VOLUME 2 □ NUMBER 7 JULY 1965

# IEEE spectrum

## features

61 Spectral lines

+ 62 **Progress in semiconductor lasers**
Benjamin Lax *An attempt is made to bridge the gap between the injection laser and more recent work on optical photon and electron pumped lasers*

+ 76 **Mariner Mars 1964 telemetry and command system**
Richard P. Mathison *An important function of the Mariner Mars 1964 project is the transmission of video and scientific data from the vicinity of Mars*

+ 85 **The role of grounding in eliminating electronic interference**
Trevor A. Robinson *A common grounding language is needed to dispel the existing confusion in the design of electronic equipment and power equipment*

+ 90 **Nuclear power in the world today**
Louis H. Roddis, Jr. *In many countries nuclear power has attained the same status as conventional fuels, with local economics dictating the final choice*

+ 98 **Information storage density**
Marvin Camras *Old and new methods of storing information are compared with theoretical limits based on compactness, speed, and total bit capacity*

+ 106 **Environmental problems in external space equipment design**
John D. Alber, Richard H. Imgram *Environmental hazards to be considered in the Lunar Excursion Module include temperature extremes, radiation effects, and micrometeorites*

+ 116 **Report on the A. S. Popov Society meeting**
F. Karl Willenbrock
Ideologies versus ideas in human progress B. M. Oliver

+ 118 **Report on the 1965 INTERMAG Conference**
Robert F. Elfant

+ 120 **Report on the IEEE–NBS Particle Accelerator Conference**
J. A. Martin, R. S. Livingston

122 Authors

32 WESCON Exhibitors

128 WESCON program

Departments: *please turn to the next page*

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

## departments

## the cover

The planet Mars, 141 million miles from the sun, has a thin atmosphere, some moisture (note white polar cap), severe temperatures, and seasonal changes. In this reproduction of a black-light mural in the American Museum–Hayden Planetarium, New York City, the dark regions may be primitive vegetation. For a description of the telemetry and command system for the Mariner Mars 1964 probe, see page 76.

VOLUME 2 ☐ NUMBER 7 IEEE spectrum JULY 1965

# Spectral lines

**More on Publications.** Although the last three Spectral Lines' articles have been devoted to the subject of the technical publications of this Institute, the Editor requests the indulgence of the readers of this feature if he again directs their attention to this subject. Although the various conventions, symposia, and meetings sponsored by the Institute play an important and unique role in the dissemination of technical information, it is of interest to note that a relatively small percentage of its members are actively involved in these meetings. While it is very difficult to obtain precise figures, it is estimated that the total attendance at IEEE meetings is 265 000 per year. In contrast, the IEEE publishes and distributes in one year approximately 2 300 000 copies of periodicals devoted to technical material. While it might be argued that this comparison is not a valid one since the publication of an article does not mean everyone reads it, it should also be pointed out that just by being in the same meeting room in which a paper is given does not mean that one has heard or understood it. The point is that the printed page carries the major share of the dissemination of technical information, which is the objective of this Institute.

What requirements should our publications meet? Perhaps the most fundamental requirement is that the papers we publish be technically sound. While the author or authors of an article certainly bear the ultimate responsibility for the validity of a paper, the Institute must have reviewing procedures that can separate with great accuracy the technically sound from the technically unsound articles. While it is probable that some mistakes will be made, keeping such mistakes to a very low percentage of the total output is an important continuing responsibility of the editors and reviewers. The proper discharge of this responsibility is of utmost importance if this Institute is to be of real service to its profession.

The next most important requirement is that we minimize the time needed for processing and printing a submitted paper. It is obvious that prompt, widespread, and accurate dissemination of information is essential for the rapid advancement of our profession. Delays in publication of new technical information can result in the unnecessary duplication of effort, wasteful use of resources, and a general slowdown in the rate of advancement. It is significant that the most rapidly developing technical fields are frequently those which have the most rapid publication facilities. Thus, if we are inefficient in our review and editing procedures, we are guilty of slowing down technical progress. It is essential that we eliminate all unnecessary delays and that we include in our reviewing processes only those procedures which are absolutely necessary to ensure technical soundness. The fact that in many of our publications the delay between submission and publication of a well-written paper is almost one year indicates that there is much room for improvement. Both administrative and technical innovations are urgently needed in this area.

It is probably not possible in an Institute having such a diversified membership as the IEEE to adopt a single procedure for reviewing and accepting papers for publication. However, it is certainly true that our technical publication procedures should be such that technically sound papers are made available to the members throughout the world with the least possible delay.

*F. Karl Willenbrock*

# Progress in semiconductor lasers

*The semiconductor laser field has progressed at a tremendous pace during the past few years. This article attempts to bridge the gap between the now-familiar injection laser and more recent work on optical photon and electron pumped lasers*

*Benjamin Lax* *National Magnet Laboratory, M.I.T.*

Many review articles[1-5] have been published on the subject of semiconductor lasers since 1962, when the first junction diode laser of gallium arsenide was developed. Each of these articles reviewed some of the same background and fundamental principles from different viewpoints and also included updated experimental and theoretical results. Perhaps the most complete and comprehensive review in this regard is the article by Burns and Nathan,[5] which also has a copious list of references up to July 1964 and thereby excuses the writer from duplicating their heroic accomplishments. In view of this, what justification is there in still another review? The writer searched his soul and, after much procrastination, decided that there was some virtue in a complementary attempt that bridges the gap between those references concentrating on the injection laser and more recent results on optical photon and electron pumped semiconductor lasers. In addition, this also provides an opportune medium for discussing those aspects of semiconductor lasers which are of particular interest to the writer and his colleagues at Lincoln Laboratory. Consequently, failure to include significant material or references presented elsewhere is purely intentional, and the writer can only hope that his sins of omission will be forgiven by his readers and critics.

## History

Soon after the development of the ammonia maser by Gordon, Zeiger, and Townes[6] in 1954, physicists began to speculate about the possibilities of solid-state masers and even lasers. Since it was known that semiconductors luminesce and emit in the optical and infrared regions, some thought was given to this possibility. Among the first to suggest a means of excitation was John Von Neuman,[7] who proposed in correspondence with colleagues an electrically pumped device very much like the present junction diode laser. However, his suggestion was not specific and was somewhat ahead of the state of the art. Another possibility of using point-contact devices for excitation was entertained by Pierre Aigrain[8] in 1957 during the course of a series of lectures at the Massachusetts Institute of Technology. He and his group began active work on this and other schemes for electrical and optical pumping of semiconductors, such as germanium and silicon, without success and reported these ideas at the International Conference at Brussels in 1958.[8]

Two other groups, one at Lincoln Laboratory and the other at the Lebedev Institute, also started independent efforts for achieving laser action in semiconductors. The Lincoln group considered a number of devices, such as the cyclotron resonance maser[9] and the interband and impurity lasers, which appeared feasible on theoretical grounds. The group at the Lebedev Institute, under the direction of Basov,[10] considered analogous devices using electrical excitation.

Many of these ideas, which were not fully crystallized, were informally discussed at the First Quantum Electronics Conference in September 1959. Some were tested experimentally and reported at the Second Quantum Electronics Conference in 1961. Although none were successful, the basic concepts were better understood and quantitative considerations were presented. It was at this meeting that Basov first proposed the possibility of exciting semiconductors by an electron beam. It was shortly after this that Bernard and Durrafourg[11] showed theoretically the necessary criterion for inverting the population in a semiconductor. The practical developments commenced when a number of investigators began to examine the emission from gallium arsenide. The most striking results were those of Keyes and Quist,[12] who demonstrated an order of magnitude increase in emission from a GaAs diode near the energy gap when they lowered the temperature to 77°K, with line narrowing and high quantum efficiency. This immediately suggested that such a structure would offer a possibility for a semiconductor laser. Independently, on theoretical grounds, Dumke[13] predicted that gallium arsenide would be most suitable for a direct transition laser and indicated pessimism about the feasibility of an indirect transition laser. The first such gallium arsenide diode laser was reported by the General Electric group[14] in the fall of 1962 and was soon followed by similar results from the groups at IBM[15] and Lincoln Laboratory.[16]

Much work has been done since then on junction diode lasers by many groups. Only highlights which appear subjectively relevant will be mentioned in this historical

Some of this work was done while the author was at M.I.T. Lincoln Laboratory, which is operated with the support of the U.S. Army, Navy, and Air Force. The National Magnet Laboratory is supported by the Air Force Office of Scientific Research.

review. One of the next interesting developments was the fabrication of an alloy diode laser[17] of $GaAs_{1-x}P_x$ which emitted in the red region. This was followed by others using InAs,[18] InP,[19] and alloys of $In_xGa_{1-x}As$,[20] and also of $InAs_{1-x}P_x$,[21] which emitted in the infrared region from 8400 to 32 000 Å (3.2 microns). Although these first lasers operated on a pulse basis, CW operation was soon achieved by several techniques, including the use of magnetic fields to lower the threshold current needed for laser amplification. This method was rather spectacular when InSb lasers were subjected to magnetic fields as high as 100 000 gauss.[22] The operation of a tunable magnetooptical laser was finally demonstrated and operated according to theoretical predictions.

Tuning of semiconductor lasers has also been achieved by varying the temperature and by subjecting the laser to hydrostatic and uniaxial pressure. Room-temperature operation of diode lasers has also been possible on a pulse basis with high threshold currents. Harmonic generation[23] from GaAs lasers has now extended the emission frequency to 4200 Å for the second harmonic. In the long-wavelength region, diode lasers of PbTe[24] and PbSe[25] have extended the spectrum to 6.5 and 8.5 microns, respectively. Large-volume bulk lasers of InSb have been operated with the radiation transverse to the diode current and also parallel to it.[26]

The next major development was the electron excitation of semiconductor lasers, which was first proposed by Basov[27] in 1961. In fact the attempts by his group at Lebedev[28] in 1963 on CdS resulted in line narrowing, which indicated possibilities for laser action by this type of pumping. The first actual coherent radiation by electron pumping was achieved in InSb and InAs by Benoit à la Guillaume and Debever[29] in July 1964. Subsequently Hurwitz and Keyes,[30] using similar techniques, succeeded in producing laser action in GaAs by electron excitation. The laser-action threshold was studied as a function of doping experimentally by Cusano[31] and theoretically by Klein.[32] The French group about the same time[33] excited GaSb as a coherent source and later produced coherent emission from tellurium, demonstrating unequivocally that a direct transition existed in this material.

A variety of applications for diode lasers have been developed. Among these has been the use of a GaAs laser to optically pump[34, 35] InSb and InAs crystals, which in turn were made to emit coherent radiation. Demonstrations of the use of GaAs diodes for communication and television transmission[36] were successful over modest distances. Radars using both low-temperature and room-temperature operation of GaAs diode lasers have shown capability of high spatial and angular resolution.[37] The GaAs laser has also been employed as an infrared source for Raman spectroscopy.[38] These lasers have been incorporated in the light of beam transistors by a group from the IBM Corporation.[39] The most recent development has been the scanatron,[40] a scanning beam semiconductor laser that, if fully developed, has a number of potentially useful applications.

**Basic principles**

In order to operate any laser, certain basic requirements have to be met. The first of these is that there must be a means of excitation or pumping of electrons from a lower to a higher level. In semiconductors a number of these techniques have been proposed so far. Only three, however, are practical: (1) electrical pumping by injection, as in a diode; (2) electron-beam excitation of hole–electron pairs; and (3) optical excitation of hole–electron pairs. The second criterion that has to be met is that the excitation process invert a sufficient number of electrons to produce sufficient stimulated emission to overcome the losses. The third requirement is that a resonant geometry be available to provide feedback so that all the photons emitted are properly phased to one another—in other words, that the radiation is coherent.

In this discussion we shall restrict ourselves to the consideration of the foregoing processes in connection with the direct transition lasers that have been operated successfully. First, however, we should explain what we mean by an inverted population in a semiconductor. The energy levels in the conduction band of a semiconductor are not discrete, as they are in atomic energy levels or in localized impurity centers in solids. The electrons are in broad bands separated by a forbidden region so that the meaning of inverted *levels* is not well defined. However, inversion can be understood physically for the following simple situation: If an intrinsic semiconductor at very low temperature is suddenly subjected to a pulse of electric or optical energy in such a way that hole–electron pairs are created, and these pairs are allowed to come to equilibrium in the valence and conduction bands, respectively, in a time that is short compared to the recombination time, then two degenerate populations are inverted with respect to one another. Electrons with a certain value of momentum $p$ in the conduction band can only make transitions downward to the valence band with the same value of $p$; that is, the selection rule states that $\Delta p = 0$ is a strongly allowed transition. From inspection of Fig. 1 we see that the photon distribution in energy $h\nu$ becomes

$$\mathcal{E}_g < h\nu < F_c - F_v \qquad (1)$$

where $\mathcal{E}_g$ is the energy gap, $F_c$ and $F_v$ are the quasi-Fermi levels of the electron and hole distributions, and $h\nu$ is the energy of an emitted photon. At $T = 0$, these Fermi energies correspond to the top and bottom of the electron and hole populations. At higher temperatures the distributions are not well defined or sharp, and there are both filled and empty states above and below the Fermi levels. Consequently, photons are not only emitted, but can also be absorbed in the range of values represented by Eq. (1). More will be said later about this. Nevertheless, under steady-state conditions when the hole–electron pairs are produced continually, the two distributions of carriers are not in equilibrium with one another, but with the lattice. Hence the Fermi distributions defined by $F_c$ and $F_v$

are still meaningful for the particular temperature of the crystal lattice.

It is interesting to note that at low temperatures the spectral distribution in an idealized interband laser shown in Fig. 1 would be determined by the condition of Eq. (1), which for a simple parabolic band would be determined by the density of states:

$$\hbar\Delta\nu = \mathcal{E}_c + \mathcal{E}_v \tag{2}$$

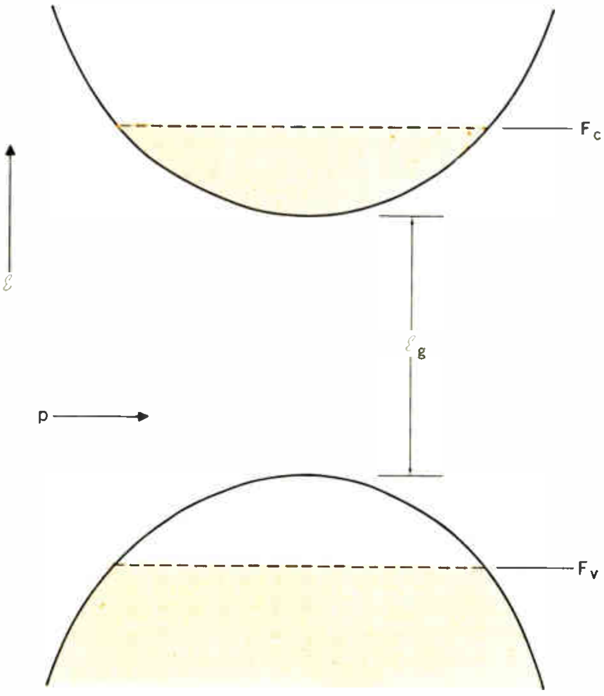
where $\mathcal{E}_c$ and $\mathcal{E}_v$ are the energies above and below the conduction and valence bands; that is,

$$\mathcal{E}_{c,v} = \frac{\hbar^2}{2m_{v,c}}\left(\frac{3N}{8\pi}\right)^{2/3} \tag{3}$$

where $m_c$ and $m_v$ are electron and hole masses and $N$ is the charge carrier concentration. If we assume a recombination lifetime of about $10^{-9}$ second, then the number of carriers excited just below threshold in a laser is typically $10^{17}$ cm$^{-3}$. This would give $\mathcal{E}_{c,v} \approx 0.01$ eV, which is a value close to that observed for the line width of the spontaneous emission of a diode laser at low temperatures. The line shape, however, would be definitely asymmetrical. In any case, this simple mechanism is not quite the correct one; other evidence indicates that interband transitions between ideal bands do not occur and that impurities play a role in determining the energy of transition and possibly in modifying the energy bands in degenerate semiconductor regions.

In an actual diode laser the two inverted regions are created in a region, near the junction, in which the large forward bias of the junction produces an overlap area between the valence and conduction bands as shown in Fig. 2. The diagram suggests that the n region is more heavily doped and hence the injection of excess hole–electron pairs is into the p region. This is the situation which has been experimentally observed, although in principle holes can be injected from a heavily doped p region into a lightly doped n region. In the present arrangement the emission would occur on the p side of the junction, where the electrons are inverted relative to the holes. The voltage applied to the junction is comparatively high—that is, somewhat larger than the energy gap—to be consistent with Eq. (1).

Fig. 1. Inverted population of electrons and holes in conduction and valence bands of a semiconductor.



Once the population is inverted, the requirement that must be met is that the stimulated emission shall exceed the losses in the laser structure. Factors that contribute to the loss of energy include (1) bulk losses of pumping energy caused by free carrier absorption; (2) reabsorption of the photon to be emitted if the laser is not at a very low temperature, and also if the population of holes and electrons is not degenerate; (3) a geometric loss factor arising because the inverted region is narrow and the luminescence overlaps lossy regions; and (4) reflection losses. We shall consider these one at a time and idealize our laser structure step by step until all of these factors are taken into account. Let us consider, therefore, the simplest configuration: a bulk laser at very low temperature in which the only loss to be overcome is that of the free holes and electrons excited into the inverted region by the pumping energy. Pumping can be done by means of optical energy, by an electron beam, or by the injection of carrier pairs, as in the bulk plasma laser. The latter two approximate this situation fairly well. Under these conditions we can determine approximately the condition for laser action by the scheme developed by Dumke[11] to predict such a possibility. He compared the negative absorption coefficient or essentially the probability of emission for a photon to the absorption by a free carrier. The coefficient for direct transition and hence the value of gain per unit length is given by

$$\alpha_g = A\sqrt{h\nu - \mathcal{E}_g} \tag{4}$$

where $A$ is a coefficient calculated theoretically and depends upon both the momentum matrix element for allowed transitions between bands and also the density of states. These quantities for GaAs are known from optical absorption data, for example. Hence for a degenerate population of carriers, $N \approx 10^{17}$ cm$^{-3}$, the maximum value of $\alpha_g$ is calculated from Eq. (3) to be

$$\Delta\mathcal{E} = h\nu - \mathcal{E}_g \approx \frac{\hbar^2 N^{2/3}}{m^*}$$

where $m^*$ is the reduced mass. For GaAs we find that $\alpha_g = 10^3$ cm$^{-1}$. This is somewhat higher than the average value, but since the density of states increases with energy and also the number of electrons are weighted, $N \sim (\Delta\mathcal{E})^{3/2}$, this is not too different from the average value. We then have to compare this with the losses for free carriers, which are given by

$$\alpha_c = \sigma\sqrt{\mu_0/\epsilon} \tag{5}$$

where the conductivity

$$\sigma = \frac{Ne^2\tau}{m^*(1+\omega^2\tau^2)} \approx \frac{Ne^2}{m^*\omega^2\tau}$$

If we use typical values, $m^* \approx 0.1\,m_0$, $\tau \approx 10^{-13}$ second, $\omega = 2 \times 10^{15}$ for GaAs, then $\alpha_c \approx 0.2$ cm$^{-1}$.

Thus we see that the free-carrier absorption is several orders of magnitude smaller than the gain from stimulated emission. In an actual laser, the stimulated transition takes place between bound donor or exciton states to acceptor states. The values for the absorption coefficient for such transitions have been measured or calculated theoretically and give $\alpha_g \approx 300\,\text{cm}^{-1}$ for a free-electron to acceptor concentration of $N = 10^{17}\,\text{cm}^{-3}$ comparable to the band-to-band transition above.

In a diode laser the width of the inverted region is usually of the order of one to several microns and is bounded on either side by degenerate material in which the free-carrier losses dominate. The electromagnetic boundary value problem has been solved for such a configuration assuming a narrow, extended, negative-conductance sheet,[42] which behaves like a dielectric waveguide. The electromagnetic field peaks up in this region and decays exponentially on either side transverse to the inversion layer. By idealizing the configuration, the growth coefficient can be calculated for TE and TM modes and the threshold current evaluated. A similar result can be obtained by approximation using a perturbation treatment, and the condition for threshold corresponds to the balance of net gain in the inverted region to the loss in the bulk on either side:

$$\alpha \approx \frac{\int_{-W/2}^{+W/2} \sigma_1 E_1{}^2\,dx - \int_{W/2}^{\infty} \sigma_2 E_2{}^2\,dx - \int_{-W/2}^{-\infty} \sigma_3 E_3{}^2\,dx}{2\int_{i_z} [\mathbf{E} \times \mathbf{h}]\,dx} \tag{6}$$

where **E** and **h** are the electric and magnetic vectors of the electromagnetic field, $\sigma_{1,2,3}$ is the conductivity in each region, and $i_z$ is the unit vector in the $z$ direction. Region 1 is the junction region of width $W$. If we assume that $E_1 \approx E_0$, $E_2 \approx E_0 e^{-\alpha_2 x}$, and $E_3 \approx E_0 e^{\alpha_3 x}$, then, assuming a skin depth penetration $\delta$ into the degenerate regions, the value of $x$ is

$$\begin{aligned} \alpha &\approx \frac{W\sigma_1 - \sigma_2/2\sigma_2 - \sigma_3/2\alpha_3}{1/\alpha_2 + 1/\alpha_3}\sqrt{\frac{\mu_0}{\epsilon}} \\ &= \frac{W\sigma_1 - (\sigma_2\delta_2 - \sigma_3\delta_3)/2}{\delta_2 + \delta_3}\sqrt{\frac{\mu_0}{\epsilon}} \end{aligned} \tag{7}$$

At threshold $\alpha \geq 0$, which means that

$$\sigma_1 > \frac{\sigma_2\delta_2 + \sigma_3\delta_3}{2W}$$

$$\text{or } \sigma_1 > \frac{\sigma_2\delta_2}{W} \qquad \text{for } \delta_2 = \delta_3,\ \sigma_2 = \sigma_3$$

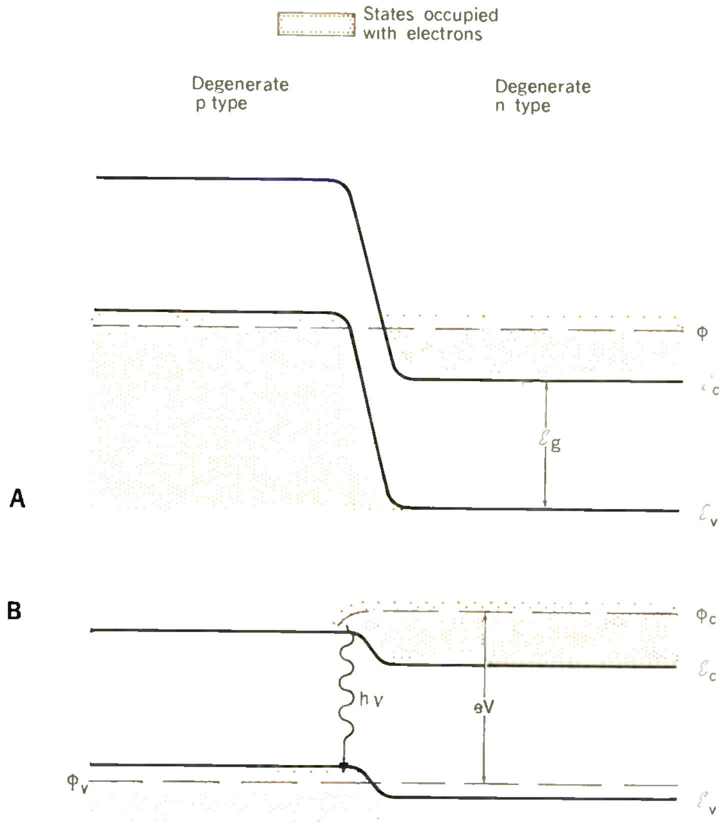
Hence the gain in the actual diode has to be larger than in the bulk by the ratio of the skin depth $\delta$ to the width of the inversion layer. This geometrical factor is usually about 10 in most diode lasers. It is now understandable why a direct transition is desirable since $\sigma_1 >> \sigma_2$, as we have shown. For the indirect transition this is not the case and at best the negative conductivity is comparable to the positive loss; that is, $\sigma_1 \gtrsim \sigma_2$. In a diode, therefore, the geometrical factor makes this situation unfavorable for achieving the laser condition for germanium, silicon, or silicon carbide for shallow impurities.

In order to obtain coherent radiation after the condition for stimulated emission is satisfied, it is necessary to provide a resonant structure. In a diode laser, the structure is a Fabry–Perot cavity, which consists usually of two parallel polished surfaces transverse to the plane of the junction, and provides a reflection of the growing wave as well as a reflection loss. If this reflection is taken into account in calculations of the threshold current $J_{\text{th}}$ as well as the bulk losses, it takes the form[43]

$$J_{\text{th}} = A\left[\bar{\alpha} + \frac{1}{L}\log\frac{1}{R}\right] \tag{8}$$

where $\bar{\alpha}$ is the average free-carrier loss within the extrinsic n and p regions surrounding the diode junction. The second term represents the reflection loss, with $L$ as the length of the diode and $R$ the reflection coefficient. The coefficient $A$ is dependent upon $\Delta\nu$, the line width of the spontaneous emission and the energy of the laser radiation. The various theoretical developments have differed somewhat, but the form of the result given in Eq. (8) has been verified by experiments[44] in which the bulk loss $\bar{\alpha}$ for relatively lightly doped materials was small and the threshold was shown to be proportional to the inverse length of the diode. This statement also suggests that the electromagnetic mode pattern is determined by the relative dielectric properties of the inverted region and the extrinsic regions near the junction. This conclusion follows from the equations for the electromagnetic problem for the traveling or guided wave along the inverted region where the value of the spread on either

Fig. 2. The energy bands present in the neighborhood of a p-n junction of a diode laser, with the applied voltage (A) V = zero and (B) V > $\mathcal{E}_g$/e. In this degenerate case, the Fermi level $\varphi$ is below the bottom of the valence band and above the top of the conduction band. (From Rediker[1])



side is given approximately by

$$\delta_2 \approx \frac{\lambda^2}{W}\,\frac{\epsilon}{\Delta\epsilon}\,\frac{1}{(2\pi)^2} \qquad (9)$$

where $\lambda$ is the wavelength in the medium and $\Delta\epsilon$ is the difference in dielectric constant. The spread may be due to free carriers or even the difference in the interband contribution near the gap, which is altered by the occupation level in the degenerate regions and the inverted region. The dispersive part of the latter can also be a contributing factor. The amazing part of the above result is that for $\lambda = 2.3 \times 10^{-5}$ cm, $W \approx 10^{-4}$ cm, and $\delta_2 \approx 10^{-3}$ cm, the value of $\Delta\epsilon/\epsilon \approx 10^{-4}$. This requirement can be easily satisfied by any one of the mechanisms above. But even before the inversion sets in, the electromagnetic field can be guided since $\Delta\epsilon > 0$ is readily attained.

The numerical values obtained from the various theoretical estimates for the threshold current yield values of $J_{\text{th}}$ of about 100 to 1000 A/cm² in semiquantitative agreement with the experimental values. In addition, the value of the vertical angle of the laser beam from $\theta \approx \lambda_0/\delta_2 \approx 0.84/10 \approx 0.08$ radian $\approx 5°$ is in agreement with the estimated value of $\delta_2$. The values (in microns) of $\lambda_0$ and $\delta_2$ are the free-space wavelengths and skin depth of the degenerate region, respectively, for the GaAs diode injection laser.

The threshold current has also been examined theoretically[45] and experimentally[46] by the IBM group. They found that the values remained at about 100 A/cm² from about 2°K to 20°K, and then increased with the relation $J_{\text{th}} \sim T^3$ beyond this temperature, as shown in Fig. 3. The theory developed by Lasher and Stern[45] accounts for this effect in terms of the reabsorption for band-to-band transition, as the temperature increase allows for empty states in the conduction and valence bands as the deviation from degeneracy increases. The temperature dependence of $T^{2.6}$ at higher temperature is reasonably close to the experimental situation. Their theoretical calculation also accounts for the line shape of the spontaneous emission of the radiation observed in the diodes below threshold.

### Recent developments

The original semiconductor diode lasers have been fabricated from III–V compounds such as GaAs, InAs, InP, and InSb, and their alloys. All of these materials are, of course, direct-gap semiconductors. Another well-known and well-investigated class of direct-gap semiconductors is the IV–VI family. These compounds include lead chalcogenides, such as PbS, PbTe, and PbSe. Indeed, with the advancement of materials preparation, the latter two now have been fabricated into diode lasers. These differ from the III–V compounds in that the optical transitions occur at the [111] edge ($L$ point) of the Brillouin zone rather than at the center $k = 0$ usually found in the other compounds. The important aspect of the development of the PbTe and PbSe lasers is that they have extended the range of semiconductor lasers to longer wavelengths of 6.5 and 8.5 microns, respectively. Furthermore, it appears from recent experiments that uniaxial and hydrostatic pressures allow tuning of these lasers further into an 8- to 14-micron "window" of the atmosphere, where useful applications for communications and military purposes are possible.

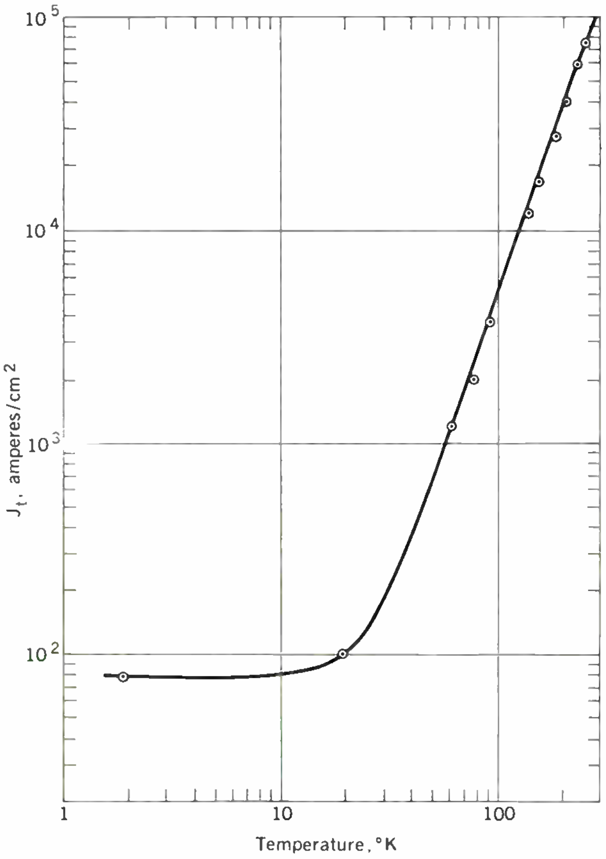
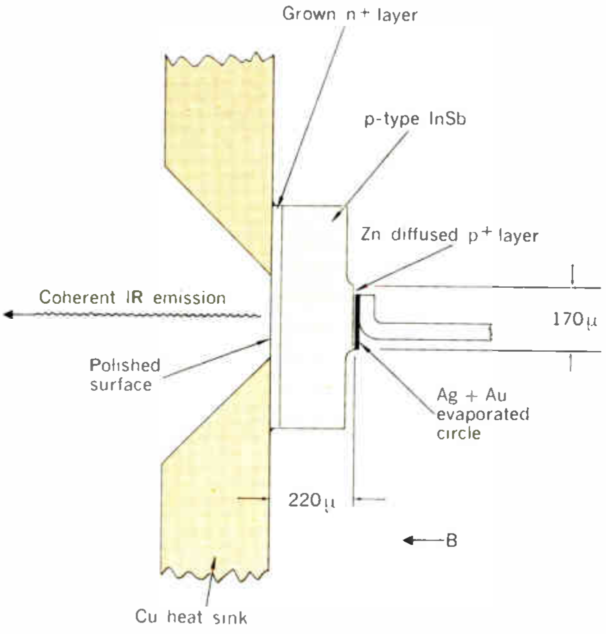


Fig. 3. Temperature dependence of the threshold current required for laser action in GaAs. (From Burns, Dill, and Nathan[46])

Fig. 4. Longitudinal injection plasma laser with the Fabry–Perot cavity parallel to the current. (From Melngailis, Phelan, and Rediker[26])

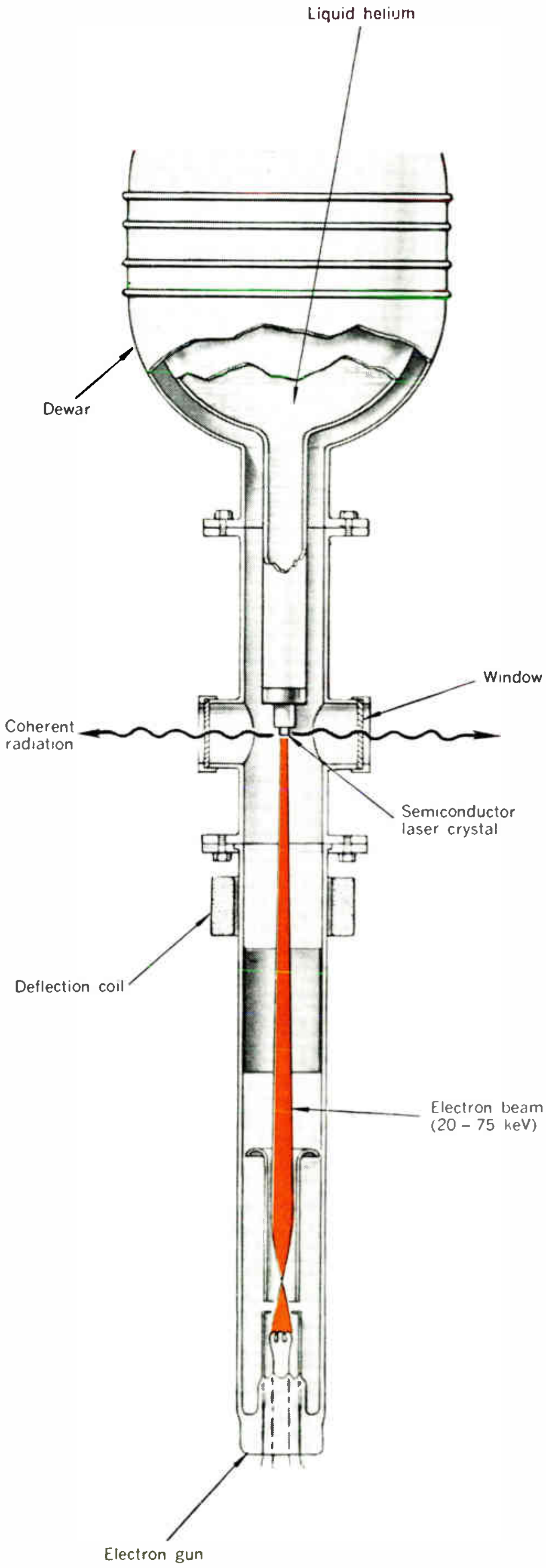


In the junction lasers the active inverted region, which is responsible for stimulated emission, is estimated to be about one micron in depth along the direction of the injecting current. The luminescent region, determined from the beam pattern, is somewhat wider and may be of the order of about 10 microns in this same direction. Recently, Melngailis, Phelan, and Rediker developed a "large-volume" injection laser in InSb in which the inverted region extends several hundred microns beyond the junction region. This innovation to the diode laser technology originates in this group's earlier work on large-volume injection-plasma diodes[17] in forward-biased $n^+$-p-$p^+$ InSb structures used for making magnetic switching and amplification devices, which they called madistors. These diode structures consist of a heavily doped n region and a lightly doped p region followed by heavily doped layers. It is the middle layer that is responsible for the unique properties of this diode. On forward bias the electrons from the $n^+$ region are injected in the p region and saturate the traps, thereby increasing the carrier lifetime from $10^{-10}$ to about $10^{-7}$ or $10^{-6}$ second. It follows that the diffusion length $L_n = \sqrt{D_n \tau}$ goes from about one micron to approximately 100 microns. The application of a magnetic field along the direction of the current increases the efficiency of emission, as indicated by the pattern observed with a mechanical scanner on an infrared lens and an InSb photovoltaic detector.

An important aspect of the large-volume laser is that for the first time a semiconductor laser has been made to produce a growing coherent wave along the direction of the current. This device, developed by Melngailis, has been called the longitudinal injection-plasma laser. Its construction, which is quite different from the usual diode laser, is shown in Fig. 4. The laser is mounted on a copper heat sink, which contains a small hole. The laser cavity now consists of an optically flat, polished surface on the $n^+$ side. The $p^+$ side is polished and covered with evaporated layers of silver and gold so that it will be totally reflecting. The layers match the circular hole so that the transverse dimension of the inverted layer is smaller than the longitudinal dimension. With no transverse reflection and losses in the noninverted adjacent regions, the gain in the longitudinal direction exceeds that of the transverse and is sufficient to produce laser action in the forward direction. An axial magnetic field also confines the injected plasma by reducing the transverse diffusion, and thereby limits the transverse dimensions of the active inverted region.

The technique of electron-beam pumping is one of the most recent developments. The principle of the electron-beam semiconductor laser is quite simple. A high-energy beam of the order of 20 to 100 keV is focused on a small semiconductor sample whose dimensions may be of the order of several hundred microns. Then, as before, two surfaces are either cleaved or polished parallel to form a Fabry–Perot resonator perpendicular to the beam. As the electron beam penetrates the bottom layer to a 5- to 10-micron depth, it creates electron–hole pairs which, at low temperatures, form degenerate populations as in the case of the injection diode. Upon the formation of this inverted population, the spontaneous emission builds up until, above a threshold current of the electron beam, stimulated coherent emission perpendicular to the electron beam emerges from the polished or cleaved surfaces of the resonant structure. The first such attempt in the

Fig. 5. Electron gun and Dewar vessel for electron-beam semiconductor laser. (From Rediker[1])



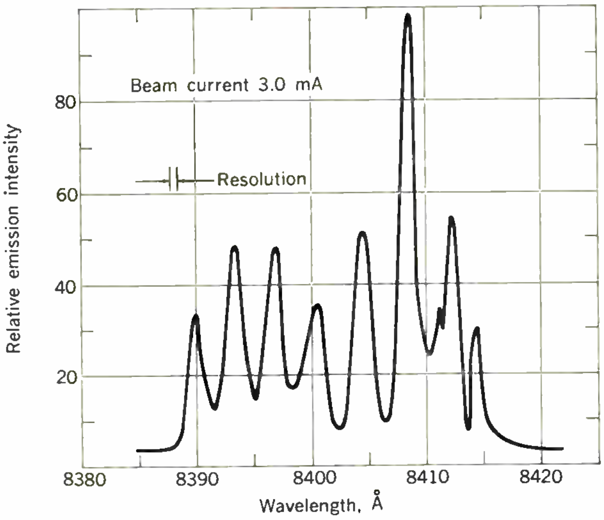


Fig. 6 (above). Laser spectra emitted from electron-beam-pumped GaAs. (From Hurwitz and Keyes[30])

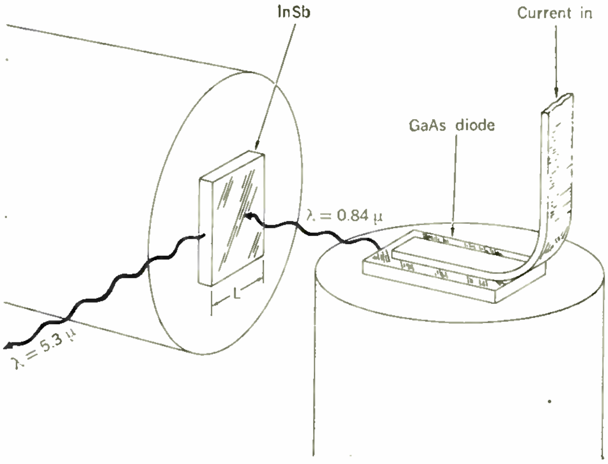


Fig. 7 (above right). Bulk InSb laser pumped optically by use of a GaAs diode injection laser. (From Phelan and Rediker[34])

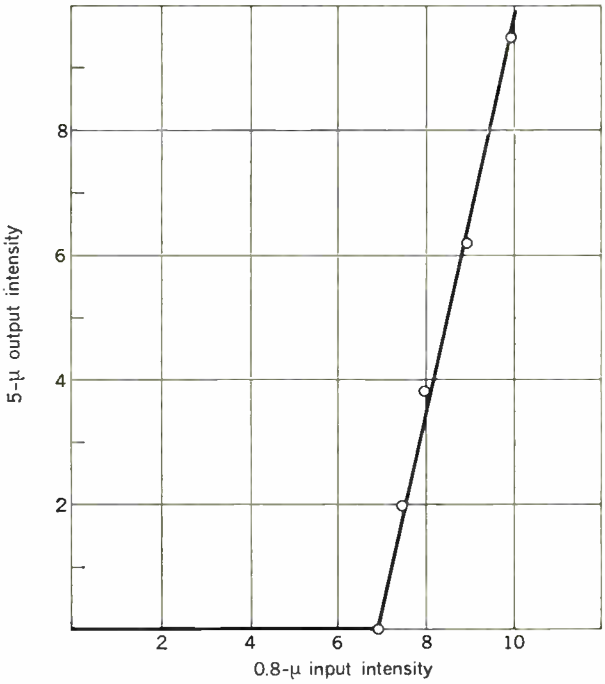


Fig. 8 (right). Relative InSb emission vs. the pumping GaAs emission. Pulses of 50-ns duration were used. On the abscissa, 10 corresponds to 20-ampere GaAs diode current and about 7000 W/cm² incident upon the InSb. (From Phelan and Rediker[34])

Soviet Union on CdS employed an electron beam of the order of 200 keV and a current density of about 1 A/cm². Although laser action was not achieved, the fact that line narrowing was observed indicated the feasibility of this technique for laser action. Shortly thereafter, Benoit à la Guillaume and Debever achieved successful laser operation in InSb and InAs with a 20-keV electron beam and threshold currents of the order of 1 mA for a beam diameter of about 150 microns. This corresponded to a current density of about 1 A/cm² or an electron concentration of $10^{17}\ \text{cm}^{-3}$ comparable to the inverted population in the junction laser. To avoid heating effects, the electron beam was pulsed for a few microseconds with a low duty cycle. It has been estimated that efficiencies of pair excitation of about 30 per cent have been achieved above threshold.

Electron-beam techniques have also been successful for GaAs, in both n and p types, with impurity concentrations of the order of $10^{18}/\text{cm}^3$. In GaAs it was found that, in addition to a threshold current, a threshold voltage of about 30 keV was necessary in order to achieve laser action. The parallel, polished faces were approximately 200 microns apart. Pulses of 0.2-μs duration at a repetition rate of 1000 per second were used. The arrangement of the apparatus is shown in Fig. 5. When the spectral properties of the emission were examined below threshold with a beam current of about ½ mA, the line width was found to be approximately 100 Å. Above threshold, however, at a beam current of 3 mA, many modes were observed, as shown in Fig. 6, in which the line width of the individual modes was approximately 2 Å with a characteristic spacing between modes of about 4 Å. The latter is in very good agreement with the theoretical value calculated for a Fabry–Perot resonator of this dimension. In addition to GaAs, GaSb has also been made to operate by the electron-beam technique by the French group. Very recently, Benoit à la Guillaume and Debever were able to excite tellurium with a 15-keV electron beam and they observed emission of two sharp peaks at 4°K of 0.334 eV and 0.336 eV, at a threshold current of about 200 μA. In this material, the emission was polarized with the electric vector parallel to the *c* axis, indicating a direct transition for this polarization. Finally, Hurwitz has achieved laser action by electron-beam techniques[48] in PbTe and PbSe at their characteristic wavelengths, with results similar to those obtained in the diode structures.

Although one of the earliest suggestions for pumping semiconductor lasers was by means of optical sources, this method was not attempted until recently. Actually, the writer has suggested, in connection with the cyclotron resonance laser, that a gas or diode laser beam focused to a spot of 100 microns or less with a power intensity of ap-

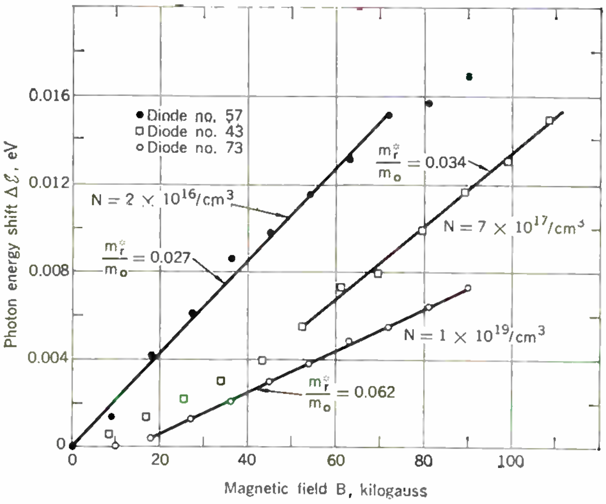


Fig. 9. The shift of the spontaneous emission line with magnetic field, for three diodes made from different n-type materials and operated near 77°K. Effective mass values are computed from the slope of the straight lines, assuming that $\Delta\mathcal{E} = \frac{1}{2}\hbar\omega_r$. Zero field photon energies at 77°K are 0.395, 0.393, and 0.381 eV for diodes 57, 43, and 73 respectively. (After Galeener et al.[50])

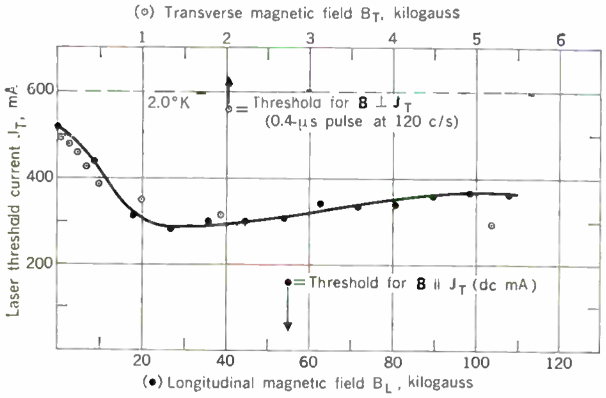


Fig. 10. Laser threshold current versus magnetic field for a CW laser diode immersed in pumped liquid helium and oriented first parallel and then perpendicular to the magnetic field. (After Galeener et al.[50])

proximately 10 watts would excite electron–hole pairs in such a way that the effect would be equivalent to that of the injection or electron-beam techniques for excitation. Indeed, by utilizing a GaAs laser in juxtaposition to a small slab of InSb as shown in Fig. 7, which was prepared as a resonant structure, the coherent radiation at 5.3 microns was emitted by the InSb, which was near 4°K. The InSb was relatively pure n-type material—that is, the impurity concentration was only about $10^{14}/cm^3$. When the energy balance was taken into account, with the appropriate corrections for reflection, the threshold corresponded to about 1000 A/cm$^2$ as one would expect of an equivalent junction diode. An interesting curve showing the relation between the output of the InSb and the input of the GaAs diode is shown in Fig. 8. Above threshold, output intensity proceeds linearly with the input energy.

### Magnetic effects

A variety of magnetic phenomena have been investigated in connection with semiconductor lasers. One of the earliest observations was the behavior of the emission from GaAs diode lasers with high magnetic fields. It was observed that the frequency increased with the square of the applied magnetic field intensity corresponding to a quadratic Zeeman effect[49] of a bound donor electron or exciton near the bottom of the conduction band. It was therefore concluded that the initial state was either of these two and that the final state had to be an acceptor state or acceptor band above the normal valence band, since the energy of the laser emission at 1.47 eV was lower than that of the intrinsic energy gap at 4°K by approximately 0.04 eV. Similar studies of emission from InAs diodes and lasers[50] were also made by Galeener and co-workers at high magnetic fields. One of the interesting observations concerning the diode emissions was that the behavior with magnetic field intensity in this small gap material was nearly linear, but the slope of the line was strongly dependent on the concentration of the doping. If one then used the model corresponding to the motion of a free carrier or an interband transition between free-carrier (Landau) states, then $\Delta\mathcal{E}_g = \frac{1}{2}\hbar(\omega_c + \omega_r)$, where $\omega_c = eH/m_c{}^*c$ and $\omega_r = eH/m_r{}^*c$ are the cyclotron frequencies for the electrons and holes. Different effective masses, corresponding to the reduced mass $m_r{}^*$, are shown in Fig. 9. The interpretation of this phenomenon is that the effective mass corresponds to that associated with the nonparabolic curvature of the energy bands in which a degenerate distribution places the Fermi level well into the band in the region of higher effective mass.

The threshold for the laser as a function of magnetic field was also studied in InAs and exhibited a behavior as indicated in Fig. 10. The curve shows the variation of threshold for the magnetic field parallel to the current. At lower magnetic fields, the threshold is reduced by a factor of two and then it slowly increases with magnetic field intensity up to about 100 kilogauss. This phenomenon is not well understood, since on a simple model one would expect the threshold to decrease monotonically. The theory does not take into account, however, the effect of magnetic field on scattering and recombination, which may then more properly explain the behavior observed. With the magnetic field perpendicular, the threshold is quickly reduced by a few thousand gauss and remains flat as the field is increased further. In this configuration, the interpretation is that the magnetic field influences the diffusion length, reducing it and thereby initially increasing the electron–hole pair concentration in the inverted region. Ultimately at sufficiently high fields near 10 kGs, the laser does not operate at all because the transverse magnetic field reduces the diffusion of carriers into the inverted region, and eventually cuts off the diodes. The diffusion coefficient is reduced according to the relation $D_{//} \approx D_0/\omega_c\tau$.

In InSb, the effect of magnetic field is even more striking. The threshold is considerably reduced, as expected from theory. In the particular diode shown in Fig. 11, the threshold is a factor of five lower beyond about 15 kGs. At low magnetic fields, the diode does not operate at all. This phenomenon is qualitatively explained by the fact that the density of states increases with magnetic field and thereby the transition probability for emission is proportionately increased. Nevertheless, at very high fields (of the order of 100 kGs) the threshold flattens out, possibly for the same reasons mentioned for the InAs laser. It is therefore apparent that the theory of this phenomenon is not completely understood.

Another striking aspect of laser emission in InSb is that the frequency shift as a function of magnetic field intensity is rather high, as one would expect from the theory. The frequency changes approximately 8 per cent as a field of 100 kGs is applied, as indicated in Fig. 12. At low magnetic fields the diode emission frequency exhibits a quadratic dependence on magnetic field indicative of a donor or exciton state. At high fields, however, the dependence is linear, with the interesting phenomenon that initially the laser operates at a higher frequency and then a lower frequency transition takes over. The splitting of these two states in InSb corresponds to the anomalous spin splitting of the conduction-band states observed in magnetoabsorption experiments. The theory indicates that the upper transition has a probability which is a factor of three higher than that of the lower one. Hence, at lower fields of 20 to 30 kGs, where both levels are populated, the upper transition level would be expected to dominate. As the magnetic field is increased, however, the population factor begins to favor the lower one so that at very high fields, only this level is populated. A similar phenomenon has been observed in PbTe and PbSe, where the pattern in both is quite similar and shows a characteristic as indicated in Fig. 13. This time, however, both the lower and upper spin states operate simultaneously, inasmuch as the transition probability for both is the same. As the magnetic field increases beyond 20 kGs, however, only the lower level continues to emit, since the higher one is no longer populated.

One other magnetic effect in connection with these lasers is that associated with the dispersive properties of interband transitions. It is known, from magnetooptical studies utilizing Faraday rotation and magnetoreflection, that the dielectric constant near the gap is strongly affected by the magnetic field. Thus, in addition to increasing the energy gap and changing the frequency of the laser, the spacing between the Fabry–Perot modes is strongly affected by the magnetic field. This pattern has been observed both in InSb and in InAs, and appears as shown in Fig. 14. As the magnetic field is increased, the laser operation jumps from one mode to the next higher one. At the same time, each mode shifts in absolute frequency to a shorter wavelength, since the dielectric constant increases because of the magnetic field.

One of the first conceptions for semiconductor lasers was a four-level system that makes use of magnetic fields.[51] It was proposed that an interband transition combined with an intraband transition between magnetic levels would produce coherent radiation in the far infrared region. The basic principle is illustrated in Fig. 15, wherein a strong source of narrow-band radiation excites electrons from magnetic level 1 in the valence band to magnetic level 1 in the conduction band. This serves to invert level 1 relative to level 0 in the conduction band, and at sufficiently high excitation there will be stimulated emission between level 1 and level 0 in the conduction band, provided that there is a resonant structure at this wavelength. This structure should be relatively easy to provide. Simultaneously there should be stimulated emission across the gap from level 0 in the conduction band to level 0 in the valence band. According to theory, this transition has a greater probability than the former, thereby preventing a bottleneck or the piling up of an electron population at the bottom of the conduction band.

Another requirement for laser action in a cyclotron resonance device is that the level spacing between 1 and 2 be different. Otherwise, stimulated transitions upward would be possible, resulting in absorption and loss and thereby preventing laser action from taking place. Fortunately, this unequal level spacing exists in InSb, in

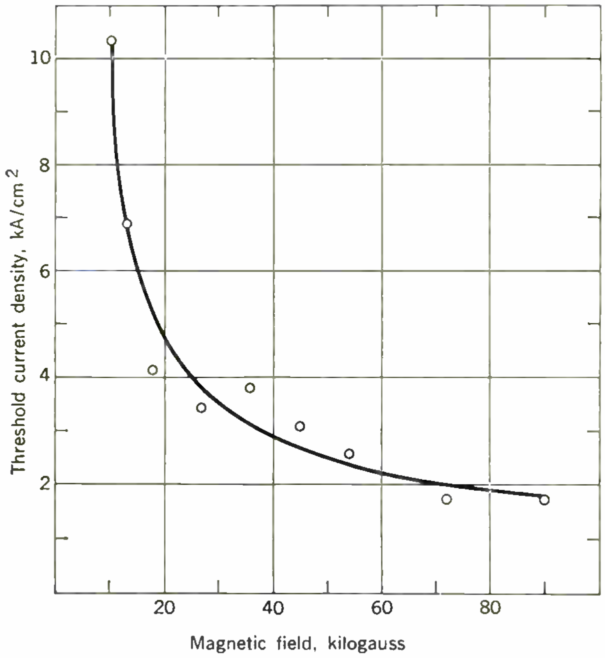
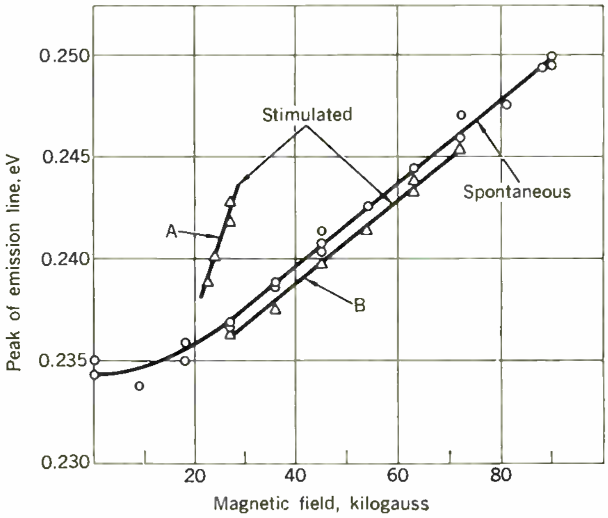


Fig. 11. Threshold current density for laser action as a function of magnetic field intensity. (From Phelan et al.[22])

Fig. 12. The energy of laser emission spectra as a function of magnetic field intensity. (From Phelan et al.[22])



which the conduction band is nonparabolic. An additional feature fulfilled in InSb is that the levels in the valence band corresponding to those labeled 1 and 0 are very closely spaced, and therefore the relaxation time between them is very short, thus providing the ideal pattern for a four-level laser. An appropriate pumping source for this cyclotron resonance laser would be either a helium–neon laser operating in its strong emission mode at 3.39 microns, or an InAs laser, which operates at 3.1 microns.

The transition just described would require magnetic fields in excess of 100 kGs, which are now readily available in either dc or pulsed form. Furthermore, we calculate that a pumping power level of approximately 10 watts, either CW or pulsed, focused into a 30-micron spot on the InSb surface, would produce $10^{17}$ to $10^{18}$ carriers per cubic centimeter, a concentration that is comparable to that obtained in electrical excitation. The calculated gain to be obtained at the cyclotron resonance transition for an emission wavelength of 20 microns would be more than sufficient (neglecting bulk loss) by several orders of magnitude to overcome the reflection loss at the surface of the Fabry–Perot resonator. It is found that $\alpha l > 1$ under this form of excitation, where $\alpha$ is the gain coefficient and $l = 3 \times 10^{-3}$ cm is the size of the cavity. Consequently, the only other loss that has to be considered is the free-carrier absorption resulting from the nonresonant upward transitions. It can be shown that the ratio of the gain coefficient to the loss coefficient

$$\frac{\alpha_g}{\alpha_L} = (\Delta\omega)^2\tau^2 \qquad (10)$$

where $\Delta\omega$ is the difference between the lower and upper cyclotron resonance levels, which in InSb is approximately $6 \times 10^{12}$. If we take $\tau \approx 10^{-12}$ second from experimental values, we find that the criterion for the gain to overcome the loss is well justified.

P. A. Wolff[52] has made a proposal for an analogous cyclotron resonance laser at longer wavelengths and lower fields of the order of 10 to 20 kGs in which he attains inversion by optical excitation of many levels above the lowest one. He expects rapid cascading downward to level 1 due to optical phonon transitions whose relaxation time he estimates to be $10^{-13}$ second. With these smaller magnetic fields, the Landau levels in the conduction band will be less widely separated, and thus the emission will be in the 100-micron wavelength range. His estimates are probably not accurate, since he ignored the effect of magnetic field on the relaxation phenomena due to either acoustical or optical phonon scattering. The latter, in particular, would show resonant behavior in a magnetic field, as predicted by Gurevich and Firsov.[53] Moreover, off resonance—where one would operate either the high-field or low-field laser—the scattering due to optical phonons is not as large as Wolff anticipates in cases where he requires large scattering to produce inversion. Finally, the acoustical phonon scattering is an order of magnitude larger than he predicts because in pure material $\tau \propto H^{-1}$ in the quantum limit.[54]

The scheme proposed here, in which we would use a monochromatic optical excitation of perhaps 100 watts peak power, would permit the cyclotron resonance laser to operate equally well in the far-infrared range using smaller magnetic fields. This pump threshold is somewhat higher than that predicted by Wolff, and the physical configuration prescribed here is, we believe, more favorable and simpler.

Fig. 13. Energy of laser emission from PbSe as a function of magnetic field intensity showing the effect of anomalous spin splitting of the conduction band states. (Courtesy of J. F. Butler)

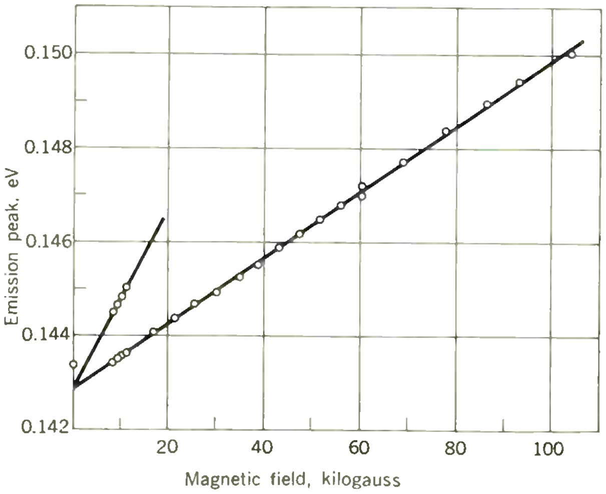
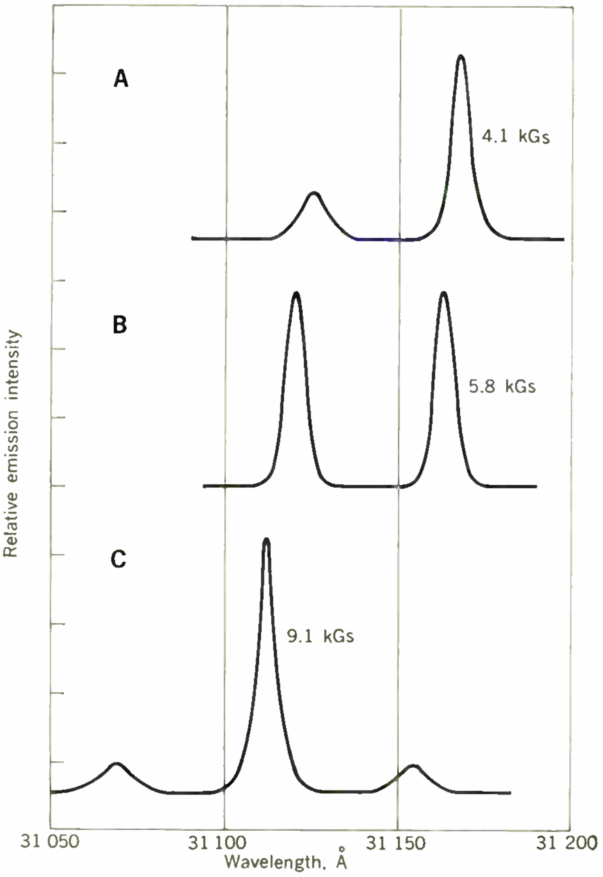


Fig. 14. Change in preferred mode of propagation and small shift of frequency as the magnetic field changes the dielectric constant in InAs diode. (From Melngailis and Rediker, Appl. Phys. Letters, vol. 2, 1963, pp. 202–204)



## Applications

One of the promising possibilities heralded for the semiconductor laser was for application in the communication field. The semiconductor laser is easily modulated electrically at modulation frequencies as high as the gigacycle range. Television signals have been transmitted over infrared beams from diodes and diode lasers of GaAs. Of course the objection to the infrared laser for transmission stems from its atmospheric absorption properties, which make it a fair-weather device. We may consider the possibility, however, of transmitting the radiation through tiny hollow waveguides, several microns in diameter, operating in the low-loss $TE_{01}$ mode. With laser diodes as repeaters, such an arrangement may provide the basis for a large-capacity communication cable.

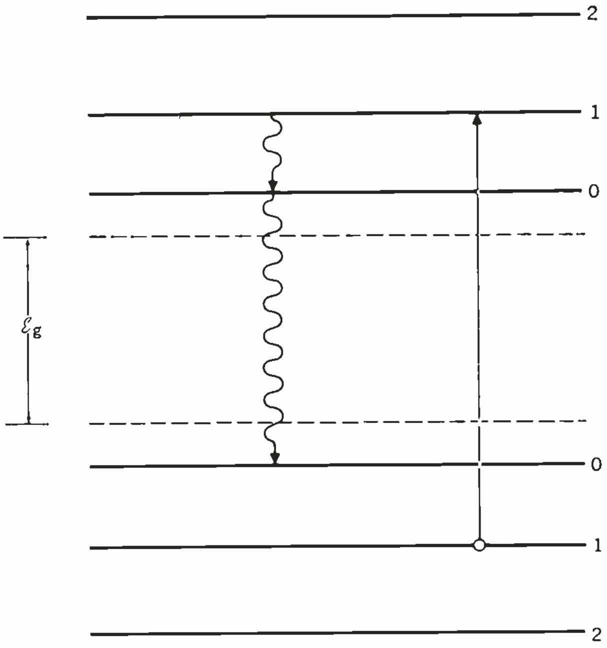
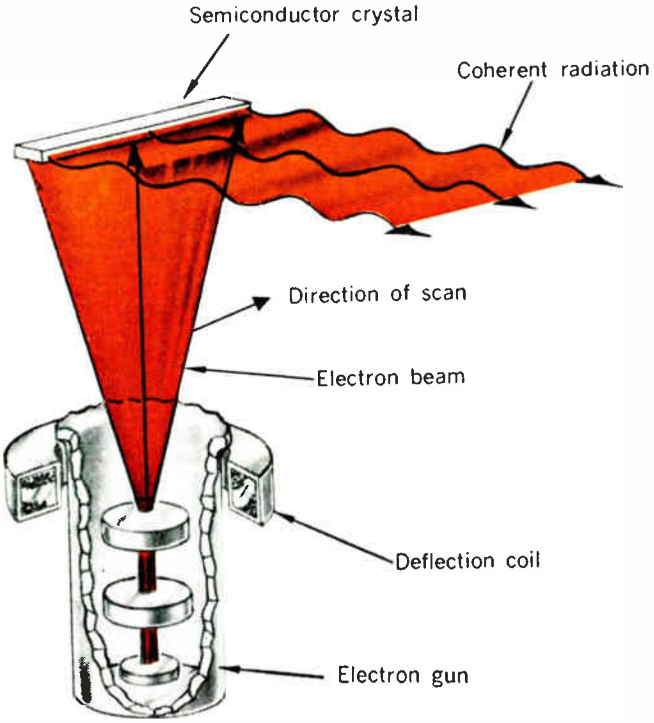


Fig. 15. Discrete energy levels created by a magnetic field for the four-level cyclotron resonance laser.

Fig. 16. Artist's sketch of elementary scanning beam laser. (From Lax[40])



The GaAs laser also has been developed as a compact transmitter for radar equipment. These diodes have been operated at room temperature with very narrow pulses of the order of 50 ns or less with output power as high as 100 watts peak. Such compact radars are capable of spatial resolution of several feet. The angular resolution with suitable optics should provide resolution of the order of milliradians or better.

The semiconductor laser has been adapted to a beam-of-light transistor in which a GaAs laser is directed at a diode collector with a somewhat smaller gap. The combination can provide current gain. In principle, such a combination can be extended to the microwave range, since the laser diodes and the collector can respond to these high frequencies and the limitation of the transit time across the base region is eliminated.

With the advent of the electron-beam pumped laser, other applications are also possible. A whole class of these is within reach when the scanatron, or scanning beam laser, is fully developed. The writer has proposed[40] two schemes for realizing this device. The first involves the adaptation of the electron-beam technique for exciting a bar of semiconducting material several centimeters in length and perhaps 100 to 200 microns wide, with polished, parallel faces, to form a Fabry–Perot resonator. The beam is then scanned across the bar and the laser light is emitted at right angles, as shown in Fig. 16, and moves with the electron beam, which can be controlled as in a cathode-ray tube. If many of these bars are arranged in a raster, the laser beam can be scanned in two dimensions, as in a television or cathode-ray tube. Through modulation of the beam, a display device or projector can be fabricated, as shown in Fig. 17. The device can also be achieved by causing the laser to emit in the forward direction. This is accomplished by polishing sheets of semiconducting material, about 5 to 10 microns thick, and coating the sheets with dielectric layers or silver and then mounting them on sapphire, in the arrangement shown in Fig. 18. The forward laser action is favored because the dielectric or silvered layers provide feedback and greater gain in this direction than in the transverse direction, in which the 100-micron width of the beam has insufficient gain for laser action. When perfected, these devices may be used for film readers, displays, page reproduction at remote distances, microscopy, a CW source with several hundred watts of output, and other applications. The feasibility of scanning a beam has already been demonstrated by Rotstein, Zieman, McNamara, and Lax in GaAs.[55] The basic principle is now established, but many technological problems remain to be solved before a practical device is realized.

The last application that will be discussed is one in which the GaAs laser has been employed as a source for studying the Raman effect in the infrared region by Chantry, Gebbie, and Hilsum.[38] This is an ingenious scheme in which the sensitivity of the technique is increased by incorporating a Michelson interferometer and the related Fourier-transform data technique for detecting

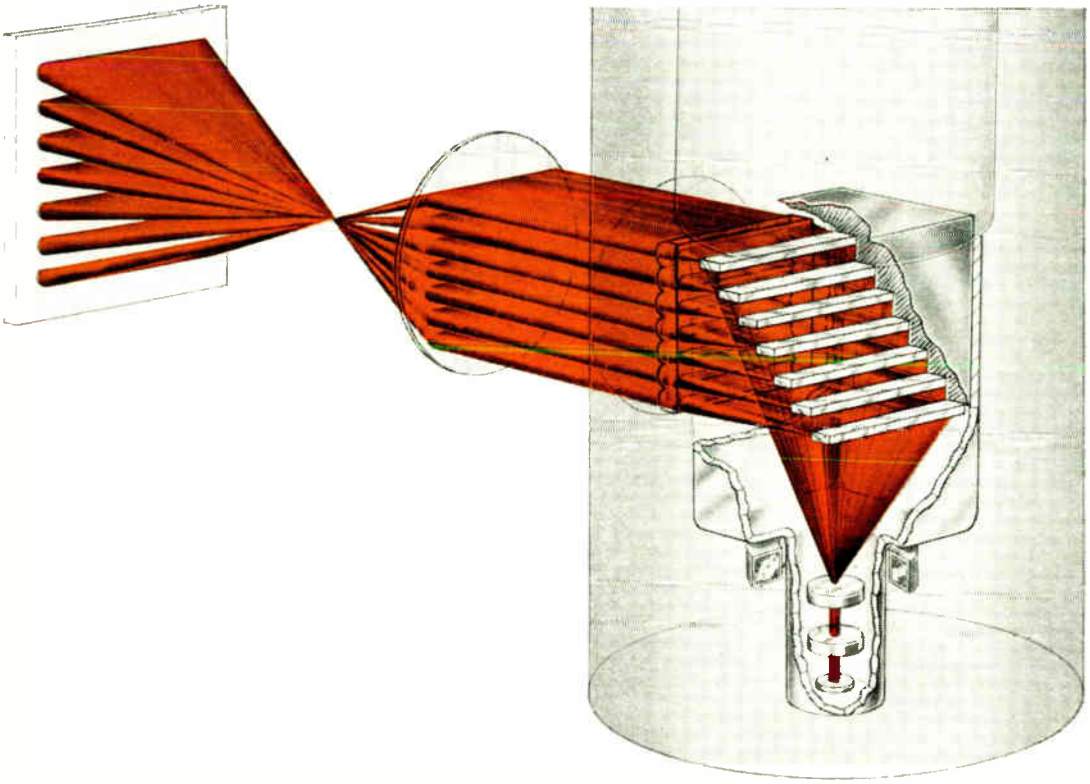
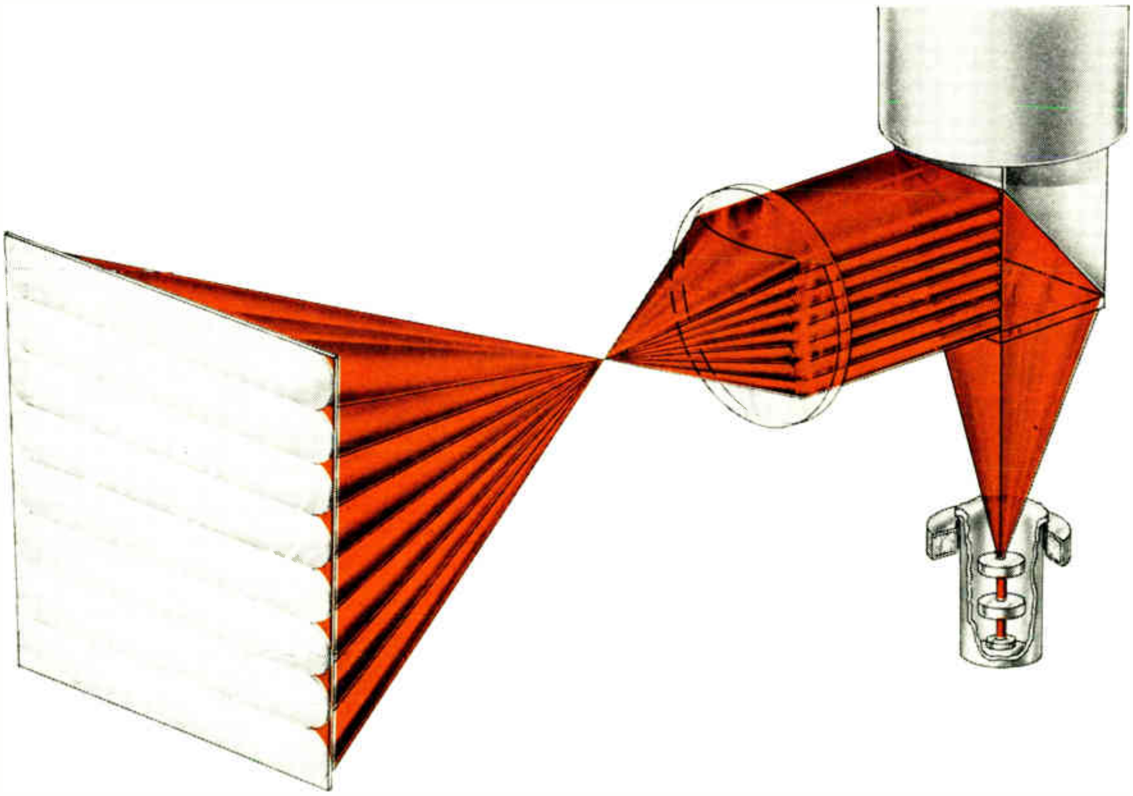


Fig. 17. Sketch of a scanatron laser device showing the multiple array of flat, polished semiconductor slabs and also lens arrangement for projecting image onto a screen. (From Lax[40])

Fig. 18. Scanatron laser device in which a forward emitting laser is employed.



the spectra in the near-infrared. The experimental arrangement is shown in Fig. 19. This technique has been successfully used to study the Raman effect in iodine. With other semiconductor lasers and increased power at longer wavelengths, it appears that Raman spectroscopy and nonlinear phenomena will be possible at still longer wavelengths in the infrared.

## Conclusion

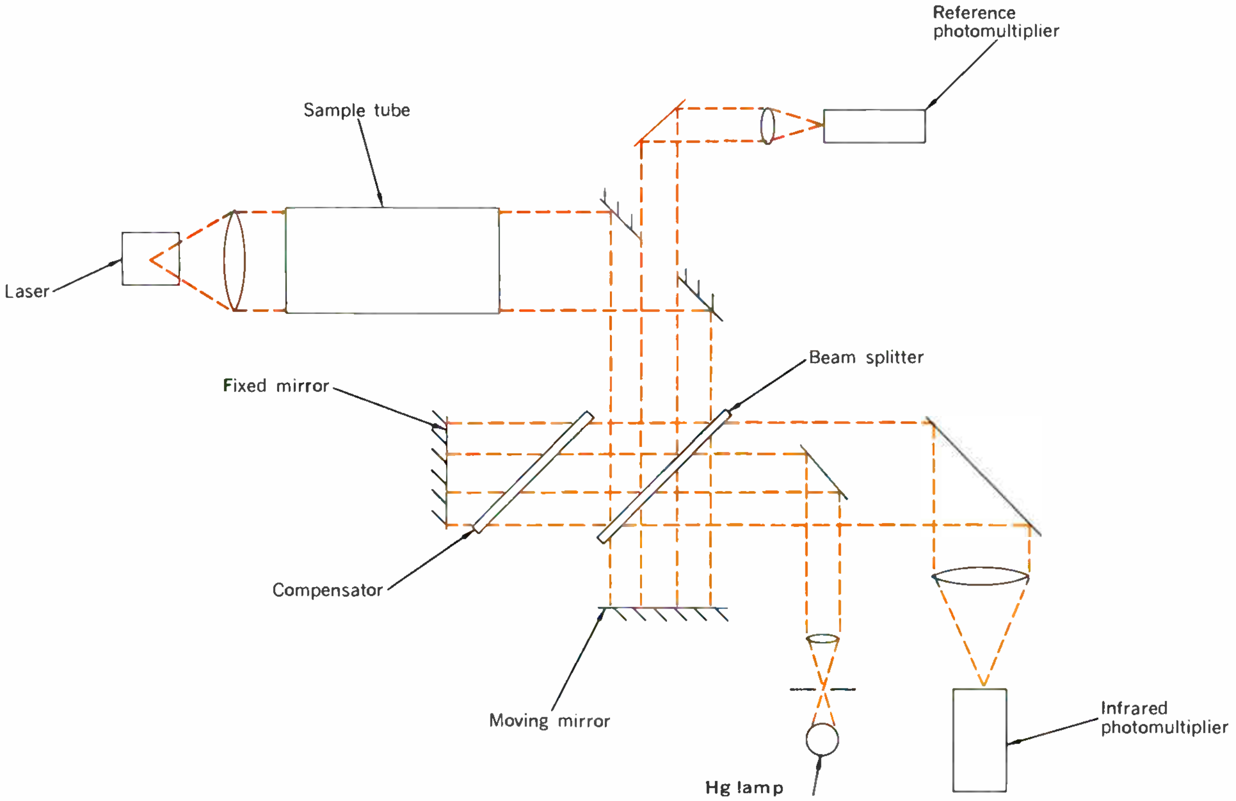
This article has represented an attempt to summarize the recent developments in the rapidly progressing field of semiconductor lasers. Within a short span of 2½ years about a dozen new materials have emerged as effective media for generating coherent radiation in the infrared from about 0.6 to 8.5 microns. If we consider the successful generation of harmonics, the range extends down to 0.32 micron, which increases the total range to more than one octave. Levels of several watts of continuous power and about a kilowatt of peak power in short pulses have been attained.

Many problems still remain to be solved in the theoretical, the experimental, and the technological areas. The basic mechanisms for guiding the waves are not fully understood. The detailed natures of the transitions involved in different lasers appear to vary and have to be identified in each material, depending upon the behavior of the laser with respect to doping level, magnetic fields, pressure, and temperature. Polarization phenomena have not been adequately explored. More realistic solutions of the electromagnetic problem may be possible if perturbation techniques are used to take into account the variation of excitation with penetration or injection depth, inhomogeneity of doping near the junction, etc. Exploration of the mechanism of excitation by electron beams and optical pumping will probably result in improved techniques.

On the applied side we have only begun to make progress. Semiconductor lasers will undoubtedly be extended further into the far-infrared region by the use of lower-gap material as well as by intraband phenomena, such as the cyclotron resonance laser. Optical lasers will be possible when better or properly controlled materials are developed in the high-gap region. These will then lead to many important applications in the field of photography and displays. Holography will then become more convenient and useful if compact and less expensive semiconductor lasers become available in the optical region. Of course, the real breakthrough we are seeking is to make such lasers more efficient at room temperature and, at the same time, to extend them to the optical and ultraviolet regions. The latter would, of course, offer interest to the chemist for the excitation of photochemical reactions.

The unique role to be played by lasers and, in particular, the semiconductor lasers, is beginning to crystallize. It is hard to predict the important and economically significant developments now, but in these very early years of this infant device it would be inappropriate to judge the field either pessimistically or overoptimistically. Those laboratories that have the talent, resources, and interest should continue to push the science and technology vigorously for the next few years or even for a decade until the basic questions are resolved. This is an exciting frontier in applied physics and technology, which is less expensive than space, nuclear physics, or astronomy. Nevertheless, the same questions of relevance, utility, prospects of scientific accomplishment, and economic worth are involved. As part of our scientific culture, we should apply the same yardstick to quantum electronics, and to lasers in particular, as to these other

Fig. 19. Interferometric spectrometer using a laser source. (From Chantry, Gebbie, and Hilsum[38])



fields. One should not adopt a short-sighted attitude by expecting the quick commercial returns that we had from transistors, computers, and other solid-state devices.

The author wishes to express his appreciation to K. J. Button for his assistance in the completion of this article, and to A. L. McWhorter and R. H. Rediker for valuable discussions on the subjects covered.

## REFERENCES

1. Rediker, R. H., "Infrared and Visible Light Emission from Forward-Biased p-n Junctions," *Solid State Design*, vol. 4, 1963, pp. 3–12; "Semiconductor Lasers," *Phys. Today*, vol. 18, 1965, pp. 42–54.

2. Lax, B., "Semiconductor Diode Lasers," *Solid State Design*, vol. 4, 1963, pp. 26–32.

3. Nathan, M. I., and Burns, G., "Injection Lasers: State of the Art," *Electronics*, vol. 36, 1963, pp. 61–65.

4. Lax, B., "Semiconductor Lasers," *Science*, vol. 141, 1963, pp. 1247–1255.

5. Burns, G., and Nathan, M. I., "*P-N* Junction Lasers," *Proc. IEEE*, vol. 52, July 1964, pp. 770–794.

6. Gordon, J. P., Zeiger, H. J., and Townes, C. H., "Molecular Microwave Oscillator and New Hyperfine Structure in the Microwave Spectrum of $NH_3$," *Phys. Rev.*, vol. 95, 1954, pp. 282–284.

7. Von Neuman, J., Unpublished manuscript.

8. Aigrain, P., International Conference on Solid State Physics and Telecommunications, Brussels, Belgium, 1958.

9. Lax, B., "Cyclotron Resonance and Impurity Levels in Semiconductors," in *Quantum Electronics*, C. H. Townes, ed. New York: Columbia University Press, 1960, pp. 428–499; "Cyclotron Resonance Maser," in *Advances in Quantum Electronics*, J. R. Singer, ed. New York: Columbia University Press, 1961, pp. 465–479.

10. Basov, N. B., Krokhin, O., and Popov, J. M., "Negative Absorption Coefficient at Indirect Transitions in Semiconductors," in *Advances in Quantum Electronics, op. cit.*, pp. 496–506.

11. Bernard, M. G. A., and Duraffourg, G., "Laser Conditions in Semiconductors," *Physica Status Solidi*, vol. 1, 1961, pp. 699–703.

12. Keyes, R. J., and Quist, T. M., "Recombination Radiation Emitted by Gallium Arsenide," *Proc. IRE* (*Correspondence*), vol. 50, Aug. 1962, pp. 1822–1823.

13. Dumke, W. P., "Interband Transitions and Maser Action," *Phys. Rev.*, vol. 127, 1962, pp. 1559–1563.

14. Hall, R. N., *et al.*, "Coherent Light Emission from GaAs Junctions," *Phys. Rev. Letters*, vol. 9, 1962, pp. 366–378.

15. Nathan, M. I., *et al.*, "Stimulated Emission of Radiation from GaAs p-n Junctions," *Appl. Phys. Letters*, vol. 1, 1962, pp. 62–64.

16. Quist, T. M., *et al.*, "Semiconductor Maser of GaAs," *Ibid.*, pp. 91–92.

17. Holonyak, N., Jr., and Bevacqua, S. F., "Coherent (Visible) Light Emission from $Ga(As_{1-x}P_x)$ Junctions," *Ibid.*, pp. 82–83.

18. Melngailis, I., "Maser Action in InAs Diodes," *Ibid.*, vol. 2, 1963, pp. 176–178.

19. Weiser, K., and Levitt, R. S., "Stimulated Light Emission from Indium Phosphide," *Ibid.*, pp. 178–179.

20. Melngailis, I., Strauss, A. J., and Rediker, R. H., "Semiconductor Diode Masers of $(In_xGa_{1-x})As$," *Proc. IEEE* (*Correspondence*), vol. 51, Aug. 1963, pp. 1154–1155.

21. Alexander, F. B., *et al.*, "Spontaneous and Stimulated Infrared Emission from Indium Phosphide Arsenide Diodes," *Appl. Phys. Letters*, vol. 4, 1964, pp. 13–15.

22. Phelan, R. J., Jr., *et al.*, "Infrared InSb Laser Diode in High Magnetic Fields," *Ibid.*, vol. 3, 1963, pp. 143–145.

23. Armstrong, J. A., Nathan, M. I., and Smith, A. W., "Harmonic Generation in GaAs Injection Lasers," *Ibid.*, pp. 68–69.

24. Butler, J. F., *et al.*, "PbTe Diode Lasers," *Ibid.*, vol. 5, 1964, pp. 75–77.

25. Butler, J. F., *et al.*, "PbSe Diode Laser," *Solid State Commun.*, vol. 2, 1964, pp. 303–304.

26. Melngailis, I., Phelan, R. J., Jr., and Rediker, R. H., "Luminescence and Coherent Emission in a Large Volume Injection Plasma in InSb," *Appl. Phys. Letters*, vol. 5, 1964, pp. 99–100; Melngailis, I., "Longitudinal Injection-Plasma Laser of InSb," *Ibid.*, vol. 6, 1965, pp. 59–60.

27. Basov, N. G., Discussion in *Advances in Quantum Electronics, op. cit.*, p. 506.

28. Basov, N. G., and Bogdankevich, O. V., "Excitation of Semiconductor Lasers by a Beam of Fast Electrons," in *Symp. on Radiative Recombination in Semiconductors.* Paris: Dunod, 1964, pp. 225–233; Basov, N. G., and Bogdankevich, O. V., "Recombination Radiation of GaAs and Ge on Excitation with Fast Electrons," *Soviet Phys. JETP*, vol. 17, 1963, pp. 751–752.

29. Benoit à la Guillaume, C., and Debever, J. M., "Effet Laser par Bombardement Electronique," in *Symp. on Radiative Recombination in Semiconductors, op. cit.*, pp. 255–257.

30. Hurwitz, C. E., and Keyes, R. J., "Electron-Beam-Pumped GaAs Laser," *Appl. Phys. Letters*, vol. 5, 1964, pp. 139–141.

31. Cusano, D. A., "Radiative Recombination from GaAs Directly Excited by Electron Beams," *Solid State Commun.*, vol. 2, 1964, pp. 353–358; Cusano, D. A., Fenner, G. E., and Carlson, R. O., "Recombination Scheme and Intrinsic Gap Variation in $GaAs_{1-x}P_x$ Semiconductors from Electron Beam and p-n Diode Excitation," *Appl. Phys. Letters*, vol. 5, 1964, pp. 144–146.

32. Klein, C. A., "Threshold Considerations for Electron-Beam Pumped GaAs Lasers," *Bull. Am. Phys. Soc.*, vol. 10, 1965, pp. 387–388; "Laser-Action Threshold in Electron-Beam Excited Gallium Arsenide," Raytheon Tech. Memo. T-631, Feb. 15, 1965.

33. Benoit à la Guillaume, C., and Debever, J. M., "Effet Laser dans l'Antimoniure de Gallium par Bombardement Electronique," *Compt. Rend.*, vol. 259, 1964, pp. 2200–2202; "Emission Spontanée et Stimulée du Tellure par Bombardement Electronique," *Solid State Commun.*, vol. 3, 1965, pp. 19–20.

34. Phelan, R. J., Jr., and Rediker, R. H., "Optically Pumped Semiconductor Laser," *Appl. Phys. Letters*, vol. 6, 1965, pp. 70–71.

35. Melngailis, I., "Optically Pumped Indium Arsenide Laser," *IEEE J. Quantum Electronics*, to be published.

36. Rediker, R. H., *et al.*, "Communications Using a GaAs Infrared Source," *Electronics*, vol. 35, 1962, pp. 44–45.

37. Dalrymple, G. F., Goldstein, B. S., and Quist, T. M., "A Solid State Room Temperature Operated GaAs Laser Transmitter," *Proc. IEEE* (*Correspondence*), vol. 52, Dec. 1964, pp. 1742–1743.

38. Chantry, G. W., Gebbie, H. A., and Hilsum, C., "Interferrometric Raman Spectroscopy using Infrared Excitation," *Nature*, vol. 203, 1964, pp. 1052–1053.

39. Rutz, R. F., "Transistor-like Device using Optical Coupling between Diffused p-n Junctions in GaAs," *Proc. IEEE* (*Correspondence*), vol. 51, Mar. 1963, pp. 470–471; Rediker, R. H., Quist, T. M., and Lax, B., "High Speed Heterojunction Photodiodes and Beam of Light Transistor," *Ibid.*, Jan. 1963, pp. 218–219.

40. Lax, B., "Scanatron—A Scanning Beam Semiconductor Laser," *Solid State Design*, vol. 6, 1965, pp. 19–23.

41. Dumke, W. P., "Interband Transitions and Maser Action," *Phys. Rev.*, vol. 127, 1962, pp. 1559–1563.

42. McWhorter, A. L., Zeiger, H. J., and Lax, B., "Theory of Semiconductor Maser of GaAs," *J. Appl. Phys.*, vol. 34, 1963, pp. 235–236.

43. Lasher, G. J., "Threshold Relations and Diffraction Loss for Injection Lasers," *IBM J. Res. Develop.*, vol. 7, 1963, pp. 58–61.

44. Pilkuhn, M., and Rupprecht, H., "A Relation Between the Current Density at Threshold and the Length of Fabry-Perot Type GaAs Lasers," *Proc. IEEE* (*Correspondence*), vol. 51, Sept. 1963, pp. 1243–1244.

45. Lasher, G., and Stern, F., "Spontaneous and Stimulated Recombination Radiation in Semiconductors," *Phys. Rev.*, vol. 133, 1964, pp. A553–A563.

46. Burns, G., Dill, F. H., and Nathan, M. I., "The Effect of Temperature on the Properties of GaAs Laser," *Proc. IEEE* (*Correspondence*), vol. 51, 1963, pp. 947–948.

47. Melngailis, I., and Rediker, R. H., "The Madistor—A Magnetically Controlled Semiconductor Plasma Device," *Proc. IEEE*, vol. 50, Dec. 1962, pp. 2428–2435.

48. Hurwitz, C., to be published.

49. Galeener, F. L., *et al.*, "Evidence for the Role of Donor States in GaAs Electroluminescence," *Phys. Rev. Letters*, vol. 10, 1963, pp. 472–474.

50. Galeener, F. L., *et al.*, "Magnetic Properties of InAs Diode Electroluminescence," *J. Appl. Phys.*, May 1965, pp. 1574–1579.

51. Lax, B., and Stevenson, D. T., "High Magnetic Field Far Infrared Radiation Sources," *Proc. 1964 Far Infrared Phys. Symp.*, Naval Ordnance Laboratory, Corona, Calif., pp. 15–22.

52. Wolff, P. A., "Proposal for a Cyclotron Resonance Maser in InSb," *Physics*, vol. 1, 1964, pp. 147–157.

53. Gurevich, U. L., and Firsov, Yu. A., "On the Theory of the Electrical Conductivity of Semiconductors in a Magnetic Field," *Soviet Phys. JETP*, vol. 13, 1961, pp. 137–146.

54. Meyer, H. J. G., "On the Low Temperature Limit of mm-Wave Cyclotron Resonance Line Widths," *Phys. Letters*, vol. 2, 1962, pp. 259–260.

55. Rotstein, J., Zieman, H. E., McNamara, F. L., and Lax, B., "A Scanned Beam Laser," to be published.

# Mariner Mars 1964 telemetry and command system

*The Mariner Mars telecommunication system was designed to transmit video data from the vicinity of Mars and additional scientific data during the flight, as well as direct and quantitative commands to the spacecraft*

*Richard P. Mathison* *Jet Propulsion Laboratory*

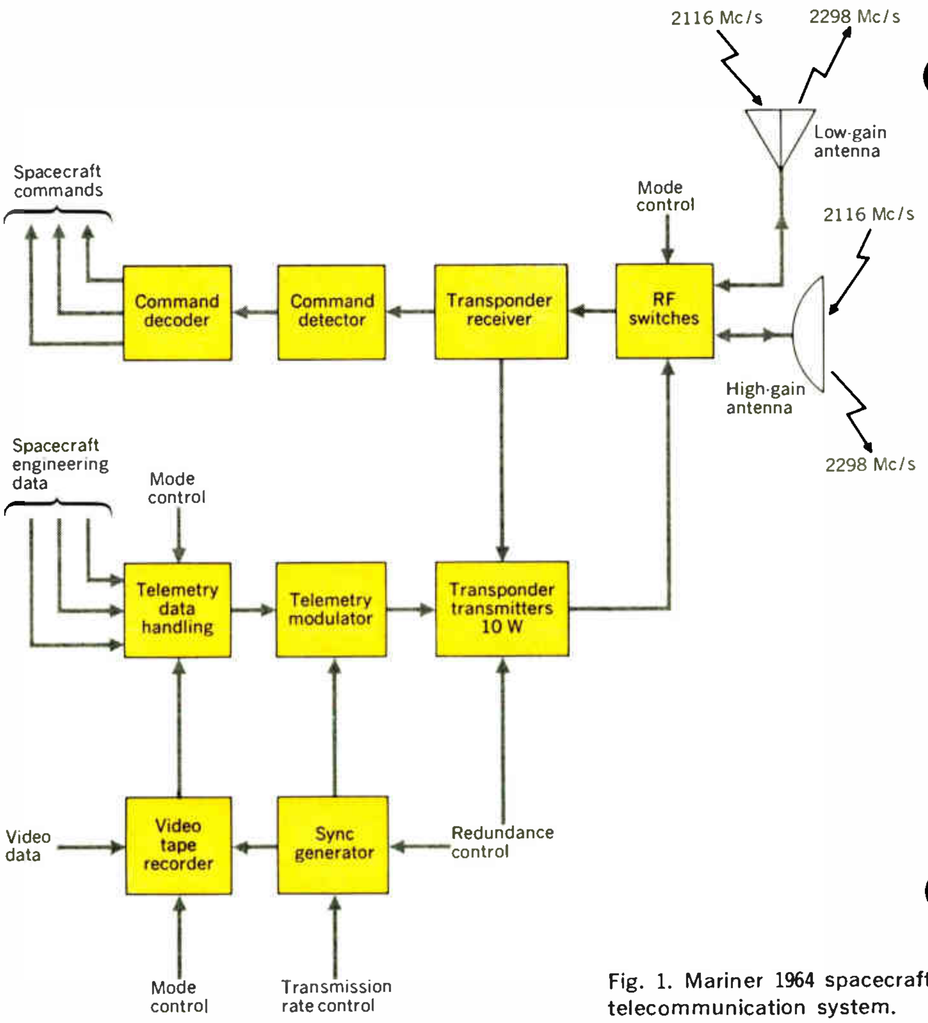


Fig. 1. Mariner 1964 spacecraft telecommunication system.

The initial objective of the Mariner Planetary–Interplanetary Program is the preliminary probing of the planets Mars and Venus by unmanned spacecraft. The probing of Venus was successfully accomplished by Mariner II. The next operation directed toward this objective is the probing of Mars by the Mariner IV spacecraft, which is now in transit.

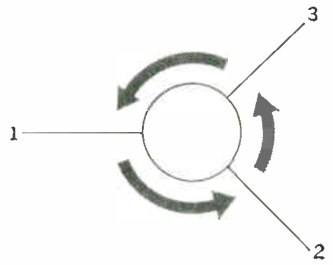
The primary objective of the Mariner Mars 1964 project is to conduct close-up, fly-by scientific observations of the planet Mars during the 1964–1965 opportunity and to transmit the results of these observations back to earth. Television, cosmic dust, and a complement of fields and particles experiments are being carried by Mariner IV. In addition, an earth occultation experiment is planned.

A secondary objective is to provide experience and knowledge about the performance of the basic engineering equipment of an attitude-stabilized fly-by spacecraft during a long-duration flight in space farther away from the sun than is the earth.

The Mariner IV spacecraft was launched on November 28, 1964. The spacecraft is fully attitude stabilized, using the sun and Canopus as reference objects. It derives power from photovoltaic cells, arranged on panels having body-fixed orientation, and a battery, which is used for launch, trajectory correction maneuvers, and back-up. The telecommunication system for the Mariner Mars 1964 Mission is comprised of spacecraft-borne equipment and the NASA Deep Space Net.[1] It is required to perform three functions: (1) tracking the position and velocity of the spacecraft, (2) telemetering engineering and scientific data from the spacecraft, and (3) transmitting commands to the spacecraft. The design of the spacecraft equipment is based upon techniques that were used for the Mariner II spacecraft.[2,3] These techniques have been extended and modified to improve equipment reliability, accommodate the increased communication range required by the Mars 1964 Mission, and utilize the characteristics of the Mars 1964 trajectories to effect simplifications in the spacecraft equipment.

Single CW radio-frequency carriers that are transmitted to and from the spacecraft are used for tracking the spacecraft and transmitting the telemetry and command information. The functional arrangement of the spacecraft subsystems utilized to accomplish this is shown in Fig. 1. For both the telemetry and command functions, pulse code modulation, phase shift key, and phase modulation (PCM/PSK/PM) techniques in combination with pseudorandom sync codes provide efficient, accurate transmission of the data over interplanetary distances.

Fig. 2. Circulator switch circuit.



| Input terminal | Normal output terminal | Reversed output terminal |
|---|---|---|
| 1 | 2 | 3 |
| 2 | 3 | 1 |
| 3 | 1 | 2 |

The telemetry portion of the system is required to transmit video data in digital form from the vicinity of Mars and both scientific and engineering data during the flight from earth to Mars. Since the rate at which the video data is gathered exceeds the capacity of the telemetry channel, data storage and playback are provided by a synchronous, endless-loop tape recorder capable of storing 20 frames of video data.

The duration of the Mars 1964 Mission is approximately eight months, in contrast to the Venus 1964 Mission, the duration of which was four months. In order to accommodate this increased equipment operating time, modest reliability improvements were incorporated within the constraints of available power and weight. These improvements take the form of better components, extensive part screening, worst-case circuit designs, and redundant elements in the telemetry modulator and transmitter.

By utilizing the unique characteristics of the Mars 1964 minimum-energy trajectories,[4] considerable savings in spacecraft weight and complexity were realized. The variation in earth, spacecraft, sun, and Canopus geometry permitted the use of a moderately high-gain antenna that is fixed with respect to the spacecraft and thus eliminated the need for antenna pointing mechanisms.

The command system provides for the transmission of both direct and quantitative commands to the spacecraft in digital form.

A detailed discussion of the scientific instruments and the data automation system that controls and interfaces with the instruments is beyond the scope of this article.[5] Moreover, details of the modulation–demodulation theory that forms a basis of the data transmission techniques have been adequately covered elsewhere[6-8] and will not be repeated here. However, estimates of expected communication performance will be included.

## Radio subsystem

The radio subsystem is required to receive a modulated RF carrier from stations of the Deep Space Net (DSN), demodulate command and ranging signals, coherently translate the frequency and phase of the RF carrier by a fixed ratio, modulate the carrier with telemetry and ranging signals, and retransmit it back to earth. It consists of an automatic-phase-control receiver, redundant exciters, redundant power amplifiers, power supplies, low- and high-gain antennas, and associated transmission and control circuits. It operates at S-band frequencies, receiving at 2116 Mc/s and transmitting at 2298 Mc/s.

As received from earth, the up-link RF signal is phase-modulated either singly or simultaneously by a composite command signal and a coded ranging signal. It is of the form

$$S_R = A(\gamma,r)\sin\left[\omega_0 t' + \phi_c(t') + \phi_r(t')\right] \qquad (1)$$

$$t' = t - \frac{r(t)}{c} \qquad (2)$$

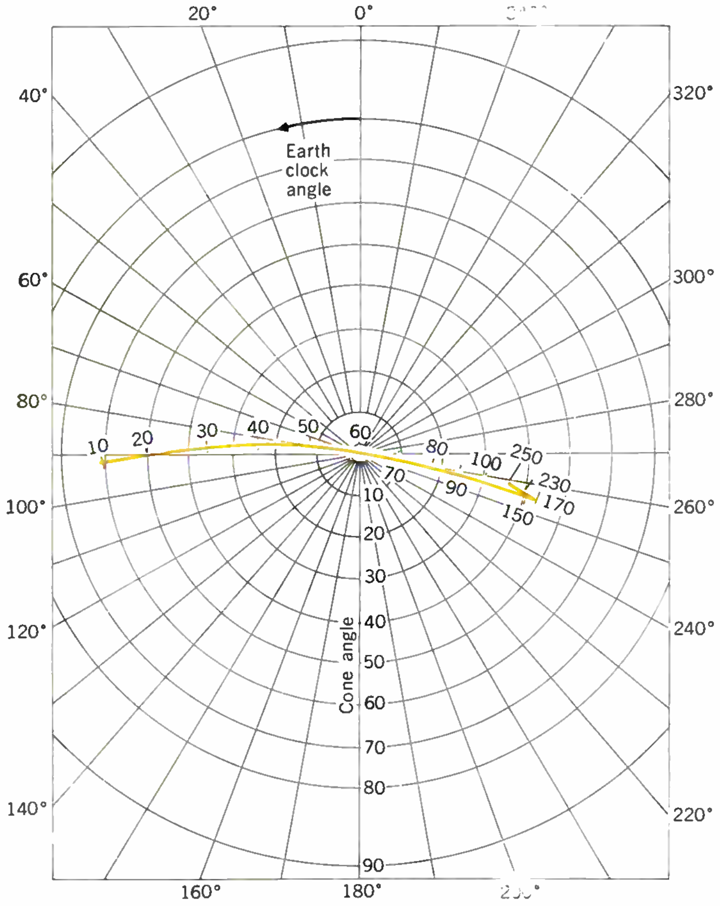


Fig. 3. Earth track vs. time. Numbers indicate days from launch; encounter = approximately 230 days.

where

$A$ is the received signal level, a function of spacecraft attitude $\gamma$ and spacecraft earth range $r$

$\omega_0$ is the carrier frequency transmitted by the DSN station

$\phi_c$ is the phase modulation by the composite command signal

$\phi_r$ is the phase modulation by the coded ranging signal

$r(t)$ is the spacecraft–earth range, a function of time $t$

$c$ is the velocity of propagation

This signal is demodulated by the automatic phase control, double superheterodyne receiver which tracks the $\omega_0 t'$ component of the carrier phase. The composite command modulation and coded ranging signals are sent to the command detector and the exciter phase modulators, respectively. When the receiver is phase-locked to the received signal, it generates for the transmitter exciter a filtered phase reference that is coherent with the $\omega_0 t'$ component of the received signal. The phase of the transmitted signal is then related to that of the received signal by a fixed ratio to within an error of less than 1 radian rms. The resulting transfer function is given approximately by

$$\frac{\theta_T}{\theta_R} = \frac{240}{221}\left[\frac{1 + (3/4B)s}{1 + (3/4B)s + \tfrac{1}{2}(3/4B)^2 s^2}\right] \qquad (3)$$

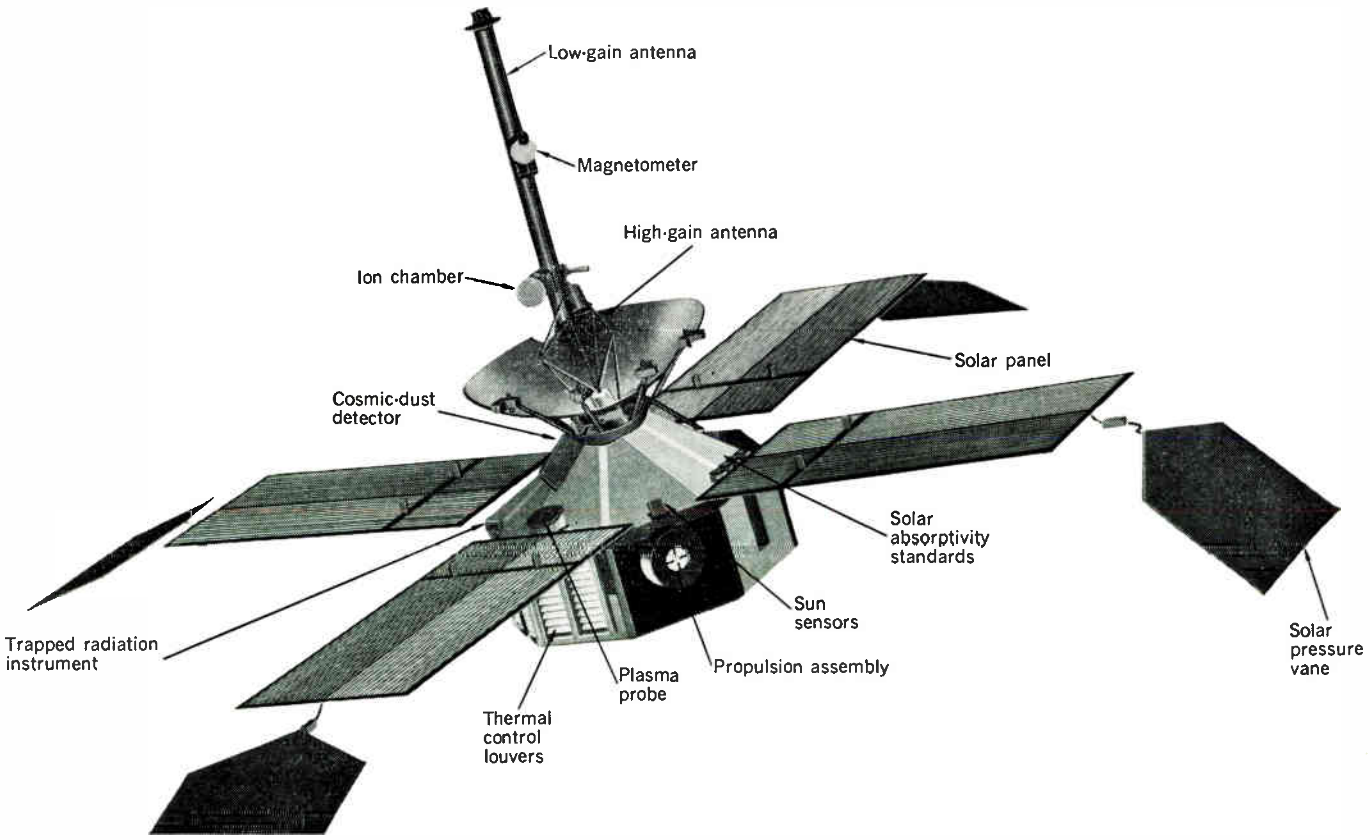
where

$s$ is the Laplace variable and $B$ is the effective noise bandwidth of the receiver phase tracking loop

$\theta_T$ is the phase of the transmitted signal

$\theta_R$ is the phase of the received signal

Fig. 4. Mariner 1964 spacecraft.



With this relationship the ground stations are provided with a signal that permits two-way Doppler tracking.[9]

The transmitted signal is phase-modulated by a composite telemetry signal and the coded ranging signal. While the telemetry signal modulates the carrier continuously, the ranging modulation can be turned on or off by ground command.

When a signal is not being transmitted to the spacecraft, transmitter frequency control is provided by an auxiliary crystal oscillator. This noncoherent mode of operation permits one-way Doppler tracking, angular position tracking, and telemetry reception by the ground stations.

In order to provide increased reliability over the Mariner II design, redundant exciters, power amplifiers, and power supplies have been incorporated in the transmitter. Each exciter consists of an auxiliary oscillator, a ×4 frequency multiplier, a phase modulator, a ×30 frequency multiplier, and an output isolator. Either exciter can be coupled to either power amplifier by a circulator switching network. Similarly, the input and output circuits of the power amplifiers are coupled through circulator switches.

The control of the switching between these elements is provided by either ground command or on-board failure detection. In the case of ground command control, the receipt of the appropriate direct command causes the control unit simultaneously to transfer the dc power from the active to the inactive element and reverse the circulator switch or switches. For the exciters, the modulation, phase reference, and mode control inputs are fed to both exciters in parallel.

In the case of switching by on-board failure detection, power monitors sample both the exciter and power amplifier RF power outputs. When an output power drops below a preset level, a gate in the control unit is enabled which allows cyclic pulses from the control computer and sequencer (CC&S) to toggle the relay driver circuit. Upon the receipt of one such pulse, the control unit transfers the dc power and RF circuits in the same manner as when a ground command is received. If the power output from the redundant element then exceeds the threshold, the gate inhibits further transmission of pulses to the driver circuit. The thresholds for enabling the gates to operate are set at 3 dB below the nominal exciter and power amplifier outputs. The cyclic pulses occur once every $66\frac{2}{3}$ hours. Thus, the maximum switching time after a failure is $66\frac{2}{3}$ hours.

Circulator switches were chosen for the control of the RF transmission paths because they appeared to offer significant reliability advantages over conventional electromechanical coaxial switches. As an RF circuit, the circulator is simply a strip-line Y connection with no moving parts. The Y is surrounded by a ferrite material that is polarized by a dc electromagnetic field. Signal flow through the device is circular, as indicated in Fig. 2. By reversing the magnetic field, the signal can be made to "circulate" in the opposite direction and hence the switching action. In the event of a loss in electromagnetic field, the circuit will function like a transmission line T, with the attendant power splitting and increased mismatch losses, but will not cause a complete loss of performance.

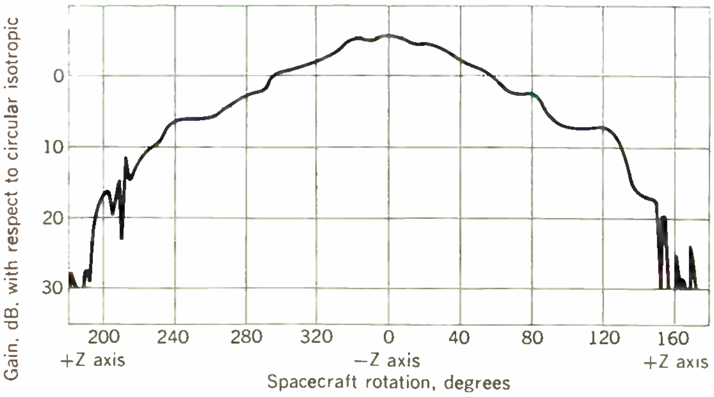


Fig. 5. Typical low-gain antenna pattern.

The Mariner Mars 1964 spacecraft uses both the sun and the star Canopus for attitude references. Sun sensors provide pitch and yaw control such that the roll axis is pointed toward the sun, and the Canopus sensor provides roll position control. With this type of attitude control the position of earth as seen from the spacecraft (the direction of the spacecraft coordinate system) varies as shown in Fig. 3 for a typical Mars 1964 trajectory.[4]

It can be seen that the locus remains within one hemisphere of the spacecraft during the entire flight and within a relatively small angular region during the later portion of the flight, from 130 days before the Mars encounter to 20 days past encounter. A comparison of this characteristic and the required minimum antenna gain vs. time-of-flight showed that the gain requirements could be met with a combination of one low-gain and one high-gain antenna, both of which were fixed relative to the spacecraft (Fig. 4). The low-gain antenna provides coverage during the first 70 to 95 days of flight, whereas the high-gain antenna fills in the remaining period until approximately 20 days past encounter.

The low-gain antenna consists of a cruciform aperture at the end of a low-loss circular waveguide, which also functions as the support structure. In order to minimize pattern distortion by reflections from the spacecraft structure, the aperture is mounted well away from the bulk of the spacecraft. As shown in Fig. 5, the antenna provides a pattern of revolution about the roll axis with a maximum gain of 5.5 dB at 2298 Mc/s in the direction of the $-Z$ spacecraft axis (oriented toward the sun) and a minimum gain of −6 dB with respect to circular isotropic over the entire $-Z$ hemisphere. The pattern at 2116 Mc/s is similar.

Since the thrust vector of the mid-course motor is perpendicular to the $-Z$ axis, the earth can be kept within the $-Z$ hemisphere while the thrust vector is pointed in any arbitrary direction. Thus, the low-gain antenna pattern also meets the requirement for providing coverage during mid-course maneuvers of unrestricted direction.

The high-gain antenna is a 46.0- by 21.2-inch parabolic reflector that is illuminated by a pair of turnstile elements. These elements are arranged so that a right-hand circularly polarized beam is projected with a maximum gain of 23.5 dB (at 2298 Mc/s) and a half-power beam width of 13.5° by 7.5°. The beam is positioned so that coverage is provided from approximately 90 days after launch until 20 days past encounter. As a result of using this design as opposed to the one-degree-of-freedom antenna that was used on Mariner II,[2,3] an estimated 50 pounds of spacecraft weight was saved by the associated reductions in structure, actuator, control electronics, and power requirements.

Three transmitting and receiving modes are available:

1. Transmit low gain, receive low gain.
2. Transmit high gain, receive low gain.
3. Transmit high gain, receive high gain.

These modes provide the required coverage during the acquisition, cruise, mid-course maneuver, and encounter phases of the flight. Selection of the proper mode is controlled by programmed CC&S commands or ground commands as a backup.

In addition, two failure mode controls are provided. First, if roll position control is inadvertently lost while the receiver high-gain mode is being used, the loss of the Canopus sensor signal automatically switches the receiver to the low-gain antenna so that command capability can be maintained. Second, if the spacecraft does not receive a signal from earth at least once between the occurrence of the 66⅔-hour cycle pulses, as signified by receiver phase lock, the control unit automatically switches the receiver from one antenna to the other after the receipt of two such pulses. The receiver is subsequently cycled between the antennas once every 66⅔ hours until phase lock is obtained. This later mode control provides partial redundance for some antenna failure modes.

Summaries of the principal radio subsystem transmission and reception parameters are given in Tables I and II, respectively.

### Telemetry subsystem: basic technique

The principal functions of the telemetry subsystem on the spacecraft are to time-multiplex engineering and scientific data samples and to encode them for efficient modulation of the spacecraft-to-earth RF carrier. The subsystem is specifically required to

1. Transduce engineering parameters into electric signals.
2. Time-multiplex (commutate) engineering and scientific measurement signals.
3. Convert engineering data samples to binary words.
4. Store digitally encoded video data.
5. Phase-shift-key a subcarrier with the binary signal.
6. Generate a cyclic, binary, pseudorandom sequence for use in synchronizing the encoding and decoding of the telemetry data.
7. Phase-shift-key a second subcarrier with the sync code.
8. Combine the two subcarriers into a composite telemetry signal.

The basic timing for the subsystem is derived from the 2400-c/s spacecraft power frequency, which is divided down to provide two subcarrier frequencies, one for data and one sync. The frequency divider is arranged to provide two data transmission (bit) rates, 33⅓ and 8⅓ bits per second (b/s). While the 33⅓-b/s rate is used during preflight check-out and the early flight phases up through a first mid-course maneuver, the 8⅓-b/s rate is used for the remainder of the flight. In-flight selection of the data rate is controlled by ground command and

**I. Spacecraft radio transmission parameters (2998 Mc/s)**

| | Transponder Low-Gain Channel | | Transponder High-Gain Channel | |
|---|---|---|---|---|
| Parameter | Value | Tolerance | Value | Tolerance |
| Total transmitter power[a] | +40.0 dBm | ±0.5 dB | +40.0 dBm | ±0.5 dB |
| Carrier modulation loss[b] | −4.1 dB | +0.9 dB<br>−1.0 dB | −4.1 dB | +0.9 dB<br>−1.0 dB |
| Transmission circuit loss[c] | −1.7 dB | +0.2 dB<br>−0.3 dB | −1.3 dB | +0.2 dB<br>−0.3 dB |
| Spacecraft antenna gain[d] | +6.0 dB | ±1.8 dB | +23.2 dB | ±1.1 dB |

[a] Ten watts nominal output of traveling-wave-tube amplifier.
[b] Based on modulation indexes of 0.809 rad peak for data subcarrier and 0.451 rad peak for sync subcarrier.
[c] Includes all circuitry between the output of the TWT amplifier and the input to the antenna.
[d] Referenced to perfectly circular isotropic pattern maximum.

**II. Spacecraft radio reception parameters (2116 Mc/s)**

| | Transponder Low-Gain Channel | | Transponder High-Gain Channel | |
|---|---|---|---|---|
| Parameter | Value | Tolerance | Value | Tolerance |
| Antenna gain (pattern maximum)[a] | +6.5 dB | ±1.8 dB | +21.8 dB | ±1.1 dB |
| Receiving circuit loss[b] | −1.0 dB | ±0.2 dB | −0.9 dB | ±0.2 dB |
| Effective system noise temperature[c] | 2700°K | +1700°K<br>−610°K | 2700°K | +1700°K<br>−610°K |
| Carrier APC noise bandwidth ($2B_{LO}$)[d] | 20.0 c/s | ... | 20.0 c/s | ... |
| Carrier threshold SNR in $2B_{LO}$ | | | | |
| Two-way Doppler tracking[e] | +3.8 dB | ... | +3.8 dB | ... |
| Command reception | +8.0 dB | ±1.0 dB | +8.0 dB | ±1.0 dB |

[a] Referenced to perfectly circular isotropic pattern maximum.
[b] Includes all circuitry between the antenna and the input to the transponder receiver.
[c] Includes contributions due to antenna temperature, circuit losses, and noise figure at input to preselector, 10 dB (+2 dB, −1 dB).
[d] Tolerance included in uncertainty of system noise figure.
[e] 3.8-dB SNR is required for +2.0-dB ground receiver degradation.

CC&S command. Either rate can be selected by ground command, but the CC&S selects only the 8⅓-b/s rate 192 days before encounter. The CC&S control is to insure that the 8⅓-b/s rate is used at encounter in the event that command capability is lost.

The square-wave sync subcarrier drives a redundant pair of pseudorandom code generators which generate a cyclic 63-bit code. A set of word gates, in turn, generates bit and word sync pulses that are used to synchronize (1) the stepping of the commutator, (2) the analog-to-digital converters, (3) the readout of data from the data automation system, (4) the readout of the event registers and timers, and (5) the playback of the stored video data. The word sync pulses occur once per cycle of the code, whereas the data-bit sync pulses occur once every nine code bits, or seven times per code cycle. Thus, each data word is seven data bits long.

In order to convey the bit and word sync timing to the ground stations for use in synchronous demodulation of the telemetry subcarrier, the code also phase-shift-keys the sync subcarrier. The resulting composite telemetry signal that modulates the spacecraft-to-earth carrier is given by

$$D(t) = V_d \left[ 1.79d\left(\frac{2f_dt}{9}\right) \oplus a(4f_dt) + X\left(\frac{f_dt}{2}\right) \oplus a(2f_dt) \right] \tag{4}$$

where

$V_d$ is the amplitude of the complex four-level wave
$d(2f_dt/9)$ is the binary telemetry data of amplitude ±1 and bit rate $(2/9)f_d$
$a(ft)$ is a symmetrical square wave of amplitude ±1 and frequency $f$
$X(f_dt/2)$ is a cyclic, binary, pseudorandom sequence of amplitude ±1, length 63 bits, and bit rate $f_d/2$
$\oplus$ represents modulo 2 addition

At the ground station, a local model of the code is phase-locked to the received code. Word gates identical to those in the spacecraft code generators then produce accurate bit and word sync pulse trains. For a detailed discussion of the technique, the reader is referred to References 6, 7, and 8.

Analog engineering measurements are sampled by a solid-state commutator that provides 100 channels, 90 of which are used for measurements and ten for synchronization points and subcommutation. These channels are divided among ten decks of ten channels each and are arranged to provide three sampling rates.

The pulse-amplitude-modulated output of the commutator is fed to two analog-to-digital converters, which convert the data samples to serial 7-bit words by a successive approximation technique. The output of the converter forms one of four data sources that comprise the telemetry modes.

Four modes of data transmission are provided for: (1) engineering data, (2) engineering and science data, (3) science data, and (4) stored video data and engineering data. In the first mode, only engineering data from the commutator, event register, event timer, and command monitor are transmitted, primarily for maneuver and check-out phases. In the second mode, both engineering and science data are transmitted in an alternating sequence of 140 engineering data bits followed by 280 science data bits. This mode is intended for most of the cruise phases. In the third mode, only science data are transmitted, as received from the data automation system. This mode is designed for use at planet encounter. In the fourth mode, stored video and engineering data are transmitted in alternating periods of approximately 9 and 1.5 hours, respectively. This mode provides for readout of the video data taken during encounter and periodic monitoring of the spacecraft performance after encounter.

Event-type signals that signify the occurrence of events such as motor-start, receipt-of-command, or solar-panels-open are accumulated as they occur in four separate registers. Each register accumulates different types of events, as shown in Table III, and holds up to eight counts before recycling. The registers are sampled in pairs at the high commutation rate in synchronism with the commutator, so that the state or count of two registers is conveyed by one 7-bit word.

An event timer measures the duration of certain events, such as the mid-course motor firing duration, by dividing the word sync rate by two and accumulating the number of pulses that occur between the start and the end of the event. This number is sampled at the medium rate also in synchronism with the commutator.

During the Mars encounter, a television subsystem which operates under data automation system control periodically generates video data in binary form. These data and the mode 3 instrument data generated at an effective rate of 10 700 b/s are organized in 516 168 bit frames, of which 504 400 are TV-related. Since this data rate greatly exceeds the 8⅓-b/s radio transmission capability at encounter, a data storage subsystem holds the data for postencounter readout.

Data storage is accomplished by an endless-loop tape recorder. This machine records binary data and sync pulses on two tracks, filling one track at a time on each of two consecutive tape cycles. Recording is started and stopped by control signals from the data automation system to coincide with the encounter data frames. In order to prevent overrecording after the two tracks are filled the first time, end-of-tape signals automatically stop the recorder after the second complete tape pass. The tape is then in the correct position for subsequent playback.

Playback, at the 8⅓-b/s transmission rate and synchronous with the telemetry bit sync pulses, is accomplished by an automatic phase control servo which con-

### III. Registered events

| Channel | Events |
|---|---|
| 1 | Pyrotechnics current pulse<br>Gyro turn-on<br>Solar panel 1 open |
| 2 | CC&S events<br>Solar panel 2 open |
| 3 | Pyrotechnic arm<br>Pyrotechnic current pulse<br>Solar panel 3 open<br>Recorder end of tape signal |
| 4 | Ground command events<br>Sun acquired<br>Solar panel 4 open<br>Scan platform unlatched |

trols the tape speed so that the recorded bit sync pulses are kept in phase with the telemetry bit sync pulses. By this means the pseudorandom sync signal allows synchronous demodulation of the recorded data at the ground stations.

In conjunction with the starting and stopping of the recorder during the record cycles, approximately 3 to 5 feet of tape are used while the machine accelerates and decelerates. No data are recorded on these segments. During the continuous playback, these blank spots provide approximately 1.5 hours in which spacecraft engineering data are inserted for periodic monitoring of post-encounter spacecraft performance. Control of this alternation between the recorded and engineering data is provided by a circuit that senses the presence or absence of data on the tape. Table IV summarizes the characteristics of the tape machine.

As with the radio subsystem, limited redundance has been incorporated in the telemetry subsystem for increased reliability. This redundance is in the form of two pseudorandom code generator analog-to-digital converter pairs which operate with parallel inputs and logical "or" coupled outputs. Only one pair operates at a time, and this pair is selected by ground command.

In addition, the commutator sequencer has been designed so that many of the possible failure modes result in a modification or "short counting" of the sequence rather than a complete stoppage. The number of channels that would be lost for a given failure depends on the location of the failed component so that a varying degree of partial success can exist. For example, a short count in a low-rate deck would not affect the higher-rate channels, while a short count in a high-rate deck could bypass a large number of low-rate channels.

Finally, redundant components such as resistors, diodes, and capacitors have been employed in the power transformer–rectifier unit. Table V lists the principal telemetry subsystem parameters.

**Command subsystem**

Commands are transmitted from DSN ground stations to the spacecraft by two subcarriers, which phase-modulate the earth-to-spacecraft RF carrier. One subcarrier is phase-shift-keyed by serial binary command words, and the other by a pseudorandom sync code in a manner similar to that used for telemetry data transmission.

The command subsystem is required to detect and decode the command words, of which there are two types: direct commands, which result in selected switch closures, and quantitative commands, which convey a magnitude and polarity for spacecraft maneuvers.

Initial acquisition is achieved by slightly offsetting the frequency of the clock at the ground stations from the average static frequency of the loop voltage-controlled oscillator (VCO). Under this condition, the local code is slowly shifted in phase with respect to the received code until the phases match. The frequency difference is made small enough so that the automatic phase control loop receives sufficient signal to acquire phase lock and the acquisition is complete.

Outputs from the command subsystem include the direct command switch closures, the quantitative command bits, bit sync pulses, alert pulses for the CC&S, and several telemetry signals. In the case of both direct and quantitative command output circuits, complete dc isola-

### IV. Video storage characteristics

| | |
|---|---|
| Record rate | 10 000 b/s |
| Playback rate (synchronous) | 8⅓ b/s |
| Storage capacity | $5.24 \times 10^6$ bits |
| Number of tracks | 2 |
| Type of tape machine | Endless loop |

### V. Telemetry parameters

| | |
|---|---|
| Type of encoding | Sampled data, digital PSK with pseudonoise sync |
| Channel requirements: | |
| Engineering measurements | 90 |
| Event counters | 4 |
| Word length | 7 bits |
| Transmission rates | 33⅓, 8⅓ b/s |
| Word error probability at threshold | 1 word in 28 (bit error probability $= 5 \times 10^{-3}$) |
| Required ST/(N/B)* for bit error probability $= 5 \times 10^{-3}$ | $7.6 \frac{\mathrm{dB}_{c/s}}{\mathrm{b/s}} \pm 0.7$ dB |
| Data channel modulation loss | −4.6 ± 0.6 dB |
| Sync channel threshold S/(N/B) | $11.0 \pm 0.5\ \mathrm{dB}_{c/s}$ |
| Sync channel modulation loss | −10.5 dB (+0.2 dB, −0.3 dB) |

*S = signal power; T = time for one bit; N = noise power; B = bandwidth.

### VI. Command parameters

| | |
|---|---|
| Number of commands | |
| Discrete | 29 |
| Quantitative | 1 address,<br>3 subaddresses |
| Modulation type | Digital PSK with PN sync |
| Word length | 26 bits |
| Transmission rate | 1 b/s |
| Command threshold criteria: | |
| Probability of correctly executing a discrete command in one attempt | >0.7 |
| Probability of completely executing a quantitative command in one attempt | >0.5 |
| Probability of a bit error in a completely executed quantitative command | $<2 \times 10^{-4}$ |
| Probability of a false discrete or quantitative command being executed when another command is sent | $<2 \times 10^{-9}$ |
| Required carrier SNR in 20-c/s bandwidth at command threshold | +8.0 ± 1.0 dB |
| Required command channel ST/(N/B) at threshold | +15.7 ± 1.0 dB |
| Command channel modulation loss | +8.5 ± 0.6 dB |
| Required sync channel SNR at threshold | +15.7 ± 1.0 dB |
| Sync channel effective noise bandwidth | +2.0 ± 0.8 dB |
| Sync channel modulation loss | +5.5 ± 0.5 dB |

tion is maintained from the interfacing spacecraft subsystems.

As an aid to acquisition and in-flight performance monitoring, the sync channel VCO frequency and in-lock signals are telemetered. For this purpose, a special counter converts the VCO frequency to a binary number which is periodically sampled by the telemetry system. Table VI lists the principal command subsystem parameters.

## Performance

The telecommunication system is required to provide tracking, telemetry, and command performance from launch to 20 days past encounter, including all of the intermediate phases. In order to reasonably assure this capability, it was desired to choose the system parameters so that the nominal received signal levels always exceeded the threshold signal levels by at least the linear sum (in dB) of the adverse tolerances.[10] This criterion has been met for all functions and flight phases, except for the telemetry for a period of 10 to 26 days, depending on the launch date.

Figure 6 illustrates, for a typical trajectory, the nominal received carrier level for the spacecraft-to-earth channel vs. time from launch. The variations are due to both the increasing range and the variable antenna gains, and it is apparent where the performance of the low-gain antenna leaves off and that of the high-gain antenna takes over.

For the diplexed tracking feed and maser ground station configuration, the nominal threshold carrier level for telemetry is −164.4 dBm at 8⅓ b/s. A comparison between this value, the nominal carrier levels, and the system tolerances (Fig. 7) shows that the design criterion has been met over most of the flight, and the extent to which it has not been met at the transition region. In the transition region, the telemetry performance may be marginal.

This situation is a result of the antenna position compromise that had to be made between mid-flight performance and postencounter performance with a relatively simple antenna design. Mid-flight performance was sacrificed to meet the 20-day postencounter requirement. Since the nominal carrier level is never less than the nominal threshold level, it is considered to be a reasonable compromise.

The nominal received carrier levels for the earth-to-spacecraft channel are shown in Fig. 8. Since the same spacecraft antennas are used for transmitting and receiving, both up and down channels exhibit similar time variations. A comparison between the command thresh-

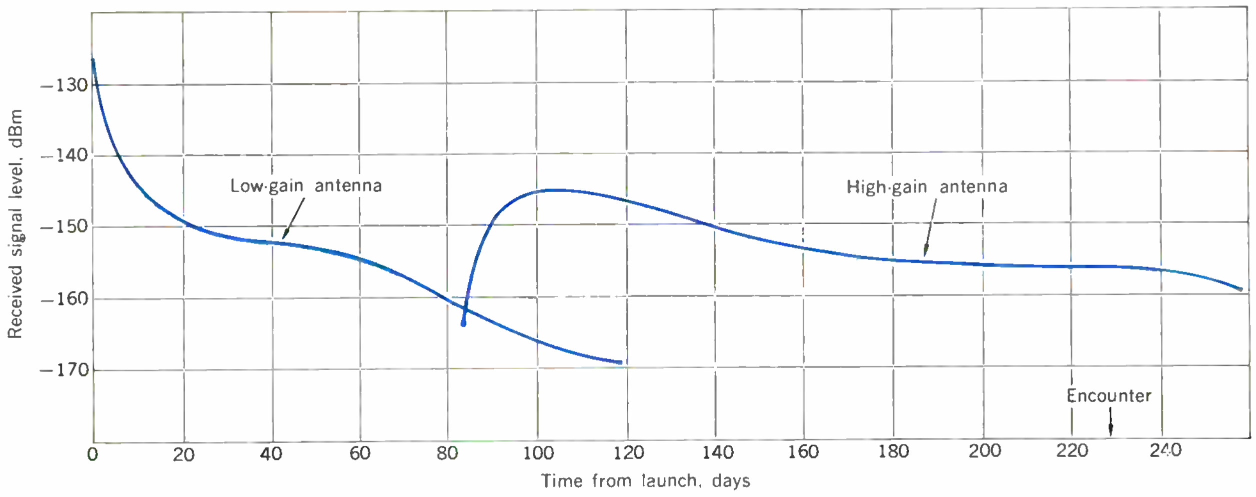
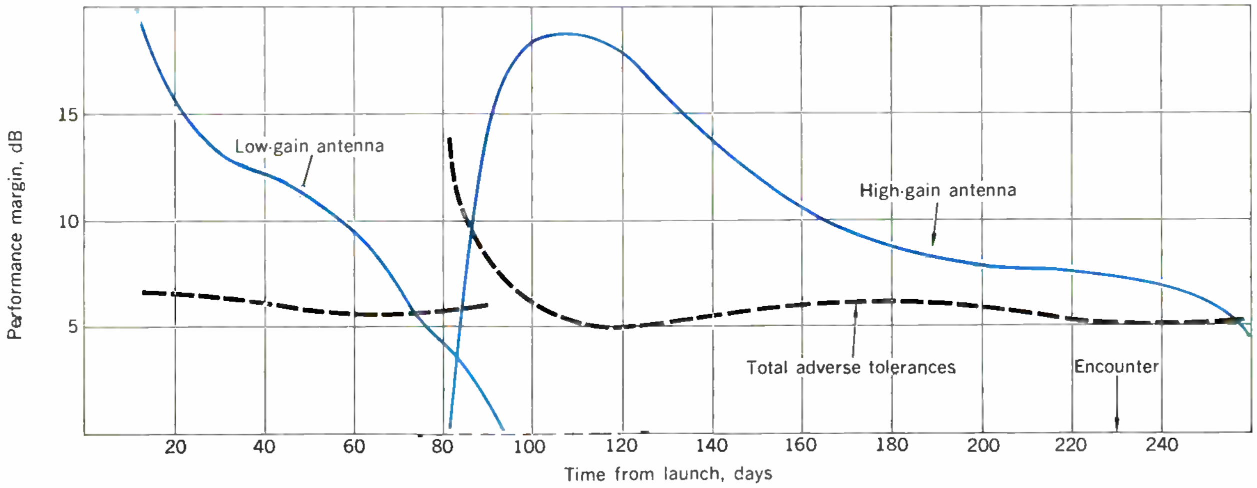


Fig. 6. Received signal level vs. time, spacecraft to earth.

Fig. 7. Telemetry performance margin vs. time. Diplexed tracking antenna with maser, 8⅓ b/s.



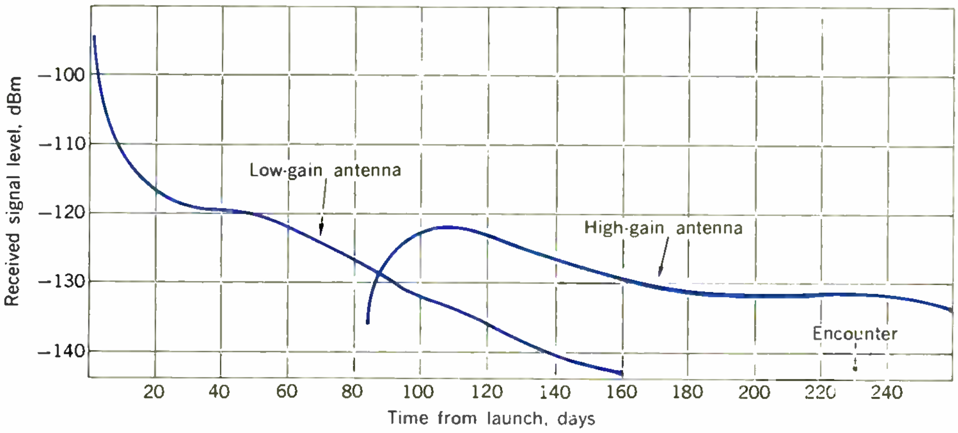


Fig. 8. Received signal level vs. time, earth to spacecraft. Ground transmission power = 10 kW, command modulation on.

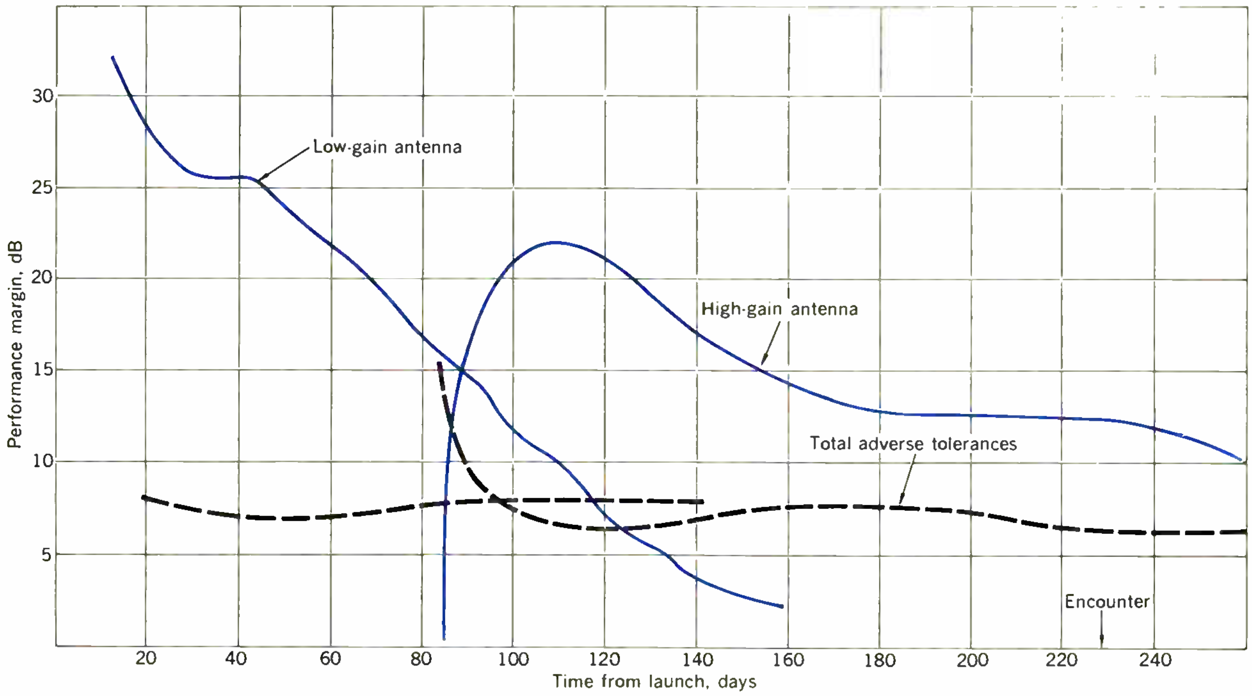


Fig. 9. Command performance margin vs. time. Ground transmission power = 10 kW.

old carrier level of −143.3 dBm, the nominal received level, and the system tolerances (Fig. 9) shows that the design criterion for command has been met for all flight phases.

## Conclusion

The Mariner Mars Mission for 1964 required a telecommunication system to provide tracking, telemetry, and command capabilities over communication distances up to 260 million km and which would operate for approximately 8 months in an interplanetary space environment. The design that has been described is an extension and modification of well-proved techniques, where the modifications included required improvements in performance and limited redundance to provide greater reliability.

This article is a condensed version of a paper presented at the Sixth Winter Convention on Military Electronics, Los Angeles, Calif., Feb. 3–5, 1965. It presents results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

The systems and techniques described were developed by the efforts not only of the author but of many other members of the JPL Telecommunications Division.

### REFERENCES

1. "System Capabilities and Development Schedule of the Deep Space Instrumentation Facility," Tech. Memo. 33-83, Jet Propulsion Laboratory, Pasadena, Calif., Apr. 24, 1964.

2. JPL Staff, *Mariner: Mission to Venus.* New York: McGraw-Hill Book Co., Inc., 1963.

3. Bryden, J. N., "Mariner (Venus '62) Flight Telecommunication System," Tech. Rept. 32-377, JPL, Jan. 15, 1963.

4. Space Programs Summary No. 37-23, vol. II, JPL, Sept. 30, 1963.

5. Space Programs Summary No. 37-24, vol. VI, JPL, Dec. 31, 1963, p. 71.

6. Martin, B. D., "The Mariner Planetary Communication System Design," Tech. Rept. 32-85 (Rev. 1), JPL, May 15, 1961.

7. Springett, J. C., "Command Techniques for the Remote Control of Interplanetary Spacecraft," Tech. Rept. 32-314, JPL, Aug. 1, 1962.

8. Springett, J. C., "Pseudo Random Coding for Bit and Work Synchronization of PSK Data Transmission Systems," presented at Internat'l Telemetering Conf., London, England, 1963.

9. Mathison, R. P., "Tracking Techniques for Interplanetary Spacecraft," Tech. Rept. 32-284, JPL, Aug. 1, 1962.

10. Mathison, R. P., "Spacecraft Telecommunications System Design," *1963 Yearbook WGLR Conf.*, Munich, Germany.

# The role of grounding in eliminating electronic interference

*Because of a basic difference in their approach to grounding, the electronics and the power engineer encounter difficulty in reconciling their respective systems. A common grounding language is needed as a first step to the acquisition of that common knowledge that can dispel the existing confusion*

*Trevor A. Robinson* *Northrop Corporation*

Probably more has been written on the subject of grounding over the past 40 years than on any other topic. Yet confusion and uncertainty still exist, particularly for electronic items that have been adapted for use with power systems. For power systems alone, the grounding techniques may be regarded as developed, codified, and covered by a specific terminology, so that definite statements can be made and understood. The electronics industry has related problems, but the differences are such that when the same vocabulary is used confusion rather than clarity often results. The best way to gain an understanding of the overall picture is first to present a brief review of power grounding techniques, and then to give a more rounded, fuller review of techniques of electronic grounding. A later discussion will deal with grounding of complexes involving both types of systems, particularly where a single specification is to be written. Correct understanding of both categories of problems is essential if a document is to be produced that will provide clear directions for two groups that often speak a different language.

This article is a plea to both power and electronics engineers to find a common grounding language and to acquire sufficient knowledge of each other's specialty so that the solutions proposed for particular problems in one field will not violate the basic requirements of the other.

### Conventional power system grounding

A typical utility power system could be divided essentially into the generating stations, transmission systems substations, primary distribution systems, and low-voltage secondary distribution systems. Grounding is used in each division for the same basic reasons. If it is done correctly to accomplish the main objectives, it is assumed by the utility distribution engineer that minor objectives will take care of themselves. The way the utility engineer thinks of his problem and the order of importance he assigns to his reasons are naturally conditioned by the system he is operating. Continuity of service and safety under all conditions are his particular concern.

To obtain continuity of service he must be assured that in case of trouble the defective line will be isolated from his system. The trouble might be an insulator failure on a high-voltage transmission line, a tree falling against a city primary feeder, or an equipment short circuit. Any such fault will cause a ground current to flow from the fault back to the associated station. At the station there must be an effective low-resistance ground connection because a low voltage and maximum current are desired. (A high resistance would, of course, pass only a small current and produce a high voltage.) The maximum current is needed so that the protective relays will function to remove the line from service in minimum time. The utility engineer thus strives to keep the controllable elements of the fault circuit at the lowest practicable impedance. Another condition that might arise could be a fault or voltage surge caused by lightning. The ground connections thus should be short and direct and of low impedance to true ground. With a poor connection, lightning discharge will build up a high local voltage before the wavefront has time to drain into the earth. Equipment could then be damaged and continuity of service again lost.

Safety, to the power man, means that no unforeseen hazard will exist during either normal or abnormal operating conditions. No circuit will flash over during a lightning surge; no equipment case or supporting structure will rise dangerously above ground potential; and no part of the building framework or the substation fence will rise above ground potential. This also implies that no two adjacent pieces of equipment or structures will reach dangerously different potentials. The power man ties all his equipments, structures, and fences together by a "safety" grounding grid network that is grounded at frequent intervals. The provisions for safety will thus reinforce the provisions for proper relay action. He is concerned with gross amounts of power. A lightning surge may reach many thousands of amperes, as will a fault current in a major system. If the grounding takes care of these needs, the chances are that it meets all the other requirements, such as drains for static electricity or suppression of radio interference.

### City power distribution systems

City power distribution systems are special. The high- and low-voltage lines and the transformers are generally all pole-mounted along the city streets. The public is intimately concerned, and so safety is the prime factor governing this grounding. It is also necessary to obtain rapid relay action to disconnect faulty equipment. If a line should fall to the ground or a truck knock over a pole, it is vital that all live parts be disconnected as soon as possible. Again, such action has to be initiated by relay operation.

Tests have shown that a ground rod driven eight feet into representative Midwest soil may have a resistance to true ground of from 50 to 100 ohms. If rods are used to obtain ground connections on distribution systems it can be seen that many must be used in parallel. Suppose that ten rods, spread over an adequate area for greatest effectiveness, are used. The resultant impedance to true earth might be 5 ohms. If a fault occurred so that a 440-volt line short-circuited to the ground system, a dangerous condition could result. The current flow according to Ohm's law would be 88 amperes. It is quite likely that the particular secondary system would be fused for 100 amperes or more; hence no fuse would blow and the entire ground system, according to Ohm's law, would be raised above true ground potential by 440 volts. This is not a good situation, and can usually be avoided by having grounding connections to water systems and other supplemental structures.

A special condition exists on the standard 120/240-volt single-phase or 120/208-volt three-phase 60-c/s secondary distribution system used for household electric service. Here all the customary requirements for safety and continuity of service exist. In addition, the lines must not carry radio interference into the home. Of course, in the first place, every effort is made to prevent radio noise, that is, high-frequency unwanted random oscillations, from getting onto the lines. Occasionally, such noise may originate in the utility equipment. This only occurs, however, from faulty equipment, as standard utility items do not cause interference when operating properly. This source of noise is easy to locate and correct. Most noise on local secondary circuits originates on some customer's premises. Again, this is usually easy to isolate, and once isolated it can be eliminated.

Few engineers today appreciate how much noise would be generated by common appliances if filters were not almost always provided by the manufacturers of these items. In the late 1920s, for example, when radio was just coming into widespread home use and oil burners with electric controls and motors, electric refrigerators, electric fans, home tools, etc., were the latest gadgets, I was working for a Midwest utility. Part of my job was to cruise around in the evening with a portable radio, listening for radio interference on the complaint of customers that a squawking noise was recurring on their radios every 12 seconds all evening long. (Of course, they always blamed the power company and not the flashing sign on the street corner.) Today, all appliances must be tested for radio noise before being marketed. If this were not so, there would probably be no acceptable television reception. In the early days, one could often "hear" on his radio the sound of motorcars being driven down the street.

To provide safety, continuity of service, and freedom from radio noise, utilities must pay particular attention to the grounding of secondary circuits. These circuits also continue on into the customers' homes. There the rigorous local safety codes have their say as to what is done. Most communities use the National Electric Code [NEC (1)] or codes that are essentially the same. These require that the neutral conductor of the incoming service be grounded at the service entrance. The metal parts of the service equipment must also be grounded. These grounds must be made to a water system if one is available and suitable (if not, a driven rod is used). Consider then a typical utility secondary section about 600 feet long, supplied by a 50-kVA stepdown transformer. The secondary runs down an alley and supplies five to ten houses. The neutral conductor will then have the grounding furnished by the utility company plus supplementary parallel grounds. Such a system probably provides a resistance to true earth of a small fraction of an ohm. Many utility companies interconnect their secondaries, and particularly the neutrals, from one area to another; the neutral wire thus becomes a ground grid wire throughout the city. Much of the credit for the proper grounding of utility systems stems from the work of S. B. Hood. He was active during the 1920s when grounding was a very "hot" topic among utility men. Figure 1 illustrates the grounding situation on a typical utility low-voltage system.

### Conventional grounding on a customer's premises

Grounding on a customer's premises is rigidly covered by local codes. If these are followed to the letter, an adequate system for safety will result. This is to be expected, as the codes are based on two requirements: safety to persons and safety to property. The latter requirement is concerned with the fire hazard (the National Electric Code is a bulletin of the National Board of Fire Underwriters). To meet the code will not necessarily ensure a complete elimination of radio noise on the system because, after all, radio noise is not a hazard to personnel nor is it likely to damage property. Conversely, grounding to eliminate radio noise could be completely successful and yet not meet the safety codes. In fact, it is quite conceivable that a purely "noise" approach to the trouble could result in practices on power systems that are hazardous. The following are two typical examples:

1. A "noise" engineer wishes to have certain small equipment power-source circuits remain ungrounded. He could enclose these circuits in a cage for safety, but would then have to ground the cage itself. This latter could be done by connecting the cage to an available power-circuit neutral. Such a practice does not meet the NEC philosophy; in fact, it would be a violation if this cage came in contact with some structure that happened to have a safety ground lead to earth.

2. A ground reference point in a radio circuit is required, and the chassis is used for this connection. This reference may also be connected to the power neutral.

Even if such a connection is made through a resistor to neutral, it is contrary to good "power" safety practice. Yet such practices have been used in radio and television systems available to the public.

The aforementioned examples show that a conflict may exist in the manner in which the two philosophies are approached. It is sometimes particularly frustrating for the electronics engineer. He may design an electronic device that works to laboratory perfection only to find that the power engineer, in adapting it for commercial use, has fouled up his grounding scheme. If the electronics engineer had a knowledge of power practice, this would not occur. It would seem to be more appropriate for the electronics engineer to learn power grounding practice than the converse. There are two reasons for this: Power grounding practice is relatively straightforward and simple; furthermore, it is compulsory. Radio-noise grounding requirements need not conflict with the former but may require additional refinements. It is, therefore, particularly important to understand the codes that may apply to a customer's premise.

Figure 1 shows how the power service may have many connections on the utility system neutral. However, for the individual customer, there is effectively a single-point ground. The code specifically requires the neutral to be grounded at the point of the service entrance equipment, but not anywhere beyond this point. Therefore, within the confines of his property, there will not be any circulating return earth currents; they must all travel by the insulated neutral conductor. The NEC recognizes two types of grounding: (1) a "system ground" for the electric system neutral or return wire, and (2) a "safety ground" for metal conduits, enclosures, or structures. The first type is used for conductors that "in normal operation" carry current; the second type is used for all parts that "in normal operation" do not carry current. A customer has but one of the former, but he may have many of the latter; in fact, in a steel building all equipment is automatically grounded by virtue of contact.

Figure 2 illustrates a typical grounding scheme a power man might apply to a flight vehicle check-out building. It is assumed there is available 60-c/s 277/460-volt three-phase Y power; and 60-c/s three-phase 120/208-volt, Y power. Four-hundred-cycle three-phase 120/-208-volt Y power is furnished by a motor–generator set. Direct current at 28 volts is furnished by a static rectifier. There will be one-point power grounding for each system: The 60-c/s power is grounded at the transformer secondary neutrals at the substation, the 400-c/s power neutral is grounded at the generator distribution panel, and the dc negative lead is grounded at the rectifier.

If other than single-point grounding is used, there will be stray return currents circulating in the areas between ground connections. Consider a scheme that might have been used in the recent past: the ac powers are correctly grounded but there are multiple grounds on the dc powers. In fact, the dc equipment components have all their negative return points connected to their respective chassis. Figure 3 illustrates how the returning direct currents may flow if the consoles are connected to power outlets by means of portable power cables that also incorporate an additional conductor for equipment safety grounding.

Suppose that an insulation failure occurs on a 120/208-volt conductor inside console 2 and that the transient short-circuit current peaks at 1000 amperes. This current must return to the substation through the safety grounding system. The dc neutral, however, is in parallel with this system and will carry current inversely in proportion to the impedances involved. It is also likely that some of the current will pass through items of dc equipment and return via the dc positive conductor to the rectifier and thence to ground. There is thus a possibility of damaging the dc equipment at the console as well as the dc power supply itself. Consider, also, another aspect that would present a normal operating condition. The dc return currents will always split up and return by all paths available, inversely as the respective path resistances. Thus stray direct current will appear in the safety grounding conductors of each power cable for each category of

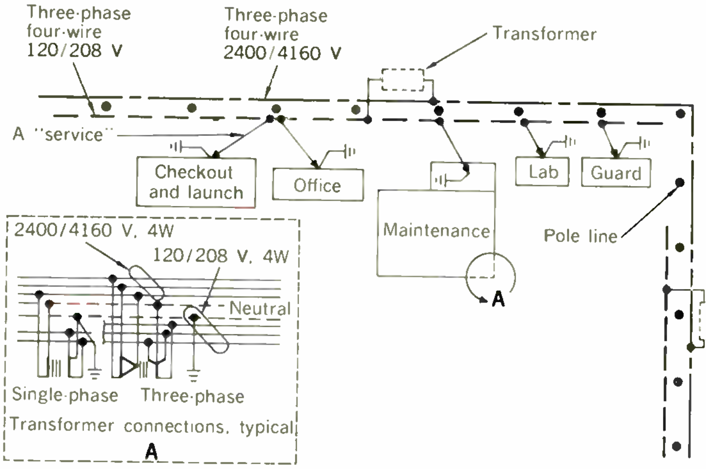
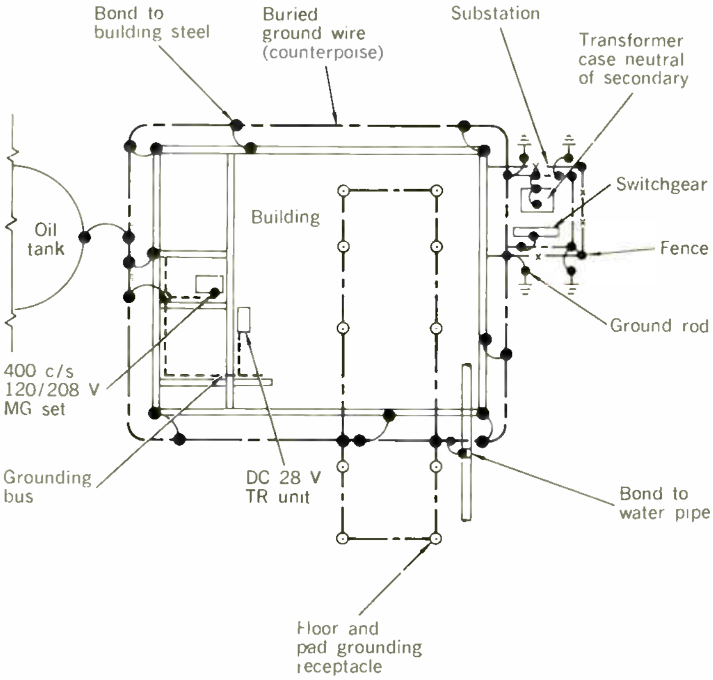


Fig. 1. Multiple grounding on utility distribution system.

Fig. 2. Illustration of a typical grounding scheme at a site.



power. If dc pulses exist, these may induce pulses into the ac circuits and possibly influence the accuracy of delicate metering circuits.

**Electronic systems**

Electronic circuits commonly imply circuits having very small currents, or circuits containing such elements as vacuum tubes, solid-state rectifiers, diodes, or transistors. True, some of these devices may produce large currents, but they are the exception. Here the types of devices implied are amplifiers, television, telemetry, computers, communication, etc., where most of the currents are measured in milliamperes and the frequency is high.

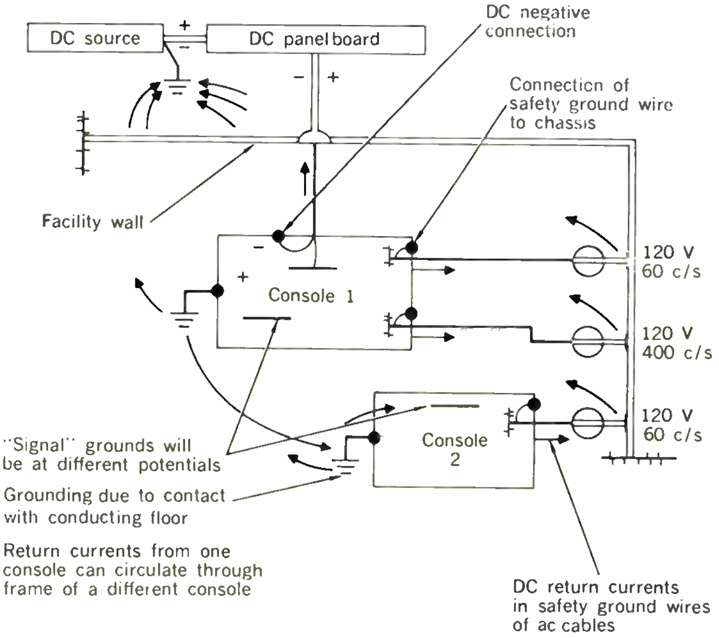
The minute currents are in the same range of magnitude in which possible interfering currents are found. As the communication engineer would say, the "signal-to-noise ratio" is not high. It is this that makes such circuits so susceptible to stray ground currents. The high frequency introduces a whole set of additional difficulties as compared with power-frequency circuits. The wavelength of 60 c/s is approximately 3100 miles; radio frequencies range from a few millimeters to a few hundred meters at most.

It can be seen that on power circuits (aside from a few long transmission lines) it is impossible to have a high-power voltage buildup on a grounded conductor. On high-frequency circuits, on the other hand, it is possible to get standing waves due to circuit reflections and high voltages on circuits that are, by power standards, well grounded. Thus electronic ground leads must be short and direct, free from loops and sharp turns, and, if possible, kept from paralleling adjacent circuits.

**Mixed systems.** Complex systems present the greatest challenge. In some systems large-current power circuits as well as minute-current and small-voltage electronic circuits exist in a confined area. There are many examples: self-contained items, such as the modern airplane, a space vehicle, a check-out van with all its power, electronics, air conditioning, and lighting; and integrated complexes such as missile check-out and launch facilities. The latter, of course, are designed for the specific space vehicle; thus the engineers responsible for the vehicle design naturally regard everything else as auxiliary equipment. They follow their own grounding philosophy. The check-out engineers design their gear to fit the vehicle and ground it, if possible, to meet the requirements imposed by the vehicle systems. All these items then are moved into a site complex built specifically for them. The complex, however, has been designed and built according to "best commercial practice" and is grounded to conform to all safety codes and power grounding practices. When the marriage takes place, the arguments start. This is followed by a long and costly process called "getting the bugs out." It is believed that much time and money could be saved if those involved in the two technologies would realize how the other man tends to think and act, and particularly if they would speak the same language.

A great deal has been written on the particular solutions for site grounding so that anyone who wants to dig deeper can readily do so (see bibliography). In all cases, it is basically a logical application of the principles already described.

Fig. 3. Multiple grounds on dc system, which cause circulating ground currents.



**Appendix**

In reading this article, it has probably become apparent that the words used to describe grounding practices are vague, and may even be misleading. Many of them fall in the "understood if known" category. This tendency, unfortunately, occurs in many of the articles on the subject. In fact, if you read several articles in quick succession, you may find the same term used with several different shades of meaning.

No science can be exact without precise measurements. Similarly, no practice can be accurately specified without precise terms. "Power" grounding terms are pretty well defined in the literature (particularly in ASA C42, where all grounding is covered in the section on "Generation, Transmission, and Distribution"), but these terms cannot always be carried over directly to aerospace practice, nor are they adequate to cover all facets of electronics application. What is needed is an updating of terminology to cover these new facets. Unfortunately, this will require "committee" action and thus take a long time. Anyone who tries to write a criteria specification for an overall aerospace complex that involves power, building, ground support equipment, and flight vehicles would welcome a precise terminology.

Some of the terms with tentative definitions are presented below. In many cases, the appropriate definition depends upon the context usage. In this case, the writer must make clear what definitions apply in his text.

*Bond, bonding.* There are 11 definitions under "bonding" in ASA C42, one of which applies to "aircraft, electrical." Probably no new definitions are needed provided the article is written to prevent misinterpretation. There have been instances where a reader associated the work with "glueing" or "fastening," which does not necessarily produce electrical continuity.

*Chassis ground.* The chassis used as a reference potential system or a connection thereto. The chassis itself may or may not be connected to earth. It may or may not be a true earth potential. (The context must, therefore, bring out whether the chassis as a "ground" is meant, or whether a connection from chassis to earth is meant.)

*Earth, earthed.* These are the British terms equivalent to "ground" and "grounded" in ASA C42. It is recommended that they be used when a connection as defined specifically in ASA C42 is meant, and where the context leaves room for doubt. In writing for a power document this may not be necessary. For an electronic document, use of the words "earth" and "earthed" are recommended.

*Earth ground.* (Same as "earth.")

*Equipment ground.* "An equipment ground is a ground connection to a noncurrent-carrying metal part of a wiring installation of electric equipment or both" [see bibliography, "National Electrical Code (NEC)" under "Codes"]. It is for the safety of personnel or for electrostatic shielding. (Also see "Safety ground" and "Static grounds.") In the aerospace industry the term "hardware ground" is sometimes encountered.

*Counterpoise.* This is defined in ASA C42 as a "system of wires or other conductors, elevated above and insulated from ground, forming the lower system of conductors of an antenna (for electrocommunication)". Webster gives "counterweight" as a synonym. Electronics men may think of a counterpoise as a system forming a "mirror image" of another system. For a power man, it may imply a bare conductor buried in the ground running along a power line. In this sense, it does form a mirror image of the overhead ground wire.

*Floating ground.* This term would imply a "reference ground" that was not earthed. It should probably not be used without some explanation.

*Ground systems of an antenna.* "The ground system of an antenna is that portion of the antenna, below the antenna loading devices or generating apparatus, most closely associated with the ground, and including the ground itself" (ASA C42).

*Impedance ground.* An earth connection made through an impedance of predetermined value. Usually chosen to limit the power short-circuit ground current to acceptable values. Same as a "resistance" ground.

*Negative ground.* The ground connection provided for the negative terminal of a dc power source. This connection may or may not go to earth. For large stationary systems, the earth is customarily used as the reference connection.

*Neutral ground.* "A neutral ground is a ground connection to the neutral point or points of a circuit, transformer, rotating machine, or system" (ASA C42). This may be the neutral of a 120/208-volt three-phase system, or of a 120/240-volt single-phase system; or, in the case of a 120-volt two-wire feeder originating at such systems, the conductor involved is that connected to the source neutral terminal. The term "neutral" is also frequently used in connection with single-phase systems originating at single-phase sources. In this case, the term "neutral" is applied to the grounded conductor.

*Reactance ground.* (See *Impedance ground;* also ASA C42.)

*Reference ground.* A reference point or bus usually close to, but not necessarily at, earth potential.

*Resistance ground.* (See *Impedance ground;* also ASA C42.)

*Safety ground.* "A ground required for human safety. This would be a low resistance earth ground, an 'equipment ground' " (NEC). It is an earth ground that is installed for this primary reason.

*Service ground.* A service ground is a ground connection to a service equipment and/or a service conductor (see ASA C42). (A "service" is the power feed running to a customer, or load, on a power utility system.)

*Signal ground.* A point within a circuit to which all signals within the circuit are referenced. It may or may not be connected to chassis, to the circuit "neutral," or to earth.

*Static ground.* A ground to reduce "static" electricity. To accomplish this, an earth connection is required. It is therefore an "earth" ground used for this primary reason. In flight vehicles, static grounds are used to connect equipment to the structure so that charges will not build up on "black boxes," etc.

*Static dischargers.* On flight vehicles, there are devices to drain off static charges ("static bleeders"). These devices will dissipate the charge on the outer surface of a flight vehicle while it is in flight and are frequently referred to as "static grounds."

*RF grounds.* Example RF line "grounded" to earth or chassis by a capacitor of low impedance for the working frequency, example radio noise filters, etc.

*Single-point grounds in electronic systems' shielded leads, etc.* Technique of insulating lead or circuit shield braids and provide a single point in system to ground. This technique eliminates "noise" currents in braids used for shielding.

"*Audio common*" *ground.* A single ground point in an audio system, such as an intercom, to ground the audio output. Technique eliminates hum being coupled into the audio output.

The author wishes to thank F. L. Lyons and D. H. Stephens for their assistance in the preparation of this manuscript.

## BIBLIOGRAPHY

The literature abounds with bibliographies and references on grounding. It is not our intention to give references that cover in depth some small area of interest only to a specialized few. A few basic references, it is felt, will be of interest to anyone working in this field. These alone should form a good basis for study of the subject.

*Codes and standards*

Bulletin no. 70 of the National Board of Fire Underwriters, National Electric Code, National Bureau of Standards, United States Department of Commerce, Washington, D. C.

Code for Protection Against Lightning, Handbook 46, National Bureau of Standards.

Grounding of Industrial Power Systems, IEEE Standard 142 (the "Green Book").

*Theoretical consideration*

Vogelman, J. H., "A Theoretical Analysis of Grounding," *IEEE Trans. on Aerospace*, vol. AS-2, no. 2, Apr. 1964, pp. 567–588.

*Application*

Clifton, W. W., "Design Your Grounding System," *Ibid.*, pp. 589–596. (This is a review based on experiences at Cape Kennedy. It shows applications to missile testing and launch vehicles.)

Robinson, T. A., "Creating the Grounding Criteria for an Overall Aerospace System," presented at AIEE Summer General Meeting, Denver, Colo., June 1962.

*Nomenclature*

American Standard Definitions of Electrical Terms, AIEE Publication, 1941 (same as ASA C42).

Robinson, T. A., "Definitions of Grounding Terms for Power, Electronics and Aerospace," *IEEE Trans. on Aerospace*, vol. AS-1, no. 2, Aug. 1963.

# Nuclear power in the world today

*Judging by reports at the 1964 international conference on atomic energy, nuclear power is now an established fact of life. The question no longer is whether it is competitive with conventional power sources but what production method will best meet economic criteria*

Louis H. Roddis, Jr. *Pennsylvania Electric Company*

In preparing a discussion of the role of nuclear power in today's world, one can choose among several alternative methods of attack. For example, one could repeat various frequently quoted cliches: "Nuclear power has arrived." "Nuclear power is competitive." "Nuclear power will take over as the prime source of energy by the end of the century." Or one could point out that, as with most situations today, there are so many "iffy" conditions in the nuclear picture that any statement could be rendered innocuous with qualifications.

Either approach to the subject could be substantiated by an arbitrary selection of material from the 739 papers presented last fall in Geneva at the Third International Conference on Peaceful Uses of Atomic Energy. An appraisal of the economic presentations indicates that different conditions, dissimilar accounting methods, and variations in the scale of power plant construction, plus currency exchange deviations, make the comparison of reactor capital and operating costs as much a matter of necromancy as of science. The situation is further clouded by the dictates of national pride, and by the prospect of lowered costs for potable water as a result of the combination of desalting operations with power generation.

In Geneva, discussions of costs so permeated the deliberations that the conference was labeled "too commercial." This tagging, as much as any other single development, indicates the growing importance of nuclear power in the world's energy markets. When engineers with their improved techniques begin edging out scientists and their theories, it is a rather strong indication that nuclear power has become a viable industry.

The Geneva conference was devoted not so much to the question as to whether nuclear power per se is competitive, as it was to the relative economic advantages of water and gas reactors. Within the limitations imposed by the conditions under which the leading types of such reactors have been developed, it is difficult to state that either has demonstrated a clear superiority on a world-wide basis. To use the vernacular of the advertising world—with gas reactors you pay now; with water reactors you pay later. That is, gas-cooled reactors have higher capital costs and are more competitive in situations where interest rates are low. Water reactors, with lower capital costs and higher fuel costs, are favored in areas where the cost of money is higher.

Advocates of pressurized water and of boiling water both offered some cogent reasons why their respective choice is best. Neither Britain's advanced magnox and high-temperature gas nor France's gas–graphite supporters succeeded in proving their claims of superiority. Canada and Sweden suggested the possibility of combining some of the better features of both in a heavy-water-moderated reactor using either a water, gas, or organic cooling system.

As a preface to reporting on the status of nuclear power in various countries, I would like to point out that the projection of a nation's image can have a vital influence on the technical path it chooses to follow. What for the scientist is discovery of the unknown and for the engineer is a means to great achievement, for the politician is a complex balance of social, financial, and international forces.

### Nuclear power abroad

England, with more installed nuclear power capacity than any other nation in the world, has concentrated on developing gas-cooled reactors that utilize natural uranium as a fuel. When its magnox series of nine stations with two units each is completed in 1969, the United Kingdom's nuclear generating capacity will be 5094 MWe.

Power costs for the ninth station, Wylfa, are forecast at 7.8 mills/kWh. This is approximately half the costs for Berkeley, the first station completed (in 1962). During the period between Berkeley and Wylfa, construction costs have declined from $493 to $280 per kilowatt. Despite these reductions the cost of nuclear power has remained approximately 20 per cent higher than that of conventional types, largely because of increasing interest rates in England. The British Atomic Energy Authority predicts that later versions of its advanced gas-cooled reactor now operating at Windscale should produce electricity at 4.8 mills/kWe as compared with 5.8 mills/kWe for the best conventional stations in operation in the mid-1970s.

The size of coal reserves and the monetary problems associated with importing fuel and equipment are fully as important as generating costs in the United Kingdom. The English have indicated a desire to add 5000 MWe of nuclear generating capacity by 1975 but have committed themselves only to the construction of two 500-MWe units at Dungeness. Bids have been received for water-cooled and advanced gas-cooled reactors as well as for an improved version of the magnox type.

France, which concentrates on gas cooling so that it can use natural uranium and thus remain independent of other countries for its fuel supply, put its first major power station in operation last year. The capacity of this station, EDF-2, nearly equals the total capacity of EDF-1 and the three dual-purpose units operating at Marcoule. France's nuclear generating capacity will again be doubled when EDF-3 becomes critical. The French are building a pressurized-water reactor in cooperation with the Belgians in the Ardennes, but their own projected reactors will be gas cooled.

The French policy is that each succeeding power reactor must show a decline in power costs, and they have indicated that they expect EDF-3 to be competitive with conventional fuel when it goes on the line in 1966. France also is trying to reduce her importation of oil and coal. Surprisingly, I found Polish coal being burned in a power plant close to Paris. And, of course, the North Sea gas field discovery is changing the whole fuel supply picture in northern Europe.

France's claim to superiority received a boost recently when Spain selected French companies to construct a 500-MWe station at Catalonia at an estimated cost of $200/kW and fuel costs of 2.25 mills/kWh. Undoubtedly these costs are lower than can be achieved in a fossil-fuel station, but other factors entered into the selection. Spain wishes to utilize natural uranium, of which it has sizable reserves, and France agreed to pay 25 per cent of the construction costs in return for 25 per cent of the power output.

Spain also has a nuclear power plant under construction, an American-type 150-MW pressurized-water station, at Zorita. Power cost estimates are considerably lower than the 7.1 mills/kWh quoted for a fossil-fuel station.

At the present time 72 per cent of Spain's power comes from hydroelectric stations as the development of steam stations has been hindered by the high cost of imported fuel. In Portugal, hydro provides an even larger part, 97 per cent, of electric production; the first Portuguese nuclear station, 250 MWe, is projected for 1975.

A somewhat similar situation prevails in the Scandinavian countries where Sweden is conducting extensive research into heavy-water reactors. A pressurized-water reactor is producing heat, as well as electricity, for an apartment building at Agesta. A 206-MWe boiling-water reactor is scheduled to be completed in 1969 at Marviken.

Italy has had a head start in nuclear power production—with three reactors, each a different type, producing 632 MWe by 1964. All have established good records for safe, continuous operation and presumably are producing power at costs lower than those for established conventional stations. (Much of Italy's fuel must be imported and therefore has a high price tag.) However, nationalization of the Italian electric industry has caused a postponement in selection of future generating stations.

Germany's progress in the nuclear field has been less than what might be expected from such a technically oriented country, partly because of conditions resulting from World War II and partly because German utilities were reluctant to proceed until the technical and economic aspects of the various concepts were more fully known. A 16-MWe boiling-water prototype has been operating at Kahl for the past two years and two other small prototypes are scheduled to become critical this year.

German utilities have contracted for five nuclear power stations; four of these will be water cooled and one gas cooled. Although they will be subsidized by the government, it is expected that these units will achieve power costs comparable to those of conventional generating stations in the same area.

Germany has launched a nuclear-propelled merchant ship and is conducting fast-breeder development at Karlsruhe with the support of Euratom.

Although Russia has announced a big nuclear program and estimates that by 1980 "tens of millions of kilowatts" will be produced, they have completed only two large power reactors. One is pressurized water and the other is boiling water; their combined capacity is approximately 300 MWe. It is reported that Russia is operating "a half dozen" dual-purpose reactors of up to 600-kW total capacity, primarily for production of plutonium. With the addition of two other prototypes, Russia's nuclear capacity is estimated to be 965 MWe, with an additional 1270 MWe under construction.

Russia's vast size and low population density, as well as national planning, militate against heavy investment in nuclear power facilities. The Russians have extensive hydro power resources that are transmitted to many parts of the country over long-distance high-voltage lines. The U.S.S.R. has more miles of transmission lines operating above 345 kV than have all the other nations of the world combined, a natural result of her geography. The country's need for local generating facilities is being met by the construction of plants capable of utilizing vast reserves of coal and oil.

Indications are that nuclear power has not become competitive in the Soviet Union but there is no question but that the Russians have the technical competence to build large nuclear power stations. One of the proofs of this competence was the unveiling at Geneva of the Romashka, a large land experimental direct-conversion thermoelectric reactor which was reported to have operated for 500 hours at temperatures of 3452°F, producing 800 watts of power.

In recent weeks the United States introduced a more sophisticated direct-conversion unit in the SNAP-10A, which is orbiting the earth as another significant achievement in space technology.

It is evident that both Russia and the United States have large staffs doing research in the direct-conversion field but much of this work is of a classified nature.

Delegates to the Third International Conference were less reluctant to discuss progress in fusion development, probably because it has been mostly of a negative nature. It was generally agreed that the problems of harnessing the power of fusion are many and complex, and nowhere near solution.

### Canada's heavy-water reactors

To return to fission, however, more recognition than at any time in the past was given at Geneva to Canada's long-time advocacy of heavy-water reactors. It has generally been recognized that heavy water offers low fuel costs with either natural or enriched uranium as a fuel. Offsetting factors have been the cost of heavy-water itself ($45/kg) and the difficulty of preventing loss through leakage. Progress has been made in overcoming both of these drawbacks.

Although Canada has abundant low-cost energy reserves, the more favorable hydro sites will be utilized within the next five to ten years. To carry present low-cost power to load centers, a substantial number of miles of EHV transmission lines are being built. The estimated cost of transmitting 1900 MW 600 miles on a 735-kV circuit is 1.4 mills/kWh.

The NPD prototype, utilizing natural uranium for fuel and heavy water as moderator and coolant, has been operating successfully for three years. Technology developed there is being incorporated in a 200-MWe reactor nearing completion at Douglas Point, and in a 132-MWe reactor in Pakistan. The Canadians expect to complete a 500-MWe twin-reactor station in Toronto by 1970 that will generate electricity for an estimated cost of less than 4 mills/kWh, which is considerably less than the cost of fossil-fuel-produced power.

Outside of the United States there are 42 nuclear power reactors operating with a net electrical capacity of 4730 MWe. An additional 25 reactors, with a net capacity of 7391 MWe, are under construction, and five, with a net capacity of 3102 MWe, are committed.

### The American picture

The United States has 16 reactors operating with a net capacity of 1157.8 MWe, and eight reactors, with a net capacity of 2884.9 MWe, are under construction. One of the latter, Hanford, became critical in December 1963, but will not produce electricity until generating equipment is completed late this year. The capacity of four nuclear stations that are definitely committed totals 2031 MWe. At least twice that capacity is involved in stations where utilities have indicated a possibility of going nuclear but are analyzing downward adjustments of coal prices before making a definite commitment.

American superiority in the nuclear power field is not evidenced unless you look behind these statistics.

The United States has compiled more operating experience with a greater variety of reactor types than any other country. It has also sold more reactors to foreign nations than has any other country. If operating reactors are compared on the basis of nation of origin rather than on the basis of operating location, it will be found that one half of the nuclear generating capacity operating and being built in the free world is of American design.

The reason for this leadership lies in the progress made under America's free enterprise system, the willingness of investor-owned utilities to conduct research and development work, the enlightened policies of the Atomic Energy Commission, and the competition among manufacturers to build a superior product at a lower price.

The effectiveness of these factors is shown in the decreasing cost of pressurized-water reactors. Connecticut Yankee, a 462-MWe pressurized-water reactor scheduled for completion in 1967, will have capital costs of $174/kW and power costs of 5 mills/kWh, as compared with costs of $220 and 9.9 mills respectively for its predecessor, Yankee, which has 40 per cent of the new plant's capacity. The first Yankee reactor, on the other hand, had three times the capacity at one fifth the cost of Shippingport, the first commercial pressurized-water reactor.

Dresden, the first commercial boiling-water reactor, cost $245/kW and produces power for 9 mills/kWh.

Oyster Creek, a third-generation boiling-water reactor scheduled for completion in 1967, will have 2½ times the capacity of Dresden and will have capital costs of $138/kW and fuel costs of less than 4 mills/kWh.

The decrease in construction costs, a factor of scale and technical experience as well as of competition for orders, accounts for the fact that every utility planning a major new generating station must consider nuclear power. However, nuclear power's gains in reducing capital costs are being offset to some extent in an area where it always has had a decided advantage—in the cost of fuel.

Faced with the threat of nuclear power, the coal industry has been able to reduce its prices substantially as a result of economies achieved by the introduction of continuous mining machines, conveyor belts, roof bolting, draglines, and other automated devices. In the past seven years the industry has succeeded in increasing the tonnage output per man more than 50 per cent; in 1963 the average production per miner in the United States was 15.19 tons, or 4.6 tons more than in 1957, the largest per-man gain ever achieved in a comparable period. I am sure it is more than a coincidence that this gain occurred during the period in which the atom began to prove its potential as a primary source of energy.

The coal operators are to be complimented for their

Within the next five years the United States hopes to land two men on the moon and bring them back safely to earth. As part of the Apollo project. the vehicle destined to make the manned landing is the National Aeronautics and Space Administration's Lunar Excursion Module (LEM).

Vital to the success of the mission are the many electronics devices carried on LEM. Those electronic devices that must operate external to the vehicular confines of spacecraft have created radically new environmental problems for hardware designers. For example, the major problem in the design of the LEM antennas is to keep them from melting—in spite of the fact that the outside temperature is absolute zero!

The irony of the situation would undoubtedly be lost on an astronaut trying to maintain a critical communication link or perform an intricate rendezvous maneuver, but it does point out the radically different problems that the advent of space travel has imposed on the design of spacecraft electronic equipment.

Equipment, such as antennas, sensors, or even an astronaut's personal communication equipment, that must operate outside the confines of the spaceship encounters bizarre environments in their most virulent forms. Temperature extremes, thermal stresses, micrometeorites, and solar radiation are sample conditions that will be encountered. A more complete listing of the characteristics of the space environment is found in Table I.

It behooves electrical engineers responsible for space electronic equipment to be aware of the dangers in this new environment. Traditionally, the environment in which electronic equipment operated has been one of the factors receiving the least attention because of its constancy and stability. This is no longer true. Today, an electronic project engineer must be a leader of a team comprising electronics engineers, thermodynamicists, structural designers, metallurgists, and mission analysts.

## Temperature

Perhaps the most striking phenomenon encountered in outer space is the wide variation in temperature that can be experienced on spacecraft surfaces and externally located equipment. Temperatures and temperature gradients not ordinarily encountered in the operation of ground or airborne electronic equipment are "ambient" conditions for spacecraft equipment. On such hardware, not suitably externally protected or housed deep within the space vehicle in a controlled environment, these temperature extremes can wreak destruction. While designers of earthbound electronics are fighting temperatures that will produce system degradation, spacecraft electronic designers may be fighting temperatures that will cause their equipment to melt.

### I. Characteristics of space environments

**Thermal radiation fields**
- Infrared solar radiation
- Planetary albedo
- Planetary and vehicle radiation
- Cold space

**Other radiation fields**
- Solar: X-ray, ultraviolet, visible, RF
- Cosmic: gamma, X-ray, ultraviolet, RF

**Force fields**
- Gravity
- Acceleration

**Material fields**
- Micrometeorites
- Engine exhaust
- Particles: protons, atomic nuclei, etc.

Fig. 1. Sources of space thermal environment.

Conversely, extreme cold may be a problem. Although low-noise-level receivers, such as liquid-helium-cooled masers, may thrive in the vicinity of absolute zero, most electronic systems are "red lined' at about −70°F. Below this temperature, the gain of many transistors is almost zero. Components such as resistors or capacitors also have limited temperature ranges and beyond these limits their values may vary significantly enough to preclude proper system operation.

On both ends of the temperature scale, the ground or airborne equipment designer has a simpler environment to contend with. In addition, the space electronics designer finds himself in a "temperature paradox." A black box cannot simply be placed in a superinsulated enclosure anymore than a human being can. All other factors neglected, both would rapidly destroy themselves because of self-generated heat. The electronic equipment must, therefore, be exposed to its environment in some manner, but it also needs a great deal of protection. The problem is not as simple as putting on or taking off a sweater, depending on whether the temperature is 30° or 70°F. The problem is to put something on and keep it on regardless of whether the temperature is −250° or +250°F. As shown in Fig. 1, many factors give rise to the temperature extremes encountered. However, these extremes must be understood before any consideration can be given to means for alleviating them.

**The temperature of space.** To put it succinctly, space is cold. A passive structure alone and exposed to space—not receiving radiant energy from the sun or planetary bodies, nor kinetic energy from collisions with other bodies—will stabilize at the temperature of space: absolute zero. This is a hypothetical case. In actuality, external equipment may get as cold as −250°F; but because of the proximity of the space vehicle, it will seldom get colder.

By itself, extreme cold would pose no extraordinary design problems. Structures can be built to withstand it. Heaters and thermal shields could be employed to keep electronic components at temperatures between 0 and 160°F. When then is cold space a problem? It is a problem when periods of extreme cold are followed by periods of extreme heat. Large thermal gradients can be developed rapidly as, for instance, when an antenna quickly warms, as it may when it is suddenly rotated into direct sunlight or a rocket exhaust. These gradients can cause deformation of a carefully machined parabolic dish or cause expansion in a critical gear chain. For example, one edge of a parabolic reflector may be exposed to a 30-second rocket exhaust, and in this small time the exposed edge could rise in temperature from −60° to 450°F. In this same time period the opposite edge of the reflector may have risen to only 120°F. The resulting thermal gradient of approximately 330° could seriously distort or buckle an antenna that required a 0.040-inch surface tolerance to maintain its performance. The result may be antenna failure, or at least temporary loss of an important communication link.

No complete solution to the actual problems imposed by cold space can be found without considering the sources of heat that will be encountered by spacecraft equipment. The largest and most important source of natural thermal energy found in the solar system is the sun.

**Solar heating.** Solar surface power density has been estimated at 65 million watts per square meter. Because solar energy, like all electromagnetic energy, decreases as the square of the distance from the source, only a fraction of energy reaches the vicinity of the earth. At the earth's orbital range, solar power density is about 1400 W/m². This energy is further attenuated by the earth's atmosphere; however, at altitudes above about 110 miles, atmospheric attenuation is negligible and 1400 W/m² (442 Btu per hour per square foot) is considered to be the solar power density available to heat space equipment. Because of the inverse-square law, however, about 92 per cent more solar heating would occur in the vicinity of Venus than on a trip to the moon. Conversely, because Mars is farther from the sun than the moon is, about 57 per cent less solar heating would occur in the vicinity of Mars; see Fig. 2.

Notice that solar energy heats space equipment. Although the sun emits X-ray, ultraviolet, and RF energy, over 90 per cent of the radiated solar energy is in the infrared and visible portion of the spectrum. Hence, thermodynamicists consider the effect of solar heat inputs as being similar to heat from a fireplace or a hot soldering iron. The effect of X rays and ultraviolet emissions must also be considered; however, since they do not generally pose any thermal problems, they will be discussed in a later section. As a physical example of these effects, consider why it is possible to get sunburned on a cloudy day. Sunburn results from the sun's ultraviolet, not its infrared, energy. Clouds attenuate the sun's infrared energy, but pass much of its ultraviolet energy. Hence, not only is it possible to get badly sunburned on a cloudy day, but there will be no uncomfortable feeling of getting warm to warn of the danger.

Another fundamental feature of solar energy is that its heating effect is limited. A piece of electronic equipment—for instance, an antenna—orbiting the earth would theoretically achieve a steady-state temperature of about 1800°F, assuming it absorbed 100 per cent of the sun's incident energy and reradiated less than 1 per cent of its accumulated energy. It would get no hotter. Actually, because no physical material has these theoretical properties, an aluminum antenna not designed to minimize solar heating or radiate to space from portions not illuminated by the sun would stabilize at about 550°F, if we assume that the antenna is not generating any internal heat. If the antenna were dissipating electric energy, perhaps in a servo drive, the temperature could be 100° or more higher.

At 550°F aluminum has lost a good deal of its strength; and although other metals such as steel can withstand these temperatures, weight penalties are incurred. It is also almost impossible to keep electronic components at reasonable temperatures when an adjacent structure is this hot for any length of time.

The magnitude of the thermal problem is now becoming more clear. On the one hand, external spacecraft equipment tends to become very cold—about −250°F. On the other hand, when this equipment is in the sun it tends to get very hot—over 500°F. What can be done? Well, for one thing, not all materials absorb solar energy equally. It is common knowledge that black cars feel a good deal hotter than white cars after both have been sitting in the sun for several hours. This happens because the majority of the sun's thermal energy is confined to a narrow band of wavelengths between 0.1 and 3 microns and the white paint absorbs much less energy in this portion of the spectrum than does the black paint.*

As an electrical engineer uses bandpass and band-reject filters, so a heat transfer engineer can make use of different thermal coatings so that only enough thermal energy is let into his equipment to balance the thermal energy that is being radiated to space. This balance between heat in and heat out can be maintained at almost any desired temperature. As an example, consider the simple heat-balance equation that may be written between an antenna, the sun, and space. If the antenna is at equilibrium, then heat in must equal heat out. Thus we have

$$Q_{\text{in}} \text{ due to the sun} = Q_{\text{out}} \text{ due to radiation to space}$$

where $Q$ = heat flow in Btu/h.

Substituting for $Q_{\text{in}}$ and $Q_{\text{out}}$, we have

$$\alpha_s A_p S = \sigma \epsilon A T_A{}^4$$

where

$\alpha_s$ = solar absorptivity (normally between 0.1 and 0.9)
$S$ = solar constant = 442 Btu/h·ft²
$A_p$ = projected area of antenna facing sun
$\sigma$ = Boltzmann constant = $0.17 \times 10^{-8}$ Btu/h·ft² (°Rankine)⁴
$\epsilon$ = emissivity of antenna (normally between 0.1 and 0.9)
$T_A$ = equilibrium temperature of antenna, °Rankine

For this discussion, solar absorptivity $\alpha_s$ may be considered to be the ratio of solar heat absorbed to solar heat incident. Likewise, emissivity $\epsilon$ may be considered to be the ratio of actual antenna heat emitted to heat that would be emitted by the antenna if it were a 100 per cent efficient emitter.

From the foregoing equation it can be seen that the antenna temperature can be controlled within a factor of about 1.7 to 1 (the fourth root of 0.9/0.1) by use of a proper surface coating to either accept or reject the sun's energy. If emissivity is also varied, the temperature can be controlled within a factor of about 2.9 to 1.

Since the frequency at which objects radiate their maximum thermal energy is dependent on their temperature, not everything radiates between 0.1 and 3 microns as does the sun. Hence, while some thermal control can be maintained by judiciously choosing a material having the desired solar absorptivity, thermal

---

* Contrary to popular opinion, color has little to do with how much or how little solar energy is absorbed. Color as seen by the eye is reflected (nonabsorbed) solar energy in the visible portion of the spectrum. Hence it is not necessarily an indication of how much infrared solar energy will be absorbed. For example, "snow-white" snow is an excellent absorber of solar energy.

balance may be upset when there are other sources of heat, such as planets or spacecraft.

**Planetary radiation.** Planetary bodies such as the earth, the moon, or Venus, while not as hot as the sun, are considerably warmer than space. These bodies may be considered to be in thermal equilibrium with the rest of the solar system. Consequently, they receive energy from the sun and exchange energy, depending on their temperatures and proximities, between each other and anything in their vicinity.

Two types of radiation from a planetary body are of importance in the thermal design of space equipment. The first is reflection of the sun's energy. This property is called albedo. It is the ratio of solar energy reflected to solar energy incident. It is important because this radiation is of the same frequency as direct solar radiation. Hence it may be added directly to a solar heat input and controlled by choice of solar absorptivity. The moon's albedo has been given as approximately 0.124. Thus a piece of gear on the sunlit side of the moon may receive

Fig. 2 (right). Solar flux density at planetary ranges out to Mars.

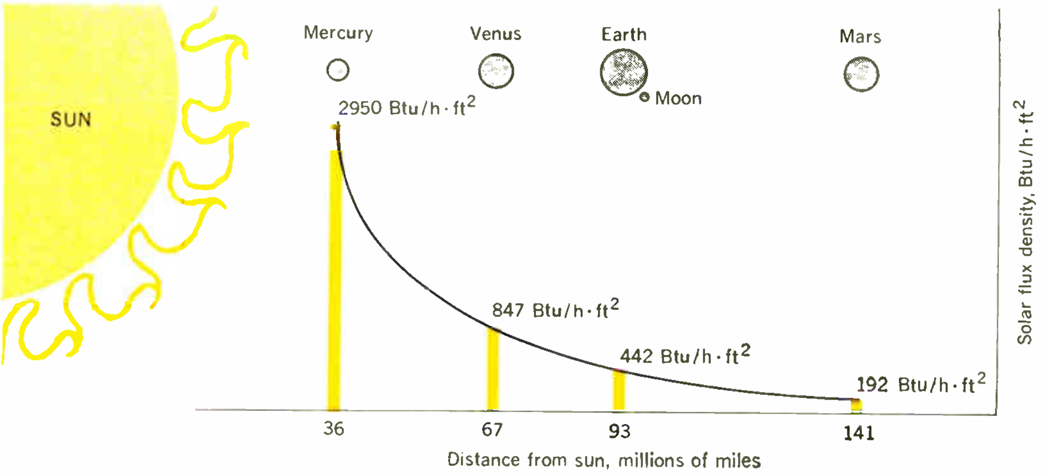
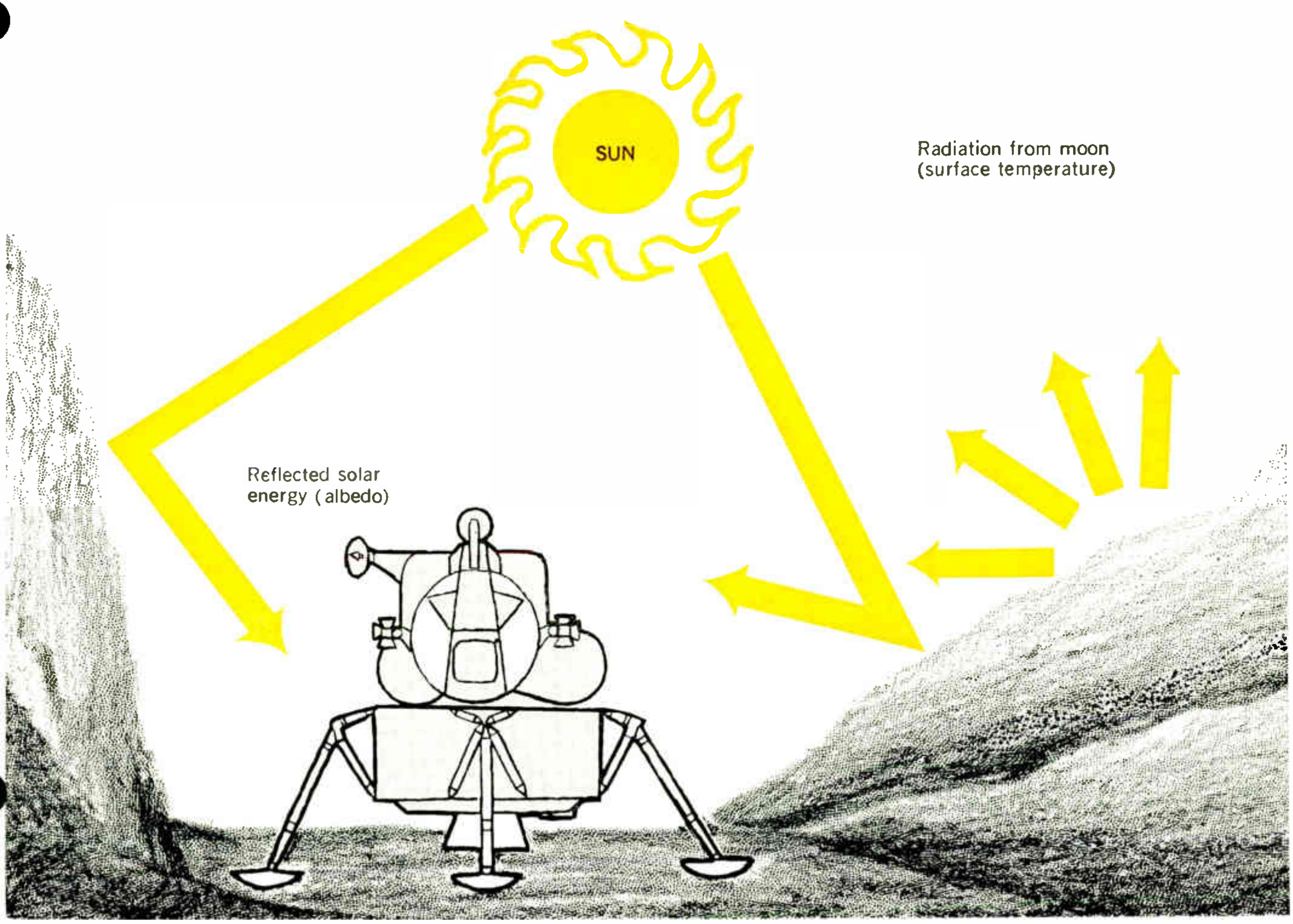


Fig. 3. Thermal environment faced by LEM on the lunar surface.



12.4 per cent more solar energy when it is on the moon.

The second important type of planetary radiation is the radiation due to the temperature of the planet itself. Because the body is not at the same temperature as the sun, it will radiate to cooler bodies and to space at a frequency dependent on its temperature. Therefore, a piece of space hardware, specifically coated to filter solar inputs, may find itself literally burning up from radiation from a planetary body. Planetary radiation is depicted in Fig. 3.

Naturally, the space hardware must be relatively close to the planetary body before it is thermally affected; how close is dependent on the size of the body and the criticality of the thermal design. From 5000 miles away, the moon subtends an angle of approximately 19° and occupies less than 1 per cent of the solid angle about the vehicle. Hence, beyond this range and most likely even at less than this range, lunar thermal effects will be negligible.

An example of a problem involving planetary heating arises in the case of a hand-held television camera for a manned lunar landing. The camera may come as close as 4 feet to the surface of the moon. Assuming a landing at high noon, the bottom of the camera will face the hot lunar surface, +250°F, while the top of the camera will face space and the sun. A passive cooling design for the camera would have to filter out solar radiation on the top and the sides (to filter reflected solar energy) of the camera case, insulate the bottom of the camera from lunar heat inputs, and radiate internally generated heat outwards from the top of the camera into space.

Such a passive thermal design can be accomplished, but the resulting camera is very sensitive to its relative orientation; that is, if the astronaut holds the camera upside down too long, things will get awfully hot inside. An active solution for the camera could be to use a "water boiler," which consists of a water-filled jacket surrounding the camera. As the jacket gets hot, the water boils off, and as long as the water lasts the camera will stay cool. It is of interest to note that since the boiling point of water depends on the pressure, the water temperature can be maintained at any desired temperature above freezing by use of a pressure valve. For example, the pressure might be set so that the water boils at about 100°F.

The camera should also operate if the landing occurs on the dark side of the moon. Here the lunar surface temperature may be as low as −250°F and there will be no solar or lunar heat inputs. Under this condition the camera must conserve its internal heat and inhibit radiation to the moon and space. This, of course, is just the opposite of what was required in the lunar day landing.

In addition to planetary bodies, there is another similar source of thermal radiation that must be considered: radiation from the spacecraft.

**Spacecraft heat transfer.** Because any externally located equipment will constantly be close to the spacecraft, there will constantly be an exchange of thermal energy between the equipment and the vehicle. Heat will radiate from the warmer body to the cooler body as in the case of planetary radiation, and, if there is physical contact between the equipment and the vehicle, there will also be conductive heat flow.

Conductive heat flow can be controlled in the structural design of the equipment–vehicle interface. The larger the interface area and thermal conductivity of the interfacing materials, the greater the conductive heat flow will be. Conversely, the effects of conduction can be minimized by using thermal insulation material or by increasing the length of the conduction path between the equipment and the vehicle. As a rule of thumb, in the thermal design of space equipment it is usually easier to decrease the effects of conduction than to increase them. An explanation is that metal-to-metal thermal conductivity (more specifically referred to as thermal contact resistance) is a function of many surface properties that are different for a vacuum environment than for a ground environment. For example, in a vacuum, a tenfold increase in the contact resistance between two metals could result from the loss of the conduction effects of air between the contacting surfaces. Thus seemingly solid bolted assemblies, having relatively large contacting surfaces, may have a surprisingly large thermal resistance between them in a vacuum. Since relatively little theoretical information or data are presently available to determine thermal contact resistance in a vacuum, it is important that any assumption made in the thermal design be verified experimentally.

The amount of radiation heat transfer between the equipment and the spacecraft will depend on the temperature difference between them. On a typical mission to the moon, the spacecraft skin may vary in temperature from −250° to 250°F. This temperature excursion is similar to that experienced by the surface of the moon. The spaceship, therefore, will impose thermal problems similar to those encountered on the moon.

As an example, consider a star tracker that is being heated by the sun on its front surface. To maintain its associated electronic equipment at a reasonable temperature, say 160°F maximum, the thermal designer may want the rear of the tracker to radiate excess heat to space. Since space is at −460°F, it would not be particularly difficult to achieve this condition. However, if between the back of the tracker and space there were a portion of the spacecraft skin at a temperature of 250°F, there would not be sufficient heat flow out of the tracker's electronic equipment, and failure could occur swiftly.

Although spacecraft are not usually designed to have skin temperatures higher than 250°F, "hot spots" occur frequently. For this reason, external electronic equipment locations must be chosen carefully. Rocket engines are perhaps the worst offenders when it comes to generating local hot spots. Nozzle temperatures may reach 5000°F and skin temperatures in the vicinity may be as high as 250°F. Rocket engines also cause thermal problems, but in a different way.

**Engine exhaust.** Perhaps the most severe heat source that may be encountered by external electronic equipment is the exhaust plume from a rocket engine. Although the equipment is generally not located in the vicinity of main engine exhausts, spacecraft such as the LEM carry auxiliary reaction control engines for vehicle maneuvering. Since these engines must fire in all directions, it is physically impossible on spacecraft the size of the LEM to locate antennas, for instance, so that they will be uninfluenced by the RCS (reaction control system) engines.

Equipment located within exhaust plumes will be heated by the impact of supersonic exhaust products as they physically strike the antenna. These exhaust products may be likened to grains of sand from a sandblasting gun. They are not very warm themselves, but they travel so fast that when they hit something, a great deal of heat is liberated.

Location of the equipment within the plume is the most important parameter in determining the heating rates from the plume. Equipment closer to the center line of the exhaust plume will be heated more than equipment that is located off to the side of the plume's center line. Further away from the nozzle, heating rates also go down.

Heating rates of about 0.1 to 0.2 Btu/s·in$^2$ (100 000 Btu/h·ft$^2$) are not uncommon on antennas located 10 feet from a rocket nozzle. This heating rate is approximately 240 times that of solar heating, and withering temperatures of about 1500°F can be reached in less than a minute. Needless to say, an antenna made of conventional materials and not designed to withstand or minimize plume heating would melt under these conditions. Even if the antenna structure is built to withstand these high temperatures—for example by the use of stainless steel instead of aluminum—external electronic devices such as servomotors, amplifiers, and tracking components must be suitably protected to prevent them from exceeding their maximum operating temperatures.

Another factor which must be considered when we deal with engine exhaust is plume torque. A single LEM reaction control engine of 100 pounds thrust will exert approximately a 50-inch-pound torque on the 24-inch-diameter parabolic antenna located 10 feet from the nozzle. Although this load is not overwhelming, it should be considered in the antenna gimbal design.

Having successfully contended with the effects of temperature, the designer must now face the problems imposed by the vacuum environment.

## Vacuum effects

Space is a vacuum far better than man has ever achieved on earth except for small limited laboratory experiments. The gas pressure beyond the vicinity of the earth (greater than about 7000 miles) is smaller than $10^{-12}$ mm of mercury, or about $10^{-15}$ of surface atmospheric pressure. At these densities and the corresponding temperatures the heat energy contained cannot significantly affect the spacecraft or its equipment.

The more insidious effects that must be examined and considered in design are those of sublimation and decomposition of materials under these pressure and temperature conditions.

Sublimation of solid materials is analogous to evaporation of liquids. At the combination of pressures and temperatures encountered in space even such metals as cadmium and zinc, which are quite commonly used for plating on earth, sublime at rates (0.040 in/yr at 250°F) that are unacceptable. On the other hand, aluminum would have to be in its liquid state (above 1490°F) to evaporate at the same rate.

One might wonder what would happen if the more volatile elements, such as zinc, were contained in an alloy. In most cases the diffusion rate through the alloy base metal is small and the net sublimation rate is much smaller than for the pure volatile element. In inorganic compounds the process can take on even more complexity because, in addition to the above process, decomposition into simpler and even more volatile components may take place.

If a volatile material is used in a vacuum enclosure that has sizable temperature extremes within the volume, some of the volatile materials at higher temperatures may sublime and roam around the enclosure, only to deposit themselves on cooler surfaces and create havoc with devices such as relay contacts.

In general, surface coatings such as paints or platings can be an effective preventative. For metallic surfaces conversion coatings such as oxides are possible. In the case of organic materials—particularly those with long polymer chains, such as polystyrene—the failure mechanism is the breakdown of long chains into smaller, more volatile components. This breakdown is partially caused by residual catalysts which, if they remain in the polymer, contribute to decomposition. Some degradation inhibitors are available.

**Property changes of materials in vacuum.** In the case of inorganic materials it appears that for the LEM mission there is not much chance of degradation of material properties. Organic materials, on the other hand, must be carefully selected so that polymers containing volatile plasticizers can be avoided. Small losses of plasticizers can cause severe embrittlement. Also, care must be exercised in the selection of foamed polymers, since they are inclined to breaking caused by internal foam pressures in a vacuum environment.

Cold welding, the phenomenon whereby adjoining metal surfaces that are clean and free from absorbed films of gas weld together into a single piece when the proper temperature and vacuum conditions exist, is another source of potential trouble if ignored.

**Lubrication.** Last but not least, the problem of lubrication in a vacuum must not be ignored. The proper treat-

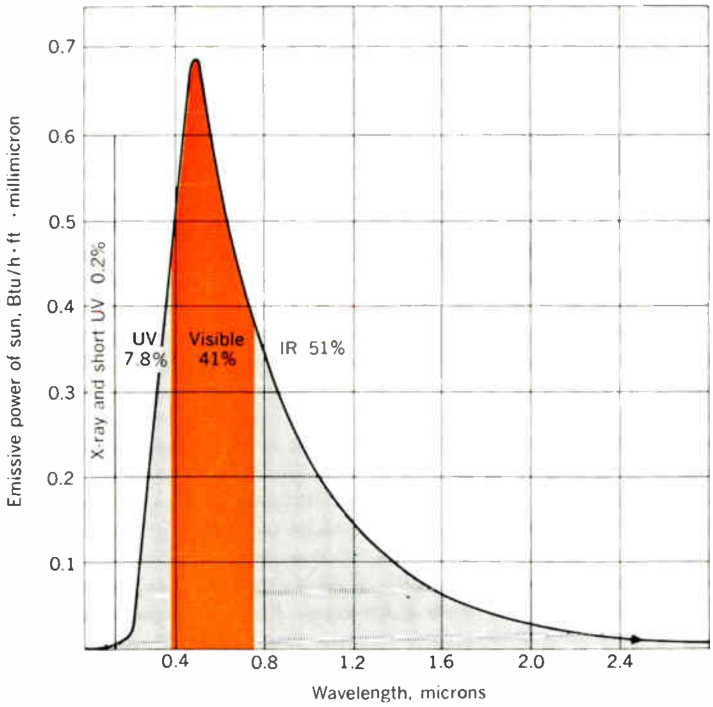


Fig. 4. Distribution of solar energy in the wavelength region below 30000 Å.

ment of surfaces that move in contact with one another—whether by enclosure, by hydrodynamic or solid lubricants, or by platings—must be assured.

### Radiation

Less than 1 per cent of the sun's radiated energy is in the X-ray, short-ultraviolet, and RF portions of the spectrum and, therefore, is not important as a source of heat. However, this radiation, especially the ultraviolet, can be harmful to electronic components, surface thermal finishes, and optical systems, and can change the structural properties of such organic materials as epoxy. Figure 4 shows the wavelength distribution of solar energy.

Since ultraviolet radiation in the range of 1 to 3000 Å will affect only the exposed surfaces of metals, metal housings will protect the electronic equipment inside from degradation—provided, of course, that the radiation does not destroy the thermal design by markedly altering the external thermal coatings.

Thermal coatings must be thoroughly investigated for their radiation vacuum properties before they are used. For example, if the emissivity of a coating should decrease from 0.86 to 0.77, a 10 per cent change, a 3 per cent temperature rise could ensue. This rise, if not anticipated, could be enough to cause component overheating. The key to success is to account for radiation deterioration in the initial design.

In the case of lenses for optical systems, two types of radiation damage are common. One concerns the thermal energy absorption of the lens, again a temperature problem; the other concerns the transmissivity of the lens, which is a measure of how well energy, such as light, is transmitted through the lens to internal sensors. Here, knowledge of the lens characteristics is most important, so that deterioration can be compensated for in the initial design. Careful selection of materials whose optical and thermal properties are critical is of prime importance, in order that the system can be initially "overdesigned" with a deterioration safety factor.

### Micrometeorites

A final item of concern in the design of external equipment is the ever-present threat of damage by meteoroids, those wanderers of space that could cause trouble for the astronaut and his vehicle.

Meteoroids, which are small particles of cometary origin, can be found in various densities and various sizes throughout the solar system. Should a spacecraft encounter meteoroids, three effects may cause equipment damage. The first is erosion, which may cause deterioration of thermal finishes, solar cells, optical systems, or even microwave radiation surfaces, such as antennas. The second effect is penetration of equipment housings or skins by the particles. This can cause internal mechanical damage to equipment similar to the effect of firing small pellets into electronic gear. The third effect is spalling of material from the inside surfaces of equipment housings, which can cause damage similar to actual penetration of the housing by the meteoroid.

Two factors must be taken into account in order to assess the probability of meteoroid encounter and to determine the amount of impact protection required. The first is that distribution of meteoroids in space is nonuniform, and the second is that meteoroids vary in composition, size, and density, as well as in relative velocity.

The largest type of meteoroids, called meteorites, need not concern us. Although these particles often exceed a gram in mass, they occur so infrequently (less than $10^{-14}$ particles per square meter per second) that a commercial airliner is just as likely to be hit by one of them as is a spaceship. A smaller class of particles, generally smaller than one gram, also exists and may be five to ten times more prevalent than the previously mentioned meteorites. Again, however, it is statistically unlikely that a spacecraft will encounter them and, in any event, missions can be planned to avoid known, relatively large concentrations of these particles.

Micrometeorites, which comprise the third category of meteoroids, are a fine dust which has been brought into heliocentric orbit by solar radiation and pressure drag. They pervade all space in varying densities. Most of these particles, which range in size from $10^{-3}$ to $10^{-13}$ gram, are in orbit around the sun, but a significant amount of them have been captured by the earth and pursue it in its travels. They orbit about the earth and slowly

**II. LEM antennas and critical environments**

| Function | Name | Type | Critical Environments | Manufacturer |
|---|---|---|---|---|
| Communication, LEM to earth | S-band omni no. 1 & no. 2 | Fixed | RCS engine exhaust | Grumman |
| Communication, LEM to command module | VHF omni no. 1 & no. 2 | Fixed | RCS engine exhaust and vibration | Grumman |
| Communication, LEM to astronaut | VHF LEM/astro | Erectable | Lunar ambient and vacuum | Grumman |
| Communication, moon to earth | S-band erectable | Erected by astronaut | Lunar ambient, vacuum, solar heating, secondary ejection | RCA |
| Communication, LEM to earth | S-band high-gain | Gimbaled | Cold space, solar heating, RCS engine exhaust, lunar ambient, vehicle, ambient, vacuum, vibration, radiation, micrometeorites | Dalmo Victor |
| Navigation, LEM to command module | Rendezvous radar (X-band) | Gimbaled | | RCA |
| Navigation, LEM to moon | Landing radar (X-band) | Gimbaled | | Ryan Aeronautical |

spiral inward. Micrometeorites are capable of causing damage, mainly by erosion, and should be considered in the design.

Several approaches to protective systems exist. One means for protecting optical surfaces is to provide shutters, which are removed for operation. This technique, however, is limited only to systems that operate at intermittent intervals and are fairly small in size. Another approach is to cover the surfaces to be protected with thin transparent films that can be removed and replaced as they become degraded by erosion. Again the disadvantages are obvious. Probably the approach most likely to succeed in maintaining surface thermal finishes is the simplest one—that is, to put the finish on as thick as possible.

It is perhaps obvious that no superior method or methods exist to protect against meteoroids. The larger meteorites, those that could cause catastrophic damage upon collision, cannot be protected against at the present time unless weighty shields are carried. Nevertheless, they pose no real threat because of their rarity. Protection against the smaller particles, micrometeorites, which can cause erosive damage to optical and similar surfaces, can in some measure be achieved. The important thing, however, is that even if 100 per cent protection cannot be attained, the possible degradation must be considered in the design.

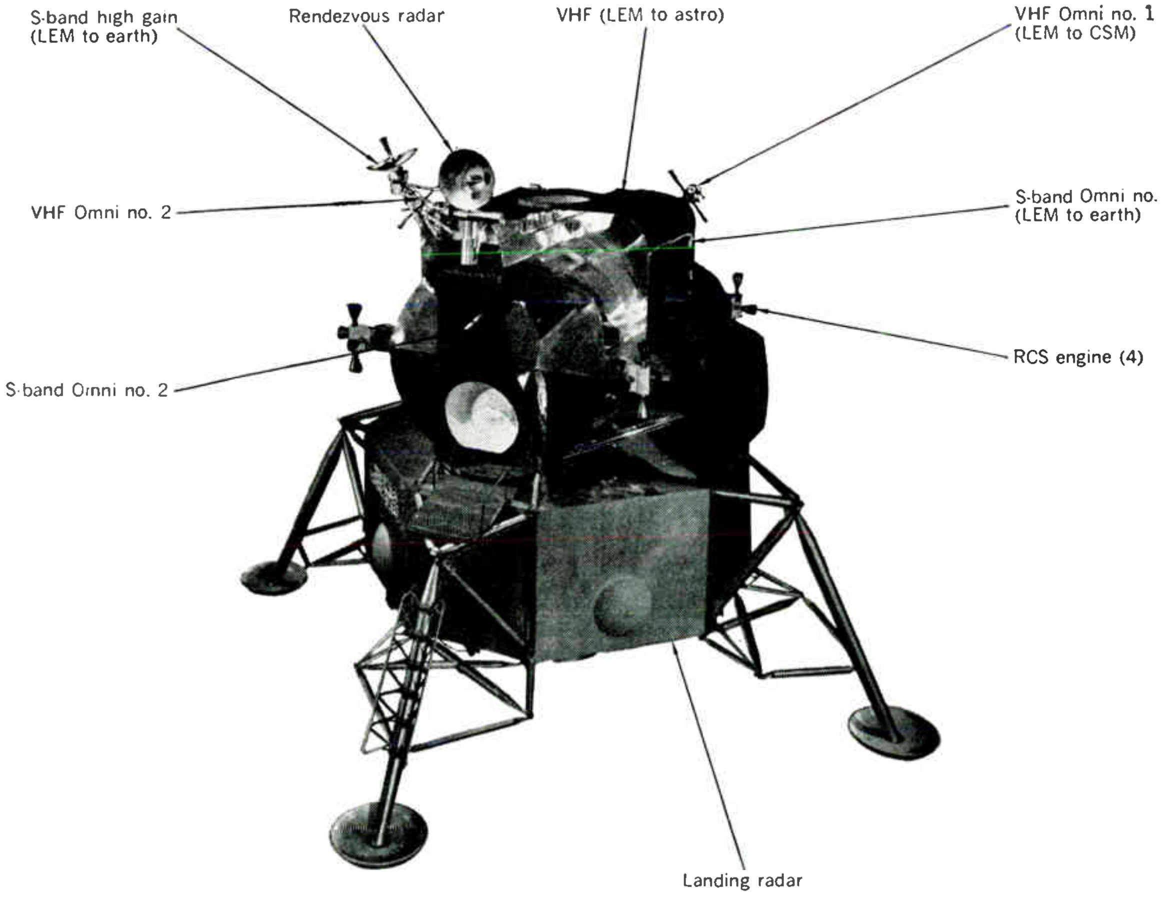
## LEM antenna system

The LEM antenna system offers several examples of external electronic components that must be designed for all the environment discussed so far. The vehicle antenna system runs the gamut of the RF spectrum from VHF communications to X-band radars, and in addition to the radiating surfaces and microwave components, the antenna assemblies include such electronic gear as motors, gimbals, servoamplifiers, and rate gyros. Finally, LEM also carries a 10-foot parabolic antenna that will be erected by an astronaut when he reaches the surface of the moon.

The initial designs of all these antennas are complete. At this stage of LEM development, however, qualified equipments have not yet been built. Undoubtedly as a result of the extensive reliability assurance and prototype environmental test program, modifications will be made. Nevertheless, the major environmental problems facing the antenna system have been met and solved. The ensuing discussion, while not a final or complete description of the antenna system, does serve to illustrate the design techniques.

A listing of the antennas and their critical environments may be found in Table II. The major problem for most of these antennas is engine plume from the LEM reaction control engines. Because all of the antennas are at different locations with respect to the engine nozzles

Fig. 5. Lunar Excursion Module antenna system.



(see Fig. 5), the antennas will all be heated at different rates; consequently the LEM antennas are a metallurgist's nightmare. Beryllium, aluminum, stainless steel, and titanium are all used—each metal being carefully chosen to do a specific job.

The ground plane and radiative elements of the VHF omnidirectional antenna (shown in Fig. 6) are made of beryllium. The antenna elements are small, light in mass, and if made of aluminum would heat up to 800°F and fail during a 30-second engine firing. Beryllium, however, has a higher specific heat than aluminum, so during the same plume firing, the beryllium antenna only gets as hot as 600°F. At 600°F, beryllium still has excellent structural characteristics and is therefore able to withstand the shock and vibration conditions that are imposed by the LEM. For similar reasons, the S-band omnidirectional antenna (shown in Fig. 7) also will be made of beryllium.

As an example of an antenna employing electronic devices, consider the high-gain S-band antenna used for in-flight communication between the LEM and the earth. The antenna, a 24-inch-diameter parabolic dish, is mounted on a boom extending from the LEM. The antenna must constantly track earth throughout all the flight portions of the LEM mission when earth is within line of sight. Once lunar landing has occurred, the exploring astronaut erects a larger, 10-foot-diameter S-band antenna, and the high-gain in-flight antenna is deactivated. However, prior to lunar lift-off, the in-flight antenna is reactivated and once again must track earth until rendezvous with the command module is completed.

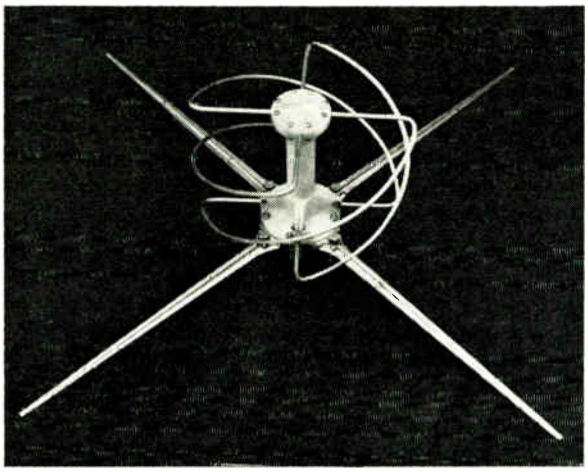
To accomplish its mission, the antenna assembly contains an integrally mounted earth sensor, which maintains earth track. Associated with the antenna's steering components are two servomotors, in addition to the associated electronic equipment and gimbal mechanisms.

During the long, three-day translunar flight, the antenna is just excess baggage, since all earth communication is via the command module. Nevertheless, the antenna's electronic equipment must be maintained between the temperature limits of −70° and 160°F regardless of the external environment.

The most critical external environments during translunar flight consist of cold space, solar heating, and the exhaust from the command module's reaction control jets. Protection against solar heating is achieved by packaging the electronic equipment in enclosures that reject solar energy while permitting the radiation of internally generated heat to space. A paint having a solar absorptivity of 0.28 and an emissivity of 0.86 will be used. The resultant antenna design is such that under the worst conditions of solar energy impingement, with 4 watts of stand-by power being internally dissipated within the servo's electronic components, the components will become no hotter than 155°F. Conversely, should the LEM vehicle constantly interpose itself between the sun and the antenna, the components will get no cooler than −60°F as long as stand-by power of 4 watts is dissipated. The decision to maintain stand-by power therefore hinges on the mission and whether or not the antenna will be exposed to solar heating.

A 45-second blast from the command module's RCS engines could easily raise the temperature of an unprotected aluminum reflector and feed to a temperature well above 1000°F, causing them to melt. All electronic components would be similarly destroyed. To prevent this, the antenna will be stowed with the reflecting surface, feed, and electronic equipment facing away from the rocket engines. The side of the antenna facing the plume will be made of stainless steel and thermally insulated from the front of the antenna. Now, even though the back of the antenna reaches temperatures of the order of 1000°F, the critical front of the reflector will be less than 100°F. Likewise, the electronic component will remain within acceptable temperature limits.

Following translunar flight, LEM will be detached from the command module and begin its descent to the moon.



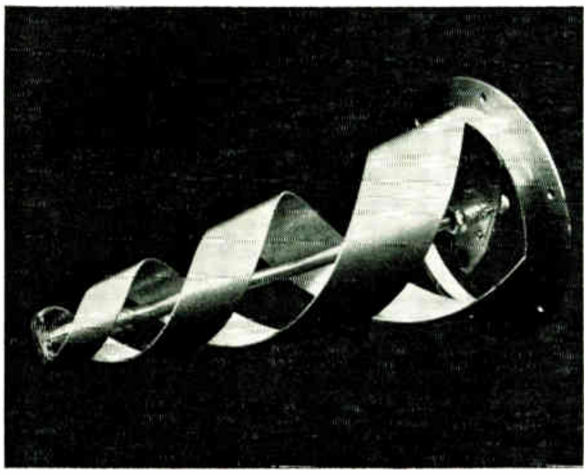


Fig. 6. VHF omnidirectional antenna used for LEM to command and service module communication. Because of RCS engine heating, the antenna will be made of beryllium. Antenna is right circularly polarized.

Fig. 7. S-band omnidirectional antenna for LEM to earth communication. Because of plume heating from command module attitude control engines, the antenna may be made of beryllium. Antenna is right circularly polarized.

The descent is expected to take about two hours, and during this time the antenna cannot be stowed and must operate. The antenna is now exposed to the LEM's RCS engines and can be dissipating 14 watts of internal power hile it is slewing. Fortunately, the LEM's control ngines will not be firing in the 45-second burst pattern of the command module. Rather, they will be pulsing at millisecond rates in order to maintain attitude control. To be safe, however, the antenna is being designed to withstand a 30-second continuous LEM firing. The antenna location has also been chosen to minimize the heating effects. Nevertheless, the radiating surface may approach 400°F and the rim of the dish may become much hotter. The dish itself, which is made of aluminum, can withstand this temperature. The rim, however, must be made of stainless steel; it is mechanically decoupled from the reflector so that it may expand and contract without buckling the dish.

Similar thermal problems occur during ascent, which nominally takes about 1½ hours from blast-off to docking. However, during lunar stay, even though the antenna is for the most part inoperative, a new environment exists due to the presence of the moon. If the landing is made at the subsolar point, which means the sun is directly overhead of the landing site, the moon will have a surface temperature of 250°F. Under these conditions even nonoperative servo electronics could exceed their maximum temperature during the 24-hour lunar stay. In addition, even though the servo electronics are protected from the LEM engine plume they increase in temperature some 10° to 15°F during a 30-second engine firing. Therefore, the temperature of the electronic equipment fter the 24-hour lunar stay must be sufficiently low to llow attitude control engine firings during ascent.

Several unique enclosures are being investigated to protect the servo electronics during lunar stay. One consists of a set of venetian blinds that have the slats drawn upward to prevent the electronic equipment from "seeing" the moon.

In addition to thermal protection, the other environments must be considered. Surface finishes that do not degrade under vacuum or radiation must be used. Insulation materials that are subject to degradation from radiation must be sandwiched between metals for protection. Electronic components are likewise protected from radiation by shielding, and special lubricants and seals are employed to protect bearings and gears.

The other LEM antennas face similar environments and are similarly protected. The one exception is the 10-foot erectable antenna. This antenna is stowed in the descent stage until it is removed from its compartment by an astronaut on the moon. Stowed, it is a cylinder, 3 feet long and 10 inches in diameter, protected from most of the environment by LEM.

After touchdown it is removed and carried by the astronaut to where there is an unobstructed view of the earth. The astronaut then unfolds the telescopic legs and releases the unfurling pin. Slowly, but surely, under the limited gravitional pull of the moon, it blossoms into a 10-foot parabolic antenna. An eyepiece will appear ear the central hub of the dish and the astronaut, peering through the sight, will aim the antenna at earth. A cable connection will be made to the antenna and moments later live television from the moon will be received at earth.

If it sounds simple, it's because it has to be. An astronaut encumbered by a spacesuit cannot be required to make a tedious assembly. The human factor is, therefore, a major consideration in the design.

The lunar environment is also a major consideration. The antenna must be mechanically sound regardless of whether the temperature is −250°F or 250°F. Solar energy must not be permitted to focus at the feed.

A mesh seems like a good choice for the reflector surface. Most of any solar energy striking the dish would pass through and not be focused at the feed. In addition, the effects of secondary ejection, if existent, will be minimized by the use of wire mesh.

Secondary ejection is an as yet undetermined phenomenon whereby meteors striking the moon will cause lunar pebbles to be thrown up. Since the lunar gravity is small, about 16½ per cent of that of the earth, these pebbles might actually be thrown up with such force that they will go into lunar orbit at altitudes of less than 6 feet. These pebbles, if they did not collide with anything, would continue to circle the moon. A mesh fabric could better withstand a collision with one of these pebbles than could a solid fabric. For one thing, the pebble might go through the mesh without hitting it. For another, if hit, the mesh would be less susceptible to tearing and surface deformation than would a solid fabric under the same tension. The mesh most likely will be a plastic, such as nylon, coated or painted with an electric conductor, such as aluminum. A mesh of this kind would be of lighter weight than a solid metal mesh and also would be less subject to stresses caused by thermal expansion and contraction.

## Conclusion

It should now be apparent that the space environment imposes unique and varied requirements on the design of electronic equipment, which are quite often outside the experience areas of the electrical engineer. What then is the best approach to a design problem involving electronic hardware to be used in space?

First, the engineer, when examining his electrical requirements, must realize that environmental factors will weigh much more heavily in the design of a device for space than they would for a similar device on earth, and indeed they might well place a limit on the achievable electrical characteristics for a particular device.

Second, the engineer must recognize the limits of his ability and be aware that engineering disciplines other than electronics will be needed to effect a satisfactory design. Consequently, the electrical engineer must actively seek help in at least the following areas:

1. Thermodynamics
2. Structures
3. Metallurgy
4. Mission analysis

Having brought these additional engineers on board, the electrical engineer must now act as team leader to insure continuity to the project.

The material used was gathered from many diverse sources; rather than acknowledge a few sources, to the exclusion of many, the authors make this general acknowledgment. In addition, they thank their many colleagues at the Grumman Aircraft Engineering Corporation—in particular, Haig A. Manoogian—for their assistance in the preparation of this article.

# Report on the A. S. Popov Society meeting

*F. Karl Willenbrock,* Editor IEEE

The Twenty-first Annual Meeting of the Popov Society was held in Moscow, U.S.S.R., May 12–15, 1965. It was attended by about 2000 persons from many parts of the Soviet Union and from 12 foreign countries, including Bulgaria, Czechoslovakia, Finland, France, the German Democratic Republic, the German Federal Republic, Hungary, the Korean Peoples Democratic Republic, the Malayan Peoples Republic, Poland, the United States, and Yugoslavia. There were 227 papers presented, of which 25 were by guests from countries other than the U.S.S.R.

In response to an invitation from Dr. V. I. Siforov, President of the Popov Society, an IEEE delegation of six members, headed by President B. M. Oliver, attended. The other delegates were Dr. Yardley Beers (U.S. National Bureau of Standards), Professor Paul D. Coleman (University of Illinois), Professor Claude E. Shannon (M.I.T.), Dr. F. Karl Willenbrock (Harvard), Professor Lotfi A. Zadeh (University of California, Berkeley). In addition, the IEEE was represented by members from a number of other countries.

The meeting was opened by a plenary session, which included three major addresses. The first speaker, Professor I. V. Brenev, discussed A. S. Popov's contributions to our knowledge of radio-wave propagation. This year is the 70th anniversary of his experiments in this field. The next address, by Professor Siforov, marked another anniversary, the 20th anniversary of the organization of the Popov Society. In this address Professor Siforov reviewed the activity of the last 20 years. The final address of the plenary session was by Dr. Oliver, who spoke on "Ideologies Versus Ideas in Human Progress." (For the speakers who did not speak in Russian, a consecutive translation method was used on a paragraph-by-paragraph basis.) Excerpts from his talk appear below.

The technical program consisted of more than 20 sections devoted to such topics as information theory; antennas and radio astronomy, waveguide components; semiconductor devices; radio receivers and amplifiers; wire communications and switching; television; plasmas; radio measurements and techniques; radio transmitters; electromagnetic wave propagation; broadcasting, acoustics, and recording; computing machines; quantum electronics; cybernetics; analog computation; history of radio; medical and biological electronics; bionics. These sections, which met simultaneously, consisted of from one to five sessions, averaging four papers per session. After the presentation of a paper a question period followed. These question periods were very active and in some cases occupied as much time as the original paper. In most cases, questions were given to the session chairman in written form during and at the end of the formal talk. When the speaker did not speak Russian, both the question and answer had to be translated, which proved to be a rather time-consuming process.

Although nearly all the papers were available in abstract form at the time of the meeting, it is, at best, very difficult to select those of particular significance. However, some comments about the subject matter of some papers will be made.

The information theory section included a paper on apparatus reliability and the possibility of using Monte Carlo methods of modeling, and also a paper on photon communication channels. In the antenna section, papers were given on the synthesis of antennas for a given field pattern and on log periodic antennas. The television section included a number of papers relating to criteria for measuring the quality of visual images.

In the electromagnetic wave propagation section, a paper described experimentally observed fluctuations in radio emission in cloudy atmospheres in the 0.8- and 0.4-cm wavelength regions. The quantum electronics section included papers on gas lasers and mode selection. A paper in the medical and biological electronics section related to a proposed method of automating microelectrode measurements of neurons, and a paper in the bionics section related to psychophysiological considerations in the design of control panels of computing machines.

From the diversity of subject matter covered, it can be seen that this meeting has some similarity to the technical program of the IEEE International Convention held in New York. Since the Popov Society now has a membership of 90 000 and 102 local chapters, it is not surprising that its principal meeting should cover a wide variety of subjects.

# Ideologies versus ideas in human progress

President Siforov, honored guests, and members of the Popov Society, let me first express my own gratitude, and that of the other members of our delegation, for the opportunity you have given us to be with you at this annual celebration of your distinguished society, and especially for the opportunity to address you and to share with you some of our ideas. I consider it a great privilege to be able for the first time in history to convey to you personally the greetings of the Institute of Electrical and Electronics Engineers and the 160 000 members it represents. While the majority of our members live in the United States, let me remind you that the IEEE is a non-national organization, with thousands of members all over the world. Let me also, therefore, invite all of you to become members, and to participate in a scientific community that knows no political boundaries. Although the ideologies of many of our members may be different from yours, we share a common interest in the field that sprang from the work of Popov, Hertz, and Marconi. Since our creative ideas are the same, in the IEEE we would find more to unite us than to divide us.

It seems to me that this is often the case: scientific truths unite people, while ideologies divide them. As scientists we are all conscious of the underlying physical laws which govern the universe. We are conscious of the majestic order of the cosmos and the insignificance of mankind in relation to it. We are conscious of the enormously complex chemistry of life and of the hereditary mechanism that we share. In the pursuit of truth we find basic commonalities among us, and only enough differences to add variety and interest to life. And yet on the political side we see these differences magnified to such an extent that we are blinded to our common needs, and our common destiny, and are in great danger of destroying ourselves. I think this danger arises because ideas sometimes become ideologies, or are incorporated into ideologies, and are assumed to have more universal importance than is in fact the case.

Because the word "ideology" has several meanings, and because I am using it in a somewhat special sense, I should give you an example of what I mean.

When Sir Isaac Newton discovered the law of gravitational attraction—that two particles attract each other directly as the product of their masses and inversely as the square of the distance between them—he found a great truth about the physical world. This one simple little equation explained so many apparently unrelated phenomena. It explained the motions of the planets in their ellipses, the ebb and flow of the tides of the sea. In short, what made the apple fall made the sun rise. He had a great unifying idea, Newton did, and although Minkowski and Einstein may have given us a deeper insight, Newton provided an excellent phenomenological description.

Many years later, Coulomb stated the law of electrostatic attraction. By then, the inverse square law, with its geometrical basis, was well established, so perhaps it required less creativity on Coulomb's part to postulate and then prove that charge played the role of mass in this case. But nevertheless Coulomb too had a great idea that is basic to all electrical phenomena.

But now suppose that the scientific world had become divided into two camps: the Newtonians who believed that Newton's law of gravitation was a universal truth covering all cases, and the Coulombians who believed only electrical forces ruled the world. Suppose in the end that there were worshippers of the masses, and worshippers of the charges, and you had to be one or the other. Suppose, in other words, that these two great scientific ideas had become *ideologies*, wouldn't science look pretty silly today? And yet doesn't this describe the political scene pretty well?

When an idea becomes hardened into an ideology or becomes distorted or overextended by the ideology, when it becomes so dominant in the minds of men that they seek to apply it where it is not valid, then that idea becomes dangerous. The danger lies not in the idea but in the context in which it is used. I am reminded of a conversation between two old friends that goes:

"How did you hurt your head?"

"My wife hit me with some tomatoes."

"That shouldn't have hurt you."

"These tomatoes were in a can."

It isn't the idea that hurts, it's the can it's in. I guess what I'm trying to say is: we must all beware of canned ideas!

Fortunately, in science, ideas seldom get misused. Ideas are tested and the exceptions noted. And so, when masses are present we remember Newton's law, when charges are present we remember Coulomb's law, and when both are present we need both laws. As my good friend John Pierce once remarked:

"A good law, holding, holds for the cases it holds for," and that is all we can expect.

Of course, not all ideologies are based on truth. There are many beliefs that have no basis whatever in reality. But it is safe to say that most of the major ideologies of the world contain some ideas that are true in certain circumstances, and that *in these circumstances* are even the most useful ideas. It is when we attempt to apply the ideas without regard for their relevance that we run into trouble.

It is when the proponents of an ideology seek to make physical law subordinate to political law that scientists rightfully object.

The trouble with ideologies is that they are not self-critical; they often reject truth and accept falsehood rather than adapt. When Galileo published his conclusions about the solar system, based on his observations with his telescope, he was tried as a heretic because the truths he had discovered conflicted with the ideology of the Roman Catholic church. Science almost suffered a disastrous setback, but in the end it was the Church that had to recant for the truth was there for all to see. Conversely, in modern times, false ideas of race and of heredity have been accepted because they were consistent with particular ideologies, but in the end truth has prevailed. In these cases ideologies have delayed human progress and caused great human suffering.

Now why are ideologies so rare in science and so rampant in philosophy, sociology, religion, and politics? I suspect it is because science has been built up from basic principles that were proven by experiment, and that each new principle has been tested for consistency before being accepted. This procedure develops a humility in the scientist with respect to his scientific beliefs. Nature is the final judge, not he. The testing of social, and especially of religious, beliefs is much more difficult. One may have to die in the attempt. In addition, and perhaps because we know so little about them, many social and psychological phenomena seem much more complex than the physical phenomena that concern the scientist.

In view of this, we, as scientists, should be at least as tentative of our political and social beliefs as we are of our newly established principles in science. We should be humble about them and ever on the alert to modify them. But alas, it seems to be true of human psychology that the *less* sure we are of a belief, the more obstinate and emotional we become in its defense.

Truth ultimately forces ideologies to adapt and in this fact lies hope for human progress. If we as scientists, and as citizens of our two countries, can help to avoid a collision of our ideologies, then we will be able to pool our ideas and work together for the future of the human race. I look forward to the day when we will no longer be competing in space, for example, but combining our efforts in this most exciting scientific adventure. I look forward to the day when I can address you, not just as fellow scientists, but also as friends in a common cause.

Let me on behalf of the Institute of Electrical and Electronics Engineers express to you of the Popov Society our admiration for your many wonderful accomplishments in electronics and the electrical sciences. May your efforts and ours continue to add to man's knowledge and ideas, and may they reduce the ideological differences that, for the present, separate us.

*B. M. Oliver, President IEEE*

Special
Conference
Report

# Report on the 1965 INTERMAG Conference

*Robert F. Elfant* *IBM Corporation*

The 1965 International Conference on Magnetics (INTERMAG) held in Washington, D.C., April 21–23, attracted more than 740 United States and foreign scientists and engineers. Of the 105 papers that were presented, more than 25 per cent were given by foreign authors.

Dr. J. J. Suozzi, the General Chairman, opened the conference. He noted that the newly formed Magnetics Group was sponsoring this and all subsequent INTERMAG conferences, and that the conference scope had been broadened to include all aspects of magnetics rather than being limited to nonlinear magnetics.

The chairman of the IEEE Magnetics Group, Dr. H. F. Storm, also addressed the meeting. He traced the history leading to the formation of the group, pointed out that the IEEE TRANSACTIONS ON MAGNETICS would be the official IEEE outlet for publishing all papers on magnetics, and urged that all scientists and engineers interested in this field join the IEEE Magnetics Group. He also announced that the 1966 INTERMAG Conference would be held in Stuttgart, Germany, on April 20–22, 1966.

S. Methfessel opened the technical sessions with a discussion of the "Potential Applications of Magnetic Rare Earth Compounds." He noted that the rare earth compounds have large anisotropies, large magnetic moments, and large optical effects. These effects only occur at low temperatures, and thus have not been fully exploited. With the advent of more sophistication in low-temperature systems, rare earth materials may play a stronger role in future technology.

Thin magnetic films received broad coverage at this year's meeting. M. S. Cohen described the use of Lorentz microscopy as a tool for investigating thin films. He indicated that Lorentz microscopy was directly applicable for measuring relative magnetization or film thickness, Curie temperatures, anisotropy constants, anisotropy dispersion, and wall motion switching.

In a three-paper series, R. Montmory *et al.*, E. Goto *et al.*, and F. J. Friedlaender *et al.* discussed the magnetic properties of exchange and magnetostatically coupled films. For the exchange-coupled films, the mechanism of coupling is similar to the mechanism that couples the spins in the film itself, whereas for the magnetostatically coupled films, the coupling mechanism is due to the stray fields resulting from the discontinuity of the spin system at the film surfaces. In either case, the coupled film pair shows interesting properties for memory applications because of its high tolerance to disturbing fields and nondestructive read properties.

A. V. Pohm *et al.* discussed a new approach to high-speed film memories. The significant aspect of this approach is the use of a ferrite keeper close to the thin film to reduce the effects of the stray fields due to the film itself. Pohm pointed out that use of the ferrite keeper results in reduced "disturb" sensitivity, reduced current requirements, and increased output signals. In addition, increased packing densities should be feasible, resulting in reduced memory delays.

J. E. Schwenker compared the properties of thin nickel–iron films plated on a wire with those of evaporated flat thin films. Schwenker found that the films had basically the same properties; however, because the plated films were thicker than the evaporated films, the plated films had a low coercivity. The measurements were made using an extremely sensitive, 1-mil-spot Kerr-effect probe.

The area of magnetic recording for home use was reviewed by M. Camras. Camras compared conventional disk recording with magnetic recording. He then reviewed the progress of magnetic video recording and indicated that video recorders in the price range of the average home owner should be available in the near future.

Magnetic recording for computer applications was reviewed by G. Bate. He concluded that because of the requirements for higher recording densities, new methods for producing recording media with thin coatings, high coercivities, and rectangular hysteresis loops are needed. Following Bate's presentation, several speakers discussed the limitations on recording density and a disagreement arose as to whether the limiting mechanism was self-demagnetization or demagnetization due to the stray fields of adjacent bits.

A new session entitled "Ferrite Microwave Devices" was introduced at this year's conference. The intent of this session was to emphasize the magnetic properties of the materials and how these properties are utilized in the microwave devices. E. Schlomann discussed the "Ultimate Performance Limitations of High-Power Ferrite Circulators and Phase Shifters." Schlomann pointed out that it is not only high power but also low losses that are

Dr. Elfant is manager of memory research, IBM Research Laboratory, Yorktown Heights, N.Y.

required for good microwave performance. He indicated that high-power materials can be obtained by most standard substitutions into the garnet lattice or by utilization of very small grain size samples, but that these techniques usually result in high losses. He suggested wider adoption of a "quality factor," which is the ratio of the power capabilities to the losses.

The field of superconductivity was keynoted by R. A. Ferrell. In his paper, "Are Room Temperature Superconductors a Possibility?," Ferrell discussed various models of superconductivity and concluded that on the basis of either a one- or a two-dimensional model, room temperature superconductivity is not theoretically possible, as had been previously suggested. On the basis of a three-dimensional model, however, a room temperature superconductor is a theoretical possibility; but the probability of finding one is quite remote.

P. Gamby and V. A. J. Maller described a superconducting continuous film store. They concluded that an in-line drive line arrangement is superior to the orthogonal connection as far as tolerances and certain disturb sequences are concerned. L. M. Troxel described a superconducting magnet used with a traveling-wave maser to obtain larger bandwidth. The bridge cell—a new superconductive memory cell for random access memories—was described by R. W. Ahrons. The essential feature of this device is that the persistent-current loop formed in the superconducting material is perpendicular to the plane of the substrate. Hence it is inherently capable of higher packing densities.

The switching properties of single and multiaperture ferrite devices were extensively discussed. Of particular value was an explanation of the quasistatic magnetization process in cross-magnetized multipath ferrite structures presented by D. S. Shull, Jr., and L. A. Finzi. This discussion was followed by a paper on the "Phenomenological Model of the Biax," by P. J. Nistler and K. J. Korowski. The general interest in this subject was so high that a special evening session was called to allow a continuation of these discussions.

In the area of magnetic amplifiers and logic devices, R. C. Barker analyzed a second-harmonic modulator. His analysis is based on the $d\theta/dt$ widening of the hysteresis loop. The analysis makes possible the comparison of amplifiers using different magnetic materials. D. I. Gordon *et al.* discussed the factors affecting the sensitivity of the ring-core magnetometer. The highest sensitivity is achieved when the self-demagnetization is smallest. The theory of a three-phase ferrosonant circuit was presented in a two-part paper by H. Kobayashi *et al.* and T. Hasumi *et al.* H. C. Bourne and T. Kusuda presented a theoretical analysis and experimental results on a three-phase magnetic amplifier.

Magnetic logic and materials were covered in a 9-paper session. An interesting paper on re-entrant hysteresis loops in a cobalt–ferrous ferrite by R. M. Glaister and I. V. F. Viney aroused great interest. D. H. Smith discussed a new magnetic shift register that employs domain wall motion. Basically, the shift register is simply a magnetic wire using a material with a pronounced re-entrant loop. A bit region is set up by an appropriate nucleating field, and it may be propagated along the wire with an applied field less than the nucleating field. The shift register described could be operated at about $10^4$ bits per second. F. R. Monforte, E. E. Newhall, and J. R. Perucca discussed the influence of material properties on the gain and threshold of a magnetic digital circuit. A small amount of $ThO_2$ inhibits the growth of large grains in the ferrite, giving an improved threshold and an increase in the amplifying capability of a balanced amplifier.

A. Apicella and J. Franks presented a new method for operating a toroidal ferrite core in an associative memory application. This method, called BILOC (biased logic core), utilizes a dc bias field which is transverse to the plane of the core in conjunction with a transverse switching field. Because of the presence of the bias field, increased NDRO signals are obtained, and it is possible to perform the "exclusive OR" function necessary for associative memories. B. E. Briley discussed a new device that uses a ferrite disk with a plurality of apertures. This device, called MYRA (myriad aperture), is proposed for read-only or mechanically changeable memories. Briley suggested that 16 words of 32 bits could be stored on one disk and accessed in 2 μs. The information is stored by the pattern of the wires passing through the disk apertures.

In two sessions devoted to combined magnetic and semiconductor devices, P. P. Biringer and P. C. Sen described a frequency tripler that utilizes linear reactors and silicon controlled rectifiers. This arrangement should have higher efficiency and be physically smaller than equivalent all-magnetic multipliers. A self-balancing dc comparator for 20 000 amperes was described by M. P. MacMartin and N. L. Kusters. Accuracies of ten parts per million when operating near rated current were reported.

The international character of the meeting was delineated by the large foreign attendance and number of papers by foreign authors. Present were representatives from such countries as West Germany, Japan, England, Israel, Switzerland, France, Poland, Holland, Formosa, Korea, and the Soviet Union.

**Correction notice**

In the article "Plasma thrustors for space propulsion" as published in the June 1965 issue of IEEE SPECTRUM on pages 36–45, an error appeared in the footnote on page 44. The footnote as printed stated, "This article is an expansion of a paper presented at the W. G. Dow (Dow Chemical Co.) Seminar: . . ." The words Dow Chemical Co. should not have appeared, as the Seminar was one held in honor of Professor W. G. Dow and had no connection whatsoever with the Dow Chemical Co.

**Special Conference Report**

# Report on the IEEE–NBS Particle Accelerator Conference

*J. A. Martin, R. S. Livingston* *Oak Ridge National Laboratory*

Specialists in the rapidly advancing field of particle acceleration met in Washington, D.C., on March 10–12, 1965, to discuss the physics and engineering aspects of a broad range of modern accelerator design and operating problems. The meeting, known as the Particle Accelerator Conference and subtitled "Engineering Problems of Accelerators and Related Devices," was the first of what may become a continuing series of national conferences on this subject. The Proceedings, scheduled for publication in the June issue of IEEE Transactions on Nuclear Science, will bear the label "First National Particle Accelerator Conference." That the conference was a success is borne out by the attendance of 760—twice the number originally expected; by the volume and tone of the comment—large and, with almost no exceptions, very favorable; and by the number of contributions—more than 230 abstracts submitted. Although nationally organized, the meeting also had an international flavor. Some 10 per cent of the participants were from other countries.

The week prior to the conference saw the Joint Congressional Committee on Atomic Energy holding hearings on long-range plans for high-energy physics. A "white paper"[1] had just been issued giving the Federal Government's proposed plans through 1981. In this atmosphere the conference proved to be particularly timely. Nowhere was this typified more clearly than in the banquet address of Prof. W. K. H. Panofsky who spoke on the relationship between accelerator builders and users and the management problems that arise with the increasing size of accelerator establishments. Professor Panofsky, Director of the Stanford Linear Accelerator Center, speaks with great authority on this topic. He is responsible for construction of the world's most expensive accelerator to date and also has made many important research contributions.

The first problem reviewed by Prof. Panofsky was the allocation of experimental resources and running time—that is, the division between the in-house staff and outside users and the question of choice of experiments. Someone must choose between important experiments that are sure to work and the more speculative but nonetheless well-conceived experiments that may yield nothing. Here he gave no answers, but remarked that layers of committees will not necessarily provide a solution. Great sensitivity is needed to yield a proper balance between conservative and liberal scientific philosophy.

The second problem discussed was the developing relationships between builder and user. There was a time, not very long ago, when there was little distinction between the two groups, but with today's large accelerators costing hundreds of millions of dollars the specialization so typical of other facets of modern society has arrived.

When an accelerator is completed, and also during its planning, a problem arises. Who is to be responsible for its management and use? At one extreme, we find complete user control with accelerator design, construction, and operation relegated to service functions; at the other, we find accelerator operation entirely by and for the benefit of its in-house staff. Both ways have been tried; neither produces a desirable state. With user control, which may result in absence of pride in accomplishment, there may be little motivation for the builder staff. Builder control, however, though suitable for small accelerators, fails by reason of being wholly inadequate for the large installations of both the present and the future.

Dr. Panofsky stated:

> "I find that the controversy now centering around the management and location of the next step in high-energy accelerators—namely, the construction of a 200- to 300-BeV proton accelerator—is much less a regional argument than a debate as to the best compromise between the responsibility and authority of users versus builders. The builders object to the downgrading of their prestige in the creation of the next generation of accelerator facilities; whereas users, with notable and laudable exceptions, would like to have the maximum control over the program and the operation of the accelerators, but the minimum responsibility for carrying out the work. In my experience it has been difficult to involve the future users of accelerators during the period of construction of the facility. Most users prefer to wait in line for their turn on the running schedule of existing accelerators rather than to participate in the planning of general-purpose accelerator facilities, or even of their own experiments, before an accelerator has actually produced a beam."

In closing he made the following important point:

> "I hope that during this discussion I have illuminated some of the questions which appear to cloud constructive builder–user relations. Each of the questions has a common feature: the problems will not be solved by extreme, one-way solutions; these problems are a feature of a new era of experimentation using large shared instruments. On the one hand, the objectives and the fundamental interest are just as 'academic' as they have always been in the past when large tools were not required; on the other hand, the technical necessities are such that traditional academic methods cannot be blindly continued. In order to solve these problems there is a clear need for the builder to understand the problems of the users, and the users, the problems of the builders."

## The conference program

Rather than to attempt a detailed summary of a conference as broad as this one we will only characterize it by

Mr. Martin is a member of the staff of the Electronuclear Division, Oak Ridge National Laboratory: Dr. Livingston is director of the Division. The Oak Ridge National Laboratory is operated for the U. S. Atomic Energy Commission by the Union Carbide Corporation.

listing the session titles to give the reader an idea of its breadth and an inkling of what to expect from the Proceedings in the June issue of TRANSACTIONS ON NUCLEAR SCIENCE. This listing will be followed by a few selected items of progress which were reported to the conference. The sessions were as follows: (A) Accelerator Component Controls and Automation Systems, (B) High Power RF Sources; Accelerating RF Structures, (C) DC Accelerators and Auxiliaries, (D) Accelerator Magnets and Power Supplies, (E) Performance of Existing Accelerators; Proposed Accelerator Designs; Facilities; Costs, (F) Beam Dynamics, (G) Safety Systems; Shielding; Radiation Effects; Radiation Protection; Vacuum, (H) Ion Sources; High Voltage Technology; Injection, (I) Beam Sensing and Handling Devices and Systems, Including Extraction, and (J) Future Accelerators.

Session (J) deserves special mention, because it consisted wholly of invited papers by some of the nation's foremost accelerator experts. The session consisted of the following papers: (1) "Sectored Cyclotrons," H. G. Blosser, Michigan State University; (2) "High Current Traveling Wave Electron Linear Accelerators," J. Haimson, Varian Associates; (3) "Meson Factories," J. R. Richardson, U.C.L.A.; (4) "The Future of Electron Synchrotrons," M. S. Livingston, Cambridge Electron Accelerator; (5) "Storage Rings," B. Gittelman, Stanford University; (6) "Super-Energy Accelerators," Lloyd Smith, Lawrence Radiation Laboratory; and (7) "Superconducting Accelerators," P. B. Wilson and H. A. Schwettman, Stanford University.

### Some glimpses of the conference

With the very high beam intensities contemplated from various accelerator improvement programs, and with the future 200- to 300- and 1000-GeV machines, the induced radioactivity problem is a very serious one indeed. Intensive studies are under way at several accelerator installations. Designs for the 200- to 300- and 1000-GeV machines are being based on scaling from present CERN and Brookhaven experience. There is a lack of adequate basic data. Some laboratories have active programs to alleviate the situation, but much remains to be done. Special equipment for remote maintenance is needed. Here the requirements for compactness and other special features prevent direct adoption of the designs available from the nuclear energy industry.

Progress continues to be made toward better voltage stability and compactness in 500-kV, multi-MeV dc accelerators. The compactness is gained both by high-frequency excitation, which minimizes the transformer size, and the use of high-pressure insulating gas, which markedly reduces clearances. In a one-MeV cascade rectifier a voltage stability of one part in 100 000 has been achieved by the use of a highly effective feedback system.

The newest large accelerators are faced with control problems of huge proportions. As might be expected, the on-line digital computer approach is being pursued with considerable diligence. The day when large accelerators are wholly automated is not far away.

At Stanford University active studies are under way relative to the use of superconducting materials for the construction of RF cavities for nuclear research devices which commonly use large amounts of RF power. Typical applications considered include electron linacs, cavities for electron synchrotrons, RF beam separators, and microtrons. Measurements of the $Q$ factor of a lead-lined cavity at 2.8 Gc/s, as a function of temperature, yield values from $\sim 10^8$ at 4°K to $10^{10}$ at 2°K. Operation below the lambda point (2.17°K) is recommended to take advantage of the superfluidity, thus improving heat transfer. At 2°K the improvement factor relative to room temperature copper is about $6 \times 10^4$. Although the savings in RF power are very large, some of the gain must be paid back in refrigeration cost and added complexity. Nevertheless, the possibility of CW operation makes the superconducting devices highly attractive.

A fairly heated controversy is developing between the proponents of gridded tubes and klystrons as to which is the best choice for accelerator applications in the UHF region. The choice of 805 Mc/s as the frequency for the high-energy end (above 200 MeV) of high-energy proton linac proposals has brought the controversy out in the open. The recent availability of gridded tubes with internal RF cavities, which operate efficiently at that frequency, has made them attractive to linac specialists who are fully versed in gridded-tube performance and circuitry. On the other hand, efficient high-power klystrons have been developed for the UHF range for other applications—for example, the BMEWS radar systems which operate at 430 and 1250 Mc/s. The klystron proponents base a large part of their argument on lower tube costs and failure rates. The ultimate tube costs for the newly developed gridded tubes are not known, since tube-life statistics can be variously interpreted. It is likely that the controversy will continue for some time. The stakes are quite high. The cost of the 805-Mc/s system for the proposed Los Alamos linac will be about $5 million.

One of the most important happenings at the conference was the announcement of successful operation of large high-field superconducting magnets. At the Argonne National Laboratory, fields of 67 kilogauss have been achieved in a 7-inch I.D. (inner coils of Nb-Ti alloy) system; AVCO reports 43 kilogauss in a 5-inch bore system (Nb-Zr alloy). Both of these coils operate close to the short-sample $I$ vs. $H$ transition curves, and demonstrate elimination of the so-called "current degradation effect" that has plagued the designers of large superconducting magnets. Both of these units use new current-stabilization techniques, which prevent a local superconducting–normal transition from propagating throughout the magnet. It can be said that the day of practical large superconducting magnets is at hand. Since the conference we have heard that one manufacturer has made a firm price quotation of $300 000 for a 25-foot-long, 8-inch-I.D., 30-kilogauss solenoid, including refrigeration system and power supply.

In closing, an acknowledgment is due the cosponsors: the National Bureau of Standards, the United States Atomic Energy Commission, and the American Physical Society. One reason for the success of the conference was the untiring efforts of the local arrangements committee: Louis Costrell, assisted by E. H. Eisenhower.

To the United States Atomic Energy Commission go special thanks for funds which enable the publication in the TRANSACTIONS of a total of 196 papers rather than the 92 that could be included in the conference program.

#### REFERENCE

1. High Energy Physics Program: Report on National Policy and Background Information, U.S. Government Printing Office, Washington, D.C., Feb. 1965.

# Authors

**Benjamin Lax** received the bachelor's degree in mechanical engineering from Cooper Union in 1941. During World War II he was a radar officer in the U.S. Army. In 1949 he received the Ph.D. degree in physics from the Massachusetts Institute of Technology. From 1949 to 1951 he was engaged in research on microwave gas discharge for the Geophysical Directorate of the Cambridge Research Center. In 1951 he joined the M.I.T. Lincoln Laboratory, where he was associated with the Solid State and Ferrites Groups. In 1948 he became associate head of the Communications Division, and was in charge of solid-state physics in several laboratory groups. He was appointed head of the Solid State Division when it was established in 1958. Since 1960 he has been director of the National Magnet Laboratory at M.I.T. In 1964 he was appointed associate director of Lincoln Laboratory. He relinquished this position in May 1965 to become a professor at M.I.T., but will continue his duties as the director of the National Magnet Laboratory.

Dr. Lax is the recipient of the 1960 Oliver E. Buckley prize of the American Physical Society. He is the author or coauthor of more than 100 technical articles, a Fellow of the American Physical Society and the American Academy of Arts and Sciences, and a member of Sigma Xi.

**Richard P. Mathison** (M) received the B.S. degree in electrical engineering from Purdue University in 1952. After two years of active duty in the U.S. Navy as a Combat Information Center officer and electronics officer, he did graduate work at the California Institute of Technology, from which he received the M.S. degree in electrical engineering in 1955.

Since 1954, Mr. Mathison has been employed by the Jet Propulsion Laboratory, California Institute of Technology, in the Telecommunication Division. His work there has chiefly been in telecommunication system design analysis. He has participated in work on anticountermeasures, the Microlock space tracking system, the Explorer satellites, and the Pioneer III and IV, Ranger, and Mariner II and IV projects. He is currently involved in preliminary investigations in connection with the Voyager deep space probe. His present position is manager of the Spacecraft Radio Section.

Mr. Mathison is a member of Eta Kappa Nu.

**Trevor A. Robinson** (SM) received the B.Sc. degree in electrical engineering in 1925 from the University of Manitoba, Canada, and in 1938, after two years' graduate work, the E.E. degree from Stanford University. Since 1954 he has been with the Northrop Corporation, where his chief concern has been with the power, rather than electronic, aspects of the aircraft and aerospace field. One major activity involved ground support equipment and launch site complexes for the Snark intercontinental missile. Those fields at the interface between aerospace and power engineering have been of particular interest to him.

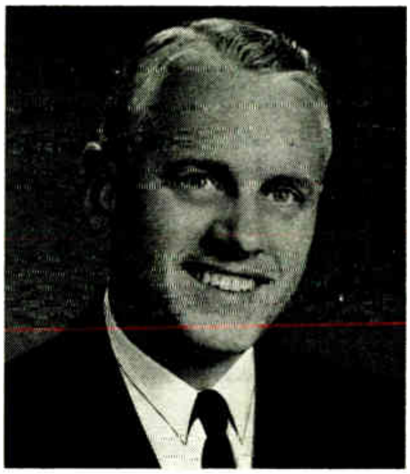
Mr. Robinson previously worked with various public electric utility companies in the city distribution and substation design areas. At one time he was an application engineer during a frequency changeover from 30 to 60 c/s in Michigan. His association with industry between 1938 and 1954 included six years with an architect–engineering firm. During this time he was engaged in a variety of consulting projects, including work for the United States Atomic Energy Commission at Los Alamos, N. Mex., and for a number of petrochemical plants.

**Marvin Camras** (F) received the B.S. degree in electrical engineering from the Armour Institute of Technology in 1940 and the M.S. degree from the Illinois Institute of Technology in 1942. Since 1940, as a member of the staff at Armour Research Foundation, he has been engaged in research work on various projects in the electronics department, including remote control, high-speed photography, magnetostriction oscillators, and static electricity.

Mr. Camras contributed developments that are used in modern magnetic tape and wire recorders, including high-frequency bias, improved recording heads, wire and tape materials, magnetic sound for motion pictures, multitrack tape machines, stereophonic sound reproduction, and video tape recording. He is a Fellow of the Acoustical Society of America and the American Association for the Advancement of Science, and a member of the Society of Motion Picture and Television Engineers, Eta Kappa Nu, Tau Beta Pi, and Sigma Xi. He received the John Scott Award in 1955 and the IEEE Consumer Electronics Award in 1964.

**Louis H. Roddis, Jr.** (SM) is president and director of the Pennsylvania Electric Company, a position he has held since September 1958. He was graduated from the United States Naval Academy in 1939, following which he was assigned to sea duty. He saw action in Pearl Harbor in December 1941. He received the M.S. degree in naval architecture and marine engineering from the Massachusetts Institute of Technology in 1944. Subsequent naval assignments included two years at the Philadelphia Navy Yard, work on the atomic weapons test at Bikini in 1946, power reactor development on the Manhattan Project, and design work on the nuclear ship *U.S.S. Nautilus*. He was assigned to the Division of Reactor Development on the U.S. Atomic Energy Commission in 1949. He resigned from active duty in 1955 and at present holds the rank of captain in the U.S. Naval Reserve.

Mr. Roddis is a member of the American Society of Naval Architects and Marine Engineers, the American Nuclear Society, the Society of Naval Engineers, ASME, ASHRAE, and Sigma Xi. He received the AEC's Outstanding Service Award in 1957. He is the author of 19 technical papers.



**John D. Alber** (M) received the B.E.E. degree from Rensselaer Polytechnic Institute in 1956 and the M.B.A. degree in management from Hofstra University in 1964. Since joining Grumman Aircraft Engineering Corporation in 1957 he has worked mainly in the field of antennas and communication systems for aircraft and space vehicles. Following an assignment in connection with the Orbiting Astronomical Observatory, he became group leader for the Lunar Excursion Module communication antennas. In that position he directed work on the definition of the present LEM antenna system parameters and their implementation into hardware design.

In 1964 Mr. Alber was selected to participate in Grumman's Engineering Professional Development Program, a two-year rotating assignment program to broaden an engineer's working knowledge in diverse technical and management areas. Currently he is in the Business Development Department, where he is engaged in long-range company planning.

**Richard H. Imgram** (M) received the B.E.E. degree from the Polytechnic Institute of Brooklyn in 1950. In his present position of assistant chief of RF engineering with the Grumman Aircraft Engineering Corporation, Bethpage, N.Y., he directs design and advanced development in the field of electromagnetics, including antennas, radomes, microwave devices, radar cross sections, and external electromagnetic interference.

Since joining Grumman in 1951, Mr. Imgram has made significant contributions to a number of aircraft and space programs, including the LEM, OAO, E-2A, A-6A, S-2E, F-111B, and Gulfstream vehicles. He is a member of the Grumman Engineering Professional Development Steering Committee. He has published several papers relating to the subject of antennas and radomes. From 1962 to 1964 he served as secretary of the Long Island Chapter of the IEEE Professional Technical Group on Antennas and Propagation. He also was editor of the 1964 PTG-AP SYMPOSIUM DIGEST.

# 1965 IEEE INTERNATIONAL CONVENTION RECORD

The IEEE International Convention Record, containing all available 1965 International Convention papers, was published in March, 1965. IEEE members may purchase any one or more parts at the IEEE member rate.

Nonmembers and libraries may place orders at the nonmember and library rates, respectively. Individuals who apply for IEEE membership at the time they place their orders are entitled to the IEEE member rate.

Clip out the order form at the bottom of this page, fill in the necessary information, and mail it with your remittance to: The Institute of Electrical and Electronics Engineers, Inc., 345 East 47 Street, New York, N.Y. 10017. When ordering, please be certain to refer to the proper column for subjects and prices.

| Part | Sessions | Subjects | Prices for IEEE Members (M), Libraries & Sub. Agencies (L), and Nonmembers (NM) | | |
|---|---|---|---|---|---|
| | | | M* | L | NM |
| * One copy at special member rate; additional copies purchased at Nonmember rate. | | | | | |
| 1 | 33, 41, 49, 53, 57, 61, 65, 71, 76, 77 | Wire & Data Communication | $ 2.50 | $ 5.75 | $ 7.50 |
| 2 | 1, 2, 9, 10, 23, 31, 36, 44, 69, 73 | Radio Communication; Broadcasting; Audio | 2.50 | 5.75 | 7.50 |
| 3 | 25, 34, 42, 50, 58, 66, 68, 74 | Computers | 2.00 | 4.50 | 6.00 |
| 4 | 5, 13, 22, 30, 37, 45, 51 | Aerospace; Military Electronics | 2.00 | 4.50 | 6.00 |
| 5 | 7, 15, 17, 54, 62, 67, 75 | Antennas; Microwaves; Electron Devices | 2.00 | 4.50 | 6.00 |
| 6 | 6, 14, 16, 24, 28, 47 | Automatic Control; Systems Science; Cybernetics; Human Factors | 1.50 | 3.50 | 4.50 |
| 7 | 3, 8, 11, 19, 20, 40, 78 | Circuit Theory; Information Theory; Basic Sciences; Electrostatic Processes | 2.00 | 4.50 | 6.00 |
| 8 | 21, 29, 38, 46, 52, 60 | Industry & General Applications | 1.50 | 3.50 | 4.50 |
| 9 | 56, 64, 80 | Power | 1.00 | 2.25 | 3.00 |
| 10 | 39, 59, 70, 79 | Components; Production; Reliability | 1.25 | 2.75 | 3.75 |
| 11 | 27, 32, 35, 48, 55, 63 | Instruments; Measurements; Industrial Electronics; Nuclear Science; Ultrasonics | 1.50 | 3.50 | 4.50 |
| 12 | 4, 12 | Bio-Medical Engineering | 1.00 | 2.25 | 3.00 |
| 13 | 18, 26, 43, 72 | Education; Management; Writing & Speech | 1.25 | 2.75 | 3.75 |
| | | Complete Set (13 Parts) | $22.00 | $50.00 | $66.00 |

**ORDER FORM** 1965 IEEE INTERNATIONAL CONVENTION RECORD

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Part | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | Complete Set |
| Number of Copies | | | | | | | | | | | | | | |
| Amount Paid | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | Total Remittance | |

Name (Please Print): ______________________ Member No. ____________

Shipping Address: ______________________________

______________________________

**Mail to:**
**IEEE**
**345 E. 47 Street**
**New York, N.Y. 10017**