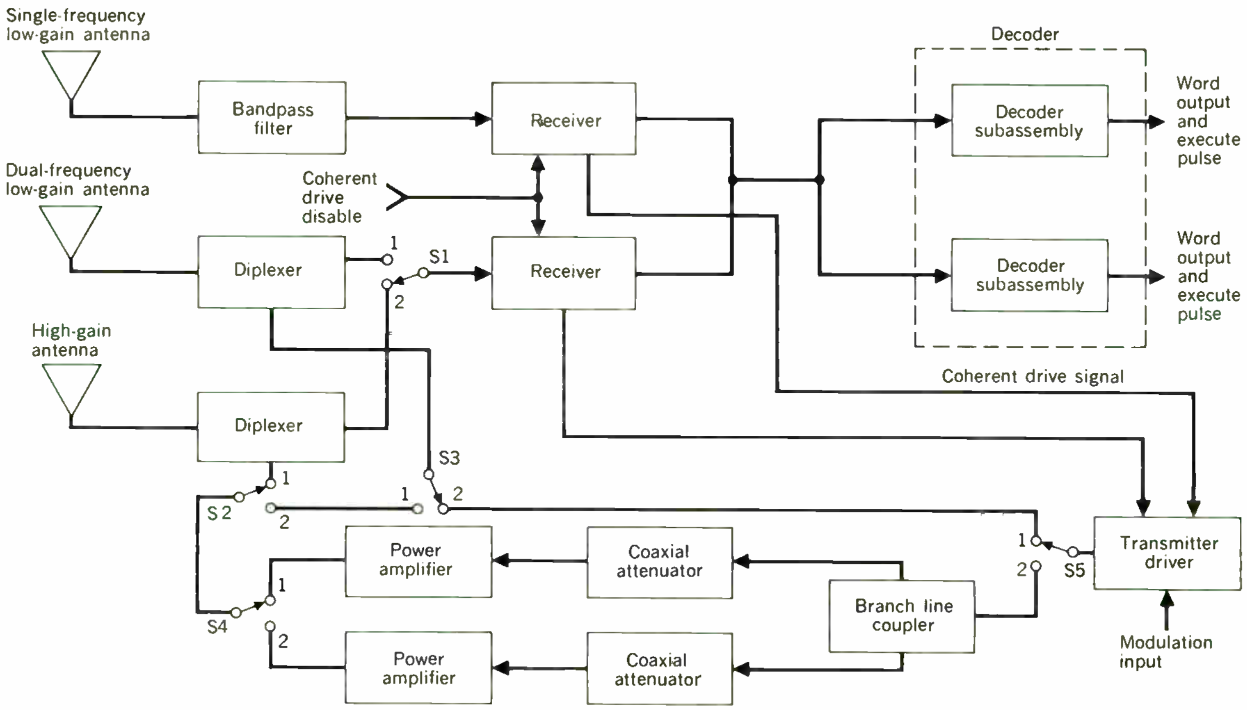
Information-sampling plans in both the Mariner IV and Pioneer VI telemetry encoders can be changed by switching modes. They also include a means for storing data for later transmission to the earth. Mariner IV was provided with a magnetic tape recorder for data storage and magnetic cores are used in the Pioneers. The four modes of data transmission from Mariner IV are: (1) spacecraft performance measurements only, (2) measurements of spacecraft performance (one third of total measurements) and scientific phenomena, (3) measurements of scientific phenomena only, and (4) primarily stored television pictures. More than twice this number of modes are provided in the Pioneer data-sampling plan.

Digital data systems are sometimes compared on the basis of modulation efficiency, which can be defined as the bandwidth in which there must exist unity signal-to-noise ratio (SNR) in the time required to transmit one bit of information at an acceptable error rate. Using the data coding, synchronization, and modulation techniques described, modulation efficiencies of about 6 cycles per bit at an error rate of 10^{-3} have been achieved with these PCM/PSK/PM systems. This value approaches the theoretical limit.

Spacecraft-borne equipment

To survive the transition through ionization pressures in the earth's atmosphere during launch, provide continuous communications during maneuvers, and assure reliable communications for months in deep space, the spacecraft-borne telecommunications equipment has been designed to be operated in alternate configurations.

Fig. 4. Communication subsystem block diagram.



The amount of redundancy and number of switching arrangements provided are based upon the requirements of the mission, individual component characteristics, and the designers' judgment.

Figure 4 shows the Pioneer VI communication subsystem switching arrangement. It may be noted that the subsystem is completely redundant, except for the coaxial switches, S1 through S5, and the transmitter driver. During the launch phase the transmitter driver is connected directly to the dual-frequency low-gain antenna and no power is supplied to the traveling-wave tube (TWT) amplifiers. One receiver is connected directly to the single-frequency low-gain antenna to insure that the spacecraft can be commanded. Shortly after separation from the launch vehicle, one TWT is turned on and the proper coaxial switch is set automatically with boom deployment to provide high-power radiation through the dual-frequency low-gain antenna. During the process of orientation a receiver and transmitter are switched to the high-gain antenna.

Similar switching arrangements are available in the Mariner IV communication subsystem.[7] However, less redundancy is provided and most switch actuations can be made by either an internally timed signal or by a command from the earth. Circulators are used instead of coaxial switches for routing the RF signals. One RF power amplifier is a planar triode in a cavity while the other is a TWT. The design includes redundant transmitters and drivers, but only one receiver.

Both the Mariner IV and Pioneer VI communication subsystems operated faultlessly during launch and maneuvers. By the end of 1965, the Mariner IV subsystem had operated perfectly for more than 9000 hours in space, and Pioneer VI had functioned properly in space for about 500 hours.

Calculated performance and flight results

During the past six years, various hypotheses portending difficulties with interplanetary communications have been advanced. Mechanisms have been proposed for the generation of multipath, rotation of polarization, deflection, attenuation, and other propagation problems with microwave signals traversing the interplanetary medium.

Although the region between the planets is often regarded as a great void, it, in fact, contains an extremely tenuous ionized gas. From the results of several radar experiments and the flights of Mariners II and IV, it may be inferred that the attenuation due to the interplanetary medium between the orbits of Venus and Mars is less than 1 dB at frequencies of about 1 to 2 Gc/s. In addition, radar experiments on the planet Venus show that any mechanism that might reverse the sense of the polarization of circularly polarized signals at S-band is less than 0.1 percent effective.

The path length of S-band signals in the region between the orbits of Venus and the earth has been shown by Mariner tracking and radar experiments to be essentially the same as the geometrical path length.

In general, measurements to date indicate that the interplanetary medium in the region between the orbits of Venus and Mars has no deleterious effects on com-

IV. Nominal spacecraft-to-earth channel performance

				Pioneer VI	
	Pioneer V	Mariner II	Mariner IV	85′ Antenna	210′ Antenna
Transmitter power, P_t, dBm	+ 37	+ 35	+ 40	+ 38	+ 38
Wavelength conversion, $\lambda/4\pi$, dB (λ in miles)	− 44	− 48	− 52	− 52	− 52
Antenna gain, dB	0	+ 20	+ 23	+ 11	+ 11
Transmitter circuit losses, L_t, dB	− 2	− 1	− 1	− 3	− 3
Spacecraft RF transmission factor, T	− 9	+ 6	+ 10	− 6	− 6
Antenna gain, dB	+ 47	+ 46	+ 53	+ 53	+ 61
Wavelength conversion, $\lambda/4\pi$, dB (λ in miles)	− 44	− 48	− 52	− 52	− 52
Noise spectral density, N/B or KT, dB per c/s	−176	−181	−181	−181	−181
Receiver circuit losses, L_r, dB	− 4	− 0	− 1	− 2	− 2
Ground RF reception factor, R	+175	+179	+181	+180	+188
Unmodulated RF distance (miles), dB	+166	+186	+192	+174	+182
Carrier:					
Modulation loss, dB	− 6	− 3	− 4	− 4	− 4
Threshold SNR or required ST/N/B, dB	0	0	0	0	0
Bandwidth, dB	+ 13	+ 13	+ 11	+ 11	+ 11
Modulation characteristic, M	− 19	− 16	− 15	− 15	− 15
Zero margin distance, dB	+147	+170	+176	+159	+167
Telemetry:					
Modulation loss, dB	− 2	− 4	− 5	− 2	− 2
Required ST/N/B, dB	+ 9	+ 8	+ 8	+ 9	+ 9
Data rate, dB	+ 9	+ 9	+ 9	+ 9	+ 9
Modulation characteristic, M	− 20	− 21	− 22	− 20	− 20
Distance, 8 b/s, dB (20 log D miles)	+146	+165	+169	+154	+162

V. Nominal earth-to-spacecraft channel performance

	Pioneer V	Mariner II	Mariner IV		Pioneer VI	
			10 kW	100 kW	10 kW	100 kW
Transmitter power, P_t, dBm	+ 70	+ 70	+ 70	+ 80	+ 70	+ 80
Wavelength conversion, $\lambda/4\pi$, dB (λ in miles)	− 43	− 48	− 52	− 52	− 52	− 52
Antenna gain, dB	+ 47	+ 45	+ 51	+ 53	+ 51	+ 53
Transmitter circuit losses, L_t, dB	− 4	− 1	− 1	0	− 1	0
Ground RF transmission factor, T	+ 70	+ 66	+ 68	+ 81	+ 68	+ 81
Antenna gain, dB	0	+ 6	+ 22	+ 22	+ 11	+ 11
Wavelength conversion, $\lambda/4\pi$, dB (λ in miles)	− 43	− 48	− 52	− 52	− 52	− 52
Noise spectral density, N/B or KT, dB per c/s	−158	−159	−164	−164	−164	−164
Receiver circuit losses, L_e, dB	− 2	− 6	− 1	− 1	− 2	− 2
Spacecraft RF reception factor, R	+113	+111	+133	+133	+121	+121
Unmodulated RF distance (miles), dB	+183	+177	+201	+214	+189	+202
Command:						
Modulation loss, dB	− 6	− 8	− 9	− 9	− 3	− 3
Threshold SNR or required ST/N/B, dB	+ 3	+ 18	+ 16	+ 16	+ 2	+ 2
Modulation characteristic, M	− 9	− 26	− 25	− 25	− 5	− 5
Distance, 1 b/s, dB (20 log D miles)	+174	+151	+176	+189	+184	+197

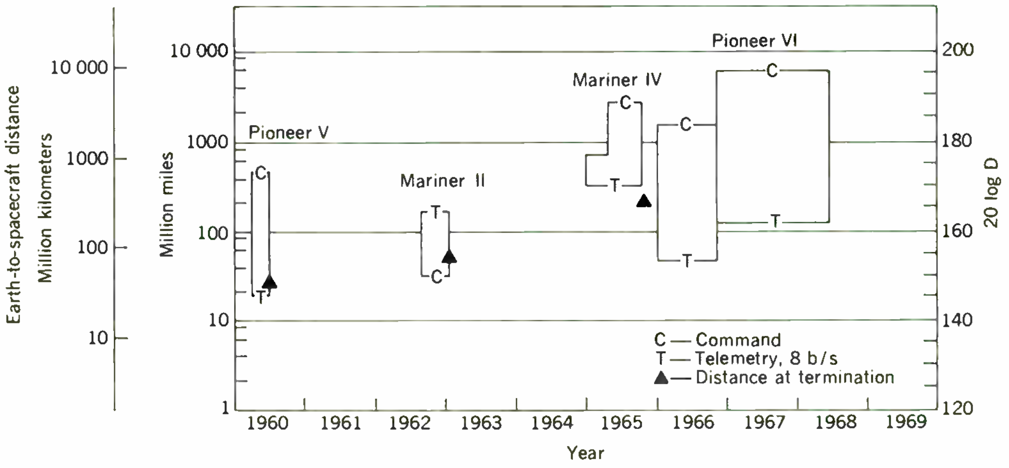
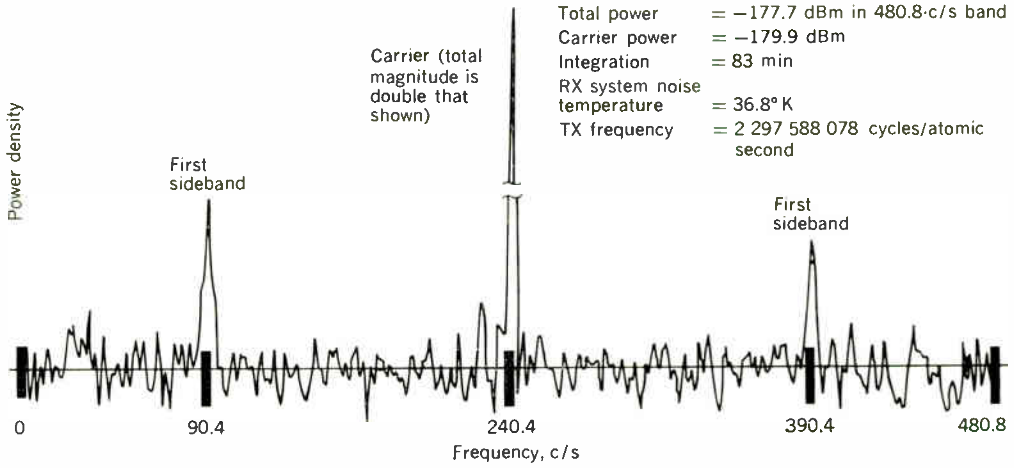


Fig. 5. Interplanetary telecommunication capability.



Total power = −177.7 dBm in 480.8-c/s band
Carrier power = −179.9 dBm
Integration = 83 min
RX system noise temperature = 36.8° K
TX frequency = 2 297 588 078 cycles/atomic second

Fig. 6. Mariner IV signal, Dec. 1, 1965.

munication with spacecraft using S-band signals and that the usual free-space formulas for calculating telecommunication system performance capability apply.[8]

Table IV contains the rounded-off values of the parameters needed to calculate the nominal spacecraft-to-earth carrier and telemetry performance. Table V shows similar values for the command link performance. With each of the parameters listed in these tables, there is an uncertainty or tolerance in the nominal value that is usually included in the design control calculations. Common practice is to provide a performance margin equal to the adverse sum of these tolerances.

Tables IV and V are arranged to facilitate the logarithmic calculation of the performance and to provide figures of merit for the transmitting and receiving equipment as well as the method of modulation. The RF transmission factor T is obtained by summing transmitter power, wavelength conversion, antenna gain, and transmitter loss parameters, with all values expressed in decibels. Similarly, RF reception factor R results from algebraically subtracting noise spectral density from the sum of antenna gain, wavelength conversion, and receiver circuit losses. The modulation characteristic M is obtained by algebraically subtracting the threshold SNR or the required modulation efficiency (ST/N/B) and the bandwidth or data rate from the modulation loss. The maximum nominal distance of communication, expressed as $20 \log D$, has been obtained by summing the RF transmission and reception factors with the modulation characteristic of the appropriate link.

Results of the calculations are summarized in Fig. 5. The horizontal lines under each flight indicate the calculated performance with no margin of the command link at 1 b/s and the telemetry link at 8 b/s. For the Pioneer V and Mariner II flights, the black triangles indicate the distance at which all signals from the spacecraft ceased, while in the case of Mariner IV the triangle indicates the distance of the spacecraft from the earth when the telemetry was purposely terminated.

It may be noted that the actual performance of the Mariner IV telemetry link is indicated to be some 5 dB below the calculated value at the time that telemetry was terminated. The reason for this apparent discrepancy is that the value of spacecraft antenna gain used in Table IV to make the calculation applies only when the main beam of the high-gain antenna points at the earth. By the time that telemetry was terminated, this beam was no longer pointing toward the earth and it was necessary to shift to the omnidirectional antenna.

Contacts with Mariner IV have been made each month since the antennas were shifted. In early January 1966, the spacecraft reached a maximum distance from the earth of about 350 million km (215 million miles), and the carrier signal with the first sideband was detected on January 3. The spectrum of the signal obtained in December is shown in Fig. 6. As the spacecraft approaches the earth in 1967, it should become possible to obtain telemetry with an acceptable error rate.

System extensions

The most immediate system extensions are the installation of 100-kW transmitters and the construction of 210-foot antennas at the ground-based terminals. This work has been partially completed, and its effect is included in the calculated capability shown in Fig. 5.

A 100-kW transmitter is in operation at the research and development site of the Deep Space Instrumentation Facility, and it was used to send commands to Mariner IV during the Mars encounter. However, this increased command capability is not yet available on a 24-hour basis.

Construction of the first 210-foot antenna is nearing completion; this structure should increase the capability by about 8 dB over the 85-foot antennas. It is expected that this antenna will be available to regain telemetry from Mariner IV and track the Pioneers around the sun during late 1966 and 1967. Plans have been made to construct these larger antennas at the other stations around the world to provide the increased capability continuously.

Further increases in the capability of communicating with interplanetary spacecraft can come when more powerful launch vehicles are assigned to these missions. With larger spacecraft, radiated power as well as antenna gain can be increased. Adaptive, or nonsynchronous, data compression techniques may be employed in future spacecraft. Through these measures it should be possible to increase the information transfer rates by a factor of more than 100 on missions into the vicinity of our neighboring planets or to extend the practicable distance of two-way communications into the far reaches of the solar system.

Conclusions

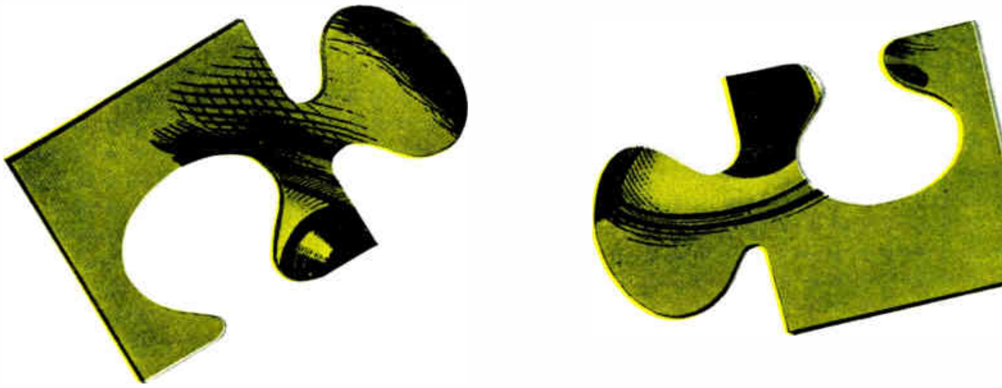
This article has reviewed the progress made in developing interplanetary spacecraft telecommunication systems and has outlined some of the basic constraints in designing systems for these missions. If the capability continues to increase during the next several years at the same rate that it has in the past six, we will soon be able to transfer large quantities of information from the surface of Mars or to communicate with spacecraft flying in the vicinities of the outer planets.

This article is based on a paper presented at the IEEE International Convention, New York, N.Y., March 21–25, 1966.

The author wishes to acknowledge the work of many hundreds of people throughout the world, particularly those at the Jet Propulsion Laboratory, TRW Systems, and Ames Research Center, in designing, developing, and operating the systems described. He is also indebted to J. A. Hunter and J. W. Thatcher for their help in reviewing this article.

REFERENCES

1. Williams, W. E., Jr., "Space telemetry systems," *Proc. IRE*, vol. 48, pp. 685–690, Apr. 1960.
2. Schneiderman, D., Reiff, G. A., and James, J. N., "Recent Mariner spacecraft design and flight," *Astronautica Acta*, vol. 11, no. 2, 1965.
3. Hansen, R. C., and Stephenson, R. G., "Communications at megamile ranges," *Microwave J.*, pp. 79–84, Jan. 1962.
4. Bryden, J. J., "Mariner (Venus '62) flight telecommunications systems," Tech. Rept. 32–377, Jet Propulsion Laboratory, Pasadena, Calif.
5. Thatcher, J. W., "Deep space communication," *Space/Aeron.*, pp. 54–63, July 1964.
6. Renzetti, N. A., "Tracking and data acquisition support for the Mariner Venus 1962 mission," Tech. Memo 33–212, JPL, July 1, 1965.
7. Mathison, R. P., "Mariner Mars 1964 telemetry and command system," *Proc. 1965 IEEE Winter Convention on Military Electronics*; also in condensed form in *IEEE Spectrum*, vol. 2, pp. 76–84, July 1965.
8. Easterling, M., and Goldstein, R., "The effect of the interplanetary medium on S-band telecommunications," Tech. Rept. 32–825, JPL, Sept. 1, 1965.



On the mechanization of

The goal of mechanizing certain creative processes in problem solving is attainable, but not in the near future. The problem is to find workable computer procedures for evolving "appropriate" representations in given problem-solving situations. The answer may lie in the extension of ideas that were developed for some elementary theory-formation procedures.

There is no general agreement on the nature of creativity or the characteristics of a creative act. Such an act is often surprising, has elements of a new approach, and is not stereotypic. Beyond these phenomenological properties, one finds that a creative act has a strong element of synthesis. It is usually associated with ill-defined goals or it involves reformulation of externally imposed goals. Some students of the human mind feel that formation of powerful imagery, abstraction to appropriate spaces, flexible associations, and rich generation of analogies are key element sof creative processes. Others feel that, in addition to these elements, some mysterious, unexplainable processes control the genesis of ideas and insights in man's creative process. If we exclude the belief in such unexplainable processes, then it is reasonable to attempt explications of (at least some) creative processes in terms of information-processing models. Explications of this type will essentially amount to advancing operational definitions of creative processes. By studying the implications of such proposed definitions, by testing them versus existing notions of creative behavior, by subsequently improving the proposed definitions, and so on, we can hope to attain a satisfactory understanding of the notion of a creative process. In this manner, there is a chance that we can arrive both at a psychological theory of creative processes and at the logical principles that underlie computer realizations of such processes.

I would like to present here some tentative ideas on an operational definition of creative processes in the general context of problem-solving processes. My comments are restricted to some of the creative processes that occur in the problem-solving activities of the physical scientist, the engineer, and the mathematician.

An important type of problem confronting the physical scientist is the formation of theories that organize empirical knowledge in certain desired ways. A common problem for the engineer is to evolve a design that satisfies desired goals. One of the problems of the mathematician is to prove theorems in a formal system. In the last decade, several procedures have been developed for solving by computer problems of these three types. Much of the present research in artificial intelligence is directed to extending the scope and power of such problem-solving procedures. It is my belief that some of the difficult problems that we are now facing in the design of more powerful problem-solving procedures are related to the problem of mechanizing certain creative processes.

Extending the power of problem-solving procedures

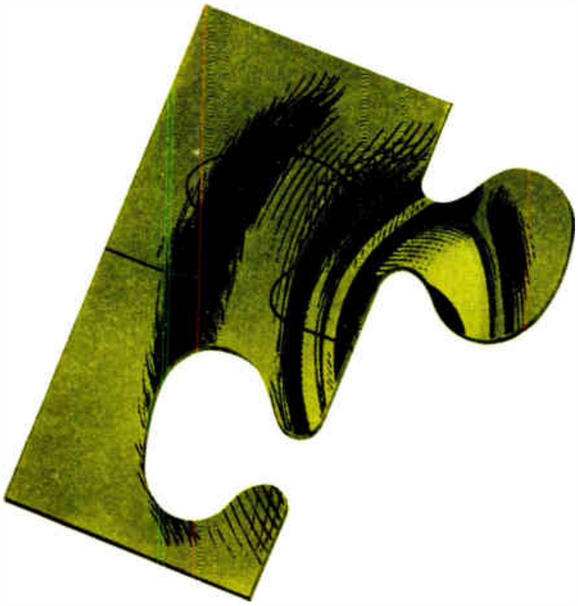
Two central notions are involved in a problem-solving procedure: first, a problem state—a description of a problem situation including goals, available resources, and intermediate results; and second, a set of relevant moves that can be applied from a state to obtain new states. The relevant moves commonly reflect the rules of the game, the rules of inference, the grammar, the available composition, etc., that can be used by a problem-solving procedure in the course of an attempt to construct a solution. In designing problem-solving procedures one must find appropriate descriptions for problem states and for transformation of states via moves. In other words, one must face the problem of defining a problem-state space; I call this the problem of representation. A closely related problem is the problem of evaluation. It involves the choice of concepts and methods for evaluating a variety of measures of progress—as well as estimates of expected search efforts—that can be associated with points in state space and also with transition between points in that space. A third major problem is that of controlling the search for a solution in state space. Here one needs overall strategies and specific decision functions for intelligently selecting problems—solving moves between problem states so that a solution can be found with as small a computational cost as possible. Most of the effort expended on machine problem solving so far has been directed to this third problem; specifically, to the study of a variety of schemes for heuristic search.[1]

In the present state of the artificial intelligence art, the designer of a problem-solving procedure is required to solve without aid from the machine the problems of choosing a state space, a basis for evaluation, and a strategy for heuristic search. The relatively intelligent behavior of the machine that solves problems in accordance with the problem-solving procedure formulated by the designer is therefore circumscribed by the choices of representation evaluation and control that are made by the designer.

Ideas on an operational definition of creative processes in the general context of problem solving are presented—restricted to such processes that occur in the problem-solving activities of the physical scientist, the engineer, and the mathematician

creative processes

Saul Amarel *RCA Laboratories*



Improvement in the power of problem-solving procedures can be achieved by making appropriate changes in the rules that control search, in the methods of evaluation, and in the modes of representation. Several attempts have already been made to adjust certain parameters automatically in local decision rules that control search, and also in evaluation functions—on the basis of statistical learning techniques.[2] However, no schemes exist as yet for general nonparametric control of the search and evaluation parts of problem-solving procedures.

In our work with proving theorems of the propositional calculus by the method of natural deduction we have developed a sequence of procedures of increasing power in order to evaluate the nature of improvements that occur at different stages of this evolution. The most spectacular improvement was obtained when the problem representation was changed in an "appropriate" way—the shift in representation has transformed the problem to one of finding appropriate closures to certain directed graphs. The new problem representation immediately suggests to human problem solvers a new, more powerful, basis for evaluation and search; the result is a much better goal-oriented, thus less inefficient, problem-solving process. We had similar experience with other relatively simple transportation scheduling problems. Indeed, the importance of "having the right point of view," "casting the problem in the appropriate form," "conceptualizing the problem correctly" has been recognized for some time by students of problem-solving processes.[3]

Finding the "most appropriate" space

I think that creative problem solving is closely related to the notion of directing the search for solution in the "most appropriate" space. More specifically, I would like to suggest that the formation of an appropriate concept of problem space, where a given problem is to be treated—in other words, the solution of the problem of representation—is a creative process. This process could also be regarded as a process of building an appropriate model. While the use of given models in problem solving has already been considered by workers in artificial intelligence,[4] the dynamic aspects of evolving an appropriate model so far have received little attention.

It is commonly asserted that by furnishing a man with convenient graphical displays of appropriate models, he will be stimulated to provide the creative contribution expected from him in his problem-solving partnership with the machine. Clearly, someone has to choose the appropriate models to be displayed in specific situations; I consider this selection of models as demanding the main measure of creativity that enters the problem-solving process. If the man's function in his partnership with the machine is not only to utilize models (that facilitate his search for a solution) but also to form and modify models during the problem-solving process, then he will indeed be exercising his creative powers. This point of view has certain implications regarding the flexibility and power of model-building languages that are needed in a genuine creative problem-solving system involving man–machine interaction. Also, it may provide a test framework for identifying the part that creative processes play in problem solving.

Creative processes and theory formation

If we start with the assumption that the function of a creative process in problem solving is the formation of an appropriate problem representation (the growth of appropriate symbolizations or of suitable models), then the mechanisms that come into play in a creative process will have much in common with theory-formation mechanisms. In the theory-formation problem of the physical scientist the objective is to construct efficiently an information structure, in terms of existing linguistic and conceptual resources, that would summarize "elegantly" and "explain" a set (usually large) of empirically obtained relationships in a given area. The information structure, because of its mode of construction, expresses the empirical knowledge in terms of existing theoretical constructs, thus incorporating the new empirical information in the main body of theoretical knowledge. It also provides an appropriate basis for solving problems and answering questions in the given area. In the model-formation problem of the problem solver the objective is to construct a theory in terms of appropriate linguistic and mathematical constructs that expresses in a convenient manner to the problem solver the properties of the problem-state space.

In our work on theory-formation processes we have studied specifically procedures for the automatic formation of computer programs in a given programming language that have to satisfy a given set of computational correspondences.[5] Candidate programs are represented in the system in terms of a language of program descriptions. The system generates program descriptions, it evaluates them over the given set of correspondences, it modifies the program descriptions, and so on, until a description with the desired computational properties is found. The crucial problem for us is to find strategies of formation that direct in an efficient way this search in program-description space. After working with certain heuristic formation procedures, where considerable "blind search" takes place, we have found that basic improvements in the power of the formation procedures can be obtained if appropriate representations of the problem-state space (in this case this is the space of program descriptions) are available to the system.[6,7] Given an appropriate mathematical model of formation space, it is possible to have a formation process that is much more efficient. Note that in the present case we are discussing the importance of using models for efficiently building models (a model of formation space is used to evolve a model for a certain computation). Hence, it is again natural to ask how it is possible to evolve in a machine an appropriate model that would guide the machine formation of another appropriate model.[4] Again, I consider the construction of such a model as involving a creative process.

Mechanization of creative processes

In general, I think that many of the ideas used in theory-formation procedures can be transferred to the mechanization of processes for evolving appropriate representations (or for recognizing that a certain formal system provides an appropriate model) in given problem-solving situations. If we agree that such representation selection processes are creative processes, then we can already envision an approach (through theory-formation ideas, which, admittedly, are still at a very early stage of development) to the mechanization of creative processes in problem solving. However, even if a general approach using theory-formation ideas is considered, the question still remains as to how to solve representation selection problems efficiently with computers. Efficient solution of representation selection problems at a certain level may necessitate the solution of representation selection problems at a higher level (as we found out from our experience with theory-formation problems); the logical complexity of the required programs and the requirements of storage and computation time may well be beyond realistically attainable systems.

In general, I think that by mechanizing the process of selecting appropriate representations for problem-solving situations we will be taking an enormous step toward advancing artificial intelligence. Furthermore, I think that the notion of creative processes in problem solving is closely related to such representation selection processes. I believe that the goal of mechanizing certain creative processes in problem solving can be achieved; but we are a long way from it now. I think that the following quotation from E. Post's diary[8] is relevant to my comments: "The creative germ seems not to be capable of being purely presented but can be stated as consisting in constructing ever higher types. These are as transfinite ordinals and the creative process consists in continually transcending them by seeing previously unseen laws which give a sequence of such numbers. Now it seems that this complete seeing is a complicated process mostly subconscious. But it is not given till it is made completely conscious. But then it ought to be constructable purely mechanically."

I find this statement remarkable because it associates the notions of visualization (symbolization or representation), of a hierarchy of such visualizations, and of conscious self-reflectiveness with the creative process; furthermore, it ends with a note of confidence about the possible mechanization of creative processes.

This article is an expanded version of the author's points of view as expressed in a panel discussion on the mechanization of creative processes that took place on May 28 at the 1965 IFIP Congress in New York, N.Y. The panel was chaired by E. A. Feigenbaum; other panelists were J. McCarthy, V. Neisser, A. Newell, G. Pask, and L. Uhr. The research discussed in this article is sponsored by the Air Force Office of Scientific Research, of the Office of Aerospace Research, under Contract No. AF49-(638)-1184.

REFERENCES

1. Newell, A., and Ernst, George, "The search for generality," in *Proc. of the IFIP Congress 65*, W. A. Kalenich, ed., vol. 1. Washington, D.C.: Spartan Books, 1965, pp. 17–24.
2. Slagle, "A multipurpose theorem—proving heuristic program that learns," in *Proc. of the IFIP Congress 65*, vol. 2. Washington, D.C.: Spartan Books, to be published.
3. Polya, G., *Mathematical Discovery*, vols. 1 and 2. New York: Wiley, 1964.
4. Minsky, M. L., "Matter, mind, and models," in *Proc. of the IFIP Congress 65*, W. A. Kalenich, ed., vol. 1. Washington, D.C.: Spartan Books, 1965, pp. 45–49.
5. Amarel, S., "On the automatic formation of a computer program which represents a theory," in *Self Organizing Systems–1962*, M. C. Youvits, G. T. Jacobi, G. D. Goldstein, ed. Washington, D.C.: Spartan Books, 1962, pp. 107–175.
6. Amarel, S., "On the formation of the concept of a transformation by computer," in *Summaries of the 1964 International Congress for Logic Methodology and Philosophy of Science*, The Hebrew University, Jerusalem, Israel.
7. Amarel, S., "Languages of representation in problem-solving procedures of the formation type," in *Proc. of the Symposium on Logic, Computability, and Automata Theory, 1965*. Washington, D.C.: Spartan Books, to be published in 1966.
8. Post, E., in *The Undecidable*. New York: Raven Press, 1965.

Special
Conference
Report

1965 URSI Symposium on Electromagnetic Wave Theory, Delft, Netherlands

H. V. Cottony *Institute for Telecommunication Sciences and Aeronomy of Environmental Science Services Administration*

From September 6 through 11, 1965, a symposium on electromagnetic theory was held at Delft, Netherlands. It was the fifth in a series of such symposia meeting at approximately three-year intervals. The principal sponsor of all of the meetings has been Commission 6 of the International Scientific Radio Union (URSI). On this occasion the cosponsors were the Nederlands Electronischen Radio Genootschap (NERG), the Technological University of Delft, and the Netherlands National Committee of URSI.

The nature and activities of URSI are familiar to those IEEE members who had been members of the IRE. For the benefit of others, who are unfamiliar with URSI, it may be appropriate to point out that it is an international organization formed for the exchange of information on new developments in the field of radio science. Its field of interest is subdivided among seven commissions within URSI, including Commission 6—Radio Waves and Circuits. The officers of URSI are selected by the National Committees, one for each nation represented in the Union. The Union meets triennially at General Assemblies, the last of which, the XIVth, was held in Tokyo in September 1963. The XVth General Assembly is to be held in Munich in September 1966. In the intervening three years, the National Committees—organized very much along the lines of the international parent body—held meetings within the countries concerned. The United States National Committee traditionally holds two meetings per year. In the past the spring URSI meeting has always been held in Washington, D.C., in April or May, and the fall meeting at various locations, but usually at a university. The last meeting of the United States National Committee was held October 4–6, 1965, at Dartmouth College in Hanover, N.H. The next one is scheduled for April 19–22, 1966, at the National Academy of Sciences, Washington, D.C.

Since 1953, in addition to the general assemblies of the entire international organization, the international Commission 6 has held meetings devoted solely to its area of interest. (The one preceding the Delft meeting was held in June 1962 in Copenhagen, Denmark.) At the Delft meeting, 132 technical papers representing contributions from 16 countries were delivered to an audience of approximately 260 registrants from 19 countries. The Organizing-Technical Committee of the Symposium found it convenient to classify the accepted papers into 11 groups, each devoted to its own specialized subject. A total of 17 half-day technical sessions were held, including a special opening session. Two sessions were run simultaneously each day except Wednesday, when no technical sessions were held. This day was left open so that the participants could visit the artificial island of Haringvliet, the central point of interest for the Delta Project, and also could become better acquainted with each other.

Opening session

The first paper of the opening session was an invited survey paper entitled "Millimeter Waves and Optical Waves for Long-Distance Telecommunications by Waveguide" by Prof. H. M. Barlow (University College, London). In his paper, Prof. Barlow mentioned the roles of microwave relay stations, satellite relay systems, and coaxial cables in long-distance communication, and

H. V. Cottony is chief of the Antenna Section of the Ionospheric Telecommunications Laboratory of the Institute for Telecommunication Sciences and Aeronomy, Boulder, Colo. This Institute was formerly known as the Central Radio Propagation Laboratory of the National Bureau of Standards.

pointed out the advantages of the waveguide type of transmission; among these is a large bandwidth. He described an experimental circuit that was installed at University College in London to operate between 30 and 40 Gc/s. An H_{01} waveguide with a copper tube lined with a dielectric was employed. Helix filters were inserted at intervals and bends up to 90 degrees using flexible corrugated brass tubes also were introduced at appropriate intervals. The loss brought about by one such bend was 0.2 dB throughout the 10-Gc/s band of the circuit for any change in direction up to 90 degrees. The use of the H_{01} waveguide in Japan for communication with and control of moving railway trains was mentioned, as was the use of a surface wave using an inverted V-section guide for the source. The problems associated with the use of a waveguide for transmission at optical frequencies were discussed.

The second of the two invited papers selected for the opening session was entitled "Experimental/Theoretical Evaluation of a Passive Communications Satellite (Echo II)" by A. Kampinsky (Goddard Space Flight Center, Greenbelt, Md.), and R. K. Ritt (Conductron Corporation, Ann Arbor, Mich.). This paper considered in some detail the work that went into the Echo II project. The reflective properties of the balloon were studied experimentally and theoretically, using a full-scale inflated balloon, scaled models, and, finally, segments of balloon material. On the basis of this study, performance of the satellite as a passive reflector was predicted. The results were used to evaluate the flight test balloon.

Other sessions

Another session was devoted to theoretical considerations of various aspects of propagation in layered, dissipative, and inhomogeneous media. The last three papers in this group discussed special cases of diffraction. [It may be worthwhile to note that the fifth paper in the group, by C. R. Burrows (U.S.), appeared in the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, vol. 13, pages 759-774, September 1965.]

Antenna theory papers were presented in a session that was based on a somewhat broad interpretation of the term. The paper by A. Kampinsky and R. K. Ritt (U.S.) on the Echo satellite was presented at the opening session and was mentioned earlier. Of the four or five papers in the aperture synthesis area, one, by P. C. J. Hill (United Kingdom) on the synthesis of cosecant radiation pattern, was of greater than average interest to this reporter. In the solution advanced in the paper, an array of equidistant elements with equal current amplitudes is made to yield approximately the desired pattern by adjustment of the phases of the currents. As a final result, a pattern closely following the cosecant contour, with nulls reduced to broad minima and sidelobes to a fluctuation in amplitude of secondary importance, was obtained.

An interesting and quite possibly an important paper was that by L. B. Felsen (U.S.) on lateral waves. In his invited paper, Dr. Felsen discussed the generation of lateral waves by reflections and refraction at the common surfaces of different media with the radiator located close to the interface. K. Iizuka (Gordon McKay Laboratory, Harvard University, U.S.) presented a paper describing a technique for modeling a medium with a layered structure. His technique involves the use of agar-agar gelatin for stabilizing the position of layers having different conductivities and other electrical properties.

A paper on high-gain back-fire array by H. W. Ehrenspeck (U.S.) was the source of lively discussion, with E. Spitz (C.S.F. of Paris, France) contributing a description of analogous work being carried on in France.

A group of 12 papers was devoted to the subject of propagation in plasmas. Since this reporter was attending another session, and since he claims scanty knowledge of the subject, the report must likewise be scanty. A review of distributed summaries reveals that three of the papers can be classified as related to the ionosphere. At least two dealt with a standard electromagnetic theory topic—diffraction—but made substantially more complex by the presence of anisotropic ionized medium. None of the papers in the group appear to be related to the subject of communication with space vehicles at reentry. However, there was a paper pertaining to this subject—"Radiation Properties of Slot Antennas on Cones in the Presence of Plasma" by M. Katzin—that was "communicated" rather than presented because the author was unable to attend the meeting.

Underground space waves were covered by five papers, including the invited one by J. R. Wait (U.S.). Only the paper by A. W. Biggs (U.S.) and H. M. Swarm (U.S.) actually dealt with underground propagation, and that one only marginally since the "ground" was antarctic ice. Three of the remaining four papers dealt with problems connected with ground-wave propagation while the fourth one, by M. Aubry and J. Cerisier (France), presented a method of computing the illumination of the ionosphere at a 100-km height by a VLF ground-based transmitter. The results of the computations appear to to be in good agreement with experimental measurements using the LOFTI satellite.

Papers devoted to boundary value problems had one common characteristic: mathematical tools employed were of more importance than the problems attacked. Other than that, there was relatively little to unify the subject matter of the papers. It is not possible at this time to pass judgment on the significance of the material presented. Some of the mathematical tools displayed may very well be of great potential importance.

The session on millimeter waves and optical waves was one of the most interesting to a member of the engineering profession. The first paper by Prof. Barlow, delivered at the opening session, has already been discussed. The fourth paper of this group, "A Circular H_{01} Low-Loss Waveguide Applicable to Long-Distance Communications" by H. M. Barlow, H. G. Effemy (United Kingdom), and S. H. Taheri (United Kingdom), was a supplement to the first one. It went into somewhat more detail as to the reasons (mostly economic) for the selection of a dielectric-coated rather than a helical waveguide, the method of making attenuation measurements, the desirable spacing of repeaters (about two miles), the technique for transmission of a TV signal, and other important but detailed points.

Another interesting paper in this group was the second, "Lasers and Optical Communications Systems" by A. E. Karbowiak (Australia). To a large extent this was a survey paper comparing various types of waveguides, metallic vs. dielectric vs. beam, especially as they function at optical frequencies. One type of waveguide—the optical microguide—was new to this reporter. As described in

I. Distribution of papers among contributing countries

	Propagation in Inhomogeneous Media	Antenna Theory	Propagation in Plasmas	Underground Waves/ Space Waves	Boundary Value Problems	Millimeter Waves and Optical Waves	Propagation in Nonlinear Media	Surface Waves and Wave Beams	Coherence Problems and Modern Optics	Multiple Scattering	Deterministic Scattering	Total
Group	1	2	3	4	5	6	7	8	9	10	11	
Algeria									1			1
Australia	1					1						2
Canada			1		1		1					3
Denmark		2	1					1		2		6
Finland		1										1
France			2	1	1	2		2	1			9
Germany		1			2			2		1		6
Hungary		1			1							2
Italy		2				1	1	2	1		½	7½
Netherlands	3	4			1			2				10
Norway							1					1
Poland	1		1									2
Sweden			1				1					2
U.S.S.R.		1			3	2	7	1		1		15
U.K.	1	1	2		1	3						8
U.S.A.	4	10	4	4	8	1	2	4	8	5	6½	56½
Total	10	23	12	5	18	10	13	14	11	9	7	132

the paper, it consists of a relatively narrow (e.g., 1 cm) extremely thin and long film of dielectric, with the long edges clamped between supporting structures. The thickness of the film is small compared with the wavelength, e.g., 5 × 10^{-8} m. The light energy is launched along the film in what is a plane surface wave mode with the plane of polarization normal to the film. The film can be curved in a plane normal to its surface with little loss but it cannot be curved in the plane of the surface. It can, however, be twisted, and by a combination of twists and bends the light energy can be guided in any direction. The microguide is strictly in the experimental stage. One of the arguments advanced in this paper for the intrinsic advantage of an optical communication system involves the low utilization of this portion of the electromagnetic spectrum. This reporter has heard this argument before and has always found it difficult to accept. The fact is that it is in greater everyday use for communication and information-gathering purposes by more people than is any other part of the spectrum. It does not require instrumentation, merely eyesight.

The papers devoted to propagation in nonlinear media may be of significance but it would be difficult to convey this significance in a report such as this. It may be of interest to note that seven of the 13 papers in this group originated in the U.S.S.R.

In the session on surface waves and wave beams, the second invited paper, by Prof. A. A. Oliner (U.S.), and the paper by J. Bach Anderson (Denmark) considered surface wave structures from the standpoint of their utility as radiators. Most of the remaining papers were more concerned with the transmission of energy in the surface-bound form. An interesting application of radially propagating wave beams to a resonator consisting of two coaxial rings was described by G. Goubau (U.S.) and F. Schwering (U.S.) in their paper, "A Mode System for Radially Propagating Wave Beams."

The papers on coherence problems and modern optics represent, possibly, the most significant stride beyond the scope of the previous symposia. While papers on coherence and partial coherence problems were presented at Copenhagen and at other electromagnetic symposia, the number of papers on this subject at this meeting, coupled with the entry into the optical frequencies, represents a notable advance in the scope of interest of Commission 6. To this reporter, the paper that best combined interest with understandability was the fourth, "Coherence Measurement in Radio Propagation" by C. Colavito (Italy) and G. d'Auria (Italy). This paper discussed the use of the interferometer technique for the measurement of coherence. In addition to theoretical discussion, experimental measurements of coherence modulus were presented. These measurements were carried out at 9375 Mc/s with two antennas in broadside, with separation between antennas varying from 0 to 80 meters. It was a happy combination of theory with experiment.

The sessions on multiple scattering—scattering at rough boundaries—and deterministic scattering had much in common. Both were held on the last day and both were under some pressure to conclude somewhat earlier than originally scheduled. The four morning and afternoon sessions originally contemplated were combined into two long morning sessions. The matter of singling out a paper in a session for purposes of highlighting is likely to be a reflection of current interest on the part of the listeners. The paper on "Scattering and Transmission of Electromagnetic Waves" by T. Hagford (U.S.) is such a case. While resembling some of the recent papers in approach, it differed in assuming two nonconducting dielectric media rather than a dielectric and a perfect conductor. It made use of a tangential-plane approximation and also assumed that the shadowing effect is not important. The scattered and emitted angular power spectra were explored for several cases of roughness and dielectric media. The results were presented graphically in the form of constant power contours. It was an interesting paper, even at the end of a week-long symposium.

Other symposia

The adjournment of the Delft Symposium completed another entry in the international record of progress in electromagnetic theory. Its value can only become evident with time. This was the fifth in the series of Commission 6, URSI Symposia on Electromagnetic Theory. The first of these events was held June 22–25, 1953, at McGill University in Montreal, Que., Canada. It was then called the Symposium on Microwave Optics. Eventually, the papers presented at this symposium, mostly in summary form, were issued as an Air Force Cambridge Research Center Technical Report AFCRC-TR-59-118(I), ASTIA Document No. AD 211499, dated April 1959. The second in this series of symposia was held June 20–25, 1955, at the University of Michigan, Ann Arbor. The proceedings of this symposium were published in the regular issue of IRE TRANSACTIONS ON ANTENNAS AND PROPAGATION, vol. AP-4, July 1956.

The third in this series was held at the University of Toronto in Canada, June 15–20, 1959. Its proceedings were reported in a special issue of IRE TRANSACTIONS ON ANTENNAS AND PROPAGATION, vol. AP-7, December 1959. The proceedings of the fourth "Symposium on Electromagnetic Theory and Antennas," held in Copenhagen, Denmark, June 25–30, 1962, were published by Pergamon Press in 1963. Current plants are for Pergamon Press to publish the proceedings of the 1965 symposium some time in 1966. Prof. John Brown of University College is to be the editor.

Of some interest is the question of what areas are being worked on by different countries. Some indication of this work is given by Table I. As might be expected, the United States leads the list with 56½ contributions or 43 percent of the total number. It made contributions to each category of papers with the least, somewhat surprisingly, to the millimeter and optical waves group. The United States contribution to the propagation in nonlinear media group also was small. The U.S.S.R., a relatively recent (1957) member of URSI, was next with a total of 15 papers. Almost half of these were in the propagation in nonlinear media group. The contributions from other countries were approximately what might have been expected.

It was a hard-working meeting, with the morning sessions beginning at 9:00 A.M. and some of the afternoon sessions continuing until 6:00 P.M. and later. Since Delft is a small city, a good many of the attendees stayed in The Hague and Scheveningen, about 12 and 15 km distant. There were compensations, however. On Monday evening, September 6, there was a reception at which the attendees had an opportunity to meet each other. On Friday evening, there was a banquet.

Delta Project

The most outstanding nonbusiness event was the visit to the Delta Project at the mouth of the Rhine and Maas Rivers. This is the second of the three mammoth reclamation projects undertaken by the Dutch. The first, involving the sealing off of the Zuider Zee from the North Sea, was completed in 1932. There is no more Zuider Zee. In its place there now is a fresh water lake named Ijsselmeer.

On February 1, 1953, a severe storm in the North Sea led to the flooding of the estuary of the Rhine and Maas Rivers causing extensive damage, including the loss of over 1800 lives. The Delta Project is intended to prevent the recurrence of such a castrophe. It involves plans to dam the three estuary channels of the Rhine and Maas Rivers, thereby preventing the tides and storms from entering. The Haringvliet Island is the keystone to the project. It is an artificial "underwater island" formed by a coffer dam encircling the site of 17 sluice gates. These gates will open to permit the drainage of the Rhine and Maas Rivers water into the North Sea at low tide and close to prevent the re-entry of the sea at high tide.

Apparently in common with many others, this reporter believed the purpose of the Delta Project was to reclaim land. That is a tertiary object. The primary purpose is to prevent the flooding of the present dry land by storms. The secondary objective is to replace the sea water in the estuary channels by fresh water from the rivers. The reclamation of arable land, "polder," from the sea is a goal only third in the order of importance. When the Delta Project is finished, in 1978, the third major project will be considered. This consists of connecting together the islands off the Friesian coast and sealing the North Sea from the intervening body of water. Completion of this project is decades away.

The successful conduct of the symposium is attributable to the efforts of many people: the authors of papers, the chairmen of the sessions, and the Organizing and Technical Committee. The chairman of the latter was Dr. F. L. Stumpers, who also is the international chairman of URSI Commission 6. The other members are H. M. Barlow (United Kingdom), H. Bremmer (Netherlands), H. A. W. Goossens (Netherlands), P. Grivet (France), A. T. de Hoop (Netherlands), O. E. H. Rydbeck (Sweden), J. P. Schouten (Netherlands), K. M. Siegel (Chairman of Commission 6, U.S.N.C.), V. I. Siforov (U.S.S.R.), B. D. H. Tellegen (Netherlands), R. Timman (Netherlands), and G. Toraldo di Francia (Italy).

The success of large meetings, such as this one, depends on the services of local arrangements committees. The very smooth and efficient operation of this symposium was attributable to the large local staff headed by Admiral H. A. W. Goossens (Royal Netherlands Navy, retired), who is the chief of technical and domestic services in the faculty of General Sciences at Delft.

The assistance by Drs. Stumpers and Wait in checking the accuracy of this report is acknowledged with thanks.