The Journal of the

BRITISH INSTITUTION OF RADIO ENGINEERS

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

VOLUME 22 SEPTEMBER 1961 NUMBER 3

THE NEXT STAGE

HAVING achieved the honour of a grant of a Royal Charter of incorporation the Institution is now pledged to work for the advancement and dissemination of knowledge in radio and electronic science and engineering, for the setting of standards of professional qualification and of good practice, and for the advancement of technical education particularly in the field of radio and electronic engineering.

Among the more important of the many activities for which the Institution is Chartered is the holding of lectures and discussion meetings. The next stage is to ensure adequate facilities for extending those activities for the benefit of members.

The present headquarters does not include a lecture theatre, and the Institution is dependent in this respect on the hiring of suitable accommodation. This often results in inconvenient arrangements, and as the demand for lectures grows with the widening of the field of radio and electronic engineering knowledge, any restriction of lecture dates cannot be countenanced.

The place of the Institution as the learned and professional society in the British Commonwealth concerning itself with radio and electronic knowledge has now been undisputably established. This being so, the Council believes that the finest way of showing appreciation for the grant of the Charter is to concentrate on the acquisition of larger premises, which in itself will be a major step toward advancing the work of the Institution. Success in this endeavour will also permit expansion of the Library and the provision of further services to the profession and industry.

The income of the Institution is mainly derived from its membership, and within the limitations of that revenue much has been accomplished. It must be recognized, however, that it would be almost impossible to envisage the acquisition of a suitable building out of membership revenue. The cost will have to be raised by support not only from members, but from the radio and electronics industry. Already several companies have given liberal support to the project.

At the same time, it is important that members themselves should feel able to give the scheme their full support. The subscription income of the Institution is now approaching £40,000 a year, and if each member were to make a contribution comparable with his annual subscription, the fund would not only be tremendously helped but would justify further support by the friends of the Institution.

Donations by way of seven-year deeds of covenant will be especially welcomed. This method gives advantages to the Institution, because it is able to reclaim income tax already paid on the sum contributed.

Will you help in achieving the next stage of the Institution's development?

INSTITUTION NOTICES

The 1961 Radio and Electronics Industry Council Banquet

The President has accepted an invitation to be Guest of Honour at the Radio Industry Council Banquet on Wednesday, 29th November, at the Dorchester Hotel, London.

Admiral of the Fleet the Earl Mountbatten of Burma, K.G., opened the National Radio and Television Exhibition in 1951. Quotations from his Address were given in the August 1951 *Journal.*

When Governor General of India, Lord Mountbatten sent a tape recorded message which was used when the Radio and Television Exhibition was opened in 1947.

Institution Premiums and Awards

The Council of the Institution announces that the following awards are to be made for outstanding papers published in the *Journal* during 1960:

CLERK MAXWELL PREMIUM

"A Proposed Space-Charge-Limited Dielectric Triode" by G. T. Wright, Ph.D. (May *Journal*).

HEINRICH HERTZ PREMIUM

"Features of Cylindrical Waveguides containing Gyromagnetic Media" by R. A. Waldron, M.A.(Cantab.) (Associate Member) (September *Journal*).

A.F. BULGIN PREMIUM

"Reflection Coefficient Curves of Compensated Discontinuities on Coaxial Lines and the Determination of the Optimum Dimensions" by A. Kraus, Dr. Ing. (February *Journal*).

ARTHUR GAY PREMIUM

"The Application of Printed Wiring to Development and Small Batch Production with Particular Reference to Television Equipment" by E. Davies (April *Journal*).

LESLIE MCMICHAEL PREMIUM

"Microwave Valves: A Survey of Evolution, Principles of Operation and Basic Characteristics" by C. H. Dix, B.Sc., and W. E. Willshaw, M.B.E., M.Sc.Tech. (August *Journal*).

LORD RUTHERFORD AWARD

"Energy Stabilization of a 4-MeV Electrostatic Accelerator using Controlled Corona Discharge" by E. C. Fellows (Associate Member) (September *Journal*).

CHARLES BABBAGE AWARD

" A Computer Storage Matrix Using Ferromagnetic Thin Films" by E. M. Bradley, B.Sc. (October *Journal*).

Programme Booklet for 1961–62 Session

The Programme of Institution Meetings in Great Britain will be ready later this month and will be sent to all members in the British Isles. It will give details of London and Local Section meetings to the end of January 1962.

Members are reminded that all meetings whether arranged by Specialized Groups or Local Sections are open to all members irrespective of their professional activity, place of residence or grade of membership. There is, however, a limited number of special Symposia at which attendance may have to be limited and/or a registration fee charged.

Circulation of the Journal

The Audit Bureau of Circulations has recently announced that the average circulation of the Institution's *Journal* for the first six months of 1961 was 8647 copies per issue. This represents an increase of 813 copies compared with the previous six months and 1076 copies compared with a year ago. These figures are based on copies sent to members and subscribers on publication and do not include later sales.

National Council for Technological Awards

In its recently published Annual Report, the National Council for Technological Awards draws attention to the large increase in the numbers enrolling for the Diploma in Technology—1770 compared with 1500 enrolled in the previous year. This covers all subjects. Nearly 5000 students were taking the course of Dip.Tech., which compares with 3800 in the previous year. Of the 3542 engineering students, 83% were works-based which compares with only 56% in technologies other than engineering. An interesting fact is that approximately 25% of the session's first year students entered these courses by virtue of holding Ordinary National Certificates.

Since starting five years ago 472 diplomas have been awarded, 309 in the last year. Electrical and Mechanical Engineering would seem to attract most students, since 110 diplomas were awarded during the past year to electrical and 84 diplomas to mechanical engineering students. The scheme does not include any distinctive course of study in electronics and/or telecommunications, although it would appear that these subjects may be included in an electrical course.

The first full year of operation of the higher qualification—Membership of The College of Technologists—has resulted in 27 applicants being registered as candidates. So far there are only three candidates registered in Electrical Engineering (including Electronics).

U.D.C. 621.396.68: 629.19

Power Supplies for Space Vehicles

By

K. E. V. WILLIS, B.Sc., A.R.C.S.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: The requirements for electrical power in space vehicles over the next five years are reviewed and estimated figures given with particular reference to power/weight ratio. Vehicle life duration and its effect on choice of supply are discussed. Problems encountered in space vehicles due to environmental conditions, e.g. temperature, radiation, zero gravity, etc., are then pointed out. The available generation methods are: thermoelectric generators; thermionic converters; the fuel cell; solar energy nuclear sources; magneto-hydrodynamic methods.

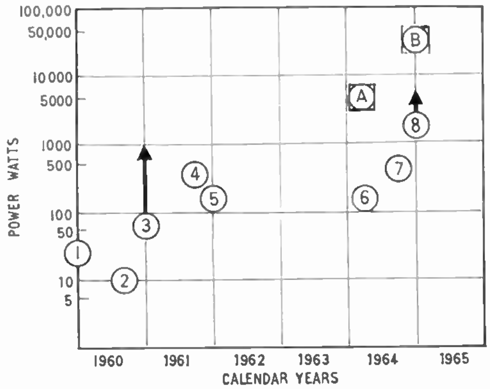
1. Introduction

The lack of a satisfactory primary source of electrical power is proving to be a major drawback in the development of space vehicles of all types other than passive satellites. Specifications for space vehicle power supplies are frequently very stringent. A particular requirement which presents most difficult problems is the demand for very long periods of operation, e.g. several months. It is the purpose of this paper to summarize more recent developments in unconventional power supplies which may lead to the development of power units suitable for space vehicle applications. Because space research is in its early stages in this country, the greater part of the work being carried out in this field is being undertaken by private firms and Government establishments in the U.S.A. What little information has been obtained from Russian launchings suggests that to a large extent they are using conventional batteries together with solar cells. Some of the more unconventional forms of power generation are still very much in the experimental stage, but with the recent increase in tempo in space research in the U.S.A., it is expected that some novel devices will be tried out during the next two years.

2. Power Requirements

Figure 1 can be used as a guide to the level of power requirements over the next four or five years, based on the American programme of firings as it stood at the end of 1959. In general, the power requirements can be split into two categories, namely, the average power requirement and the short-term peak demand. With some vehicles there are also special power requirements during launch which may differ considerably from the demands during orbit or free flight. The outstanding feature of Fig. 1 is that to fulfil the programme which it illustrates, power supplies giving a peak output of 4 kW and an average power of the order 1 kW for at least fourteen days will be required as early as 1964. At the other end of the scale a *Ranger* space craft for lunar landings requires only 0·2 of a watt but for periods up to three months in duration. By 1970 a megawatt capability may be required if manned lunar-landing vehicles are to be developed. Unless the space vehicle can utilize solar radiation as a power source, then it must carry with it all fuel supplies for power generation over the prescribed period. Such considerations are of the greatest importance inasmuch as they have a profound effect on the weight/power ratio of the vehicle, the fuel representing so much dead-weight during launching. If, as a rough guide, one assumes that to place 1 lb weight of satellite in stable orbit a launching vehicle of 1000 lb weight is required, then the significance of this dead-weight factor is obvious. This emphasizes the necessity to miniaturize instrumentation and associated power supplies as far as possible. Such difficult power requirements tended to channel early work into the development of nuclear power systems. The long half-life of certain nuclear fuels is admirably suited to the problem of operation over extended periods for which certain space vehicles are designed. However, the weight-to-power ratio of nuclear systems, including all ancillaries, is unattractive, being of the order 1 lb weight per watt in generators in the 300-watt class. In larger nuclear units such as the American SNAP-8, which is a nuclear reactor, a weight/power ratio of 50 lb per kilowatt can be achieved. However, there is much engineering work to be done on nuclear devices for space and, in general, they are unsuited to lower orders of power generation. Furthermore, there is international concern over the possibility of the nuclear contamination of space, with the result that more recently there has been a change in emphasis from nuclear systems to those which make use of solar energy.

† National Research Development Corporation, London, W.1.



1 Pioneer space probe: 20 to 30 W.
2 Scientific satellites launched from *Scout* and *Delta* vehicles: 10 W.
Mercury 1-man capsule: 70 W average power, 1000 W peak.
Nimbus meteorological satellite: 250 W.
Ranger spacecraft launched from *Atlas-Agena* vehicle: Lunar landings 0·2 W for 1 to 3 months; 115 W average demand during flight.
6 Venus probe: 170 W.
7 Mars probe: 500 W.
8 *Saturn* booster to launch manned circum-lunar flight. Average 1·2 kW for 14 days. Peak 4 kW.
A SUNFLOWER 3 kW turbo-electric generator using solar heat. Lithium hydride heat storage.
B SNAP-8. A 30 kW nuclear reactor system.

Note.—Items 1–8 represent requirements while A and B are typical systems under development to meet some of those requirements.

Fig. 1. Space vehicle power requirements based on U.S.A. programme. (1959)[1]

Table 1

Typical Features of the Successive Environments of a Space Vehicle[2]

Environment	Conditions
Prelaunch:	Shelf life Temperature Delivery and handling Salt atmosphere Sunshine Oxidation Atmospheric pressure Humidity Rain Shock
Launch:	Temperature Vibration Acceleration Pressure change Acoustic noise
Orbit:	Temperature Low pressure Particle radiation Solar radiation Meteorite impact Zero gravity Ionized gases Orientation and spin
Landing on Celestial Body:	Temperature Deceleration Shock Pressure Atmospheric gases Solar radiation Gravity

3. Environmental Problems

Many factors must be considered by engineers in designing power supplies for space applications, in particular the environmental conditions which are encountered during pre-launch, launch, orbit and landing on other celestial bodies. Table 1 shows typical environmental parameters. One does not need to emphasize the fact that the reliability factor of the equipment will be of paramount importance. Ground servicing, particularly simplicity in making last minute component changes on the launching pad in the event of difficulties arising during the count-down, should be taken into consideration during the design stages. Other unusual environmental conditions which the designer will encounter are those of zero gravity, exposure to atomic radiation and ionized gases, and meteorite impact. Operation under conditions of zero gravity may pose particular problems in the mechanical design of the unit, while radiation hazards will call for some form of screening which must adversely affect the weight/power ratio. In general, the designer of a power unit for a space vehicle has little to learn from experience gained in developing power supplies for operation in a ground environment. A completely new approach to the problem is required, supported by novel engineering techniques, and the trend has shown that the physical chemist and solid-state physicist may have more to offer than the radio engineer in the basic research phase. Once the choice of power supply has been made and feasibility proved, then conventional engineering techniques should be applicable to such problems as layout, packaging, miniaturization and reliability. The development of new materials for power unit construction suited to space environmental conditions is very important, and all the devices at present being investigated depend on successful materials research for their fulfilment.

Other problems which should be mentioned arise out of the somewhat inflexible form factor forced upon the designer of instrumentation for space vehicles. It is perhaps unfortunate that the pay-load has to occupy the space in the nose-cone region of the rocket where usually there is a need to preserve aerodynamic shape and a circular cross-section. Thus, provision must sometimes be made for instrumentation elements to be extended into more suitable positions after launching. This is particularly true of radio antennae and outriggers supporting batteries of solar cells, sensing elements, etc.

Having outlined some of the difficulties which beset the designer, possible sources of electric power applicable to space vehicles will now be discussed. All are based on the principle of direct generation of electricity, that is the conversion of some form of energy into its electrical counterpart in a single step, thereby avoiding the losses which occur in processes requiring several stages of conversion, e.g. a central power generating station which converts fossil (or nuclear) fuels into heat, uses the heat to generate steam to rotate a turbine which in turn drives a conventional generator. The losses in these intermediate steps are such that the overall efficiency of such an operation is of the order of 35%. Another important point to be taken into account is that any form of electrical generator which is a heat engine will be subject to the "Carnot Limitation" with consequent effect on overall efficiency. The highest efficiency of generation is therefore likely to be achieved by the use of a device which is not a heat engine and which converts some form of available energy into electrical counterpart in a single step. However, due to the special problems which operation in a space environment impose, it may not be possible to select the most efficient process. Choice may be governed by other factors associated with the type of space vehicle and the mission for which it is designed.

The methods which will now be described all generate low-voltage direct-current power. While this may be admirably suited to transistor circuits, any requirement for high voltages must be met by using several elements in series or by converting to alternating current, transforming and rectifying. The latter method is unattractive since it introduces additional losses. It may, however, be the best approach if volume (and sometimes weight) reductions can thus be achieved.

4. Thermo-electric Generators

The subject of thermo-electric power generation has attracted interest for the past several years. A pioneer in this field was the Russian physicist, Ioffe.[3] The interest in this type of generator stems from the fact that there are no moving parts, and also because sometimes the device can be inserted in the effluent of some other generating device, thereby making use of heat which might otherwise be wasted. The basis of all thermo-electric generators is the simple thermocouple. If two dissimilar metals A and B are joined in a closed loop, and one junction heated while the other is cooled, then current will flow in the loop, a phenomenon known as the "Seebeck Effect". The reason for this current flow is generally ascribed to the movement of electrons from one end of a thermoelectric material to the other when heat is applied. It is assumed that the positive charges remain fixed in position, and the result is the appearance of a potential difference between the ends of the material.

The effect has been known for many years but only recently has it been applied to electrical power generation. This was because earlier thermo-electric materials yielded efficiencies of only 1% or so but with the advent of semi-conductors it was found possible to increase this figure very considerably so that the device has now become a valuable addition to the family of electric generators. In a negative type semi-conductor, electrons flow away from the heated surface. In a positive type, they flow towards the heat. When two are yoked together in a thermoelectric couple the electrons flow around them and through an external circuit. (See Fig. 2(a).)

In general, semi-conductors for thermo-electric work are less pure than those used in transistors (and therefore cheaper) and also less so than those needed for solar cells.

Thermo-electric materials can be assessed by their "figure of merit" Z;

$$Z=\frac{S^2}{\rho K}$$

where S is the Seebeck coefficient (or electric power in volts/degree)

ρ is the electrical resistivity in ohms/cm

and K is the thermal conductivity in watts/cm/degree.

The higher the value of Z, the better the material for thermo-electric purposes. Z can be increased by choosing a high value of S and relatively low values of ρ and K, so we are looking for materials with low electrical resistivity and thermal conductivity. Metals have low resistivity but high thermal conductivity. Also their thermo-electric powers are generally very low. Insulators suffer from very high resistivity values and are consequently poor thermo-electric generators. Semi-conductors appear to be well suited to such applications however, particularly since their physical properties can to some extent be adjusted by doping which varies the number of free electrons present in the material. Work is proceeding to control the thermal conductivity of such materials. The properties S, ρ and K are dependent

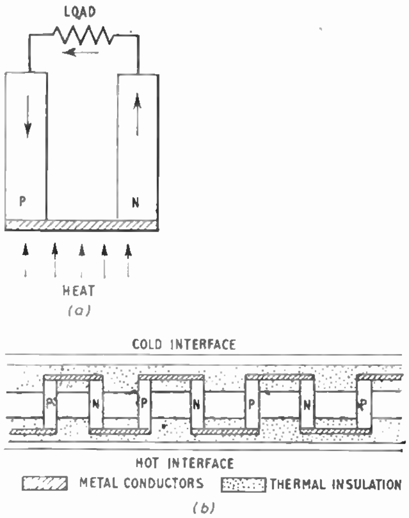


Fig. 2. (*a*) Basic semi-conductor thermo-element.
(*b*) Practical arrangement of thermo-element junctions.

on functions of temperature, so that the figure of merit Z is also a point function and tends to fall off rapidly on either side of the optimum temperature. To allow operation over a wide range of temperatures, one can fabricate the thermo-electric elements from several different materials so that each segment operates at the temperature which gives the highest output. The same sort of effect can be produced by selective doping of various parts of the semi-conductor element.

Thermo-electric voltages of the order 200 μV per deg C are readily achievable in practice with semi-conductors which have a free electron density of the order 10^{16} per cubic centimetre. Electrical conductivity is directly proportional to free-electron density and therefore output voltage varies inversely with this parameter.

Typical materials for thermo-electric use are zinc-antimony, lead telluride, bismuth telluride and germanium telluride. A high value of mean atomic weight is necessary for a good value of Z, and the material should have predominantly co-valent bonding. The best thermo-electric materials also have a crystal structure with a high degree of symmetry.

The thermo-electric performance of many semi-conductors falls off rapidly with temperature and above about 1000° C they become almost useless. A simple explanation of this is that at high temperature both electrons and positive ions migrate in equal numbers and there is no resultant potential difference. Thus a new class of materials is required for operation at higher temperatures. In the region of 2000° C the so-called "transition metals" (lying near the centre of the periodic table) are suitable.[4] The thermo-electric generator is a simple heat-engine and is therefore subject to the usual efficiency limitations of such systems.

The manufacturing techniques associated with semi-conductor materials present quite serious problems since these substances are not ductile. Most of the elements are prepared in the form of rods or cylinders and their ends are joined by metal strips whose function is only to carry electric current from one leg to another. These connections are important, however, since contact resistance must be minimized, and there is a tendency for the electrode material to diffuse into the semi-conductor and contaminate it. (See Fig. 2(b).) The constructional details vary according to the particular manufacturer, but in general the elements are arranged as series-connected thermocouples whose materials have been chosen and treated (doped) so that their voltages are additive. A typical arrangement uses a vertical stack of n-type and p-type semi-conductor cylinders with alternate hot and cold junctions.

The application of thermo-electric generators to nuclear reactors is being actively pursued. Over the next decade we may see them used in association with fuel-elements (hot junction) to generate electricity directly from the pile heat. The materials are naturally subjected to very high irradiation levels and Westinghouse in Pennsylvania are investigating this aspect of environmental testing.

Practical results to date, all obtained in the U.S.A., are:

(a) Westinghouse Electric Corporation.—A 500 watt unit for U.S. Navy. Operates at 1200° F (hot junction) and 50° F (cold junction). Heat supplied by burning kerosene. Efficiency 10%.

(b) Transitron Corporation.—A lead telluride generator for the temperature range 0–1000° C. 14% efficiency.

(c) Martin Company.—A 5 watt thermo-electric generator for an automatic weather station (4 V d.c.). Heat supplied by nuclear fuel (strontium 90 in form of strontium titanate pellets (1 lb wt.)).

(d) Minnesota Mining & Manufacturing Co.—This company can offer elements for experimental work. 7·8% efficiency can be obtained at a 1000° F temperature differential. A 10 watt unit is on the market (gas-burning) and a 500 watt gasoline burning unit is under construction for the U.S. Navy.

(e) Semi-Elements, Inc.—Offering gadolinium selenide in powdered granular form at about £2 per gram in U.S.A. It gives a Seebeck output of 500 μV per degree, but this varies with doping, and the resistivity also. Can be processed into rods.

The improved performance obtainable with semiconductor materials is also applicable to the reverse Seebeck Effect, i.e. the Peltier Effect. If an electric current is passed through these materials one junction becomes heated and the other cooled. The effect is proving valuable in the development of small cooling units which could have application in space vehicle instrumentation.

5. The Thermionic Converter

The thermionic energy converter is a diode having a high work-function cathode and a low work-function anode. The cathode is heated by the heat source to a temperature at which electrons are emitted. These electrons pass to the anode and return to the cathode via the external circuit, resulting in the generation of electrical power in the load. (See Fig. 3.) The optimum output voltage corresponds to the difference in work functions of the cathode and anode and is generally higher than the working voltage of a single thermo-electric generator (e.g. 0·5 to 3·0 V). Since not all the available heat is converted into electrical energy, the surplus will appear in the anode structure and must be removed by cooling. The device offers efficiency figures well in excess of those obtainable from thermo-electric generators and it is anticipated that values in the range 25–30% will ultimately be achieved in practical units.

There are, however, several problems to be solved before the device can be put to practical use, namely:

(a) Space-charge effect which, as in conventional electron-tubes, limits the electron flow.

(b) Materials having the correct electrical, mechanical and thermal properties, must be found for cathode, anode and envelope.

(c) Technological problems, arising from the manufacture of the device and dependent to a large extent upon the application, must be overcome.

The principle of this device has been known for many years; in fact Edison observed the effect in his early work on electric lamps. Only during the past four or five years have intensive efforts been made to develop the system, mainly due to the search for a simple, robust, direct-conversion device for insertion into nuclear reactors. Its lack of moving parts is an attractive feature. The generator, being a heat engine, is subject to the Carnot Limitation mentioned earlier. The working fluid in this case consists of electrons.

Some workers in this field liken the device to a "plasma thermocouple", stipulating that the cathode is the hot junction and the anode the cold junction. In practice the plasma may be more complex than a simple cloud of electrons since positive ions may be introduced into the anode-cathode space.

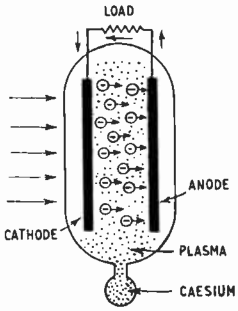


Fig. 3. Basic thermionic diode using caesium for space-charge neutralization.

The negative space charge produced by the electrons in transit between cathode and anode tends to limit the current flow with a resultant reduction in efficiency. If caesium ions are introduced into the envelope, these are effective in neutralizing the space charge. For the device to work efficiently a cathode work function of the order of 3·9 V is required, and this necessitates heat source temperatures in the region of 2000° C for high conversion efficiencies. Work is in progress to reduce the operating temperature when the caesium ion technique is adopted. Another way of minimizing space-charge effects is to reduce the cathode-anode spacing. This introduces constructional problems, but close-space diodes have been manufactured with only a 0·0001 in. gap between electrodes. Such diodes tend to have rather low efficiencies, 6–10% being a typical value.

The anode must be made from a low work-function material if high efficiency of operation is required. Surfaces covered with barium, caesium or lithium have work-functions as low as 1·8 V, while semiconductors offer even lower values.

The cathode should be a good electron-emitter and at least part of the surface should have a work function of 3·9 V or higher. The evaporation rate of the cathode surface must be low to give a long operational life. Tantalum is a good choice of cathode material although multiple work function materials (e.g. uranium carbide, thoriated tungsten), have interesting possibilities. Table 2 lists suitable anode and cathode materials.

Table 2

Materials suitable for use in Thermionic Converters[5]

	Material	Work function (volts)	Temperature range (°C)
Anodes	BaO/SrO on Ni	1·0	—
	Cs on AgO	0·75	—
	Cs on WO	0·75	—
Cathodes	Ba impregnated W	1·7	900–1200
	Th on W	2·55	1800–2000
	Cs on W	1·7	1400–1600

The conversion efficiency rises rapidly with increased cathode temperature but this also increases the evaporation rate with the result that a compromise must be adopted between efficiency and operational life in any design. This type of generator starts to become a practical proposition at temperatures where the thermo-electric devices begin to fail.

Table 3 shows the performance of typical thermionic converters at their present state of development. Note rise of efficiency with cathode temperature.

Table 3

Performance of a Typical Thermionic Converter[5]

Cathode temperature °C	Current density (amps/cm²)	Power output (watts/cm²)	Efficiency %
900	0·1	·0·07	3·0
965	1·2	0·84	17·9
1020	2·6	1·96	24·7
1070	5·2	3·64	29·1
1130	9·0	6·3	32·0

The above figures were obtained using a barium impregnated tungsten cathode and an anode of BaO/SrO on nickel.

The constructional problems associated with thermionic generators are very similar to those encountered in the manufacture of electron tubes of comparable power ratings. The basic difference is the method by which the cathode is heated, and one of the major design problems is the method of supporting the hot cathode. The cathode lead not only supports the cathode element but also thermally isolates it from its output terminal and acts as the current carrier. These three functions are to some extent incompatible and again a compromise solution must be sought.

The General Dynamics Corporation in the U.S.A. has recently produced a thermionic converter of 90 watts capability with 10% conversion efficiency at a temperature of 2000° C. The tube contains caesium and is heated by its own nuclear fuel element consisting of uranium carbide and zirconium carbide. Under the sponsorship of the N.R.D.C., Professor D. Gabor at Imperial College has developed a thermionic diode using argon or an argon-mercury mixture instead of caesium which is a very corrosive material. One of the great advantages of the Gabor device is that it can be made to generate alternating current. Predicted efficiency figures for the device are of the order 30% for d.c. generation and 20% for a.c. generation.[6, 7]

6. The Fuel Cell

The fuel cell [8, 9] is an electro-chemical device in which the chemical energy of a conventional fuel is converted directly into low voltage direct current electrical energy. The fuel cell is not a heat engine and its efficiency theoretically will approach 100%, although in practice it is unlikely that values above 60–70% will be achieved.

Fuel cells are not new devices but have been known since 1839 when Sir William Grove demonstrated their feasibility. There are many major problems to be solved before fuel cells become available for general use but over the last five or six years the tempo of research work on these cells has increased so markedly that several manufacturers, particularly in the U.S.A., are approaching the time when such devices can be placed on the market. The basic problem defined by Grove more than 100 years ago still exists today. In any fuel cell using gaseous fuels, the problem is one of a three-phase boundary; the liquid electrolyte has to be maintained in close contact with a gaseous fuel and a solid electrode, the function of which is to conduct the electricity out of the cell. (See Fig. 4.) It is obvious that the electrode is of vital importance since it must not only retain the liquid electrolyte, but also allow the gaseous fuel to come into close contact with it. Research work on suitable electrodes is therefore vital to the success of the whole fuel cell art.

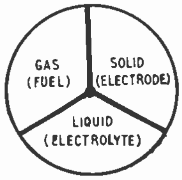


Fig. 4. Three-face boundary problem in fuel cells.

To understand the action of a fuel cell it is convenient to consider a cell operating on hydrogen and oxygen using a caustic potash electrolyte. To oversimplify the problem one can consider the reaction which takes place to be the reverse of the electrolysis of water. When an electric current is passed through water it is split up into hydrogen and oxygen; by the reverse process, if hydrogen and oxygen are combined in a cell the result is a flow of electric charge across the electrolyte, this forming an electric current in an external circuit. In this particular case, hydrogen atoms become ionized on passing into solution, yielding an electron to the electrode and a positive ion to the electrolyte. At the other electrode, which is in an atmosphere of oxygen, the hydrogen ions are neutralized by taking an electron from the electrode and the free hydrogen combines with the available oxygen to form water. The net result is a loss of positive charge by the hydrogen (fuel) electrode and a gain of positive charge by the oxygen (oxidant) electrode (see Fig. 5). To obtain useful powers and efficiencies, steps must be taken to speed up the

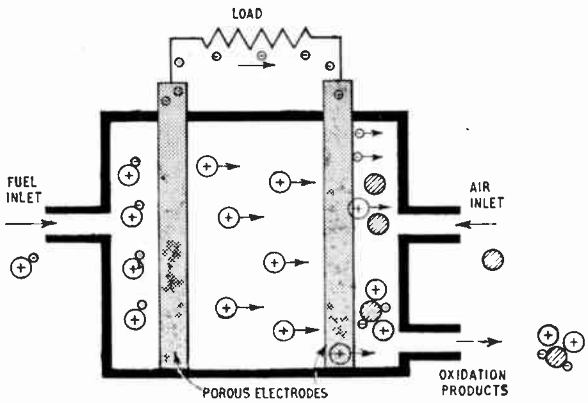


Fig. 5. Representation of fuel cell mechanism.

reaction at each interface and to contrive some means of removing the water formed by the reaction in order to prevent dilution of the electrolyte. The problem is acute and is demanding the time and efforts of some of the best electro and physical chemists in the world today, and research and development in fuel cells is costing millions of dollars per annum in the U.S.A. alone. The device has attracted a good deal of interest in military circles since a fuel cell operating at full power emits no noise and generates no toxic or unpleasant fumes of any sort. They are, therefore, well suited to operation in enclosed spaces such as underground, in under-sea vessels, in aircraft and in any environment where silent operation is a necessity.

In England, fuel cell research and development is being devoted mainly to cells operating on hydrogen and oxygen, or air and hydrocarbon fuels such as propane, methane or kerosene. This work, however, represents only a small part of a very comprehensive programme on fuel cells of many different types, mainly in the U.S.A. and in Western Germany. Single fuel cells deliver an output at approximately 1 V under no-load conditions, falling to about 0·75 V under optimum load conditions at which an operating efficiency of 60–70% is obtainable. This indicates that the device must be considered as having a high internal resistance as compared with a conventional lead-acid battery, but it has the advantage that the voltage on load will be maintained for as long as fuel and oxidant gases are supplied to the cell. By connecting several single cells in series, a battery of suitable voltage (e.g. 24 V) can be constructed capable of delivering several hundred amperes. The Bacon-type fuel cell, for example, which operates on pure hydrogen and oxygen will deliver 1500 amperes per square foot of electrode.[10,11]

A good deal of heat is generated during the chemical reaction in a fuel cell. This depends to a large extent on the type of cell but, in general, small fuel cells may require to be supplied with heat to maintain the chemical reaction. There is a transition point above which a cell becomes self-sustaining due to the heat of reaction and the I^2R losses within the cell. Large cells will require cooling systems to dispose of the heat generated.

For space applications the use of fuel cells requires that supplies of fuel and oxidant gases be carried, and this affects both the weight to power ratio and the life of the power supply. To increase the operating life, considerable scientific effort in the U.S.A. is being directed towards the development of a regenerative fuel cell. This cell uses a closed-cycle system and fuel is re-generated by making use of thermal (e.g. solar), photochemical or radio-chemical energy forms.

The operation of fuel cells in zero-gravity environments can also introduce difficult mechanical problems. However, there is much scope for further development and it is to be expected that fuel cells will play an important part in space power generation over the next decade.

7. Solar Energy

For space vehicles operating within the solar system it is possible to use solar radiation as a source of energy to provide power for built-in instrumentation. The radiation from the sun is electromagnetic with a wavelength distribution consistent with black-body emission at 6000° K. This radiation is made up of about 3% ultra-violet, 45% visible and the remaining 52% in the infra-red portion of the spectrum. Solar energy is a form of radiant energy and, thermodynamically speaking, is in the "intermediate" range of energies. Table 4 gives examples of energy forms of various levels and whereas it is easy to convert "high-level" energy into lower forms, the reverse process is more complex.

Table 4

Conversion of Energy

Energy level	Energy form		
High	Mechanical	Electrical	Potential
Intermediate	Chemical	Radiant	Kinetic
Low	Thermal	High temperature Low temperature	

This simple method of expressing the second law of thermodynamics indicates that although solar radiation can be converted into low-temperature thermal energy, this transformation will occur at less efficiency than when mechanical energy is converted into thermal form.

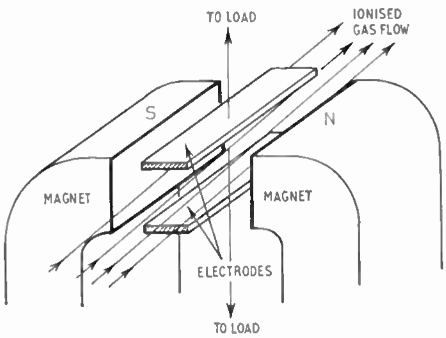


Fig. 8. General arrangement of a magneto-hydrodynamic generator.

Unfortunately the temperatures required to produce effective results with this type of generator are so high (e.g. 3000° C) that progress is almost entirely dependent upon the development of new materials. However, the AVCO-Everett Laboratories in the U.S.A. have been actively engaged in this area of research for some years and have developed an m.h.d. generator which has produced 10 kW for short periods. In this country work is proceeding in the research laboratories of the C.E.G.B. at the C. A. Parsons & Co. Ltd. Nuclear Research Centre in Newcastle-on-Tyne and elsewhere, but it will be some years before practical m.h.d. generators become available for general use.

Space applications of m.h.d. generators must await further progress in solving the basic materials problems and general engineering development of the systems. Proposals for space units have already been put forward by General Electric in the U.S.A. who have demonstrated a 1 kW generator which has operated for a few seconds. This Company has also proposed utilizing the rocket motor exhaust as a source of ionized gas for m.h.d. generation during the boost phase, and a system which uses solar power to heat the gas for operation during orbital conditions. Although complex, the m.h.d. technique is attractive since it offers high energy conversion efficiency (e.g. 60%) compared with certain other generators.

10. Inverters

Since the devices described in this paper all generate direct-current electricity, conversion to alternating-current will require the development of lightweight, efficient inverter systems suitable for space applications. It has been reported recently that the Lockheed Missile and Space Division has developed a solid-state three-phase 400 c/s 115 V inverter providing 500 W on full load when energized from a 22 to 29 V d.c. input. The unit, which is designed for a space environment, is 85% efficient from 30% to 100% full load and 70% efficient at 10% of full load. There is much scope for further research in this field.

11. Acknowledgment

The author is indebted to the National Research Development Corporation for permission to publish this paper.

12. References

1. William Beller, "Space is no place for untried apu's", *Missiles and Rockets*, p. 75, 29th May 1961.
2. G. E. Starkey, "Fuel Cells for Astronautic Application". American Rocket Society publication, September 1960.
3. A good general work on Thermo-electricity is the book "Semi-conductor Thermoelements and Thermoelectric Cooling" by A. F. Ioffe, published by Infosearch.
4. Since the reading of this paper, information has been published by Westinghouse Electric Corp., Pittsburg, U.S.A., relating to the use of samarium sulphide and cerium sulphide as thermo-electric materials up to 2000° F. These compounds exhibit high temperature stability and good thermo-electric efficiency at such temperatures, but the materials are still in a developmental stage.
5. Karl G. Hernqvist, "Thermionic converters", *Nucleonics*, **17**, pp. 49–53, July 1959.
6. D. Gabor, "A new thermionic converter", *Nature*, **189**, No. 4768, pp. 868–72, 18th March 1961.
7. "Direct generation of electricity". *N.R.D.C. Bulletin*, No. 18, April 1961.
8. G. J. Young (Editor), "Fuel Cells". (Reinhold, New York.) (general reference.)
9. "Fuel Cells—Power for the Future". (Fuel Cell Associates, P.O. Box 157, Cambridge 38, Mass., U.S.A.) (general reference.)
10. F. T. Bacon, "The fuel cell: power source of the future", *New Scientist*, **6**, No. 145, pp. 271–4, 27th August 1959.
11. F. T. Bacon, "Fuel cells: will they soon become a major source of electrical energy", *Nature*, **186**, No. 4725, pp. 589–92, 21st May 1960.
12. Charles D. La Fond, "Solar mechanical engine to give 15 kW", *Missiles and Rockets*, pp. 24–8, 15th August 1960.
13. "Isotopic Power". (Martin Nuclear Division, Baltimore 3, Md., U.S.A.)

Manuscript first received by the Institution 1st June 1961 and in revised form 19th June 1961. *(Paper No. 663.)*

CONVENTION RECORD

Although participation by overseas scientists and engineers has always been a feature of Brit.I.R.E. Conventions, this year's fully justified its description as an "International Convention", with no fewer than eleven of the forty-four papers originating from overseas. Seven papers came from the United States, two from the U.S.S.R. and one each from Canada and Australia.

The attendance at the Convention was similarly international, and in addition to the countries represented by authors of papers, a welcome was extended to scientists and engineers from Belgium, Finland, France, Holland, India, Israel, Italy, Kenya, Norway, Pakistan, South Africa and Western Germany.

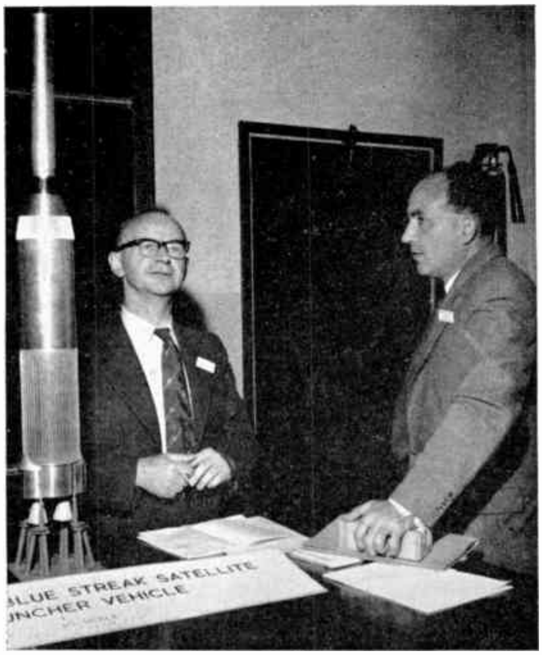
The Chairman's Address

The Chairman of the Convention Committee, Mr. Ieuan Maddock, O.B.E., M.Brit.I.R.E.,† welcomed delegates in the following terms:—

"The subject of this Convention is the most challenging and exciting field which has confronted the Radio and Electronic Engineer in recent times. This is the first conference of its kind to be held in Britain and I am confident that it will do much to encourage the Industry to expand their endeavours in this new field. The organizers of the conference have had this aim in mind at all stages, and this is reflected in our programme.

"Great Britain and the Commonwealth have already played a conspicuous role in the use of Radio Techniques in Space Research—I refer to the work on radio astronomy and in satellite tracking—and we shall be hearing about some of this work during the convention. In the important field of instrumented artificial satellites and space probes, we are still in the embryo stage and it is one of the main aims of this convention to educate and, I hope, to stimulate radio and electronic engineers so that we can show the same invention and skill in this new field that we have displayed in so many other branches of science and technology.

"We thought that it was important to emphasize the severe constraints which face the radio and electronic engineer in satellite work—the limitations of weight, size, power demand, the environmental difficulties, and the heavy premium on reliability—and we have contributions on all these aspects. It is also important that the Radio Engineer should be aware of the scientific uses of satellites and the problems he may be called on to solve for his scientific colleagues. Thus, there are a number of papers from research workers in the Universities, many of whom are engaged in preparation for Britain's first venture into space—the *Scout I* project. This will be a satellite containing British conceived and instrumented experiments to be put into orbit by an American launching system. We look forward to the success of this project and hope that it will be the pioneer in a steadily expanding endeavour in Britain, the Commonwealth and Europe.



The Chairman of the Convention Committee, Mr. I. Maddock, and Mr. D. J. Lyons, examine the 1/30th scale model of the Blue Streak satellite launcher which stood in the lecture theatre throughout the Convention. Mr. Lyons, who gave the opening paper at the Convention, is in charge of the Guided Weapons Department at the Royal Aircraft Establishment, Farnborough.

"Only two countries have direct experience of satellite work to date, and I am sure we all acknowledge and salute the brilliant achievements of our colleagues in America and in Russia. Clearly, no convention of this kind would be complete without a contribution from these experienced workers and I am glad that we have the good fortune of having several important papers from both America and Russia. I am sure that they will not consider that I am inhospitable if I say that I look forward to the time when we will be giving them some healthy competition.

†Representative of the United Kingdom Atomic Energy Authority on the Royal Society's British National Committee on Space Research.

Two authors of papers, Dr. R. L. F. Boyd, reader in physics at University College, London, and Dr. S. J. Bauer of the Goddard Space Flight Center, U.S.A., compare details of their work on electron density measurements.

"It is one of the misfortunes of modern scientific research that it is immensely expensive. Nuclear physics, plasma physics, low temperature studies, solid state physics, biology—all call for vast expenditures on large facilities, yet even in comparison with these, space research appears very costly, and the man who has to shoulder this burden, the taxpayer, must be convinced that some practical value will accrue from this work. Thus, although the title of our convention refers to 'Space Research' we have deliberately included contributions which deal with the more practical aspects of communications and navigation satellites. It is here that the challenge lies for the Industry—a challenge which our President stressed in his recent Address.†

"I have referred mainly to communication satellites because they are a natural focal point for the skills of the radio and electronic engineer. We must not ignore, however, the fact that possibly the most important environment to man is the atmosphere, and our knowledge of the upper atmosphere, particularly the outer ionosphere, is sparse. This is an area where the radio engineer has always played a vital role, and with the aid of rocket and satellite borne instrumentation, he can now help to unravel the intricate physical phenomena which occur at very great heights. This will be of importance to meteorology, communications and high altitude flying, besides yielding information of general value to the physicists.

"What of the future? Astronomy is one of our oldest sciences, but the advent of the 'space age' and radio techniques has given it a new impetus. The next few decades will see a rapid upsurge of our knowledge of the sun, the stars, the cosmos and very mechanisms of creation. Clearly, the future holds great prizes for the scientist, but what of the layman, what can he expect from these great developments. Already we are in sight of satellite systems which will enable us to circle the globe with a reliable and extensive network of communications. Trans-oceanic television links also appear feasible. Satellites may provide accurate and reliable navigation references for the future and also they will provide world-wide data on weather conditions—an essential facility when we enter the era of supersonic aircraft flying above the atmosphere. These and other developments demonstrate that we cannot ignore this challenge and we must be ready to play our part at all stages and to share in the results."

Daily review

The first Session comprised survey papers covering the subjects to be dealt with on the following three days. These proved particularly helpful to delegates

Mr. J. M. Bridges, Director of Electronics in the Office of the Director of Defense Research and Engineering, Washington, reading his paper in Session I.

† Published on pages 473-6 of the June *Journal.*

since they introduced some of the basic problems to which the more specialized papers would try to provide answers.

In the evening many delegates gathered in the Old Lecture Room at Christ Church to see a programme of technical films on subjects connected with the Convention theme. These showed some of the successful firings of British research rockets (*Skylark* and *Black Knight*).

The second day of the Convention dealt with Satellite Engineering papers, grouped under the headings of Systems and Components respectively. Particular interest was aroused by the paper by Mr. John D. Nicolaides, Technical Director of Astronautics of the U.S. Bureau of Navy Weapons in which he described economical launching techniques.

The session on components for use in satellites led to a useful discussion on the problems of reliability and how this might be achieved by microminiaturization. The provision of electric power to operate the electronic equipment in satellites was covered in three valuable papers and attention was given to the environmental effects which add to an already difficult problem.

The second day's programme terminated with a film show, featuring three unorthodox launching projects being developed by the U.S. Navy, the W.R.E. testing facilities at Woomera and the American Project Transit.

Friday, 7th July, was concerned with Communication Satellites. The reception by ground stations of minute signals by parametric amplifiers and masers was first discussed, as well as some of the basic engineering problems. In the afternoon, papers giving specific proposals for satellite systems were presented and the discussions during this session were most useful. Delegates were especially interested to hear the comments made by the President, Admiral of the Fleet the Earl Mountbatten of Burma, K.G. The President had earlier been shown round the Clarendon Laboratory in which the Convention Sessions were being held.

In the evening, delegates attended the Convention Banquet held in the Hall of Christ Church.†

Saturday was the day when the physicists and the radio astronomers came into their own. The papers during the morning were given by physicists who have designed experiments for research in outer space and their papers posed the problems to be solved and in most cases were accompanied by descriptions of the equipment produced by the electronic engineer to do this. It was particularly interesting to be able to compare the results obtained by workers in the U.S.S.R. with those obtained in the United States, notably in the determination of local electron density in the upper ionosphere. Several papers described the instrumentation which is to be carried in the first Anglo-American satellite, *Scout 1*, when measurements will be made of solar x-radiation and cosmic rays. Techniques for carrying out ultra-violet astronomy to be flown in a later *Scout* satellite attracted considerable attention. In this session a description was given of the Canadian Defence Research Board's topside sounder satellite.

After the Christ Church Banquet the President and Sir Lynn Ungoed-Thomas, Q.C., discuss progress on the Charter Petition.

The final session of the Convention was devoted to "Techniques in Radio Astronomy" and the first paper described how radio astronomy can be carried out from rockets and satellites. A description was then given of the new Australian 210 ft radio telescope. A notable discussion on radar measurements of the planet Venus carried out to determine the value of the astronomical unit (the diameter of the Earth's orbit) and the period of rotation of Venus followed. Earlier this year, workers in both the United States and in the U.S.S.R. measured these

†A report of the proceedings and of the speeches made by the President and by the Dean of Christ Church appeared in the August issue of the *Journal*.

parameters and arrived at different results. The difference in the astronomical unit was comparatively small, but the Russian scientist, Professor Kotelnikov, obtained a value of 9 to 11 Earth days for the period of rotation. Mr. Malling and Dr. Golomb at the Jet Propulsion Laboratory, California Institute of Technology, put forward the figure of 225 Earth days. The detailed discussion which followed these papers was particularly useful from the point of view of the contributions made by British radio astronomers from the Jodrell Bank Radio Astronomy Laboratories. Jodrell Bank made measurements of these disputed quantities, but unfortunately at the date of the Convention these had not been fully evaluated. The debate was thus a "tie", and both sides retired to re-calculate their results! It is hoped to give a full report of this discussion in the next issue of the *Journal.*

Institution Conventions usually take place during a heat wave and 1961 was no exception! Mr. P. M. Thompson (Associate Member) and a U.S. delegate, Mr. J. R. Huynen, obviously appreciate the iced soft drinks which were provided instead of the traditional tea during the afternoon break.

Over the three and a half days some 300 delegates took part, and for most of the sessions the Lecture Theatre in the Clarendon Laboratory was packed to its capacity of nearly 200 people. Indeed, a closed circuit television link was called into use on several occasions.

This was the first Convention which has been held on this subject in Britain and its proceedings will serve as an authoritative record of progress in this important new field up to 1961 and provide an inspiring challenge for the next decade.

Other Events

Throughout the period of the Convention a display of working and static exhibits relating to space research and communication satellites was on view in the top floor laboratory of the Lindemann building. The eye of those entering was immediately attracted to the nose cone assembly of a *Skylark* research rocket constructed by the Royal Aircraft Establishment containing an ultra-violet telescope and associated telemetry equipment. Telemetry was also the subject of a working exhibit by E.M.I. Electronics, in which both missile and ground apparatus were shown. Models of proposed satellites were displayed and photographs supplementing several of the papers were mounted on boards around the room.

A number of delegates took advantage of the offer of the Administrator of the Laboratory, Dr. A. J. Croft, to see some of the research work of the low temperature physics, nuclear physics and radio-frequency spectroscopy groups; the latter included the optical maser.

Acknowledgments

This record would be incomplete without acknowledging the great help given by Col. D. V. Hill, the Steward of Christ Church, and his staff, and by Dr. A. J. Croft, Administrator, and his colleagues and staff of the Clarendon Laboratory, all of whom contributed to the smooth running of the Convention.

Members of the Institution's staff added to their duties the midnight production of a daily *Convention News*, which recorded the previous day's happenings and the next day's additional events. This service—a feature of Brit.I.R.E. Conventions—was much appreciated by delegates.

U.D.C. 629.1

The Advantages of Attitude Stabilization and Station Keeping in Communications Satellite Orbits

By

W. F. HILTON, D.Sc.†

AND

B. STEWART, M.Sc.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: The power demands for broadband transmission of multi-channel telephony are far greater than those for telemetry. Satellite communications repeaters might require inputs of the order of hundreds of watts and the weight and cost of solar cells to provide this power is considerable. Attitude stabilization is one method of economy in that it makes possible the utilization of satellite transmitting aerials of high gain. On the other hand, provision of stabilization equipment involves a payload weight penalty which might otherwise be allocated to additional power generators. There exists, for given aerial gain, a satellite mass for which the radiation flux at a given distance is the same whether the satellite is stabilized or not. The paper examines the dependence of this "critical mass" on aerial gain and satellite lifetime and shows that there is an optimum vehicle size for which stabilization pays the greatest dividend.

The advantages and limitations of spin stabilization as a means of attitude control are discussed and the paper concludes with a comparison of random and station keeping (velocity controlled) satellites, the latter being possible where attitude stabilization is employed.

1. Introduction

It is now certain that in the next ten years the world's cable and h.f. radio telecommunications systems will be supplemented by a communications satellite system sponsored by American industry, and perhaps also by a second system sponsored by telephone organizations in other countries including the United Kingdom.

Diversity of route is very welcome as affording the best insurance against interruption of service by the forces of nature and the unreliability of man-made equipment.

Quite correctly, interruption of service is considered to be a most undesirable phenomenon in telecommunications, as it carries with it an uncertainty of when that service will be restored. However, mobile satellites differ from cables, h.f. radio and stationary satellites in this respect, as any interruptions of service will be of known and finite duration, and calls may safely be "booked" during this period, with certainty of connection within an hour or two at most, when the next satellite will be in service.

Having briefly established the point that communications satellites will soon be in use, we will examine the desirability of attitude stabilization of a satellite and the associated question of station keeping.

This paper discusses in detail the use of attitude stabilization to economize transmitted power, and the types of communications satellite to which it applies with advantage. The performance to weight ratio of any component is taken as an arbitrary constant, so as to permit the use of the method with component efficiencies differing from those assumed here.

Stabilization equipment takes account of valuable payload and where only small powers are required attitude control is not necessarily advantageous. For this reason the term critical mass is introduced, and defined as the mass of a satellite which, to an observer at a given distance from it, appears to be radiating the same power whether attitude stabilized or not. In other words that part of the payload which is allocated to stabilization equipment would, if replaced by power generation equipment, provide increased radiated power sufficient to compensate for the loss of antenna gain.

The analysis is treated in a general manner, the control system design not being discussed at all. However, certain parameters in the relationships are taken from an actual design.[1] It must be understood that the critical mass concept does not take account of two other important factors, namely cost and reliability. Ignoring development expenditure, a critical payload is expected to cost less when stabilized, as power generators (solar cells), are currently very expensive. On the other hand, a stabilized satellite is less reliable and will fail completely if the control system fails. Further, if subject to secular disturbing

† Hawker Siddeley Aviation Ltd., Welkin House, Charterhouse Square, London, E.C.1.

torque its life is limited. These factors must be kept firmly in mind when a system undergoes preliminary evaluation.

Spin stabilization is a method of obtaining some aerial directivity without incurring a weight penalty. An active but much simplified control system can be employed in conjunction with spin in certain orbits to obtain the maximum possible gain. This form of stabilization is discussed in some detail in Section 4, as it is probable that a global network based on designs using either this principle or gravity stabilization will eventually be set up. Several American proposals similar in concept have already been published.[2, 3] Velocity control and hence station keeping is also much simplified.

However, a number of reasons why spin stabilization might not find early application are also discussed and these lead to an examination of orbits suitable for fully attitude stabilized designs.

Finally, if a satellite is stabilized then, of course, station keeping becomes possible. Random systems require large numbers of satellites if a high probability of coverage is to be maintained and it may well be that the saving in booster costs which velocity control allows, more than compensates for both limited satellite life and loss of reliability.

2. Sub-systems of a Communications Satellite

In order to obtain a relationship for critical mass it is necessary to express the mass of each sub-system in terms of either the total mass of the satellite or the radio input power. The satellite is broken down into the following sub-systems:—

(1) Solar cells and supporting structure (called the "array").

(2) Radio transmitting and receiving equipment (including batteries, voltage regulators, etc.).

(3) Equipment canister.

(4) Stabilization equipment including sensors.

(5) Stabilization propellant and tanks (if relevant).

It is assumed that the satellite is rigid, i.e. that the array has no special orientation with respect to the sun.

2.1. *Solar Cell Array*

In this analysis the supporting structure is considered not to form part of the equipment canister. Therefore it is reasonable to take the mass of the array proportional to the d.c. power output of the cells.

Thus

$$M_A = K_A P_o \qquad \ldots\ldots(1)$$

where M_A = mass of array
P_o = d.c. output
K_A = constant.

2.2. *Radio Equipment*

Curves for both array mass against d.c. output (Fig. 1) and radio equipment weight against radio input power (Fig. 2) are reproduced from ref. 1. That paper recommends, for reliability and ease of amplifier manufacture, that the transceiver should be subdivided into a number of smaller units called unit repeaters each with their own power pack. Therefore, neglecting the mass of the aerial (small for the large carrier frequencies essential in broad band radio-telephonic communication), it is clear that the mass of the equipment will be proportional to radiated power and thus to input power.

Thus

$$M_R = K_R P_i \qquad \ldots\ldots(2)$$

where M_R = mass of transceiver
P_i = input to transceiver
K_R = constant.

2.3. *Equipment Canister*

If the repeaters are uniformly distributed within the canister then the volume of the latter is proportional to the mass of the repeaters. For constant skin thickness the mass of the canister (M_c) is proportional to its volume to the power two-thirds.

Therefore

$$M_c = K_c'(M_R)^{2/3} = K_c'(K_R P_i)^{2/3} = K_c(P_i)^{2/3} \qquad (3)$$

where K'_c = constant
and K_c = constant.

2.4. *Stabilization Equipment*

In principle there are two methods of stabilizing an orbiting vehicle, by reaction jets, which necessitate on-board propellant, and by inertial or moving mass techniques. However, if some of the disturbing torques acting on the vehicle are secular (accumulative) then a system based on the latter principle would, in general, need to be augmented by the former.†

For simplicity the systems are considered separately here.

2.4.1. Pure jet reaction system

The mass of the propellant distribution pipes is proportional to their length, that is, to the cube root

† In certain cases where a temporary change of reference axes can be tolerated, the tendency of inertial systems to saturate can be obviated without the utilization of reaction jets. Another method, employing a bar magnet to provide desaturation torques, also has been suggested.[4]

of the transceiver mass

$$M_p = K'_p(M_R)^{1/3} = K_p(P_i)^{1/3} \qquad \ldots\ldots(4)$$

where M_p = mass of the pipes

K'_p, K_p = constants.

To the degree of approximation taken here the remainder of the equipment, excluding the tank, can be assumed of constant mass (M_s) and constant power requirement (P_{ST}).

2.4.2. Inertial system

Control by rotating flywheels is most frequently suggested as an inertial device and will be considered here. These flywheels, for economy of weight, should be of annular construction with mass concentrated along the rim.

In order to determine the mass of the system in terms of vehicle mass, it is necessary to assess disturbing torques and their relation to vehicle weight.

It is useful to digress slightly and list them here.

The principal torques acting on a satellite are due to:

(1) residual aerodynamic moments

(2) gravitational gradient

(3) solar pressure including periodic solar disturbances (flares)

(4) Earth's magnetic field

(5) meteoric bombardment

(6) apparent torques arising from rotation of reference axes.

Torques (1) and (3) can be overcome by ensuring that the centre of pressure and centre of gravity of the satellite are as near to one another as possible.

Torque (2) gives rise to a torque which tends to align the axis of minimum inertia with the local vertical. For a stabilized body with an earth seeking antenna it can be overcome by ensuring that the centre line of the latter and the axis of minimum inertia are coincident. In circular orbits this torque can be employed to provide passive stabilization. However, to prevent tumbling, injection errors must be extremely small.

Torque (4) is important for spinning satellites and is considered later.

Torque (5) must be treated statistically: the force exerted can be taken as proportional to surface area.

For earth-seeking satellites in elliptic orbits torque (6) will be predominant. Even in nominally circular orbits, the torques due to (6) may well overpower the combined effects of (1) to (5).

For this study (6) is taken to be applicable, hence a correcting torque proportional to satellite inertia is required.

Now the polar moment of inertia of an annular flywheel (I_F) is given approximately by $I_F = 2\pi a^3 t d\rho$ where a, t, d, ρ, are its radius, rim thickness, rim depth and density respectively.

Thus, if t, d are proportional to a, $I_F \propto a^5$

Therefore

$$(M_F)^{5/3} \propto I_F$$

where M_F is the mass of the flywheel.

For a given angular acceleration of the flywheel the reaction torque developed is proportional to I_F.

Now it is required that the reaction (correcting) torque be proportional to satellite inertia.

Hence $\quad I_F \propto I_V \propto M_F^{5/3}$

where I_V is the vehicle inertia.

For the type of satellite under discussion (rigid and non sun-seeking), shape can be assumed invariable. Hence I_V can be expressed as a function of vehicle mass. For a cylindrical satellite of uniform density

$$I_V = \tfrac{1}{2}\pi r^4 l\rho, \qquad M_V = \pi r^2 l\rho$$

where r, l are its radius and length respectively

M_V = mass

ρ = density

Hence

if $\quad r \propto l, \quad I_V \propto (M_V)^{5/3}$

This result also holds for a uniform sphere.

Hence, taking these vehicle shapes as representative:

$$(M_F)^{5/3} \propto I_F \propto I_V \propto (M_V)^{5/3}$$

or $\quad M_V \propto M_F$

The power consumed by the stabilization equipment, in this case, is almost entirely motor input power which varies with flywheel inertia. One can write

$$P_{ST} = K_{ST} M_V^{5/3} \qquad \ldots\ldots(5)$$

where K_{ST} = constant

Taking motor mass $M_m \propto M_F$ then:

$$\left.\begin{aligned} M_F &= K_F M_V \\ M_m &= K_m M_V \end{aligned}\right\} \qquad \ldots\ldots(6)$$

where K_F, K_m are constants.

As before the remainder of the equipment can be considered of constant mass M'_S.

2.5. *Propellant and Tank*

This item of equipment forms part of the pure jet reaction system only as it has been assumed that methods described in 2.4.1 and 2.4.2 above are not employed together on one vehicle.

Now the couple required for a given response is proportional to $(M_V)^{5/3}$. Thus

$$\text{thrust} = \frac{\text{couple}}{\text{typical dimension}} \propto (M_V)^{4/3}\dagger$$

Thus, as propellant required is proportional to life (T say), a relation can be written down:

$$M_E = K_E T M_V^{4/3} \qquad \ldots\ldots(7)$$

where M_E = mass of propellant

K_E = constant

For a spherical tank, mass of propellant $\propto d^3$ (d = diameter) and mass of tank $\propto d^2 t$.

Now the tensile stress in the shell $= \dfrac{pd}{4t}$

where p = internal pressure

and t = thickness.

Thus for constant stress and pressure $d \propto t$.

Hence mass of tank $\propto d^3 \propto$ mass of propellant, and eqn. (7) can be taken to include the tank.

2.6. *Variation of M_V with Life*

The propellant tank needs to be situated near the centre of gravity of the vehicle in order not to affect the control dynamics seriously. Thus the inertia changes negligibly as propellant is expended and the thrust required for a particular manoeuvre will be constant throughout the lifetime of the satellite.

3. Critical Mass

The mass of a jet stabilized vehicle is given by

$$M_V = M_A + M_R + M_c + M_s + M_E + M_p \ldots\ldots(8)$$

From Section 2, eqns. (1), (2), (3), (4), (7)

$$M_A = K_A P_o$$
$$M_R = K_R P_i$$
$$M_c = K_c(P_i)^{2/3}$$
$$M_p = K_p(P_i)^{1/3}$$
$$M_E = K_E T M_V^{4/3}$$
$$P_o = P_i + P_{ST}$$
$$M_s = M_s \text{ (constant)}$$

Substituting in (8)

$$M_V - K_E T M_V^{4/3} = (K_A + K_R)P_i + K_c P_i^{2/3} + K_p P_i^{1/3} + M_s + K_A P_{ST}$$

With the same notation the mass of an unstabilized satellite is given by:

$$M_V = M_A + M_R + M_c$$

i.e. $$M_V = (K_A + K_R)P_i + K_c(P_i)^{2/3}$$

Therefore if the gain of the stabilized satellite antenna is G, then in order to appear to an observer at a given distance that it is an isotropic radiator of power KP_i (K = transmitter efficiency) the transmitted power must in fact be KP_i/G.

That is, the critical mass is the value of M_V which satisfies simultaneously the following equations. It is, of course, a function of G.

$$M_V - K_E T M_V^{4/3} = (K_A + K_R)\frac{P_i}{G} + K_c\left(\frac{P_i}{G}\right)^{2/3} + K_p\left(\frac{P_i}{G}\right)^{1/3} + M_s + K_A P_{ST} \ldots(9a)$$

$$M_V = (K_A + K_R)P_i + K_c(P_i)^{2/3} \qquad \ldots\ldots(9b)$$

Similarly, the relations for an inertially stabilized vehicle are:

$$M_V = M_A + M_R + M_c + M_F + M_m + M_s' \quad \text{(stabilized)}$$
$$M_V = M_A + M_R + M_c \quad \text{(unstabilized)}$$

giving

$$(1 - K_F - K_m)M_V - K_A K_{ST} M_V^{5/3} = (K_A + K_R)\frac{P_i}{G} + K_c\left(\frac{P_i}{G}\right)^{2/3} + M_s' \ldots(10)$$

and $$M_V = (K_A + K_R)P_i + K_c P_i^{2/3}$$

to be satisfied simultaneously.

Figure 3 shows the variation of critical mass with gain using parameters derived from Figs. 1 and 2 for a reaction jet stabilized vehicle. Figures 1 and 2 are reproduced from reference (1). Curves are shown for satellite lives of 1 and 5 years and the efficiency parameters have been assumed as follows:

$$K_A = 1{\cdot}25$$
$$K_R = 0{\cdot}69$$
$$K_c = 1{\cdot}0$$
$$K_p = 2{\cdot}1$$
$$K_E = 0{\cdot}0067$$
$$M_s = 15$$
$$P_{ST} = 15$$

Mass is in pounds, power in watts, and time in years.

Similar curves can be drawn for an inertially stabilized vehicle.

It can be seen that the curve for $T = 5$ possesses a minimum at $G = 2$ dB; this result is perhaps surprising at first sight, as one would expect stabilization

† To be consistent with 2.4.1 one should write:

$$\text{thrust} = \frac{\text{couple}}{\text{typical dimension of canister}} \propto \frac{(M_V)^{5/3}}{(M_R)^{1/3}}$$

However, the above approximation simplifies the algebra and is sufficiently accurate for most practical cases.

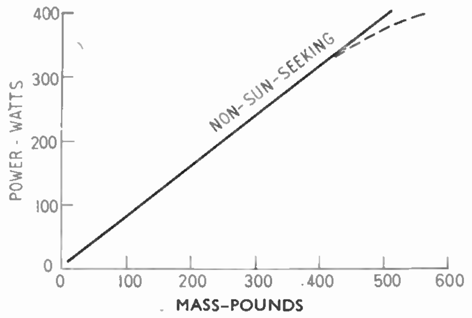


Fig. 1. Power/weight relationships for a solar cell array using 12% efficient cells with glass covers.

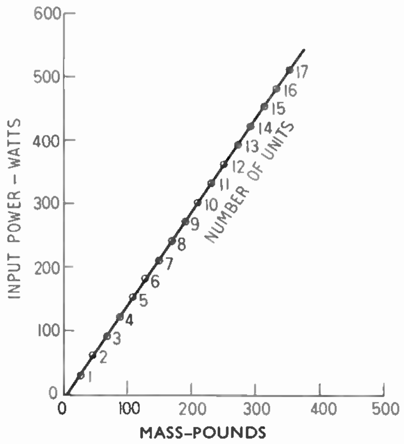


Fig. 2. Input power/weight relationships for 1 watt repeaters.

to pay high dividends at high payload weights. However, examination of equations (9) will show that this is not the case.

Equation (9a) can be rewritten:

$$(K_A+K_R)\left(\frac{P_i}{G}\right)+K_c\left(\frac{P_i}{G}\right)^{2/3}+K_p\left(\frac{P_i}{G}\right)^{1/3}$$
$$=M_V-K_E TM_V^{4/3}-M_s-K_A P_{ST}\ldots(11)$$

The right hand side vanishes for two non-zero values of M_V and, from (9b) these correspond to non-zero values of P_i. Hence for the above equation to be satisfied, $G \to \infty$ in both cases. Physically this result is easily explained. For small satellites radiating low powers the "dead weight" or invariant part of the stabilization equipment can account for a large percentage of the payload. The weight of this equipment might well exceed that of the additional power source required to compensate for the loss of antenna gain. In the limit, when the r.h.s. of (11) vanishes the entire payload is taken up by the attitude control system. For large satellites there exists an upper limit on payload for a given set of values of the parameters. This arises because the mass of propellant plus tank is a function of the vehicle inertia and increases at a greater rate than the mass of the vehicle itself. For the constants listed above and a required life of 10 years these limits can be obtained by equating the r.h.s. of (11) to zero.

i.e. $$M_V-0{\cdot}067M_V^{4/3}-33{\cdot}75=0$$

giving $$M_{V(\min)}=44 \text{ lb}$$
$$M_{V(\max)}=3250 \text{ lb}$$

Obviously the actual values calculated here are dependent on the assumptions that have been made and the values of the constants will change during the development of a system. Further the relationships might need to be modified, for example, if satellite shape is a function of size and weight.

Also in cases where propellant weight is a significant part of the total, variation of vehicle inertia with life will have to be considered. However, it is hoped that the analysis has indicated the importance of a thorough preliminary investigation and brought to light some of the factors affecting the choice of an optimum satellite design.

For example, Fig. 3 shows that for a given set of parameters, there is an optimum critical mass (point A) where, in the case of satellites subject to disturbances proportional to their inertia, attitude stabilization offers the greatest advantage. More generally, the results indicate that even for antennae of little directivity, an economy in weight can be obtained by employing attitude control.

The foregoing analysis, although in principle quite general, in fact considers the case of an earth pointing vehicle with three-axis control. The parameters used in the example are taken from a design study for a repeater for use in inclined elliptic orbits. It will be seen later that such a stabilization system is

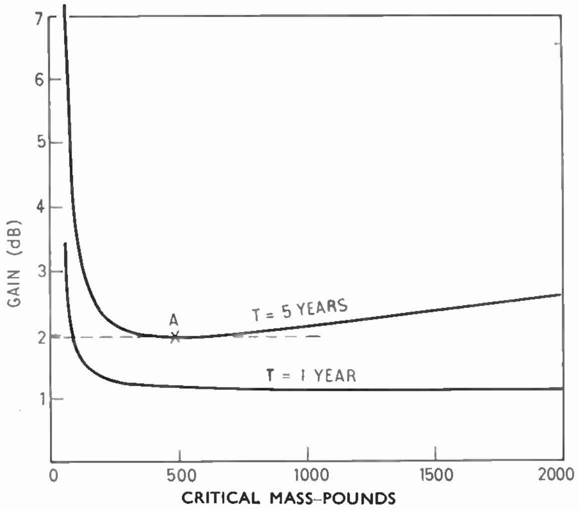


Fig. 3. Variation of critical mass with gain. If payload weight and required gain are plotted and found to be above the curve corresponding to the design life then stabilization pays.

necessary if the chosen communications orbits are of this general type which, although introducing relatively complex problems in attitude control, nevertheless present fewer difficulties from the standpoint of injection into orbit. The actual design of the equipment including the basic dynamics is not within the scope of this paper. It has been considered in some detail elsewhere.[5]

However, a sophisticated stabilization system is not essential for communications satellites, as some improvement over the performance of unstabilized vehicles can be attained by much simpler means. For example, in the case of circular orbits, gravity stabilization is possible. Another technique allowing an improved power output for a given payload, is the sun-seeking solar cell array which, by suitable design of the panels, can be made to "weathercock" under the pressure of the sun's light. These systems are termed "passive" and are not subject to the general analysis of Section 2. Finally, there is a third method which can fall into either classification—spin stabilization. In its passive form this method is the easiest to implement and has already found wide application in the American space programme. It will be shown later that the system is eminently suitable for satellites in equatorial orbits, and as these orbits are ideal for communications, a somewhat detailed discussion is perhaps worthwhile.

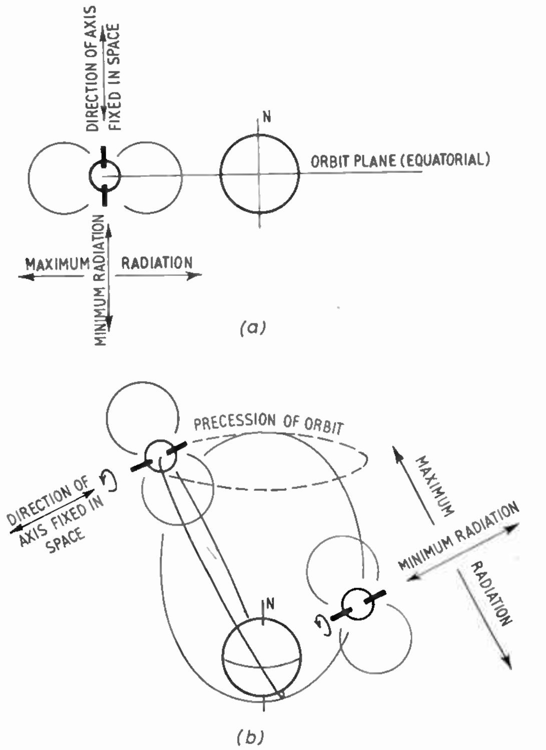


Fig. 4. Spin-stabilized communications satellite with axi-symmetric lobe pattern.

4. Spin Stabilization

By imparting spin to a body about its principal axis of either least or greatest inertia a certain measure of attitude control is obtainable namely the directional invariance of the spin axis. This type of control, in the absence of large disturbing torques, requires no power and, as such, is described as passive stabilization.

If uni-directional aerials were required for a particular mission (e.g. interplanetary travel) this type of stabilization would be ideal in that the spin axis could be aligned with the centre line of the earth-pointing main radiation lobe. However, for inter-continental communications the lobe is required to point to the earth and, if its centre line is coincident with the spin axis, this necessitates precession of the satellite and hence the adoption of an active control system.

Further, the thrust programme presents technical difficulties as the couple required to cause this precession is fixed in direction and therefore varying with respect to axes fixed in the satellite. As the jets producing the couple would be fixed to the vehicle, an impulsive or time varying thrust programme would be essential.

Nevertheless, some effective gain can be obtained by orienting the vehicle with its spin axis normal to the orbit plane and utilizing an aerial having an axi-symmetric lobe pattern (Fig. 4(a)).

For a satellite having this lobe pattern and moving in a circular orbit, the intensity of radiation received on the ground is constant and independent of the spin. For example, if the satellite's antenna has a field equivalent to that of a half-wave dipole then at points on the earth's surface lying in the orbit plane, the transmitter will appear to have a gain of 2 dB. If the orbit is circular, height 6000 miles, the maximum gain that can be obtained while ensuring world coverage from a pattern possessing axial symmetry is 4 dB. These patterns are easily produced in practice, and represent a further point in favour of the stabilized satellite as the problem of engineering an isotropic radiator is formidable when the transmitter is large compared with the wavelength.

However, although the spin axis of the satellite is fixed in direction, in general the normal to the orbit plane is not.

Due to the earth's oblateness and the consequent lack of spherical symmetry in the gravitational field, near-earth orbits suffer two major perturbations[6]—a rotation of the orbit plane about the earth's axis, known as precession of the nodes, and secondly, a rotation of the apse line (or major axis) in the plane of the orbit. It is the former perturbation that prevents the use of the above technique in all but polar and near-equatorial orbits. Inclined orbits precess and the situation shown in Fig. 4(b) arises; after a precession of 90 deg there exists a satellite position where the earth "looks at" a region of minimum radiation in the lobe pattern. This perturbation is proportional to the cosine of the angle of orbit inclination to the equator and thus vanishes for polar orbits. For very small inclinations precession, though non-zero, is of little consequence as variations in the normal to the plane so caused are small.

4.1. *Stability of Rotation*

As stated earlier, a spin imparted about a principal axis of least or greatest inertia stabilizes a rigid body to the extent of fixing the direction of its spin axis in space. This statement holds subject to the conditions that the body is infinitely rigid (no internal dissipation of energy), and is subject to no external disturbing forces. No real body satisfies the first condition and if the second held there would be no point in stabilizing the satellite at all.

When random disturbances are present (Section 2) and the vehicle is subject to strain, only rotation about the axis of greatest inertia is stable and therefore, in an actual design it must be coincident with the axis of spin. In practice the other two moments of inertia will be equal. Any precession of the spinning body caused by external disturbances can be readily damped out.[7] Of the torques listed in Section 2.4, the effects of (1), (2), (3) and (5) can be minimized by careful design and (6) is not applicable. It is estimated that a spin rate of 50–100 rev/min would be sufficient to stabilize a communications satellite effectively for several years. The behaviour of a spinning body in the presence of (4) (the earth's magnetic field) is discussed below.

4.2. *Effect of the Earth's Magnetic Field*

The existence of an external magnetic field is particularly troublesome in the case of spin-stabilized bodies. Most important is the effect of the field on the rate of spin. This is essentially different from other phenomena that cause disturbances in that a torque is generated by the spin itself and acts so as to decrease the spin. This is a well known phenomenon associated with any conductor rotating in a magnetic field. The derivation is detailed in Ref. 8. Except for satellites containing unsaturated ferro-magnetic materials the decay rate in a uniform field is exponential with a time constant which can be predicted with fair accuracy from the known physical constants of the components, the orbit and the vehicle orientation. The decelerating torque varies as the square of the local magnetic field and the sine of the angle between the spin axis and the local field vector. Thus the torque is very small for satellites in equatorial orbits with spin axis normal to the orbit plane and near maximum for polar orbiting satellites similarly oriented with respect to their plane.

The earth's magnetic field varies as the inverse cube of the distance from the earth's centre and the torque varies as the inverse sixth power of the distance. Figure 5 shows the rapid change in half life with height for a vehicle with the following design parameters moving in a polar circular orbit. The satellite is assumed to be a homogeneous sphere with a thin aluminium skin.

In M.K.S. units:

skin thickness = 0·001 m
mean density of satellite = 200 kg/m^3
radius = 0·5 m
bulk resistivity = $2{\cdot}8 \times 10^{-8}$ ohm m (aluminium)

For simplicity a mean value of the field at the earth's surface is taken, and the intensity at distance r from the earth's centre is given by

$$B = 3 \times 10^{-5} \left(\frac{r_o}{r}\right)^3 \text{ webers/m}^2$$

where r_o is the earth's radius
r is the radius vector to the satellite.

The curve is derived from the relationship

$$\tau = 7{\cdot}84 \times 10^{-2} \left(\frac{r}{r_o}\right)^6, \text{ where } \tau = \text{time constant}$$

It can be seen that orbits of altitude of less than 3–4000 miles are unsuitable, with purely passive stabilization, if satellite lives of more than 2 years are envisaged. For example *Vanguard II* (perigee and apogee heights of 350 and 2000 miles respectively) has a decay time constant of 72 days. This rate is unacceptable for long life communications satellites

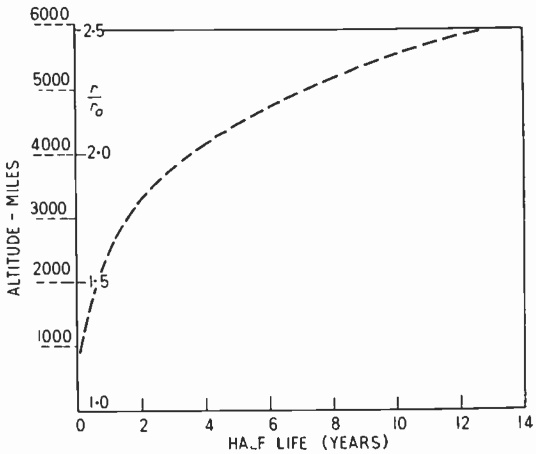


Fig. 5. Circum-polar orbits—spin axis normal to orbit plane.

and would demand an active spin rate system on board including a supply of propellant.

4.3. *Station Keeping*

Station keeping presents no dynamical problem where spin stabilization is employed as a ready reference exists for orbital velocity correction. As the spin axis of the satellite is normal to the orbit plane, it is possible by setting the reaction jets normal to the spin axis, to ensure that the thrust vector lies in the orbit plane. If, in addition, an infra-red horizon sensor is attached to the vehicle at 90 deg to the jet nozzle (nozzles) then by initiating an impulsive thrust at the instant the sensor passes through the local vertical, the coincidence of the thrust and satellite velocity vectors is assured. This, of course, is only strictly true for circular orbits but assuming velocity correction is made at closest approach of the satellite the technique is applicable to eccentric orbits. The disadvantage of the method is an engineering one. If secondary perturbations are to be avoided the thrusts must be of very short duration. A deviation of thrust vector from velocity vector of ± 15 deg is reasonable which allows a maximum thrust period, for a spin rate of say 50 rev/min of 0·1 sec. If a change in velocity of 10 ft/s is required approximately 1400 impulses from a jet of 1 lb thrust are needed for a payload of 200 kg, taking 27 minutes to complete. The minimum velocity increment corresponding to one impulse is 0·007 ft/s which is of the order of accuracy required for a satellite of 5–10 years life. However, a controlled thrust of 1 lb lasting 0·1 sec is difficult to achieve and it may be necessary to employ smaller thrust units and effect the larger velocity corrections over long periods of time.

4.4. "*Active*" *Spin Stabilization*

The principle of spin stabilization can be employed in an active control system having only 1 deg. of freedom. Such a system is very much simpler, lighter and more reliable than three-axis control yet for communications purposes, it provides all the advantages of the latter when used in designs intended for equatorial or polar orbits.

The active system utilizes a flywheel with its axis coincident with the satellite axis of symmetry and which replaces the main body, as the angular momentum store, thereby maintaining the axis fixed in direction as before.

The outer shell of the vehicle, with aerial rigidly attached to it is rotated with respect to the flywheel by an electric motor the rotor of which can be considered integral with the flywheel and the case (stator) rigidly attached to the outer shell (or main body). As the total angular momentum of the vehicle is constant, the absolute angular velocity of the main body can be fixed at any desired value by an appropriate adjustment in rotor speed; in particular, if this angular velocity is made equal to that of the radius vector from the earth's centre to the satellite, it can be arranged to have the antenna earth seeking.

A further sophistication, possible for equatorial orbiting satellites where the sun is never more than $23\frac{1}{2}$ deg from the orbit plane, is a system incorporating both an earth seeking antenna and a sun seeking solar cell array. This is a logical extension of the previous technique, the satellite being divided into three coaxial components, a flywheel, an aerial structure and a solar cell panel. In this case two motors are required, one controlled by a sun seeker and the other by a horizon sensor.

Designs such as these minimize the station keeping problem because the velocity correction jet can be permanently aligned with the vehicle path and thrusting continuously if required.

It follows that a general analysis, similar to that in Section 2, can be applied to ascertain whether advantage can be gained from the above refinements to a basically simple spin stabilized communications satellite.

For reasons of space the relationships are not derived here but it is clear that at high orbiting altitudes when the maximum useful gain is large compared with that of an axi-symmetric lobe (13 dB gain at 6000 miles altitude), these design features should be seriously considered. This is specially important for high circular orbits where payload is severely restricted.

4.5. *Limitations of Spin Stabilization*

Undoubtedly spin stabilization permits a simple and therefore reliable attitude control and velocity correction system. Further, analysis shows that equatorial circular orbits—possible orbits for this type of stabilization—are suited to global communication requirements and especially to Commonwealth communications. There are, however, a number of reasons why satellites of this form are unlikely to be first in the communications field.

Briefly these are:

(1) Limitation in choice of orbit (equatorial or polar) where reasonable antenna gain is required.

(2) As there are no equatorial sites in existence, equatorial orbits are not possible at the present time.

(3) Difficulty of putting large payloads into high circular orbits with present boosters. (The need to carry fuel to orbiting altitude for injection purposes imposes this restriction.)

(4) Problem of achieving injection accuracy increases with height. (High orbits are essential for long life because of the magnetic effect.)

For these reasons a system of station keeping satellites to be discussed later, is based upon inclined elliptic orbits.

5. The Advantages of Station Keeping

Whatever form of attitude stabilization is employed on a particular design additional complexity is involved in the provision of a velocity correction system. However, if major corrections are made at injection, and by means of third stage vernier controls, the extra weight of propellant required in augmenting the attitude control equipment with this sub-system should not be very great.

It has been shown that a spin stabilized vehicle can derive a velocity correction reference directly from its horizon sensors. Furthermore, in the case of earth pointing satellites, it can be shown that yaw displacement can be determined without reference to an external source from the yaw-roll cross coupling that results from the rotation of reference axes. These techniques might obviate the need for additional sensing equipment and in any event lead to the conclusion that, where an attitude stabilization system is employed, velocity correction might well be attempted. In view of this a comparison of random and station keeping systems has been made for specific communications links.

6. Random Systems

Some investigations of random distributions of repeater have already been undertaken, notably on a system of 3000 miles altitude polar circular orbits.[2] It was found that for a particular link (Newfoundland—Hebrides) 10 and 19 satellites are required to give 90% and 99% probability of service respectively and further that approximately 50 are necessary for worldwide communication. Ground repeaters are, of course, imperative for the longer links.

An analysis of the number of repeaters needed in an absolutely random distribution over the globe to serve a system of five ground stations has been completed. In practice the distribution will not be random unless launch direction is random as orbit inclination does not vary greatly with life, but the results are useful in that they set a maximum to the number required. The link chosen was New York—London—Aden—Singapore—Sydney.

Typical results for 6000 miles altitude circular orbits are:

Number of Satellites	*Probability of Service*
36	95%
50	99%
68	99·9%

A station-keeping system designed specifically for Commonwealth communications has been presented by Dauncey.[9]

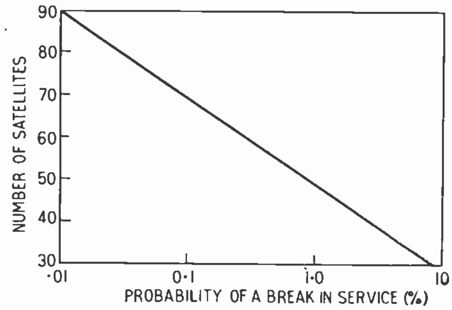


Fig. 6. 6500-mile altitude equatorial orbit.

This paper proposes a ring of near equatorial ground stations distributed around the earth into which stations of higher latitude feed their intercontinental telephone traffic. This system, employing eight satellites in 6500 miles altitude equatorial circular orbits provides continuous 24-hour service. A similar system requiring fewer ground repeaters has been taken as a basis for direct comparison of station keeping and random satellites. The ring ground stations in this case are:

Nairobi—Calcutta—Sydney—Fanning Island—Jamaica—Ascension Islands—Nairobi. These form a closed loop, Fanning and Ascension Islands with little traffic of their own being included solely as repeaters.

The numbers of satellites moving in the above orbit for given probabilities of service are plotted in Fig. 6. A reasonable probability to take as a comparison with a station keeping system is 99% and for this value the curve shows that 49 satellites are required.

7. Station Keeping Systems

The most frequently proposed system in this class is one based on the 24-hour stationary orbit. This is an equatorial circle of height 22 300 miles and a body moving in this orbit circles the globe once in 24 hours and hence appears stationary to an observer on the earth's surface. At this orbiting height three satellites would provide continuous service to all inhabited regions and in theory would not require steerable ground tracking antennae. However, the injection accuracies required to utilize this property successfully are considerable and almost certainly beyond the current state of the art. Another equatorial system which will present fewer launching problems once a suitable site exists is the subject of the proposal discussed above. An immediate comparison is offered here as Dauncey[9] has shown that eleven station keeping satellites are sufficient for continuous 24-hour service. In this case velocity correction allows a reduction of nearly four-fifths in the number of launches and hence, ignoring capital investment in sites, a reduction of the same amount in launching costs.

Yet another system based on inclined elliptic orbits has been proposed[10] which can be set up using existing launching sites and in this case the orbits have been chosen to favour the Northern hemisphere where the demand for communication is greatest.

The constants used in Section 2 have been taken from a satellite designed specifically for this system and the same reference gives the following values for the orbit parameters:

Period = 6 hours
Perigee height = 300 st. miles
Apogee height = 12 600 st. miles
Orbit inclination = 63·4 deg
Latitude of apogee = 63·4 deg N

At an inclination of 63·4 deg one of the principal perturbations caused by the oblateness of the earth, the rotation of the apse line, vanishes. Therefore the latitude of apogee remains constant and it is this property that makes the orbit of particular interest for communications purposes. Studies have been carried out in two cases—firstly with apogee at 63 deg N to serve the world's principal centres of communication, *viz.* places of latitude greater than 45 deg N, and secondly, for a family of orbits with apogees on the Equator for Commonwealth communications.[11]

As an example of the coverage obtainable in the first case it is found that four satellites in four orbits of period six hours each, spaced at 90 deg intervals around the earth's axis give 24-hour service between all points above 50 deg N and within 100 deg longitude separation and that five satellites are sufficient for all pairs of stations above 45 deg and within 100 deg longitude separation. Thus if two "hops" are acceptable a continuous service is possible between every pair of stations above 45 deg N and this is sufficient to feed the internal networks of North America, Europe and Japan.

8. Conclusion

It appears that attitude stabilization pays handsome dividends in economy of weight for communications satellites even in cases where only small antennae gains can be tolerated. There is, of course, a reliability penalty, but this will be minimized when satellite systems that can employ spin stabilization are in existence. Finally, apart from the advantages of antennae gain, attitude control makes possible velocity correction and hence station keeping, which, it has been shown, allows a reduction of perhaps 80% in the number of repeaters required in a commercial communications system.

9. References

1. B. Buss and J. R. Millburn, "A proposal for an active communications satellite system based on inclined elliptic orbits", *J. Brit.I.R.E.*, **22**, pp. 209–25, September 1961.
2. J. R. Pierce and R. Kompfner, "Transoceanic communications by means of satellites", *Proc. Inst. Radio Engrs*, **47**, March 1959.
3. B. Miller, "Hughes communication satellite detailed", *Aviation Week*, 73, 12th December 1960.
4. J. J. Adams and R. F. Brissenden, "Satellite Attitude Control Using a Combination of Inertia Wheels and a Bar Magnet", NASA TN. D-626, November 1960.
5. B. Stewart and P. A. E. Stewart, "Dynamics and Engineering of Satellite Attitude Control Systems". Paper presented at the European Space Symposium, London, June 1961.
6. D. G. King-Hele and D. M. C. Walker, "Methods for predicting the orbits of near earth satellites", *Brit. Interplanetary Soc. J.*, 17.
7. "*Tiros I* spin stabilization", *Astronautics*, June 1960.
8. Yu. V. Zonov, "On the Problem of the Interaction between a Satellite and the Earth's Magnetic Field". U.S.S.R. Academy of Sciences, 1959.
9. S. R. Dauncey, "Ground Stations for Equatorial Circle". Hawker Siddeley Aviation Internal Report.
10. Extract from *The Aeroplane* on Hawker Siddeley Aircraft Ltd. 1st Brochure on Space.
11. W. F. Hilton, "Communications Satellite Systems suitable for Commonwealth Telecommunications". Paper presented at B.I.S. Symposium on Communications Satellites, London, 12th May 1961.

Manuscript received by the Institution on 19th June 1961.
(Paper No. 664.)

DISCUSSION on

"Communications Satellites"

Dr. W. F. Hilton: I would like to question the rejection of s.s.b. in the paper by Messrs. Wright and Jolliffe.† While this is quite valid for a link between two points, communications satellites open up the possibility of area communication between ten or more ground stations through each satellite. If s.s.b. transmission is used from ground stations to satellite, the combined signals can be separated as required by simple selective receivers located one at each receiving site. F.m. would present a more difficult problem, or else demand many radio links in the satellite. The choice is therefore a difficult one and not to be dismissed lightly.

Messrs. W. L. Wright and S. A. Jolliffe (*in reply*): We did not reject s.s.b., but simply stated its limitation. Pulse code has the advantage over s.s.b. in that power can be exchanged for bandwidth in the former system, but this flexibility is not possible in a s.s.b. system.

Dr. Hilton has rightly called attention to an important feature of s.s.b., i.e. its suitability for use in systems in which several transmitters in different localities can simultaneously engage a given satellite using a common repeater. We agree that f.m. is not well suited to this "area" use, since it would be necessary to employ separate r.f. carriers for each ground station transmitting f.m. and to allocate to each a separate frequency band. The use of a common satellite repeater is possible, but technically difficult since intermodulation problems place severe limitations on permissible phase non-linearity.

Pulse code systems may be designed so that they share with s.s.b. the facility of area servicing by arranging groups of pulses, each representing the total information corresponding to a particular area, to form part of the total combined spectrum on a frequency division multiplex basis.

Mr. J. A. Taylor: Regarding the attitude stabilization equipment described by Messrs. Buss and Millburn,‡ what accuracy of stabilization is envisaged?

Secondly, what the ratio of "jet-off" to "jet-on" time has been assumed if the average power for operating the valves is 5 watts?

Finally, what causes the "peak loads of a few milliseconds"?

Messrs. B. Buss and J. R. Millburn (*in reply*): We expect to obtain an accuracy of ± 1 deg; allowance was made for this in the comparison of aerial radiation patterns in Fig. 5. The "jet-on" to "jet-off" time varies widely as the satellite moves round its elliptic orbit, but is always very small, i.e. attitude correction is not applied continuously but in relatively infrequent short bursts. Further details are given in reference 21. Large peak loads of a few milliseconds duration may occur if several pairs of jets operate simultaneously.

The power taken by the solenoid valves could well be several orders lower than the figure we have assumed, but by allocating 5 W for this purpose we have made sure that sufficient power is available for reliable operation at all times.

Mr. C. R. Ditchfield: In Fig. 6 there is a large difference between curves C and E and I question the assumptions involved for the following reasons:

(*a*) 0·5 dB is a larger loss than one would expect from good system design; i.e. a contribution of 30° K from the feeder is excessive.

(*b*) 1 dB noise factor for a maser corresponds to a noise input temperature of 75·1° K. This is very much larger than the 2·5° already achieved by B.T.L. using a travelling wave maser.

Thus any conclusions based on the precise quantitative form of curve E are, I think, open to question.

Messrs. Buss and Millburn (*in reply*): The loss of 0·5 dB in the feeder allows for the possibility of placing the maser in a more accessible position than at the focus of the 100 ft dish. This may be desirable in a commercial communications system, where the periods available for maintenance are determined by the orbital pattern and may not coincide with good weather conditions.

A generous allowance for maser noise was included in Fig. 6 because the r.f. bandwidth of about 110 Mc/s quoted in Section 8.2 is several times larger than that of any existing maser (so far as we are aware). Extension of operating bandwidth is almost certain to be accompanied by an increase in effective noise temperature, though the amount of the increase may be somewhat less than the value we have taken.

(*Communicated.*) Since the Convention we have realized that there was a discrepancy between the 15–20° K of Fig. 6 and the 1 dB quoted in the text of the preprint. We agree with Mr. Ditchfield that 1 dB is excessive, and have now altered this to 0·3 dB to agree with the value given in Fig. 6. The channel capacity of the satellite is not affected by this correction, as Section 8 was based on Fig. 6 in the first place.

Dr. E. K. Sandeman: Why is it proposed to use f.m. instead of p.c.m., which taking into account peak rather than mean powers requires less than half to one-third times the mean r.f. power and so reduces the required weight and size of the satellite by an appreciable factor?

Messrs. Buss and Millburn (*in reply*): We considered only f.m. for the proposed satellite for two reasons:

(*a*) Preliminary studies of p.c.m. (which were admittedly limited to p.c.m./f.m.) seemed to indicate that the difference in power requirements was insignificant.

(*b*) We were not aware of any operational p.c.m. systems capable of handling many hundreds of circuits,

† W. L. Wright and S. A. Jolliffe, "Optimum system engineering for satellite communication links", *J. Brit.I.R.E.*, **22**, 1961. (*To be published.*)

‡ B. Buss and J. R. Millburn, "A proposal for an active communication satellite system based on inclined elliptic orbits", *J. Brit.I.R.E.*, **22**, pp. 209–25, September 1961.

whereas multi-channel f.m. has been used for many years on microwave point-to-point links and the techniques involved are well known.

There is a tendency for communication satellite proposals to be based on techniques which are at a very early stage of development. We feel that communication satellite engineers would do well to follow the example set by the designers of the first transatlantic telephone cable, and base their proposals on techniques and equipment which have a proved performance and life.

Mr. L. Malling: I seriously question the use of vacuum tubes for satellite communication applications; I suggest the correct approach is all solid-state devices.

Messrs. Buss and Millburn (*in reply*): We do not agree that solid-state devices are necessarily more reliable than vacuum tubes for space applications. Vacuum tubes can be made sufficiently rugged to withstand launching stresses and, once established in orbit, are working in an almost ideal environment; solid state devices, on the other hand, are susceptible to radiation damage and may well have a shorter life than the cathode of a vacuum tube.

Mr. B. Rees: I would like to ask why such a highly elliptic orbit has been chosen by Messrs. Buss and Millburn.

Messrs. Buss and Millburn (*in reply*): An elliptic orbit was chosen for two main reasons:

(*a*) The payload which a given booster can place into an elliptic orbit of apogee height x is larger (by about a factor of two for our example) than for a circular orbit of altitude x.

(*b*) The reduced velocity of the satellite at apogee, compared with that in a circular orbit of the same height, means that longer communication time can be obtained between a given pair of ground stations.

The chosen orbit is not so "highly elliptic as the 300-mile perigee and 12 600-mile apogee suggest. The eccentricity is only 0·6 for the true distances to the focus (i.e. the centre of the earth) at perigee and apogee are 4300 miles and 16 600 miles respectively, as shown in Fig. 3.

We do not claim that elliptic orbits are necessarily the best for communication purposes, but we have attempted to show in this paper that the apparent disadvantages of elliptic orbits (such as the change in angle subtended by the earth) can be overcome.

Admiral of the Fleet the Earl Mountbatten of Burma, K.G. (*President*): I would like to ask Dr. Hilton† if he feels that the equatorial circle is the best orbit, and should we not waste time in setting up elliptical orbits from Woomera, but utilize our existing investment in Christmas Island and Gan?

Dr. W. F. Hilton (*in reply*): I do not agree unconditionally that the equatorial circle is the optimum orbit. More people live in the Northern Hemisphere, and a service between 60° N and 30° S would be the maximum requirement. In spite of the investment in buildings, we should also need a liquid oxygen plant and launching pad. Some £10 millions of Woomera's £93 millions would need to be duplicated on the equatorial site.

In addition, the equatorial circle is not ideal from the point of view of ground relay stations. In order that a telephone subscriber may be connected to any other subscriber, it is not sufficient that they can both see one satellite in the system, unless they both see the same satellite. Relay ground stations must be provided which always see two satellites.

Now, for an equatorial circle, most of these ground relay stations will be within 20° or 30° of the equator, some, in places like Christmas Island or Ascension Island, would be quite uneconomic by themselves.

Other orbits demand different "ring mains" of ground relay stations. Thus an orbit plane at 63 deg with apogee over 20° N covers 60° N to 30° S, and has an optimum "ring main" at about 47° N. Therefore many of the relay stations can be chosen in revenue earning positions such as London, New York, Vancouver and Tokyo. However, if all ground relay stations are restricted to Commonwealth territories an equatorial orbit appears much more favourable.

Surely, however, the choice of orbit must await a definite proposal for the telephone requirements from the G.P.O.?

Mr. J. D. Barr: It would appear possible to stabilize the attitude of a satellite in a circular orbit by exploiting the gradient of the gravity field. If the satellite is made tall in relation to its width, then it will oscillate about the vertical with a long period. One method of damping the oscillation was recently demonstrated at the G.W. Department of the Royal Aircraft Establishment. Their proposal also involved azimuth stabilization by making all three of the principal moments of inertia different.

I would like to ask Dr. Hilton whether the advantages of simplicity and weight saving could perhaps outweigh the disadvantages of a circular orbit?

Dr. W. F. Hilton (*in reply*): Gravity stabilization is certainly possible in principle for satellites moving in circular or near circular orbits. The condition for the principal moments of inertia is $B > A > C$ when A, B, C are the roll, pitch and yaw moments respectively. If passive dampers are employed, the complete stabilization system can be made passive and therefore reliable.

However, there are disadvantages. The control torques that can be realized by this method are extremely small. For example, a vehicle having moments A, B, C of 200, 300, 100 slugs feet^2 respectively, moving in a 5,000-mile altitude circular orbit will experience a yaw stabilizing torque of $2{\cdot}3 \times 10^{-7}$ lb ft per degree displacement. Although this value is sufficient to overcome natural disturbing torques (provided the vehicle is precisely symmetrical in shape), it is not large enough to ensure stability in the presence of torques caused by thrust misalignments should station-keeping (i.e. velocity correction) be attempted.

A further point is that active damping might be required in many applications, as passive dampers need to be fairly massive to have reasonable time constants.

† W. F. Hilton and B. Stewart, "The advantages of attitude stabilization and station keeping in communications satellite orbits", *J. Brit.I.R.E.*, **22**, pp. 193–202, September 1961.

U.D.C. 523.164: 629.19

Proposed Satellite Techniques for Performing a High Resolution Survey of the Radio Sky at Medium Wavelengths

By

R. C. JENNISON, Ph.D.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: The reception pattern of a short dipole above the F layer is considerably modified by refraction in the ionized medium and has the form of a cone with sharply bounded edges modulated by a fine interference pattern. The properties of this reception pattern are discussed and suggestions are made for satellite experiments based on this phenomenon. It is shown that a variety of techniques are available to give resolving powers of any desired value down to a fraction of a degree.

1. Introduction

Certain features of the radio sky at frequencies of the order of a few megacycles are to be expected by extrapolation from surveys at higher frequencies. It is likely that radiation from the corona of the galaxy may predominate whilst the galactic plane will appear as a dark absorption band. The well-known radio stars will be insignificant, as well as the quiet sun. It is not known if other features exist which do not appear on surveys at higher frequencies, and in this connection the accidental discovery of emissions from Jupiter at 20 Mc/s may be quoted as an example of the possibility of observing unexpected phenomena in the many unexplored octaves of lower frequency.

Reber and Ellis[1] have endeavoured to perform surveys of the sky at medium wavelengths by ground observations through "holes" in the ionosphere. The results from these experiments are not easy to interpret though they yield an order of magnitude for the total flux from the galaxy.

It is clearly of importance to perform a survey of the sky at low frequencies with aerials of high resolving power and without the need for the rare, difficult and complicated propagation to ground stations through the earth's ionosphere.

A high gain aerial at frequencies of the order of 1 Mc/s would be an enormous structure many miles across and could not be deployed from a satellite in the foreseeable future. Interferometer systems between parts of a satellite connected by long wires and between separate satellites can be conceived, but the separation would have to be many miles before fine beams could be achieved. Problems ensue in controlling the orientation of such systems during the lifetime of the satellite.

The situation appears very different when consideration is given to the effect of placing the aerial in the region of decreasing electron density above the F layer of the ionosphere. Even a short dipole may be shown to acquire a high gain and narrow beamwidth in these circumstances.

2. The Directivity Pattern of an Aerial above the Ionosphere

The gradual decrease in the electron density modifies the directivity pattern of an aerial immersed in the medium. This situation was first investigated by Haselgrove, Haselgrove and Jennison[2] who showed that the beam pattern is focused into the direction of the vertical and is modulated by fine interference fringes which increase in amplitude towards the edge of the pattern where the termination of the beam is extremely sharp. The above paper indicated how the variations of signal strength could be obtained for any orbit or ionosphere by digital computation; it originally contained a number of errors and obscurities which have since been amended.[3] A more rigorous wave treatment confined to an investigation of the modification of the directivity pattern of an aerial inclined at various angles above the F layer has since been performed by Budden.[4]

These papers indicate that if radiation is emitted or received at an angle θ to the vertical by an aerial in the ionosphere, the corresponding angle far beyond the ionosphere will in general be different due to refraction. If this second angle is θ' then, assuming only that the ionosphere has symmetry about the vertical axis through the satellite, the relative power gain of the aerial at angle θ is given by

$$\frac{\sin\theta\, d\theta}{\sin\theta'\, d\theta'}$$

which may be written

$$\frac{\cos\theta'}{n\sqrt{(n^2-\sin^2\theta')}}$$

where n is the refractive index of the ionosphere at the satellite.

† Nuffield Radio Astronomy Laboratories, Jodrell Bank.

If, as an approximation, the earth is considered to be flat, the aerial beam is confined to a cone whose semi-angle θ'_0 is determined solely by the refractive index in the vicinity of the satellite, the semi-angle θ'_0 being given by $\sin \theta'_0 = n$. The radiation at the edge of the cone is that associated with the horizontal radiation at the aerial. This approximation is reasonable for an aerial deeply immersed in the ionosphere and operating on a frequency just above the plasma frequency but it breaks down for higher frequencies or more distant orbits. The limitation in the angle of radiation or reception is naturally equivalent to an increase in the overall gain of the aerial and useful gains of ten or twenty may be obtained in this way before ionospheric irregularities cause serious distortion.

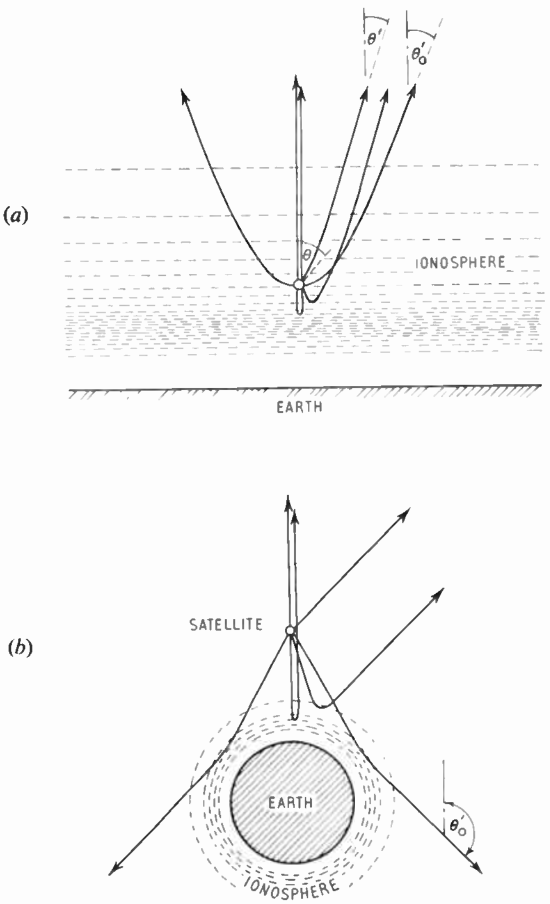


Fig. 1. (a) Ray paths to a satellite aerial slightly above the maximum of the F layer when the receiver is tuned to a frequency slightly higher than the ambient plasma frequency (flat earth and plane stratified ionosphere approximation). A ray leaving the satellite at angle θ emerges at an angle θ' at infinity. The limiting ray emerging at θ_0' leaves the satellite horizontally ($\theta_0 = 90°$).

(b) The corresponding ray paths when the satellite is at a greater distance from the earth and the flat earth and plane stratified ionosphere approximations are no longer valid. The limiting ray emerging at θ_0' no longer meets the satellite horizontally.

The average gain of an isotropic aerial when immersed in the ionosphere (assumed to be flat and plain stratified) is given by the ratio of the solid angle of the reception cone of the combined aerial ionosphere system to the solid angle, 4π, of an isotropic system in free space.

Thus if θ'_0 is the semi-angle of the reception cone when the aerial is immersed in the ionosphere the average gain

$$G_{av} = \frac{2}{1-\cos \theta'_0}$$

but the refractive index $n_0 = \sqrt{(1-f_0^2/f^2)}$ where f_0 is the ambient plasma frequency and f is the operating frequency, whence we obtain $\cos \theta'_0 = f_0/f$ and

$$G_{av} = \frac{2}{1-f_0/f}$$

Typical values of f_0/f, θ'_0 and G_{av} are given in Table 1.

The analysis by Haselgrove, Haselgrove and Jennison also showed that the aerial directivity pattern is modulated by interference fringes resulting from the two possible ray paths to the aerial, one direct and one reflected in the ionosphere beneath the satellite. Furthermore, the aerial gain tends to increase towards the edge of the pattern and the final edge is very sharply bounded. Ray treatments show an infinity at the boundary but Budden has shown that this is resolved by a wave analysis and the open circuit aerial voltage

$$V = lK_1\mathrm{Ai}(\zeta_0)$$

for broadside radiation from a horizontal dipole and

$$V = lK_2\mathrm{Ai}'(\zeta_0)$$

for end-fire radiation from a horizontal dipole where K_1 and K_2 are slowly varying functions involving the refractive index, the gradient of refractive index and the angle θ, l is the length of the aerial, Ai is an Airy integral and Ai′ its vertical derivative.

If the aerial is at a large distance from the earth such that the electron density, and therefore the refractive index, is constant or only slowly varying with distance, the directivity pattern is no longer dictated by the conditions in the immediate vicinity of the satellite, apart from the primary directivity pattern of the aerial. In these circumstances the flat earth approximations no longer hold, the rays defining the edge of the reception cone no longer meet the satellite horizontally and the cone semi-angle can exceed 90 deg. The earth and its ionosphere still affect the directivity pattern as rays may still be reflected and refracted at the appropriate levels so that the pattern retains its fringe modulation and contains a sharply cut "hole" or shadow in the direction of the earth.

In Table 1 a list is set out in which the semi-angle of the reception cone and the aerial gain are given as a function of the refractive index for a flat earth and plain stratified ionosphere. In Table 2 maximum reception angles are quoted for two sets of data on electron densities at large distances from the earth when a fuller analysis has been used.

Table 1

$f_0/f = \cos\theta_0'$	$n = \sqrt{\left(1 - \frac{f_0^2}{f^2}\right)}$	$\theta_0' = \sin^{-1} n$	Average gain $= \frac{2}{1 - f_0/f}$
1	0	0	∞
0·98	0·2	11° 31′	100
0·9	0·4359	25° 50′	20
0·8	0·6000	36° 50′	10
0·7	0·7141	45° 35′	6·7
0·6	0·8000	53° 10′	5
0·5	0·8660	60° 00′	4
0·4	0·9165	66° 25′	3·3
0·3	0·9539	72° 30′	2·85
0·2	0·9798	78° 30′	2·5
0·1	0·9950	84° 20′	2·2
0·0	1·0000	90° 00′	2·0

Table 2

f(Mc/s)	2	3	4
θ_0' degrees at 5000 km	123·7 124·2	128·1 126·1	130·6† 127·4‡
θ_0' degrees at 10 000 km	141·7 144·0	144·7 145·0	146·5† 145·6‡

† Based on the distribution of electron density determined by Berning (1960).
‡ Based on the distribution of electron density determined by Al'pert *et al.* (1958).

In each case a background electron density of 10^2 has been used in extrapolation of the profiles. A background density of any value from zero to 10^3 produces negligible difference in the cone angles.

3. Practical Systems

The modification of the primary directivity pattern referred to in the previous section may be used in a number of practical systems to achieve a high resolving power at medium frequencies.

The principal properties of the combined aerial-ionosphere reception pattern are:

(i) The restriction of the radiation into a narrow cone when the refractive index approaches zero.

(ii) The modulation of the reception pattern by a system of interference fringes (approximately concentric circles in horizontal cross-section).

(iii) The very sharply bounded edge of the pattern, preceded by an enhancement due to focusing at the limb.

Property (i) would enable gains and beamwidths at least comparable to those of Yagi antennas to be obtained at medium waves, but the beamwidth is very susceptible to variation in electron density and would continually change throughout the orbit of a satellite. It would be possible, in principle, to monitor the ambient electron density or plasma frequency and hence predict the cone angle. A more ambitious proposal which could only be applied to radiation having an approximately constant spectrum, as may well be the case for cosmic radiation, is to control the frequency of the receiver by means of a servo system to keep it always operating at a frequency where the refractive index is a constant slightly in excess of zero. This system would preserve a constant beamwidth although the frequency of reception would be continuously varying. The frequency could be monitored and passed back, after reduction to a low data rate, by the telemetry system to give information on the electron density of the ionosphere along with the output of the receiver giving a crude sky survey on a further low data rate channel.

Property (ii), the modulating fringe system, is only of use for c.w. transmission or reception, though for bandwidths up to a maximum of about 1 kc/s the fringes near the edge of the reception cone may still be distinct. The zero-order fringe is situated at the edge of the reception cone and whilst higher orders, further towards the centre, would rapidly integrate out over finite bandwidths, the position of the zero-order fringe itself moves and hence also smooths out the fringes to give a constant output with less sharp edges as a result of the change in the semi-angle of the cone with change in frequency. The fringe system appears, therefore, to be of little use to radio astronomy although it may be that it will find uses in the field of communications or for investigating the ionosphere itself.

(iii) The very sharply defined edge of the reception cone presents several possibilities for practical techniques. The intrinsic resolving power of this edge is typically a few minutes of arc. This figure would hold in practice if the system were monochromatic and the ionosphere were perfectly smooth. The effect of a finite bandwidth is to round off the edge to an extent which is simply determined by the variation in semi-angle of the reception cone over the same frequency range. If the satellite carrying the elementary aerial is a few thousand kilometres from the earth and the operating frequency is about $2\frac{1}{2}$ to 3 Mc/s, the effect of a bandwidth of 100 kc/s is to reduce the resolving power to about 20 minutes of arc, whilst a bandwidth of 1 Mc/s would still give a resolving power of the order of three degrees.

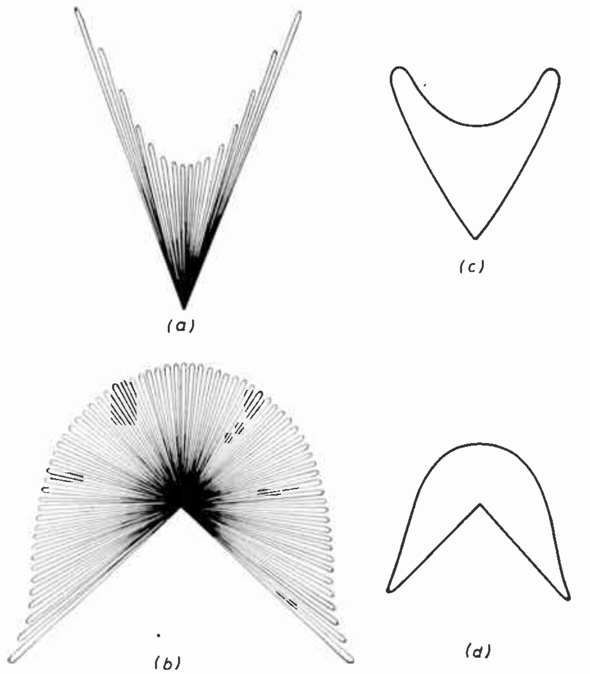


Fig. 2. (a) The directivity pattern corresponding to the conditions in Fig. 1 (a).
(b) The directivity pattern corresponding to the conditions in Fig. 1 (b).
(c) The effect of a finite bandwidth upon the directivity pattern of Fig. 2 (a).
(d) The effect of a finite bandwidth upon the directivity pattern of Fig. 2 (b).

The effect of ionospheric irregularities is not easy to assess as very little is known about the small-scale distribution of electron density above 500 km. It is probable that if the operating frequency is sufficiently high and the distance from the earth sufficiently great the effect of irregularities would not be troublesome. Even under more adverse conditions the possible effect of irregularities does not preclude the use of this principle in a satellite technique as the obtainable resolution would be very high and small sources, if such exist, might make their presence known by the production of scintillations in the received signal.

Two techniques are proposed which make use of the sharp edge of the reception pattern. In the first of these, the output of the receiver after detection is passed through a bandpass filter with a lower frequency limit of several minutes per cycle and an upper frequency limit adjusted to suit the fastest rate of change of signal that could be caused by radiation from a point source entering the edge of the reception cone as the satellite orbits. This frequency is a function of orbit and bandwidth and may lie between one cycle per second and one cycle per minute. The application of the bandpass filter removes the large background level of signal and the remaining higher Fourier components, corresponding to the edge of the reception pattern, may be further amplified prior to connection to a low data rate telemetry channel. The resolving power with this technique is determined by the bandwidth and the high pass limit of the filter whilst the large bandwidth and time constants reduce the background noise ripple and enable a high sensitivity to be obtained. The direct output of the detector, without differentiation or further amplification, may be monitored on a further low data rate channel.

The second proposal, making use of the sharp edge of the reception cone, is to commutate the receiver rapidly between two frequencies separated by a few tens or hundreds of kilocycles. The output of the detector is commutated in synchronism so that the reception in the two corresponding states is compared. The difference signal, corresponding for a flat spectrum to the signals from a narrow ring in the sky, is further amplified and passed to the telemetry as a low data rate signal.

The latter technique could be combined with that employing the band pass filter provided that a suitable switching rate were chosen, and both could independently feed separate telemetry inputs. Similarly the proposal making use of the narrow reception cone where the refractive index approaches zero, could be incorporated in the same satellite by employing a separate receiver on a lower frequency.

4. References

1. G. Reber and G. R. Ellis, "Cosmic radio frequency radiation near one megacycle", *J. Geophys. Res.*, **61**, p. 1, March 1956. G. Reber, "Solar activity cycle and spread F", *J. Geophys. Res.*, 63, p. 869, December 1958.
2. C. B. Haselgrove, J. Haselgrove and R. C. Jennison, "Ray paths from a cosmic radio source to a satellite in orbit", *Proc. Roy. Soc.*, A, **261**, p. 423, 1961.
3. C. B. Haselgrove, J. Haselgrove and R. C. Jennison, Corrigendum, *Proc. Roy. Soc.*, A. (In publication.)
4. K. G. Budden, "The edge focusing of a radio signal received from outside the earth by an aerial within the ionosphere", *Proc. Roy. Soc.*, A., **263**, 1961.

Manuscript received by the Institution on 3rd July 1961 (Paper No. 665).

U.D.C. 621.396.946: 629.19

A Proposal for an Active Communication Satellite System based on Inclined Elliptic Orbits

By

B. BUSS†

AND

J. R. MILLBURN,
B.Sc. (*Graduate*)†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July, 1961.

Summary: Satellites in the 24-hour stationary orbit will have to be stabilized in attitude relative to the earth, and kept in the correct station in their orbit, if their full potential is to be realized. If attitude stabilization and station keeping can be achieved, there are many advantages in using inclined elliptic orbits; this paper describes the design of a communication satellite for use in a 63-deg orbit with apogee 12 600 miles and perigee 300 miles, within a weight limitation of 550 lb. The optimum frequency for a commercial communications system, in which due allowance is made for adverse weather conditions, is shown to be about 2 kMc/s. The optimum distribution of weight in the satellite is obtained by considering the W/lb factors for various types of repeaters and solar cell arrays. It is shown that a design based on a semi-sun-seeking concept, in which rotation about the earth-satellite axis is used to maximize the output from the solar cell array, could have a radiated power of 12 W and a capacity of up to 1000 telephone channels.

1. Introduction

An optimum communication system may be defined as one which provides the required service with maximum reliability at minimum overall cost. The difficulty of achieving this ideal with communication satellites is illustrated by the variety of the proposals submitted to the Governments of this country and the U.S.A. during 1960–61.

In this paper one particular aspect of the problem is considered: the design of the satellite itself (and associated ground equipment where relevant) to make best use of a given orbit. The orbit chosen by way of example has the following characteristics:

Apogee	12 600 miles
Perigee	300 miles
Inclination of orbit plane to equator	63°
Latitude of apogee	63°N
Period	6 hours

The properties of elliptic orbits have been presented in detail elsewhere,[1] but an outline is given in Sections 2 and 3 so that the overall picture can be seen. Factors affecting the choice of operating frequency are discussed in Section 4, and various types of radio repeater are examined in Section 5. Mechanical design is considered in Section 6. The integration of the radio equipment, attitude controls, and power supplies into a complete design is described in Section 7, and the performance to be expected from this design is derived in Section 8.

† Hawker Siddeley Aviation Ltd., Astronautics Group, Welkin House, London, E.C.1.

Throughout the paper the necessity of providing a reliable public service has been kept to the fore. Thus, proved techniques and components have been used wherever possible, and the performance figures quoted are those which should be achieved under the most unfavourable conditions.

2. Orbits

2.1. *Coverage Required*

The basic difference between a telephone cable and a communication satellite is that the former provides a permanent link between the limited number of places actually on the cable route, while the latter is potentially capable of providing links between *all* the places which can see the satellite simultaneously. Communication satellites make the ultimate goal of 24-hour world-wide coverage a real possibility.

This does not necessarily mean that the coverage should be uniform.[2] Figure 1, which is based on data for 1956, shows that most of the world's population and telephones are in the Northern Hemisphere. It is also clear from the figure that although there is a large undeveloped market between the equator and 40°S, the demand for service outside the range 60°N to 40°S is not likely to be very great, and is virtually zero at the poles.

2.2. *Circular Orbits*

Most of the existing proposals for communication satellite systems are based on either polar circular orbits at a height of 2–5000 miles,[3] or the equatorial circular "stationary" orbit at 22 300 miles[4, 5]; in

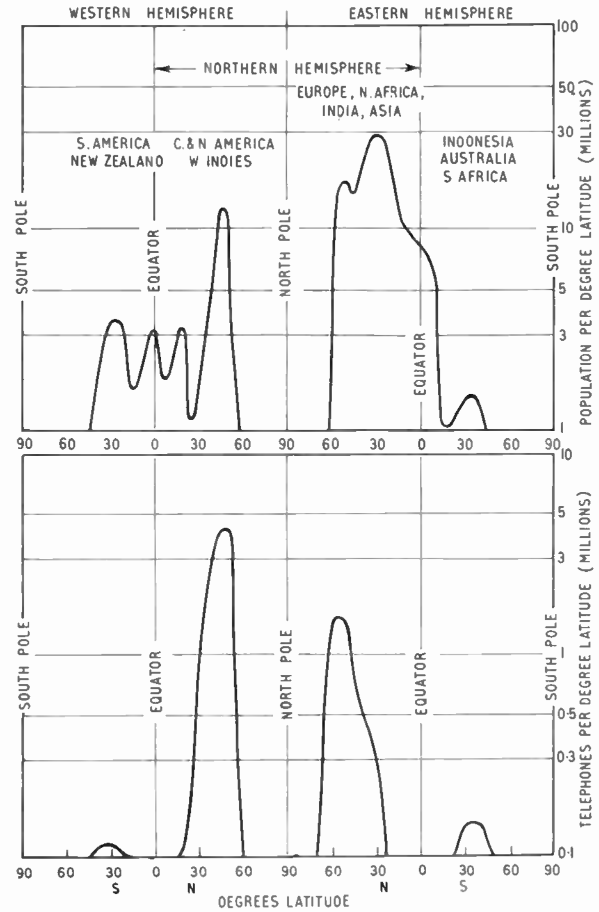


Fig. 1. Distribution of World population and telephones. (Excluding U.S.S.R. and China)

fact, it is sometimes stated that these are the only types of orbit worth consideration. In view of the lack of demand for a service at the poles, systems based on low polar circles are generally inefficient, and require large numbers of satellites. The equatorial stationary orbit, on the other hand, offers the possibility of a good service in the required regions with only three satellites.

There is no doubt that this orbit will eventually find many applications in future communication systems, but for immediate use it has a number of disadvantages:

(a) The payload which can be placed in this orbit by currently available boosters is very small.

(b) Establishment of an equatorial orbit, in the absence of an equatorial launching site, is difficult.

(c) Attitude stabilization and station keeping are essential if the "stationary" feature of the orbit is to be realized and fully exploited.

(d) The round-trip delay of 0·6 seconds may place impossible demands on the performance of the echo suppressors in the system, though opinions differ on this point.

2.3. *Elliptic Orbits*

It is evident that (a) and (d) would be overcome if a satellite could be placed in an equatorial circular orbit with a period of 24 hours at a height of 5–10 000 miles. This, of course, would require a continuous thrust, far larger than could be obtained by extracting power from the medium surrounding the satellite. An approximation to a stationary satellite at these heights can be obtained, however, by making use of several satellites in highly elliptic orbits with a period of less than 24 hours. At apogee, a satellite moving in an orbit of this type moves slowly relative to the earth, and may even appear to oscillate about the zenith when seen from a point on the equator directly below it. Furthermore, if the orbit plane is inclined to the equator, the apogee can be placed over the Northern Hemisphere, where the immediate demand for more telephonic communications is greatest. An inclined orbit also overcomes objection (b).

It is shown in reference 1 that the perturbations of elliptic orbits caused by the earth's equatorial bulge can be used to keep the apogee in the required region. In particular, if the inclination of the orbit plane to the equatorial plane is 63°, the latitude of apogee remains constant. Studies of common viewing times from New York and London have shown that a single satellite in a 6-hour period elliptic orbit at 63 deg to the equator, with an apogee height of 12 600 miles, could provide a service between these points for about nine hours per day. Four satellites suitably phased could give almost continuous coverage to all places above 40°N, and a usable, though intermittent, service to places as far south as Australasia.

Provided that attitude stabilization and station keeping can be achieved, the use of three or four satellites in inclined elliptic orbits instead of the equatorial circular stationary orbit enables a good service to be given to the northern hemisphere, with the advantages of reduced range, increased payload, and use of existing launching sites.

3. Satellite Weight

3.1. *Booster Performance*

The development of a launching rocket and its associated ground support system is an expensive and lengthy business, and it seems almost certain that any commercial applications of space will have to rely on boosters which are either already in existence or being developed for other purposes. As the weight of the payload which can be placed in a given orbit is very dependent on booster performance, it is necessary to decide which booster is to be used before the satellite can be designed.

For the purpose of the study reported in this paper, a booster based on the predicted performance of a *Blue Streak/Black Knight*/third stage combination was assumed, but the principles of the design procedure given in Sections 4–8 apply to any booster of similar size.

The variation of payload with apogee height for the selected booster is given in Fig. 2. At 12 600 miles, corresponding to four orbits per day, the payload is about 550 lb.

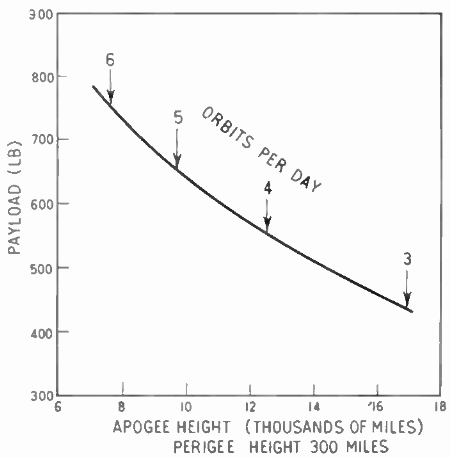


Fig. 2. Variation of payload with apogee height for *Blue Streak/Black Knight*/5000-lb 3rd stage in 63-deg orbit, launched from Woomera.

3.2. *Weight Distribution*

The optimum distribution of weight in the satellite, for maximum information handling capacity in the communication system, is a basic design problem which does not appear to have received much attention in the literature of space communications.

It has been shown in a previous paper[6] that with a satellite weight of a few hundred pounds and an altitude of several thousand miles a marked advantage can be gained by the use of attitude stabilization, even though the equipment necessary to achieve this may require 20% of the payload weight. The total weight of the satellite may therefore be broken down in to the following main divisions:

(a) Radio repeater equipment.
(b) Solar cell arrays and supporting structure.
(c) Batteries and power distribution system.
(d) Telemetry and command system.
(e) Attitude controls and station keeping equipment.
(f) Basic structure, including thermal controls.

For a given orbit and satellite lifetime, (e) is almost constant and (c) bears a fixed relation to (a). Telemetry and command (d) should be small compared with (a) in an operational system and may be disregarded in a preliminary analysis. The basic structure (f) should also be small compared with (a), and is determined largely by the third stage geometry. The major variables are therefore (a) and (b).

Maximum radiated power will be obtained when the weight of the satellite is so distributed that the power output of the solar cells is just matched to the power input required by the radio equipment. In order to achieve this, the magnitudes of the watts/pound factors for various types of radio repeaters and solar cell arrays must be determined.

4. Choice of Carrier Frequency

4.1. *The Radio "Window"*

Proposals for allocation of frequencies for various earth/space purposes were submitted to the C.C.I.R. by the I.A.F. in 1959,[7] and were considered at the IXth Plenary Assembly. It was decided that allocation of wide bands for communication satellite systems would be premature; consequently, until the situation has been clarified by the 1963 conference it is necessary for individual designers to consider all the factors involved and then choose what appears to be the optimum frequency.

The upper and lower limits of the frequency range which can be used for earth–space–earth communications are set by the atmosphere and ionosphere respectively.[8] There is no sharp cut-off at the high frequency end, but atmospheric losses become appreciable at 5–10 kMc/s and reach prohibitive values at 25–50 kMc/s; at the other end of the range the minimum usable frequency varies widely with location, angle of elevation, and sunspot activity, but for reliable day-to-day communications it is unlikely to be lower than 100 Mc/s. To determine the optimum within this range for a particular system it is necessary to take into account the aerials on the satellite and on the ground, satellite transmitter efficiency, receiver noise, atmospheric and galactic noise, and absorption by rain.

4.2. *Satellite Aerials*

Attitude stabilization implies the ability to keep the satellite pointing in the required direction, and hence makes possible the use of aerials with some gain. Eventually, the development of larger boosters will make it possible for the satellite to carry sufficient equipment to keep high-gain aerials pointing towards individual ground stations. For the time being, however, it will be assumed that the aerials are fixed to the body of the satellite, which is directed towards the earth by means of reaction jets controlled by infra-red sensors. Suitable sensors have already been produced in the U.S.A., and improved versions without rotating parts are under development.[9]

The maximum gain which can be used with this system is limited by the requirement to cover the whole of the earth seen from the satellite. In the case of circular orbits this is a constant; for the 24-hour orbit, for instance, it is about 18 dB. With elliptic orbits, the angle subtended by the earth changes as the satellite travels round its orbit, and for the proposed 6-hour orbit the angle varies between 28 deg and 140 deg (Fig. 3). Although at first sight this appears to be a serious disadvantage, in practice the satellite would not be used near perigee as it would be too low to be visible simultaneously from stations more than a few hundred miles apart. Studies of common viewing time from New York and London have shown that in general the satellite would not be used at altitudes less than 4000 miles, at which height the angle subtended by the earth is 60 deg. The simplest approach would therefore be to use an aerial with a -3 dB beamwidth of 60 deg, corresponding to a maximum gain of about 10 dB.

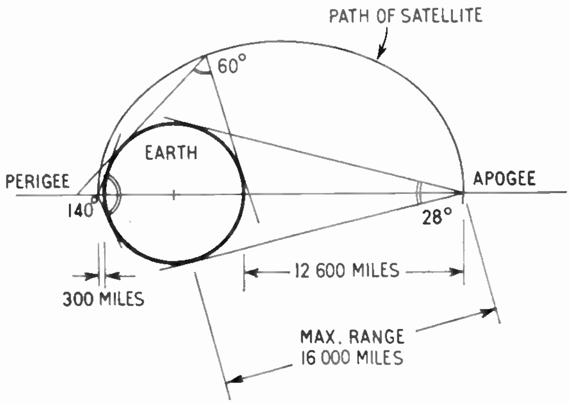


Fig. 3. Geometry of elliptic orbit with 300 mile perigee and 12 600 mile apogee (period 6 hours).

However, it will be seen from Fig. 3 that the maximum slant range to the satellite at an altitude of 4000 miles is about 7000 miles, which is only 0·44 of the range at apogee; the signal is consequently five times greater. In fact, a satellite transmitter aerial radiation pattern can be specified such that the power received on the ground at maximum range remains constant as the satellite travels round its orbit. The required pattern is shown on a linear intensity scale in Fig. 4. When the satellite is at apogee, the distance to the part of the earth immediately below it is appreciably less than the maximum slant range, which accounts for the dip in the pattern between ± 14 deg.

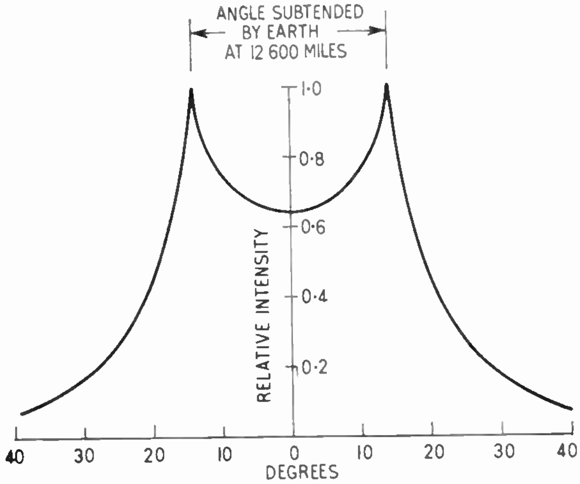


Fig. 4. Radiation pattern for constant signal intensity at maximum range.

The shape of the pattern outside ± 14 deg is not very different from that obtained with conventional microwave aerials. Figure 5 shows the required pattern superimposed on the theoretical pattern from a uniformly illuminated rectangular aperture with sides 1·67 wavelengths long, which has a half-power beamwidth of 30 deg and maximum gain of 15·3 dB. It will be seen that the required intensity is exceeded out to angles of ± 23 deg, and although there is a 3 dB discrepancy at ± 30 deg there is ample power in hand between 0 deg and ± 15 deg to allow the aerial designer to compensate for this. For the purpose of calculating received signal power the aerial gain has therefore been taken as 12 dB at the maximum range of 16 000 miles; this figure applies to both the transmitting and receiving aerials on the satellite and is independent of frequency, provided that the aerial dimensions are compatible with those of the satellite as a whole.

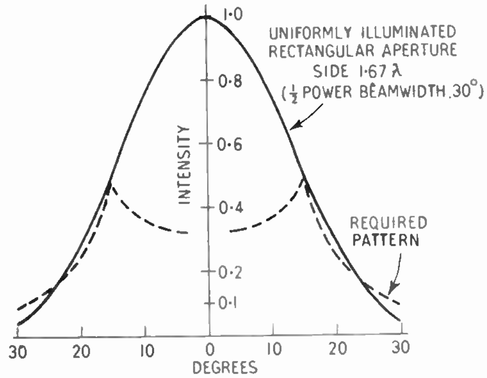


Fig. 5. Comparison of required pattern with that produced by uniform illumination of a rectangular aperture.

4.3. *Ground Aerials*

The power P_{RG} entering the ground receiver, in the absence of fading is given by

$$P_{RG} = \frac{P_{TS} G_{TS} A_{RG}}{4\pi R^2} \qquad \ldots\ldots(1)$$

where P_{TS} = satellite transmitter power,
G_{TS} = satellite transmitter aerial gain,
A_{RG} = ground receiver aerial effective aperture,
R = distance between satellite and ground receiver.

The effect of using the aerial pattern given in Fig. 4 is to make G_{TS}/R^2 constant. If P_{TS} is constant, it follows that P_{RG} is proportional to A_{RG} and is independent of frequency. A_{RG} should therefore be as large as possible. The tolerance on the aerial dimensions, which must be adhered to if its effective area is to be comparable with its physical area, is normally about $\pm\lambda/16$ and is independent of size, but clearly becomes more difficult to meet as the size and/or the frequency are increased. Experience with large radar aerials suggests that the maximum diameter of parabolic reflector which can be used at 10 kMc/s is about 30 ft. If the frequency is reduced to 2 kMc/s the usable diameter goes up to 100 ft, corresponding to an increase in P_{RG} of 10 dB. Although in theory there is a further increase in usable diameter as the frequency is reduced, 100 ft is probably near the limit for commercial use, and it seems reasonable to place an upper limit of about 2 kMc/s on the frequency of the satellite-to-ground link.

The figures quoted above take into account wind loading and temperature distortion as well as static deflections. A communication system for public use cannot afford to lose operating time through adverse weather conditions.

For the ground-to-satellite link similar considerations apply, the power P_{RS} entering the satellite receiver being given by

$$P_{RS} = \frac{P_{TG} G_{TG} A_{RS}}{4\pi R^2} \qquad \ldots\ldots(2)$$

where T, R, G, S indicate transmitter, receiver, ground and satellite respectively.

Since $G_{TG} = \dfrac{4\pi A_{TG}}{\lambda^2}$, and $A_{RS} = \dfrac{G_{RS}\lambda^2}{4\pi}$

eqn. (2) may be written in the form

$$P_{RS} = \frac{P_{TG} G_{RS} A_{TG}}{4\pi R^2}$$

As before, G_{RS}/R^2 is constant, so that P_{RS} is proportional to $P_{TG}A_{TG}$ and is independent of frequency. In this case, however, P_{TG} is not subject to the restrictions of weight and available power which limit the transmitter power on the satellite, so it is not essential that A_{TG} should be as large as possible. On grounds of cost, there is some advantage to be gained by reducing the aerial diameter to, say, 30 ft instead of 100 ft. The frequency of the ground-to-satellite link is then not restricted by the aerial problem, and can be chosen to suit the requirements discussed below.

4.4. *Noise*

An earth–space–earth communication system differs from point-to-point microwave links on the ground in two important respects:

(a) in general, there is only one repeater in the system even when the ground stations are many thousands of miles apart;

(b) the repeater is completely inaccessible once it has been placed in orbit.

Clearly, the system should be designed in such a way that the repeater is as simple and reliable as possible. Fortunately, it follows from eqns. (1) and (2) that the most critical part of the system from the point of view of signal/noise ratio is the ground receiver. Since the ground stations will have to incorporate expensive aerial systems and control gear, the provision of low noise maser receivers and refrigeration plant will have little effect on the overall cost of the communication system. This would not be true in the case of a multi-hop point-to-point link, where the cost of the electronic equipment at each station is a much larger fraction of the whole.

Assuming that the ground receivers incorporate masers, the noise level at the ground receiver input will be determined largely by external factors, especially cosmic noise and atmospheric absorption noise. Curve A of Fig. 6 shows that these factors set an absolute lower limit to the effective aerial temperature of about 2·5°K, but this figure can only be achieved if the aerial has infinite directivity and is pointing towards a source of minimum cosmic noise at 90 deg elevation. For the purpose of designing a reliable communication system, it is necessary to take into account the worst conditions which are likely to be encountered (provided that they do not occur for a negligibly small fraction of the operating time). It is generally accepted that the minimum angle of elevation which should be used is about 7½ deg, as below this the noise contributed by the radiation from the earth into the main sidelobes of the receiving aerial becomes excessive. Taking D. C. Hogg's calculated values for atmospheric noise at this elevation,[10] together with the maximum cosmic noise (excluding the sun), the minimum noise temperature is shown by curve B to be about 20°K at 3–5 kMc/s.

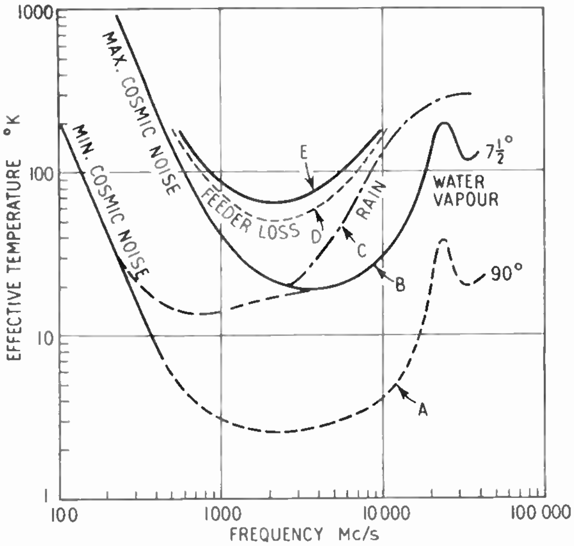


Fig. 6. Factors affecting receiver aerial noise temperature.

This figure will be increased in practice by the effect of rain, feeder losses, and the receiver noise factor. The absorption of microwaves by rain and other forms of precipitation was considered in detail by J. W. Ryde,[11] who showed that the path attenuation in dB/km is nearly proportional to the precipitation rate. Figure 7 shows the effective aerial temperature due to absorption by rain at 300°K, assuming no contribution from other sources, using data from Fig. 7 of Ryde's paper. In this country, rain of a few mm/hr may persist for many hours, but the path length in rain of this type at the minimum angle of $7\frac{1}{2}$ deg is unlikely to exceed 5–10 km (corresponding to a cloud base of about 1000 ft). Curve A shows that at frequencies below 10 kMc/s the effect is negligible. Moderately heavy rain of 10–20 mm/hr over a similar path length does occur for an appreciable time, however, and curve B shows that this must be taken into account at frequencies greater than 4 kMc/s. Curve C corresponds to conditions never encountered in this country, but illustrates the drastic effect which tropical storms could have on aerial temperature for short periods.

Referring back to Fig. 6, curve C shows the effect of rain at 100 km mm/hr on curve B, using the relation

$$T_e = \alpha T_0 + (1-\alpha)300 \qquad \ldots\ldots(3)$$

where T_e = total effective temperature,

T_0 = temperature in the absence of rain,

$1-\alpha$ = fraction of signal lost by absorption in the rain.

The disadvantage of using frequencies above 3 kMc/s is evident.

So far it has been assumed that all the power intercepted by the aerial is fed into the receiver. Even if the maser could be placed at the focus of the aerial, however, there would be some losses due to imperfect matching, and in practice the problems of routine maintenance and all-weather working may make it desirable to mount the maser in a more accessible position. In this case, it may be difficult to avoid a loss of less than 0·5 dB in the feeders, which, being at 300°K, adds a further 30° to the effective aerial temperature, resulting in curve D of Fig. 6. If the maser itself has a noise factor of 0·3 dB, the final effective aerial temperature is that shown in curve E.

From this it can be seen that the frequency which gives least noise under the worst conditions is about 2 kMc/s.

4.5. *Satellite Transmitter Efficiency*

The signal/noise ratio at the input to the ground receiver will not necessarily be greatest when the operating frequency is chosen for minimum noise. If the satellite transmitter efficiency increases as the frequency is reduced, there may be some advantage in using a frequency lower than 2 kMc/s.

It happens that 0·5–2 kMc/s is a frequency range in which the efficiency of many disc-seal triodes changes rapidly. Specific examples taken from the manufacturer's literature, however, are not really significant, as it is unlikely that any existing type of valve would be used in a satellite without modification. All that can be said with certainty is that the deterioration of signal/noise ratio as the frequency is reduced need not be so large as Fig. 6 suggests.

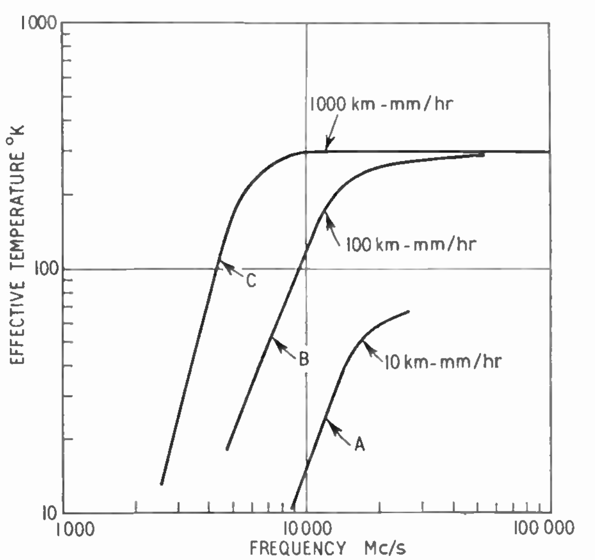


Fig. 7. Effective aerial temperature due to absorption by rain at 300°K.

Even if the effect of transmitter efficiency is disregarded, it is evident that the "optimum" frequency range is a very broad one, reaching from about 700 Mc/s to 7 kMc/s for a 3 dB change in signal/noise ratio. In practice it seems likely that the operating frequency will be determined by the problem of interference to or from existing services.

4.6. *Interference*

Denoting transmitter and receiver by suffixes T and R, and the satellite, ground station, and existing service by S, G, E respectively, the four cases to be considered are

(1) S_T to E_R

(2) G_T to E_R

(3) E_T to S_R

(4) E_T to G_R

In the frequency range 700 Mc/s–7 kMc/s, the principal wideband services are, in order of increasing frequency, television broadcasting, tropospheric scatter links, surveillance radars, and point-to-point links. As all these services use different bandwidths, types of modulation, and aerials, a full treatment of the interference problem is too complex to be considered here, but a few specific examples are given below.

In a typical communication satellite system S_T and G_T will be about 10 watts and 10 kW respectively. It can be seen at once that case (4) is likely to be troublesome if the existing service is anything but a point-to-point link. Suppose E_T is a Band V television station with an e.r.p. of 100 kW; this is 25–30 dB larger than the e.r.p. of S_T, so that even if G_R is not pointing towards E_T the signal picked up by the minor sidelobes of the G_R aerial will be very much larger than that due to S_T in the main beam. With S_T at 10 000 miles, and E_T at 100 miles (by no means the worst case), the signal from E_T in the sidelobes will probably be 40 dB above that from S_T in the main beam. Although E_T will presumably be amplitude modulated it is unreasonable to expect the ground station receiver to cope with an interfering signal of this magnitude. This seems to rule out the use of frequencies below 900 Mc/s, unless a clear channel can be provided.

It is more profitable to consider interference between the satellite system and a microwave point-to-point link. Here, the radiated power of E_T is of the same order as S_T (about 10 W) and, except for mobile links, the aerials associated with E_T and E_R have narrow beams which are fixed in direction. Cases (1) and (3) are clearly not important, but both (2) and (4) have to be considered.

If G_T is in the main beam of E_R and just within line-of-sight range (say 50 miles maximum), E_R will receive almost equal powers from E_T and from the minor sidelobes of G_T. If G_T points directly at E_R it will, of course, completely override E_T, but the 7½ deg horizon should eliminate this possibility. If G_T is more than 10 deg off the E_T–E_R line, the difference in power levels seen by E_R should be at least 30 dB which should be adequate for suppression of adjacent channel interference if both the satellite system and the point-to-point link use f.m. A small amount of care in siting G_T should therefore dispose of case (2), provided that the point-to-point link aerials have no subsidiary lobes comparable with their main lobe.

Case (4) is more difficult to satisfy, for E_T is bound to have a big advantage over S_T in range. Assuming that G_R is more than 10 deg off the E_R–E_T line, and about 50 miles from E_T, so that the only link between them is via minor sidelobes at each end, G_R will see in effect ($E_T - 60$ dB) at 50 miles and S_T at between 4000 and 16 000 miles. In the worst case, the signal from S_T will be just comparable with that from E_T. In order to obtain an adequate differential G_R will therefore have to be sited beyond line-of-sight range of E_T.

The theoretical attenuation due to diffraction round a smooth spherical earth beyond the radio horizon is about 2 dB/mile at microwave link frequencies.[12] Thus a distance of 10–20 miles beyond the horizon from E_T should be adequate under normal conditions. Although abnormal propagation conditions will reduce the attenuation considerably, in many cases it will be possible to make use of the additional attenuation provided by natural obstacles. While there is no doubt that an exclusive allocation of frequencies for communication satellite systems is highly desirable, it does seem that frequency sharing with a microwave point-to-point link may be just possible if considerable care is taken over the siting of the ground stations.

4.7. *Summary of Factors affecting Choice of Frequency*

A frequency of about 2 kMc/s for the satellite/ground link appears to be near the optimum from all points of view.

For the ground/satellite link there may be some advantage initially in using uncommitted channels in Bands IV or V, as has been specified for the N.A.S.A. experimental satellites[13]; this enables relatively inexpensive aerials and transmitters to be used. For a complete system, however, it seems inevitable that lack of bandwidth will force the ground/satellite link into the microwave region. It was suggested in Section 4.3 that the transmitter aerial need not be more than 30 ft in diameter (because the power of G_T can be at least 30 dB greater than S_T) and if this size of aerial is used the frequency can be raised to about 7 kMc/s. For the purpose of the present study it is therefore assumed that the ground/satellite link will operate at a frequency in the 6 kMc/s microwave link band.

5. Satellite Radio Equipment

5.1. *Types of Repeater*

Assuming that the design of the satellite equipment should be based on well-tried techniques and components, there are three basic types of repeater to be considered.[14]

(a) receiver and transmitter back-to-back,

(b) non-demodulating repeater with i.f. amplification,

(c) non-demodulating repeater with all the amplification carried out at radio frequencies.

Type (a) has found little application in point-to-point links, except those of a temporary nature, because of the distortion introduced by the de-modulation and re-modulation processes. It may be useful for earth–space–earth communications in those cases where it is necessary to change the nature of the modulation in the satellite, in order, for example to conserve bandwidth on the ground/satellite link by using a.m. while power is exchanged for bandwidth on the satellite/ground link by using f.m. Communication satellite schemes of this type have been proposed,[15] but will not be considered further in this paper owing to lack of operating experience.

Type (b) has the advantage that the i.f. amplifier can use transistors throughout, but against this must be set the disadvantage that the bandwidth is limited to a few tens of megacycles by the performance of currently-available v.h.f. transistors. If the repeater is required to have a bandwidth of hundreds or even thousands of megacycles, as has been suggested for some communication satellite systems,[3] type (c) is preferable, as the satellite repeater can then consist of a number of travelling wave tubes in series. This arrangement is rather inefficient, however, as even the tubes operating at low signal levels will consume several watts. The efficiency could be improved by using one travelling wave tube in a multiple reflex circuit, but this system has yet to be proved in practice.

Two important points which must not be overlooked when considering the type of repeater to be used are the ability to work with more than one pair of ground stations, and the reliability of the system. Although a single wide-band repeater appears to provide the most flexible system for multiple ground station operation, the presence of many signals of different amplitudes and frequencies places severe restrictions on the non-linearity which can be tolerated. Also, varying Doppler shifts on the signals from different ground stations result in inefficient use of the repeater bandwidth. To meet the reliability requirement, if a wide-band repeater is used it will be necessary at least to duplicate most of the equipment. If the radio equipment is divided into, say, a dozen quite separate type (b) repeaters, not only will the reliability be improved but also the problems of linearity and Doppler shift will be reduced. Against this must be set some loss of flexibility in operation, though it will still be possible to work six pairs of ground stations simultaneously, if desired, without frequency sharing.

The relative importance of these considerations is largely a matter of opinion, but it seems that the multiple-repeater system is likely to be preferable to the wide-band system for "first generation" satellites.

5.2. *The Non-demodulating I.F. Repeater*

A basic block diagram of this type of repeater is shown in Fig. 8. As the main reason for the use of a v.h.f. intermediate frequency is that it enables most stages in the repeater to use transistors, there is no point in using an r.f. preamplifier. The limited output

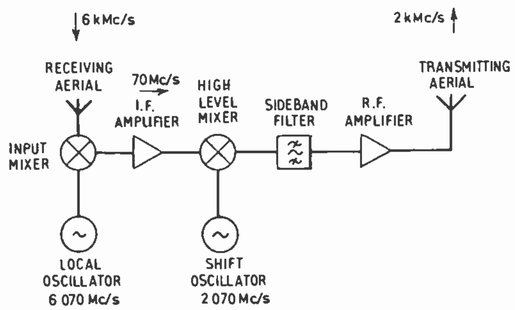


Fig. 8. Non-demodulating i.f. repeater.

which can be obtained from the high level mixer, however, makes an r.f. output stage necessary; this may be either a triode or a travelling-wave tube. Together with the two oscillators, which at the suggested frequencies will probably be a klystron and a triode respectively, the repeater will therefore incorporate three thermionic valves.

The choice of a triode or a travelling-wave tube for the output stage depends on the relative importance of gain, efficiency, weight and life. The best currently-available triodes are capable of giving an output of 1 W and gain of 20 dB at 2 kMc/s with a probable life of 10 000 hours, and the total input required to obtain this is 10–14 W. For the same power input, a

1-W travelling wave tube can have at least 10 dB more gain, at the expense perhaps of some increase in weight. A gain of 20 dB implies that the high level mixer has to produce an output of 10–20 mW (allowing for some loss in the sideband filter) in order to make full use of the d.c. power supplied to the r.f. stage, if this is a triode. It is doubtful whether conventional crystal mixers can reliably produce more than 2–3 mW, so it seems that if the extra weight is not excessive there is some advantage in using a travelling wave tube. If higher output powers are required, the advantage is much greater, as the gain of a t.w.t. can be increased as its power output is increased, so that the r.f. input can remain constant. There does not appear to be any prospect of increasing the gain of triodes in this way, so for output powers greater than about $\frac{1}{4}$ W it follows that two or more triodes in series would have to be used, resulting in a drastic reduction in efficiency.

Most current types of t.w.t.s are provided with heavy magnets and coupling devices, and a typical 1-W amplifier may weigh 10 lb. Such weights are not essential, however, for it has recently been shown[15] that if light weight is of primary importance, a $2\frac{1}{2}$-W 2 kMc/s t.w.t. can be made within a weight limitation of 8 ounces (226 grammes). When this is compared with the weight of the resonant circuits associated with two triodes, it will be seen that even if the above figure is thought to be rather optimistic the t.w.t. is a more attractive proposition than the triode.

5.3. *Bandwidth and Radiated Power*

The limitations on available power, imposed by the weight of the satellite, mean that a modulation system which enables power to be exchanged for bandwidth will have to be used. The maximum usable bandwidth is then set by the threshold signal/noise ratio. It is shown in Section 8 that for the proposed orbit, aerial gains, receiver noise factor, etc., the maximum usable bandwidth will be of the order of 10 Mc/s per watt of radiated power from the satellite. Thus, i.f.-type repeaters may be expected to have a radiated power of 1–5 W; similarly, wide-band repeaters should have a radiated power of 10–50 W.

5.4. *Efficiency and Weight of the Unit Repeater*

A block diagram of the proposed satellite radio equipment, based on the above conclusions, is shown in Fig. 9.

The "unit repeater" is self-contained, and is electrically connected to the rest of the satellite equipment at only 3 points (excluding telemetry connections): the 24 V d.c. power supply, the r.f. input at 6 kMc/s, and the r.f. output at 2 kMc/s. The number of these unit repeaters and their power output, is determined by consideration of their efficiency and weight.

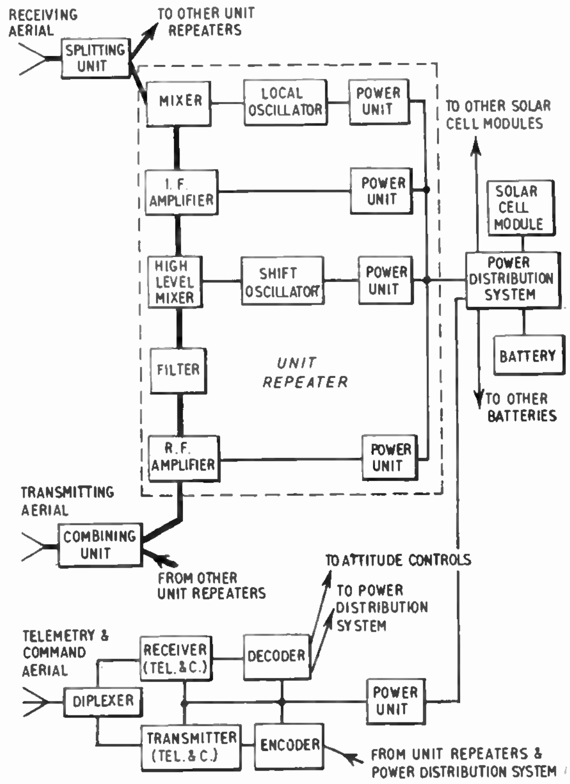


Fig. 9. Block diagram of satellite radio equipment.

For this purpose, it is convenient to divide the complete radio equipment into three parts:

(1) weight and input power fixed, comprising the aerials and feeders;

(2) weight and input power independent of radiated power, but proportional to number of unit repeaters, comprising the two oscillators, mixers, i.f. amplifier, and sideband filter;

(3) weight and input power proportional to both radiated power and to number of unit repeaters, comprising the r.f. amplifier and battery.

An estimate of the input powers and weights of parts (1) and (2), and of part (3) for radiated powers of 1 and 5 W, is given in Table 1. These figures are, of course, very dependent on the assumptions made about valve efficiencies, materials, and type of construction, but they are thought to be representative of what can be achieved today.

The powers shown against "battery" assume that a maximum of one-ninth of any orbit is in the earth's shadow.

The relation between input power and weight for repeaters with 1 W and 5 W outputs is shown in Fig. 10.

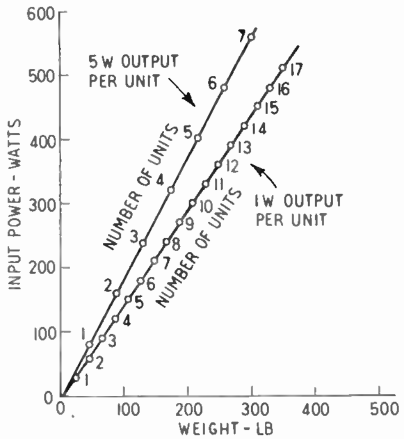


Fig. 10. Input power/weight relationships for 1-watt and 5-watt repeaters.

Table 1

Estimated Input Power Requirements and Weights of Satellite Radio Equipment

Part	Item	Input (watts)	Weight (pounds)
1	Aerials and Feeders	—	4
2	Splitting and Combining Units	—	$\frac{1}{2}$
	L.O. and P.U.	5	3
	Mixer	—	$\frac{1}{2}$
	I.F. and P.U.	2	$1\frac{1}{2}$
	H.L. Mixer and Filter	—	1
	S.O. and P.U.	5	3
	Total	12	$9\frac{1}{2}$
3(a) (1 watt)	R.F. Amplifier and P.U.	15	8
	Battery	3	3
	Total	18	11
3(b) (5 watt)	R.F. Amplifier and P.U.	60	25
	Battery	8	$7\frac{1}{2}$
	Total	68	$32\frac{1}{2}$

6. Mechanical Design Considerations

6.1. *Power Supplies*

Electrical power for the satellite radio equipment can be obtained either from a source carried on the vehicle or by intercepting solar radiation. The only method considered to be feasible at present is conversion of solar radiation into electrical energy by means of silicon voltaic cells, and the watts/pound factors for various types of solar cell arrays are discussed below. It should be borne in mind, however, that when alternative sources become available (e.g. nuclear thermionic convertors, or solar boilers) the optimum distribution of weight in the satellite may be quite different.

6.1.1. Characteristics of solar cells

The reliability of solar cells for satellites in low orbits has been demonstrated by their use on, for example, *Vanguard I*, which has now been transmitting for over 3 years. Cells with efficiencies in the range 8–14% are now available commercially.

The efficiency of solar cells in space may be reduced by collisions with micrometeorites or high-energy particles. Protection against damage by the high-energy protons in the inner Van Allen belt is impracticable, but a glass cover 0·07 in. thick is sufficient to reduce the chance of damage by micrometeorites or the electrons in the outer Van Allen belt to a very small value.[16] The glass also increases the emittance in the infra-red region, and thus helps to reduce heating of the cells, though it may be necessary to include an ultra-violet filter as well. As the cells are made from single crystals of silicon with little inherent mechanical strength, they will be cemented to a light alloy honeycomb backing structure weighing about 0·7 lb/ft^2. The total weight of the backing, cells, and glass covers is then about 2 lb/ft^2.

The power output from the solar cells depends on the intensity of the illumination, and also on the load connected to the cells. The optimum load (for maximum power) depends on the effective intensity of the illumination; this means that if the cells are mounted on a curved surface, the nominal efficiency of the cells will only be realized if each cell is connected to a different load. The configuration of an array, therefore, has a considerable effect on the watts/pound factor.

6.1.2. Solar cell arrays

If the cells are mounted on a flat plate which is always maintained at right angles to the direction of the sun, the full efficiency of the array can be obtained (if the load is the optimum value). Taking the solar constant in free space near the earth to be 0·14 W/cm^2, the power available from this "sun-seeking" array is 7·5 W/lb for cells of 12% efficiency.

In practice a fully sun-seeking array cannot be used in conjunction with earth-seeking radio aerials, unless there is a flexible link between these two parts of the satellite. Even so, the stabilization problems are formidable, in view of the mutual reactions of the two parts of the satellite on each other.

If, instead of a stabilized flat plate, the array takes the form of a sphere, the power output for cells of constant efficiency will be reduced by a factor of four.

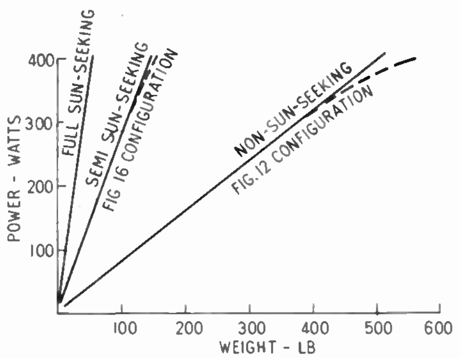


Fig. 11. Power/weight relationships for various types of solar cell array.

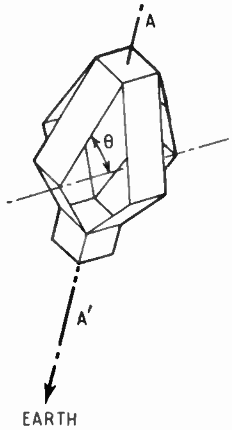


Fig. 12. Non-sun-seeking type of array for a 63-deg orbit.

But, as mentioned above, this can only be achieved if the cells are all working into different loads. If the load is fixed, the power output will be reduced by a further factor; alternatively, the cells can be switched so that only those receiving, say, not less than 90% of the maximum illumination are used at any given moment. In the latter case, the effective projected area will be only 1/20th of the total surface area, and the power output will be 0·37 W/lb.

An approximation to the isotropic form, which is more convenient mechanically, is the eight-sided array shown in Fig. 12.[17, 18] The angle θ can be selected to give maximum available power in a given orbit, and it has been shown that for the 63 deg orbit θ should be 60 deg. This type of array produces 0·8 W/lb; the output from only one panel is used at any given moment.

The watts/pound factors for fully sun-seeking and eight-sided arrays are plotted in Fig. 11. As the arrays increase in size, there will be a tendency for the lines to curve as shown, due to the necessity for more complex structures, but the extent of this curvature should be small for the size of the arrays under consideration.

6.1.3. The "semi-sun-seeking" array

The weight of the eight-sided array can be reduced by a factor of four (for a given power output) if advantage is taken of the fact that rotation about the axis A—A′ does not affect the earth stabilization system. Thus, if six of the panels are removed and the array is rotated about A—A′ so that the remaining panels face in the general direction of the sun, the power output can be increased to 2·8 W/lb. If the equipment canister is moved to a central position, the array takes the form shown in Fig. 13. The power/weight relationship for this configuration is also plotted in Fig. 11.

6.1.4. Battery

A storage battery is required to supply power for the radio equipment and attitude controls when the solar cell panels are in the earth's shadow. Computations have shown that a maximum of one-ninth of the proposed orbit can be in shadow, and the

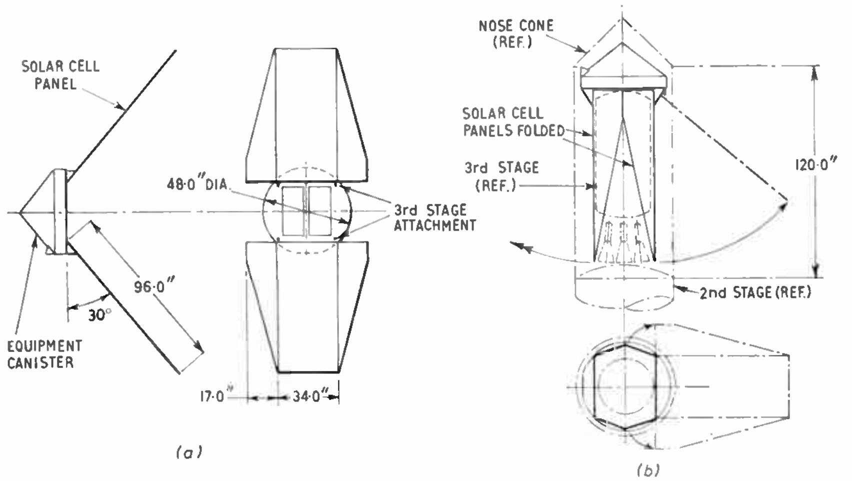


Fig. 13. Semi-sun-seeking configuration.

powers given in Table 1 allow for this. The weight has been taken as 0·15 lb/W hr, corresponding to a practical figure for nickel-cadmium cells. This may be reduced by a factor of 2 or 3 if silver cadmium cells are used instead.

6.2. *Space Limitations*

The space available for the satellite and third stage is determined by the following factors:

(a) The diameter must be not greater than that of the previous stage, and must fair into it smoothly.

(b) The height must not be so large that excessive bending moments are produced in the previous stages. Also, the height must be compatible with the dimensions of existing test and launch facilities.

(c) Provision must be made within the outline dimensions decided by the above factors for an outer shield, which will be jettisoned before orbital height is reached.

For a launch vehicle based on *Blue Streak/Black Knight*, the above considerations result in outline dimensions for the third stage and satellite of about 48 in. diameter by 120 in. long, excluding the nose cap. The latter would probably increase the overall height by a further 20 in.

A method of folding an eight-sided array of total area 220 ft^2 into this space has been described elsewhere.[17, 18] The way in which the semi-sun-seeking configuration of Fig. 13(a) can be folded is shown in Fig. 13(b). Apart from the increased watts/pound the semi-sun-seeking configuration has a number of mechanical advantages, namely

(a) Thermal control of the array and of the equipment canister can be carried out independently.

(b) The solar cell panels are in tension during the ascent phase.

(c) The effects of vibration during the ascent phase can be minimized by adding side panels, as shown, which not only provide increased stiffness but also make a useful contribution to the available power.

(d) The erecting mechanism is simple and does not involve rotation through angles exceeding 90 deg.

(e) The satellite is in the correct orientation relative to the earth when it is injected into orbit.

(f) The positions in the orbit at which station-keeping corrections will be made can be determined before launching.

The equipment canister consists mainly of a circular base frame, to which the solar cell panels are hinged, carrying the radio and other equipment on its upper face. The weight of this structure is determined mainly by the outline dimensions given above, and by the total weight of the satellite; it has been estimated at 50 lb for a diameter of 48 in. and total weight of 550 lb.

6.3. *Materials*

The environmental conditions prevailing outside the earth's atmosphere are rather different from those experienced on earth, and as the majority of existing materials have been conceived to withstand only the conditions within the earth's atmosphere, it is necessary to choose the materials for use in a satellite with care.

The effects arising from the space environment, although inter-related, can be considered under four headings: (1) temperature control, (2) meteorite and particle erosion, (3) high vacuum and (4) radiation.

The subject of materials for satellites has been discussed at length in numerous published papers (e.g. ref. 19). Full simulation of environmental conditions and realistic life testing are impossible on earth, but extensive ground testing can establish initial trends, which can then be confirmed or modified by the information transmitted back to earth by suitably equipped satellites.

The successful launching of over 50 satellites into orbit already has shown that the present knowledge of materials is sufficient to ensure orbital lives of several months (or years, in a few cases), but further investigation of materials and treatments will obviously be required. Details of proposed programmes for the development of satellite materials have been published elsewhere.[20]

6.4. *Attitude Controls and Station Keeping Equipment*

A reaction jet system for attitude control and station keeping has been described in a previous paper.[21] The total weight, including sufficient propellant for a lifetime of one year, for a 550 lb satellite in the proposed orbit will be about 85 lb. This figure is based on the use of propane for the propellant.

A longer life could be obtained in theory by using electrically driven flywheels instead of reaction jets for attitude stabilization, but this method has not been considered here as initial studies suggest that the reaction jet system is likely to be more reliable. It is possible, of course, that operating experience may cause this decision to be reversed.

The power required to operate the solenoid valves of the reaction jet system has been estimated at 5 W, averaged over the orbital period. Peak loads somewhat greater than this will be required occasionally, but these will be of only a few milliseconds duration and can be supplied by the storage battery.

The infra-red earth sensors and their associated electronic equipment will consume very little power, but a total of 10 W has been assumed, to allow for the inclusion of redundant equipment. It is particularly important that the attitude stabilization system should not fail before the radio equipment.

6.5. *Optimum Distribution of Weight*

From the information given above, it is now possible to determine the best distribution of weight between the six major divisions listed in Section 3.2.

The basic structure, and attitude control and station keeping equipment, together weigh 135 lb and consume 15 W. Telemetry and command equipment has not been considered in detail here, as it is so dependent on the amount of information on the operation of electronic devices (and on the environment of space) which will have been obtained by the time this satellite is launched. A weight of 25 lb and average power consumption of 20 W have been taken, on the assumption that the satellite will form part of an operational communication system, but it is probable that a much greater proportion of the 550 lb total would be allocated to this purpose in earlier rounds.

Of the 550 lb total weight, 390 lb is therefore available for the radio equipment (including batteries) and solar cell arrays. The latter must provide 35 W in addition to that required by the radio. This information, and the watts/pound curves of Figs. 10 and 11, are combined in Fig. 14.

It can be seen that if the radio equipment is based on 1-W repeaters, the satellite should contain twelve; alternatively, if 5-W repeaters are used, it should

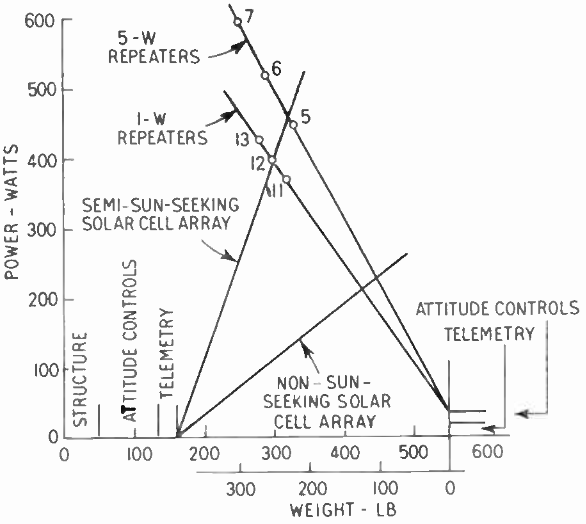


Fig. 14. Distribution of weight in a satellite of total weight 550 lb.

contain five. The areas of the solar cell arrays required will be 70 ft^2 and 83 ft^2 respectively.

Both of these solutions are compatible with the outline dimensions and semi-sun-seeking configuration. The radiated power, and hence information-handling capacity, is greater by a factor of approximately 2 in the case of 5-W repeaters. Nevertheless, the advantages of multi-station working and increased redundancy favour the use of many small repeaters. The proposed design described below is therefore based on twelve 1-W repeaters, with the weight distribution given in Table 2.

Table 2

Weight Distribution of Proposed Satellite

Equipment canister structure	50 lb	9%
Attitude control	85 lb	15½%
Telemetry and command	25 lb	4½%
Solar cell arrays	140 lb	25%
Radio repeaters and batteries	250 lb	46%
Total	550 lb	100%

7. Description of Proposed Satellite

The general arrangement of the equipment canister is shown in Fig. 15, and an artist's impression in Fig. 16.

The base frame is a honeycomb structure 4 ft in diameter. The hinges for the solar cell panels, and the attachment points to the final stage of the booster, are located on the lower face (viewed in the launch position). Also on this face are most of the nozzles for the attitude-control and station-keeping system, and the thermal-control shutters. The latter are controlled by a bimetallic spring system, which causes surfaces of high emissivity to be exposed when the internal temperature of the canister rises.

The upper face of the base frame carries the twelve unit repeaters, the batteries, power distribution system, aerials, and combining and splitting units. The travelling-wave tubes are mounted radially, to withstand the high launch accelerations, and are connected at each end to a thermal ring main. Most of the heat generated in the travelling wave tubes is conveyed from the ring mains to a heat exchanger, through which the propellant for the attitude control system passes, and is removed from the satellite when the reaction jets are used.

Above the base frame is an auxiliary structure which carries the propellant tank, sun sensors, telemetry and command equipment, and the rest of the reaction nozzles. An overall fairing, through which the aerials project, covers this equipment and that mounted on the upper face of the base frame.

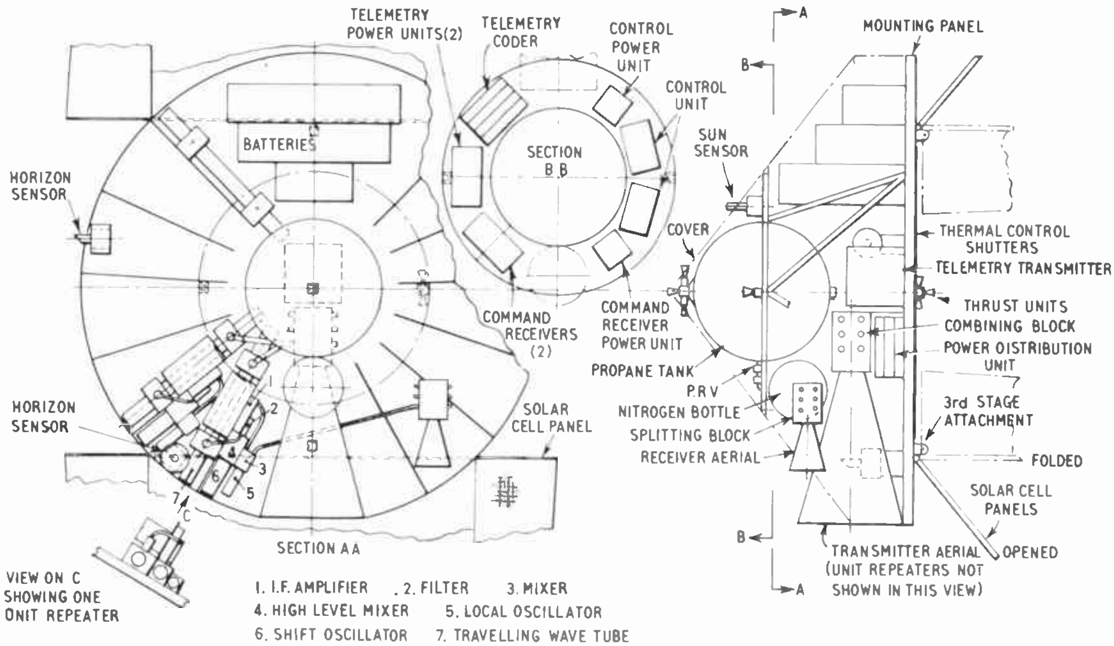


Fig. 15. Equipment canister of proposed satellite.

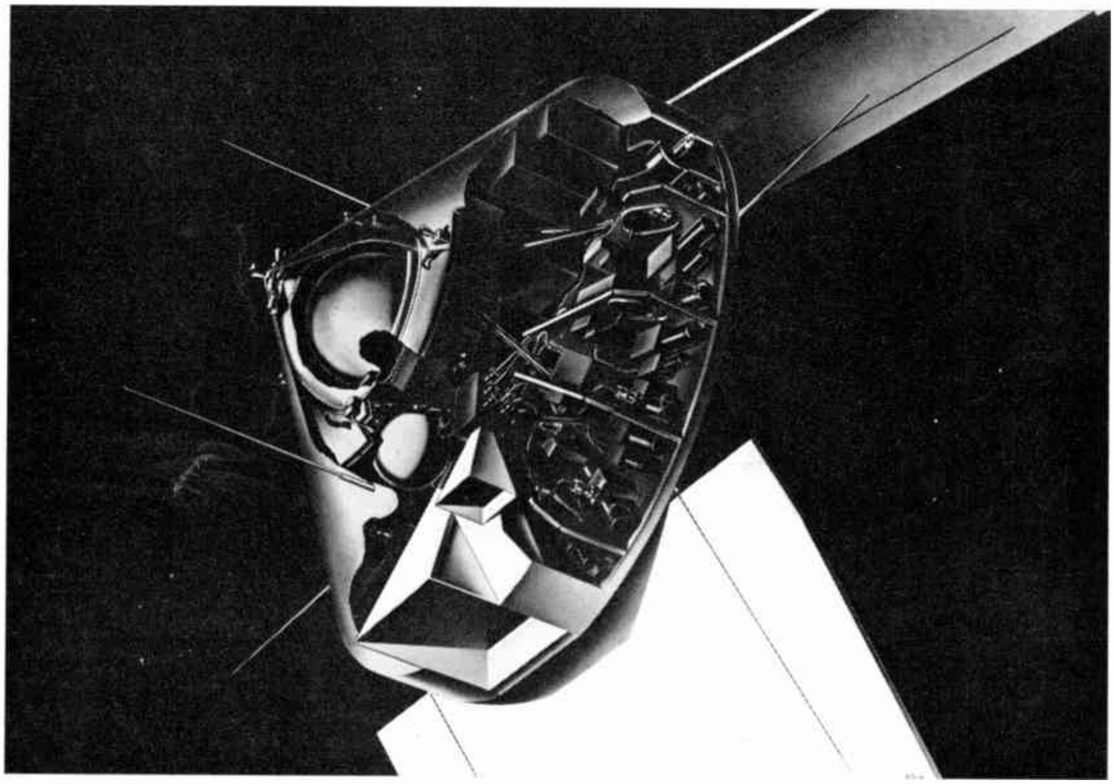


Fig. 16. Artist's impression of proposed satellite.

The solar cell panels are 8 ft long, and from 5·66 ft to 2·83 ft wide when fully extended. In the stowed position they rest against stops on the wall of the final stage tank. Separation is effected by firing the four explosive bolts which attach the base frame to the final stage; these are designed to prevent any shrapnel damage to the satellite. At the same time, the array is erected by spring-loaded struts, and is locked in the extended position by spring-loaded plungers so that the angle between the panels is 120 deg. This figure is a compromise between the requirements for maximum available power and minimum interference with the aerial radiation patterns. As the array is already erecting when separation takes place, the possibility of fouling the final stage during this process is greatly reduced.

The radio equipment is essentially the same as shown in the block diagram of Fig. 9. The unit repeaters, batteries, and solar cell panels are connected together by the power distribution system in such a way that, if all the equipment is functioning normally, the satellite is divided into twelve electrically self-contained parts. In the event of failure of one or more components, the power supplies can be redistributed by command from the ground to make best use of the remaining operational equipment.

8. Performance

8.1. *Signal Power at Ground Receiver*

The power entering the ground receiver is given by eqn. (1) of Section 4.3,

$$\text{where } P_{TS} = 1 \text{ W}$$

$$G_{TS} = 12 \text{ dB}$$

$$A_{RG} = \frac{\eta . \pi d^2}{4} = 4000 \text{ ft}^2 \ (\eta = 50\%)$$

$$R = 16\,000 \text{ miles} = 8{\cdot}5 \times 10^7 \text{ ft}$$

i.e. the ideal received power is 7×10^{-13} W.

This figure will be reduced in practice by losses due to atmospheric absorption, polarization effects, feeder attenuation, and aerial misalignment.

Considering these factors:

(a) Absorption by the atmospheric gases at frequencies between 2 and 6 kMc/s can be neglected.[11] Absorption by rain of 10 mm/hr over a 10 km path (corresponding to curve B of Fig. 7) is also negligible at 2 kMc/s; it amounts to about 1 dB at 6 kMc/s, however, and must be allowed for in the ground/satellite part of the link.

(b) With unstabilized satellites, and frequencies at which ionospheric effects are appreciable, it is customary to use linear polarization at the satellite and circular at the ground. This normally introduces a fixed loss of 3 dB. However, the proposed semi-sun-seeking satellite will rotate only once per orbit at a predictable rate, and it should be possible to follow the rotation of the plane of polarization of the received signal at the ground. In this case no allowance for polarization losses is necessary.

(c) Feeder attenuation, as mentioned in Section 4.4, may be appreciable if the maser is not mounted at the focus of the receiver aerial. Assuming that standard waveguides are used, the attenuation and mismatch losses will probably amount to 0·5 dB at 2 kMc/s.

(d) On point-to-point links, fading occurs during anomalous propagation conditions because changes in atmospheric refraction result in an effective misalignment between the transmitter and receiver aerials. If the ground receiver of the satellite system is fitted with auto-tracking facilities, fading due to this effect should be eliminated, provided that the total refraction in the atmosphere does not exceed $7\frac{1}{2}$ deg and that the accuracy of tracking is well within the beamwidth of the aerial. This latter point may be of considerable importance if the apparent direction of the satellite changes rapidly by amounts comparable with the beamwidth. In the absence of operating experience on communication satellite systems, it seems reasonable that a fairly large factor for reduction of signal due to aerial misalignment should be allowed for. For the purpose of this paper, a factor of 6 dB has been assumed, corresponding to a misalignment of about $\pm \frac{1}{4}°$.

It is perhaps worth mentioning that if refraction and other effects make it impracticable to follow the satellite with a 100 ft dish to better than $\pm \frac{3}{4}°$, fades of 30 dB will be experienced and the whole concept of the communication satellite system will break down.

Taking 0·5 dB from (c) and 6 dB from (d), the available signal power at the receiver input will be reduced from 7×10^{-13} W to $1{\cdot}6 \times 10^{-13}$ W, per watt of power radiated from the satellite.

8.2. *R.F. Bandwidth*

From Fig. 6, the receiver noise temperature at 2 kMc/s is about 80°K, so that the receiver noise power per cycle is $1{\cdot}1 \times 10^{-21}$ W. It can be seen at once that if a signal/noise ratio of about 50 dB in an audio bandwidth of 3 kc/s is required, one watt of radiated power is barely adequate for a single channel if a.m. is used. This confirms that "first generation" satellites will have to use f.m. or phase modulation for the satellite/ground part of the link.

The maximum r.f. bandwidth which can be used is set by the f.m. improvement threshold. If this is taken to be a carrier/noise ratio of 12 dB, from the figures given above the maximum usable bandwidth is 9 Mc/s per watt of radiated power.

The proposed design incorporating 12 unit 1-W-repeaters will therefore require 12 frequency allocations in the 2 kMc/s band and 12 in the 6 kMc/s band, each about 9 Mc/s wide; the total r.f. bandwidth will be about 220 Mc/s. These figures assume that the best possible channel signal/noise ratio has to be obtained with the limited power available.

8.3. *Channel Signal/noise Ratio*

Assuming that the ground stations will be connected to the national networks by links conforming to C.C.I.R. recommendations, the telephone channels should preferably be assembled in the standard frequency division multiplexed form, occupying the basebands given in Table 3.

An r.f. bandwidth of 9 Mc/s will accommodate a multi-channel f.m. signal with a peak deviation Δf_M and highest modulation frequency $F_{\max}$ given by

$$9000 = 2(\Delta f_M + F_{\max}) \qquad \ldots\ldots(4)$$

The signal/noise ratio for the worst channel is given by[21]

$$\frac{B}{2b} \cdot \left(\frac{\Delta f}{F}\right)^2 . \, C/N \qquad \ldots\ldots(5)$$

where B = r.f. bandwidth = 9000 kc/s

b = a.f. bandwidth = 3 kc/s

Δf = peak deviation corresponding to single channel of mean frequency F

C/N = carrier/noise ratio at input = 12 dB.

The relation between Δf for a single channel and Δf_M for the multichannel signal depends to some extent on assumptions made about speech characteristics. The values given in column 4 of Table 3 are derived from Fig. 3 of reference 21.

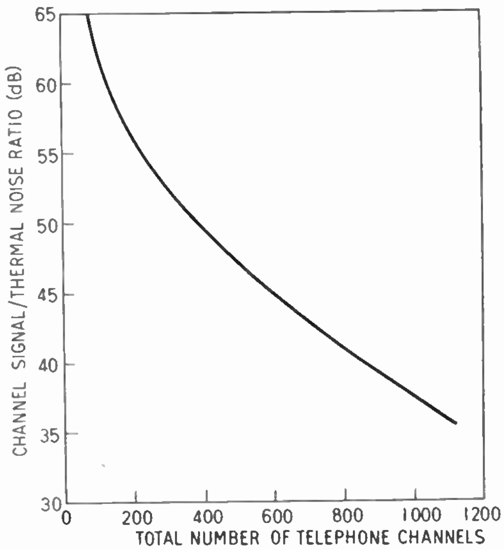
Signal/noise ratios corresponding to numbers of channels between 12 and 180 per unit repeater are given in column 5 of Table 3. Assuming that two-way telephone channels require a separate repeater for each direction, the total number of channels which a single satellite can handle is six times the figures in column 1. This is shown in graphical form in Fig. 17.

Table 3

Performance of Unit Repeater with 9 Mc/s R.F. Bandwidth

No. of channels	Baseband kc/s	Δf_M kc/s	$\Delta f_M/\Delta f$	S/N dB
12	12–60	4440	6·3	65
24	12–108	4390	6·9	59
36	12–156	4340	7·2	55
60	12–252	4250	7·5	51
96	12–396	4100	8·4	46
120	12–552	3950	8·7	42
144	60–636	3860	9·1	40
180	60–800	3700	9·6	37

No allowance for pre-emphasis was made in calculating the signal/noise ratios for Table 3. An improvement of about 3 dB by this means is theoretically possible.



Note: S/N does not include pre-emphasis and psophometric factors.

Fig. 17. Performance of satellite with twelve 1-watt repeaters.

It will be seen that a total of about 400 two-way channels can be handled if a signal/noise ratio of 50 dB, corresponding to C.C.I.R. recommendations, is required. If the signal/noise ratio is relaxed to 40 dB, which is still a reasonable quality by commercial standards, the number of two-way channels can be increased to 1000.

9. Conclusions

In this paper an attempt has been made to show that communication satellites based on current radio techniques, and capable of being launched by existing boosters, should be able to make a really useful contribution to the telephone needs of the world. Of course, much development work remains to be done, especially on the attitude control and station keeping equipment, before a world-wide communication satellite system for public use can be established, and it has yet to be shown that communications equipment can withstand the environment of space for long periods. There is no doubt, however, that the advantages to be gained by using communication satellites to supplement h.f. radio and submarine cables will justify the expenditure of considerable effort on overcoming the difficulties involved.

10. Acknowledgments

The design of a satellite is essentially the work of a team rather than individuals, and the authors are pleased to acknowledge the assistance given by their colleagues in the Advanced Projects Group in the preparation of this paper. The authors also wish to thank the Board of Hawker Siddeley Aviation Limited for permission to publish the paper.

The above proposals are the result of studies carried out before the formation of the British Space Development Company Ltd., and are not necessarily representative of the latter Company's thoughts on Communication Satellites.

11. References

1. W. F. Hilton and S. R. Dauncey, "Communications satellite orbits", Paper presented at XIth Congress of the International Astronautical Federation, Stockholm, August 1960.
2. W. F. Hilton, "Communications satellite systems suitable for Commonwealth communications", Paper presented at a B.I.S. Symposium on Communication Satellites, 12th May 1961.
3. "Frequency Needs for Space Communication", Testimony and Exhibits of the American Telephone and Telegraph Company. F.C.C. Docket No. 11866, July 1960.
4. L. Pollack, "Radio communication using earth-satellite repeaters", *Elect. Commun.*, **36**, No. 3, pp. 180–188, 1960.
5. Comments of Radio Corporation of America and R.C.A. Communications Inc. before the F.C.C. Docket No. 13522, March 1st, 1961.
6. W. F. Hilton and B. Stewart, "The advantages of attitude stabilization and station keeping in communication satellite orbits", *J. Brit.I.R.E.*, **22**, pp. 193–203, September 1961.
7. A. G. Haley, "The 1963 Extraordinary Administrative Radio Conference on Space Communications", *Telecommunication Journal*, **28**, pp. 9–20, January 1961.
8. "Factors Affecting the Selection of Frequencies for Telecommunication with and between Space Vehicles". C.C.I.R. Report No. 115. Published in Vol. III of the Documents of the IXth Plenary Assembly, Los Angeles, 1959.
9. C. V. Stanley and T. P. Dixon, "Horizon scanner has no rotating members", *Electronics*, **34**, No. 2, p. 104, 13th January 1961.
10. D. C. Hogg, "Effective antenna temperatures due to oxygen and water vapour in the atmosphere", *J. Appl. Phys.*, **30**, No. 9, pp. 1417–19, September 1959.
11. J. W. Ryde, "Attenuation and radar echoes produced at centimetre wavelengths by various meteorological phenomena". "Conference on Meteorological Factors in Radio-Wave Propagation", pp. 169–189 (Physical Society, London, 1946).
12. W. J. Bray, H. G. Hopkins, F. A. Kitchen and J. A. Saxton, "Review of long distance radio-wave propagation above 30 Mc/s", *Proc. Instn Elect. Engrs*, **102 B**, pp. 87–95, January 1955 (I.E.E. Paper No. 1782R).
13. "Project Relay will have year's orbit life", *Aviation Week*, **74**, No. 5, pp. 93–101, 30th January 1961.
14. Symposium of Papers on Microwave Links. *Proc. Instn Elect. Engrs*, **99**, Pt. III, No. 61, pp. 245–6 and 289–90, September 1952.
15. B. Miller, "Hughes communication satellite detailed", *Aviation Week*, **73**, No. 24, pp. 52 *et seq.*, 12th December 1960.
16. J. M. Denney, "Radiation Damage in Satellite Solar Cell Power Systems", American Rocket Society Report No. 1295-60.
17. "Industry and Space—A Franco-British Study of European Potential." Hawker Siddeley/S.E.R.E.B. Joint Report, February 1961.
18. "Hawker Siddeley's communication satellite", *Aeroplane*, **100**, No. 2577, pp. 258–60, 10th March 1961.
19. F. J. Clauss, "Surface effects on materials in near space", *Aerospace Engineering*, **19**, No. 10, pp. 16 *et seq.*, October 1960.
20. M. A. Steinberg, J. L. Fox and B. W. Augenstein, "Proposed ground and flight program to develop space age materials", *American Rocket Soc. J.*, **31**, No. 2, pp. 228–36, February 1961.
21. B. Stewart and P. A. E. Stewart. "Dynamics and engineering of satellite attitude control systems", Paper presented at European Space Technology Symposium, London, June 1961.
22. J. W. O. Halina, "Design of f.m. radio relay systems for multi-channel telephone service", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-4**, No. 3, pp. 3–13, October 1956.

Manuscript received by the Institution on 8th May 1961 (Paper No. 666).

Points from the discussion on this paper are included on pages 203–4 of this *Journal*.

# of current interest . . .

Radio Research 1960

The report of the Radio Research Board for 1960* describes the Minitrack facilities recently opened at Winkfield and operated in association with NASA as "a valuable asset to space research at the Station and to the national effort in this field". (The Winkfield installation is dealt with in some detail in the Convention paper by Dr. B. G. Pressey which is published in the August *Journal.*)

Preparations are also in hand to collaborate with Canada and the U.S.A. in investigations on the ionosphere from above, by means of satellites in orbit at the appropriate heights: the so-called "top-side" sounding experiment referred to in another Convention paper ("The Canadian Defence Research Board Topside Sounder Satellite" by R. C. Langille and J. C. W. Scott). In addition an experiment has been initiated in which use will be made of an apparatus to be launched by rocket in order to explore the radio-wave field above a low-frequency transmitter on the ground, and so obtain information about the lowest region of the ionosphere.

The report shows that the work of the Station is in progress of being re-oriented so as to apply about fifty per cent of the Station's effort to space research problems. Signals have been received from a number of satellites on various frequencies and measurements made on Doppler frequency-shift, Faraday fading and signal-strength scintillations. These data have been used in studies of the ionosphere. Predictions of the position of all satellites of interest to observers in the U.K. have continued to be issued. The flow of observations required for predictions has increased, particularly with the launching of larger satellites such as the communications balloon-satellite, *Echo*.

As a result of the knowledge of ionospheric behaviour that has been built up at the Radio Research Station over many years, a "once for all" book referring to world-wide conditions in those parts of the ionosphere which control long distance communications is being prepared. Research into some of the remaining problems in both long distance and short distance transmissions in the field of propagation at very high frequencies through the troposphere, has been actively pursued during the period under review. An important feature of the Station's work in this field continues to be direct study of the refractive index structure of the lower atmosphere.

The analysis of atmospheric radio noise data obtained during the International Geophysical Year has been continued and further studies of local lightning flashes made. The necessity for assessing the interference to be expected from man-made noise at any given receiving location and in particular the need to find a "quiet" site for the Station's projected radio telescope have required experimental studies of this kind of noise in the v.h.f. and u.h.f. band.

* "Radio Research 1960". Published for D.S.I.R. by H.M.S.O. Price 2s. 6d.

Electronic Telecontrol Equipment in the Fuel and Power Industries

For a number of years close co-operation between the Coal, Electricity and Gas Industries in connection with radio communication and frequency planning problems has been achieved through the Joint Radio Committee of the Nationalized Fuel and Power Industries.

This Committee has long realized that the three industries have much in common in regard to the remote indication and control of plant at unattended points over both wire and radio links. As the basic requirements of the separate industries are virtually identical, although related to the indication and control of different types of plant, the Joint Radio Committee has prepared a Basic Specification for Electronic Telecontrol Equipment which sets out the common requirements. In this way it is hoped to rationalize the production of suitable equipment and avoid the necessity to design equipments on an individual basis for specific applications.

The Specification provides information on the minimum facilities required and guidance on the techniques preferred, e.g. the use of items such as cold cathode tubes, transistors, magnetic cores etc., rather than electro-magnetic relays and hot cathode valves, but it intentionally avoids laying down rigidly the exact manner in which any facility should be provided.

The equipment is intended to provide a highly flexible and readily extensible telecontrol system capable of keeping pace with the rapid growth of electricity and gas distribution; as such it would no doubt be of use to other large distribution systems. A number of outstations can indicate to a central station the state of their plant by means of two-state and analogue measurements which are digitized for transmission; control of the plant from the central station is catered for but may be omitted if desired. Automatic alarm features are included and the central station equipment is arranged to give print-out for data logging if required.

Copies of the Basic Specification may be obtained from Mr. G. D. Turton, Secretary, Joint Radio Committee of the Nationalized Fuel and Power Industries, c/o Central Electricity Generating Board, Bankside House, Sumner Street, S.E.1.

U.D.C. 621.396: 629.19

Some International Aspects of Satellite Communication System Planning

By

Captain C. F. BOOTH, C.B.E.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th-8th July 1961.

Summary: The work of the organizations concerned with international planning of telecommunication systems is described. The prime need for satellite communication systems is the allocation of frequency bands, preferably on an exclusive basis, but some sharing with low-power radio-relay systems may be necessary. Other aspects requiring international agreement include the minimization of interference with other services, and the modulation and base-band characteristics.

1. Introduction

Much has been said and written about the technical aspects of satellite communication systems, including the orbits to be used, the design of satellites and ground stations and so on, and it is not the intention in this paper to discuss the detailed design of such systems. However, it may be noted that the Post Office fully appreciates the potentialities of this field and is actively engaged in planning and development work. Comprehensive studies are being made with the Royal Aircraft Establishment on the design of satellite communication systems, and a ground station is being built in Cornwall for experimental tests across the Atlantic on Project RELAY in collaboration with the National Aeronautics and Space Administration.

To the present there has been little discussion on the international aspects of planning satellite communication systems. These may present some difficult problems and the present paper is devoted to this theme.

The international aspects are important for four main reasons:

Firstly, a satellite system requires the use of radio; it follows that its utilization of the radio spectrum must be such as not to produce interference to, or suffer interference from, other radio services.

Secondly, a satellite system is essentially a medium for world-wide communication, it must therefore meet the diverse communication requirements of many countries.

Thirdly, a satellite system should meet the transmission performance standards already agreed internationally for long-distance circuits; these refer to noise, stability, transmission delay, etc.

Fourthly, it must be possible readily to interconnect the long-distance circuits provided by a satellite system with the national networks of the countries it serves.

It is perhaps not too much to say that the success of satellite communications will depend in large measure on the effectiveness with which the international planning is carried out.

It is fortunate that there is in existence an organization—the International Telecommunication Union,[1,2] an organ of the United Nations—which has already demonstrated its usefulness as a forum for the achievement of international agreements on the design, planning and operation of telecommunication systems of all kinds. In this work the I.T.U. is assisted by two technical advisory committees: the International Radio Consultative Committee (C.C.I.R.) and the International Telephone & Telegraph Consultative Committee (C.C.I.T.T.). The former is concerned with the study of technical and operating questions relating specifically to radio communication and with the issue of recommendations on them. The latter deals with similar questions and recommendations concerning line transmission, as well as with the integration of line and radio systems.

It is clear that the C.C.I.R. will be closely concerned with the achievement of international agreements in respect of the use of radio by satellite communication systems. It is of interest to note here that the C.C.I.R. has already produced[3] preferred technical characteristics for international line-of-sight radio-relay systems, and that these are widely used both for links between countries and for links in national networks. These characteristics are:

(a) overall transmission performance for telephony, telegraphy and television;
(b) baseband characteristics;
(c) intermediate-frequency characteristics;
(d) modulation characteristics;
(e) patterns of radio-frequency usage;

† Deputy Engineer-in-Chief, Post Office.

(f) control, monitoring and supervisory system characteristics.

Similar studies will have to be made in the near future for satellite communication systems.

It would appear that the C.C.I.T.T. will be concerned primarily with the overall transmission performance of satellite systems and with their full integration into the general network of communications. This has already been achieved in the case of conventional radio-relay systems, which are fully integrated with national cable systems.

For completeness brief reference must be made to the valuable work on the scientific aspects of space which is being carried out by other organs of the United Nations, including the International Scientific Radio Union (U.R.S.I.)[2] and the Committee on Space Research (C.O.S.P.A.R.). These bodies are organizing effectively in the scientific plane the fundamental studies on which the design of practical communication systems must be based.

2. Aspects Requiring International Agreement

It will be of interest to consider those aspects of a satellite communication system on which international agreement is necessary. First and foremost there is the question of the radio frequency bands to be used—a matter of fundamental importance. The selected frequency bands must be suitable not only from a technical point of view, but they must also allow for the present and planned utilization of the radio spectrum.

From a technical point of view, and in the light of present knowledge, it would appear that the frequency range from about 1000 to 10 000 Mc/s is suitable for satellite communication systems. This takes account of the relatively high level of cosmic noise at frequencies below about 1000 Mc/s and of atmospheric noise (including rain noise) at frequencies above some 10 000 Mc/s. However, the range from 1000 to 10 000 Mc/s is already being used for a variety of radio services, some of which, such as high-power radar and tropospheric-scatter services, would be incompatible with satellite systems because of the interference they would produce. In fact, at the last Administrative Radio Conference (Geneva, 1959), at which the whole frequency range from 10 kc/s to 40 000 Mc/s was allotted to the various types of service, the band which is now of interest to space communications, 1000 to 10 000 Mc/s, was more or less used up for other services.[4,5] It should be noted that at that time there was insufficient knowledge of the frequency requirements for space communication systems to consider allocations for them, and it was possible to allocate only relatively narrow frequency bands for space research purposes. In the frequency range of interest these comprise the following:

1427–1429 Mc/s (2 Mc/s)	2290–2300 Mc/s (10 Mc/s)
1700–1710 „ (10 „)	5250–5255 „ (5 „)
8400–8500 Mc/s (100 Mc/s)	

The use of these bands for space research is on the understanding that such use does not cause interference to other services using the bands. These allocations are of course totally inadequate for space communication systems. However, it was agreed at Geneva that an Administrative Radio Conference[6] should be called in 1963 to consider frequency allocations for space communication systems. The work done then will clearly lay the foundations on which such systems will be built.

Agreement in 1963 to allocate relatively wide exclusive bands for the purpose would, to say the least, be most difficult to achieve. Fortunately there are reasons to believe that, given suitable planning and proper system co-ordination, satellite systems could share frequency bands with low-power line-of-sight radio-relay systems without mutual interference.

The frequency bands which might be considered are : 3700–4200 Mc/s and 5925–6425 Mc/s.

These bands are widely used by many countries for low-power radio-relay systems in accordance with the recommendations of the C.C.I.R., and it is thought that they offer the prospect of international agreement for use on a shared basis by satellite communication systems. However, in the long term, some additional frequency space may well be required for satellite systems, e.g. between 6425 Mc/s and about 10 000 Mc/s.

Within the designated bands it will be desirable to define, on an internationally agreed basis, patterns of radio frequency channel usage so that satellites may co-operate effectively with ground radio stations in various countries. In order to minimize mutual interference these satellite channels might be interleaved with the radio-frequency channels of line-of-sight systems. The frequency pattern of satellite channels will have to take into account not only the technical characteristics of each satellite system, but also the possibility that, in the course of time, more than one world-wide system may be established.

Agreements will also be necessary on the modulation and baseband characteristics employed by satellite systems. In these respects also, as has already been indicated, a parallel exists with line-of-sight radio-relay systems, where preferred characteristics have been established by the C.C.I.R. for international use. Economy in the use of the spectrum, and the choice of optimum modulation techniques, will also be essential.

Studies are required of the interference aspects between satellite systems and other radio services, e.g. line-of-sight radio-relay systems, sharing the same frequency bands and work is in hand in this field. These studies need to include, for example, the protection ratios necessary for the services concerned and propagation studies related to the frequencies to be used. Such studies are necessary to enable the locations and frequencies of satellite system ground stations and the stations of other radio services to be planned on an interference-free basis. This matter has international as well as national aspects since there is a possibility of country-to-country interference unless the planning of such stations and services is properly co-ordinated.

There are a number of operational aspects of satellite systems to be considered on the international plane, including the following:

—the locations of satellite ground stations, including relay stations, from the standpoint of traffic flow and interconnection with existing and planned national communication networks;

—the sub-division of satellite circuit capacity to meet the different communication requirements of the various countries;

—the methods to be used for satellite tracking and control, and the organization of such facilities on an internationally co-ordinated basis;

—the procedures required for satellite-to-satellite switching in non-synchronous systems;

—the arrangements to be made for the maintenance of satellite systems, including the replacement of satellites on failure.

In the foregoing the all-important financial aspects, including the sharing of the costs of provision and maintenance and traffic revenues have been omitted. Such matters will be in large measure the concern of Telecommunication Administrations of the participating countries.

3. Procedures for Arriving at International Agreements

On technical matters special responsibility attaches to Study Group IV of the C.C.I.R., which is responsible for the preparation of recommendations relating to the preferred technical characteristics of space systems. An interim meeting of this Study Group is to be held in the U.S.A. in the Spring of 1962 in preparation for the Xth Plenary Assembly of the C.C.I.R. in early 1963. One of the most important issues for the Study Group meeting will be the question of the pattern of radio frequency usage, and, with several other countries, the United Kingdom is actively concerned in formulating its proposals.

It may perhaps be useful here to explain that the work of the C.C.I.R. Study Groups begins on the national plane and is generally organized by the Telecommunications Administrations of the countries concerned. Each National Study Group usually comprises representatives of the Telecommunication Administrations, other Government Departments, scientific bodies, industrial associations, broadcasting organizations and operating agencies. The national Study Groups prepare draft recommendations which are then considered by the international Study Group, and the final outcome is then used as a basis for arriving at international agreement at the Plenary Meetings of the C.C.I.R.

The decisions of the Administrative Radio Conference in 1963, which is to consider space systems, will be based on the technical recommendations of the C.C.I.R. and this underlines the importance of these recommendations and the need for detailed and careful study in their preparation.

4. Conclusions

Although this survey is brief and perhaps incomplete, it is hoped that it has served to focus attention on some of the important international aspects which Telecommunication Administrations are now considering in their study of satellite systems.

In conclusion, it may be said that the development and introduction of satellite communication systems will require the closest possible international co-operation, planning and agreement, if the full potentialities of this new medium are to be realized.

Fortunately, there are in existence the appropriate international organizations for achieving a large measure of this co-operation, and it rests with all concerned to use, as effectively as possible, the tools that are to hand.

5. References

1. International Telecommunications Convention Geneva, 1959. Article 4, Purposes of the Union. Article 13, International Consultative Committee.
2. R. L. Smith-Rose, "International radio organizations—some aspects of their work", *J. Brit.I.R.E.*, **18**, pp. 631-9, November 1958.
3. W. J. Bray, "The Standardization of international microwave radio-relay systems", *Proc. Instn Elect. Engrs*, **108B**, pp. 180–200, March 1961.
4. Radio Regulations, Geneva, 1959, Article 5, Section IV, Table of Frequency Allocations.
5. C. F. Booth, "New radio regulations", *J. Instn Elect. Engrs*, **6**, pp. 431–3, July 1960.
6. Radio Regulations, Geneva, 1959, Recommendation No. 36, Relating to the Convening of an Extraordinary Administrative Radio Conference to allocate Frequency Bands for Space Radiocommunication Purposes.

Manuscript received by the Institution on 13th July 1961 (Contribution No. 37).

APPLICANTS FOR ELECTION AND TRANSFER

As a result of its meeting on 22nd August the Membership Committee recommended to the Council the following elections and transfers.

In accordance with a resolution of Council, and in the absence of any objections, the election and transfer of the candidates to the class indicated will be confirmed fourteen days after the date of circulation of this list. Any objections or communications concerning these elections should be addressed to the General Secretary for submission to the Council.

Direct Election to Member

VEVERS, Lieutenant Colonel John, R.E.M.E. *London, S.E.18.*

Transfer from Associate Member to Member

FAIRFAX, Commander Osmond Maurice, R.N. *Bath, Somerset.*
JONES, William Anderson. *Glasgow.*

Direct Election to Associate Member

BAGLEY, Geoffrey Charles. *Ironbridge, Shropshire.*
BARKER, Flight Lieutenant John Edward, R.A.F. *London, N.8.*
COWAN, Leslie Frank, B.Sc.(Eng.). *Burnham, Buckinghamshire.*
HOW, Richard Edward. *Harrow, Middlesex.*
KARANDIKAR, Narayanan Shankar. *London, S.W.11.*
KING, Captain Reginald Alfred, R. Sigs. *Balcombe, Australia.*
LEETHAM, Bernard Alfred. *Great Malvern, Worcestershire.*
MOORE, Wing Commander Charles Desmond, R.A.F. *Jordans Buckinghamshire.*
SHAW, Henry James. *Carshalton, Surrey.*
STEVENSON, Edwin Charles. *Cheshunt, Hertfordshire.*
WORTH, Desmond Frederick Joseph, B.Sc.(Eng.). *London, S.W.17.*

Transfer from Associate to Associate Member

GRENFELL, Commander William Edward, R.N. *Konigswinter/Rh, West Germany.*

Transfer from Graduate to Associate Member

DUFFEY, Eric. *Liverpool.*
DWIVEDI, Captain Munesh Bal, M.Sc., Indian Sigs. *Poona, India.*
EVANS, Peter Rowland. *Wembley, Middx.*
LOVERING, Roy Thomas. *Maidenhead, Berkshire.*
McKAY, Eric Martin, B.Sc. *Berkhamsted, Hertfordshire.*
OGDEN, William Robert. *Edinburgh*
WISE, Philip Noel. *Stevenage, Hertfordshire.*

Transfer from Student to Associate Member

MARWICK, David Baikie. *Wantage, Berks.*
MUNROE, Duke Gray. *Accra, Ghana.*

Direct Election to Associate

GORHAM, Horace Henry James. *Watford, Hertfordshire.*
HEWITT, Donald James. *Ashford, Middlesex.*
SHEPPARD, Charles Vincent Francis. *London, S.E.24.*
STEWART, Ritchie. *Willerby, East Yorkshire.*

Direct Election to Graduate

ASHTON, Vernon. *Woodley, Berkshire.*
BEWS, Michael Howard. *Liverpool.*
BOWMAN, John Singleton. *Manchester.*
CODD, Paul Oliver. *Letchworth, Hertfordshire.*
COLLINS, Joseph John Domnic, B.E. *London, S.E.26.*
CROSBY, Edward Arthur. *Liverpool.*
EVANS, Michael Oliver, B.E. *South Croydon, Surrey.*
GEORGE, Christopher Richard. *Stevenage, Hertfordshire.*
GILHAM, Alan Edward. *Morden, Surrey.*
GODDARD, Fred Keith Scott. *Woodley, Berkshire.*
GRIGSON, Flight Lieutenant Michael William, R.A.F. *Woking, Surrey.*
HALL, Captain Harold Edward, R.E.M.E. *Wilmington, Kent.*
*HAUGHEY, Peter. *Wirral, Cheshire.*
HENDERSON, Frank. *Brockworth, Gloucestershire.*
HONEY, Denis William, B.E. *Halewood, Lancashire.*
LANG, James. *Liverpool.*
LESTER, Roy Stanley. *London, S.W.11.*
LOVELADY, Leo George. *Liverpool.*
RICKETT, Thomas Walter. *Fulbourn, Cambridgeshire.*
TOULSON, Edward. *Sutton Coldfield, Warwickshire.*

Transfer from Associate to Graduate

MASTRONARDI, Flight Lieutenant Edward John, B.A., M.C. R.C.A.F. *Ottawa, Canada.*

Transfer from Student to Graduate

DALE, Collis Seymour. *Thetford, Norfolk.*
ELLIOTT, Brian Charles. *Reading, Berks.*
EVANS, John Hadley. *Cardiff.*
FRASER, William Morrison. *Weston-super-Mare, Somerset.*
HARTNELL, Desmond. *Stanmore, Middlesex.*
KARTHIKEYAN, Muthukmaraswamy, B.Sc. *Madras.*
KENT, Derek Wilfred. *Enfield, Middlesex.*
LING, Shun Ki. *Hong Kong.*
TALMACIU, Josef. *London, N.4.*
TSAPPARELLI, Louis Christos. *London, N.W.5.*
TURNBULL, Herbert Johnston. *Bishop Auckland.*
TURNER, Derek Walter. *Spalding, Lincs.*
WADDINGTON, Bryan Richard. *Farnham, Surrey.*

STUDENTSHIP REGISTRATIONS

The following students were registered at the 22nd August meeting of the Committee.

AHMAD, Jameel. *Lahore, West Pakistan.*
ANDREWS, Herbert William. *Bexley, Kent.*

BIRLEY, Roy. E. *South Brisbane, Australia.*
BIXBY, Bryan John. *Hatfield, Hertfordshire.*
BONFIELD, Christopher George. *Wantage, Berkshire.*

COULBER Roger, B.Sc. *New Malden.*

DAGGER, Flying Officer Alfred, R.A.F. *Weston-super-Mare, Somerset*

EDWARDS, Pilot Officer Derek Anthony Harding. *Downpatrick, N. Ireland.*

FULTON, John Michael Lewis. *Singapore.*

GARDINER, Bernard John. *Newbury, Berkshire.*
GAUNT, John Richard. *Fiji.*
GREEN, John Brian. *Sidcup, Kent.*
GUPTA, Suresh Kumar. *London, W.C.1.*

HUSBAND, Anthony John. *Dorchester, Dorset.*

LIEW TIEW GUAN, Ernest. *Singapore.*
LOW HUN KHING. *Kuala Lumpur.*

MIDDLETON, Leonard, B.Sc. *Winnipeg, Canada.*

NAMJOSHI, Madhusudan Waman, B.Sc. *Bombay.*

PADMANATHAN, Muruguppillai, B.Sc. *Point-Pedro, Ceylon.*

SEN JWAK LEE. *Singapore.*
STEVENS, John Edward. *Braunton, Devon.*

THOO FOOT SIONG. *Kuala Lumpur.*
THURSTON, John William. *Gosport, Hampshire.*

WALKER, William Rodney. *Sydney, Australia.*
WHITEHEAD, John Bentley. *Thetford, Norfolk.*
WILSON, Arnold. *London, E.4.*

* Reinstatement.

U.D.C. 531.383: 629.19

Inertial Systems in Space Vehicles

By

M. A. V. MATTHEWS†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: This paper discusses the use of inertial elements in space vehicles for attitude control and navigation. Some limitations of inertial systems in this environment are described and it is shown that inertia alone is not adequate for most space applications. However, the use of inertia to bridge the gaps between information from other sources can simplify the requirements for other sensors in the satellite, and also on the ground tracking equipment. This is illustrated by some examples and by discussion of the characteristics of inertial components suitable for a space vehicle.

1. Introduction

Inertial systems are those which operate by measuring the deviations of a frame of reference fixed in the vehicle from an inertial frame. The sensing elements in inertial systems are gyroscopes (for measuring rotation of the frame of reference relative to an inertial frame) and accelerometers (for measuring translation of the frame of reference). Such elements, when associated with a suitable computer, can be used to determine the attitude and position of the vehicle relative to a given inertial frame, or to any other frame of reference whose motion relative to an inertial frame is known (for example, an earth fixed frame). In this paper the application of such methods to space vehicles will be discussed.

It will be shown that inertial systems will not often be used in a self-contained way: they are more likely to be associated with other methods of measurement, including radio measurement. This paper is confined, however, to giving an account of what inertial systems can contribute to the overall design of space vehicles, without attempting to assess whether the same result could be achieved by other means, or whether, in a particular instance, the contribution is worth the size, weight, and power consumption of the inertial system.

2. Attitude Reference

2.1. *Properties of Gyroscopes*

The basic element in attitude determination is the gyroscope. This consists of a mass which is spun by a motor at a high angular velocity about an axis fixed in itself. This spin axis is mounted in gimbals so that it has one or more degrees of freedom relative to the body of the vehicle. The fundamental dynamic law which determines the behaviour of gyroscopes is expressed by the vector equation which states that, relative to an inertial frame,

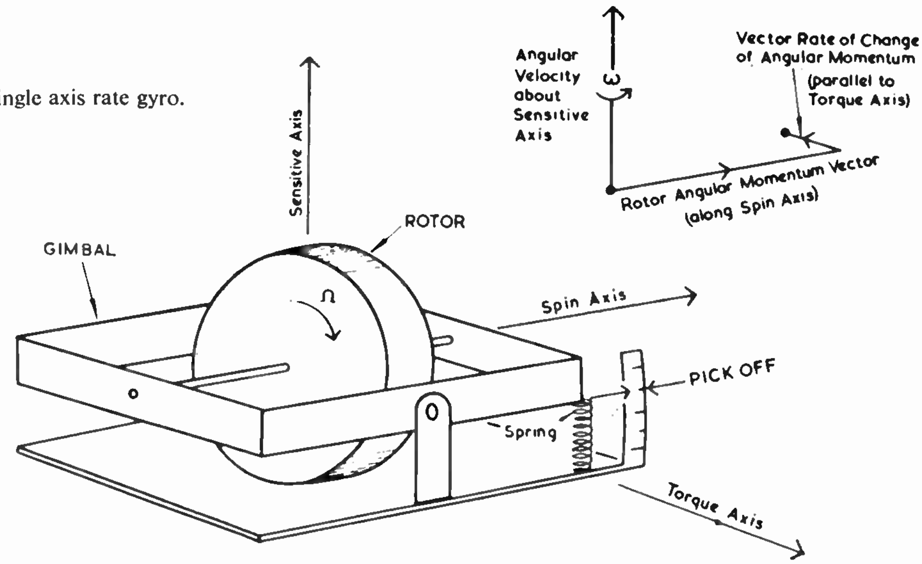
torque = rate of change of angular momentum.

† Ferranti Limited, Electronic Systems Department, Edinburgh.

Since the spin rate and therefore the angular momentum, of the gyroscope is high, any substantial change of direction of the spin axis, relative to the inertial frame, represents a large change of the angular momentum vector, which can only be brought about by a large torque × time product. Thus the gyro spin axis has an inherent tendency to maintain its direction in an inertial frame. An attitude reference could be constructed by mounting the gyro in a gimbal system giving the spin axis complete freedom to point in any direction relative to the vehicle. Such a "free gyro" would tend to maintain a fixed direction in an inertial frame, but it would develop errors because of torques due to friction etc., which would eventually cause it to wander off the required direction. These unwanted torques are a limitation on all gyroscopes. One of the objectives of gyro design is to make the unwanted torques as small as possible, and to make the angular momentum as high as possible, so as to minimize the wander rate.

The free gyro has various disadvantages and is not suitable as a high accuracy reference. Instead the *rate gyro* is used. In this instrument a measurable torque is applied to keep the spin axis in a nearly constant attitude relative to the outer case of the instrument. In a simple instrument the torque may be applied by a stiff spring. The rotation of the gyro relative to the case is then small, and proportional to the torque applied by the spring. The torque applied is proportional to the rate of change of angular momentum which depends on the angular velocity of the spin axis. Thus the measurement of torque determines the angular velocity of the spin axis which is nearly equal to the angular rate of the instrument case. To simplify the engineering the torque applied is usually measured about one axis only; such an instrument is called a single axis rate gyro, and measures the angular rate about one axis only (Fig. 1). A variant is the *integrating rate gyro* in which the integral of the applied torque is measured. This is proportional to the angular displacement of the spin axis.

Fig. 1. Single axis rate gyro.



2.2. *The Stable Platform*

A measurement of vehicle attitude could be made by mounting three rate gyros to measure the components of angular velocity about three mutually perpendicular axes in the vehicle. Integration of these three components would give the change of attitude of the vehicle since the instant when integration began (which might be an instant when attitude was known, for example just before take off). This method is simple but it has the disadvantage that the gyros have to be able to measure rates up to the maximum angular velocity of the vehicle; this may limit the attainable accuracy.

In cases where the vehicle attitude is stabilized throughout the period in which the inertial system is in use, this is not a serious limitation, but when the inertial system is initially set up before take off and has to function through the boost period, when there may be large changes of vehicle altitude, then it is necessary to use a *stable platform.* This is mounted in a system of gimbals so as to have complete angular freedom in the vehicle; servo motors on each axis can drive the platform to any required position (Fig. 2). Three rate gyros or integrating rate gyros are mounted on the platform to measure its angular rates about three mutually perpendicular axes corresponding to the gimbal axes. The servo motor on each gimbal axis is driven by a combination of angular velocity and integrated angular velocity about that axis, derived from the gyros, so as to maintain the stable platform in a fixed attitude relative to an inertial frame. Alternatively additional terms may be fed to the servo motors to cause the platform to rotate in a chosen way and remain aligned with a frame of reference whose rotation relative to an inertial frame is known (for example a frame of reference fixed in the earth).

2.3. *Monitoring the Stable Platform*

Errors in measuring the torques applied to the rate gyro will cause the platform to slowly drift from its correct position. Thus the platform will be a good attitude reference soon after it is set up, but its accuracy will deteriorate with time. If a good attitude reference is needed for prolonged periods, then some other source of attitude information will be needed to correct the accumulated platform errors. The advantage of using the platform at all in the circumstances

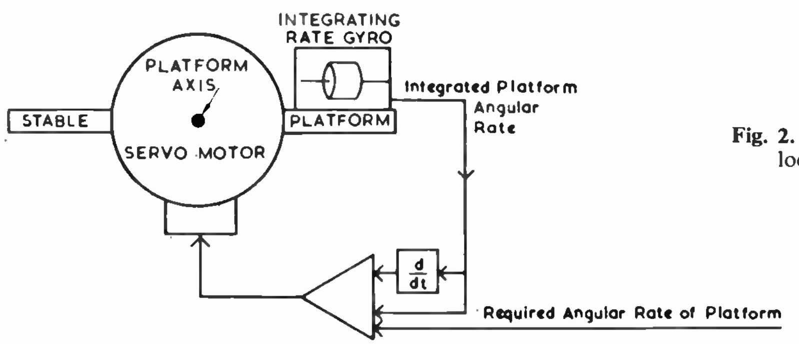


Fig. 2. Stable platform servo loop, one axis only.

is that the other attitude sensor may be intermittent or noisy. If it is intermittent, then it can be used to correct the platform when it is available, and the platform then provides a continuous reference. If it is noisy the platform is given an angular velocity proportional to the difference between the attitudes indicated by the other sensor and the platform. The gain of this control loop is made low so that the platform follows the attitude sensor with a long time constant, and the noise is smoothed out. This process is known as monitoring the platform.

2.4. *The Requirement for an Attitude Reference in a Space Vehicle*

2.4.1. Attitude control during boost

A space vehicle is placed in a required orbit by controlling the magnitude and direction of the thrust applied to it during the boost period. An attitude reference is therefore needed so that any error in the direction of the thrust vector can be detected and corrected. An inertial reference is very suitable for this purpose since most of the thrust will be applied within a few minutes after lift off, and in this time the inertial reference will not have had time to wander far from its correct position. The accuracy required of the inertial reference for this purpose will depend on the method of velocity control adopted, but will generally be quite low in systems where there is closed loop speed control.

2.4.2. Attitude control when in orbit

The requirement for attitude control when in orbit will depend on the satellite mission; the accuracy required can vary from 0·1 seconds of arc for an astronomical satellite down to, say 5 deg, required for keeping a low gain antenna directed at a receiving point on the earth. With the best available present-day components gyro wander rates are of the order of 0·01 deg/hour; thus an unmonitored platform will be in error by $\frac{1}{4}$ deg after one day, and $7\frac{1}{2}$ deg after a month. Since most satellites will require to operate for long periods to justify the high cost of launching them, the accuracy of an unmonitored platform will be inadequate for most purposes. Where an inertial system is used at all it is likely to be monitored by some other form of attitude sensor, and the justification for using inertia will be that it enables the other sensor to be simplified or that it improves the accuracy. If the other sensor is only capable of detecting small attitude errors, the inertial system will also have the function of pointing it near to the right direction.

If the vehicle is to be attitude stabilized, i.e. is to maintain a fixed direction in an inertial frame or a frame of reference which only moves slowly relative to an inertial frame (e.g. earth, moon or sun stabilization), then it may not be necessary to use a platform having freedom of movement relative to the satellite. The gyros could be fixed to the vehicle and the correcting torques applied directly to the vehicle. This represents a saving in complexity and weight, but it has the disadvantage that owing to the large changes of attitude during boost, the inertial system cannot be set up before take off, and cannot therefore provide a reference at the beginning of the orbit. This may or may not be serious depending on the nature of the monitor, and the speed with which it is required to stabilize the vehicle. On the other hand, if a movable platform is used, an error is introduced due to the need to measure platform attitude relative to the vehicle. For some applications, the best solution may be to use a movable platform which is locked to the vehicle at the beginning of orbiting. If the vehicle requires to have a known but varying attitude relative to an inertial, or nearly inertial frame, then a movable platform must be used, unless the accuracy requirements are low.

2.4.3. Earth stabilized vehicles

Many satellites will need to be stabilized in a frame of reference in which one axis points at the earth's centre (i.e. vertically downwards), another is horizontal in the orbital plane, and the third is perpendicular to the orbital plane. For such vehicles it will be simplest to use attitude sensors which are operated by energy received from the earth. Examples of such devices are horizon scanners which can work on visible light or at infra-red, and instruments which find the centre of the earth's disc by comparing the energy, received in pairs of light or infra-red sensitive elements which are arranged to receive energy in a ratio dependent on attitude. We may note first of all that any devices working on visible light are likely to give very different indications depending on whether the earth's surface visible from the vehicle is entirely, partly, or not at all, in daylight. Such devices could clearly be simplified if they need only operate in one of these conditions, for example full daylight.

This would be made possible by use of an inertial system whose drifts would be corrected, by the earth sensor in the full daylight condition only, but which would provide a continuous attitude reference. This consideration is mainly applicable to vehicles operating at distances from the earth which are too great to allow infra-red to be used. With either light or infra-red, however, there will be errors at any instant, since variations of weather and geography will result in non-uniformities in the emission of energy from the earth. For a vehicle which is continually changing its position relative to the earth such errors will be random with little or no fixed bias. They can therefore be smoothed by the use of an inertial system, with a consequent improvement in the accuracy of the reference. The sensors just described will define an axis through the earth's centre. To define fully the frame of reference it is also necessary to determine the direction of one of the other two axes. The axis at right angles to the orbital plane is a space fixed axis and can therefore be defined by a star tracker, which will need to be put on initially by reference to the stable platform.

2.4.4. Sun stabilized vehicles

Similar devices could be used in a vehicle which is stabilized to have one axis pointed at the sun. There is also the possibility of a particularly simple reference in which the difference in temperature between two identical, insulated, patches on the satellite's surface is used as a measure of the difference in the angle of incidence of sunlight on them. Such a device will have a long lag due to the time required to attain thermal equilibrium after a change of attitude. It could not therefore be used on its own as an attitude reference, but it could be used as a monitor to an inertial system which would give the necessary short term response.

2.4.5. Space stabilized vehicles

If the vehicle is to be stabilized in an inertial frame, the best attitude reference will be the direction of known stars. A star tracker will need to be initially pointed at the chosen star, and this can be done by using an inertial system, which will provide an accurate reference at the beginning of the orbit. Subsequently the star tracker can monitor the inertial system, which can then provide a reference if it is desired to change the attitude of the vehicle and point the star tracker at a different star. The arrangement might be useful in a satellite for astronomical observation in which it is desired to point a telescope at a number of stars in sequence. In this case it should be noted that a single star tracker will provide attitude information about two axes only. To provide full attitude information at any instant two trackers would be necessary. If there is an inertial system this can be monitored about two axes by a single tracker and left unmonitored about the third axis. When the star tracker is transferred to a different star there will now be attitude information about two different axes and the accumulated errors about the previously unmonitored axis can be corrected. Thus if the star tracker is not required to follow a single star for too long the use of an inertial system enables the second tracker to be eliminated.

2.4.6. Antenna control

Where the vehicle has a directional antenna which must be pointed at a ground station, then an attitude reference can be obtained by using a static split or conical scan system in the antenna which will establish the direction of the ground station. However, there will usually be periods in which the ground station will be obscured, and there may be a need to direct the antenna at other stations during these periods. An inertial system in the satellite can be monitored by the radio measurement when the antenna is receiving signals from a ground station, and can provide information for pointing the antenna when the ground station first comes into view.

3. Inertial Navigation

3.1. *The Accelerometer*

An accelerometer consists of a mass mounted so as to have one degree of freedom of translational motion relative to the instrument case. A measurable force is applied to the mass in the direction in which it is free to move so as to maintain the mass in a nearly fixed position relative to the outer case of the instrument. In a simple instrument, this force could be supplied by a stiff spring and measured by the small displacement of the mass relative to the outer case. Thus, the instrument measures acceleration along one axis. It should be noted that the accelerometer is also sensitive to gravity, and therefore if the instrument is used in a gravitational field, it will measure the difference between the acceleration along its axis and the component of gravity along its axis. Thus, if the accelerometer is mounted in a body in free fall, its measurement will be zero.

3.2. *The Inertial Platform*

If three accelerometers are mounted with their sensitive axes pointing along the three axes of a stable platform, they will measure the components of acceleration of the platform and therefore of the vehicle in which it is mounted in platform axes. If the platform is inertially stabilized then (if allowance is made for gravity) integration of the accelerometer outputs will give the vehicle velocity, and a second integration gives position. If the platform is not inertially stabilized, then corrections for platform rotation derived from the rate gyros must be computed and added to the accelerometer outputs before integration. Just as the platform attitude will lose accuracy with time owing to the integration of small errors in the rate gyros, so the measurements of velocity and position will lose accuracy due to the integration of accelerometer errors. Again similarly to attitude monitoring, the inertial measurement of velocity and position can be monitored by other measurements of velocity and position, which may be intermittent or noisy, to give an accuracy greater than that due to either measurement by itself.

Unlike the rate gyros, the accelerometers are not measuring a quantity which is nearly zero. Thus, errors which are proportional to acceleration (i.e. scale factor errors) are significant and under conditions of high acceleration greater errors will accumulate in the direction in which the acceleration is applied. These scale factor errors cause errors in velocity and position in a given interval of time which are proportional to the changes in velocity and position relative to a free fall trajectory (i.e. a condition in which the accelerometer readings are zero). There are also zero errors which cause accumulated errors in velocity and position proportional to time squared. Attitude errors of the platform will also contribute to the navigational errors since they will cause speed and position errors proportional to the changes of speed and position in a similar manner to accelerometer scale factor errors.

3.3. *Inertial Navigation for Space Vehicles*

When a space vehicle is in a free fall orbit its future path is almost entirely determined by its present position and velocity. The trajectories of present space vehicles consist of long periods of free fall orbiting and short periods in which accelerations are applied; the simplest instance is the earth satellite which after a short period of boost, enters an indefinite period of free fall orbiting. The navigational problem is therefore to determine the velocity and position of the vehicle in the course of, and particularly towards the end of, each period of acceleration with enough accuracy to apply the necessary thrust control. The accuracy required in each dimension will depend on the type of orbit and the use to which the vehicle is to be put, and so it will be best to discuss a particular example.

3.4. *The Station-keeping Earth Satellite*

We assume a satellite is to be placed in a circular orbit, whose orbiting period is to be controlled to a high degree of accuracy.

The plane of the orbit does not need to be so accurately defined, nor are small departures from a circular orbit very important. The orbital period is a function of the energy of the satellite, which must therefore be very closely controlled. From the well-known orbital equations we can deduce that the proportional error in orbital period, $\delta T/T$ is given by

$$\frac{\delta T}{T} = \frac{3\delta V_x}{V_x} + \frac{3\delta Z}{Z}$$

where $\frac{\delta V_x}{V_x}$ and $\frac{\delta Z}{Z}$ are the proportionate errors in forward velocity and distance from the earth's centre. The errors in downwards and sideways velocity δV_y and δV_z (both of which should be zero), do not contribute to the error in orbital time unless they are of an order of magnitude greater than δV_x. Assuming inertial control from take off, and accelerometer errors of $10^{-5}g$ (zero error) and 0·01%, which are about the best that can be expected from present day instruments, then the scale factor error contributes 0·01% of V_x to δV_x and 0·01% of $(Z-R_0+\frac{1}{2}gt^2)$ (i.e. the height climbed relative to a body in free fall starting at the same point on the earth's surface) to δZ (R_0 is the radius of the earth and t the boost time). Thus there is an error of more than 0·03% in orbital time using inertial control. This will accumulate an error of about 500 deg per year for a 2 hour orbit. This is clearly too great for the purpose. Thus inertial control by itself is inadequate.

If we assume that V_x can be established by radio measurement to 0·25 ft/s, and Z to 10 ft, then the contributions to $\delta T/T$ are 0·003% and 0·0006% respectively. The first of these represents an error of about 50 deg per year, which still seems excessive. If a higher accuracy is required, then it will be necessary to measure the time taken by the satellite to perform the number of orbits which it should perform in 1 day (assuming there should be a whole number of orbits in a day, the satellite would pass near its launching point on the earth's surface after a day; this would simplify the task of establishing its position). Following the measurement, correction is applied by giving the satellite an impulse corresponding to V_c the change in forward speed required to correct the orbital time. Should this change in forward speed be measured and controlled by reference to an inertial system, then there will be an error due to accelerometer zero error being integrated over the time in which the impulse is applied (the scale function error will be small if the impulse itself is small). If we assume that the impulse is provided by a constant thrust acting for a time t, and there is a possibility of an error δt in controlling the thrust time, then there will be a proportionate error $\delta t/t$ in the impulse. There will be another error $10^{-5}gt$ due to the accelerometer zero error. If t is chosen to equalize these errors, i.e. to minimize their resultant, then the resultant proportionate error in V_c is of the order of 0·025 $(\delta t/V_c)^{\frac{1}{2}}$ and the proportionate error in change of orbital time will be 0·075 $(\delta t/V_c)^{\frac{1}{2}}$. Thus as the accuracy becomes high and the correcting impulse becomes small, the proportionate accuracy with which the impulse can be applied is reduced through the combined effect of inaccuracy in controlling thrust time and the accelerometer zero error. Eventually there will come a point where there is no advantage in using inertial measurement, since the same proportionate accuracy of impulse can be obtained by the "open loop" method of burning a quantity of fuel which should give the required impulse. If $\delta t = 0{\cdot}1$ seconds (say) and $V_c = 0{\cdot}25$ ft/s (corresponding to radio measurement of speed during boost) then the proportionate error in change of orbital time will be 5%. If $V_c = 2{\cdot}5$ ft/s corresponding to inertial measurement of speed during boost, the proportionate error in change of orbital time will be 1·5%. In the two cases the remaining error after the correcting impulse is 0·00015% (or 2·5 deg/year) and 0·00045% or 7·5 deg/year in the orbital period. Thus unless the "open loop" method is accurate to better than 1·5% the use of the inertial system seems justified for the other values assumed. It should be remembered that even if inertial navigation is not used, some form of attitude reference will be required to ensure that the impulse is applied in the right direction. However, the proportionate error in change of orbital time due to an attitude error θ can be easily shown to be θ^2. Thus for a proportionate error of 1%, θ can be as high as 6 deg. This low accuracy should be relatively easy to obtain with or without an inertial platform.

Thus a more accurate control of orbital time can be obtained by using radio measurements than by using an inertial system by itself for controlling the velocity during the boost period, though in both cases a measurement of orbital time after boost followed by a correcting impulse is required. We now enquire into whether there is any advantage in using both radio measurements and an inertial system in the boost period.

(a) Radio measurement of velocity to the required accuracy is only possible if Doppler methods are used, which means that any one ground station can only measure the component of velocity along the line joining the ground station to the satellite. To measure the resultant satellite velocity it is therefore necessary to use three ground stations which should be placed such that the lines joining each to the satellite are as nearly at right angles to each other as possible. Furthermore it is necessary to establish satellite position so that the directions of the three measured components of satellite velocity are known. Satellite height must be known in any case, for satellite energy to be known. Thus three range measurements are also needed. This implies that three range measurements are also needed. This implies that three further ground stations must be provided, or else the three velocity measuring stations must also be capable of measuring range.

(b) For control of velocity perpendicular to the orbital plane (i.e. keeping it zero) inertial accuracy is quite adequate. Since there should be no acceleration in this direction the only significant error is the accelerometer zero error of, say, $10^{-5}g$. This will build up a velocity error at the rate of 0·02 ft/s per minute of boost time. Thus for any likely boost period this velocity error, which only results in a slight tilting of the orbital plane, will be quite insignificant. The positional error in this direction will be 1 foot after 1 minute and 25 feet after 5 minutes. Such an error will not contribute significantly to the error in radio measurement of satellite forward velocity and height. Thus the use of inertial measurement perpendicular to the orbital plane enables one ground station for measurement of velocity and position to be eliminated. For maximum accuracy the remaining two stations should lie in the orbital plane, but in practice it may not always be possible to arrange this.

(c) It may not always be possible to place the ground stations so that accurate radio measurements of speed and height can be made at the end of the boost period. For example, if the satellite is at a height of 1000 miles the ground stations should be 2000 miles apart for maximum accuracy of measurement (Fig. 3). There may be geographical or political difficulties in achieving such a separation. A possible solution to this difficulty is to inject the satellites into an elliptic orbit chosen such that its apogee is at the same height as the desired circular orbit, and to apply an extra boost at the apogee to put the satellite into the circular orbit (Fig. 4). In this case the satellite will be at a lower height at entry to the elliptic orbit and therefore the ground stations can be nearer together (and also will not need so much power). The change of speed applied at the apogee can be small enough to be measured inertially. For example, if the satellite is at 250 miles height at entry to the elliptic orbit whose apogee height is 1000 miles, and the point of entry is at 90% from the apogee then its speed at the apogee will be 7·5% lower than that of a circular orbit at a height of 1000 miles. Thus the error due to measuring the impulse inertially will be 0·01% of 7·5% of the speed for the circular orbit. This results in an error of 0·0023% in orbital period, which is of the same order as the error due to an 0·25 ft/s error in a radio measurement of speed.

(d) To measure a Doppler frequency to a high accuracy it is necessary to count a large number of cycles. This inevitably introduces a lag into the velocity measurement. This lag makes it difficult to achieve precise velocity control particularly at high accelerations, and therefore special measures must be taken to keep the counting period as short as possible. With an inertial system, which is monitored by the Doppler measurements, up-to-date velocity is

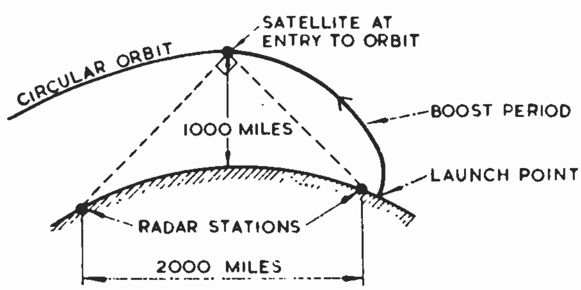


Fig. 3. Radar measurement at entry to orbit.

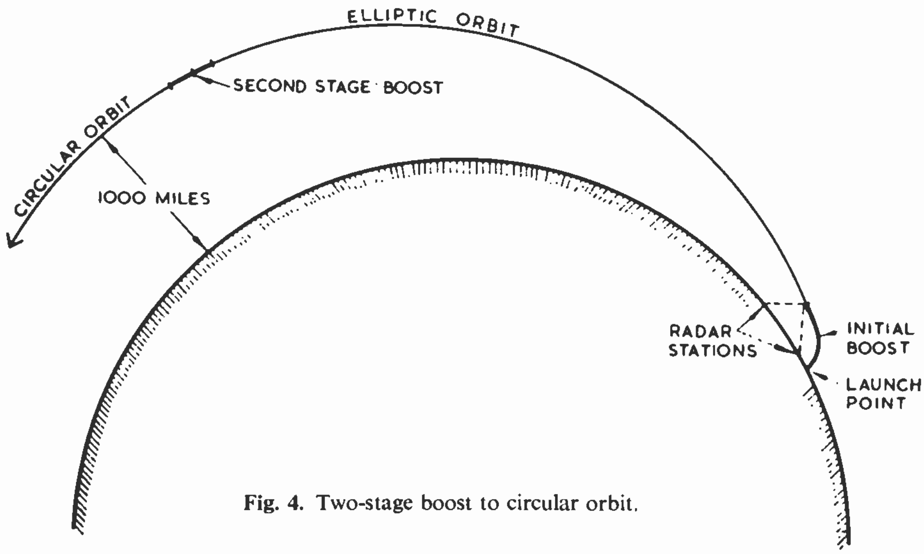


Fig. 4. Two-stage boost to circular orbit.

obtained, and there is no need for the counting period to be particularly short. Assuming a maximum acceleration of $10g$ and an accelerometer scale factor error of 0·01%, the maximum lag which can be tolerated on the Doppler count for an accuracy of 0·25 ft/s is 8 seconds. If, on the other hand, we assume that precise velocity control will not be attempted until the acceleration has been reduced to, say, $2g$, then a lag of as much as 40 seconds can be allowed.

(e) We have already noted that the position of the satellite must be established by radio measurement in order to measure its velocity accurately. It would be a considerable simplification of the ground stations if the Doppler measurement could be abandoned and velocity obtained by differentiation of position. This velocity would certainly need to be smoothed and this could be done by using an inertial system. Suppose the positional measurement has a random error of 10 feet, and that in a final phase of precise velocity control the acceleration is limited to $2g$, and a smoothing time constant of 30 seconds is used. Then the error due to the positional measurement is 0·33 ft/s and the steady state error due to the accelerometer scale factor is 0·19 ft/s. In addition there will be the error accumulated in the previous phase of high acceleration (say $10g$) which will be reduced exponentially with a time factor of 30 seconds. This error will be about 0·6 ft/s at the beginning of the low acceleration period and will be reduced to 0·2 ft/s after about 30 seconds. Thus it should be possible to control velocity using range measuring radars only to an accuracy of about half of that obtained using range and velocity measuring radars. Quite modest improvements in the accuracy of the range measuring radar, or of the inertial system, over the figures quoted here could make this system competitive in accuracy with the Doppler system.

We may summarize this discussion of the station keeping communications satellite as follows:

To obtain an accuracy of orbital time of about 0·00015% the boost should take place in three phases, in each of which there is a use for an inertial system.

(1) The satellite is accelerated into an elliptic orbit. Its speed at entry into this orbit is controlled to an accuracy of 0·25 ft/s using radio measurements. The inertial system can be used:

(a) To control the velocity out of the orbital plane

(b) To simplify the Doppler radar design by allowing longer counting times

(c) Possibly to enable the range and velocity measuring radars to be replaced by range only radars.

(2) At the apogee of the elliptic orbit (i.e. about ¼ of the orbit from the launch point) the satellite is accelerated into a circular orbit. Speed is controlled to an accuracy of 0·25 ft/s, using an inertial system, or alternatively a different ground installation from that used in the initial boost.

(3) The satellite remains in orbit for one day, in which an accurate measurement of orbital time is made from the ground. After this a final correcting impulse is applied which must be controlled to 3% either by the inertial system, or by the "open loop" method.

Finally the inertial system may be used in the control of attitude, once the satellite is in orbit.

We conclude that it has not been found that an inertial system is essential; nevertheless it can make a contribution to the system in several different ways, enabling other associated equipments to be simplified, and therefore it may well be profitable to include it.

4. Advantages and Disadvantages of Inertial Navigation

The example of the station keeping satellite brings out the advantages and disadvantages of inertial navigation. It cannot control large changes of velocity to a high accuracy because of scale factor errors; neither can it control very small impulses to a high proportionate accuracy because of zero errors. Nevertheless it has the advantage that its operation is quite independent of vehicle position. Thus it is useful in all cases where the whole of the boost trajectory is not suitably placed for measurements from the ground. This advantage will become increasingly important as the range of space vehicles increases, and therefore the difficulty of measurement from the earth increases also. At very great distances from the earth, navigation is likely to consist of positional determination by observation of the relative directions of celestial bodies, combined with inertial measurement of impulses applied to the vehicle to correct its velocity.

The other advantage of using an inertial system is that it can be combined with other methods of measurements, so that the requirements on the other methods can be relaxed. It is no longer necessary for the determination of velocity and position to be continuous and up to date, and in some cases this could mean a considerable simplification of the equipment required.

An important consequence of using an inertial system is that the navigational computation may have to be done in the vehicle rather than on the ground. When the vehicle velocity and position is entirely measured from ground stations, then it is reasonable to do the computation on the ground and merely send commands to alter thrust up to the vehicle. When there is an inertial system in the vehicle, however, the navigational computer must work on data some of which is obtained on the ground and some in the vehicle. Because of power restrictions in the vehicle it is easier to send data from the ground to the vehicle than vice-versa. Thus it may be advisable to put the computer in the vehicle. For earth satellites this is a disadvantage because of the size, weight, and power consumption of the computer, but for long range vehicles it is probably essential in any case.

5. Inertial Components

In recent years there has been much development of small inertial components of high accuracy in the United States. These have been developed for missile, aircraft, and submarine application, but the requirements of high accuracy, light weight, and long life apply to the space vehicle field.

5.1. *Gyroscopes*

To reduce the wander rate to a minimum, stray torques on the gyro must be made as small as possible. In precision instruments this is done by floating the gyro. The gyro rotor spins in a container which is immersed in and supported by a liquid. This container, known as the floated element, rests on bearings attached to the outer case of the instrument. The floated element is designed to have an average density equal to the density of the liquid, so that there is practically no load on the bearings and frictional torques are reduced to a minimum. If the gyro is moving relative to the outer case the liquid exerts a viscous torque proportional to angular velocity of the floated element relative to the outer case. This torque must also be proportional to the angular velocity relative to an inertial frame, by the fundamental property of gyroscopes. Thus the angular displacement of the floated element as measured by an electrical pick-off is proportional to the integrated angular velocity of the instrument, relative to an inertial frame. The instrument is therefore an integrating rate gyro. A miniature instrument of this type can be 2 in. in diameter by 3 in. long, and weigh under 1 lb. The random drift will be of the order of 0·01 deg/hour. There will also be drifts of greater magnitude than this due to thermal effects and to mass unbalance, but these need not affect the accuracy of a platform. The thermal drifts can be avoided by running the gyro for a sufficiently long period before use, so that thermal equilibrium is achieved, and the mass unbalance drifts, which are acceleration sensitive, can be measured in the laboratory and corrected by signals from the platform accelerometer.

5.2. *Accelerometers*

Precision accelerometers also use an electrical pick-off whose output is fed to an electromagnetic coil system to restore the pick-off output to zero. The current in the restoring coil is then proportional to acceleration. Such an instrument can measure 2 in. × 1·4 in. × 1 in., and weigh 4 oz. It will have a zero error of the order of $10^{-5}g$ and a scale factor and linearity error of about 0·01%.

5.3. *The Platform*

An inertial platform using components of this sort, and having 4 gimbals to allow full vehicle attitude freedom could measure about 10 in. × 10 in. × 8 in. and weigh about 15 lb. Under steady running conditions, the power consumption would be about 25 watts. On initial erection peak powers up to 200 watts would be required. For many space applications full attitude freedom would not be required, this would enable the number of gimbals to be reduced from 4 to 3, which would effect some saving in size, weight, and power consumption.

6. Conclusions

We may conclude that an inertial system can be used as an aid to the attitude control and navigation of a space vehicle. It will not often be capable of performing these functions by itself, but if used in conjunction with other instruments, it can allow the design of the other instruments to be simplified. Thus the question of whether an inertial system is justified for a particular space vehicle can only be answered by an overall assessment of the control system including the other instruments and general tracking facilities. This task has not been attempted in this paper, but we may say, in a general way, that the relative advantage of using an inertial system will increase as the sizes of space vehicles and the distances they travel increase.

7. Acknowledgments

The author wishes to thank Messrs. Ferranti Limited for permission to publish this paper.

8. Bibliography

1. C. S. Draper, W. Wrigley and J. Horaka, "Inertial Navigation (Pergamon Press, London, 1960).
2. D. E. Adams, "Inertial methods in hybrid navigation systems", *Journal of the Institute of Navigation*, **13**, p. 61, January 1960.
3. R. P. G. Collinson, "Inertial navigation 1960", *Flight*, 19th February 1960.
4. W. Cawood, "Some design problems in inertia navigation" *Journal of the Royal Aeronautical Society*, **62**, p. 704, October 1958.
5. C. F. O'Donnell, "Inertial navigation" (Parts I and II), *Journal of the Franklin Institute*, **266**, pp. 257–78, 373–402, October and November 1958.
6. P. Schuerb, "Introduction à la Navigation par la Inertie," *Navigation (Paris)*, No. 15/16, 1956. (M.O.S. Translation Ref. TIL/T 4761.)

Manuscript received by the Institution on 14th June 1961. (Paper No. 667.)

The G.P.O. and Satellites for Communications

The British Post Office, in studying problems associated with communications satellite systems, is taking the following features as desirable targets:

(i) Continuous 24-hour service should be available for telegraphy and telephony between all ground stations provided in connection with the system.

(ii) World-wide coverage should be provided in association with the existing cable and radio links.

(iii) Provision should be made for the satellites to carry about 1000 telephone channels and one or two television links.

(iv) The system should be capable of carrying the traffic in the minimum number of hops and the bulk of the traffic in one, i.e. by bringing into use only one satellite for the operation.

Much further study will be required before it is known whether these targets can be achieved. In general however the number of satellites required in a system would depend on the type of orbit adopted and its height, on whether the satellites are position-controlled in this orbit, and to some extent on the number and location of the ground stations. For example, with position-controlled satellites in an equatorial orbit a system consisting of 12 satellites would need to be put into orbit at a height of between 7000 and 8000 nautical miles, whereas for a system with 16 satellites an orbital height of between 5000 and 6000 nautical miles might suffice.

The launching of satellites for communications purposes is one of several important uses of a launcher based on the *Blue Streak* rocket, developed in the way envisaged by the European Launcher Development Organization. The technical and other problems associated with the development of a commercial satellite communications system will clearly need full discussion and co-operation with European and Commonwealth countries as also with the United States. Some of the *international* planning aspects which will also be involved are discussed in the Convention paper by Capt. C. F. Booth elsewhere in this *Journal*.

In the White Paper on Post Office Prospects 1961/62 (Cmnd. 1327) reference was made to collaboration between the Post Office and the United States National Aeronautics and Space Administration in tests of satellite communications across the Atlantic. The Post Office is providing an experimental ground radio station on the Lizard, Cornwall, in preparation for experimental tests of satellite communications next year. These tests which will cover the transmission of speech, telegraphy, and television across the Atlantic, were referred to in the *Journal* for May 1961 (page 408).

Many technical, operational and economic questions will require to be studied before a commercial satellite communications system can be established. The information and experience that will be gained from the North Atlantic tests in 1962 will, however, be of great value in future planning.

DISCUSSION

on

"The Change-of-State Crystal Oven"*

Mr. A. Russell†: I was very interested in Mr. D. J. Fewings' paper on the change-of-state crystal oven as I was unaware that any work was being done on this in Britain. In 1957 several attempts were made at the National Engineering Laboratory to find a suitable filler for an operating temperature around 35° C.‡

The first substance tried was sodium sulphate (Glauber salts), and this showed good latent heat of fusion properties but very little expansion/contraction change at the melting point. Unfortunately, the salts become unstable after a few melt/freeze cycles.

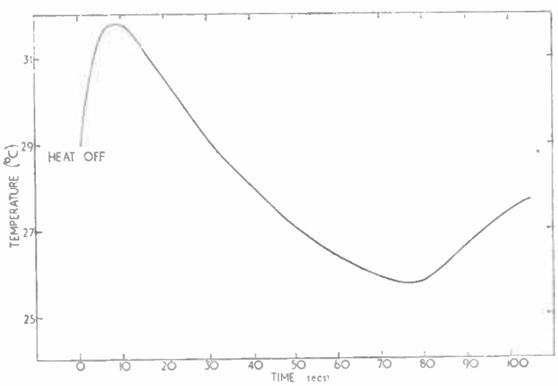


Fig. A. Cooling curve for cinnamyl alcohol.

Initial tests with cinnamyl alcohol were very encouraging and showed (Fig. A) remarkable latent heat of fusion properties; the cooling curve clearly shows this at the freeze point of 26° C. Expansion though was practically linear over the range and unsuitable for operating a micro-switch at the melting point.

2-methyl naphthalene was the most promising if care was taken to avoid contamination from the container and all air removed. The container was coated with silicone varnish and the mass cooled from the bottom upwards to remove all air. Figure B is a typical heating curve and shows how the temperature levels out at the m.p. (36° C) while the chemical gives rise to a greatly increased expansion rate; the section within the dotted lines represents an expansion change of 0·025 inch for very little change in observed temperature.

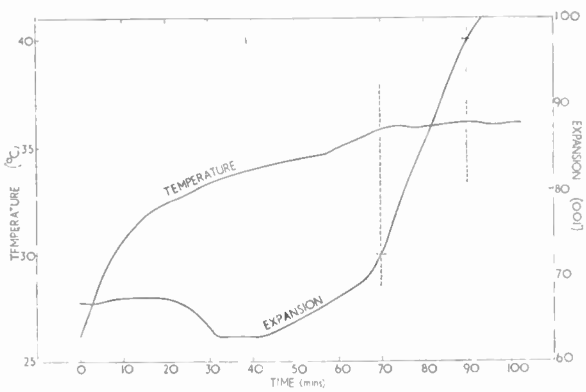


Fig. B. Heating curve for 2-methyl naphthalene.

Mr. D. J. Fewings§ (*in reply*): The use of substances other than naphthelane for obtaining temperatures lower than 80° C has of course interested me since crystals of differing types have different temperatures of minimum temperature coefficient. There is no lack of suitable substances having melting points in the region 70-80° C. Apart from naphthalene there is decane 1 : 10-diol, m.p. 73° C, and 1 : 2 : 4 : 5 tetramethyl benzene, m.p. 78° C. These two substances expand by 12% and 24% respectively on melting and would be eminently suitable for a bellows controlled oven.

At lower temperatures, however, especially in the region of 50° C, which would be particularly suitable for some widely used crystals, the expansion rate of the substances tested so far have been disappointing. Three examples are benzophenone m.p. 47° C, thymol m.p. 49° C, and camphene m.p. 49° C. Of these camphene has the largest expansion but this only amounts to 2·5%. We are still hoping to find something more suitable.

* *J. Brit.I.R.E.*, **21**, pp. 137-42, February 1961.

† Metrology Division, Department of Scientific and Industrial Research, National Engineering Laboratory, East Kilbride, Glasgow. *Letter received by the Institution on 27th April 1961.*

‡ Previously, work on this subject was done at a much higher temperature. See for instance, "Fusion heat stabilizes crystal", *Electronics*, **29**, No. 5, p. 208, May 1956; "Precision temperature control using m.p. of chemical", *Product Engineering*, **27**, No. 3, p. 206, March 1956; "A precision crystal oven", McFarlane and Metz, *Proc. Nat. Electronics Conf.*, **12**, 1956.

§ Research Division, Marconi's Wireless Telegraph Company Limited, Chelmsford, Essex.

U.D.C. 539.16: 629.19

Radiation and other Environmental Effects on Satellites

By

R. INNES, B.Sc., Ph.D.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961

Summary: The radiation environments encountered by a satellite according to its orbit are summarized. The breakdown of high-energy radiation on meeting matter is described, indicating the mechanisms and nature of damage caused to equipment. An argument is presented to show that screening against hard radiation is inefficient and uneconomical. Estimates are given of the life of a transistor—the most susceptible to damage of normal electronic components—in a satellite traversing the lower Van Allen belt. Means of protection against the erosive low energy radiation are discussed and requirements to meet launching conditions are mentioned.

1. Introduction

It is generally appreciated that once a satellite is launched little can be done in the way of equipment maintenance, hence it is essential that all apparatus in a satellite be of the highest standard of reliability possible. In a complicated electronic system, equipment may be duplicated and changeovers arranged, but the overall life is still a function of the life-span of each portion; and of course a chain is no stronger than its weakest link. Much has been done to develop highly reliable equipment for operation in inaccessible areas, for example in trans-oceanic cables, but in a satellite we have additional environmental hazards to contend with. A physical description of the nature of these extra hazards is given here to assist in designing reliable equipment for use in satellites.

Satellite orbits in general will lie in the exosphere which we may loosely define as the transition region between the Earth's atmosphere and inter-planetary space. Unfortunately the precise composition of the exosphere with regard to the number of neutral atoms, the number and energy of ionized gas atoms, and the number and energies of free electrons, seems in some doubt; but as the analyses of further satellite experiments proceed, a fuller picture will emerge. It is believed that the exosphere consists of a high temperature gas with a proportion of high energy ($\sim$ 50 eV) particles of direct solar origin—the solar wind. The pressure at an altitude of 500 km[1] is about 3×10^{-8} mm Hg with a gas particle velocity equivalent to a temperature of 2000° K, corresponding to about $1{\cdot}4 \times 10^{7}$ atoms/cm^3. Lad[2] gives the number of solar wind ions as 10^3/cm^3 with an average energy of 50 eV in normal conditions, but the number and particle energy may each increase a hundredfold following a solar storm. All satellites will be exposed to the full solar spectrum, except of course when in the shadow of the Earth or Moon. The Sun, a black body radiator, pours out energy giving a flux of 2 cal/cm^2/min at a distance of one Astronomical Unit. About 1% of this is composed[1] of photons with energies greater than 4 eV; few of these have energy greater than 6 eV. An earth satellite will therefore receive about $1{\cdot}4 \times 10^{14}$ photons/cm^2/s with $E > 4$ eV. Most of the solar energy is in the region of the visible spectrum and is therefore useful for energizing silicon solar cells. The infra-red radiation only produces heat.

The Earth with its magnetic field acts as a huge cyclotron maintaining high energy charged particles in circular areas roughly in the equatorial plane. The first and most important region is the inner Van Allen belt[3]; this is a torus lying in the equatorial plane around the earth, with its centre at an altitude of 3 600 km and extending North and South to latitudes of 20 degrees. Van Allen summarized the radiation at the centre of this lower belt as follows:

(a) Electrons of $E > 20$ keV, maximum unidirectional intensity $\sim 2 \times 10^9$/cm^2/s/sterad.

(b) Electrons of $E > 600$ keV, maximum unidirectional intensity $\sim 1 \times 10^7$/cm^2/s/sterad.

(c) Protons with $E > 40$ MeV, omnidirectional intensity $\sim 2 \times 10^4$/cm^2/s.

Further out is a second zone with a large number of electrons but few high energy protons. Van Allen's summary of the radiation at the centre of this zone at an altitude of 16 000 km is:

(a) Electrons of $E > 20$ keV, omnidirectional intensity $\sim 1 \times 10^{11}$/cm^2/s.

(b) Electrons of $E > 200$ keV, omnidirectional intensity $\sim 1 \times 10^8$/cm^2/s.

(c) Protons with $E > 60$ MeV, omnidirectional intensity $\lesssim 1 \times 10^2$/cm^2/s.

† Marconi's Research Laboratories, Great Baddow, Essex.

(d) Protons with $E<30$ MeV, no significant information.

Arnoldy, Hoffman and Winckler[4] found that the outer region was not stable with time; a solar storm appeared to break up this zone, some electrons being transferred to constitute a third band half way between the inner and outer regions.

From the environmental hazard aspect, three distinct types of orbit will be recognized. In the first category are those orbits with apogees of less than 1100 km altitude: such orbits, if above 500 km altitude, are in a gas density low enough not to seriously impede the flight of the satellite. The second category of orbit is that leading to the satellite spending appreciable time in the Van Allen radiation belts. Here the atmospheric density is decreasing to a level beyond that obtainable in a vacuum laboratory; but there are high energy electrons and protons which constitute—certainly in the inner belt—a serious radiation hazard. The third category of orbit is that outside the radiation belts, i.e. above about 60 000 km, where high energy radiation density is small, being truly cosmic in nature. Although the atmospheric density is very low—approaching that of inter-planetary space—conditions are still governed by the earth's magnetic field. The geostatic satellite orbit lies in the outer edge of the second region. The environment surrounding space probes will, in general, be in the third category.

To assist in asssessing the effect of the various forms of radiation on a satellite, the breakdown of high energy radiation on meeting matter will be described first, indicating the nature of the damage which may be caused. The effect of this damage on satellite equipment and possible protective measures are discussed. The effect of the low energy solar wind and e.m. radiation are then considered.

2. The Radiation Spectrum and its Interaction with Solid Matter

The radiations of interest to us here cover a very wide range of energies. Topping the high end are the truly cosmic particles with energies up to 10^7 MeV, then there are the Van Allen particles with energies from 700 MeV down to about 20 keV, lower still are the solar wind ions with energies in the region of 100 eV, then there is the direct solar radiation concentrated around the 2 eV level and the infra-red which at a wavelength of 50 μm has an energy of only 0·025 eV. Radiation may take one of three forms:

(1) a positively or negatively charged particle,

(2) a neutral particle with high velocity, and

(3) quanta of electromagnetic waves which in the high energy region are called γ-rays.

All these radiations are in definite packets of energy and it is convenient to call each packet a "particle" whether we think of it as having real mass or whether it is a quantum of e.m. radiation.

2.1. *The High Energy Region*

Radiations having energies in excess of 100 MeV are quite capable, on colliding with an atomic nucleus, of shattering the nucleus. The nuclear debris from this miniature explosion may consist of charged fission fragments of various masses, neutrons and γ-rays; these particles will share the energy carried in by the incident particle. The highest energy Van Allen particles, being protons, are capable of dissipating some of their energy by ionization, i.e. they scatter electrons from their path, but the nuclear fission process dissipates their energy much more drastically. Protons, electrons, neutrons, α-particles and γ-rays in the energy range of 1 to 100 MeV are capable of causing nuclear disintegration, but as we descend the energy scale, scattering mechanisms become the main process of energy dissipation and almost the sole process below 1 MeV.

Many of the high energy interactions lead to the emission of γ-rays; there are three interesting scattering mechanisms for these rays. The first is pair production wherein a γ-ray of energy greater than 1·02 MeV is converted to a pair of electrons, one with the normal negative charge and the other with an equal positive charge. 1·02 MeV of the incident γ-ray's energy is required to provide the mass of the electrons and any extra energy appears as the kinetic energy of the electrons. The positron, however, will sooner or later meet an electron and the "annihilation" will produce a γ-ray of energy 1·02 MeV or greater. This change and change back again is not, however, without effect as there are two important results. The first is that this mechanism allows γ-rays to dissipate energy in a smaller volume of matter since electrons, being charged particles, cause much more dense ionization than the original γ-ray. The second effect is that in order to conserve momentum when pair production takes place, it must do so in the close vicinity of another particle, e.g. an atomic nucleus. This nucleus may receive sufficient energy to cause its displacement in the crystal lattice.

The second and third mechanisms are Compton scattering and photoelectric emission. In the former a γ-ray collides with an electron and imparts some of its energy to it—the majority of electrons are given a direction similar to that of the incident photon. In the photoelectric effect virtually all the incident photon's energy is transferred to the electron. Again to conserve momentum, the atom from which the electron was ejected may be given sufficient energy to displace it from its lattice position.

2.2. *The X-Ray Region*

Around the nucleus of a normal atom is a number of layers of electrons in orbit. The first layer containing only two electrons is called the K shell, the next layer, which can contain up to 8 electrons, is the L shell, then the M shell and for the heavier atoms the N shell. In stable conditions these shells are filled up in order except for the second outer shell in the paramagnetic elements. If an electron is removed from an atom, e.g. a K electron, the atom is raised to a metastable state of higher energy. This position is only temporary as an electron from an outer shell will fall in towards the nucleus and in so doing will emit a quantum of e.m. energy $h\nu$. This electron may come from one of the orbital shells or from outside the atom and the frequency of the resultant radiation depends on which shell the electron comes from. If an electron falls in from the L shell it will leave a gap there, and an electron from the M shell—assuming the atom has one—will fall in from M to L again emitting a small quantum of radiation. The maximum energy for any one transition is when an electron falls from outside into the K shell and in this case it will emit the characteristic K radiation for that element. This maximum energy—which is in the region of 10 keV for elements in the middle of the periodic table—will be that required to remove a K electron from an atom. Transitions involving the intermediate shells give rise to the soft X-ray spectra.

2.3. *The Visible Spectrum Region*

The adjustments in the outer electron shells involve low energy levels of the order of a few electron volts and this corresponds to radiation in the vicinity of the visible spectrum: the definite energy quanta involved in transitions between levels give rise to the well known spectral lines.

In any normal compound two or more atoms combine in such a way that the total number of electrons in the outer shells is sufficient and only sufficient to complete these shells. In ionic compounds the spare electron or electrons are transferred to the other atom which is short of that number. When the atoms of a compound contribute a few electrons each to complete a shell the chemical bond is covalent wherein the outer electron shell is shared between the ions of the molecule. In all compounds there is a mixture of both types of bond but salts are mainly ionic and organic compounds mainly covalent.

Many chemical compounds on forming give out heat; this is because the outer electrons of the constituent atoms on completing an electron shell arrangement have a lower energy state than the separate atoms; the energy difference is liberated as heat. There are a large number of energy levels associated with chemical composition and with the forming of crystal structure, most solids being crystalline in nature. The energy levels associated with chemical combination cover the region from the ultra-violet to the near infra-red while the small energy differences associated with fine crystal structure give rise to the infra-red spectrum of the material. In the far infra-red photons and phonons are freely interchangeable in the crystal lattice and hence radiation may be converted to heat.

2.4. *Neutrons*

An important particle product of nuclear disintegration is the neutron. Its energy is purely kinetic, ranging in value from the few MeV region right down to thermal energies of 1/40 eV. Fast neutrons can cause nuclear fission but those with energy less than about 1 MeV normally dissipate their energy through scattering by straightforward mechanical collision. As neutrons are 1838 times heavier than electrons they carry more momentum for a given energy, hence a neutron of only several hundred eV energy can impart sufficient momentum to an atomic nucleus to displace the atom from its position in the crystal lattice. The probability of a high energy neutron scattering several atoms during its life is therefore quite high. When a neutron has dissipated its energy until it is in the thermal region the probability of the neutron being absorbed by a nucleus rises again steeply. The new isotope formed by adding a neutron to a nucleus is generally unstable and will suffer radioactive decay at some later time, so liberating considerably more energy than that carried in by the neutron. If the neutron is not captured by a nucleus within about 12 minutes on average it suffers spontaneous decay into a thermal energy proton and a β-ray of energy 782 keV.

2.5. *Displacements*

It is convenient here to elaborate on the mechanisms of atomic displacement. The energy binding an atom in its place in a crystal lattice is in the region of 25 eV and this much must be given to the atom to permit it to move from its position. For a neutral particle to displace an atom it must achieve direct collision, whereas a charged particle on closely approaching a nucleus can cause displacement through electrostatic attraction or repulsion. Similarly electrons may be scattered by particles and γ-rays. The 25 eV energy given to the atom must however be to the atom as a whole and therefore is ainmly due to the momentum of the nucleus. Energies under about 100 keV given to an orbital electron merely raise the atom to an electrical metastable state and contribute little to the atom's momentum. In the Appendix it is shown that an electron must be ejected with high energy—of the order of 1 MeV—to have sufficient momentum recoil to cause displacement of an atom. Electron recoil displacements are therefore effected only by high energy radiations, whereas neutrons of only moderate energy, fission fragments, and already displaced atoms of quite low energy can cause lattice displacement of atoms. The Varley[18] mechanism is exceptional in allowing atoms to be displaced on the rare occasions of their multiple ionization by radiation of several keV energy.

2.6. *Summary of Means of Dissipation of Radiation in Matter*

Very high energy particles cause incidental ionization, but also dissipate their energy by shattering nuclei, so producing an increased number of particles (the term particle includes γ-rays) with energies of several MeV. The passage of these particles through matter may cause the scattering of atomic nuclei both directly and by recoil from Compton scattering and from the photoelectric ejection of an electron from one of the shells around these nuclei. Further nuclear disintegration may arise from interaction of these secondary particles with matter causing further multiplication of energized particles but except in the unlikely event of fission reducing the average energy per particle. Below about 1 MeV particles are in general not capable of causing nuclear fission and so dissipate their energy by scattering mainly electrons and only a few nuclei. The scattered electrons having energies of several keV are capable of ionizing atoms by removing even the K shell electrons. When particle or photon energies are below the 10 keV region they are then incapable of deep ionization but can disturb the middle and outer electrons, there being an increase in the number of excited atoms. The deeper electron shells will quickly fill leaving increased ionization of the outer shells. The molecular bonding of compounds, which may of course be disturbed by any of the higher energy particles, may still be upset by radiation of a few electron volts energy. The ionization energy being now spread over a large number of atoms dissipates itself into the crystal structure energy, thereby being converted from photons to phonons producing an increase in temperature.

3. The Nature of Radiation Damage to Materials

The criterion of damage for the present purposes is the alteration of any property of the material to an extent which would interfere with its reliable use in a particular application. Thus a slight conductivity caused by radiation induced ionization in the base board of a printed circuit carrying several milliamperes may be of no significance, while the same conductivity in the target of a vidicon or similar tube might prove quite disastrous. Much radiation damage may be annealed out of substances in time; thus Sakiotis *et al.*[5] have shown that microwave ferrites return to very nearly their original condition in some six months after being irradiated in a nuclear pile. For the present application, however, any damage which lasts for more than a few hours will be considered permanent damage: atomic displacement damage falls into this category. Ionization will cause a form of damage which may be classified as temporary because it is self healing in a time varying from 10^{-12} seconds to minutes; the importance of ionization damage will depend on the rate of ionization as a steady state should be reached when the rate of repair just balances the rate of excitation. Permanent damage will occur when an atom is transformed from one element into another so introducing a chemical impurity: the number of such transmuted atoms will be comparatively small and their effects will be outweighed by other damage.

3.1. *Electrical Effects*

When an atom is knocked out of place in a crystal two defects are formed:

(1) a hole in the crystal lattice and

(2) an extra atom jammed into the lattice at another point—an interstitial atom or ion.

This pair of defects together form a Frenkel defect. At the atomic gap in the lattice there will be an excess of charge, its sign depending on whether the displaced atom formed an anion or a cation. If the nearest neighbours round the atomic gap have a surplus of positive charge the gap will act like a free positive ion and so may trap an electron into hydrogen type orbits round it: this arrangement is called an F centre. The interstitital atom will behave similarly. If the defects have excess negative charge an ordinary electron-like hole may still orbit them in the surrounding medium again giving rise to a hydrogen type spectral series. At both defects, however, the energy difference between the orbit levels is reduced by the dielectric constant of the medium surrounding the defect. Accompanying the decrease in energy differences, energies being inversely proportional to the square of the dielectric constant, the effective radii of the orbits are increased in the ratio of the dielectric constant.

When the concentration of these F centres is so high that there is strong overlapping of electron orbits we may have conductivity due to the drift of the orbital electrons. Before this high concentration is reached, however, these centres may form donors or acceptors so converting the material into a semiconductor. The energy gap between the top of the full band and the acceptor levels, or between the donor levels and the conduction band will depend on the parent material but will be less than the intrinsic energy gap for the original undamaged material. Also although the generation of F centres in itself may not give rise to excessive conductivity at room temperature it will reduce the energy required to raise electrons into the conduction band so that ionization induced conductivity (temporary damage) will be greater after an accumulation of displacement damage (permanent). F centres may diffuse slowly through a crystal lattice and if a vacancy and interstitial meet—in a material composed of one element—the lattice will be restored to normal: this is annealing. If F centres drift to crystal dislocation edges they may be trapped and modified; close pairs of defects form yet another type of centre. In a semi-conductor, e.g. silicon in a transistor, all these centres have the ability to scatter conduction electrons or holes and so reduce carrier mobility in the material.

3.2. *Magnetic Effects*

The elements iron, cobalt and nickel are magnetic because the N electron shell has started to fill before the underlying M shell has its full complement of 18 electrons, leaving the magnetic spins of the d electrons in the M shell not paired off. In the metallic state of these elements there exist within the crystal structure, magnetic domains wherein all the spins of the atoms are aligned. (The spin of an atom is the algebraic sum of its electron spins and is therefore that of the unpaired electrons.) As each electron spin has a magnetic moment, in a ferromagnetic material where the unpaired spins line up, the magnetic moments will add so producing the macroscopic magnetic moment. Ferrites, which are formed by reacting mixtures of the oxides of the ferromagnetic elements with other metal oxides, exhibit ferrimagnetism in which the spins of neighbouring atoms are antiparallel, but without equality between opposing moments, so that there is a non-zero resultant magnetic moment.

If the odd atom here and there is knocked out of its place in the crystal lattice there will be a lack of spin and hence magnetic moment at the gap. In ferromagnetics this will cause a slight decrease in the total magnetic moment, but in ferrimagnetics it will mean that inatomic moment is left unpaired and according to whether it is a parallel or antiparallel spin which is removed, the magnetic moment will be either decreased or increased. Now the element manganese is not ferromagnetic because in the metallic state the atoms are too close together and the M shell *d* electron orbits overlap—by pushing the atoms apart as in the magnetic compound manganese nitride the magnetic nature of manganese is demonstrated. It would thus seem plausible that an iron atom forced into an interstitial position may be too close to its neighbours and hence both it and its neighbour may fail to exhibit ferromagnetism. This would cause a reduction in total magnetic moment for both ferro- and ferri-magnetics. The chances of a ferrite increasing or decreasing its magnetic moment will depend on the relative probabilities of scatter of the atoms from the various positions in the crystal lattice. Only a small change, however, would be expected in the magnetic moment of a material until the radiation bombardment was high by pile standards.

In practical applications of magnetic materials, other than in the microwave region, the useful basic property is the ability to change the direction of magnetization thereby causing a large change of magnetic flux; furthermore the value of applied field required to cause this flux change is of great significance as the permeability is closely related to it. The macroscopic reversal of flux is caused by the growth of those magnetic domains whose moments are parallel to the applied field at the expense of domains otherwise orientated. The boundary wall between domains moves at right angles to its plane on the expansion of a domain, and the various sources of resistance to domain wall movement determine the magnitude of applied field required to cause flux change. Radiation induced defects are expected to increase the resistance to domain wall motion and hence expand the material's hysteresis loop in the H direction (X axis).

3.3. *Chemical Effects*

We have seen that high energy particle radiation will break down in matter to produce a large amount of low level ionization. It requires only a few electron volts to transfer an electron back to its parent anion so producing two neutral atoms and causing in effect chemical decomposition. Now if one of these atoms has sufficient energy and if it is near the surface of the material then there is the possibility of the atom evaporating from the material: the likelihood of surface damage is moderate but in general the neutral atoms will be unable to move from their positions and will recombine chemically, emitting one photon or several small photons in stages leaving the original chemical intact and dissipating the energy in smaller quantum jumps which in the infra-red may be converted to heat. Thus bulk chemical damage in inorganic materials is unlikely.

Plastics, however, are organic in nature and are composed in general of large molecules. Low level ionization may have sufficient energy to break a long organic molecule and the parts of this molecular chain may not rejoin: alternatively several such free molecular portions may rejoin to form a longer molecule than in the original material. This latter process called ionization polymerization is used to advantage in the manufacture of plastics, but when carried to excess produces a brittle product. This would have little harmful effect on satellite engineering in general but would cause serious damage to plastic based magnetic recording tape. The reaction rate of polymerization processes is highly temperature dependent and investigations into the resistance to radiation of plastics should always be carried out at the highest temperature to be expected in the application. The displacement of an atom by a knock-on collision will of course also break a molecule.

The physical properties of plastics may thus be changed but there is also the possibility of short chain molecules having sufficient vapour pressure at ambient temperature to evaporate—bearing in mind that satellite environment is a good vacuum. External parts of a satellite will be exposed to the full ultra-violet portion of the solar spectrum where photon energies are sufficient to ionize organic molecules and the solar wind particles have both the energy and momentum for atomic displacements. We may therefore expect the erosion of the external surfaces of satellites to be more severe when plastics are used in the construction of the skin.

3.4. *Structural Effects*

High-energy radiations can cause changes in the structural properties of inorganic material. Uranium, for example, expands considerably[6] after a long spell in a reactor and graphite hardens noticeably after prolonged irradiation. Such effects are unlikely to disturb satellite equipment as the radiation densities in space are too low. There are a few materials, however, where crystalline phases are in delicate balance; these could undergo changes with moderate irradiation, e.g. Fleeman and Dienes[7] have studied the white to grey tin transformation. Such materials would also be characterized by undue sensitivity to mild temperature treatment and are exceptional.

4. Effect on Equipment

Having outlined the breakdown of high energy radiation and the mechanisms of radiation damage, the question arises of the exposure required to cause malfunction of satellite electronic equipment. Knowing the orbit, radiation densities expected, and the detailed layout of components and materials in a particular satellite it should be possible to calculate the absorption of radiation in the equipment and hence estimate its reliable life. Such a calculation would be very elaborate and not worthwhile as the correlation of theoretically predicted damage and that found experimentally is not good—in nuclear pile experiments most of the error arises from the lack of knowledge of the precise energy spectrum of the radiation incident on the test sample. Some estimate of damage must be made, however, and we find Denney and Pomeroy,[8] Lad,[2] and Heeger, Nisbet and Happ[9] have estimated 2 days, 8 months and 10 years respectively as the life of a transistor at the centre of the inner Van Allen belt. From measurements of radiation damage in atomic reactors[10] one may conclude with fair certainty that of normal electronic components semi-conductor devices are the most susceptible to damage, hence we may concentrate attention on the expected decay of transistors as governing the reliability of electronic equipment.

Although there is a lack of experimental information on radiation damage to transistors, an estimate of transistor life may be made as follows. The number of displacement centres expected to cause noticeable change in transistor properties are given by Dienes and Vineyard[6] as $5 \times 10^{14}/\text{cm}^3$ and $3{\cdot}8 \times 10^{15}/\text{cm}^3$ in silicon and germanium respectively. These centres reduce the mobility of charge carriers by deflecting them, and by trapping curtail carrier life; this would lead to reduced gain and poorer frequency response. Smoluchowski[11] found that 350 MeV protons generated about 5000 centres/cm^3 in a sodium chloride crystal. Now the binding energies and scattering cross-sections of sodium and chlorine ions in NaCl will differ from those of crystalline silicon and germanium; but because the primary radiation used was high energy protons, we may consider his results more pertinent to the present problem than the results of atomic pile experiments. The Van Allen protons have energies ranging from 40 to 700 MeV, the average energy being 100 MeV. We would therefore expect 5000/3·5 centres/cm^3 to be generated by the average Van Allen proton, and as there are 2×10^4 protons/s/steradian then the number, n', of centres generated per second per cm^3 is:

$$n' = 2 \times 10^4 \times 4\pi \times 5000/3{\cdot}5 = 3{\cdot}59 \times 10^8$$

Therefore expected life of a silicon transistor is

$$5 \times 10^{14}/3{\cdot}59 \times 10^8 = 1{\cdot}4 \times 10^6 \text{ seconds or } 16{\cdot}2 \text{ days.}$$

An orbit of considerable interest for communications purposes is an elliptical polar one crossing the South pole at an altitude of about 800 km and the

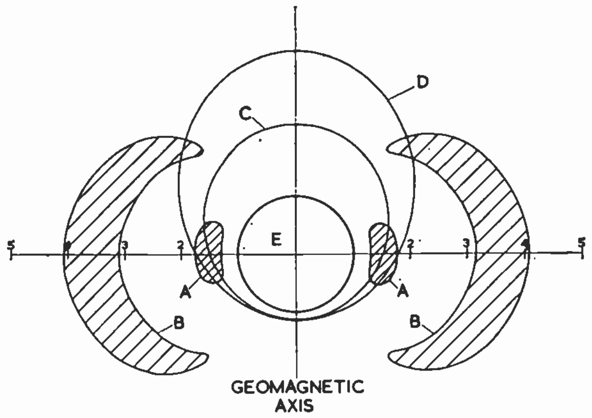


Fig. 1. Satellite orbits and the Van Allen belts.
A—Inner Van Allen belt.
B—Outer Van Allen belt.
C—5000 mile altitude polar apogee orbit.
D—10 000 mile altitude polar apogee orbit.
E—The Earth.

North pole at some 8000 km. In such an orbit a satellite will be usable by two radio terminal points in the Northern hemisphere for an appreciable part of each orbit period. Such a satellite, however, would spend 15% of its time in the inner Van Allen belt (see Fig. 1). Transistor life has been estimated at the centre of this belt, but the radiation density falls off at the edges of the belt and we may consider the average density to be half that at the centre. For the orbit with a 8000 km altitude at a polar apogee, we have an "out of radiation factor" of 40/3 giving 215 days as the effective life of a silicon transistor in that orbit.

It is obvious that the above computation is more a guess than an accurate calculation but it does indicate the seriousness of the problem. Results of irradiation of silicon transistors in the Brookhaven reactor[10] show that a bombardment by $4{\cdot}56 \times 10^{12}$ neutrons of energy greater than 300 keV would halve an original current gain of 50. On this basis and assuming that Van Allen protons had the same effect, energy for energy, as neutrons, we would have a life of 0·8 days for a silicon transistor at the centre of the inner Van Allen belt. Neutrons, however, lose their energy by mechanical collision only whereas protons and γ-rays dissipate their energy by ionization also: hence the results of high energy proton experiments are more applicable to the present problem.

The likelihood of the Varley mechanism being more effective in ionic crystals than in co-valent ones has been ignored: also annealing, although slow acting at normal temperatures, may have appreciable effect over a year. From the above discussion it may appear preferable to use germanium rather than silicon devices on the grounds of susceptibility to radiation damage—this factor would have to be weighed carefully against the better temperature behaviour of silicon components. Furthermore Keister and Stewart[12] have shown that there are large variations (three decades) in damage susceptibility from one type of semi-conductor device to another. Although such tests in an atomic reactor do not give a good value for the life of a transistor in the radiation belts, the ratios of the lives in the reactor should approximate to the ratios of the lives of the same types of component in orbit. To be certain of exposing each device to precisely the same energy spectrum, transistors etc. would have to be irradiated without their normal encapsulation as the case material is liable to give protection, especially from thermal neutrons.

High energy particle bombardment will spoil the permeability of high-μ materials. Alley[13], for example, found Supermalloy to be highly susceptible to radiation damage. Irradiation may thus cause eventual failure of miniature transformers which rely on a narrow, low coercive-force hysteresis loop, e.g. as used in magnetic amplifiers. The exposure time at the centre of the inner Van Allen belt would probably be years before transformer failure occurred but the change of hysteresis loop characteristics may seriously upset the calibration of a flux-gate magnetometer if such were flown for long periods in the radiation belts to record variations of the magnetic field.

Coincident current matrices of magnetic cores are convenient information stores, particularly where random access is required, and as there are no moving parts they are well suited to operation in a vacuum where frictional difficulties may be encountered with other storage systems. Core stores are therefore suitable for use in a satellite where information is recorded around the orbit for rapid transmission to ground on interrogation. The coincident current storage core, however, rely on having a well-defined coercive field for their operation, and radiation induced modifications to the coercive force would seriously upset such a storage system, making it unreliable in a satellite which spends much of its time in the inner Van Allen belt.

Neither the outer Van Allen belt nor the intermediate belt described by Arnoldy, Hoffman and Winckler[4] have been discussed because the energies, penetrating power and total radiation densities in these belts are very much smaller than in the lower belt. Pfann and Roosbroeck[14] have shown that 630 keV electrons in germanium and about 300 keV ones in silicon can cause atomic damage. Electrons of these energies exist in the outer radiation belts but solar cells—the most exposed components—will be protected from erosion by thin slips of silica and these will absorb the energy from the incident electrons.

5. Screening

Having established the possibility of radiation damage, the idea of screening must be considered. A screen to be effective must not only stop the incident particle but also absorb any secondary radiation emitted. About 1·7 cm of Pb will stop a 100 MeV proton but will not greatly attenuate γ-rays. Consider Fig. 2; with no shielding present the proton P1 is trapped in the equipment and creates some damage. With screening the proton P2 which would have penetrated the the equipment is stopped in the shield, but on stopping will emit lower energy radiations in all directions. Protons P3, 4 and 5 which would have missed the equipment altogether are stopped in the shield and will reradiate lower energy quanta—a preponderance of secondary radiation will have a direction similar to that of the incident particle. The net result is that the equipment may receive as much damage from the increased γ-radiation from the shield as it would have received from the passage of the original protons, especially as the latter will dissipate only a small fraction of their energy, on average, while traversing a transistor for example. As

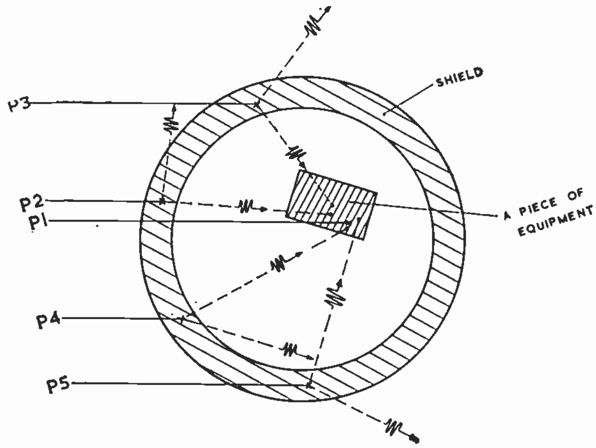


Fig. 2. Illustrating the efficiency of screening.

it is well nigh impossible to provide complete protection, a shield against high energy particles does not therefore appear to justify its large part of a payload; on the other hand a moderate thickness of metal will shield against medium and low energy radiations.

6. Allowances for Radiation Damage

The trapping of high-energy particles and the ensuing energy breakdown will result in ionization in materials normally considered as insulators. Capacitors, having a large area of thin insulating medium, will show this ionization as a temporary lowering of their leakage resistance. From atomic pile experiments[10] it appears not unreasonable to expect the insulation resistance of capacitors to fall by a factor of 10 or more during transit of the inner Van Allen belt. Again susceptibility to radiation damage varies according to the structure of the capacitor. Thus all high impedance circuits should be designed to operate satisfactorily with this loss of insulation.

As the rate of decrease of transistor current gain on irradiation is directly proportional to the original current gain[10] it would be desirable to use only circuits requiring low current gain transistors. The high frequency response of transistors will also fall with irradiation and therefore semi-conductor devices should not be required to operate near their normal high frequency cut-off.

As gallium and arsenic are both heavier atoms than silicon, gallium arsenide[15] solar cells would have greater resistance to radiation damage than silicon ones. The efficiencies of GaAs and Si cells should be similar; the former may therefore be preferable in fonditions of dense irradiation.

7. High Vacuum and Low-energy Radiation

In any useful orbit a satellite will be in a very hard vacuum which will encourage evaporation from all open surfaces. Any orbit centred on the Earth is going to remain in the outer region of the solar gas envelope, and hence will be bombarded by ions of oxygen, nitrogen and hydrogen with energies of the order of 50 eV. The satellite body will also be exposed to the full solar spectrum. Both the particle and e.m. radiations are going to increase the evaporation rates by giving individual atoms or molecules sufficient energy to break the atomic "surface tension"; in the case of chemical compounds these low energy radiations can cause chemical decomposition and at least some of the products are liable to be volatile. Hence we may expect all exposed surfaces to suffer from erosion and change of chemical nature, the extent depending on the actual material.

Most of a satellite's electronic equipment can be placed in metal containers which provide protection from all low-energy radiation. It has been suggested[2] that evaporation could cause changes in components particularly, for example, in the thin film resistors of an evaporated microcircuit. Equipment can be run effectively at normal pressure by potting in a rigid expanded foam plastic which has the additional advantage of improving rigidity at little expense in weight. There are two points requiring care here. One is that although the metal case may be made of thin material to save weight, foamed polyurethanes for example, lose an appreciable part of their strength at 100° C, so leaving a smaller mechanical safety factor —a satellite's temperature may eventually rise owing to erosion of the thermal-radiation-balancing outer layer of silicon monoxide. Secondly if there is a small leak in the container the air will diffuse out and at low pressures the structure may have appreciable dielectric loss due to high frequency discharge effects.

On launching it is only a matter of about two minutes before the rocket carries the satellite to an altitude of, say, 33 miles where the pressure is only 1mm Hg and therefore negligible from mechanical aspects. All containers, other than those designed to remain permanently pressurized, should have egress holes sufficient to allow the air to escape under a safe differential pressure. Just after launching the pressure inside the satellite will be a few millimetres of mercury which is ideal for plasma discharge. All high voltage supplies with exposed conductors are therefore liable to short circuiting for a few minutes by a low resistance. Such design details can of course be checked easily and cheaply in the laboratory.

8. Conclusions

From our admittedly incomplete knowledge of damage caused by high-energy radiation as generated in atomic piles and by high voltage machines, and from the measurements of the radiation density and its spectrum around the Earth, it appears that component failure due to radiation damage is not unlikely in a satellite whose intended working life is several years when that satellite spends an appreciable part of its time in the inner Van Allen radiation belt. A radiation shield of tolerable weight is unlikely to greatly reduce the number of high-energy primary particles traversing a satellite's equipment, but is likely to increase the amount of γ-irradiation so that the decrease in damage to electronic equipment would be small. Thus a shield against high-energy particles does not seem to justify its large part of the payload. Solar cells cannot be effectively protected from hard radiation. To improve the reliability of equipment intended for operation in the radiation belts a judicious choice of transistor types may be made and circuits designed to afford wide tolerance to transistor parameters. The latter feature is also required to cope with the large temperature variations in a satellite.

All satellites orbit in a hard vacuum and are exposed to bombardment by low-energy ions and the full solar spectrum. Sufficiently hard vacua for most purposes may be created in the laboratory to simulate space, the solar spectrum can be well simulated by a high pressure xenon discharge, ionization may be produced by x-rays, and ionic bombardment produced by direct gas discharge: with these tools a wealth of work awaits us to investigate the resistance to erosion of metallic and insulating materials, evaporation rates and mechanical reliability of specific equipment designs. When further quantitative information is available from laboratory research on the effects of environment, an elaborately equipped satellite may then be designed with greater confidence as a radio communications link.

9. Acknowledgment

The author wishes to thank the Director of Research, Marconi's Wireless Telegraph Co. Ltd. for permission to publish this paper.

10. References

1. Smithsonian Physical Tables.
2. R. A. Lad, "Survey of Materials Problems Resulting from Low-Pressure and Radiation Environment in Space", N.A.S.A. Tech. Note D.477, November 1960.
3. J. A. Van Allen, "The geomagnetically trapped corpuscular radiation", *J. Geophys. Res.*, **64**, p. 1683, November 1959.
4. R. L. Arnoldy, R. A. Hoffman and J. R. Winckler, "Observations of the Van Allen radiation regions during August and September 1959", *J. Geophys. Res.*, **65**, p. 1361, May 1960.
5. N. G. Sakiotis, E. I. Salkovitz and A. I. Schindler, "The Effects of Nuclear Irradiation upon the Microwave Properties of Certain Ferrites", U.S.N.R.L. Lab. Rep. 5321, June 1959.
6. G. J. Dienes and G. H. Vineyard, "Radiation Effects in Solids". (Interscience Publishers, New York.)
7. J. Fleeman and G. J. Dienes, "Effect of reactor irradiation on the white-to-grey tin transformation", *J. Appl. Phys.*, **26**, p. 652, June 1955.
8. J. M. Denney and D. Pomeroy, "Radiation damage and transistor life in satellites", *Proc. Inst. Radio Engrs*, **48**, p. 950, May 1960.
9. A. J. Heeger, T. R. Nisbet and W. W. Happ, "An estimate of transistor life in satellite applications", *Proc. Inst. Radio Engrs*, **47**, p. 991, May 1959.
10. Eighth Quarter N.E.E.P. Report.
11. P. Smoluchouski, "Effects of nuclear irradiation on ionic crystals", *Proc. Internat. Conf. Peaceful Uses of Atomic Energy, Geneva*, 7, August 1955.
12. G. L. Keister and H. V. Stewart, "The effect of nuclear radiation on selected semiconductor devices", *Proc. Inst. Radio Engrs*, **45**, p. 931, July 1957.
13. R. E. Alley, "In pile measurements of radiation effects in magnetic materials", *J. Appl. Phys.*, Suppl. 30, p. 284, April 1959.
14. W. G. Pfann and W. Van Roosbroeck, "Radioactive and photoelectric *p-n* junction power sources", *J. Appl. Phys.*, **25**, p. 1422, November 1954.
15. D. N. Nasledov and B. V. Tsarenkov, "Spectral characteristics of GaAs photocells", *Sov. Phys. Solid State*, **1**, p. 1346, March 1960.
16. E. E. Klontz and K. Lark-Horovitz, "Displacements produced by electron bombardment of germanium", *Phys. Rev.*, **86**, p. 643, 1952.
17. F. Seitz and J. S. Koehler, "Displacement of atoms during irradiation", *Solid State Physics*, **2**, p. 305, 1956.
18. J. H. O. Varley, "A new interpretation of irradiation induced phenomena in alkali halides", *J. Nuclear Energy*, **1**, p. 130, December 1954.

11. Appendix

A few elementary examples are given here to illustrate the magnitudes of the energies involved in the displacement of an atom from its position in a crystal lattice, by radiation bombardment.

A nuclear particle will transfer a maximum of energy to an atom when a central direct collision with the nucleus occurs. In this case the particle will bounce off the nucleus and return along its original path with a velocity only slightly lower than its incoming one, on account of the faction of energy transferred to the nucleus. The energies and momentum of the particle and atom before the collision may be equated to their total energies and momentum after the encounter; on solving these two equations we have:

$$E_{\max} = \frac{4mME_i}{(m+M)^2} \qquad \ldots\ldots(1)$$

where $E_{\max}$ is the energy transferred to the atom of mass M, by the particle of mass m with original energy E_i. The energy binding either a Si or a Ge atom in a crystal lattice is about 30 eV, e.g. see Klontz[16], and as Si is very nearly 28 times and Ge 72 times heavier than a neutron or proton, it is straightforward to show that the minimum energy a neutron must have to displace a Si or Ge atom is 225 eV and 555 eV respectively.

As an electron is about 1838 times lighter than a neutron it is obvious that a much higher energy electron would be required to dislodge an atom. A high energy electron travels with a speed approaching that of light and relativistic effects have to be considered in a calculation involving such an electron, e.g. see Seitz and Koehler.[17] For a relativistic particle of much smaller mass than the target nucleus, eqn. (1) becomes

$$E_{\max} = \frac{2m}{M} \cdot \frac{(E_i + 2mc^2)}{mc^2} \cdot E_i \qquad \ldots\ldots(2)$$

where c is the velocity of light and m is now the mass of an electron. When relativistic effects have to be considered it is convenient to measure energies in units of $1\ mc^2 = 511$ keV. For example a 511 keV electron in a head-on collision with a nucleus of silicon and a nucleus of germanium will transfer 59·6 eV to the Si and only 23·2 eV to the Ge. An electron of this energy would therefore be incapable of displacing a Ge atom from its crystal position, whereas it not only would displace a Si atom but almost gives the atom sufficient energy to displace another Si atom. This demonstrates the difference in susceptibility of Si and Ge to radiation damage.

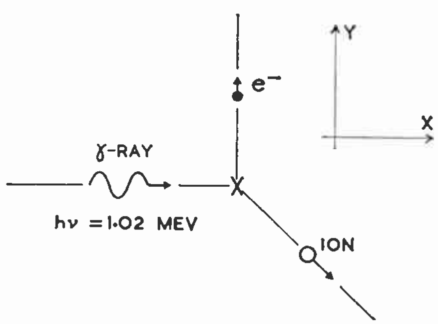


Fig. 3. Photoelectric effect.

A particularly simple example of the photoelectric effect is illustrated in Fig. 3, where an incident photon travelling in the X direction ejects an electron in the Y direction. As the electron has no velocity component in the X direction and the γ-ray gives up all its energy in the encounter, the incident photon's momentum must be absorbed by the ion. Similarly as the γ-ray has no component in the Y direction the ion must absorb all the electron momentum recoil. We will assume that the incident γ-ray has just a little over 1·022 MeV and gives the ejected electron an energy of $2\ mc^2$ making the effective electron mass 3 times its rest mass. The momentum of the γ-ray is

$$m = \frac{h\nu}{c} = \frac{E}{c} = 5{\cdot}44 \times 10^{-17} \text{ g cm/s}$$

The kinetic energy and the momentum of the electron are respectively

$$E = mc^2\left(\frac{1}{\sqrt{1 - B^2}} - 1\right) = 1{\cdot}022 \text{ MeV}$$

$$m = \frac{mBc}{\sqrt{1 - B^2}} = 7{\cdot}73 \times 10^{-17} \text{ g cm/s}$$

where B is the electron velocity divided by the velocity of light.

Therefore total momentum transferred to the ion = $9{\cdot}43 \times 10^{-17}$ g cm/s. In silicon this corresponds to an energy of 59·2 eV and in germanium to an energy of 23·0 eV. The Si atom would therefore be displaced from its lattice position while the Ge one would not. This example again demonstrates the difference in susceptibility to radiation damage between these two elements.

Manuscript received by the Institution on 9th June 1961.
(Paper No. 668.)

U.D.C. 537.591

Cosmic Ray Measurements in the U.K. Scout I Satellite

By

Professor H. ELLIOT, Ph.D.,†

J. J. QUENBY, Ph.D.,†

D. W. MAYNE‡ and

A. C. DURNEY, B.Sc.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: The main purpose of this experiment is to make accurate measurements of the primary cosmic ray energy spectrum and of the way in which this spectrum changes as a result of modulation by the interplanetary magnetic field. The cosmic ray intensity will be measured as a function of latitude and longitude using Geiger-Muller counters, which respond to the total flux of primary particles, and a Cerenkov counter, which responds only to the heavy primaries with a charge of six or greater. An outline of the physical background to the experiment is given, together with a description of the sensor and its associated electronics.

1. Experimental Objectives

The main purpose of this experiment, to be flown in the *Scout* satellite *U.K.1*, is to make accurate measurements of the primary cosmic ray energy spectrum and of the way in which this spectrum changes as a result of modulation by the interplanetary magnetic field. There are at present several alternative models of this field, and in order to distinguish between the various possibilities, much more refined data are necessary than those now available from the observation of secondary cosmic ray intensity variations deep in the atmosphere and fragmentary information from balloon ascents.

In the present experiment it is proposed to investigate the cosmic ray spectrum by using a Cerenkov detector to measure the intensity of heavy nuclei ($Z \geqslant 6$) as a function of latitude. The energy spectrum can then be determined from the known values of the minimum energy which a particle must have in order to penetrate the earth's magnetic field at a given latitude. The advantages of carrying out measurements in this way are: (a) it avoids the invalidation of cosmic ray measurements by inadvertent detection of Van Allen particles; (b) it avoids the introduction of uncertainties by albedo particles scattered back from the atmosphere.

The proposed orbit for *Scout 1* has an inclination of 52 deg, an apogee of 600 statute miles, and a perigee of 200 statute miles.

The detector will sweep through the cosmic ray energy spectrum, in the range 3 GV to 15 GV, four times on each orbit, giving a virtually continuous check on its variation with time. In addition, it is intended to carry out aeroplane surveys at the same time so that the intensity distribution in the atmosphere can be uniquely related to the primary spectrum. It is hoped by this means to obtain a sufficiently accurate relationship for the primary spectrum to be determined at any future time simply by an aeroplane survey.

A large and a small Geiger counter will be included in the instrument package. On those parts of the orbit where the satellite is outside the Van Allen belts the large Geiger will give a measurement of the energy spectrum of the primary cosmic ray protons and also bursts of solar protons in the same way as the Cerenkov detector measures the heavy particle spectrum. Thus it is hoped to be able to compare the proton and heavy particle spectrum as a function of time. Inside the Van Allen belts the large Geiger counter will saturate but the small Geiger counter will be able to count a much higher particle intensity. It is hoped to obtain data on the time variations of the trapped radiation from the small counter rate.

2. Operation of the Sensor

The Cerenkov detector consists of a hollow perspex sphere, 4 in. in diameter, together with an E.M.I. 6097 photomultiplier of 2 in. in diameter which looks into a hole cut in the surface of the sphere. Cerenkov light flashes produced by cosmic rays in the wall of the sphere are detected by the photomultiplier. The pulse output of the tube is fed to a discriminator which only accepts pulses corresponding to primary nuclei of $Z \geqslant 6$ and rejects smaller background pulses due to lighter nuclei and the Van Allen particles. Output from the discriminator is fed to a chain of eight binaries and the contents of this store are sampled by both the data storage and direct telemetry encoders in the satellite.

In-flight calibration is provided since the Cerenkov counting rate is sensitive to the overall gain of the

† Imperial College of Science and Technology, London.
‡ McMichael Radio Ltd., Slough, Bucks.

system. An α-particle source embedded in a scintillator will be used to provide standard light flashes for the photomultiplier. These light flashes are smaller than those produced by the passage of heavy nuclei through the perspex sphere and will not normally pass the discriminator. When the binary store is filled by cosmic ray counts, an electric switch will reduce the attenuation of the photomultiplier output pulses to allow the α-particle pulses to trigger the discriminator. After these calibration pulses have filled the store, the switch will return the attenuation once again to the level for heavy nuclei. Thus a standard counting rate is provided for a fraction of the time on the telemetry so that the system gain may be continuously monitored.

The complete instrument package is located on the spin axis at the forward end of the satellite with the perspex sphere protruding outside the satellite surface and the photomultiplier and associated electronics just inside. Primary heavy nuclei can arrive at the sphere from the forward 2π solid angle where the amount of shielding is in general less than 1 g/cm^2 but are absorbed by the satellite structure in the backward 2π solid angle. It is advantageous to locate the Cerenkov detector on the spin axis since the direction in which it sees cosmic rays only changes slowly in time with respect to the earth's surface and the geomagnetic field. The effect on the cosmic ray intensity of both the absorption by the earth and the deflection by the geomagnetic field must be taken into account in the analysis of the results.

The Geiger counters are placed in the instrument package with structure shielding of the order of 2 g/cm^2 from the outside in the forward direction. The effective length and diameters of the counters are respectively 5 cm and 2 cm for the larger and 1 cm and 0·3 cm for the smaller. An additional 1 mm of lead is put round the small counter. Both counters feed in parallel a chain of 14 binaries and the contents of this store are read periodically by both the data storage and direct telemetry encoders. If the counting rate of the large Geiger tube exceeds the capacity of the binary store, an integrating circuit turns it off with the result that only the smaller counter remains operative, so increasing the dynamic range of the system.

3. Sensor Electronics

A block diagram of the electronic circuitry is shown in Fig. 1. Two input power supply lines are required: 200 mW at -9 ± 1 V and 100 mW at $-6{\cdot}5 \pm 1$ V. The -9 V line is fed to a d.c. converter producing a very stable -6 V line which is used to supply the e.h.t. converter, the emitter follower, the discriminator, and the gate circuits.

The d.c. converter which incorporates a corona stabilizer supplies the 60 megohm dynode chain of the photomultiplier with $+1000$ V, and this e.h.t.

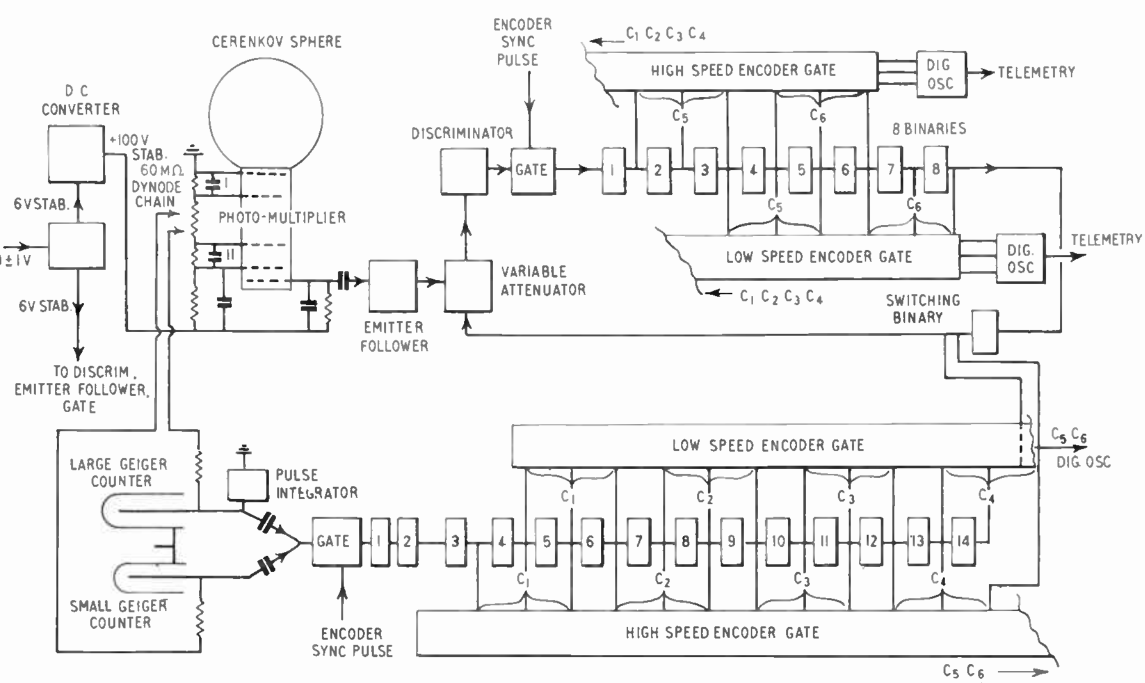


Fig. 1. Block diagram of the electronic circuits.

voltage was designed to have a ± 4 V variation for a ± 1 V variation on the -9 V input supply. Pulses from the photomultiplier are fed via an emitter follower and an attenuator to a discriminator. The discriminator level is designed to be stable to $\pm 3\%$ for ± 1 V variation on the -9 V input supply, and the attenuator is switched to permit measurement of either the calibration source or the cosmic ray rate.

The discriminator output goes via a gate circuit to the chain of eight binaries. The outputs of the first six binaries are fed to the input gates of the high speed encoder, while the outputs of the last six binaries are fed to the input gates of the low speed encoder. When a particular information channel is to be sampled, the gate circuits feed three of the binary outputs to a digital oscillator which takes up one of eight different frequency levels, depending on the state of the three binaries. This frequency output is then put on the satellite tape recorder in the case of the low-speed encoder, and goes to modulate the telemetry transmitter in the case of the high-speed encoder. Thus to read the binary store two channels are required on both the data storage and the direct telemetry systems.

While the low-speed encoder is sampling the binary store, it is arranged that the gate circuit placed after the discriminator prevents any output reaching the first binary. This is done because the state of the binaries may change during the time taken to sample both binary sets.

An electronic switch working from the output of the last binary will change the value of the two-position attenuator and successively feed the cosmic ray counting rate and the calibration pulse rate into the telemetry systems. When the binary store is filled, the flip-over of the last binary changes the state of a switching binary, the output of which determines the value of the attenuation. In order to tell which attenuation value is being used, the switching binary output is also fed to both the low- and high-speed encoders as one of the three binary elements in information channel C4, the other two elements being taken from the Geiger counter store.

Appropriate voltages for the Geiger counters are tapped off the dynode chain. The anode to cathode voltage of the larger counter is made to fall below the working voltage at counting rates in excess of that which the data encoder can handle. Pulse outputs from the two counters are mixed and fed via a gate circuit to a chain of fourteen binaries. Numbers 3 to 13 are read by the high speed encoder and Nos. 4 to 14 are read by the low speed encoder. The design of the binaries and the method of sampling by the encoders are similar to the case of the Cerenkov channel. Thus four channels of both the data storage and the direct telemetry systems are required. The gate circuit stops extra pulses being fed to the store while the low speed encoder samples the four sets of binaries.

4. Engineering Aspects

4.1. *General Considerations*

The basic requirement was to design the cosmic ray analysing equipment so as to form a compact package suitable for launching in a research satellite. The objectives in this work were (1) reliability, (2) economy in power consumption, (3) low weight, and (4) small size. The first two objectives were of paramount importance; the last two were subordinated to them. In addition, the equipment was required to work over a temperature range of $-10°$ C to $+60°$ C.

The circuitry can be considered to comprise four basic sections: the power unit, the Cerenkov channel, the Geiger channel, and the data storage unit (Fig. 2).

Silicon transistors, rectifiers and Zener diodes are employed throughout. Solid electrolyte tantalum capacitors are used for high values of capacitance and paper capacitors for medium values. For high working voltages and for small capacitances ceramic capacitors are used. The resistors are of the cracked-carbon high-stability type. A total of 121 semi-conductors are employed of which 65 are transistors. The total power consumption is less than 300 milliwatts.

The whole equipment is housed in an aluminium container measuring approximately 9·5 in. high and 6·5 in. in diameter. Its weight is 5·7 lb.

4.2. *Power Supplies*

A number of factors influenced the selection of a power supply voltage. Since much of the circuitry is concerned with microsecond pulses, it is desirable to keep impedances low and a low supply voltage is helpful in this respect. Furthermore, a low supply voltage is conducive to power economy, a factor of paramount importance in satellite instrumentation. Finally, the two factors, drift and circuit accuracy, have to be considered. Since these factors often depend on the stability of transistor parameters, it is important to keep the signal voltages high compared with the magnitude of the changes due to transistor variations. With these points in mind, the voltage selected from the power supplies available was 6·5 V. This, however, was an unstabilized supply, varying by about ± 1 V. It was immediately apparent that a far more stable supply was needed for the discriminator, emitter follower, and e.h.t. converter and accordingly a 9 V line as well was selected to supply a voltage stabilizer giving a stabilized 6 V, controlled to $\pm 1\%$. The design of this stabilizer is largely conventional, embodying a silicon Zener diode voltage reference, a long-tailed pair d.c. amplifier, and a 2-stage series

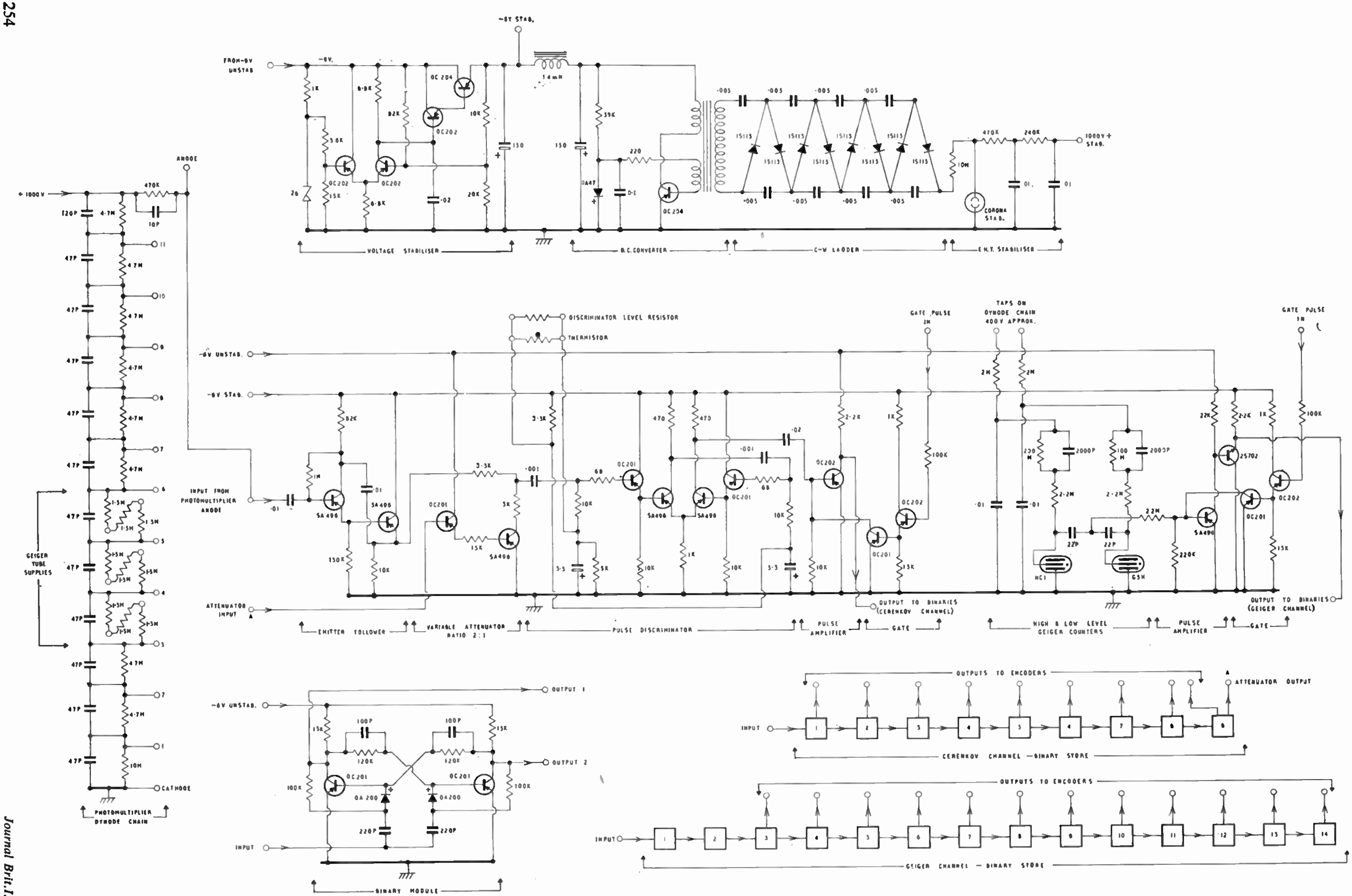


Fig. 2. Circuit of the cosmic ray analyser.

circuit, i.e. four transistors in all. Compensation is also applied to produce a higher stabilization ratio than that which could normally be obtained.

The e.h.t. converter is fed from this stable 6 V line. It is a single-transistor convertor using a ferrite-core transformer and producing 150 V at the secondary. The secondary winding drives a voltage-multiplying Cockcroft-Walton ladder of eight silicon diodes to produce a 1200 V e.h.t. supply. The high oscillation frequency of 16 kc/s was chosen for the convertor in order to keep the size of the capacitors in the ladder small. This high frequency also reduces the size of the filter required to isolate the e.h.t. convertor from the rest of the circuitry connected to the 6 V stabilized line. The filter consists of a 1·5 mH ferrite-core inductor and two 150 μF capacitors included in the supply lead to the convertor. A low secondary voltage followed by voltage multiplication was chosen to enable conservatively rated silicon rectifiers to be used, and to ease the problem of secondary-winding insulation. The efficiency of the convertor, including the voltage multiplier, i.e. from 6 V input to 1200 V output, is 75%. A corona stabilizer is connected across the 1200 V line, so giving a stabilized 1000 V supply maintained to within 0·4%. The photomultiplier dynode chain is connected across this 1000 V line; the Geiger tube supplies are derived from this chain.

4.3. *Cerenkov Channel*

The photomultiplier generates negative-going pulses of microsecond duration and of a few hundred millivolts amplitude at an impedance of about 500 kΩ. To reduce this impedance to a more convenient level, a two-stage emitter follower was designed, using high-speed silicon transistors, to give an input impedance in excess of 1 mΩ, with sufficiently low input capacitance to handle these pulses with a minimum of distortion. The output impedance is a few kilohms.

The emitter follower feeds into a switched attenuator, which is simply a resistor switched in and out of circuit by a transistor. Since in one condition this transistor is required to present an open circuit to fast pulses, a very low collector capacitance is called for and the type of transistor used was chosen with this in mind.

The discriminator is of the emitter-coupled monostable type and a pre-selected resistor is employed to set the discrimination level. This is necessary since each equipment is individually calibrated and adjusted to allow for such variations as the photomultiplier gain.

The pulse amplifier and gate consist of a grounded-emitter limiting amplifier, with its input shunted by a transistor which is switched, via another transistor, by the gating pulse.

4.4. *Geiger Channel*

Two Geiger tubes are used, one of high and the other of low sensitivity. Their outputs are connected in parallel and the circuit is arranged so that in high radiation regions the high sensitivity tube passes sufficient current into an integrating network to reduce its supply voltage to below the operating plateau. The two inputs are fed directly into a pulse amplifier and gate similar in design to those in the Cerenkov channel.

4.5. *Data Storage Unit*

The data storage section is composed of 22 binary stages, 8 for this Cerenkov channel and 14 for the Geiger channel. These are of conventional form but designed to operate with maximum power economy from one supply line only, i.e. the −6·5 V unstabilized supply, each binary stage consuming 2·4 mW. Here, to some extent, a compromise is made between economy and maximum operating speed. Reduction of current consumption leads to higher impedances and with given transistors a lower maximum speed results. Most of these binaries will operate up to 30 kc/s, although 10 kc/s is considered to be the reliable maximum for production modules. Since the maximum pulse rate expected in this equipment is 1000 per sec there is an ample safety margin.

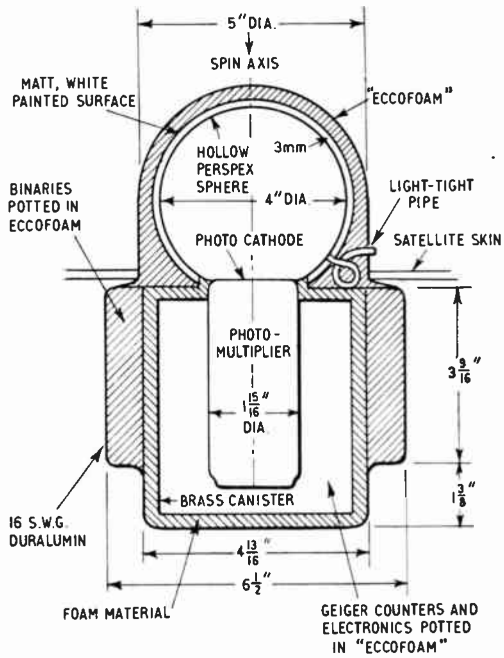


Fig. 3. Mechanical arrangement of the analyser.

4.6. *Mechanical Considerations*

It is considered that, when using conventional components, the greatest component density is achieved by using a "sandwich" form of construction. In this arrangement, the components are strung between two parallel printed-circuit boards, care being taken to choose components of approximately equal lengths. This form of construction has been adopted as far as possible in this equipment. Since there are 22 identical binary units, these were treated as individual modules in "sandwich" construction, each measuring approximately 1 inch square by $\frac{3}{4}$ inch deep.

The basic mechanical arrangement of the analyser shown in Fig. 3 is determined by the photomultiplier, which must be resiliently mounted. Furthermore, since it was considered undesirable for high voltage leads to pass through the resilient material, the d.c. convertor and associated circuitry have been mounted with the photomultiplier. Long leads at high impedance are also undesirable and so the emitter follower is likewise included with the photomultiplier as are the Geiger tubes and their immediate circuitry. All these units are housed in a perforated brass cylinder divided into compartments to screen the high impedance circuits from the e.h.t. convertor. The whole is floated in neoprene foam. The remainder of the circuitry is embodied in three separate toroids around the cylinder, one of "sandwich" construction and two carrying the binary modules.

A detachable dome assembly, housing the perspex sphere, fits over the cathode end of the photomultiplier.

All the electronic circuitry is encapsulated in "Eccofoam"—a proprietary rigid foam-in-place polyurethane resin which contributes to the rigidity of the structure.

Aluminium was chosen as the material for the container, and since the instrument took a cylindrical form, of three different diameters with a hemispherical end, spinning was selected as the method of manufacture. This avoids welded or riveted seams and produces a continuous metal structure without the expense of machining from the solid.

The external finish above the satellite skin is a high polish ready for the application of the correct thermal coating.

Manuscript received by the Institution on 7th June 1961.
(Paper No. 669.)

U.D.C. 523.164.8362: 1.396.93

Long Distance Communication via the Moon

By

P. A. WEBSTER, B.Sc. (Eng.)†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: Three experiments in Moon reflection using 1 kW v.h.f. transmitters and the Jodrell Bank radio telescope are described. The results are discussed in terms of practical long-distance communication circuits using the Moon as a passive reflector. Consideration is given to the optimum frequency, type of modulation, aerial size, etc., for economic utilization of such a circuit.

1. Early History

Engineers started considering the possibility of communicating in space a very long time ago: in February 1927 an article was published in *Radio News* entitled "Can we radio the planets?".

The first actual attempts to reflect signals from the Moon were made before World War II in the U.S.A. by the Naval Research Laboratory, but these were unsuccessful. The first successful attempt seems to have been in Hungary in 1946 using war-time radar equipment and this was followed by further work in Australia, England and the U.S.A.

So far, all the tests had been carried out using radar pulses and the first recorded message, using slowly keyed c.w., was sent from Cedar Rapids, Iowa to Sterling, Virginia, in 1951[1] using a frequency of 418 Mc/s.

In 1958, speech echoes were obtained by Trexler[2] at the Naval Research Laboratory in the U.S.A. and then by Professor A. C. B. Lovell at the Jodrell Bank Experimental Station of the University of Manchester. Recordings of the historic "Hello" echoes were broadcast by the B.B.C. when Professor Lovell delivered a series of Reith Lectures.

It was largely the success of the "Hello" echoes which aroused the interest of the author's company and resulted in discussions with Professor Lovell with a view to assisting him in a series of experiments to explore the possibility of utilizing the moon as a reflector for long distance communication.

2. The Three Experiments

It was agreed that the first experiment would be made with a frequency modulated 201 Mc/s 1 kW transmitter and narrow-deviation receivers of standard commercial design but with a low-noise preamplifier. Both transmitter and receiver were readily obtained, being available for normal commercial purposes, but were slightly modified to reach the allocated frequency of 201 Mc/s compared with their normal upper limit of 174 Mc/s.

The equipment was installed at Jodrell Bank as shown in Fig. 1. The transmitter was hauled up into "Red" tower where a low-loss cable from the aerial tower was already terminated. The receivers—there were three versions of differing bandwidths for comparison purposes—were also installed in "Red" tower. In the case of the receivers, cables with very low loss were avoided by placing the low-noise pre-amplifier on the aerial tower itself.

The hybrid ring and aerial relay were used together to isolate the receiver from the transmitter (an important precaution!) and the hybrid ring was also used to reverse the sense of the circular polarization between transmit and receive. The reason for the change in polarization will be discussed later.

Actual control of the transmitter and receiver was by land-line from a room in the main control building. Provision was made for speech modulation at three alternative levels of peak deviation, for c.w. keying of the transmitter, for tape-recording the received signals, and for making measurements of a.f. tone modulation.

In addition to the proposed tests of reflections back to Jodrell Bank, the U.S. Air Force Research Center at Bedford, Massachusetts had agreed to co-operate by listening with their 84-foot aerial. A corresponding set of receivers and a low-noise preamplifier were accordingly sent to the U.S.A.

The installation was completed in May 1959 and the next moonrise awaited. Unfortunately, at first, no reflections were received and a hurried check of the equipment revealed that the valve in the receiver preamplifier was burnt-out. During the initial setting-up the transmitter had been keyed-on whilst the 50-V supply to the aerial relay was disconnected, thus proving that the isolation of the hybrid ring, some 20 dB, was inadequate without the relay.

† Pye Telecommunications Ltd., Cambridge.

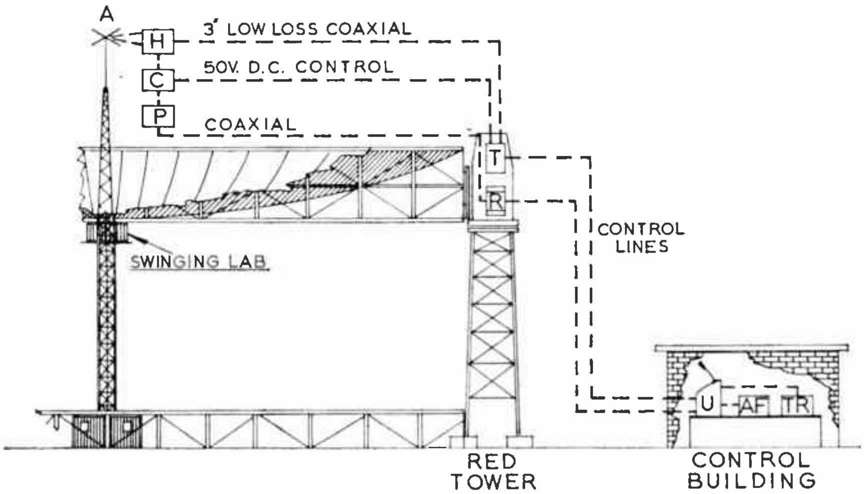


Fig. 1.

Block diagram of first experiment.

A circularly polarized aerial.
H hybrid ring.
C coaxial relay.
P low-noise preamplifier.
T 1 kW transmitter.
R receivers.
U control unit.
AF audio oscillator.
TR tape recorder.

Having overcome this initial difficulty, intelligible speech echoes were received consistently back at Jodrell Bank. In the U.S.A., however, the speech was mostly unintelligible. Continuous wave signals were then sent and a recording of a small sample of these from the American receiver is shown in Fig. 2.

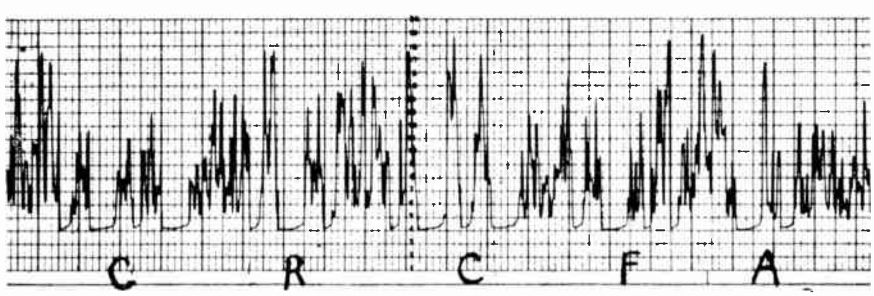
A second experiment was then planned to supplement the first. For this, it was decided that amplitude modulation would be used despite the additional problems of installing a transmitter of twice the size and weight. The original frequency of 201 Mc/s was no longer available and the nearest available frequency, 162·4 Mc/s, was chosen. Since the receiver cable to "Red" tower could not be used, the receivers were installed in the "swinging lab", a small hut pivoted directly beneath the aerial "dish".

The equipment was installed in April 1960 and speech reflections were immediately obtained. This time, the quality of reception was better with less difference between the predicted and measured signal/noise ratio figures. This enabled more accurate measurements to be made, the results of which will be discussed later.

All the equipment used for this second experiment was left at Jodrell Bank until early in 1961 when Dr. J. L. Pawsey of the Commonwealth Scientific and Industrial Research Organization in Sydney, Australia, offered to co-operate in a further experiment, using a recently completed 60 ft aerial.

The first experiments had shown that speech communication could not be expected with a 60 ft aerial and the present equipment. The possibility of telegraph communication was therefore considered. It was calculated that telegraph communication was just possible if a pair of "f.m. telegraph channel units" were used in conjunction with the radio equipment of the second experiment. The "f.m. telegraph channel units" which were kindly loaned by the Telephone Manufacturing Company, consisted of a tone transmitter at 1620 c/s which could be keyed plus and minus 30 c/s. At the receiving end, the 1620 c/s signal was first filtered by a 100 c/s bandwidth filter. This was followed by an amplifier, limiter, discriminator and finally a polarized relay.

One of the channel units was sent out to Australia, together with a receiver and low-noise preamplifier.



Note.—The right-hand end of the paper record emerged first and hence the morse characters read from right to left.

Fig. 2. Pen recording of c.w. signals received in U.S.A. in second experiment.

These were installed at the receiving site and the telegraph unit coupled to a teleprinter. Similarly, a teleprinter and telegraph unit were added to the Jodrell Bank equipment and in February all was ready.

Teleprinter signals were sent, and received in Australia sufficiently well to record a few phrases and short sentences but fading in between times to an unworkable level. These results will also be discussed later.

3. Characteristics of Signals Reflected from the Moon

Examination of signals reflected from the Moon reveals four main characteristics:

(a) A very rapid fading—fairly severe and on a time scale in the range 0·1 to 2 seconds.

(b) A slower fading on a time scale in the range 10 minutes to 2 hours.

(c) A limited useful bandwidth.

(d) A delay time of approximately 2·5 seconds.

That the rapid fading and the slower one were separate effects was first recognized by Kerr and Shain in 1950[3] as a result of their experiments in Sydney, Australia. They were, however, unable to explain the reason for the slow fading but they offered an explanation of the rapid fading. They pointed out that the surface of the Moon was "rough" and a signal reflected from it would therefore contain components which would vary in their relative phase relationship. Also, it was well known to astronomers that, to an observer on earth, the Moon appeared to rotate from side to side in a complex manner. This effect is called "libration". Thus, the signal will not only contain components of differing phase relationship, but these will be constantly changing because of the libration, resulting in the rapid fading generally termed "librational fading".

The rapid fading was now accounted for. It is an unfortunate effect as it causes complete cancellation of the reflected signal although for very short periods. However, the correlation of fading minima is lost with frequency differences as little as tens of kc/s and a frequency diversity system could largely eliminate the problem.

A reason for the slower rate fading was eventually given in 1954 by Murray and Hargreaves[4] at Jodrell Bank. They showed that the plane of polarization of signals passing through the ionosphere is rotated by the Earth's magnetic field. The degree of rotation varies with conditions in the ionosphere and causes fading, called "Faraday fading", when received on a linearly polarized aerial. Fortunately the effect can be largely eliminated by using circular polarization which was used at Jodrell Bank in the tests described. On reflection, circularly polarized signals are reversed in the sense of polarization and therefore when transmitting with clockwise polarization, the receiving aerial must be polarized anticlockwise and vice-versa.

The third characteristic, namely the limited useful bandwidth, is again caused by the uneven nature of the Moon as a reflecting surface. It was originally thought that this would cause such a large scattering in time that speech modulated signals would be rendered unintelligible. That this was not the case was proved when Trexler obtained the speech echoes mentioned in his paper in 1958.[2]

The reason for this difference is that 50% of the reflected power originates from the centre of the visible surface of the Moon, a disc of only one-tenth the diameter of the Moon with a variation in distance from the Earth (due to the Moon's curvature and undulations in the surface) of only five miles.

Thus, if we consider 90 deg phase shift in modulation sidebands to be the maximum acceptable, the upper modulation frequency limit will be about 2·3 kc/s. The effect is generally termed "demodulation".

The fourth characteristic of the reflected signal, while obvious, is nevertheless important since it eliminates the possibility of two-way telephony in the usual sense. Simple arithmetic translation of the speed of light and radio signals, 186 000 miles per second, over a round trip averaging 470 000 miles gives a delay figure of 2·5 seconds. In terms of the delay in replying over a telephone this would be doubled to 5 seconds. Thus, while it is possible to carry out a slow simplex conversation on an "over to you" basis, full duplex commercial telephony connected to the public network would scarcely be practical. No such limitation would apply to telegraph or telex working.

4. Results of Measurements

The main parameters of the equipment for the first experiment are shown in Table 1.

A measurement of signal strength was made by taking the mean of one hundred echoes and found to be 0·46 μV. This corresponds to an aerial-to-aerial circuit loss, taking into account a 3 dB two-way feeder loss, of 172·2 dB, which is in good agreement with the calculated figure 174·3 dB, based on the method given in the Appendix.

Dr. J. V. Evans of the Jodrell Bank Experimental Station reported the results of tone measurements made in the first experiment[5] and Table 2 shows the results of signal/noise ratio measurements using a tape recording of a large number of echoes. To obtain these figures, a wave analyser was used to

Table 1

Parameters of First Experiment

Transmitter:	
Frequency	201 Mc/s
Transmitter power	1 kW
Modulation	f.m., narrow deviation with peak limiting adjustable $\pm 2\frac{1}{2}$ to ± 15 kc/s
Jodrell Bank Aerial:	
Diameter	250 ft
Gain	42 dB at 201 Mc/s
Beamwidth	$1\frac{1}{4}$ deg to 3 dB points (Moon subtends approximately $\frac{1}{2}$ deg)
Polarization	circular
Aerial at Bedford, Mass.:	
Diameter	84 ft
Gain	$32\frac{1}{2}$ dB at 201 Mc/s
Beamwidth	4 deg
Polarization	circular
Receivers (U.K. and U.S.A.):	
Modulation	f.m., narrow deviation
Bandwidth	Three versions: (*a*) $\pm 2\frac{1}{2}$ kc/s to -6 dB (*b*) ± 5 kc/s to -6 dB (*c*) ± 10 kc/s to -6 dB
Noise factor	6 dB with preamplifier

Table 2

Signal/Noise Ratio Measurements using 1 kc/s Modulation Frequency, 1st Experiment

Deviation	Signal/noise ratio $\pm 2\frac{1}{2}$ kc/s receiver	Signal/noise ratio ± 5 kc/s receiver
1 kc/s	6 dB	—
2 kc/s	2 dB	1 dB
3 kc/s	6 dB	10 dB
5 kc/s	—	13 dB†

† One very strong echo.

compare the level of 1 kc/s tone with the level of noise in a narrow band at 1·5 kc/s, and a correction factor applied to allow for the noise in the 3 kc/s a.f. bandwidth of the receivers. It was found that errors in the speed of the recorder, coupled with the short duration of echoes (2·5 seconds) and the narrow bandwidth of the wave analyser, caused the measurements to be subject to quite large errors.

Comparison of these results with the signal/noise performance of the receivers themselves (Fig. 3) shows a discrepancy of some 6 to 10 dB. Some of this discrepancy was undoubtedly due to "demodulation" on reflection and Dr. Evans attempted to make some further assessment of this by measuring the apparent loss of modulation level with increasing frequency of modulation. These results are shown in Fig. 4 and may be compared with the predicted curve (Fig. 5) by Grieg, Metzger and Waer.[6] Dr. Evans also points out that Fig. 4 refers to amplitude modulation

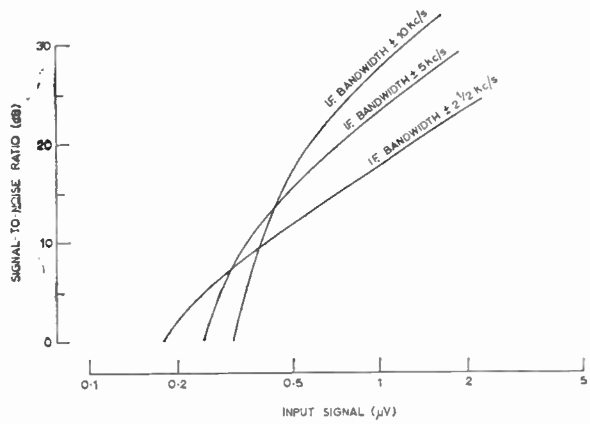


Fig. 3. Signal/noise ratio of the f.m. receivers. Deviation = $\frac{1}{2}$ i.f. bandwidth. Modulating frequency = 1 kc/s.

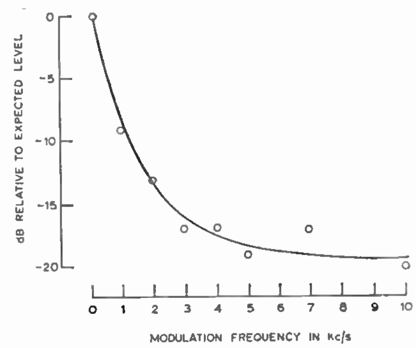


Fig. 4. Variation of level with modulation frequency, first experiment.

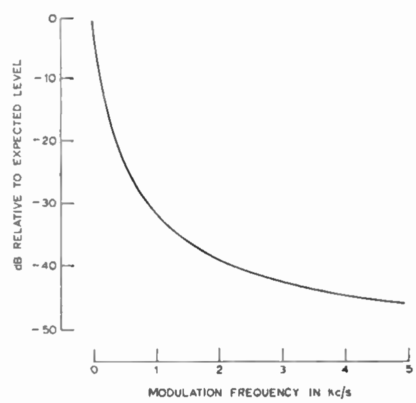


Fig. 5. Variation of level with modulation frequency, predicted by Grieg, Metzger and Waer.[6]

whereas his results are for frequency modulation. However, the theory for frequency modulation with very low modulation indices is very little different because of the similarity of the r.f. spectra, and the large discrepancy is considered to be due to the higher proportion of reflected signal which originates from the central area of the Moon than was originally supposed. Allowing for considerable measurement errors, Dr. Evans estimates the cut-off frequency of modulation to be 5 kc/s.

The second experiment differed from the first only in the following respects:

(a) The frequency was 162·4 Mc/s.

(b) Amplitude modulation was employed.

(c) Receivers of only one i.f. bandwidth (± 5 kc/s) were used.

(d) An automatic timer to key from transmit to receive at 3 second intervals was added to simplify tone measurements.

The signal/noise ratio using 100% modulation at 1 kc/s was measured to be 10 ± 2 dB, taking the mean level of a large number of echoes. This compares with 13 dB which is calculated from the receiver input carrier power to noise power ratio and indicates a system bandwidth of 1 kc/s to the half power point.

Dr. Evans was this time able to plot the system response with increasing modulation frequency much more accurately.[7] He used a pen recorder in conjunction with a wave analyser to record the mean level of each of 240 echoes at 10 different modulation frequencies. The modulation was set to 75% in each case and the result corrected at the higher frequencies to allow for the effects of the receiver i.f. bandwidth. Measurement below 300 c/s was not possible because the transmitter was designed solely for speech purposes.

These results are plotted in Fig. 6 together with the estimated mean curve. The curve shows an even greater discrepancy with the predicted results than was

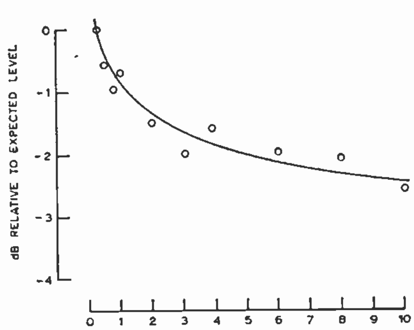


Fig. 6. Variation of level with modulation frequency, second experiment.

the case in the first experiment and indicates that bandwidths as large as 10 kc/s could be used.

Dr. Evans decided against attempting distortion measurements as the particular radio equipment had 7% distortion at 75% modulation, which is too high for accurate radio path measurements.

In the third experiment, teleprinters were used in conjunction with f.m. telegraph channel units, as already described. Laboratory checks of this system showed that reliable results could be obtained down to a level as low as 0·08 μV. This compared with an estimated signal of 0·11 μV for the received level on a 60 ft aerial.

As mentioned earlier the results were poorer than had been hoped for, but this was probably due, mainly, to the fact that the receiving aerial in Sydney was linearly polarized. This would account for a further loss of 3 dB in addition to the degradation caused by the reintroduction of Faraday fading. A further loss of signal is liable to arise due to refraction in the earth's atmosphere when working at the low elevations necessary for very long distances.

It is hoped that tests can be made in the future with a circularly polarized aerial in Australia and, perhaps, a parametric amplifier to reduce receiver noise still further. The author believes that with the same equipment these two changes could improve teleprinter communications to a reliability in the region of 95%.

5. Practical Moon Circuits

The experiments made so far have clarified the nature of the problems involved and lead one to consider the best parameters for a practical Moon circuit.

Certainly one cannot consider erecting aerials the size of the Jodrell Bank radio telescope for the purpose. An aerial of, say, 30 to 60 ft diameter is much cheaper and easier to use; aerials less than 30 ft would not effect much economy as the tracking mechanism would then be the major cost factor.

What, then, is required to obtain a useful performance with an aerial in the range 30 to 60 ft? For a given aerial size, transmitter power and other relevant parameters, the circuit loss reduces by 20 log (frequency), i.e. 6 dB per octave, as the frequency is increased. This law applies up to the point where absorption in the atmosphere begins and Fig. 7 illustrates the overall effect. The curve of Fig. 7 can be modified further, by corrections for feeder losses, noise factor, etc., as well as for the maximum available transmitter power. Figure 8 has been drawn on this basis using the latest available figures but excluding parameters of extremely expensive equipments such as maser amplifiers or very large transmitters. It

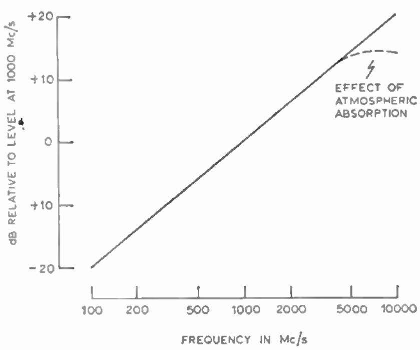


Fig. 7. Relative received signal level for constant conditions with small directive aerials.

should be pointed out at this juncture that there is an upper limit of useful aerial size at any given frequency because the beamwidth could become less than the angle subtended by the useful portion of the Moon. Also, very large aerials could not be manufactured with adequate precision for very high frequencies. However, the range suggested of 30 to 60 ft is quite satisfactory up to 10 000 Mc/s and the optimum shown in Fig. 8 is 5000–6000 Mc/s.

The following example has been calculated using the optimum frequency in order to indicate the order of performance that it is possible to achieve:

Aerial size	50 ft dia. (transmit and receive)
Frequency	5 kMc/s
Transmitter power	5 kW
Receiver noise factor	3 dB
Effective bandwidth	$2\frac{1}{2}$ kc/s
Total feeder losses	3 dB

Using the formulae given in the Appendix the aerial-to-aerial attenuation is 175 dB, and the carrier/noise power ratio is 26 dB.

In considering the performance that could be obtained with such a link, it is necessary to choose a particular form of modulation. The experiments carried out have enabled a comparison to be made of low deviation f.m. and double-sideband a.m. Also, a recording is available of an American Moon reflection (at 403 Mc/s) of a single-sideband a.m. transmission.

The speech quality of the reflection using f.m. was poorer than that with a.m. and the variation of signal/noise ratio greater. It is not possible to utilize a modulation index much in excess of 2 as the loss of correlation of sidebands increases rapidly with frequency spacing[7] and, for most speech modulation frequencies, only the first sideband would be effective.

Consider also the case when one of the two first sidebands is lost instantaneously, leaving the carrier and the other sideband. In the case of f.m., very severe distortion will result—a similar situation to mistuning an f.m. receiver—compared with a 3 dB drop of level and relatively slight distortion in the case of a.m.

Comparison of double-sideband a.m. with single-sideband is also very interesting. Single-sideband suppressed carrier signals will be directly affected by the librational fading and correlation factors whereas double-sideband signals are affected in a more complex manner. The relative phase of the sidebands will vary and will cause cancellation in the demodulation process occasionally. On the other hand, complete loss of one sideband will not cause the intelligence to be lost. It would be expected that, as with normal multipath d.s.b. transmissions, distortion would result when fading of the carrier, without the sidebands being affected, causes apparent over-modulation. In the case of the second experiment however, when d.s.b. was used, there was little evidence of this occurring.

A further factor affecting s.s.b. with suppressed carrier is that slight variation in the Doppler shift (the amount is directly proportional to frequency) will cause fluctuations in the apparent modulation frequency. It is believed that this was the chief cause of loss of intelligibility in the American s.s.b. transmission.

The differences between double and single-sideband working require further study to assess the effects more accurately, but it may be that there is little, if any, overall advantage to be obtained with s.s.b.,

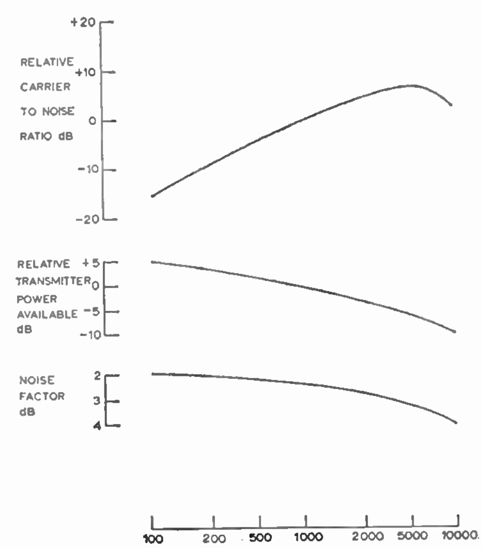


Fig. 8. Variation of received carrier/noise ratio, corrected for available noise factor and transmitter power.

especially at frequencies near the optimum of 5000 to 6000 Mc/s where the technical difficulties are very great.

Systems of pulse modulation using pulse techniques are worth investigation since the results of the second experiment indicate a useful bandwidth as high as 10 kc/s. It is hoped that further work can be carried out to assess the merits of possible methods, but in the meantime there seems little doubt that simple double-sideband a.m. yields a performance at least as good as the other modulation systems in general use and is easy to apply.

In the experiments at v.h.f., Doppler shifts were so small as to be well within the passband of the receivers. Is this so at microwave frequencies, knowing that the shift is directly proportional to frequency? The formula for calculating the Doppler shift on a Moon-reflected path is complex as it must take into account the geometry of the Moon's orbit as well as the position of the radio stations. However, Grieg, Metzger and Waer[6] show the order of magnitude to be $\pm 1{\cdot}5$ cycles per megacycle. Thus the Doppler shift at 5000 Mc/s will be 7·5 kc/s which would be of no consequence in the d.s.b. receiver.

Finally, it is necessary to assess the extent of fading. It has already been mentioned that Faraday fading can be virtually eliminated by circular polarization, thus leaving only the librational short-term fading. Accurate estimates of the distribution are not available, but a figure of 4 to 5 dB for 50% of the time is estimated for the v.h.f. tests. It is uncertain how the librational fading characteristic will change as the frequency is increased and further measurements are necessary. However, it is reasonable to suppose that the time scale will reduce at the higher frequencies and the effect on speech transmissions may well be reduced.

Referring back to the hypothetical link at optimum frequency, a comparison can be made with the performance obtained in the second experiment. Using double-sideband a.m. there is an improvement in carrier/noise power ratio which, coupled with a shorter time-scale for the librational fading, should produce a speech performance in the class of "good communication" quality. Telegraph transmission should also be "good" with, say, twenty circuits multiplexed.

To obtain any further improvement in speech reliability with the same order of transmitter power, etc., it would be necessary to employ frequency or space diversity. Because of the high cost of aerials, the frequency diversity method would have obvious advantages.

In the case of telegraphy, frequency diversity might well be obtained by using the channels in pairs. This is quite possible since Evans[7] has shown that the upper frequencies of modulation in the a.f. range are largely uncorrelated in their fading characteristic.

6. Conclusions

The experiments made have confirmed the path loss, received signal and carrier/noise power ratio calculations of Moon-reflected radio paths and have given a measure of the fading and bandwidth performance.

By using frequencies near the optimum for minimum circuit loss it will be possible to achieve useful speech or teleprinter operation without excessive aerial sizes or transmitter powers, especially if a parametric amplifier is employed in the receiver.

Moon reflected circuits have the disadvantage of too long a transmission time for commercial telephony compared with artificial satellites. However, the total path attenuation is not excessive, for example it is 5 dB less than for the American *Echo* satellite at 1000 miles, and it has the advantage of a much lower angular velocity than non-synchronous satellites with a consequently simpler tracking problem. The actual track of the Moon and hence the actual time it is available in different parts of the world can be found from the *Nautical Almanac*. A table of "mutual viewing periods" of the Moon for representative places is given in Table 3 to indicate the order of period which can be expected. Because of the high level of noise radiated from the Sun, it is necessary

Table 3

Typical Mutual Viewing Periods of Moon

Place	Maximum daily hours	Minimum daily hours
CALCUTTA	8 h 30 min	3 h 54 min
ADELAIDE	3 h 35 min	1 h 52 min
CENTRAL NIGERIA	13 h 44 min	9 h 46 min
MONTREAL	10 h 18 min	4 h 05 min
LUSAMBO (Congo Republic)	12 h 40 min	9 h 03 min
CAPE TOWN	11 h 57 min	9 h 59 min
SINGAPORE	6 h 55 min	3 h 15 min
CENTRAL VENEZUELA	9 h 46 min	5 h 53 min

The figures in the tables were calculated for the year 1959 and are corrected to exclude the period when the elevation of the Moon is less than 7 deg. In all cases the periods are based on circuits from London.

to exclude any periods when the Moon is within about five degrees of the sun as viewed from the receiving site.

7. Acknowledgments

The tests were made possible by the enthusiastic co-operation of Professor Sir Bernard Lovell, F.R.S., and two members of his staff, Dr. J. V. Evans and Dr. J. Thompson.

In addition, a valuable contribution was made by the U.S. Air Force Research Center at Bedford, Massachusetts and by the Commonwealth Scientific and Industrial Research Organization in Sydney, Australia.

Finally, the author is indebted to the Directors of Pye Telecommunications Limited for permission to publish this paper, and to Mr. J. J. Rudolf for advice on the microwave aspects of Moon measurements.

8. Bibliography

1. P. G. Sulzer, G. F. Montgomery and I. H. Gerks, "An u.h.f. Moon relay", *Proc. Inst. Radio Engrs*, **40**, p. 361, March 1952 (Letter).
2. James H. Trexler, "Lunar radio echoes", *Proc. Inst. Radio Engrs*, **46**, pp. 286–92, January 1958.
3. F. J. Kerr and C. A. Shain, "Moon echoes and transmission through the ionosphere", *Proc. Inst. Radio Engrs*, **39**, pp. 230–42, March 1951.
4. W. A. S. Murray and J. K. Hargreaves, "Lunar radar echoes and the Faraday effect in the ionosphere", *Nature (Lond.)*, **173**, pp. 944–5, 15th May 1954.
5. J. V. Evans, Internal Report, Jodrell Bank Experimental Station of the University of Manchester, June 1959.
6. D. D. Grieg, S. Metzger and R. Waer, "Considerations of Moon-relay communication", *Proc. Inst. Radio Engrs*, **36**, pp. 652–63, May 1948.
7. J. V. Evans, "Radio Communication using Moon-reflected Signals", Internal Report of the Massachusetts Institute of Technology, Lincoln Laboratory, December 1960.

9. Appendix

The Calculation of the Performance of a Circuit using the Moon as a Reflector

The aerial-to-aerial circuit loss, L_A can be calculated from

$$L_A = 2L_{FS} - G_T - G_R - G_M + L_a$$

where L_{FS} = free space loss, dB.
G_T = transmitting aerial gain, dB.
G_R = receiving aerial gain, dB.
G_M = gain of Moon as passive reflector, dB.
L_a = atmospheric absorption, dB.

In terms of commonly used units:

$$L_{FS} = 36{\cdot}6 + 20 \log_{10} d + 20 \log_{10} f.$$
$$G_T \text{ or } G_R = -52{\cdot}2 + 20 \log_{10} D + 20 \log_{10} f.$$
$$G_M = 76 + 20 \log_{10} f.$$ (This is based on a reflection factor of 0·03 as measured by Dr. J. V. Evans at Jodrell Bank.)

where d = distance in miles (mean value for Moon = 235 000).
f = frequency in Mc/s.
D = aerial diameter in feet.
L_a can be determined approximately from Fig. 7 of the paper above 4000 Mc/s. Below this frequency, the value can be taken as zero.

Simplifying the equation still further:

$$L_A = (2 \times 36{\cdot}6 + 2 \times 52{\cdot}2 - 76) + 40 \log_{10} (235\,000) - 20 \log_{10} D_T - 20 \log_{10} D_R - 20 \log_{10} f + L_a$$
$$= 316{\cdot}4 - 20 \log_{10} D_T - 20 \log_{10} D_R - 20 \log_{10} f + L_a$$

In many cases it will be necessary to calculate the carrier/noise power ratio. Let this be C_N in dB, then

$$C_N 204 - L_A + 10 \log_{10} P_T - F - 10 \log_{10} B - L_f$$

where P_T = transmitter power in watts.
F = receiver noise factor in dB.
B = effective bandwidth of receiver in c/s.
L_f = total feeder losses in dB.

Manuscript received by the Institution on 3rd May 1961. (*Paper No. 670.*)

U.D.C. 621.375.9 : 621.396.946

The Maser and its Application to Satellite Communication Systems

By

P. HLAWICZKA,

B.Sc.(Eng.), B.Sc.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: The principles of operation of the maser are reviewed with emphasis on its ability to give low-noise amplification. The properties of the comb structure travelling wave maser are discussed and related to the microwave "space window", with special reference to an S-band amplifier which has been constructed. The performance of a receiver, incorporating a maser or a parametric amplifier, is analysed as a function of aerial noise temperature. The basic elements of a maser installation are described with particular regard to the problems of attaching it to an ultrasensitive receiver incorporating a fully steerable, highly directive low-noise aerial. Some consideration of possible cost is included.

1. Principles of Maser Amplifiers

The operation of a maser device can only be explained in terms of the properties of atomic and molecular particles, and their interaction with electromagnetic waves, based in turn on the principles of quantum mechanics. The properties and interactions relevant to the maser can be summed up under two headings.

(i) Atomic and molecular particles tend to exist naturally at discrete energy levels.

(ii) They can jump from one energy level to another under the influence of an electromagnetic wave of the right frequency. In the process they give up to the wave, or absorb from it, a discrete quantum of energy, depending on whether they dropped to a lower level or were lifted up.

The last statement contains the germ of the maser principle. All that must be done is to arrange for a great number of particles to exist at a higher energy level and to expose them to the stimulating influence of an electromagnetic wave, when they will drop to a lower energy level giving up electromagnetic power to the wave, and thus amplifying it. To achieve the interaction, the gap between the energy levels of the particles must be related to the frequency of the wave by the equation

$$\Delta E = hf$$

where $h = 6{\cdot}6 \times 10^{-34}$ joule-seconds is Planck's constant.

To realize a maser, particles that occupy energy levels separated by amounts proportional to microwave frequencies according to the equation above must be found. Secondly, favourable conditions must be arranged for the particles to interact with microwave signals and amplify them. Unfortunately such conditions are not found to exist naturally.

Paramagnetic ions provide energy levels suitable for use in solid state masers. The ions are included as trace impurities in host crystals which themselves are non-magnetic. The ion used most successfully so far is chromium, Cr^{3+}, diluted by the alumina (Al_2O_3) lattice, resulting in a pink ruby crystal. The energy level scheme provided by this substance is shown in Fig. 1 as a function of the d.c. magnetic field

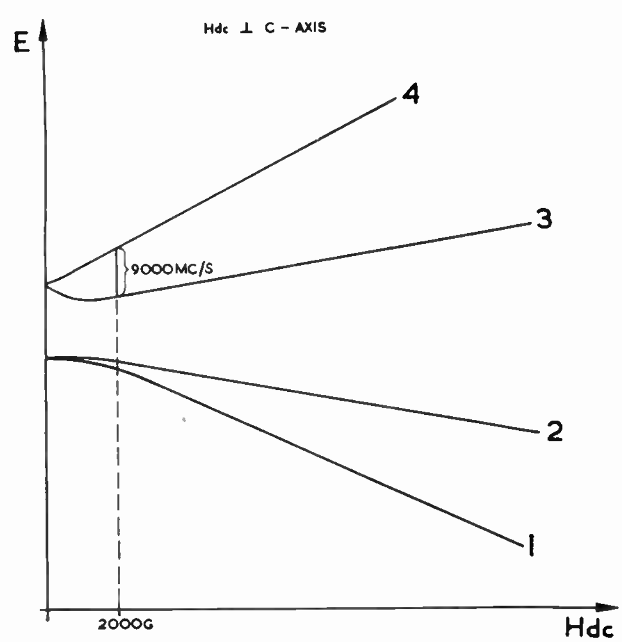


Fig. 1. Energy levels of Cr^{3+} in ruby as a function of the applied d.c. magnetic field.

† Marconi's Wireless Telegraph Company Ltd., Chelmsford.

applied to the crystal in a definite direction relative to the crystal axis. The vertical axis of energy should be visualized as being calibrated in terms of frequencies.

Under natural or equilibrium conditions the paramagnetic ions are distributed among the levels according to the Maxwell-Boltzmann exponential law shown in Fig. 2. The lower levels have greater populations than higher levels.

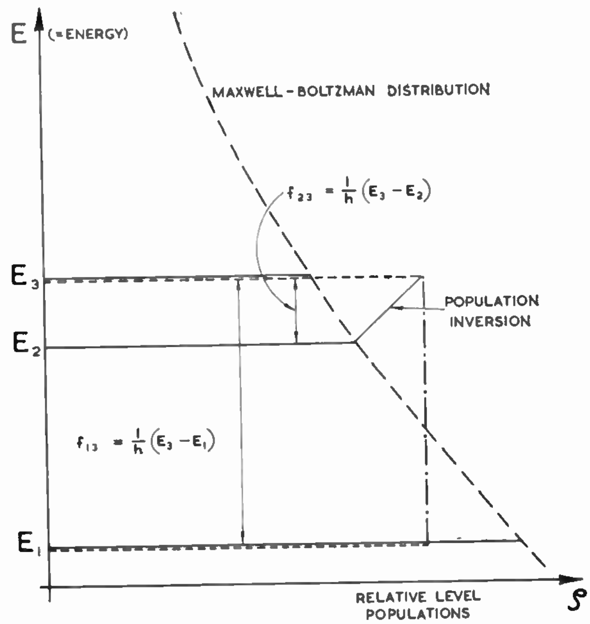


Fig. 2. Principle of the three-level maser $\Delta E = hf$.

The interaction of a microwave signal with ions distributed over a discrete energy level system is governed by the law of equal transition probabilities for both upward and downward transitions, performed under the stimulus of the signal. Hence, a microwave signal applied between levels E_2 and E_3, under equilibrium conditions, will cause more upward than downward transitions, due to the greater population of the lower level. As a result net absorption of microwave energy will be observed. To achieve amplification, the natural state of affairs must be upset and an excess of ions must be made available in level E_3. This is accomplished by applying a strong microwave signal of the right frequency to cause transitions between levels E_1 and E_3. This intense wave will tend to "pump" ions from level E_1 to E_3, leaving E_2 unaffected, until E_3 has a more numerous population than E_2. A weak signal of frequency f_{23} will now be amplified.[1, 2]

The phenomena described above will only take place if the paramagnetic crystal is cooled to a very low temperature. For this reason masers are usually operated in a bath of liquid helium (4·2° K at 760 mm Hg). Otherwise the favourable population distribution could not be realized because rapid relaxation processes would frustrate the work of the pump radiation.

The maser amplifier can be constructed in the form of either a one port cavity together with a circulator, or a travelling wave structure. The former has a very narrow bandwidth and suffers from considerable gain instability, for which reasons it is unlikely to be of any use for communication purposes. The travelling wave maser (t.w.m.) is much more stable and its bandwidth is only limited by the linewidth of the paramagnetic resonance. For ruby crystals this exceeds 50 Mc/s. It is possible to construct slow wave structures capable of operation over much wider bands—several hundred megacycles. A maser based on such a structure could be made tunable over the entire band, by adjusting the value of the d.c. magnetic field and the frequency of the pump, in accordance with the energy level diagram.

It should be borne in mind, however, that the actual t.w.m. bandwidth decreases with increasing gain according to the expression[3]:

$$B = B_m \sqrt{\frac{3}{G_{dB} - 3}}$$

where B = maser bandwidth.
B_m = paramagnetic resonance linewidth.
G_{dB} = maser gain in dB.

The maser is essentially a low level device. It will not handle linearly signals of the order of 10^{-6} W or more.

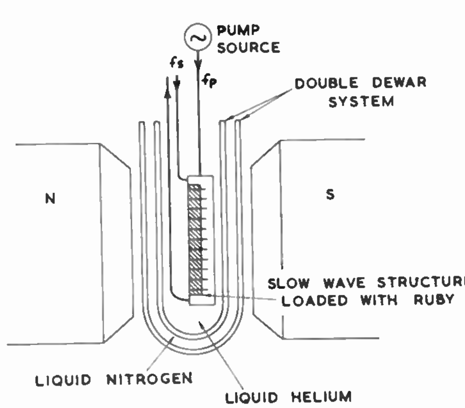


Fig. 3. Schematic diagram showing the essential elements of a travelling wave maser.
f_s = signal frequency. f_p = pump frequency.

The t.w.m. is thus a practical amplifier, to be used in situations where its complexity and high cost are balanced by the advantage offered by its extremely low noise figure. Figure 3 shows schematically the essential elements of a t.w.m., while Fig. 8 is a photograph of an actual laboratory equipment.

2. Low Noise Property of the Maser

The extremely low noise figure of a maser amplifier is due primarily to the low temperature environment in which the device is operated. Measurements of the noise temperature, made on travelling wave masers, yield a figure of about 10° K.[3] It is further estimated that the noise generated within the t.w.m. structure itself contributes only about one-quarter of this figure. The remainder is made up by the input line, about 30 in. long, one end of which is at room temperature, the other inside the helium bath. The noise figure would suffer serious deterioration if an additional length of line were to be connected before the input port of the maser. It follows that a t.w.m. must be mounted directly in the focus of an aerial, if the full advantage of its low noise figure is to be realized.

3. Comb Structure Maser and its Relation to the Microwave "Space Window"

The most successful t.w.m. so far constructed and tried is based on the comb structure, shown diagrammatically in Fig. 4. Figure 9 is a photograph of a

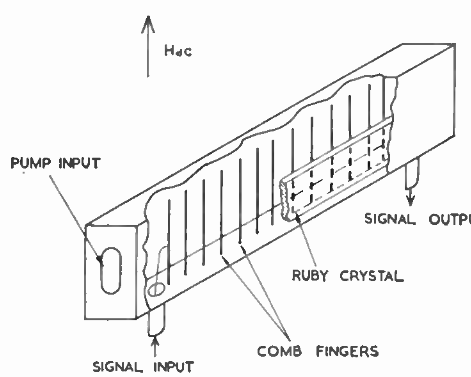


Fig. 4. Diagrammatic view of the comb structure maser.

fabricated S-band t.w.m. assembly, ready for insertion into a liquid helium dewar. The input and output ports take the form of 50-ohm coaxial cables, the amplifier being matched to this impedance at both ends over the operating band.

The gain of the t.w.m. is proportional to the length of the structure. At S-band we expect to provide a gain of 30 dB over a bandwidth of 20 Mc/s, using a structure about 9 in. long. For lower frequencies the structure becomes larger and difficulties arise in connection with the size of the magnet that must be provided, as well as with the helium bath. At higher frequencies the structure is smaller, and fabrication

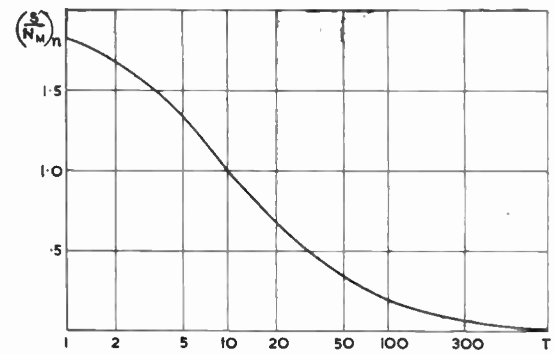


Fig. 5. Normalized signal/noise ratio for a maser receiver, plotted as a function of the aerial noise temperature.

Maser noise temperature: $T_M = 10°$ K.

problems appear, due to the fine tolerances that must be observed. As a rough guide it can be accepted that below 2000 Mc/s the t.w.m. becomes unwieldy, while above 10 000 Mc/s fabrication presents difficulties. Within this band the comb structure maser is comparatively easy to construct and operate.

As it happens, the microwave "space window" coincides approximately with this band, making it possible to derive full advantage from the maser's low noise figure. Below this band galactic noise increases rapidly, while above it atmospheric absorption provides a rather high background of sky noise.[4] Within the "window", sky temperatures remain in the neighbourhood of 10° K for aerial elevations of 20 deg or more above the horizon.[5]

The usefulness of a maser preamplifier drops off rapidly with increasing aerial noise temperature. Here aerial noise temperature is meant to include all sources of noise preceding the input port of the maser. Figure 5 displays this effect graphically, by showing the normalized signal/noise ratio, as a function of aerial temperature, for a receiver system incorporating a maser.

Signal/noise ratio:

$$\frac{S}{N_M} = \frac{S}{kB(T+T_M)} = g(T)$$

where N_M = total maser receiver noise power,
T_M = maser noise temperature (= 10° K),
T = equivalent aerial noise temperature.

Normalizing with respect to $T = 10°$ K:

$$\left(\frac{S}{N_M}\right)_n = \frac{g(T)}{g(10)} = \frac{10+T_M}{T+T_M}$$

The diagram emphasizes the fact that a maser must be installed with a highly directive aerial to offer any advantage.

4. Comparison of the Maser and Parametric Amplifier

The parametric amplifier is a much simpler (and cheaper) device than the maser, although it also requires a source of pump power. With an equivalent noise temperature of about 100° K it comes nearest to the maser as a low noise preamplifier. Its noise figure is sufficiently high to make negligible the effect of a reasonable length of feeder, connecting it to the aerial, so that installation problems are reduced. The aerial itself need not be so highly directive as one used with a maser.

To obtain a graphical assessment of the relative receiver performance, using a maser or a parametric amplifier, it is useful to plot a comparison of the signal/noise ratios for the two cases, as a function of the aerial noise temperature.

System noise power using a maser

$$N_M = kB(T + T_M)$$

System noise power using a parametric amplifier

$$N_P = kB(T + T_P)$$

where T_P = parametric amplifier noise temperature (= 100° K).

Forming now a comparison between the signal/noise ratios for the two cases one finds

$$\frac{S/N_M}{S/N_P} = \frac{T+T_P}{T+T_M} = f(T) = \frac{M}{P}$$

This function is plotted in Fig. 6. It is clear that the advantage derived from using a maser, as compared with a parametric amplifier, diminishes rapidly with increasing background noise. It is up to systems engineers to decide at what point the advantage of reduced transmitter power (and weight) or increased range outweighs the extra complication and cost of a maser.

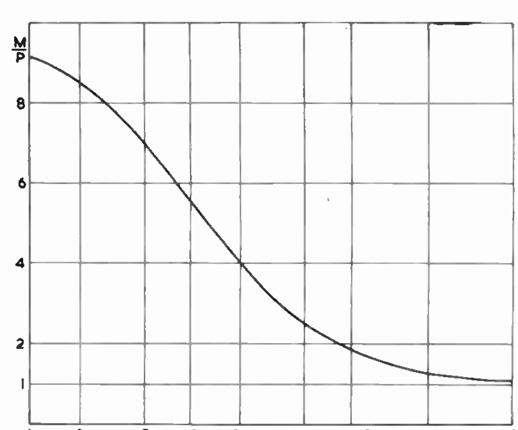


Fig. 6. Comparative receiver performance, using a maser or parametric amplifier, as a function of aerial noise temperature.

Maser noise temperature: $T_M = 10°$ K.
Parametric amplifier noise temperature: $T_P = 100°$ K.

5. The Maser Installation and its Probable Cost

The basic components of a maser installation can be listed as follows:

(i) the maser structure itself,

(ii) liquid helium bath in which the structure is immersed,

(iii) magnet, setting up paramagnetic energy levels ($10^3 - 10^4$ Oe),

(iv) pump source (~0·1 W), with its power supply, tunable, if a tunable maser is required.

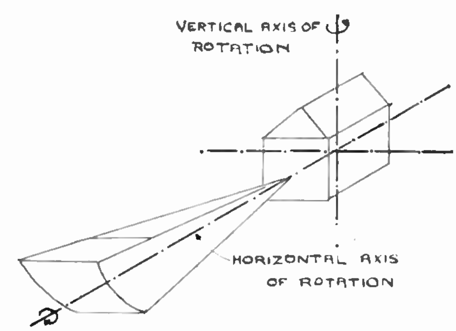


Fig. 7. Schematic view of a fully steerable cornucopia mount.

To receive signals from moving satellites all these elements must be assembled and mounted in or very near the focus of a fully steerable aerial. It is evident that installation difficulties will be substantially diminished if an aerial is used whose focus remains stationary, relative to the housing containing the receiver equipment. The reflecting horn aerial is of this type. It can be mounted horizontally with its focus inside a hut. Full steerability is provided by rotating the aerial about its horizontal axis (passing through the focus), and by rotating it, together with its mounting, about a vertical axis (Fig. 7). In the process the focus does not move relative to the hut, so that quite elaborate equipments can be easily installed and operated inside. Moreover, a maser installed at the apex of a horn will not constitute an obscuration of the radiation path as would a bulky installation at the focus of a parabolic dish. The latter would have a serious effect on the radiation pattern by increasing side lobes, resulting in a higher effective aerial temperature. The cornucopia possesses the advantages outlined in addition to its highly directive and wide band properties.

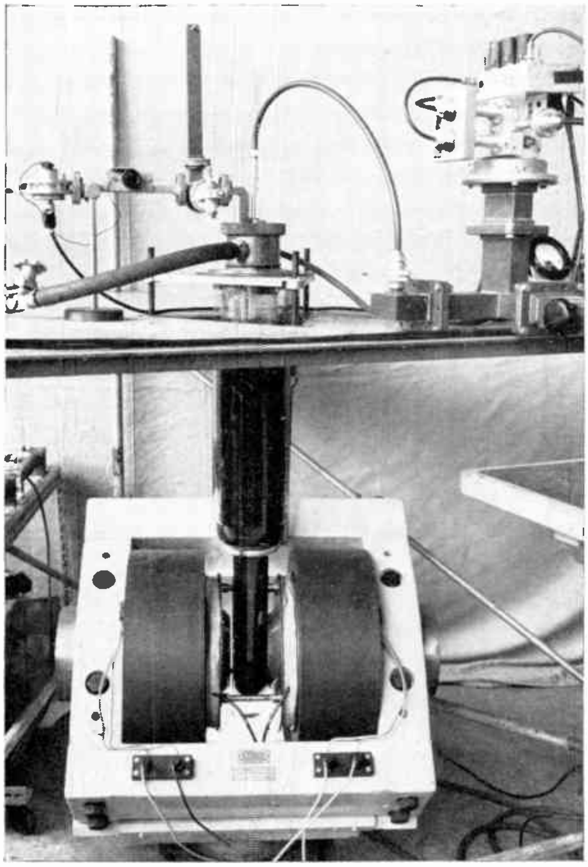


Fig. 8. Maser assembly in the laboratory, showing the electromagnet and double dewar liquid helium system.

The liquid helium refrigeration system may be one of two types. First, there is the conventional double dewar system as shown in Figs. 3 and 8. It has to be topped up periodically (say, every 24 hours) from a storage vessel, so that at least one trained technician would have to attend regularly. The initial cost of this system is comparatively small amounting to about £2000. Against this running costs are high. About 2 litres of liquid helium would be consumed daily, at £5 per litre. Adding to this the cost of transport to and from a fairly remote liquefier and the services of a technician, a sum of £5000 per year is easily reached.

The second type of cooling system is a closed, properly regulated, miniature helium liquefier, permanently containing the maser structure. Equipments of this type are under development and it appears that their cost is likely to be of the order of £10 000. Should they prove reliable and not requiring frequent and expensive maintenance, they would present a more economical proposition than the conventional system, for prolonged continuous operation. In addition, a maser using the closed system would be much easier to install with an aerial having a moving focus.

The d.c. magnetic field required by the maser may be provided by an electromagnet or a permanent magnet. The former is more bulky and must be provided with a stabilized current supply of several amperes capacity. The latter is smaller but permits operation on a fixed carrier frequency only. If it is desired to change the operating frequency over a band, an electromagnet will have to be used. The cost of either type would be of the order of £1000.

A less conventional means of setting up a magnetic field has been tried,[6] utilizing the superconductivity of niobium at liquid helium temperatures. A circulating current of many amperes can be established, in a solenoid made of niobium wire, with the help of a simple circuit and a small battery, and it will maintain itself undiminished for considerable periods of time. Since a low temperature medium is available with a maser anyway, this attractive method may prove to be the best solution.

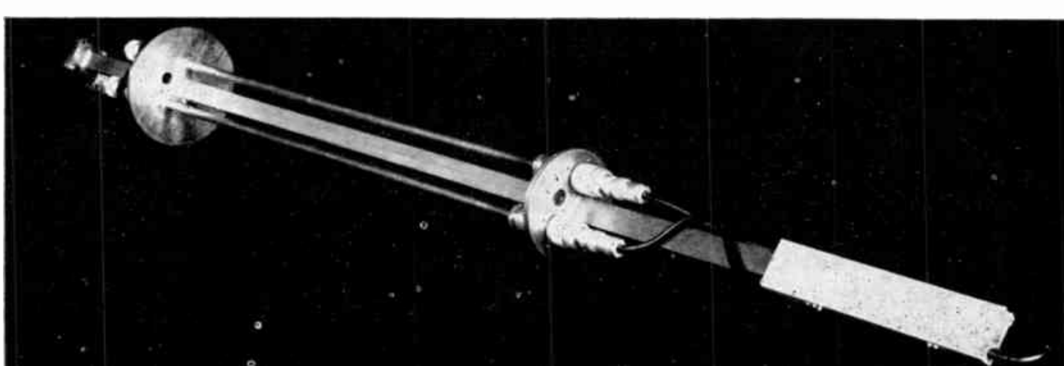


Fig. 9. Travelling wave maser assembly, ready for insertion into a liquid helium dewar.
Note the coaxial input and output lines and the rectangular waveguide carrying pump power into the structure housing.

The pump source may affect the gain stability of the maser in two ways. If insufficient pump power is provided, the populations of levels 1 and 3 may not become completely equalized. Under such conditions slight variations in pump power output will cause fluctuations of maser gain, as this depends directly on the excess population in level 3 over level 2. The pump power level will, therefore, have to be stabilized or made sufficiently high to ensure saturation. Pump frequency fluctuations are another cause of gain instability, as they shift the pump power away from the paramagnetic resonant frequency, and absorption between levels 1 and 3 is reduced. Hence frequency stabilization of the pump oscillator will be necessary. Bearing in mind these complications the cost of a pump source is unlikely to be below £1000.

6. Conclusion

The travelling wave maser is comparatively easy to construct and install on centimetric frequencies corresponding to the microwave "space window". Under these conditions its low noise figure can be put to good use, as a result of low sky noise, provided it is installed in the focus of a highly directive aerial, whose elevation above the horizon exceeds 20 deg. Installation and operating problems will be substantially reduced if an aerial with a stationary focus is used. The cost of a maser installation including a closed refrigeration system is likely to be of the order of £20 000. The use of a double dewar liquid helium system would reduce this figure substantially, but running costs might be higher.

7. Acknowledgments

Thanks are due to Dr. R. J. Benzie for suggesting a number of improvements to the text. The paper is published by permission of the Director of Research, Marconi's Wireless Telegraph Co. Ltd.

8. References

1. N. Bloembergen, "Proposal for a new type solid state maser", *Phys. Rev.*, **104**, p. 324, October 1956.
2. "Low noise for microwave communication", *Engineering*, **191**, p. 400, 17th March 1961.
3. R. W. De Grasse, E. O. Schulz-DuBois and H. E. D. Scovil, "The three-level solid state travelling-wave maser", *Bell Syst. Tech. J.*, **38**, p. 305, March 1959.
4. D. C. Hogg, "Effective antenna temperatures due to oxygen and water vapour in the atmosphere", *J. Appl. Phys.*, **30**, p. 1417, September 1959.
5. R. W. De Grasse, D. C. Hogg, E. A. Ohm and H. E. D. Scovil, "Ultra-low-noise measurements using a horn reflector antenna and a travelling wave maser", *J. Appl. Phys.*, **30**, p. 2013, December 1959.
6. S. H. Autler, "Superconducting electromagnets", *Rev. Sci. Instrum.*, **31**, p. 369, April 1960.

Manuscript received by the Institution on 15th June 1961.
(Paper No. 671.)

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied. Information on translating services will be found in the Institution publication "Library Services and Technical Information".

RESONANT BACKWARD-WAVE OSCILLATOR

A paper by a German engineer has recently described a resonant backward-wave oscillator with a periodic delay line for the frequency range of 30 to 38 kMc/s. The electron gun of the tube produces a cylindrical electron beam with a current intensity of 15 to 20 mA which is focused through the beam zone of the delay line by means of a permanent magnetic field. The rugged delay line is only 8 mm long. The end facing the cathode is short-circuited, its other end is tightly coupled to a rectangular waveguide section of variable length by way of a matching element and a vacuum window. The frequency of the tube is adjusted outside the vacuum envelope by means of a tuning slide which terminates the waveguide section. The generated r.f. power is extracted through the gap between the slide and the wide face of the rectangular waveguide. Corresponding to the dispersion curve of the delay line, which is terminated without reflection, the line voltage must be adjusted with the tuning slide as well. In the tunable range the tube gives a c.w. output of more than 200 mW with an efficiency of 1 to 1·5%.

"A resonant backward wave oscillator for the frequency range 30–38 Gc/s", F. Gross. *Archiv der Elektrischen Ubertragung*, **15**, pp. 227–33, May 1961.

WIDEBAND PARAMETRIC AMPLIFIERS

To obtain broad bandwidth, staggered operation of doubly-resonant variable-capacitance parametric amplifiers in the 6 Gc/s band has been studied by engineers in the Nippon Telegraph and Telephone Corporation. A doubly-resonant amplifier has been achieved by inserting a resonant stub into a waveguide cavity. The experimental results, especially the improvement of bandwidth characteristics by adjusting the phase difference between the pumps of two parametric amplifiers connected in cascade are reported. For a gain of 29 dB, a bandwidth of 55 Mc/s has been obtained.

"Staggered operation of doubly resonant parametric amplifiers", S. Hamada and H. Mukai. *Review of the Electrical Communication Laboratory (NTT)*, **9**, pp. 22–5, January/February 1961. (In English).

INFORMATION THEORY AND AIR TRAFFIC CONTROL

A paper by a German radar engineer shows that the information rate of the radar information from aircraft is extremely small. He contrasts this figure with the high information rate which is handled by the complex three dimensional radar systems which are being planned. It is suggested that a solution to the problem of air traffic control may lie in a much simpler radar system when the aircraft flying in the controlled space observe certain flight rules.

"The problem of civil air traffic control on the basis of the information theory", H. Meinke. *Nachrichtentechnische Zeitschrift*, **14**, pp. 273–8, June 1961.

THERMOELECTRIC COOLING

A laboratory cooler has been developed by a Canadian company which uses standard thermoelectric modules and can provide temperatures down to −20° C. It will hold an 800-ml beaker and keep the contents below 0° C. A simple modification enables the unit to be used as an oven, providing temperatures up to 100° C. The paper describes the design and discusses the solutions to problems of insulation and power supply.

"Thermoelectricity permits new design concepts in cooling equipment", J. Keane. *Canadian Electronics Engineering*, **5**, pp. 48–50, May 1961.

IMPROVEMENTS IN S.S.B. TRANSMISSION

A system has been proposed for single sideband transmission which is based on the analysis of a speech signal into two components which determine its amplitude and relative phase angle. The amplitude on its own contributes no intelligence. On the contrary, the whole intelligence is conveyed by the value of the phase angle cosine. This characteristic allows the generation of an s.s.b. signal limited in amplitude to the extent that its dynamic range from average to full modulation is reduced to 1 dB. It also permits non-linear amplification of this signal. Comparative tests show that, everything else being equal, the received intelligibility of a complex signal transmitted through a constant amplitude s.s.b. system is notably greater than through an s.s.b. system where the signal amplitude is not limited. This gain can reach the equivalent of a 13 dB increase of transmitter power.

"Constant level speech in single sideband transmitters", J. Daguet and K. Gilabert. *L'Onde Electrique*, **41**, pp. 498–509, May 1961.

BENT CIRCULAR WAVEGUIDES

A bent circular waveguide for the transmission of H_{01}-mode waves has been designed at the Heinrich Hertz Institute, Berlin. This bend contains a periodically arranged dielectric with surfaces perpendicular to the axis of the waveguide and the operation depends on the anisotropic dielectric properties of the waveguide loading. The method of measuring the modes excited and present at the output of the bend is based on a measurement of the magnetic field at the wall of the waveguide. For this reason the electro-magnetic energy is coupled through a circular aperture in the wall into a rectangular waveguide with a crystal diode. The axis of this waveguide extends through the centre of the coupling hole. At the output end of the dielectric loaded bend with a radius of curvature of 300 mm, and angle of 100 deg and a waveguide diameter of 50 mm, the measured amplitude ratio is 1 : 0·41 between H_{01} and E_{11} waves at 9315 Mc/s when a pure H_{01}-mode wave is launched. Without loading, but otherwise identical conditions, an essentially pure E_{11} wave emerges.

"The propagation of H_{01}-mode waves in bent circular waveguides", G. Morgenstern. *Nachrichtentechnische Zeitschrift*, **14**, pp. 300–7, June 1961.

THE SECAM COLOUR TELEVISION SYSTEM

There is considerable interest among television engineers in the French system devised and developed by Mr. Henri de France known as the memory sequential transmission process, or SECAM. In a recent paper two French engineers give an example of a decoder and describe the functions provided. The examination of the performance obtained, compared to that of the N.T.S.C. system, claims substantial superiority of the SECAM as regards quality and economical running.

"The SECAM colour systems compared against the N.T.S.C. system", P. Cassagne and M. Sauvanet. *Annales de Radioélectricite*, **64**, pp. 109–21, April 1961.

F.M. RECEPTION

The sum of two frequency-modulated signals approximately follows the phase of the stronger signal. Consequently, for reception of disturbed f.m. signals, during the moments when the disturbance is stronger than the desired signal, it is virtually only the disturbance that is detected. By taking into account the "anomalous" interference thus detected, a familiar formula for the signal-to-noise ratio is described in a Dutch paper which can be extended to the case of a relatively high interference level. To do this, it is necessary to know the statical behaviour of the amplitude and that of the phase of a noise disturbance. Both are derived from a somewhat simplified model of such a disturbance. A formula is found which describes the effects satisfactorily.

"F.M. reception under conditions of strong interference", J. van Slooten. *Philips Technical Review*, **22**, pp. 352–60, 1960–61.

CRYSTAL HOLDERS

Broadband microwave measurements often require low level crystal detectors with high sensitivity and low, uniform v.s.w.r. A recent Canadian paper describes two coaxial holders for cartridge type crystals similar to the 1N23B. Measurements were made over a frequency band 1000–10 000 Gc/s. The first holder has a relatively low and uniform v.s.w.r., while the second holder has a higher sensitivity over the lower part of the band. The design emphasizes simplicity of construction and the use of modified commercial coaxial connectors.

"Simple holders for crystal detectors feature low v.s.w.r., high sensitivity", A. Staniforth and J. K. Pulfer. *Canadian Electronics Engineering*, **5**, No. 7, pp. 31–4, July 1961.

CERAMICS FOR MICROWAVES

A survey by a research professor at the University of British Columbia indicates some of the ways in which low loss ceramics may be employed in u.h.f. electron tubes and other microwave devices. The applications envisaged are still no more than proposals and must remain so until adequate evidence on reliability has been obtained from comprehensive laboratory tests. On one hand, there are theoretical advantages, which offer yet further avenues for speculation. On the other hand, there is a lack of information on practical matters and an entirely new technology has to be developed. Perhaps the most surprising feature is the very wide range of microwave devices for which ceramic techniques can be considered. Two only are discussed here, namely, an electron linear accelerator and a millimetre wave travelling wave tube, design studies on which are at present being carried out at the University of British Columbia.

"New uses for ceramic materials in microwave tubes", G. B. Walker. *The Engineering Journal of Canada*, **44**, No. 7, pp. 46–50, July 1961.

PULSE-SLOPE MODULATION

The characteristics of pulse-slope modulation have been explored at the Indian Institute of Technology, Khargpur, in experiments on the distortion, crosstalk and noise characteristics. Maximum harmonic distortion is about 2% for slow cut-off rate of the medium and about 5% for sharp cut-off filters. It is found that slicers introduce more distortion and non-linearity in modulation. Crosstalk ratios are better than those in p.a.m. for slow cut-off rate and improvement in crosstalk may be affected by simultaneous introduction of h.f. and l.f. cut-offs. Output S/N ratios show considerable threshold effect and are approximately proportional to the square root of the video bandwidth.

"Transmission characteristics of pulse-slope-modulated signals through band-limited systems", J. Das. *Indian Journal of Physics*, **35**, No. 5, pp. 245–60, May 1961.

COLOUR TELEVISION CONVERSION CIRCUITS

The conversion of an N.T.S.C. colour signal into one with a different chrominance carrier has been described in a paper by a German engineer. A double mixing process is required and only two out of 16 possible signal combinations can be used. A practical design of the converter is described. Tests have shown that the picture quality after conversion of the chrominance carrier is not substantially reduced in comparison with the quality of pictures transmitted directly. Only the reduced resolution as a result of bandwidth restriction in the luminance channel of the converter is noticeable. Further tests have revealed that practically no crosstalk interference is noticeable in normal receivers when the bandwidth for the colour TV signal is made larger at the transmitting end that it is intended at the receiving end because of restrictions of the video bandwidth.

"The conversion of an NTSC colour tv signal into a signal with a different chrominance carrier", H. Gorling. *Nachrichtentechnische Zeitschrift*, **15**, No. 7, pp. 336–44, July 1961.