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 OCTOBER 1961 NUMBER 4

TAKING ADVANTAGE OF OPPORTUNITY

ALTHOUGH the Institution arranges special meetings or Conventions during the summer months, the regular meetings for members in London and Local Sections in Great Britain are planned for the period from October to May. The beginning of autumn is therefore an appropriate time to look at the programme of meetings and to note the opportunity provided by the Institution for its members to discuss current work and ideas and to keep abreast of development in fields other than their own. The booklet giving details of the meetings arranged in Great Britain up to the end of January shows that the papers to be read and discussed cover a very wide range of subjects catering for the general as well as the specialized interests of the radio and electronic engineer.

The one- or two-day symposiums arranged by the Institution in London and at local centres during recent years have proved to be of great value to members. During the coming session opportunity will be available for members to attend symposiums on data transmission, electronic aids for the handicapped, digital differential analysers, adaptive optimizing control, industrial electronics and electronic techniques in non-destructive testing. Subjects of evening meetings include microwave valves, colour television, inertial navigation, radio astronomy, echo-location, reliability, tunnel diodes, and nuclear instrumentation.

In the past the programme of London meetings was mainly the responsibility of the Programme and Papers Committee. The formation of specialist groups has, however, provided the opportunity for highly technical matters to receive due attention in a way which was difficult to arrange under the old, and more central, control. To the five Groups which have operated during the past two years, Council has now authorized the addition of an Industrial Electronics Group and arrangements are being made for the inaugural meeting to be held during the 1961–62 session. Thus almost every member will find his special professional interests covered.

The radio and electronic engineer must, however, keep in touch with fields other than his own. The Programme and Papers Committee is well aware that this becomes difficult with the increasing specialization of the Institution's proceedings. It has therefore arranged several "general meetings" during the session which, because of their breadth of approach, will be of interest and value beyond the needs of the specialist.

The 1961–62 Session will therefore be one of great activity, and with a programme of 80 meetings, 35 of which will be held in London, it is hoped that every member will attend at least one meeting.

In the editorial entitled "The Next Stage" in the September *Journal* mention was made of the desirability of the Institution possessing its own lecture theatre. This brief review of the coming session emphasizes the vital need for a headquarters building appropriate to the standing and activity of the Institution. It is clear that, if the Institution is to make the most of its opportunity as the newest Chartered body, and if the facilities for members are to continue expanding in the way which we have come to expect, this problem will have to be solved in the near future.

A. D. B.

B

INSTITUTION NOTICES

Letters Patent

The Home Office has now delivered to the Institution the Letters Patent containing the Grant of the Charter of Incorporation. This takes the form of a seven-page document engrossed on parchment, bearing the impression of the Great Seal and dated 31st August, 1961.

The Cost of the Brit.I.R.E. Journal

During recent months the cost of printing and materials has steadily increased, and in addition postage rates have been raised with effect from the beginning of October. The impact of these increased costs on the enlarged *Journal* has lead the Institution to decide, with reluctance, that the subscription rates to libraries of universities, industrial organizations, etc., which have remained unchanged since 1957, will be increased to £5 10s. for one year (12 issues). These rates take effect at once for new subscribers and renewals of existing subscriptions.

The charge to non-members for single copies of the *Journal*, which has been 7s. 6d. since 1945, is increased to 10s. 6d. per copy. Members will continue to be able to obtain single copies to complete volumes for binding at a cost of 5s. each.

The Institution and Non-Destructive Testing

Testing is one of the most important features of industrial production. In the radio and electronics industry, electrical testing can provide many of the answers. In heavy industry, however, other non-destructive testing methods have been developed and this is a field of activity of great importance.

The Institution has recognized this by its membership of the British National Committee which in turn is a member of the Standing Committee for International Co-operation in the field of Non-Destructive Testing. An article on the British National Committee has been contributed by the Institution's representative, Dr. A. Nemet (Member), and appears on page 311 of this *Journal*.

The West Midlands Section of the Institution is now organizing a one-day symposium on this subject which will be held on 6th December at the Wolverhampton and Staffordshire College of Technology, Wulfruna Street, Wolverhampton. The programme will include papers on eddy currents, magnetic flux leakage, photo-electrics, ultrasonics, x-rays and signal data processing. Further details will be published in the November *Journal*; requests for registration forms should be sent to the Local Hon. Secretary, Major C. W. Weech, M.Brit.I.R.E., 5, Shelton Fields, The Mount, Shrewsbury.

The Secretary's Visit to the U.S.A.

The General Secretary of the Institution, Mr. Graham D. Clifford, is now in Canada where he is attending meetings of members, visiting Universities, Government establishments, etc.

He will start his tour of the United States on 20th October and his itinerary includes visits to Chicago, New York, Princeton, Philadelphia, Los Angeles, San Francisco and Washington. Mr. Clifford returns to England towards the end of November.

The Council of Scientific and Industrial Research

The Minister for Science, Lord Hailsham, has appointed five new members to the Council for Scientific and Industrial Research. The new members will serve for a five-year term and include Mr. Leslie H. Bedford, C.B.E. (Member), a Vice-President of the Institution. Mr. Bedford who was President of the Institution in 1948–50, is Director of Engineering, Guided Weapons Division, English Electric Aviation Ltd.

Proposed Convention in India

During his recent visit to London, Major-General B. D. Kapur, Chairman of the Institution's Indian Advisory Committee, discussed the possibility of an Institution Convention being staged in Delhi or Bangalore in February, 1963.

It is proposed to follow the pattern of Brit.I.R.E. Conventions by having a number of sessions. There will not, however, be a single theme, the proposal being that each session might deal with various forms of communication and industrial electronics in those particular fields which offer earliest promise of development in India.

Members expecting to visit the Far East in the early part of 1963 are invited to communicate with Major-General B. D. Kapur, B.Sc., M.Brit.I.R.E., Chief Controller, Defence Research and Development, Ministry of Defence, New Delhi, India, or Mr. F. W. Sharp, Assistant Editor, *J. Brit.I.R.E.*, 9, Bedford Square, London, W.C.1, if the member is able to offer or suggest a paper or wishes to attend the Convention.

Index to Volume 21

The Index to Volume 21 of the *Journal* (the first half of 1961, January to June) has now been prepared and copies are being sent with this issue to all members and subscribers.

Members are reminded that they may send their *Journals* (six issues plus index) to the Institution for binding. The charge for this service is 16s. 6d., postage and packing extra (Great Britain 3s.; other countries 4s.).

U.D.C. 629.19

Critical Engineering Factors in the Design and Development of Space Systems

By

J. M. BRIDGES†

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

Summary: The paper deals with the basic concepts underlying the design and development of systems for space operations. The importance of reliability and weight as system parameters, particularly in regard to cost, performance and development time schedules are discussed. The peculiar needs of a space system are shown to affect the design philosophy, the ground test programme and the developmental approach to an optimum system. Some suggestions are made regarding the techniques and procedures to be used in design and development to ensure that the system will operate satisfactorily on its mission in the space environment. The paper also points out certain work that is needed in basic research before we can design the reliable space systems that will enable our conquest of space.

Introduction

There are within the U.S. space programme many interesting developments in research and engineering that have both military and scientific objectives. From the military viewpoint, our current interests in space relate primarily to communications, navigation, surveillance and early warning of ballistic-missile firings. Civilian interests, of course, are directed more toward scientific exploration and commercial communications. Our programmes abound in new and unique technical problems of concern to people in communications and electronics work. These problems range over a very broad spectrum that includes communications, telemetry, guidance and control, infra-red and physics of space and the upper atmosphere. Common to all these programmes, however, are engineering problems associated with obtaining high reliability, long life and lightweight electronic devices for space vehicles.

I have elected to discuss some of the critical engineering factors associated with designing and developing spaceborne electronic systems to meet these objectives. Although I realize that there may not be as broad interest in this particular area as in other aspects of space communications-electronics, I feel that it is in this somewhat less glamorous area of space technology that I can contribute most to this meeting.

Reliability and Life

I am convinced that the limiting technical factor in our conquest of space may eventually be the lifetime of the equipment and devices carried in space vehicles.

† Director, Office of Electronics, Office of the Director of Defense Research and Engineering, Washington, U.S.A.

This may be evidenced differently in the various space missions. For example, with respect to an earth-satellite communications system, the lifetime of the satellite and its equipment directly influence the system's economic and operational feasibility. For interplanetary missions or deep-space probes, which may involve flight times of months or even years, the equipment's operating lifetime may be become a physical limitation. In manned space missions, the maintenance of human life can be directly related to equipment reliability.

Reliability and lifetime have important economic implications with regard to orbital systems. Let us consider, for illustration, a system in which six satellites must be maintained in orbit to perform a certain function—perhaps world-wide commercial communications. How does the cost of establishing this system relate to the reliability or lifetime of the equipment carried in the satellites?

Let us assume that the system requires (1) a 90% probability that its six satellites will be functioning properly in orbit at the end of one year, (2) a 75% probability of attaining orbit per launch and (3) an interval of one week between launchings. Then, if each satellite had a Mean-Time-to-Failure of one year, we would have to launch 23 satellites. (A one-year Mean-Time-to-Failure, or m.t.t.f., is a 37% probability that a system will operate for one year.) If the operating life of the satellite and its equipment could be extended to a two-year m.t.t.f., we would only need to launch 16 of them; in other words, doubling the m.t.t.f. of the satellite would save seven launchings. At current prices of $5 to $10 million a shot, this would represent a saving of $35 to $70 million. Obviously, on this basis, it is good economy to spend fairly large sums of money on research and engineering to improve system reliability.

For deep space probes or interplanetary flights, the requirement for long equipment life is not merely a matter of economy; it is essential for mission accomplishment. Using current propulsion systems, a one-way trip to Mars would take about 250 days. To attain a 99% probability of completing the mission, with all sensing and communications devices still working satisfactorily, would require a m.t.t.f. of some 25 000 days, or approximately seven years.

I believe these examples clearly indicate the economic and practical importance of reliability in our space efforts. They also show that for space operations, we must gain an equipment lifetime several orders of magnitude greater than is needed for any other application.

Now let us consider some of the research and engineering problems we encounter in seeking to reach the degree of reliability that will satisfy the new and rigorous requirements of space missions.

For many years I have been deeply interested in the serious problem of obtaining greater reliability in military electronics and have taken an active part in setting up reliability-improvement programmes in my country. As a result of this experience, I have become intensely concerned about this aspect of our space programmes.

Although reliability in military electronics has improved greatly over the past several years, we still lack components—and methods for testing them as well—of a quality and design that will enable us to build electronic devices that will operate over the time period demanded by some advanced space programmes now being developed or planned.

In considering the engineering approaches to the solution of these problems, we must first recognize the rather fundamental differences between space electronics and electronics for other purposes. It is because of these differences that we must develop and employ new design philosophies, new test methods and new components.

The most important feature unique to space electronics is that equipment must operate for very long periods without manual adjustment or maintenance. This requirement alone dictates the application of fundamentally new design concepts and components. In designing most electronic systems for military uses, we can make engineering compromises between reliability and such factors as length of mission, the amount of preventive maintenance that is practical or economically feasible, and the degree to which operating adjustments can be made during the mission.

In airborne electronics, for example, where mission time is on the order of a few hours, it is feasible to schedule routine or preventive maintenance, and some operating adjustments can usually be made during the mission to optimize performance. Under these conditions, equipment can be designed with a m.t.t.f. of a few hundred hours, which gives a fairly high probability that the mission will be successful. Overall equipment life can be related to such factors as military or technological obsolescence and the economic balance between depreciation and replacement cost, instead of the mission's length.

In space applications, these opportunities for compromise are not present, and the primary reliability requirement is long operating life. The need for reliability in guided missiles bears a little closer resemblance to that of space vehicles—but not too close. Missiles are expendable; there can be no adjustments during the mission; and some of the environmental conditions (such as shock, vibration and temperature) approximate those of a space vehicle's launching phase. But the mission lengths of guided missiles differ radically from those of space vehicles, the missile flights being measured in seconds or minutes, while space missions extend to thousands of hours or even years. So the chief reliability requirements for missiles are an extremely high probability of satisfactory operation for a short interval of time, long storage life and quick check-out before firing.

In some i.c.b.m.s, reliability must more nearly approach that of a space system. The *Minuteman*, for example, is being designed to operate after being continuously in its silo for periods as long as a year with only routine checks. But even one of these missiles can fail without being permanently lost to the force, because the operating personnel will have been warned of its failure. There would be a real loss only if the missile were to fail just prior to an order for its launching in an all-out general war. So we see that, in contrast to most space ventures, long life in missiles is not entirely controlling. The *Minuteman* philosophy of continuous operation, however, dictates an objective m.t.t.f. of around 25 000 hours. If its objectives are attained, this missile-development programme will contribute many new components and design and test techniques that will be applicable to space programmes—at least to those of the first generation.

Designing for Reliability

A second difference between space and other electronics applications exerts a significant influence on design for reliability. This is the fact that space systems are produced in relatively small numbers, as compared with most other types of electronic devices, and it is my guess that they always will be.

In designing electronic devices that are to be produced in quantity, the developer has tended to rely—perhaps too heavily—upon the possibility that design deficiencies can be corrected or inadequate components replaced during production, or that his design mistakes can be rectified by field modifications during the equipment's service life. In many of our guided-missile programmes (*Nike* and *Terrier*, for instance), missiles have been produced in quantities of several hundred for developmental firings. Designers have used the information gained in this production experience and the numerous early firings to improve equipment design and to test its reliability.

The practice of rectifying design deficiencies during production on the basis of data obtained from quantity test firings has to a certain extent relieved the engineer of pressure to do a thorough initial job of design, and unfortunately it has tended to delay recognition of the need for thorough reliability engineering.

In the design of space vehicles and their equipment this practice is not valid because there is no production; also, the high cost of launching makes it impractical to depend on that procedure for determining design errors or testing equipment reliability. The system and its components must be correctly designed in the first place and its performance and reliability proved, as far as possible, before the first launch is attempted.

This means, of course, that the designer must place much more emphasis on those phases of engineering and test that are vital to obtaining and measuring reliability. When a space system's initial requirements are being established, reliability—particularly operating lifetime—must be considered just as thoroughly as any other performance characteristic. In this early phase of the programme a quantitative reliability goal must be established that is consistent with the functional objectives of the space mission. Once defined, this reliability requirement must be adhered to throughout the programme as rigorously as any other established performance characteristic.

Test Programmes

To this end, the most advanced reliability techniques must be applied throughout all phases of the programme—initial system studies, design, development and construction. This involves, first, the mathematical prediction of system reliability, using known parts-failure rates. These calculations will reveal critical areas in which redundancy is required or where new long-life components will have to be developed. Then, as system development proceeds, the predicted reliability should be checked by periodic design reviews and analyses. These activities should be supported by a well-designed test programme—both at ambient conditions and simulated launch and space environments—that will determine quantitatively whether the over-all design is meeting the reliability objectives for the system. The tests should encompass all levels of system assembly, from materials through component parts and sub-systems to the entire space-borne system.

Enough samples should be tested at each level of system assembly to provide statistically valid assurance of the entire system's Mean-Time-to-Failure. Very likely this will require the construction of several more sub-system assemblies or even complete systems—than are needed to meet the firing programme. This is expensive, but (as I pointed out earlier), the cost of constructing several payloads for test purposes may be considerably less than the cost of a single aborted launch or the successful launch of a "dead" vehicle. I also appreciate that test facilities, particularly those needed to simulate space environment, are extremely costly. However, a laboratory or a contractor will not be able to stay in space R. and D. for long without these facilities—no longer than he would stay in the business of designing shipboard electronics without shock and vibration testers.

If the reliability analyses and ground tests indicate that final system reliability is substantially below design goals, the entire programme should be reconsidered. If the situation is bad enough, it may be necessary to revise the schedule so that better components can be developed or the design improved; it may be necessary to reduce the system's complexity to satisfy reliability requirements. It may be justifiable to place in orbit a space system with marginal life characteristics in order to test the feasibility of the fundamental system concepts. We are deluding ourselves, however, if we continue too far along toward an operational system or a tactical commitment before we are reasonably certain that we can achieve the space-vehicle life needed to make the system economically and operationally feasible. This is especially true of an orbital system in which several satellites must be simultaneously kept in orbit. I am not sure that this programme philosophy is too well recognized in the management of some of our space programmes.

Future Research on Reliability

Before I leave the subject of reliability in space electronics, it might be well for me to mention a few areas in which research is sorely needed to gain the knowledge and provide the components that will enable us to design these highly reliable equipments and systems and, as I have advocated, to predict and measure reliability.

First, we need to accelerate and expand our efforts to develop extremely long-life component parts and materials optimized for the space environment. We need an expanded research and measurement programme, aimed at providing a better understanding of the total space environment and its effects on components and circuits. We need a lot of research on the failure mechanisms of electronic components so that we can predict, even before a component has been tested, what its life characteristics will be under certain conditions of use. We need to develop techniques for performing accelerated life tests, particularly for those components with a very low failure rate—0·001% per 1000 hours or better. With present methods these tests require too large a sample and too long a testing interval.

We need research to find better methods for quantitatively specifying and measuring reliability of complex systems for which an extremely long m.t.t.f. is required. This, incidentally, is a need that applies not only to space programmes, but to some of our very complex radar and command and control systems.

We need more research on techniques of self-healing for components and circuits, advanced methods employing redundancy, fail-safe techniques and self-organizing systems. We need a lot of R. and D. work to develop better techniques for isolating component parts and critical sub-assemblies—possibly entire electronic systems—from environmental extremes imposed by temperature cycling, radiation and acceleration.

Finally, we need much more effort on the development of techniques for sealing and lubrication in high vacuum.

Weight Reduction

Now there is another engineering problem closely associated with reliability that exerts a particular (if not unique) influence on space programmes. It is the critical matter of equipment weight in space applications, a factor of great importance because it is directly related to rocket thrust. Not only does this affect the cost of launching, but perhaps the time scale for attaining a certain space objective.

I have no precise figures relating the cost of launching rockets and boosters to payload weight, but you may gather some idea of the monetary value of weight reduction from some other pertinent figures.

It is estimated that, for an i.c.b.m. such as *Atlas*, the ratio of payload weight to gross weight of the rockets on the launching pad is somewhere around 80 to 1. It is also estimated that the cost of such a rocket system is around $40 per pound. One can assume from these figures that each pound saved in payload weight would result in a saving of around $3200 in the cost of launching a space vehicle. Admittedly these are rough figures, but they certainly illustrate the economic value of reducing system weight.

Perhaps a more concrete illustration is the current programme of weight reduction in our *Transit* navigation satellite. The present satellite weighs around 190 lb and requires a *Thor-Able-Star* rocket combination to inject the vehicle into a 600 nautical mile orbit. Engineering redesign is now underway to reduce the weight of the vehicle from 190 lb to a weight such that the cheaper *Scout* rocket can attain a 600 nautical mile orbit. This weight reduction is being accomplished through further miniaturization of the electronics package, optimization of mechanical construction and improvement in efficiency of the solar power supply. If this objective can be met the use of the *Scout* rocket will represent savings of many millions of dollars in maintaining an operational navigation system using this satellite.

Savings of this magnitude clearly justify the expenditure of a substantial amount of money to advance the day when techniques of microminiaturization solid-state and so-called molecular electronics are available for space applications. In fact, both the requirements I have discussed—extremely long mission life and light weight for space systems—clearly point out the need for a greatly accelerated research programme in these promising new areas of electronics technology.

In its rapid development, the space age has made unprecedented demands upon the electronics technology that we are not fully prepared to meet. For example, to satisfy the requirements of a satellite having 24 000 electronic parts and a m.t.t.f. of one year, the average failure rate of the components must be of the order of 0·0005% per 1000 hours. Contrast this with the average failure rate of currently available electronic parts—0·1% to 0·01%. Until we have electronic components with the desired life characteristics, space requirements can only be filled by applying the principles of redundancy and component derating. And both these techniques add weight, which calls for greater rocket thrust and so costs more money. In some of our space programmes we are actually faced with the necessity for making a functional compromise in order to reduce the systems' complexity and weight to a point where the vehicles can be launched by the available thrust.

Standardization of Components

The last point I would like to make has to do with standardization.

It is going to cost a lot of money to carry on the research programmes in reliability and integrated circuits and to develop long-life components. These activities must be supported by all agencies that have space programmes.

Furthermore, since there will be relatively few space vehicles in our total national programme, it will be mandatory to use the specially developed long-life components to the maximum extent. This is absolutely essential to create enough of a demand to justify a parts producer in setting up the engineering, production and test facilities required to make a specialized long-life part. Widespread use of these parts by all developers of space equipment will also save money; let me cite an example.

In a recent programme to develop an extremely complex airborne electronic system, the contractor made a major effort to standardize parts through the entire system, which involved four different contractors. The system contains 60 000 electronic parts, but the resulting design called for only 400 different types of parts. If there had been no parts standardization, this system would have had at least 1000 different kinds of parts. At a cost of $5000 for testing a single part type, it is estimated that the standardization achieved a net saving of around $3 million.

In space applications the pay-off in standardization is not so much the dollars saved as the improvement in the quality and reliability of parts that will result from the development and production of fewer types.

Conclusions

Let me quickly recall the points I have been trying to bring out:

First, I outlined the economies in space programmes that can arise from improved life characteristics. Next, the fundamental differences between spaceborne and other electronics exert a significant influence upon design philosophy and methods. These are primarily that space systems require extremely long mission life and unattended operation, have very small production requirements and are very costly to test in actual operation.

I have emphasized the necessity of doing a correct engineering job the first time, placing the same emphasis on requirements for reliability that other performance characteristics receive. If engineering analysis and reliability testing during system development show that reliability goals are being missed by a substantial margin, the programme's schedule should be changed to allow the development of better parts and improved design.

I have also emphasized the extreme importance of weight reduction of space payloads, spoken of the need for an accelerated and expanded programme of research on solid-state and micro electronics.

Now I believe that I can summarize all I have said in this statement:

The ability of the electronics scientist and engineer to provide electronic devices that have the operating lifetime and the weight required for space missions is probably a more critical factor in our conquest of space than our ability to produce thrust or develop guidance techniques. My personal opinion is that our future in space exploration depends fundamentally upon our progress in the new area of technology in micro and molecular electronics.

Manuscript received by the Institution on 5th July 1961 (Address No. 31).

APPLICANTS FOR ELECTION AND TRANSFER

As a result of its meeting on 28th September 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

ABERCROMBIE, Stanley Douglas. *Tadley, Hampshire.*
TRIER, Peter Eugene, M.A.(Cantab.). *Reigate, Surrey.*

Transfer from Associate Member to Member

CLARK, Wing Commander Ian Birket, M.B.E., R.A.F. *Devizes, Wilts.*
HUGGINS, Peter. *Lichfield, Staffs.*
MASKELL, Wing Commander Nelson William, R.A.F. *B.F.P.O.33.*
THOMAS, Lieutenant Colonel Dennis, M.B.E., R.Sigs. *Beckenham, Kent.*

Direct Election to Associate Member

*COURT, Frederick Charles. *Cookham Village, Hampshire.*
CRIBB, Major Anthony, B.Sc., R.E.M.E. *London, S.W.1.*
HEYWOOD, Peter John, B.Sc. *Westhoughton, Lancashire.*
HUSSEY, Elliott Douglas Francis. *Walton-on-Thames, Surrey.*
IVES, Cecil George. *Uxbridge, Middlesex.*
*JAGIR SINGH, Wing Commander, B.A. *Amritsar, India.*
KHALIL, Anis Yacoub, B.Sc. *Kuwait, Persian Gulf.*
KINDER, John Dennett, B.Sc.(Eng.). *Bournemouth, Hampshire.*
MARSHALL, Jeffrey. *Crowthorne, Berkshire.*
MILES, Robert John. *Romsey, Hampshire.*
OAKLEY, Captain Denis Frederick, B.Sc.(Eng.), R.Sigs. *Pinner, Middlesex.*
PARSONS, Anthony Philip. *High Wycombe, Bucks.*
*VENKATESWARAN, Squadron Leader Gopalakris-Hnayyar, I.A.F. *Kozhikode, South India.*
WORTHINGTON, Walter John. *Rochford, Essex.*

Transfer from Graduate to Associate Member

BLAKE, Bernard Herbert. *Southampton.*
CHAMBERS, John Albert. *London, N.W.9.*
CHRISTENSEN, Svend Aage. *Paris, France.*
CLARKE, Captain George David, R.E.M.E. *B.F.P.O.34.*
DIVER, Norman William. *Eastwood, Essex.*
EDMONDSON, John, M.A.(Cantab.). *Dollard Des Ormeaux, Canada.*
GRIFFIN, Ronald Frederick. *London, S.E.12.*
HAWKINS, Arthur Goodwin. *Letchworth, Hertfordshire.*
KLIMEK, Andre Mathieu. *Burgess Hill, Sussex.*
O'CONNOR, Bartholomew John. *Dublin.*
PANTHAKY, Jal-Khurshed. *Coulsdon, Surrey.*
WHITE, Nigel John. *Wells, Somerset.*

Transfer from Student to Associate Member

McALLISTER, John Smillie. *London, W.5.*
NEUMANN, Shimon Siegfried. *Ramat Can, Israel.*

Direct Election to Associate

BIRLISON, Flight Lieutenant Ralph Kenneth, R.A.F. *Weston-super-Mare.*
CAMERON, Duncan Hamilton. *Greenock, Scotland.*
HENEGAN, Robert Paul. *Virginia Water, Surrey.*
JOHNSTONE, William Thomas. *Stafford.*
STEELE, Gerald Frederick. *New Malden, Surrey.*

Transfer from Student to Associate

BATES, Gordon Harry. *Aylesbury, Bucks.*

Direct Election to Graduate

ALDRIDGE, Clement Horace. *Malvern, Worcestershire.*
ASHTON, Alfred William. *Orpington, Kent.*
BENISTON, Patrick Thomas. *Newbury, Berkshire.*

Direct Election of Graduate (*continued*)

BESWARWICK, Edward Terence. *London, S.E.12.*
BILLINGS, Keith Hugh. *London, S.E.22.*
BREWER, Michael John. *Leicester.*
CARROLL, Roy. *Liverpool.*
COOKE, Tony Alfred. *Rochford, Essex.*
DESBOROUGH, Colin Leonard. *Harwell, Berkshire.*
DICKSON, John Mabon. *Hockley, Essex.*
DUNCALF, Alan Jesson. *Bolton, Lancashire.*
GREEN, Peter John. *Calcutta, India.*
GRUNDY, James Anthony. *Halifax, Yorkshire.*
HALL, Ronald Basil. *London, S.E.18.*
HARRISON, Victor. *Bolton, Lancashire.*
HEMSLEY, David William. *Pinner, Middlesex.*
HUTCHINSON, Cyril. *London, S.E.2.*
JEFFERY, Raymond Harold. *Ruislip Manor, Middlesex.*
JESTY, Richard. *Felstead, Essex.*
KEIR, Alexander Stuart. *Perth.*
KING, James Brian. *Stanmore, Middlesex.*
KNAPMAN, Ian Wallace. *Plymouth, Devon.*
KUPPUSAMY, Visvanathan. *Perak, Malaya.*
LEE, George Bryan. *Wallasey, Cheshire.*
LORD, John. *Disley, Cheshire.*
MARLOR, James. *Oldham, Lancashire.*
MURRAY, George Robert, B.Sc. *Canterbury, New Zealand.*
PAGE, Norman Walter. *Chelmsford, Essex.*
PEGGS, Bryan Desmond. *Leverstock Green, Hertfordshire.*
PENNEY, Sidney Roy. *Havant, Hampshire.*
PIMBLETT, Joseph Arthur. *Wigan, Lancashire.*
POLLITT, Thomas. *Manchester.*
PRICE, Nevill Penry. *Teddington, Middlesex.*
PROCTOR, Eric James. *Bletchley, Buckinghamshire.*
ROBBINS, David. *Bushey, Hertfordshire.*
ROBERTS, John Winston. *Basingstoke, Hampshire.*
SCHOFIELD, Colin David. *Cheltenham, Gloucestershire.*
SHEDDEN, Donald Arthur. *London, S.W.18.*
SHEIKH, Lieutenant Mohammad Jaffar, B.Sc. *Karachi, Pakistan.*
SHIELDS, James. *Whitehaven, Cumberland.*
SULSH, John David. *Crayford, Kent.*
WREN, James Frank. *Bexley Heath, Kent.*

Transfer from Student to Graduate

ASHEN, David John. *London, E.17.*
BLAIR, Desmond McGavock. *Ampthill, Beds.*
CHAN, Ping Cheung. *Hong Kong.*
GALLAGHER, Thomas. *London, E.12.*
GREEN, Martin Richard. *Farnborough, Hampshire.*
GRICE, William Henry. *London, W.13.*
HOARE, Anthony Michael Edward. *Reading, Berkshire.*
HUGHES, Terence. *Weybridge, Surrey.*
IRISH, Reginald Tom. *Wallington, Surrey.*
JONES, Denis Charles. *London, E.17.*
KAPOOR, Onkar Nath. *Ludhiana, India.*
KHANNA, Kishan K., B.E.(Elect.). *Bombay.*
MEADOWS, William Rowland. *Camberley, Surrey.*
OEI DJONG HAY. *Southampton, Hampshire.*
RICHES, Maurice Herbert. *Chester.*
ROCHE, Francis Gilbert. *Enfield, Middlesex.*
ROGERS, George Brian. *Weston-super-Mare, Somerset.*
TOMASZEWSKI, Eligiusz Eugeniusz Ezechiel. *New Malden, Surrey.*
TUCKER, Jeremy MacIntyre Crutchlow. *London, N.W.11.*
WOOD, Peter Albert Harold. *Cheshunt, Hertfordshire.*

STUDENTSHIP REGISTRATIONS

The following students were registered at the 28th September meeting of the Committee.

ABREU, Joseph Francis. *London, S.E.27.*
ABDUL HAMAD, B.Sc. *Karachi, Pakistan.*
ADEYINKA, Jacob Ore. *Lagos.*
AGASHE, S. N., B.Sc. *Aurangabad, India.*
AHERN, Brian Henry. *Portsmouth.*
AKPAN, Nnamso Udo. *Lagos.*
ANDERSON, William. *Belfast.*
BILLEY, Cleveland W. G. *British Guiana.*
BARDOS, Peter Andrew. *London, W.2.*
BOWSFIELD, Ralph Clement. *Watford.*
BROWN, George Alfred. *Stevenage, Herts.*
CHAMPNEY, Philip D'Arcy. *Newbury, Berks.*
CHANG MENG TENG. *Singapore.*
CHIVERS, Ernest Richard John. *Chippenham.*
COOPER, Robert Alfred. *Leighton Buzzard.*
CORR, Patrick Joseph. *Belfast.*
COUWENBERG, V. W. *Northmead, Australia.*
CRANE, Peter Watson. *Pewsey, Wiltshire.*
CUNNINGTON, Herbert John. *Brisbane.*
DIXON, Donald C. S. *St. Austell, Cornwall.*
EDOHO, Joe Mark. *London, S.E.27.*
EDWARDS, Clement. *London, N.4.*
ELLIOTT, Francis Hamilton. *Belfast.*
FAGENCE, David Alan. *Sutton, Surrey.*
FIGUEIREDO, Adolf Benito S. *London, W.1.*
FOLEY, Edward Thomas. *Dublin.*
FREEMAN, Dudley George. *London, E.7.*
GANNA, Jonathan Ebun. *London, S.W.10.*
GOULDING, Royston David, B.Sc. *Lagos.*
GRIFFITHS, George Henry. *Edinburgh.*
HAMBELTON, David R. *New Malden, Surrey.*
HARRIS, Jeffrey S. E. *Chippenham, Wilts.*
HON WAH LEE. *Hong Kong.*
HUGHES, Ian Reginald. *Beckenham, Kent.*
JOHNSON, Ernest Arthur. *Winnipeg.*
JONES, Peter Arthur. *Malvern, Worcs.*
KAINE, Donald, B.Sc. *Reading.*
KAMATH, Shashikant G. *Mysore, India.*
KAUSHAL, Sub-Lieutenant Dev Dutt. *India.*
KELLY, Gerald Victor. *Gosport, Hampshire.*
LAXMAN, Minsaram. *London, S.W.7.*
LEUNG, Yuk Pui. *Hong Kong.*
MARCHESE, Paul Victor, B.Sc. *Ruislip.*
MONACHAN, Daniel. *Hebburn-on-Tyne.*
MUKHOPADHYAYA, Biswanath. *Nadia, India.*
NAIDU, Dinkar Jayagopal. *Bombay.*
NICHOLLS, Geoffrey Alan. *Kumasi, Ghana.*
PACE, Darryl Brevin. *Kings Lynn.*
PAGET, Frederick G. *Newry, Ireland.*
PARULKAR, Subhash R. *Maharashtra, India.*
PATHMANATHAN, T. S. O. *London, S.E.18.*
PERRIN, Roy Laming. *Manchester.*
PAINE, John A. *Little Chalfont, Bucks.*
PRYKE, Henry Charles. *Sidcup, Kent.*
QUEK, Sunny Teo Chuah. *Hitchin.*
RANDALL, James William. *Portsmouth.*
*RICHARDS, William M. *Weston-super-Mare.*
RIGBY, Gerald Malcolm. *Wallasey, Cheshire.*
RIVERS, Colin Walter. *Plymouth.*
ROBINSON, William. *Grimsby.*
SALTER, Martin Thomas A. *Croydon, Surrey.*
SARDANA, Om Prakash. *London, N.W.6.*
WANG, Saw Thang, B.Sc. *Singapore.*
SCHUSTER, Mark. *London, E.5.*
SHARMA, Chandra Prakash, B.A. *New Delhi.*
SKERRY, Christopher Scott. *Portsmouth.*
SPINKS, John Edward. *Basingstoke.*
STEVENS, William Henry. *Prudnoe-on-Tyne, Northumberland.*
THESEIRA, Maximillian S. *London, S.E.6.*
TREVETT, Eric Henry. *Pembroke, Bermuda.*
TUCHERLEY, Riadh Mustafa. *London, W.2.*
VENKATARAMIAH, Dandinashura Venkateshiah. *Bangalore.*
WALSH, Patrick William. *New Malden.*
WHAN TONG, Hubert C. *London, N.W.6.*
WILLERTON, James. *Malvern, Worcestershire.*
WORMS, Michil P. *Camberley, Surrey.*
ZWETSLOOT, Marcel G. A. *'s Hertogenbosch, Holland.*

* Reinstatements.

U.D.C. 621.396.69

Techniques of Microminiaturization

By

D. H. ROBERTS, B.Sc.†

AND

D. S. CAMPBELL, B.Sc.†

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

Summary: Two approaches to micro-circuits are described and their merits discussed. In the case of thin film circuits it is shown that resistors and capacitors can be made in conjunction with one another over a fairly wide range of values. The use of semi-conductor *p-n* junctions as circuit elements is then discussed and one tentative design of an integrated circuit is given. Low temperature devices are also described. The importance of micro-circuit techniques for space applications is discussed.

1. Introduction

There has been a great deal of discussion in the past few years about microminiaturization. It is the opinion of the present authors that in many fields the important advantages to be gained from some of the techniques used will be—

Improved reliability

Low cost

Simplicity

Speed (for computing applications).

However, when we consider electronic equipment for a space vehicle it is immediately apparent that reduction in size and weight is of prime importance, so it is on this aspect that emphasis will be laid in this paper.

There have been three approaches to microminiaturization:

(a) The R.C.A. micromodule.

(b) The evaporated film micro-circuit.

(c) The semi-conductor solid circuit.

The first of these will be considered no further because it is not considered to go far enough in size reduction, nor to offer the other four advantages referred to above. Accordingly we shall now describe briefly the techniques used and results achieved in (b) and (c), and in addition discuss some more recent developments in low temperature elements.

2. Thin Film Circuits

No attempt will be made to give a complete picture of the field of evaporated components. However, it is hoped to give a summary of the position with regard to resistors and capacitors, with some mention of the range of values and performance that can be expected.

2.1. *Substrates*

Before discussing resistors and capacitors in detail, it would, perhaps, be as well to consider the substrates used. A large number of substrates have been examined by different workers, but of them all, glass seems the most suitable for many applications. The main advantage that glass offers is its extreme

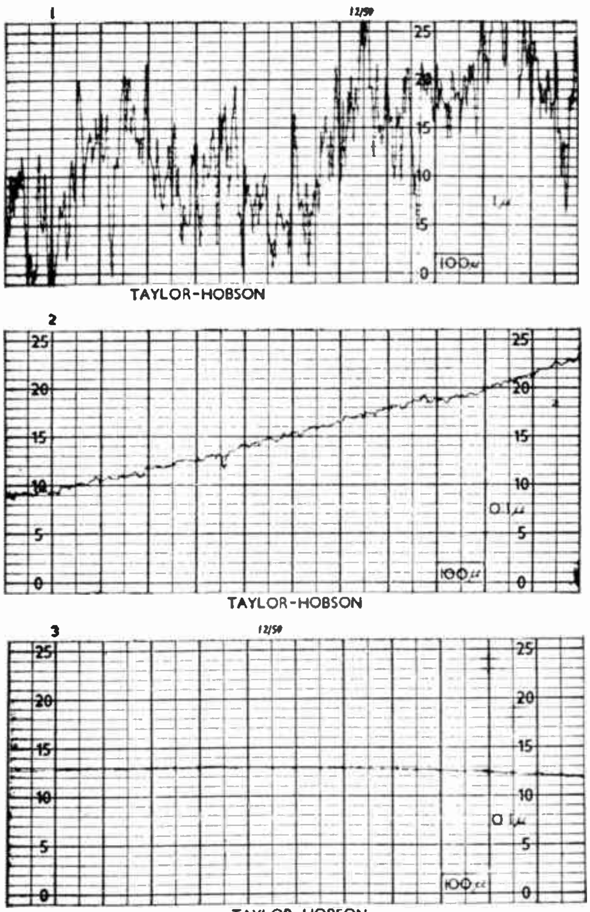


Fig. 1. (*a*) Surface profile of ceramic.
(*b*) Surface profile of silicon (polished).
(*c*) Surface profile of microscope slide.

† The Plessey Co. Ltd., Caswell Research Laboratories, Towcester, Northants.

flatness, on a micro-scale, without any special processing.

Figure 1 illustrates this. Figure 1(a) is the surface profile of a ceramic designed for use with evaporated components. It is relatively rough, with a surface roughness of approximately 1 μm in amplitude. Such a surface can be used for resistances if a large area of evaporated surface can be used (since this will average out the roughness), but not for capacitors. The second example is of a polished silicon surface, that has been finally ground with $\frac{1}{4}$ μm diamond paste. The roughness is of about 200 Å amplitude and this is acceptable. Presumably this surface could be made even flatter by electrolytic polishing.

Figure 1(c) is of a normal microscope slide. Such a surface is extremely flat. Provided that the surface is suitably cleaned to remove dust and other debris, most materials can be evaporated on to it. In some cases, the type of glass is important (e.g. with silicon monoxide,[1] where pyrex is better than soda glass) and in others, the temperature at which the substrate must be held or raised, necessitates the use of fused quartz. However, all drawn glasses and fused quartz have the same quality of surface finish.

Very thin glass plates can easily be obtained, (microscope cover glass is approximately 0·1 mm thick) and this means that components occupying very little volume can easily be fabricated.

Adhesion of materials to glass is usually reasonably good, particularly if there is an oxide formation that causes a chemical bond with the glass.[2] Measurements in the authors' laboratory have indicated that even where this oxide formation does not arise (e.g. with some dielectrics) very high values of adhesion energy are found.

Other substances have been examined. Mica is a natural one to consider and this is certainly suitable for dielectric evaporation where thicknesses around 1000 Å or greater are needed. Nichrome resistors are, however, much thinner than this, and as a result are affected by the formation and movement of cleavage steps in the mica due to the stress of the film. This is made apparent by a gradual increase of resistance with time.

2.2. *Resistors*

2.2.1. Materials

Various materials have been used for the deposition of resistors. Amongst these one of the most widely used is that of the vacuum deposition of nichrome from a source at 1600° C.[3] Careful control of the thickness of the film, if it is deposited on glass, is necessary to obtain a sufficiently high value of ohms per square. This is illustrated in Fig. 2 where the film resistance is plotted against thickness. Since 300 Ω/square is a suitable film resistance value this has to be achieved by using a control square—that is a glass square complete with electrodes on either end, the resistance of which is continuously monitored during the evaporation process. The wattage dissipation of such films is about 1 W/cm^2.

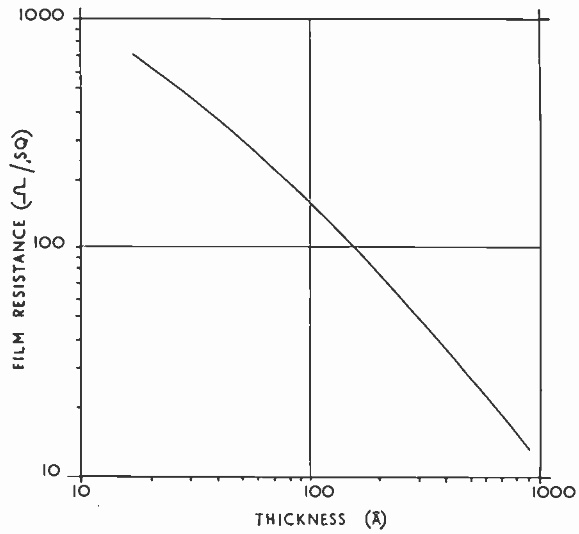


Fig. 2. Variation of film resistance with thickness (after Alderson and Ashworth[3]).

Films of metal can also be deposited on relatively rough surfaces, but in this case it is not easy to make very small components. Tubular resistors and potentiometers are manufactured by such a process.

The temperature coefficient of nichrome resistors is usually small. However, to achieve low values it is necessary to bake the resistors at 350° C for half an hour subsequent to deposition. This process is usually carried out *in situ* in the vacuum chamber, but it can be performed subsequent to the evaporation and at a lower temperature. Figure 3 shows the variation of resistance with temperature on a 60 Å thick film on glass, corresponding to a temperature coefficient of -40 parts in 10^6 per deg C.

Films of alloys based on nichrome have also been used as some of these have higher resistivities. An example of these is "Karma 331" (Driver-Harris) with a composition of Ni 74%, Cr 20%, Fe 3%, Al 3%. This gives a stable resistor at 400 Ω/square.

Resistors can be made by other methods. The sputtering[4] or chemical deposition[5] of stannic oxide with or without small additions of antimony or indium oxide can give films with resistances up to 300 Ω/square, according to Holland.[6]

Resistors can also be made by the anodization of evaporated or sputtered tantalum films. These films which are initially approximately 1 μm thick, are anodized until the resistance is up to 3000 Ω/square. The temperature coefficients of such resistors are, however, higher than in the case of NiCr films.

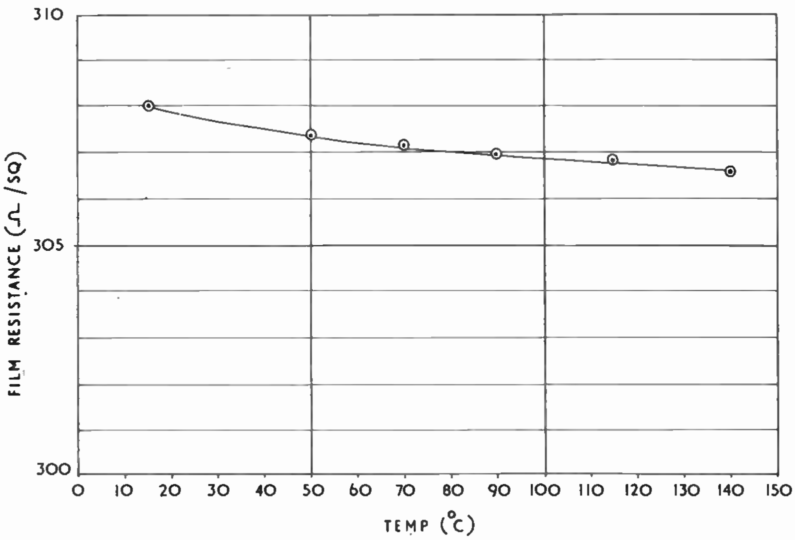


Fig. 3. Variation of film resistance with temperature.

2.2.2. Adjustment to value

As noted above, in the evaporation of nichrome resistors it is usual to monitor the resistance during deposition by means of a control square. However, to obtain the resistance value required it is often necessary to deposit the nichrome in a particular pattern or to deposit the nichrome and then cut the deposit to value afterwards.

Photoetching techniques[7, 8] are widely used in this field. They can be used either to etch an appropriate pattern in a thin sheet of copper to act as a mask, or to etch the nichrome layer after it has been deposited. The latter process is often more convenient as it avoids the necessity of constructing two masks and registering them, if the pattern needed is such that it cannot support itself when cut in copper foil.

Table 1 summarizes the processes needed to print a resistor pattern of nichrome and a connector pattern of nichrome overlaid with copper. It is assumed that large scale drawings of the resistor and connector network and of the connector pattern alone have been prepared and photographed on to a suitable plate. The required lines of the resistor and connector pattern will be black, and the connectors alone will be transparent. Registration marks will be included.

Resistors prepared by the above process can be expected to be correct to $\pm 5\%$. The limitations are imposed by the change of resistance encountered in the baking process subsequent to the deposition and also by the slight non-uniform deposit that results from a semi-point source. However, sources have been designed to overcome this latter difficulty, at least over a 4 in. square area.[9]

The photoprinting processes and also the limitations on tolerance can be largely avoided by cutting the resistors to value after deposition. Arc, electron beam[10] or mechanical methods can be used to cut an initially uniform deposit to the required value. The cutting process consists of removing slots of metal alternately from opposite edges of the metal film so as to produce a meandered pattern similar to that which can be produced by a mask. In a current production process for producing metal film resistors on ceramic tubes, mechanical removal of a

Table 1

Summary of Processes for the Printing of a Resistor-connector Pattern

1. Evaporate copper over the whole substrate with the substrate at 350° C.
2. Cover with lacquer and photoresist and contact print the resistor-connector patterns on to the surface.
3. Develop and etch in dilute ferric chloride. This removes the copper from the unexposed regions. Remove the developed photoresist in acetone.
4. Evaporate nichrome over the whole surface with the substrate at 350° C until a resistance of 300 Ω/square is shown on the monitor square.
5. Evaporate copper on to the slide with the substrate still at 350° C.
6. Cover the slide with lacquer and photoresist and contact print the connector pattern on to the surface.
7. Develop and etch in dilute ferric chloride. This removes both layers of copper in the unexposed region, and the nichrome layer as well where it is underlayed with copper.
8. Remove the developed photoresist in acetone.

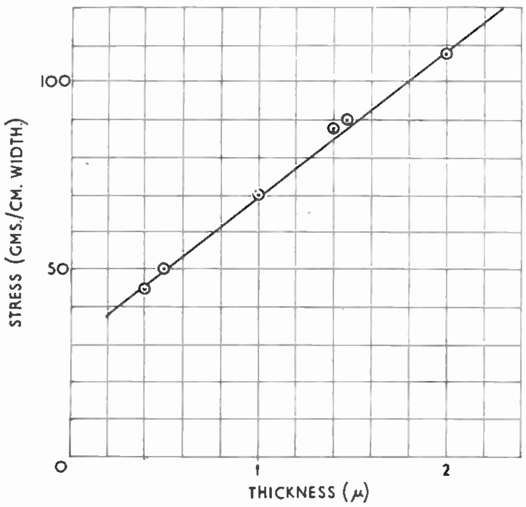


Fig. 4. Stress vs film thickness for zinc sulphide films. Substrate temperature 110° C; rate of evaporation ~ 15 000 Å/min.

spiral of metal from the tube wall can result in resistors correct to ±0·5%.

2.3. *Capacitors*

Evaporated capacitors can be made by initially depositing a layer of aluminium on to a cold substrate, then a layer of some dielectric, and finally a counter electrode of aluminium. Aluminium is an advantage as an electrode material as it is then possible to burn out weak points in the dielectric film without shorting the electrodes together. If a discharge occurs before the proper breakdown field is reached, it is found that the aluminium top electrode, provided it is no more than 1000 Å thick, will melt away from the discharge point. This leaves a small hole in the electrode, but no current path through the film.

The dielectric itself can be a low or high permittivity material. Low permittivity materials are usually those used in optical work. There is, however, one important limitation. It is found that deposited films are usually in a state of stress, the amount of stress depending on the temperature of the substrate.[11] Most metals are found to deposit in a state of tensile stress[12] and this also applies to a large number of dielectrics (e.g. magnesium fluoride). Films in tensile stress have the tendency to craze, especially if they are thick, and so satisfactory dielectric films can best be made from materials that deposit in compressive stress. Zinc sulphide[11] and silicon monoxide[1] are two such materials that have been widely used. The electrical properties of the silicon monoxide films depend very much on the rate of evaporation of the material and on the residual atmospheres.[13, 1] Zinc sulphide appears to be slightly more reproducible but in both cases the thickness of deposit is limited by the total stress built up by the film. In the case of zinc sulphide this limit appears to be at 5 μm for a substrate temperature of 50° C. Figure 4 shows the variation of stress with film thickness of zinc sulphide. Temperature coefficients are usually fairly high. Siddall[1] gives values between +100 and +400 parts in 10^6/deg C up to 200° C depending on the conditions of deposition. Figures 5 and 6 show curves obtained with zinc sulphide for variation of capacitance and tan δ with temperature. Breakdown fields are at approximately 1 mV/cm—values up to 1·5 mV/cm have been obtained.

The maximum capacitance per unit area that one can expect with such films is around 0·03 μF/cm². To deposit capacitors with higher values of capacitance per unit area it is necessary either to build up multi-layer stacks by repeated evaporation, or to deposit layers of higher permittivity. Both these approaches are being studied at present. Multi-layer stacks are found, in general, to have breakdown characteristics substantially inferior to those of a single layer film. Work on high permittivity materials has included

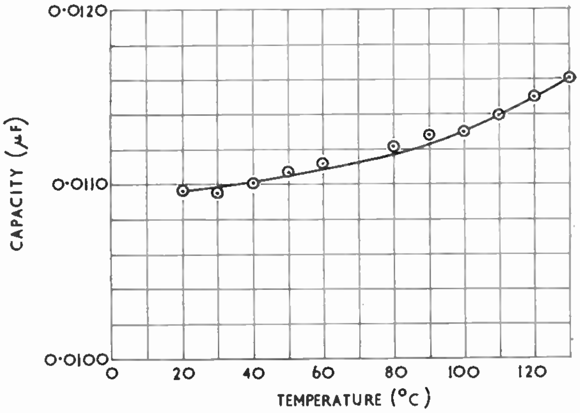


Fig. 5. Capacitance vs temperature for zinc sulphide film capacitor (4000 Å thick).

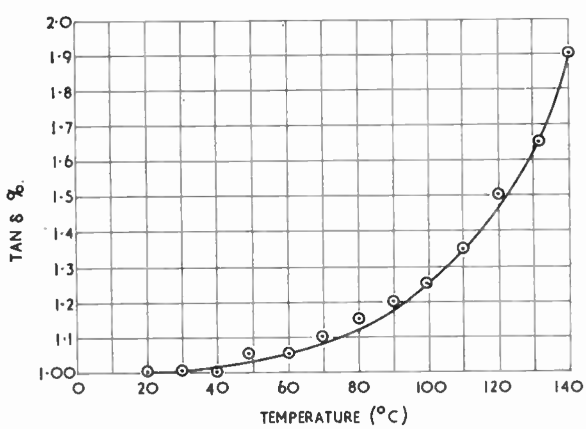


Fig. 6. Tan δ vs temperature for zinc sulphide film capacitor.

reactive sputtering from composite cathodes of lead and titanium to form films of lead titanate, and the electron bombardment evaporation of barium titanate.

Small size capacitances are not easily constructed. With films 5 μm thick a capacitance of 0·003 μF per cm^2 is possible. Taking a square millimetre as a convenient minimum size, this means that 30 pF is a useful minimum figure.

Micro-capacitors can be made by other methods. The anodization of aluminium or tantalum[14] in the form of foils or evaporated or sputtered films can give oxide layers with a capacitance per unit area up to 0·12 $\mu F/cm^2$ for tantalum (formed at 100 V, 100° C). Formation at a higher voltage gives a thicker film and consequently less capacitance per unit area (e.g. 0·03 $\mu F/cm^2$ if Ta formed at 350 V, 80° C). Such capacitors have tan δ values of 0·7% and temperature coefficients of capacitance of +250 parts in 10^6 deg C at room temperature. The anodization of sputtered films is said to yield the best results.

Chemically deposited films are also of use for making micro-capacitors. One of the most promising of these is formed by decomposing ethyl silicate on to a substrate at a temperature greater than 600° C.[15] Such films are said to have very high resistivities ($\sim 10^{18}$ ohm cm) compared with evaporated films (10^{15} ohm cm).

2.4. *Combinations of Resistance and Capacitance*

2.4.1. Conventional circuitry

Conventional combinations of R and C together with transistors give circuits that can occupy very little space. Typical examples are logic elements in computers and amplifiers that do not require large value capacitors. An i.f. amplifier designed with such components has recently been described by Black.[16] Such circuits can use ceramic transformers in the place of the usual wire wound components.[17] An audio amplifier, evaporated on to a normal microscope slide using nichrome resistors and zinc sulphide capacitors, is shown in Fig. 7. This is a two-stage device with a gain of 30 dB over the frequency range 1 kc/s to 80 kc/s.

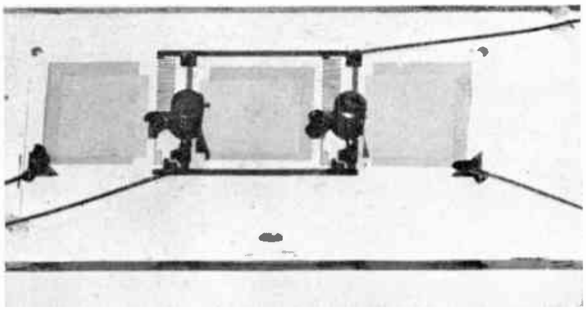


Fig. 7. Audio amplifier.

2.4.2. RC networks

Circuits are possible that make use of the distributed nature of the capacitors. In this type of circuit, a layer of conductor is first deposited on the substrate, followed by a layer of dielectric, and finally the resistor network is deposited on top of the dielectric. Thus there is capacitance to the initial metal deposit all the way along the resistors. In practice, it may be necessary to reverse the process and deposit the resistors first, due to the substrate heating that is necessary, adversely affecting the dielectric. Figure 8 shows the arrangement.

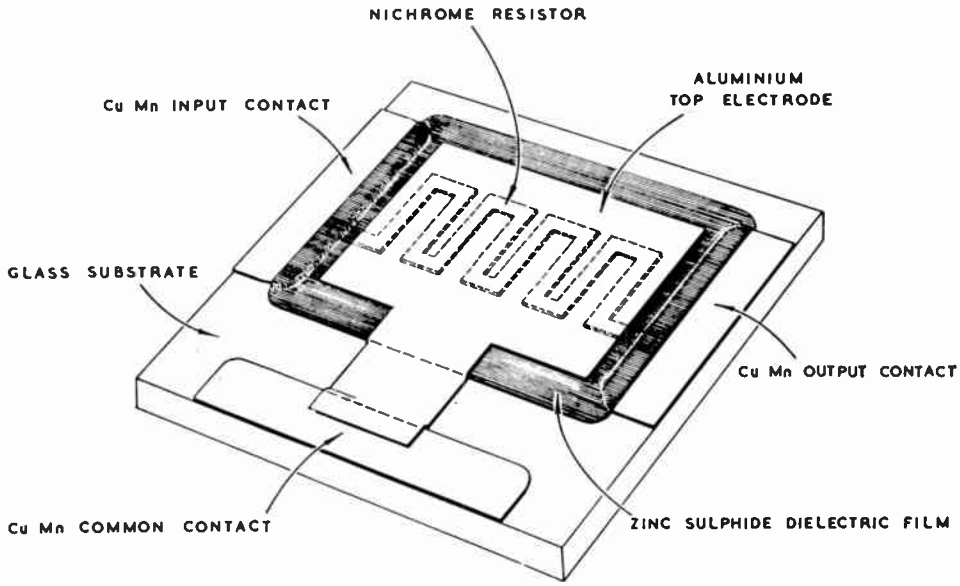


Fig. 8. Evaporated distributed RC network.

Circuits employing distributed components of this type have been described by Hager.[18] Applications have included RC oscillators and multivibrators.

2.5. *Other Components*

The evaporation of other circuit components has been considered from time to time. Inductances are difficult if more than a few micro-henries are required since it is necessary to enclose a certain amount of volume. Small size units are at present manufactured by normal winding techniques; 500 μH can be manufactured in a volume $2 \times 2 \times 3$ mm.

In some cases it is possible to replace coils by ceramics, as in the example of the i.f. amplifier mentioned above. Filter networks can be made up in this way with ceramic discs,[19, 20] a 470 kc/s filter can be made in a tube 3½ cm long by 0·8 cm diameter.

Research into the deposition of other components, especially transistors, is being undertaken by various groups. These structures will be considered in more detail in the next section.

2.6. *Discussion*

Table 2 summarizes the range of values that can be obtained with evaporated components.

Table 2

Range of Values Possible with Evaporated Components

	Resistors		Capacitors	
	Nichrome	Anodized Tantalum	Low *k* materials	Anodized Tantalum
Value	1Ω–1MΩ Slightly higher with Karma. Higher still with SnO_2	Up to 10MΩ	0·03 μF/cm² Difficult to make areas < 50 pF	0·03 μF/cm² with thick films. Down to 0·12 μF/cm² with thin ones
Temperature coefficient parts in 10^6 per deg C	± 25	Probably −300	+250 at 20° C rising as temperature rises	+250 from −200° to +170° C
Breakdown			1 MV/cm	≃ 5 MV/cm

One of the chief advantages that such techniques of construction offers is that they are clean and hence reproducible. Furthermore there should be an increase in reliability as a considerable number of component connections are eliminated.

However, problems still remain to be solved, problems such as the construction of inductances and variable *R* and *C*. The advent of a flat transistor that could be easily inserted in evaporated circuits would be one considerable help. Such things will come, so it is obvious that such techniques have a very promising future.

3. Integrated Single Crystal Circuits

3.1. *Justification*

In the face of the possibilities of fabricating flat film circuits by means of the techniques described above, there are three main reasons for attempting an alternative approach:

(a) *Active Elements.*—One of the most troublesome aspects of the evaporated film technique is the insertion of transistors or diodes and the establishment of the appropriate contacts. It does seem that in the solid circuit approach to be outlined below this will be less of a problem.

(b) *Ultimate Simplicity and Cost.*—Because, as will be shown below, many of the required passive elements can be made using the same well established solid-state diffusion techniques[21] as are used in mesa transistor production, the solid circuit approach avoids the need for new development (e.g. evaporation of dielectrics) and offers the long-term possibility of the passive elements associated with a transistor being formed at no extra cost compared with the same transistor in an orthodox package.

(c) *Reliability.*—By virtue of the complete lack of data, the hoped-for improvements in reliability must remain open to doubt, but nevertheless it seems reasonable to hope that the reduction in the number of fabrication stages and techniques, and also in the number of dissimilar materials used—and hence of interfaces between different materials—will in time lead to an overall improvement in long-term reliability.

3.2. *Choice of Material*

It is necessary to make a choice between germanium and silicon for solid circuit. Germanium is, in fact, rejected for the following reasons:

(a) *Temperature Limitations.*—Germanium devices tend to become inefficient at temperatures in the region of 80° C compared with 150° C for silicon. As it does seem that the ultimate limitation on component packing density will be set by the difficulty of removing the internally generated heat this is an important factor.[22]

(b) *Bulk Resistivity.*—Germanium can only be prepared with a room temperature resistivity of 50 ohm/cm compared with up to 10 000 ohm/cm in practice (and 220 000 ohm/cm in theory) for silicon. When considering the construction of resistors in the bulk crystal at values up to, say, 1 megohm, the aspect ratios required by germanium would be excessive.

(c) *Reverse Currents.*—Because of the smaller energy gap and hence larger intrinsic carrier concentration at a given temperature, germanium *p-n* junctions exhibit far larger reverse currents,[23] and these would greatly degrade a *p-n* junction capacitor and add to the circuit noise.

(d) *Available Experimental Techniques.*—The ease with which the oxide masking[24] of diffusion can be applied to silicon in contrast with germanium is a major factor in allowing multi-junction circuits requiring different junction depths in the same single crystal to be fabricated in a single diffusion schedule.

(e) *Surface Passivation.*—The need to achieve at least short-term stability of device characteristics so that any final protection is applied to a large number of functional blocks as a whole is fairly obvious. The same silicon oxide film mentioned above for use in the control of diffusion can also be applied to the passivation of surfaces. There is no similar technique available for germanium. More will be said on the virtues of the oxide surface passivation when we discuss transistors below.

For the above reasons therefore, silicon has been selected as the first solid circuit material, although one hopes that in the future, with improvements in materials technology, a material such as gallium arsenide, with its virtues of higher temperature of operation and theoretically higher frequency cut-off for transistors of the same geometry, will be used to fabricate solid circuits.

3.3. *Choice of Experimental Techniques*

There are three ways of making such *p-n* junctions in wide use at the moment.

(a) *Grown junctions*—made by doping during crystal growth.

(b) *Alloyed junctions*—impurity introduced by recrystallizing from a molten alloy phase.

(c) *Diffused junctions*—the desired impurity is introduced on to the surface of the specimen (usually in the vapour phase) and proceeds into the specimen by solid state diffusion.

There is yet another way of making *p-n* junctions which is the best of all and will become increasingly important when more is known of the necessary technology. This is the vapour phase crystal growth in which a volatile silicon compound (e.g. trichlorsilane) is decomposed on a silicon substrate in such a way that the resulting deposit is a perfect single crystal whose impurity content can be controlled by mixing suitable gaseous impurities with the trichlorsilane.

Many of the devices to be described below could be made by either alloy or diffusion techniques. However, where a diffused layer is to be used as a resistor there is no simple alloy equivalent. In addition, it is strongly believed that by the very nature of the physical process involved, solid state diffusion is essentially far more reproducible and controllable than alloying.

For the present then, diffusion is accepted as being the more powerful tool, alloying to be kept on the side line for any odd cases where it is necessary to make a final junction after other junctions have been diffused, the point being that the low alloying temperatures (~650° C compared with ~900–1200° C for diffusion) will not induce further unwanted diffusion.

3.4. *The Uses of* p-n *Junctions as Passive and Active Elements in Solid Circuits*

3.4.1. Resistor

A silicon filament with a diffused *p-n* junction and two ohmic contacts applied to the diffused layer as shown in Fig. 9 is one of several ways of obtaining a

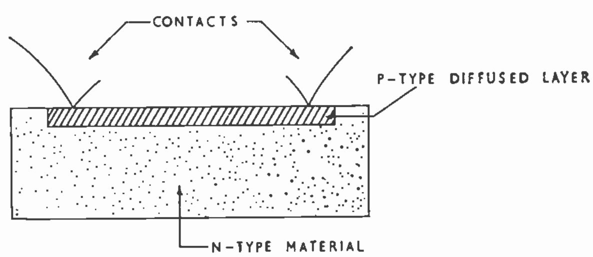


Fig. 9. Fixed resistor.

fixed resistor. Surface concentrations on diffused material are typically 10^{16}–10^{20} atoms/cm^3. The lowest value corresponds to a sheet resistivity of 1500 ohms/square (*p*-type) and 500 ohms/square (*n*-type) for a layer 10 microns thick.

All silicon resistors possess a temperature variation of resistance owing to the temperature variation of mobility, the number of free carriers being largely independent of temperature from say −200° C to +200° C other than in exceptional cases such as gold doped material. The mobility variation gives a positive temperature coefficient given by

$$\rho \propto T^{2 \cdot 3}$$

for *p*-type silicon, or 0·77%/°C at 300° K

$$\rho \propto T^{2 \cdot 6}$$

for *n*-type silicon, or 0·87%/°C at 300° K

These figures are based on lattice scattering and would be reduced by impurity scattering (for which $\rho \propto T^{-1 \cdot 5}$) in lower resistivity material. Given future development the balance of impurity and lattice scattering in diffused resistors may well give improved temperature coefficients. In any case, it is frequently

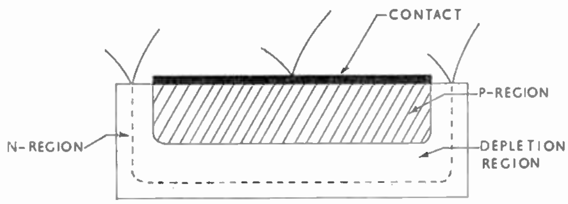


Fig. 10. Variable resistor.

the ratio of two resistor values that counts so that these large temperature variations are less serious than appears at first sight.

3.4.2. Variable resistor[25]

A variable resistor between the outer electrodes is shown in Fig. 10 in which the resistance control is achieved by varying the spread of the depletion region by means of the applied potential between the inner and either outer electrode.

3.4.3. Variable capacitor[26]

Since any normal *p-n* junction behaves as a voltage-variable capacitor this type will be considered first. A *p-n* junction acts like a parallel plate capacitor with plates whose spacing is that of the space-charge layer, w. The capacitance per unit area is given by

$$C = k/4\pi w \text{ e.s.u.} = 1{\cdot}058/w \text{ pF/cm}^2$$

where $k = 12$ = dielectric constant of silicon.

The width, w, and through it the capacitance is in general a function of the applied voltage. The calculation of w has been discussed by Giacoletto and yields the result

$$w = 2{\cdot}64 \times 10^3 \, (V/N_0)^{\frac{1}{2}}$$

where N_0 is the net impurity concentration in atoms/cm^3 on the high resistivity side of the junction, and V is the total potential across the junction. For $N_0 = 10^{15}$/cm^3 we obtain for the capacitance

$$C = 12\,700/V^{\frac{1}{2}} \text{ pF/cm}^2$$

This capacitance is dependent only on the total numbers of ionized impurities and is thus substantially independent of temperature.

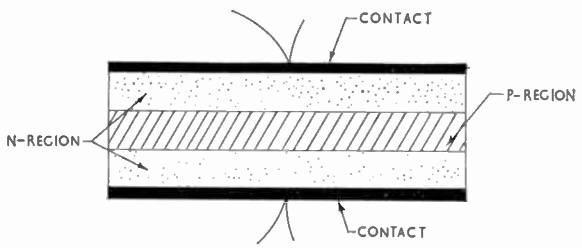


Fig. 11. A.c. capacitor.

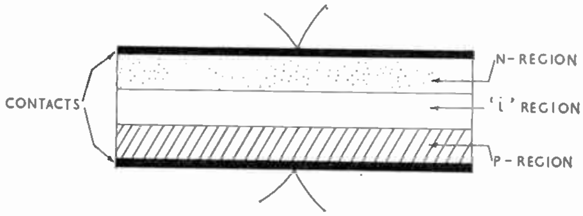


Fig. 12. Fixed capacitor.

3.4.4. A.c. capacitor

If a potential of more than $\sim 0{\cdot}5$ V is applied in the forward direction to a *p-n* junction it becomes highly conducting and not a very high quality capacitor. However, if two such capacitors are used back to back in an *n-p-n* structure as shown in Fig. 11 this defect is overcome, and furthermore the voltage dependence of the overall capacitance is somewhat reduced.

3.4.5. Fixed capacitor

If two low resistivity layers of opposite types are separated by a very thin high resistivity layer, i.e. in a p^+-i-n^+ diode as shown in Fig. 12, the capacitance will now be appropriate to the width of the high resistivity layer and independent of reverse bias voltage. Unfortunately the capacitance per unit area will be a good deal less than one can obtain for a simple *p-n* junction.

3.4.6. Distributed RC networks

It rapidly becomes obvious in designing solid circuits that the simplest conceptual approach, that of replacing individual circuit elements piecemeal, is not necessarily the simplest from the viewpoint of fabrication or the most economical in space.

If we consider the device shown in Fig. 13 we can see that between terminals 1 and 2 we have resistance, but the isolating junction produces a capacitance to terminal 3 which is distributed along the length of the resistance.

Initial structures were made in which the d.c. resistance was 11 kΩ and the total lumped capacitance

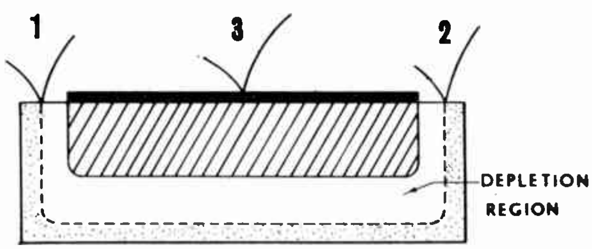


Fig. 13. Distributed RC network.

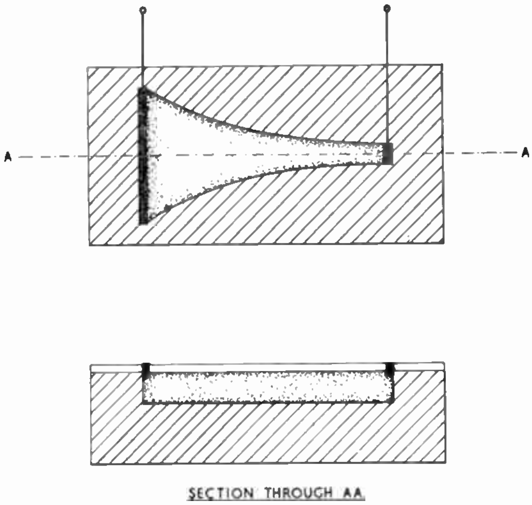


Fig. 14. Exponentially shaped distributed RC network.

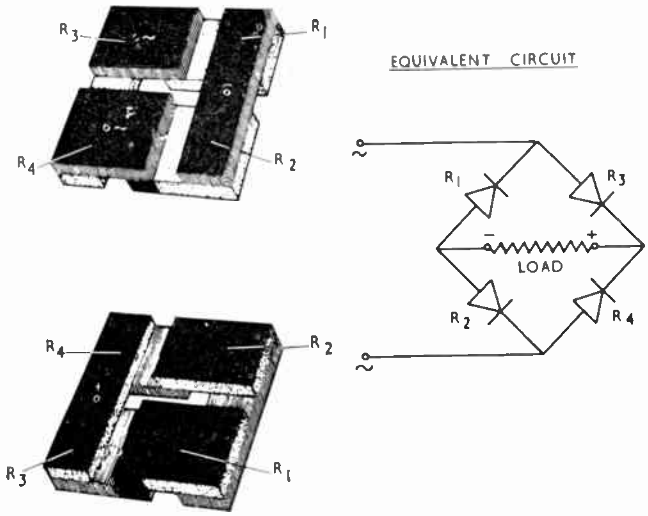


Fig. 15. Solid circuit full-wave bridge rectifier.

~1000 pF. Measurement of voltage transfer coefficient and phase shift as a junction of frequency were made and fitted theory. Such elements were used to make a pentode phase shift oscillator which could be tuned from 300 to 600 kc/s by means of the applied reverse bias on the junction.

As used in this circuit the device offers the following advantages:

(a) One small component replaces six.

(b) The frequency for 180 deg phase shift is voltage dependent.

(c) The attenuation (~21 dB) is less than that for the three-step lumped network (~29 dB).

However, this value of attenuation is too high to permit the widespread use of this circuit in transistor oscillators or amplifiers. Accordingly attention has been paid by B. L. H. and R. B. Wilson of these laboratories to the shaping of distributed RC networks with a view to obtaining 180 deg phase shift with low current attenuation. The current transfer function for an exponentially tapered network in which the RC product is invariant, has been obtained analytically, and some numerical examples indicated that current attenuation factors as low as 6 should be easily realized in practice.[28] Such a tapered network is shown in Fig. 14.

It is possibly worth emphasizing here that the orthodox three lumped R's and C's could be made as such in a solid circuit. The approach described above, however, yields far more useful characteristics and illustrates an important principle, that one should design the solid circuit to perform that required function, not necessarily to contain the same orthodox components it will replace.

3.4.7. Diodes

Obviously according to the detailed impurity concentrations in the neighbourhood of a *p-n* junction, and the geometry of the junction, we can use the appropriate *p-n* junction as Esaki diodes or 2 kV power rectifiers. Two points to be stressed in relation to solid circuits are the ease of making an array of diodes with one common terminal, by diffusion techniques, plus the fact that the emitter-base diode of a double-diffused mesa transistor is quite adequate, when repeated elsewhere in the same block of silicon, to fulfil the role of diode inputs in NOR or AND circuits.

3.4.8. Full wave bridge rectifier

In Fig. 15 is shown a silicon full wave bridge rectifier in which only a single diffusion is necessary, followed by ultrasonic shaping and plating of contacts.

3.4.9. Transistors

As was stressed earlier, only diffused transistors are being considered at the present. The virtues of the silicon mesa transistor are by now well known, not the least of which is the fact that one achieves anything up to several hundred working transistors from one diffused slice. One of the big criticisms of the solid circuit approach has been the argument that if the yield of one device is η then the yield of m devices in one block will be η^m. The fallacy lies in the implicit assumption that the same individual number of independent operations will be performed. Now as we have seen it is possible to diffuse many hundred transistors simultaneously and identically. The big remaining snag is the etching, washing, surface treatment, and application of contacts without harming the surface. It is accordingly of interest to

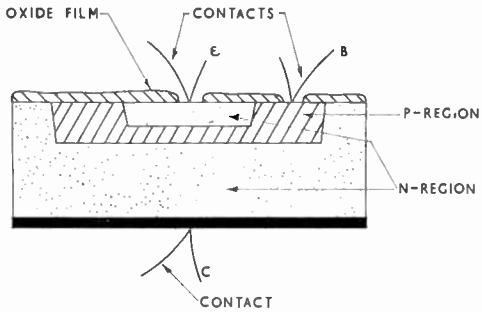


Fig. 16. Planar transistor.

show in Fig. 16 a sketch of a planar transistor structure in which a 4000 Å thick oxide film grown on the silicon surface as part of the diffusion schedule is *not* removed from the surface in the junction regions, but is left in position to give temporary, at least, surface protection. In this way no junction etching, washing or other treatments are necessary—and even lower leakage currents are reported than in normal mesa devices.

3.4.10. Inverter

In Fig. 17 is shown a simple transistor inverter, comprising one transistor and two resistors, in which the resistors are made during the transistor fabrication.

3.4.11. Direct-coupled transistor logic

A typical d.c.t.l. circuit is shown in Fig. 18. This has been made on one piece of silicon 20 mm × 1 mm × 0·125 mm. The rise and decay times were < 50 mμs.

3.4.12. Phase-shift oscillator

One of the many possible examples of component layout and construction is given in Fig. 19.

3.5. *Future Techniques*

Using the epitaxial process[27] a thin film of silicon can be grown from the vapour phase on a single crystal substrate under such carefully controlled conditions that the deposit is also a single crystal oriented with the substrate. Already this technique has been used to grow a high-resistivity silicon film on to a low resistivity substrate, so that if this is now used to make a double diffused mesa transistor one can now achieve the ideal low collector capacitance, high collector break-down voltage, and high dissipation of the mesa and the very low collector resistance and subsequent good bottoming characteristics of the MADT.

4. Low Temperature Devices

4.1. *The Cryotron*

It was first suggested by Buck[29] that one aspect of superconductivity might have practical use in computer circuits. For all superconductors the presence of a magnetic field in excess of a certain threshold value (H_c) restores normal conduction. His device consisted in the first instance of a superconducting tantalum wire surrounded by a few turns of niobium wire through which could be passed a current sufficiently high to generate a magnetic field H_c for tantalum, so switching the tantalum to its high-impedance state. This is analogous to a relay, except that the high impedance is far from infinite. By suitably arranging the geometry it is possible to achieve current gain greater than unity, and hence to use the output of one cryotron (as this device is called) to control another.

Cryotrons can be used in many logical circuits such as AND and OR gates, matrix switches, and shift registers.[29, 30] The speed limitation of this type of device is set partly by the thermal time constant of the rise and fall in temperature produced by Joule heating during switching, and partly by eddy-current delays limiting the speed with which the transition spreads through the wire. Both these effects can be reduced by the use of thin films on suitable substrates, and the future application of cryotrons is most certainly tied to the thin-film techniques—either evaporation or electrodeposition. Currently most laboratories are concentrating their effort on thin-film technology, endeavouring to achieve greater understanding of the

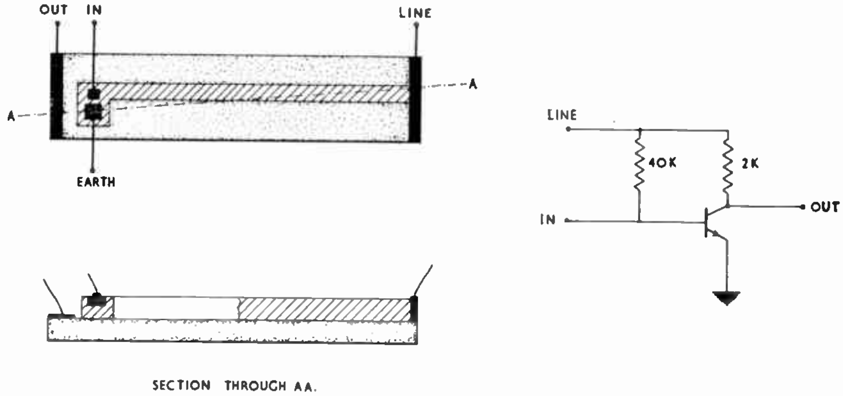


Fig. 17. Integrated solid circuit inverter.

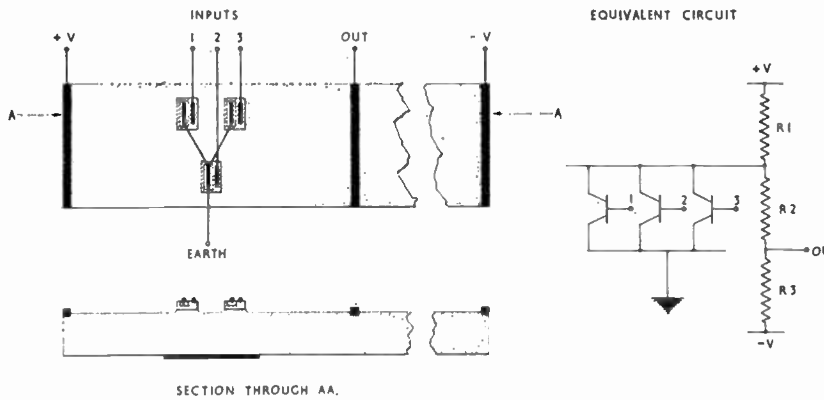


Fig. 18. A direct-coupled transistor logic NOR integrated solid circuit.

influences of substrate material and temperature, degree of vacuum used, deposition rate and so forth, on the structure and the superconducting behaviour of evaporated metal films.

4.2. *The Tunneltron*

More recently Giaever[31] has shown how if an insulating layer ~30 Å thick is interposed between two different superconductors the resulting *I-V* plot is similar to the well known Esaki diode insofar as there is a region of negative resistance.

At the authors' laboratories all successful devices have been made by evaporation techniques. An evaporated film of aluminium has an oxide film ~30 Å thick grown on its surface by exposure to air. A second evaporation of about 1 mm square of lead is then made over the aluminium. Contacts to the films are made with soldered contacts to a copper electrode system previously evaporated on to the glass. It is necessary to reduce the temperature to ~1° K with this device to obtain a negative resistance region. However, work is now proceeding on the niobium/niobium oxide/lead system which should work at 4·2° K.[32]

4.3. *Uses of Low Temperature Devices*

By combining cryotrons with tunneltrons it does appear that a useful range of active elements can be made. These are no doubt initially of use in data processing and handling but experience with the Esaki diode has shown that such negative resistance devices can perform a useful role in the communications field. The big drawback of such devices is obviously the need for cryogenics. However, there does seem to be every possibility that with the likely improvements in the design of helium liquifiers, their use in space vehicles will become feasible. On the credit side we have the cheap nature of the actual elements, the very large packing density—both achievable by the fabrication technique and permissible on the basis of dissipation—and the high degree of reliability likely to be associated with such devices.

5. Discussion

There are two reasons why we believe that some of the developments described above are important in the context of space vehicles, namely

reduction in size and weight,
reliability.

From the point of view of size, these techniques speak for themselves. For example, the transistor logic circuit had an equivalent component packing density of $\sim 4 \times 10^6$ components per cubic foot, and this was without deliberately trying to make it small. However, when we discuss such values of the packing density it is important to recognize that the limitation is no longer set by the fabrication techniques but by the dissipation requirements of the internally generated

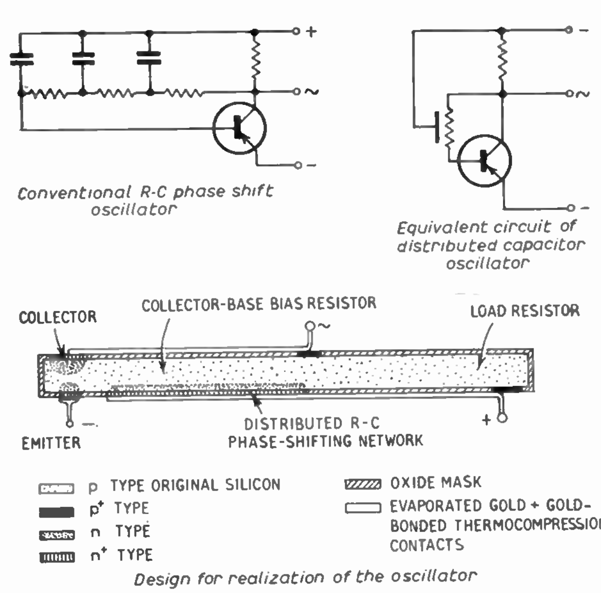


Fig. 19. Solid circuit: phase-shift oscillator.

heat. Because of this it is advisable to look for low power devices, and this is where Esaki diodes, and the low temperature elements could become very important. It is also worth saying that improved semiconductor technology, as exemplified by the planar transistor, is making it possible to use orthodox devices at lower power levels (because β is still high at 1 μA, for example).

On reliability, there is as yet no data, nor can there be any until the devices discussed above have been made in large numbers and adequately life tested. However, because of the nature of the techniques used, high reliability is to be expected.

6. Acknowledgments

The authors wish to thank their colleagues for many helpful discussions, in particular Mr. B. L. H. Wilson. They also wish to thank the Plessey Co. for permission to publish this paper.

7. References

1. G. Siddall, *Vacuum*, **9**, p. 279, 1959.
2. C. Weaver and R. M. Hill, *Phil. Mag.*, **4**, pp. 253–66, 1959.
3. R. H. Alderson and F. Ashworth, *Brit. J. Appl. Phys.*, **8**, p. 205, May 1957.
4. Anon., *Nature*, **169**, p. 829, 1952.
5. H. A. McMaster, Brit. Pat. No. 632, 256, 1942.
6. L. Holland, "Vacuum Deposition of Thin Films", p. 493. (Chapman & Hall, London, 1958.)
7. G. W. Dummer and D. L. Johnston, *Proc. Instn Elect. Engrs*, **100**, Part III, p. 177, 1953.
8. G. W. Dummer and D. L. Johnston, *Electronic Engng*, p. 459, November 1953.
9. K. H. Behrndt. Proc. 7th Natl. Vacuum Symp., p. 137, 1960.
10. W. Opitz, *Proc. 2nd Symposium on Electron Beam Processes*, Boston, U.S.A., March 1960, pp. 32–44.
11. H. Blackburn and D. S. Campbell. To be published.
12. H. P. Murback and H. Wilman, *Proc. Phys. Soc.*, **B.66**, pp. 905–10, November 1953.
13. G. Hass, *J. Amer. Ceram. Soc.*, **33**, p. 353, 1950.
14. R. W. Berry and D. J. Sloan, *Proc. Inst. Radio Engrs*, **47**, pp. 1070–5, 1959.
15. L. Pensak, *Phys. Rev.*, **75**, pp. 472–8, 1949.
16. J. R. Black, *Proc. Nat. Electronics Conf.*, Chicago, October 1960, pp. 211–9.
17. A. Lungo and K. Henderson, *Conv. Rec. I.R.E.*, 1958, pp. 235–42.
18. C. K. Hager, *Proc. Electronics Comp. Conf.*, Philadelphia, May 1959, pp. 195–203.
19. D. R. Curran and W. J. Gerber, *Proc. Electronics Comp. Conf.*, Philadelphia, May 1959, pp. 160–5.
20. D. S. Campbell and A. M. MacSwan, *I.E.E. Conf. on Components*, June 1961. To be published.
21. F. M. Smits, *Proc. Inst. Radio Engrs*, **46**, p. 1049, 1958.
22. I. M. Early, Solid State Circuits Conference, Philadelphia, 1960.
23. D. Dewitt and A. L. Rossoff, "Transistor Electronics", p. 76. (McGraw-Hill, London, 1957.)
24. Frosch and Derick, *J. Electrochem. Soc.*, **104**, p. 547, 1957.
25. R. M. Warner, W. H. Jackson, E. I. Doulette and H. A. Stone, *Proc. Inst. Radio Engrs*, **47**, p. 44, 1959.
26. L. J. Giacolletto and J. O'Connell, *R.C.A. Rev.*, **17**, p. 68, 1956.
27. J. E. Allegretti and D. J. Shombert, *Electronics*, 2nd December 1960, p. 55.
28. B. L. H. Wilson and R. B. Wilson, *Proc. I.R.E.* To be published.
29. D. A. Buck, *Proc. I.R.E.*, **44**, 482, 1956.
30. A. E. Slade and H. O. McMahon, *Proc. Eastern Computer Conference*, p. 115, 1957.
31. I. Giaever, *Phys. Rev. Letters*, **5**, 147, 1960.
32. P. Townsend and J. Sutton, *Proc. Phys. Soc.* To be published.

Manuscript first received by the Institution on 17th March 1961, and in revised form on 19th June 1961 (Paper No 672.)

U.D.C. 621.396.969.181.4: 523.42

Radar Contact with Venus

By

Prof. V. A. KOTELNIKOV †

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

Summary: In April 1961 radar contact with Venus was established in the Soviet Union. A transmitter and receiver at about 700 Mc/s were used in conjunction with an analysing and integrating system which enabled both the range of the planet and the Doppler spread in the returned signals to be determined. From the range data a new value of the astronomical unit is found; from the Doppler spread it is deduced that the maximum rotation period of the planet is eleven days. Venus is found to re-radiate about 10% of the radio energy intercepted by the disc.

In April 1961 radar contact with Venus was made in the Soviet Union. The purpose of this experiment was to determine more precisely the Astronomical Unit (the semi-major axis of the Earth's orbit), as well as to determine the rotation period for Venus and to obtain data on the structure of its surface.

In this experiment the transmitter frequency was about 700 Mc/s. The power flux density was 250 megawatts per steradian, which gave 15 watts on the surface of Venus. The transmitted waves had circular polarization, while the receiving aerial was linearly polarized.

The transmitted signal consisted of square pulses, 128 or 64 milliseconds long with spaces of the same duration between them. Sometimes a pulse of the same duration was transmitted instead of the space, but at another frequency.

Corrections were introduced in the signal and modulation frequencies used in transmission to account for the Doppler shift caused by a change in the distance from the Earth to Venus and also by rotation of the Earth. The frequencies of the transmitter, its modulation and the frequencies of the receiver heterodyne oscillators were derived from a precision crystal oscillator having a stability greater than one part in 10^9.

Transmission was carried out during the time for the signal to travel from the Earth to Venus and back again (about 5 minutes). During about the same period of time thereafter, the equipment was switched for reception.

A simplified diagram of the transmitter is given in Fig. 1. The frequency divider oscillations control key K_2, modulating the signal; key K_1 is used to start and stop transmission, and works from the timer which has an accuracy of up to 1 ms.

† Director of the Institute of Radio Technology and Electronics, Academy of Sciences, Moscow.

The incoming signals were received by a superheterodyne receiver having a semi-conductor parametric amplifier. A counting-down process in the receiver was so arranged that the reflected signals should produce a frequency of 743 c/s at the output of the receiver if Venus did *not* rotate. This signal together with all noise introduced was recorded on magnetic tape in the band 420 to 1020 c/s. A sine-wave of 2000 c/s was also recorded on this tape in order to provide a time scale when reproducing.

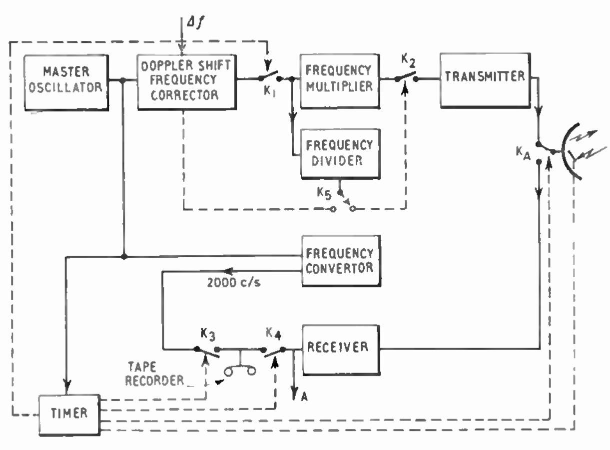


Fig. 1. Block diagram of the radar equipment.

Recording of this sine-wave was started exactly at the instant at which the 5-minute series of reflected signals was calculated to arrive, and was stopped at the end of the series of pulses. This served to indicate by how much the actual time of travel of a signal to Venus and back again differs from its calculated value.

After the transmission cycle has terminated key K_1 is opened, and key K_A connects the antenna to the receiver. The polarization of the antenna is changed and the receiver output is connected to the tape recorder together with the 2000 c/s time-marking oscillations. For monitoring purposes, the action of the keys was also recorded against the time marks, by means of a galvanometer oscillograph.

A frequency analyser (Fig. 2) was used for analysing the signals recorded on the tape. In it the signals from the tape recorder go to ten filters F_1–F_{10}, each having a pass band of 60 c/s. The filter outputs go to electronic relay circuits; pulses at a recurrence frequency of 1000 per second are also applied to these circuits

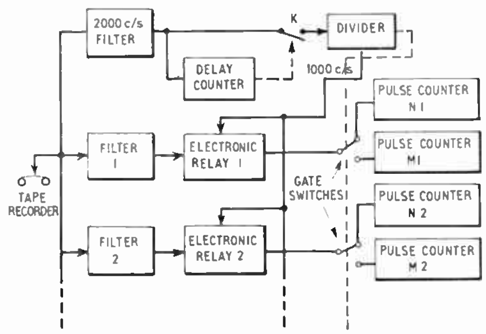


Fig. 2. Block diagram of the signal analyser.

from the divider. If the amplitude of the oscillations at the filter output exceeded a certain threshold level, the appropriate relay circuit passed the pulses to a corresponding gate switch. If it was less than this level, the pulses were blocked. The switches admitted the pulses either to pulse-counter M or N.

Figure 3 shows how the electronic relay circuit functions. 1 represents reflected signals drawn ideally without any noise superimposed; 2 shows amplitude of the signal plus noise; 3 pulses at the output of the circuit; and 4 represents the functioning of the switch. The pulses falling in the shaded areas go to counter N, while the rest go to counter M.

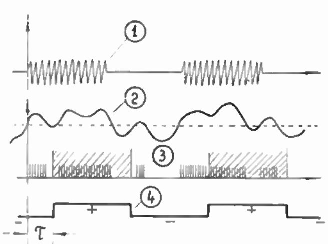


Fig. 3. Shape of signals in the analyser.

The 2000 c/s oscillations recorded were selected by a filter of that frequency as shown in the diagram and passed to the delay counter. After counting out the number of cycles corresponding to a given delay time τ, the delay counter closed switch K, through which the output of the 2000 c/s filter passed to divider. This divider, which has two outputs, then started to give out 1000 c/s pulses to relay circuits, and also to operate the gate switches at a rate equal to that of the signal modulation (about 4 c/s).

If the delay time τ was such that the switch turned on counters N during the time a signal arrived and counters M when there were none, the difference between the readings of these counters $n-m$ over a sufficiently long period will be positive; moreover, this positive difference will increase with the signal power.

By playing back this recording several times with different delay times we obtain the difference as a function of τ. One such function using the output of the 6th filter (this filter accepts signals which do not have a Doppler shift due to the rotation of Venus) is presented in Fig. 4. Here X marks the points obtained on the basis of observations on 18th April 1961, and O those on 19th April. The solid line on the drawing is the theoretical curve when no noise or signal distortions caused by reflection from different points on Venus exist. The period of a modulation cycle here amounted to 256 ms.

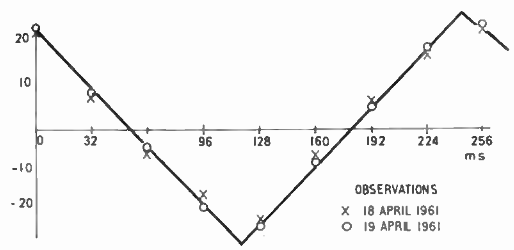


Fig. 4. Correlation function of the received signal.
x — 18th April 1961
o — 19th April 1961

One can estimate the delay of the signal and, consequently, the distance to Venus and the value of the astronomical unit, from the horizontal positions of the maxima of the curve in Fig. 4. According to preliminary data the astronomical unit was found equal to 149 457 ± 130 p thousand kilometers where p is an integer. The term 130 p results from the fact that variations in the delay greater than the modulation period, that is, over 256 ms, are not detected by this method of determining distance since it cannot distinguish between integral cycles of the modulation frequency.

When Venus rotates, the signals reflected from the different points on its surface gain an additional Doppler shift in frequency.

Examples of reflected signal energy distributions in the filters for 18th, 19th and 20th of April are shown in Fig. 5. The number of the filter is marked in the X-direction, while in the Y-direction we plot the difference in the readings of the counters N and M for the given filter divided by the standard deviation of this difference due to noise. As is seen from the drawing, this quantity for one of the days amounted to 5 in filter No. 6, and to 7 when summed over three days.

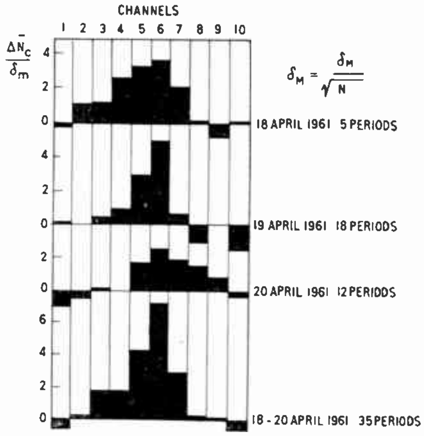


Fig. 5. Reflected signal spectrum.

On the basis of Fig. 5 and similar diagrams for other days we determined the maximum frequency shift in the reflected signals caused by the rotation of Venus and from this, the speeds that caused it. These speeds turned out to be about $\pm$ 40 m/s. If we assume that the entire surface of the planet reflected and that its axis of rotation was perpendicular to the radiation direction, this speed corresponds to a period of revolution of 11 days. If the spin rotation axis were at an angle of 60 deg to the radiation direction (according to Kuiper data), this period should be about 9–10 days. If the entire surface did not participate in reflection, this period should be even less.

Changes in the reflection spectrum of Fig. 5 from day to day can be explained as follows: On 18th April the surface of the half of the planet going away from us (filters 2, 3, 4, 5) was rough and caused dispersion giving reflected signals, while the surface approaching us (filters 8, 9, 10) was smooth and did not reflect back signals. On 19th April this smooth surface occupied the centre of the planet facing us. This led to the strong reflections coming through the 6th filter (since the centre of the planet gives a small Doppler shift) and to the weak radar reflections from the sides of the planet, which go to the other filters. On 20th April this smooth surface moved to the edge and so disappeared. As a result, the signal energy in filters 2, 3 and 4 was small, while the centre of the planet facing us, and the edge approaching us (filters 5, 6, 7, 8, 9, 10) were occupied by a dispersive surface.

Calculations and special measurements showed that the analyser used developed a signal quite near optimum and enabled signals hundreds of times smaller than the noise (in the 60 c/s band-width) to be detected during a period lasting tens of minutes.

In order to exclude systematic errors, the modulation sign of the transmitted signal was changed every operating period, as was the sign of the difference in the readings of counters N and M.

The analyser was used in conjunction with the magnetic tape recorder when the reflected signals were arriving, and the two were therefore connected in parallel.

In order to determine the power of the signals reflected from Venus, they were compared with the power from radio star Cassiopeia A. We estimate that Venus reflected back to the Earth from its entire surface on the average about 10% of the signal energy impinging on it.

At present not all of the material has been processed completely, and this report should therefore be regarded only as a preliminary one.

Manuscript received by the Institution on 5th July 1961. (Contribution No. 37.)

The 1961 Farnborough Exhibition

Radio and electronics formed a prominent feature at this year's Flying Display and Exhibition of the Society of British Aircraft Constructors which was held at Farnborough Aerodrome, Hampshire, between 4th–11th September. Overshadowing the outside display of larger equipment was a *Blue Streak* missile, the proposed launcher for the European satellite project about which Brit.I.R.E. members will be familiar from the papers read at the Convention in Oxford. Several organizations showed evidence of their planning for communication and other satellite systems but it will obviously be some years before space technology appears in more than a very minor role at this Exhibition.

The manufacturers of guided weapons were this year allowed to divulge more of their activities but comparatively little of this easement of security restrictions applied to engineering details of associated electronic guidance control and telemetry equipment. However, many components and smaller assemblies shown appeared to have been developed for this field.

With the growing concentrations of aircraft on the world's air routes, the importance of even more accurate navigation, particularly in the approach stage, is obvious and air traffic control and landing aids were thus the real keynote of the Exhibition from the radio and electronic engineer's point of view.

The Air Traffic Control Experimental Unit of the Ministry of Aviation installed at the Exhibition an experimental operations room in which a radar simulator was used to study and analyse a.t.c. problems and test new ideas and equipments. Closed circuit television was employed for display of flight data to the radar controllers and other special equipment included a radar recorder to provide a background of radar tracks on the p.p.i. displays.

Marconi's showed a ground radar display unit in which both radar signals and synthetically produced information are displayed on one 12-in. p.p.i. tube. The compactness achieved called for special circuit techniques in initiating the high speeds required in the symbol writing process.

The Decca Navigator, which operates by comparing the phase difference of radio signals transmitted from the ground, and makes use of the result to indicate a given position on a specially-drawn hyperbolic chart, can now operate with standard charts and maps which have no distortion. The digital computer is used between the airborne Decca receiver and Flight Log, which gives a pictorial display of the aircraft's track. The information from the receiver is expressed in hyperbolic co-ordinates which the computer converts into rectilinear terms, which means that the track information can be drawn on charts which have no distortion.

A new high-accuracy Doppler-effect automatic v.h.f. direction finder was demonstrated by Standard Telephones & Cables. Its design departs from current long base-line d.f. practice by using rapidly-rotating aerial systems which give a true Doppler effect as opposed to the simulated Doppler effect used in most other long base-line systems.

Weather radar covering the requirements of practically every type of aeroplane was shown by Ekco Electronics. Transistorization keeps the weight and bulk of this equipment remarkably low. The standard weather radar consists of a scanner, transmitter-receiver and indicator and weighs less than 60 lb. For executive aircraft, a smaller and lighter scanner is used that brings the installed weight down to 46 lb. The maximum range of the standard weather radar is 150 miles, covering an azimuth sweep of 180 degrees. The scanner can be deflected to give a "map-painting" effect, or it can be tilted upwards to "look" at clouds above the altitude at which the aircraft is flying.

A notable Ferranti exhibit was a new stable platform for the navigation system used in high performance military and civil aircraft. This carried three accelerometers and is stabilized by three flotation type gyros. Synchros on the gimbals supply attitude and heading information; inertial velocity is derived from the integration of accelerometer outputs. The platform was shown mounted on a rocking table simulating a manoeuvring aircraft.

A new constructional technique has been developed by Decca Radar for data handling and display systems. It is known as "environment stabilization" and all circuits and transistors are maintained at a constant internal temperature to within $\frac{1}{4}$ deg C and at controlled humidity. Liquid cooling is combined with a mechanical design which achieves good thermal insulation. Under test, the system has handled ambient temperature variations between 0° C and 45° C with relative humidity ranging from 10% to 90%. During the reliability tests, mains voltage varied up and down by 15%, but stability of the equipment held the display to an error of less than 0·2% overall and less than 0·08% for short term measurements. An overall reliability of 99·8% is claimed and it is stated that the new concept enables power consumption to be reduced by a factor of 50, while physical size goes down by a factor of approximately eight.

Considerable interest was shown in a solid state parametric amplifier and mixer developed by Marconi's for use with a 600 Mc/s surveillance radar. It utilizes a variable capacitance diode and an "up-conversion" system is adopted in which the pump frequency is about 15 times higher than the signal. The pump source is a klystron oscillating at 9200 Mc/s.

U.D.C. 621.396.969.181.4: 523.42

Radar Measurements of the Planet Venus

By

L. R. MALLING, B.Sc.†

AND

S. W. GOLOMB, Ph.D.†

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

Summary: The Jet Propulsion Laboratory established contact with the planet Venus using a planetary radar system from 10th March to 10th May 1961. Using both velocity and range data, a new value of 149 598 500 km was determined for the astronomical unit which is presently accurate to 500 km. Further data reduction is expected to improve accuracy towards 150 km. The rotation of Venus has been determined by spectral analysis as ~ 225 days.

A super-sensitive receiving system was employed with a 10 kW transmitter operating at 2388 Mc/s. The receiver range capability is a 50 mW signal located on Venus at 30 × 10^6 miles. The data received included (1) received signal level, (2) power spectrum of the Venus-reflected signal, (3) Venus-Earth velocity and (4) Venus range. The first two used open-loop receivers; the second two used closed-loop automatic tracking receivers. All significant radio frequencies were derived from an atomic frequency standard.

1. The Experiment

At intervals of about 584 days, the planet Venus passes through inferior conjunction and approaches within 25 million miles of the earth. A maximum separation of 162 million miles occurs at superior conjunction. The inverse fourth-power law expressed by the radar range equation and the present state of the art make it most practicable to attempt radar contact with Venus during the few weeks just before and after each inferior conjunction. For the year 1961 this period is defined between 10th March and 10th May.

The decision was made to modify the Goldstone radar system, originally designed for tracking the *Echo* satellite, and reflect a continuous-wave signal off the surface of Venus. The necessary capability was to be obtained by using a super-sensitive receiving system rather than transmitting with extremely high power.

Some of the high performance characteristics specified and achieved for the Venus radar experiment were the following:

(i) A high-gain, low-noise receiver antenna feed which, in conjunction with the 85-ft parabolic reflectors, provided 54 dB gain but with a noise temperature of no more than 30° K.

(ii) A 2388 Mc/s maser amplifier, capable of two months of nearly continuous operation, with a noise temperature of 20 to 30° K.

(iii) A low-noise post-maser parametric amplifier, with a noise temperature of 300° K.

(iv) An ultra-stable transmitter crystal oscillator slaved to an atomic frequency standard, stable to 1 part in 10^9 over the 10 hours or so per day that Venus is visible, and stable to 1 part in 10^{10} over the 6-minute interval required for a radio signal to go from Earth to Venus and back.

(v) A receiver local oscillator, stable to 1 part in

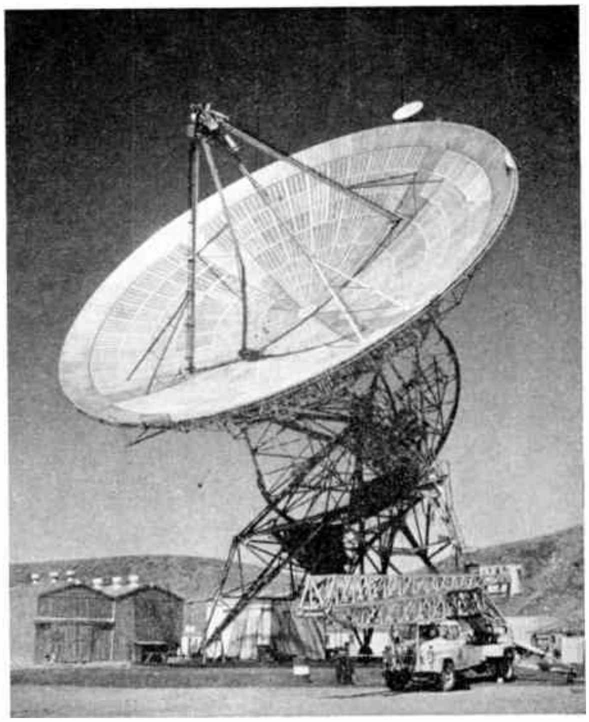


Fig. 1. The receiving antenna.

† Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, U.S.A.

10^9, which can be programmed in accordance with predicted data to follow the Doppler-shifted return signal. With these specifications, the range capability of the receiver was extended so that a 50 milliwatt omnidirectional signal located on Venus could be detected at a 30 million mile range. Figure 1 is a picture of the receiving antenna showing the mounting of the maser and parametric amplifier.

On 10th March 1961, the Goldstone transmitter was aimed at Venus for the first time. Several minutes later, the receiver was also pointed at Venus. During the 68 seconds of signal integrating time, one of seven stylus recordings was seen to deviate significantly and remained centered at a new mean signal level until the transmitter was deliberately allowed to drift off Venus, nearly an hour later. A typical recording for 29th

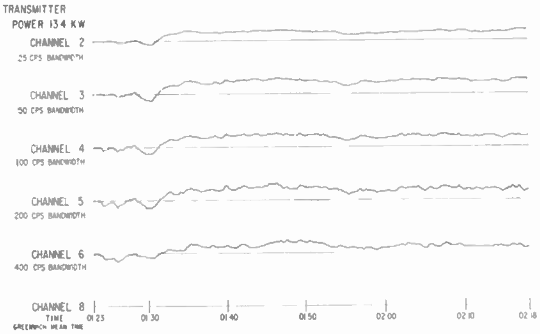


Fig. 2. Non-synchronous open-loop receiver signal level data. 29th March 1961 (P.S.T.).

March is shown in Fig. 2. Six and one half minutes later (the time of flight) the stylus recording reverted to its previous mean value. For positive confirmation, this experiment was promptly repeated, and with the same result. The first real-time detection of a radar signal returning from Venus had been accomplished.

2. Description of the Equipment

The 85 ft transmitting and receiving antennas are located 7 miles apart, with the receiving antenna shielded by intervening mountains. The desert location is shown in Fig. 3 for the transmitter site. The transmitter has a c.w. output of 12·6 kW at a frequency of 2388 Mc/s which, with an 0·35 deg conical beam, illuminates Venus with 10 watts. Table 1 gives the radar system parameters. About one watt of this is re-radiated into space and produces a signal of 10^{-20} watts, or -170 dBm, at the input terminals of the receiver. For a receiver bandwidth of 1 c/s and a system temperature of 60° K with the antenna aimed at Venus, the receiver threshold is -181 dBm, providing a typical signal/noise ratio of 11 dB. The contribution of Venus to the system temperature was measured as being approximately 0·7° K.

Several different receiver configurations were utilized to determine received-signal levels, the power spectrum of the Venus reflected signal, Venus-Earth velocity and Venus-Earth range. Closed-loop, synchronous automatic-tracking receivers obtained both velocity and range data. Open-loop receivers, with predicted-Doppler local-oscillator tuning, provided signal strength and range data, spectrum analysis and reflectivity characteristics. A non-synchronous open-loop receiver block diagram is

Table 1

Venus Radar System Parameters

Unmodulated transmitter power (12·6 kW)	+ 71 dBm
Transmitter antenna gain	53·8 dB
Transmitter line loss	0·3 dB
$\left(\frac{\sigma}{4\pi R^2}\right)$ at 31 million miles	− 84 dB
Power intercepted by Venus	+ 40·5 dBm
$\left(\frac{\lambda}{4\pi R}\right)^2$ at 31 million miles	−255 dB
Receiving antenna gain	53·5 dB
Maximum received signal level	−161 dBm
Apparent reflection and propagation loss	9 dB
Typical received signal level	−170 dBm
Receiver threshold ($T = 60°$ K, $BW = 1$ c/s)	−181 dBm
Typical signal/noise ratio	11 dB

Fig. 3. The transmitter site.

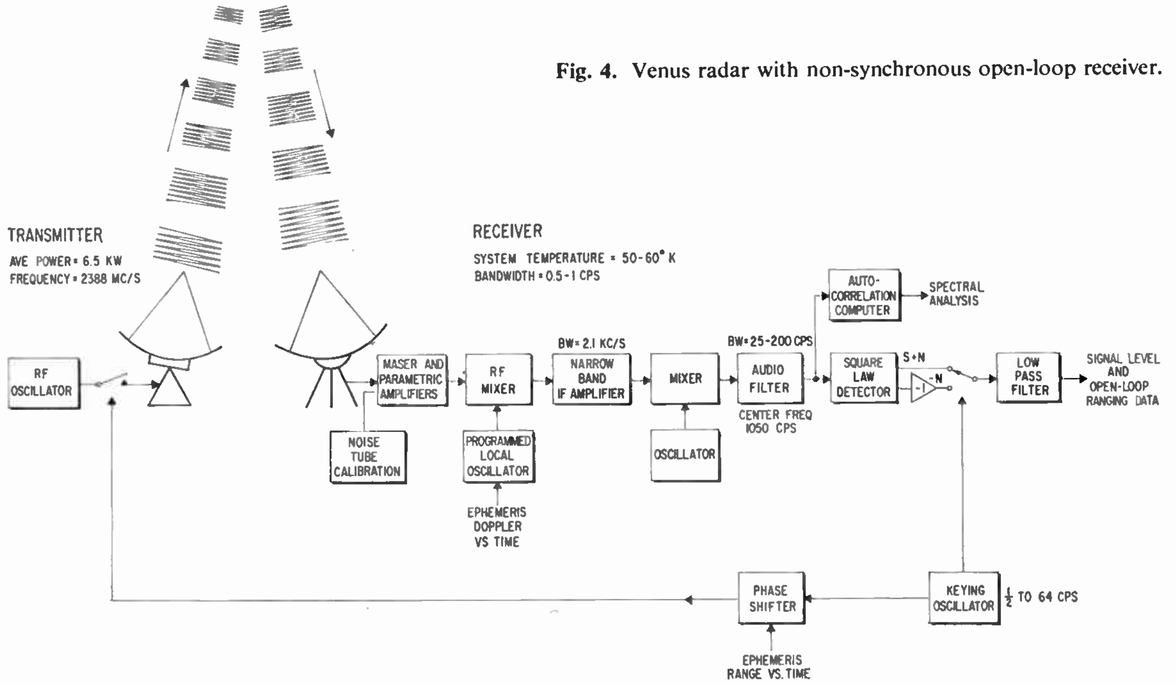


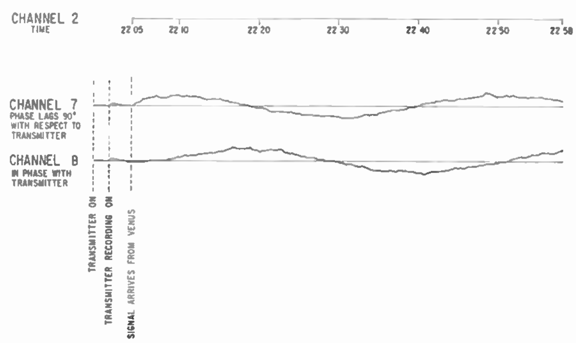
Fig. 4. Venus radar with non-synchronous open-loop receiver.

shown in Fig. 4. The transmitter is keyed with a low-frequency square-wave which also provides a phase-controllable reference signal for the receiver. The output of the low-pass filter provides both signal level and ranging data. An auto-correlation computer preceding the square-law detector permits spectral analysis of the reflected signal on a long term basis. Spectrum analysis was also accomplished in 1½ minutes using a signal digitizer and recorder connected to the output of the i.f. amplifier.

The velocity of Venus with respect to Earth was measured using a c.w. transmitted signal and a synchronous closed-loop automatic tracking receiver with a tracking bandwidth of 5 c/s.† This configuration permits direct comparison of the Venus reflected signal with the transmitted frequency. Using a non-synchronous receiver and a closed-loop coded-ranging system accurate instantaneous ranging data could be accumulated.

3. Data Obtained

The principal result of the experiment was the more accurate determination of the Astronomical Unit by measurement of the range and velocity of Venus relative to the Earth. Three independent determinations of the A.U. were obtained with a best number at the present time of 149 598 500 km, which is considered to be accurate to 3 parts in 10^6, or 500 km. Further data reduction, it is hoped, will reduce this to ± 150 km.



Keying Frequency 32 c/s
Predetection Bandwidth 100 c/s
Received Signal Level −168 dBm

Fig. 5. Non-synchronous open-loop receiver ranging data. 6th April 1961.

A typical recording of the Venus reflected signal obtained with a non-synchronous open-loop receiver for determining range is shown in Fig. 5. Relative to a perfectly conducting sphere, Venus was shown to have a reflectivity of 10% at 2388 Mc/s. A sample

† This mode of operation has been discussed in the literature. See M. H. Brockman, H. R. Buchanan, R. L. Choate and L. R. Malling, "Extra-terrestrial radio tracking and communication", *Proc. Inst. Radio Engrs*, **48**, pp. 643–54, April 1960.

spectral analysis of the reflected signal is shown in Fig. 6, where the 3 dB bandwidth is 9 c/s. The spectrum appeared quite stable over the two-months' observation period and, assuming that the axis of rotation is not pointed directly at the earth, Venus appears to rotate very slowly, possibly with a rotation period of 225 Earth days.

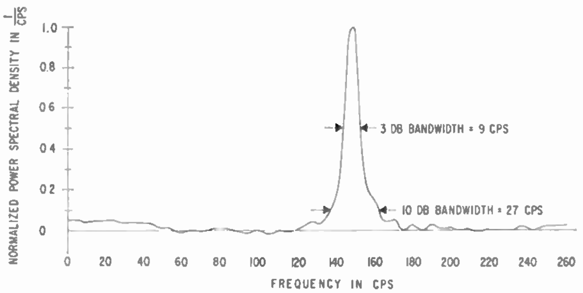


Fig. 6. Spectral analysis of Venus reflected signal using special auto-correlation computer with non-synchronous receiver. Integration time 1 hr 23 min. 21st April 1961.

The accuracy of the velocity measurement is 1 part in 10^5 without further data smoothing. The width of the closed-loop range gate was 8·2 ms, and it is believed that the accuracy of the time of flight measurement was about 1 ms for a round trip propagation time of 300 000 ms.

Date analysis and reduction began as soon as the first data became available. Complete recordings of all data obtained have been collected, and this raw data is being published in its entirety. It is hoped that numerous hypotheses concerning Venus can be confirmed or refuted on the basis of these recordings, over and above the conclusions discussed in this paper.

Manuscript received by the Institution on 29th June 1961. (Contribution No. 38.)

DISCUSSION

Dr. S. W. Golomb: My first reaction to reading the Soviet press release on Venus radar in *Izvestia* on 16th May was that we should congratulate our Russian colleagues on the discovery of a new planet. It surely wasn't Venus! It was some 30 000 miles closer than Venus and rotating 25 times *faster*.

To-day Professor Kotelnikov has cleared up part of the mystery. To the figure 149 457 thousand kilometres for the astronomical unit which appeared in *Izvestia*, he has added the term $\pm 130\ p$ thousand km, for a suitable integer p. With $p = 1$, the figure 149 587 can be reconciled with the value obtained by J.P.L. as well as Lincoln Laboratory and Jodrell Bank. The decision we reached at J.P.L. two years ago, when first designing our ranging systems for deep space measurements, to leave no range ambiguities, however, gross, unresolved, definitely paid off here. As it turned out, 130 000 km was bigger than the uncertainty in the A.U., and in picking the most reasonable value on the basis of previous experiments, the Russian choice was originally 0·1% in error. So much for the A.U. The problem about the discrepancy in rotation rate requires more elaborate comment.

First, allow me to digress briefly on the history of the Goldstone system. When J.P.L. was building *Explorer I*, the Western World's first successful satellite (Jan. 1958) and the first of the highly successful series which discovered the Van Allen belts, we had a very sensitive receiver system called Microlock. During the summer of 1958, the receiver antenna was installed at Goldstone to track our intended escape shot, *Pioneer III* (Dec. 1958). In Mr. Pardoe's paper† this, the first *Juno II* launching, was listed as a failure, but it was a remarkably successful failure Although the *Jupiter* booster underperformed slightly, so that an escape orbit was not quite achieved, all the high speed stages as well as the payload electronics worked perfectly, and the five-day trip up to an altitude of 70 000 miles and down again, resulted in the best profile of the radiation belts that Van Allen has ever obtained to date. In March 1959, the *Pioneer IV* shot was successfully tracked on its escape orbit out to several Moon-orbit-diameters, when, as anticipated, its transmitter power was exhausted. The sensitivity of the Goldstone receiver at that time was about −120 dBm.

† G. K. C. Pardoe, "The engineering aspects of satellites and their launching rockets", *J. Brit.I.R.E.*, **22**, pp. 145–160, August 1961.

The Goldstone transmitter was built during the second half of 1959 in preparation for Project *Echo*, and the receiver sensitivity was improved to −150 dBm. During 1960, this two-antenna system was used to bounce signals and voice off *Echo*, the Moon and even a desk-sized *Courier* satellite. Voice via the Moon was communicated to Woomera, and since *Echo* travelled from West to East, it was the Goldstone antenna system that had to *acquire* the *Echo* balloon, while the Bell Laboratories antennas at Holmdel were slaved in angle to the output of the co-ordinate converter data from J.P.L.

For the Venus experiment, the transmitter power was up some 30%, the frequency was more than doubled, and an overall receiver sensitivity of better than −180 dBm was obtained—one of the most sensitive receivers at any frequency, for any purpose, anywhere in the world. The very first time the system was tried on Venus, on 10th March 1961, there was a clear-cut indication that success had been achieved. With only a 68 second time constant in the post-detection filter, when Venus was brought into the beam of the receivers, their output could be seen to move off its old centre value and climb to a new value, within the duration of the time constant, and remain there until Venus was removed from the beam, when it dropped back again just as quickly. Almost every single day, many hours per day, for the next two months, contact with Venus was established, with literally hundreds of hours of data gathered, representing four major experiments and a few minor ones as well.

The most surprising and characteristic feature of our experiment was the narrowness in the bandwidth of the returning signal. Using the extra flexibility of a two-antenna system, for spectral analysis we had the transmitter turned on and off during alternate seconds, keeping the receiver always aimed at Venus. Processing the even and the odd seconds separately, we obtained spectra for pure noise and for signal-plus-noise. The difference of these two spectra gave us our very narrow signal spectrum, which by no stretch of the imagination contains Doppler spread in excess of 81 c/s at our 2388 Mc/s frequency, however far down in the noise. The 81 c/s spread would correspond to Doppler spread from the limbs of a "synchronous" Venus, rotating on its axis once every Venus year of 225 Earth days. While our frequency was much higher than the Russian frequency, the results of Lincoln Laboratory at a slightly lower frequency than the Russian one fully confirm our conclusion of the absence of Doppler spread. If one were to chop the signal-plus-noise spectrum into vertical strips, the resulting diagram would resemble Professor Kotelnikov's Fig. 5. The possibility that his experiment failed to remove all the noise power from the

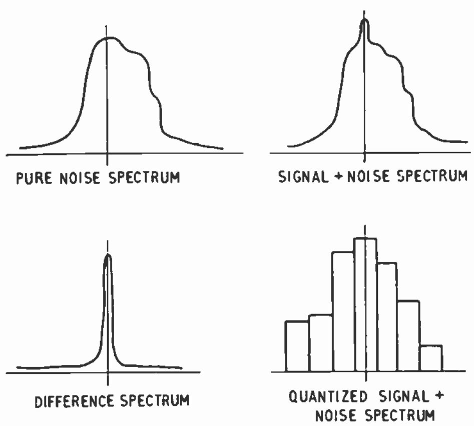


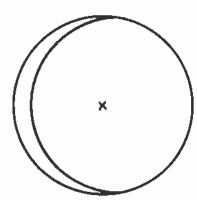
Fig. A.

"signal" spectrum should, therefore, be investigated (see Fig. A). Also, there are a variety of component malfunctions that have the apparent effect of broadening the received spectrum, but I am certainly not aware of any hardware problems that would *narrow* the spectrum. Another possibility is that when the wrong astronomical unit value is used (say $p = 0$ instead of $p = 1$), what is really a Doppler *shift* due to relative orbital motion between Earth and Venus may be misinterpreted as Doppler *spread*, suggesting a non-existent rotation.

One of our "minor" experiments involved polarization reversal to measure the surface roughness of Venus. The result, far from indicating a highly polished Venus, indicated a roughness very similar to that of the Moon, and I have never heard anyone call the surface of the Moon particularly smooth. If a Venus of that roughness were rotating as rapidly as the Russian experiment suggests, the non-coherent spectral return would certainly not be too far buried in the noise, and we are convinced we would have detected it.

A few comments about the precision we hope for in the A.U. may be of interest. We actually had several independent measurements, two of them computing the Earth-Venus round trip time very accurately in light-microseconds, one computing the instantaneous Earth-Venus relative velocity from the Doppler shift in the returned signal, and one computing the cumulative change in distance between Venus and Earth over a several hour period by tallying Doppler cycles cumulatively. Using the very accurate angle measurements compiled by optical astronomical observations, we converted each of these experiments into a value for the A.U., getting numbers in the range of 149 598 000 km to 149 599 200 km, each experiment consistent with itself to a few kilometres. However, by the time one examines sixth, seventh and eighth decimal places, such problems as the conversion from light seconds to kilometres and the propagation properties of the interplanetary medium become quite significant. Also, near inferior conjunction Venus has an apparent diameter of about 1 minute of arc, of which only a narrow crescent is visible. If the ephemeris angle to the "centre" of the Venus disc is in error by as much as 1 second of arc, differences of more than 1000 km can arise. A careful recheck and recomputation of the Venus ephemeris have enabled us to bring all our measurements together at the value 149 598 000 $\pm$ 500 km, where the uncertainty of 3 parts in 10^6 is not experimental fluctuation, but uncertainty in other physical constants and conversion factors. We have as a goal for ourselves the ultimate resolution of the A.U. to one or two parts in 10^7, but the computations involved are quite prodigious.

Fig. B.
Venus: crescent and centre.



Finally, some brief comments on the outlook for the future. Radar with Venus is 65 to 70 dB harder than with the Moon. Illuminating Venus with 10 kW is comparable to illuminating the Moon with a milliwatt. Think of turning on your flashlight, covering it with your hand, pointing at the Moon, and trying to detect the difference in lunar brightness! An extra 65 or 70 dB beyond Venus would make it possible to establish radar contact with most of the smallest and remotest bodies in the solar system—Neptune, for example, and either of Mars' tiny moons.

With only 12 to 16 dB beyond our present system, we could follow Venus all the way around its orbit, track

Mercury and Jupiter all around their orbits too, and follow a sizeable portion of the orbit of Mars. However, it is only fair to warn you that Pluto presents a special problem. The round trip delay time is almost exactly twelve hours, so that if you aim your transmitter at Pluto now, by the time the signal returns you will have rotated clear around to the other side of the earth!

My concluding proposal is definitely "blue-sky". I suggest we rush out and turn our transmitter on the planetary system, if any, of the star Arcturus. It is true that successful detection of the returning reflection is some 280 dB harder than for Venus. However, the round trip delay is 66 years, and surely by that time our grandchildren will have no trouble at all with the incoming signal.

Dr. J. H. Thomson: As the Jodrell Bank Venus radar system has already been described† I will confine myself to a very short account. Unlike the two systems we have just been hearing about, it is a fairly conventional pulsed radar system at 408 Mc/s, with a peak power of 60 kW in 30 millisecond pulses at a repetition rate of 1 pulse per second. Unlike the Russian system, the Doppler shift is removed at the receiver rather than at the transmitter. As the signal/noise ratio expected was less than unity it was necessary to use an integrating device to examine the receiver output to detect the echo. The output of the final intermediate frequency amplifier which had a bandwidth of 67 c/s was applied to a linear detector; a variable frequency pulse output was then produced, linearly dependent in frequency on the detected voltage. The integrator sampled this pulse output at eight adjacent positions on the time-base, each sampling period being 30 ms. The outputs from corresponding sampling periods on successive sweeps were counted in eight separate counting channels. Thus if only noise is present the counts in the eight channels will show random fluctuations about a mean, but if an echo is present the count in one channel will show a significant rise from the mean.

During the observations the range of Venus was changing by as much as a sampling period in a few minutes so an automatic system was employed to alter the position of the sampling channels to keep the echo always in one channel. Echoes were first detected on 8th April as soon as the equipment was functioning correctly and continued to be detected till observations ceased on 25th April.

By 8th April, both J.P.L. and Lincoln Laboratory had been obtaining echoes for some weeks, and the Lincoln value of the parallax was available to us by private communication. The echo obtained agreed almost exactly with the Lincoln value, and also with the J.P.L. result when that became available. Nevertheless a complete search was made over the full range of possible values of parallax, covering unambiguously the full spread of the astronomical results. Our value for the A.U. is 149 600 000 ± 5000 km.

It is interesting to compare the four radar results now available. The Lincoln and J.P.L. results agree together within their limits of error, and are well under the large umbrella of the Jodrell Bank limits of error—which at ± 5000 km are ten times as great as those on the admirable J.P.L. result we have just heard. Even allowing for the ambiguity now revealed in the Russian work, the corrected value of 149 587 000 km is still about 11 000 km from the mean of the three western results, well outside the limits of error, and it appears therefore that a real but small disagreement still exists.

As to the width of the spectrum of Venus, our pulsed equipment with its wide radiated spectrum is fundamentally rather unsuitable for such an investigation. Nevertheless a frequency analysis of the many tape recordings taken during the observations may shed some light on this matter.

Prof. V. A. Kotelnikov: I should like to reply to the remarks of Dr. Golomb.

The supposition that the enlarged spectrum of reflected signals in the experiments on radar contact with Venus carried out in the Soviet Union was due to an inaccurate value of the Astronomical Unit used in the preliminary calculation is not correct. Indeed, the Doppler shift of the frequency during the time of radar contact amounted to 20 000–30 000 c/s for the signals used. Therefore, a discrepancy in the Astronomical Unit of 1 part in 10^3 from its real value would result in an error in the Doppler shift of only 20–30 c/s. Moreover, this Doppler shift would be systematic. Evidently this cannot explain the enlargement of the spectrum by about ± 200 c/s.

The enlargement observed in the spectrum of the signals reflected from Venus also cannot be due to stochastic or systematic measurement errors, since all necessary measures were taken to exclude them.

Special measurements, in which signals imitating the attenuated transmitter signals were applied to the input of the receiver, did not result in an enlargement of the spectrum.

I believe that the discrepancy in the character of the reflected signal spectrum obtained in the U.S.A. and the U.S.S.R. can be explained as follows:

Let us assume that when sending radar signals to Venus a large part of the signal energy is reflected from the point nearest the radar set (just as would be the case for reflections from a shiny sphere), and moreover, that the dispersed reflection from all of the remaining surface is due to irregularities. Then, the reflections from the nearest point will give a narrow spectrum, while the reflections from the rest of the surface due to the Doppler shift of the frequency caused by the rotation of Venus will give a wide spectrum. The narrow spectrum in our measurements must fall in the sixth filter. From the energy distribution among the filters obtained in our experiments we can deduce that the energy in the sixth filter is about of the same order of magnitude as the total energy in all the other filters. Thus, for purposes of orientation we can assume that the energy in the narrow and wide parts of the spectrum are about equal.

Comparing the width of the narrow part of the spectrum (which according to the data obtained in the U.S.A. amounts to 9 c/s or in terms of our frequency 2·6 c/s) with the wide part (which according to our data amounts to

† J. H. Thomson, J. E. B. Ponsonby, G. N. Taylor and R. S. Roger, "A new determination of the solar parallax by means of radar echoes from Venus", *Nature* (Lond.), **190**, No. 4775, pp. 519–20, 6th May 1961.

400 c/s), we see that the spectral density of the energy in the wide part of the spectrum must be about 400/2·6 $\simeq$ 150 times smaller than in the narrow part.

In the Soviet Union a preliminary analysis of the spectrum was carried out with filters having a bandwidth of 60 c/s. As a result, the narrow component of the spectrum could not be detected. Apparently most attention in the U.S.A. was given to the narrow component, while the wide component of the spectrum having a much smaller spectral density could have been undetected.

The spectral density curve given in the paper by Mr. Malling and Dr. Golomb does not refute what was said. Indeed, beyond the bandwidth of 30 c/s it contains irregularities, whose ordinates comprise 1/30th of the maximum value even if we assume that the abscissa should be taken to be slightly curved as Dr. Golomb indicated. Such a value of the spectral density is even larger than the approximate value of the spectral density of the wide part of the spectrum obtained above.

In the future we will attempt to reproduce the narrow part of the spectrum, also, using the magnetic tape recordings that we have. Maybe it would be possible to reproduce the wide part of the spectrum in the U.S.A. using the recordings that they have.

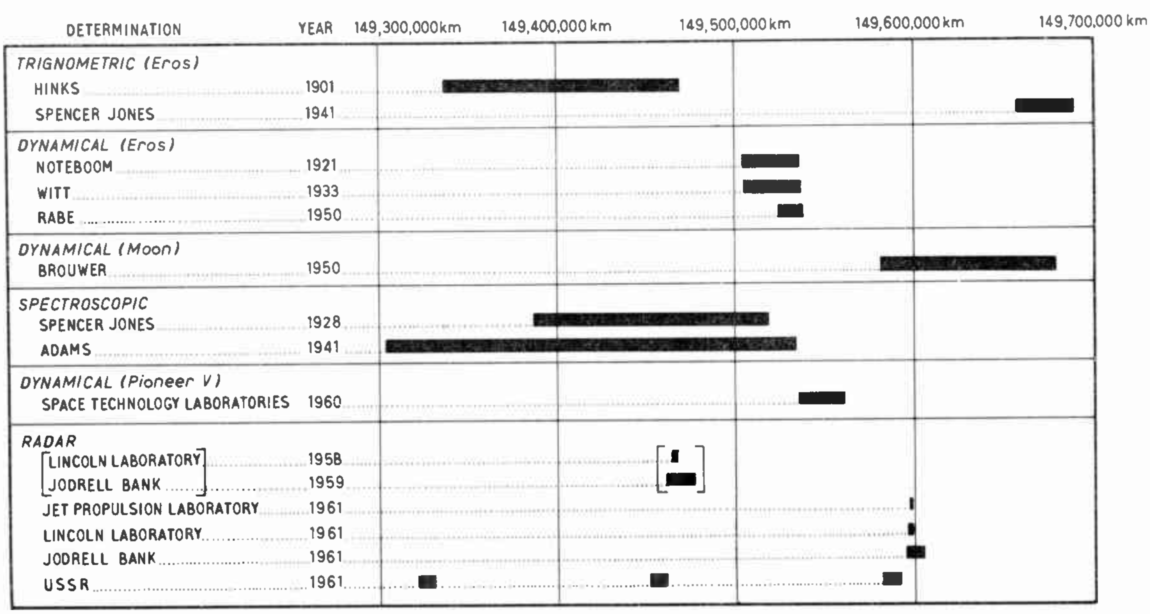
It would be also very interesting to attempt to obtain the form of the spectrum using the recordings of the signals reflected from Venus, which were obtained at Jodrell Bank.

I hope that these investigations will confirm the suppositions that I have made.

Dr. S. W. Golomb: I would very much like to thank Professor Kotelnikov for all the interesting information he has given us about the Soviet Venus-radar experiment. Apparently my suggestion of signal-plus-noise instead of signal spectrum does not underlie our discrepancy. Neither, however, does Professor Kotelnikov's suggestion about the total off-centre power in our experiment, since originally we used seven consecutive filters each 200 c/s wide, which when adjusted for wave-length is exactly parallel to the Russian's use of 60 c/s filter; and we detected *no* signal power except in the centre filter. I think Professor Kotelnikov will join me in urging our colleagues at Jodrell Bank to speed the reduction of their spectral data and thereby assist us in resolving our conflicting conclusions.

DETERMINATIONS OF THE ASTRONOMICAL UNIT

In an article in *The Times Science Review*, Autumn 1961, entitled "The Scale of the Solar System", Dr. J. H. Thomson of the Nuffield Radio Astronomy Laboratories, Jodrell Bank, discussed recent determinations of the astronomical unit by radar techniques and their relation to values obtained by other methods. By kind permission of the Editor of *The Times Science Review* and Dr. Thomson, the useful tabular presentation of the various results given in the article is reproduced here as an appendix to the discussions which took place at the Institution's Convention.



The principal determinations of the astronomical unit since 1900. The lengths of the blocks corresponding with each determination show the limit of error as determined by the workers concerned. The two radar results enclosed in square brackets have since been repudiated by their authors. The reason for ambiguity in the recent Soviet determination (here presented by three blocks) is explained by giving different values to the integer p in Professor Kotelnikov's final result.

Fourth International Conference on Medical Electronics

This year the Conference held under the auspices of the International Federation for Medical Electronics was held at the Waldorf Astoria, New York, from 16th–21st July and over 1600 delegates attended.

Nearly 300 papers were presented in 36 sessions, necessitating the use of four conference halls simultaneously. The timing of papers was very well controlled, making it possible for delegates to change over to other sessions. Nevertheless, a number of papers were not worthy of presentation and held the delegates for sufficient time to prevent them moving elsewhere. Conference organizers should always maintain a certain standard and resist employers who do not feel expenses for delegates' participation are justified unless a paper is presented.

The papers in general dealt with further developments in the subjects covered at previous conferences. As might be expected there was a considerable emphasis on computer techniques from the American side, covering diagnosis, mathematical models, control systems and the storage and retrieval of data. The biological effects of microwaves was also dealt with in some detail, a subject which has had little attention in this country.

In addition there were six workshop sessions for informal discussions. These were very popular in spite of a tendency to be slow in "warming up" due to the large attendance.

The Exhibition covered three rooms and contained about 80 exhibits. The American companies prominently featured miniature transistor devices such as cardiac pacemakers. A number of companies demonstrated instruments for remote monitoring of physiological data from the patient either by line or radio and this approach shows a trend which may well follow in this country. The Japanese industry was present in some strength with 11 companies, but British industry was not represented apart from one display by an associated company.

In the Scientific Exhibition, however, the positions were reversed. Only five organizations participated, but two of these, The Biological Engineering Society and the Post Graduate Medical School of London, were from Great Britain, the B.E.S. having seven separate displays. The popularity of this exhibit emphasized a deficiency and future conferences could encourage the display of apparatus, not only as an alternative to papers, but in some cases to supplement them.

As can be seen, the programme completely filled the period of the Conference. To summarize, it was a complete conference with everything covered, but it might have been improved with a little less quantity, leading to higher quality and leaving more time for delegates to meet colleagues and to discuss subjects of common interest.

W. J. PERKINS

The 1961 Radio Show

Despite the fact that the number of exhibitors and the area of the exhibits were both slightly down on previous years, the National Radio Show 1961 attracted 385 925 visitors, an increase of nearly 40 000 over last year and the highest attendance recorded since the Show moved to Earls Court in 1951.

The outstanding feature of this year's Show was the demonstration to the public for the first time of colour television, using 405-line N.T.S.C. system and the shadow mask tube display device. The demonstration was given by the British Broadcasting Corporation and included live, closed circuit, as well as film. Six displays were used, each of them compared with a monochrome display of the same picture. Colour television was not demonstrated on any of the manufacturers' stands.

On the stands manufacturers anticipated the necessity for a change to a higher line standard by demonstrating switchable standard receivers, or receivers which could be converted to u.h.f. and new line standard.

Sound radio has not vanished from the Show or taken a back-seat as in previous years, but rather has surged forward again with the manufacturers' stands displaying many transistorized models.

The General Electric Company and Radio and Allied Industries Group did not take part in the National Radio Show but held a separate show during the first week of the National Radio Show at the new Horticultural Hall. Here they put on a display of colour television comparing N.T.S.C. with the SECAM standard, although large crowds made it difficult to make objective comparisons.

Considering the National Radio Exhibitions over the past years, it is disappointing to note the withdrawal of many component manufacturers, thus lessening its interest to the engineer. However, the public support which the National Radio Show has secured must justify its continuing in its present form. The other exhibitions of more interest to the professional engineer are more and more sub-divided into their specialist categories —components, instruments, computers, scientific apparatus, aeronautical radio, audio equipment, etc. One wonders more than ever whether there are not too many exhibitions in one year.

U.D.C. 681.142 : 621-52

A Method of Using a General-Purpose Computer in an On-Line Time Sharing Application

By

A. ST. JOHNSTON,
B.Sc. (Member)†

Presented at a meeting of the Computer Group held in London on 3rd March 1960.

Summary: When a computer is coupled to a real time operation as in process control, the work load on the machine may fluctuate widely depending on operating conditions. External phenomena demanding action will occur essentially at random with respect to the machine. To cope with peak conditions the machine must be adequately fast, and thus will have idle time when activity is low. The use of time sharing techniques or more properly equipment sharing, allows the computer to utilize any idle time by having alternative jobs of lower priority, higher priority jobs having an interrupt facility on the machine. An engineering and programming method has been evolved of applying a computer to a process control system for on-line applications.

1. Introduction

The normal usage of a scientific or business computer is off-line in that the machine operates in its own time and only the overall completion of the job has to be related to real time. The more complex data processing takes longer and conversely, if the programme sequence is simple, the machine proceeds to the next step immediately.

However, in a data logging or control application in industrial plant, the process calls the tune as regards the speed and timing of operations. For example, in data logging, a number of transducers, such as thermocouples, may be scanned at five points per second; this scanning rate must be maintained whether the values scanned are within the alarm limits or not. However, the amount of work that the computer has to do when a point is within the limits is smaller than under alarm conditions. The machine's internal operating speed must be sufficient to carry out the largest amount of calculation necessary in the given time interval, and so, if as is the normal case, the machine is processing points within the alarm limits, there will be idle time between scans.

The machine may also be asked to type out a log of some of the points every hour or on demand; this operation certainly takes seconds and minutes if complex, but it is essential that the scanning of the points for alarm carries on at the five per second rate. The operator may demand a summary print-out or wish to alter an alarm level by the manual switches and this must be interlaced with the scanning cycle. To make possible this form of operation, whereby the machine appears to be doing several things simultaneously, a system of priorities must be established and an engineering method by which the machine can test what must happen next must be evolved.

2. Method Adopted for use with an 803 Computer

The 609 system is a data logging or control system for on-line industrial plant operations which uses, as its central processing unit, a standard 803 computer. Figure 1 is a block diagram showing how the standard computer fits into the overall system.

The standard computer has a paper tape reader as one input, a set of manual keys for another and an output channel by means of a paper tape punch. The order code comprises 64 instructions referring to functions between the 4096 word core store and the accumulator.

The "channel 2" input and output channels can be pre-switched to a number of different alternatives; this makes it possible to connect many different types of peripheral equipment such as card readers, etc., to the central processor. For on-line working additional inputs may come from several other sources:

(i) An analogue-to-digital converter, which can be switched in units of 128 channels to almost any number of d.c. inputs by means of relays, thus connecting the analogue digital converter to the appropriate transducer. This enables the equipment to select transducers in any order and read the point value into the computer to better than 0·1% of the full scale.

(ii) A mains-driven clock which produces a pulse every 15 seconds and has contacts which enable the time to be read digitally in minutes and hours.

† Elliott Brothers (London) Ltd., Borehamwood, Hertfordshire.

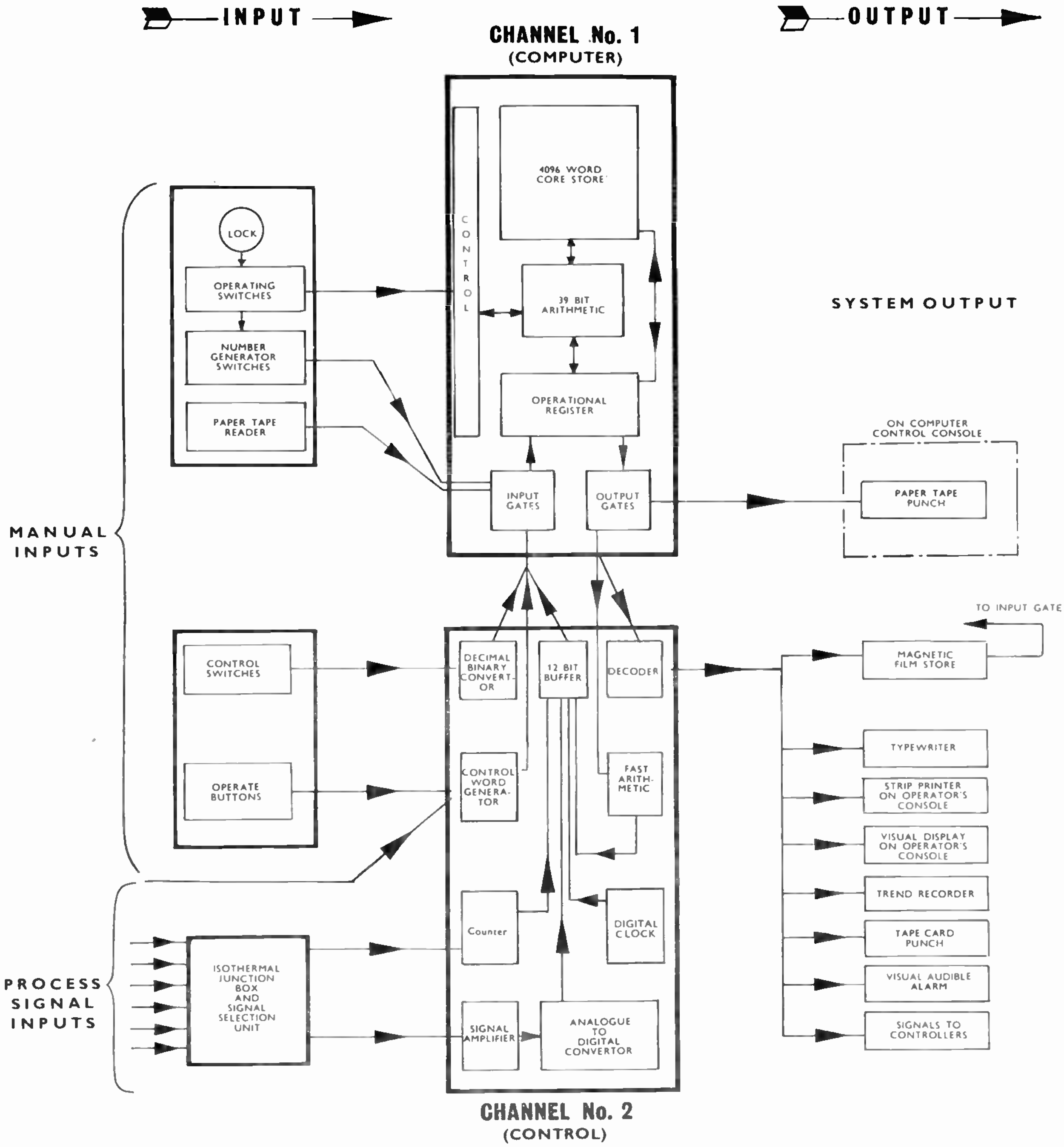


Fig. 1. Flow diagram for the 609 process control system.

(iii) Operator's switches to select manual operations (from 99 possibilities), point number up to 999, and a value up to 9999. Note that unlike the computer manual keys which can alter the program, these can be read into the machine only by a program, and hence are safe, in that the operator cannot disrupt the program or the sequence by maloperation. An "obey manual" command key initiates the operation once set up.

Additional outputs for process work may include:

(i) A logging typewriter.

(ii) Strip printer or printers on which alarms, summaries, and all operator value changes are recorded.

(iii) Digital display lights giving the value selected by the operator and a feed back of the point number selected.

(iv) Incremental outputs with analogue conversion to shaft rotation from which standard control signals (0–0·5 V a.c., 3–15 lb/in^2, etc.), for altering set points of conventional controllers in a fail-safe manner may be derived.

2.1. *Busy Line Operation*

Almost any type of peripheral equipment has its own inherent speed which cannot be a precise multiple of the computer's own internal timing.

There are a number of methods of coupling the one to the other. If the computer is faster than a particular piece of equipment—say 25 characters per second paper tape punch—then the simplest method of all is for the machine, having given a punch instruction, to wait until the punch has completed its cycle. After a known time interval or preferably by a signal back from the punch to the computer, the computer is "released" and carries on. This means that all operations whether connected with the punch or not are held up while the punch is in operation. To allow the machine time to carry out other operations and still get maximum speed out of the punch, a frequently used system is the "busy line" method. Here, having given a punch instruction a "punch busy" line is set as well as giving the punch its go signal. The machine is not held up on any other instruction and while the punch is operating on the character just signalled, the computer can be calculating the next character for punching. When it has completed the calculation it gives the next punch order; if the "punch busy" line is set the machine will now be held up until the "cycle completed" signal from the punch re-sets the busy line and frees the machine. If the punch completes its cycle before the calculation is complete, the machine is unaware of the busy line operation, but in this case the punch will not run at full speed.

This arrangement makes it safe from the programming point of view, as it does not matter to the programmer—apart from the effective punching speed—whether the punch cycle is completed or not. To achieve the maximum punch speed, however, he must ensure that he gives the next punch instruction before the punch has completed its cycle, so that the machine is waiting for the punch—albeit only for a small portion of the punch cycle. This is a satisfactory method for most normal computer applications where the program is only taking into account simple calculations in the intervals between punching. During the main part of the calculation the punch is not required to be in operation, and so no attention has to be paid to the relationship between calculation and punch instruction timings. The busy line system is that employed in the 803 computer and is adequate for normal off-line scientific or business type of calculations.

When used as a part of an on-line system the requirements are rather more complex, as it is important that in most applications "continuous" scanning of plant variables is maintained to detect possible alarm conditions, the scanning rate is typically an average of 5 points per second. The basic program reads the points successively at approximately 200 millisecond intervals and tests for alarm conditions. If the plant is not in alarm this sampling and testing may take about 50 milliseconds out of the 200 milliseconds period. All other operations must share the remaining 150 milliseconds, jobs taking their appropriate priority.

2.2. *Control Word*

To carry out priority assessment on a number of different possible functions in order to share the time available, it is necessary that the machine—or programmer—can be informed of the state of every device. In the 609 this is done by using one word, i.e. a maximum 39 digits, as a control word. Instead of the digits in the word having increasing significance in the binary code, each digit has its own meaning independent of the other digits in the word. To pursue the example already quoted, a special digit in the control word would be a 1 when the tape punch is busy and a 0 when it is free to accept another instruction. If the punch and the tape reader were the only two users of the control word, the program could test whether either were busy by reading in the control word to the accumulator and testing for zero. If non-zero, then by collating or repeated shifting and testing, the program can identify the state of either

Table 1

Code for the Control Word Digit

Digit	Significance
3	Obey Manual Command
4	Printer Motor Bar Busy
10	15 Second Interval Clock Pulse
11	Clock Busy
14	Input Relays Busy
15	Tape Reader Busy
16	Punch Busy
17	Output Character Register Busy
18	Typewriter Busy
19	Parity Failure

piece of peripheral equipment. By this means the program can repeatedly test the state of the outside world and if any other action is impossible reverts to the main program without being held up.

The layout of the control word digit for one application is given in Table 1. It can be seen that there are a number of quite "independent" functions which are co-ordinated in the one control word.

Actually, in the normal on-line application the tape reader and punch are of minor importance, but it is essential to know, for example, that the obey manual command key has been depressed by the operator, as the machine must then read the command description switches to determine the operator's request and take appropriate action depending on the priority allotted to it by the programmer. Alternatively, when the machine has completed a minor scanning cycle which need not be precisely timed, the machine regains synchronism with real time by testing whether the 15 second interval pulse from the mains driven clock has been received. The machine is programmed so that it will always be ahead of the clock and other additional jobs of lower priority can be done while waiting for the 15 second pulse.

It is not necessary for the program to have to test for all control word bits at every stage, for example in the case of those digits indicating busy state on the output equipment, unless the appropriate output order has previously been given, it is a waste of time to examine the control word for a busy bit. However, the manual command and 15 second interval digits may appear substantially at random to the program and therefore require regular inspection. The manual command is highest priority for reading if not for action, in that, if the manual command key has been depressed, the program must determine which of the 99 manual commands is desired. The type of demand then determines the priority for subsequent action by the machine.

A number of installations have been made using this control word method. Automatic priority interrupt circuits are built into the 803B machine but the automatic interrupt does not replace the control word operation in all cases as the control word system allows the programmer a more direct control which is useful in certain cases.

3. Programming Approach

3.1. *Main Program*

The basic rhythm of the machine in an on-line application is its scanning cycle, and it is vital that this is kept to a minimum so that the maximum "spare" time is available for flexibility in the programming of the lower priority jobs. However, the basic alarm cycle may become complex in itself particularly if there are several different types of alarm for each point, such as rate, as well as absolute value. Also the job specification may require a history to be kept of the last so many readings for each point. Careful consideration of flow diagram is necessary to avoid putting so many tests for priority in series that time is wasted in doing this, when in fact some of the alternative jobs may be mutually exclusive. For example, the printing of an alarm report excludes the machine from printing a summary or log. The flow diagram must therefore differentiate carefully between AND and OR jobs and avoid unnecessary tests. Interrupts of the program can only come in certain places; thus, although the alteration of a set point by manual command takes priority over the printing of alarms, the machine must complete the line (or lines) relating to a single alarm, before accepting the manual alteration of set point, and the print-out of the new value which accompanies the alteration, after which it may revert to printing the outstanding alarms.

3.2. *Alternative Programming*

When all is normal in the plant it is desirable that the minimum time is taken over the alarm organization that is the "basic program loop". This can be achieved by storing the last value of a point and testing before anything else whether the new value is identical. If this is so, then all the above complex alarm routine can be omitted and the majority of the time before reading the next point used on lower priority jobs such as demand print-outs or routine logging. Even here if all is normal, and it is to be hoped that this is the usual plant condition, even routine logging and manual demands will be required only infrequently, in which case the machine really has "time to spare". Other factors will now determine whether any use can be made of this time. The most important of these is storage capacity. As ever, in the use of computers the programmer has to foresee all possible combinations of happenings, and have programming routines available to carry them out. These routines will occupy a considerable number of locations in the machine store, even if they are not being used at this moment and because of the number of transfer orders involved in these routines it is essential that they are all held in immediate access locations. In the 803/609 there are up to 4096 39-bit words available and so under normal plant conditions, if the machine has time to spare, it is probable that there will be enough storage capacity left over to be able to hold an alternative program to which it can revert in idle milliseconds. This program must also be immediately accessible if full utilization of the time to spare is to be made, as the machine must return to the main program every 200 milliseconds or less in order to scan.

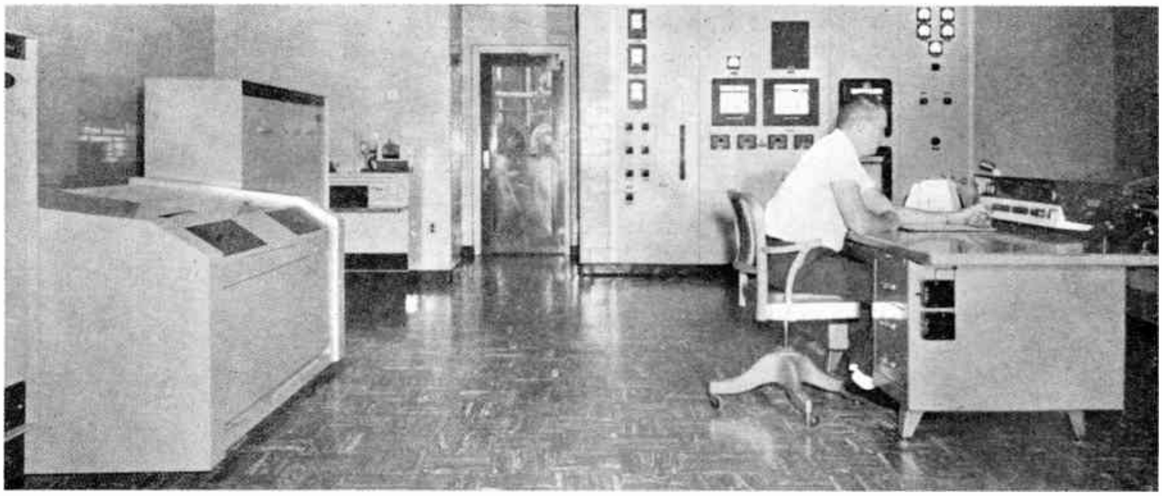


Fig. 2. A 609 installation at Roy S. Nelson Station, Gulf States Utilities, Louisiana.

An alternative program could in theory have nothing whatsoever to do with the process, and could even be a payroll. However, it is clear that, as the machine only operates on one order at a time, to the machine this is one big program and, although it is possible to have store lock-outs and other artifices, these two programs *might* interact. Even if this interaction happened because of an operator's error, by the alternative program operator this would be of no consolation to the plant manager. What is reasonable to consider is there being a number of alternative programs available which may be stored on magnetic film and can be brought into the spare storage space by the process operator's manual command switches on his existing control panel. These programs would all have been tested comprehensively before on the same machine in off-hours or on a similar machine off-line. They would almost certainly relate to the information gathered by the machine from the plant with possibly some new paper tape or card entries containing external factors which would enable efficiencies or complex analyses to be completed by the machine in its spare time. Heat cycle efficiencies in a power station are calculated by one 609 installation in the U.S.A.

4. Indirect Control of Plant

Most installations made in the U.S.A. envisage that the next step towards computer control is done by keeping the human operator in the control loop with "advice" from the machine.

A steel rolling mill installation in the U.K. has been made using this indirect control method. In this application the inputs are relatively few, coming from measuring counters in the cut-up line equipment and the operators' manual controls. The calculation is considerable compared to that involved in a logging job and the output on numerous numerical displays is used as control data by the cut-up operators. The machine uses the measured length of the billet rolled with reference to the customer's required lengths and advises the operators where to cut for minimum waste. In fact the machine will only be working for a few seconds every few minutes as first installed. Time-sharing, which is used domestically inside the machine to deal with the various inputs, may later be applied on large scale to do a similar type of operation for other processes in the plant. As the first job can justify the machine financially, time-sharing here allows other jobs to be done virtually for the price of the extra input and output equipment; however, as mentioned earlier, sufficient internal storage capacity must be available and although only 1024 words are needed for the first application, the full 4096 store would have to be fitted to allow the extension of the machine's usage to other "simultaneous" processes.

5. Direct Control of Plant

The use of a computer for direct control of continuous processes in, for example, the chemical and power generating plants is hindered by several factors, apart from natural conservatism and reluctance on the part of potential users. It is difficult to justify large expenditures when the inaccuracies of measurements may make it impossible to know for certain whether the machine is effective or not; or, as in some plastic processes, although the potential savings may be enormous and measurable, the mathematics of the process are not known sufficiently to be able to apply a computer at all. As the potential savings in the latter type of process are so large, machines are being installed—particularly in the U.S.A.—to gather information at the same time as carrying out alarm, scanning and logging functions, so that in the course of time the user is able to build up his knowledge of the processes. By using the alternative programming technique, programs can be derived from experience that are potentially able to increase plant efficiency. In the first instance these may be applied to the plant by a human operator altering the set points of the existing controllers, but if the random fluctuations in the processes are frequent it will soon be obvious that the machine should alter the set points itself. Facilities and equipment are available already to do this in a fail-safe manner, which is, of course, essential.

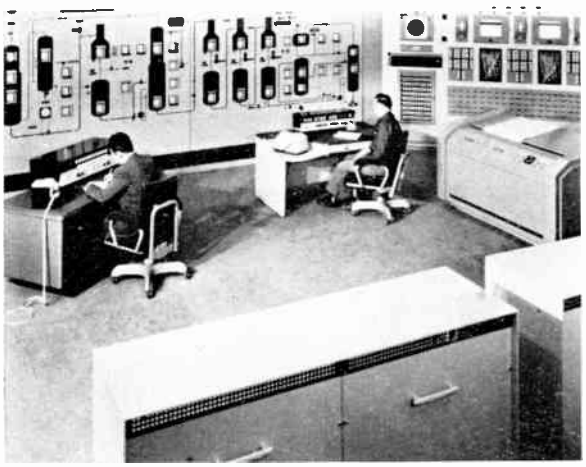


Fig. 3. Another on-line computer installation in the U.S.A.

A convenient common language output of the machine for this application is in the form of shaft rotations. The machine can incrementally change the shaft position in either direction but in the case of failure the shaft maintains its previous position and the plant will remain operating safely though probably not optimally. The shaft position can be converted to suit the local controller, whether it be electronic or pneumatic.

Several 609 systems have been installed in the U.S.A. on this basis and are expected to act in a purely data gathering and digesting role for several years before sufficient knowledge (and confidence) is available to allow the closing of the optimizing loop automatically.

Two power station installations have also been made in the U.S.A., and in this country an installation has been made by U.K. Atomic Energy Authority, as well as the steel plant mentioned above. Figures 2 and 3 show typical installations in the U.S.A.

6. The Future

The long-term target of computer control is to have sufficient experience to be able to design new plant which does not have the reservoirs and storage vessels that amount to such a large proportion of most existing plants. It should be possible to use the computer's ability to calculate at high speed and time share many control points so that the control of fast reacting processes may be done directly by the machine. The machine now becomes an essential part of the control loop itself. This would make possible not only increased through-put but, more important, considerable savings in the installation cost of the process plant itself, and of the part product that currently fills the plant before it can start to produce a useful output.

7. Conclusion

Although the digital computer has been in existence for over ten years, its application to industrial process control is in its infancy. The paper describes the operation of a system of a type currently being installed both in this country and the U.S.A. At the present time the hardware is more advanced than the knowledge of how to apply it, but the next ten years will see a new era of plant control established.

Manuscript first received by the Institution on 3rd March 1960 and in final form on 25th March 1961 (Paper No. 673).

The Growing Importance of Non-Destructive Testing

Non-destructive testing (n.d.t.) is a term which was quite unfamiliar to most people before the second World War. It was introduced to describe various types of test methods which were originally applied to the detection of flaws in metals, for instance, blow holes in castings, cracks in forgings, and cracks, lack of fusion, etc., in welds. Of these methods, radiography with x-rays was applied to industrial problems as early as 1919. Others, gradually evolved, include magnetic crack detection, eddy current testing, surface examination with the aid of penetrant oils and ultrasonic echo testing.

The development of each of these techniques during the last 20 years has had two important effects. Firstly, it was realized that the methods could be useful for solving many more problems in the engineering industry than metal testing. Secondly, technical progress has led to closer association of physicists and engineers engaged in this work, resulting in scientific meetings, in the formation of specialized societies, and finally in the organization in 1955 of the First International Conference for Non-Destructive Testing. The British National Committee for Non-Destructive Testing is a direct sequel to this Conference.

As regards development of n.d.t. techniques, it was realized that they can be extremely useful for various problems encountered in many industries. For instance, in the plastics industry they may be used for the detection of voids, and checking inclusions, both intentional and accidental; in the radio and television industry, for checking valve and tube assemblies; in the electrical component industry, for testing for flaws, homogeneity and the analysis of primary materials. The term n.d.t. is being continuously widened to include many methods better known to electrical engineers as "quality control". In fact, the border line today between "non-destructive testing" and "quality control" does no longer exist, as the interpretation of these terms varies from industry to industry and often overlaps.

The electronic engineer is concerned in the way in which n.d.t. in its widest sense can help his own technology. At present, the main significance of the method is in the manufacture of components, but it is useful in the testing of assemblies, connections and potted circuits.

To mention some examples, in capacitor manufacture we are concerned with the measurement of the dielectric constant, loss factor, insulation resistance and breakdown strength of the dielectrics, as well as with testing of oxide films and electrode material. Most of these tasks are based on the measurement of conductivity and capacitance. Radiography may be used to test for defects and dimensional control of concealed measurements. The fundamental methods used are often conventional but their application for efficient production testing will depend on the design of automatic recording and sorting equipment. The records obtained may be fed to suitable computers to facilitate statistical quality analysis and control. In some cases closed loop automatic control is already feasible.

For the manufacture of chokes and transformers, n.d.t. methods may be applied to magnetic sorting of the core material. In microwave tube manufacture, radiography is extensively used for electrode design and assembly. Military packaged stores can also be conveniently checked radiographically.

Ultrasonic methods are being applied for testing of brazed and soldered joints; the limitation here is often in the size of the probe that can be applied.

For the measurement of thickness, a variety of n.d.t. methods are available, some of which lend themselves to automatic production sorting and control. Thus, metal foil, plastic and paper sheet for capacitor manufacture are tested by radiation, pneumatic and eddy current methods, all of which have been developed for continuous plant-use both for measurement and control, and have the advantage of being "non-contacting". Similarly, insulating films on metals and metallic films on insulators can be measured continuously.

These are only a few of the applications of n.d.t. methods, namely those which are of interest to the electronic engineer. Many more applications have been found in the field of heavy electrical engineering.

It must be realized that only comparatively few physical principles have hitherto found application in n.d.t.-plant instrumentation. New principles are, however, being tried by the physicist, and some of the instruments evolved in the laboratory find their way eventually into the production plant. The development of instruments used in the laboratory for production is a slow process, yet its importance for the user- and for the instrument-industries is beyond question. For instance, thermal and electrical conductivity measurements are among those which have hardly yet been used in plant instrumentation.

Interpretation

The problem of interpretation is one of the major difficulties in non-destructive testing. The electrical engineer usually finds it easier to formulate his acceptance standards than the mechanical engineer or the metallurgist, who are concerned with castings and welds. Nevertheless, it is important to realize that, in general, the more sensitive a method becomes, the more difficult it is to know what to do with the test result in practice. This may often lead to a healthy tussle between the production- and the inspection-departments.

Professional Societies

Non-destructive testing as a part of engineering inspection was co-ordinated in this country in 1935 by the foundation of the Joint Committee on Materials and Their Testing. This Committee had a membership of some 22 societies most of which were only marginally interested in non-destructive testing.

In 1941, the Institute of Physics formed their Industrial Radiology Group which for the first time brought together all interested parties in this field. These meetings assumed considerable importance during the war in the testing of armaments, ships and aircraft. This group (since 1954 under the name of the Non-Destructive Testing Group), is still very active, and for the last few years it has been the Institute's policy to specialize on the more fundamental aspects of material testing and the principles employed in n.d.t. In this respect it can be regarded as a pioneer society in the art.

The engineering interests in this country are looked after by the Society of Non-Destructive Examination and the Non-Destructive Testing Society of Great Britain (both founded in 1954). The former holds informal meetings 5–6 times a year (at present mainly in London), followed by talks and discussions. Its aim is to bring together people responsible for n.d.t. laboratories and managements, and to discuss the engineering, economic and production aspects of n.d.t. as well as techniques.

The Non-Destructive Testing Society of Great Britain organizes lectures and discussions, mainly on techniques, in many centres throughout the country. This Society has produced the first specialized Journal in this country: *The British Journal of Non-Destructive Testing*. This is valuable as a forum for British contributions on this subject as hitherto these were scattered in many journals or appeared in the well-established American periodical.

Apart from these three specialized societies, a number of other institutions, including of course the Brit. I.R.E. who dealt with the subject at the 1954 and 1957 Conventions on Industrial Electronics, have maintained their interest in n.d.t. An opportunity for bringing all these interests together occurred after the First International Conference on Non-Destructive Testing.

International Co-operation

The First International Conference on Non-Destructive Testing came about mainly through the efforts of Professor Homes, of the University of Brussels, and was held in Brussels in 1955. Its members were mostly individuals as only a few societies and countries were represented. Although its organization was far from perfect, and the co-ordination of the subjects covered naturally primitive, nevertheless, over 500 members met and the Conference was very successful. It demonstrated not only interest and will to co-operate, but disseminated knowledge of techniques, applications, economy and new requirements. It clearly showed also three defects, namely, lack of national co-ordinating bodies in many countries, and wide variations in both acceptance standards and in technical competence.

To progress matters further, national co-ordinating bodies were necessary and this was recommended in a resolution of the first conference to all participating countries.

The British N.D.T. Societies took their cue from this resolution and with the approval of the existing Joint Committee for Materials and Their Testing, the British National Committee for Non-Destructive Testing was formed in 1957. This Committee, under the present chairmanship of Dr. L. Mullins, now comprises more than 20 societies interested in n.d.t. and is an active body meeting regularly, several times a year. Its objects are to promote discussion, to assist member institutions in the presentation of papers and symposia and in the national co-ordination of these events. It represents the United Kingdom in the field of n.d.t. in international meetings.

The Second International Conference was held in Chicago in 1957 ; in March of this year the third conference took place in Tokyo and the importance of the meeting to engineers from all over the world is fully established.

The Fourth Conference is being arranged by the Institution of Mechanical Engineers and will take place from 9th–13th September 1963.

The forthcoming London conference promises to be large and will include all aspects of Non-Destructive Testing, from basic physics to applications in manufacturing industries, transport and building. Papers will be limited to 5000 words and synopses must reach the organizing committee by February, 1962. Overseas papers should be submitted, in the first instance, to the National Committee of the country of origin where such a Committee exists. The British Institution of Radio Engineers is directly concerned in the organization of this Conference as a member of the British National Committee; members of the Institution who intend to submit papers to the Conference are invited to contact the Technical Committee at 9 Bedford Square. The London Conference will bring together many electronic engineers interested in this technique and will provide an opportunity for British universities, technical colleges, scientific societies and industry to be host to the engineers of the world.

A. NEMET.

U.D.C. 621.385.832

On the Geometric Transformations at the Foundation of all Cathode-Ray Tube Design

Part 1—Transformations which are Conformal with respect to Space Charge †

By

HILARY MOSS,
Ph.D., (*Member*) ‡

Summary: Starting from very fundamental postulates, and by use of scaling theory, this paper deduces the relationships between all the basic parameters of any c.r.t., such as overall length, screen diameter, spot size, beam voltage, beam current, cathode loading, grid drive, etc., etc. The theory is exact within its postulate framework for transformations in which the beam angle θ is invariant. On the other hand it is proved that for *any* transformation even if aberrations are present, provided *only* the beam angle diminishes, then the derived tube will have a performance *at least* equal to that predicted.

Practical illustrations of the use of the theory are included, showing how easily we may deduce many relationships of importance in c.r.t. design work. An Appendix justifies the validity of the basic assumption—that the deflection defocusing is proportional to $f(\lambda)z\theta$. It is shown that with proper interpretation this expression applies both to magnetic and electric deflection.

List of Symbols, Notation, Definitions

V Generic symbol for voltage. Volt-velocity of electron beam unless otherwise qualified.
I Current in focused beam striking screen, i.e. *beam* current.
I_c Total cathode current.
z Distance from centre of deflection to screen centre.
βz Distance from focusing lens centre to screen centre.
r_i Radius of electron beam at centre of deflection.
βr_i Beam radius at beam trimming aperture in final focusing lens.
r_s Radius of undeflected electron spot on screen.
λ Scan angle. Angle between tube axis and beam after deflection.
x Deflection on screen measured normal to beam axis, for scan angle λ.
e/kT Electron charge/Boltzmann constant × cathode temperature.
θ Semi-convergent angle of beam.
ρ_c Cathode loading (emission density).
ρ_0 Peak density in focused spot.
k Scaling factor. k always positive.
a Index of scaling factor.
n Index of scaling factor.
D Diameter of modulator hole on triode.
α Semi-angle of rays emerging from triode.
V_d Grid drive from cut-off.
V_c Grid cut-off voltage.
u Distance from crossover to final focusing lens (Fig. 3).
M Geometric magnification of electron gun, $M = \beta z/u$ for gun of Fig. 3.
m Index of scaling factor describing cathode loading ratio between derived tube and prototype. Eqn. (9).
g Index of scaling factor describing voltage ratio between derived tube and prototype. Eqn. (13).
p A constant equal to zero for transformation at constant spot size, and equal to unity for transformation at variable spot size. Eqn. (11).

Notation:

The suffix 1 applied to any quantity means that this quantity relates to the prototype tube.

The suffix 2 applied to any quantity means that this quantity relates to the derived tube.

Thus the form $D_2/D_1 = k^n$ means that the ratio of modulator hole diameters on derived and prototype tubes is k^n. An alternative notation used is simply $D \times k^n$. Similarly for any parameter.

† The exact meaning of "conformal" will become clear after study of the text. It means that all the transformations treated in this part of the paper are accurate even if the electron beams are very dense and are subject to significant space charge forces. A system of transformations can be worked out, which seem of practical value, but which would *not* be accurate in the presence of appreciable space charge. These are not treated here, but may form the subject of Part 2.

‡ Westinghouse Electric Corporation, Elmira, New York, U.S.A.

Definitions:

An equi-angular transformation is defined as one in which θ does not change.

A bi-directional transformation is defined as one in which k can assume *any* positive value from 0 to ∞.

A mono-directional transformation limits k either by the condition $k > 1$ or $k < 1$ depending on the value of a, so that always $\theta_2/\theta_1 < 1$.

The beam radius, r_s, means the radius at which the current density has fallen to any defined fraction of its peak on-axis value. What fraction is immaterial, provided it is constant in both prototype and derived tubes.

1. Introduction—Objectives and Limitations

Workers striving to advance the design of cathode-ray tubes may press their attack on two fronts. Firstly, they may study the performance of the basic tube components—essentially lenses—with a view to reducing aberrations. Such an approach, if successful, would enable wider beams to be employed—so increasing the spot density or perhaps permitting operation at lower voltage or reduced cathode loading. Any progress here must be regarded as of a rather fundamental nature and can clearly be applied "across the board" to improve the performance of many electron-optical devices.

Although this fundamental approach is obviously appealing, we must recognize two drawbacks at the outset. The record shows that progress along these lines has been very slow and arduous. Furthermore, even if any advance is achieved, it leads to no clarification of the overall design problem. The cathode-ray tube designer is still left lost in a sea of detail, and successful work on lens aberrations gives no clue as to how these components may be assembled in a manner calculated to optimize the overall tube performance.[1]

The second and quite distinct line of attack is the one chosen in this paper. Basically, it depends on the methods of scaling and similitude. As a very trivial example, consider the operation of reducing every dimension of a cathode-ray tube in constant proportion, maintaining constant all applied voltages at corresponding points. It can be shown quite rigorously that all the beam angles are preserved in this transformation.[2, 3] The fluorescent spot size decreases in the same proportion while the total current remains constant, so that the density in the focused electron beam rises as the square of the linear scaling factor.

In this paper we shall study such scaling changes but of a more sophisticated nature. The operation of scaling may be defined in the following manner. We start with some prototype cathode-ray tube and then "perturb" some of its dimensions or operating parameters so changing its performance. This establishes a derived tube. The process of conversion from prototype to derived tube will be termed a "transformation". It will be found that an infinity of scaling transformations exist because it so happens that there are fewer defining equations than there are variables. To limit the field of study, therefore, we have to make some arbitrary decisions about how certain of the variables are to change. These must be based on experience in the light of the changes sought.

Transformation theory finds its most obvious use in deriving some new tube designed to satisfy a new set of operating conditions. However, the whole process has a much more important purpose, for by comparing the relative characteristics of the derived and prototype tubes, one gains a clear insight into the ultimate basis of c.r.t. design. Embedded in the relationships will be found the connections between such things as cathode loading, tube length, beam voltage, spot size, etc. Problems of the following type are readily resolved. Given a tube of fixed screen diameter, how may we design another tube of the same screen diameter, but having a spot k times as large, giving the same writing speed? What are the operating characteristics of this tube? Innumerable questions of this nature can be solved by transformation theory. Obviously, therefore, this subject occupies a central position in the technique of cathode-ray tube design. It is surprising that it has so far received so little attention in a formal way, although in a broad sense some of the transformations to be discussed are well known empirically.[3] It is the purpose of this paper to place them on a proper scientific foundation.

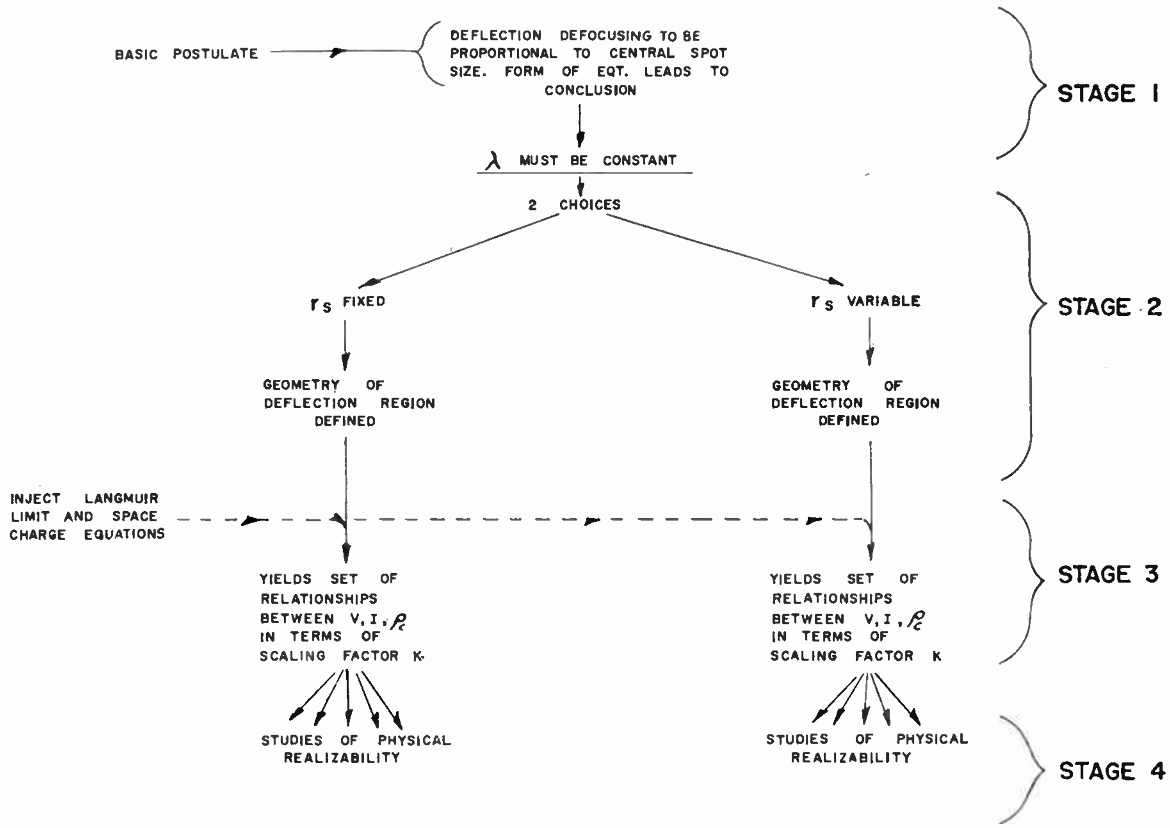
Transformation theory is the tool *par excellence* in the study of overall design. Largely it derives its power from a convenient ability to ignore all detail and to concentrate only on what is essential. This is a universal characteristic of all methods based on such things as scaling and dimensional homogeneity.

2. The Overall Logical Progression

Table 1 shows the logical flow diagram for our procedures. A basic assumption throughout the whole paper is that on any derived tube the ratio of deflected spot size to central spot size is the same as it is in the prototype, for corresponding points on the screen surfaces. Mathematically, this simply means that if we transform the central spot size by multiplying it by k, we also simultaneously transform the linear increase in spot size on deflection (the "deflection defocusing") by the same quantity k. Using this fact together with a very well-founded expression for the deflection defocusing in terms of scan angle, beam width and beam length, it is proved that the scan angle λ must be constant. This conclusion also depends on foreknowledge of the form of the equation re-

Table 1

Logical Flow Diagram



lating to the influence of space charge on the spot size. This constitutes Stage 1 in Table 1.

Stage 2 in the reasoning recognizes that the central spot radius r_s may remain fixed or may be variable. These choices uniquely define the geometry of the deflecting region. To carry the argument further, we now inject two equations which dominate the beam performance. The first is D. B. Langmuir's well-known density limit equation.[4] The second is the space charge limit equation.[5, 6, 7] These two equations connect the three variables, beam voltage V, beam current I, and the cathode loading ρ_c. Since there are only two equations with three variables, it is obvious that an infinity of solutions exist. To proceed further, we therefore make arbitrary choices on how any one of these variables may change, whereupon the remaining two become defined. These choices are made in the light of experience to give transformations which seem of practical value. All this constitutes Stage 3 on the reasoning in Table 1. Finally, for each of these solutions we must investigate their physical realizability. This constitutes Stage 4.

In what follows we shall study these four stages in detail. We shall then comment on their significance and show how important conclusions in the whole field of c.r.t. design can be reached through this mode of reasoning.

2.1. *Stage 1*

The starting point of the whole analysis is an expression which will, with reasonable accuracy, relate the essential geometry of the deflecting region with the deflection defocusing. Figure 1 shows what constitutes this essential geometry. We suppose that the

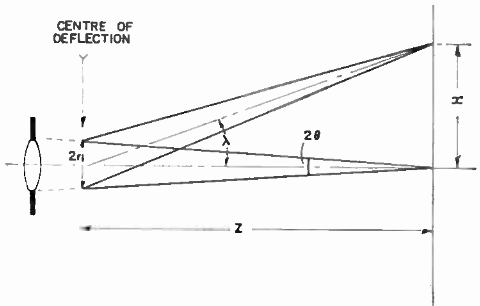
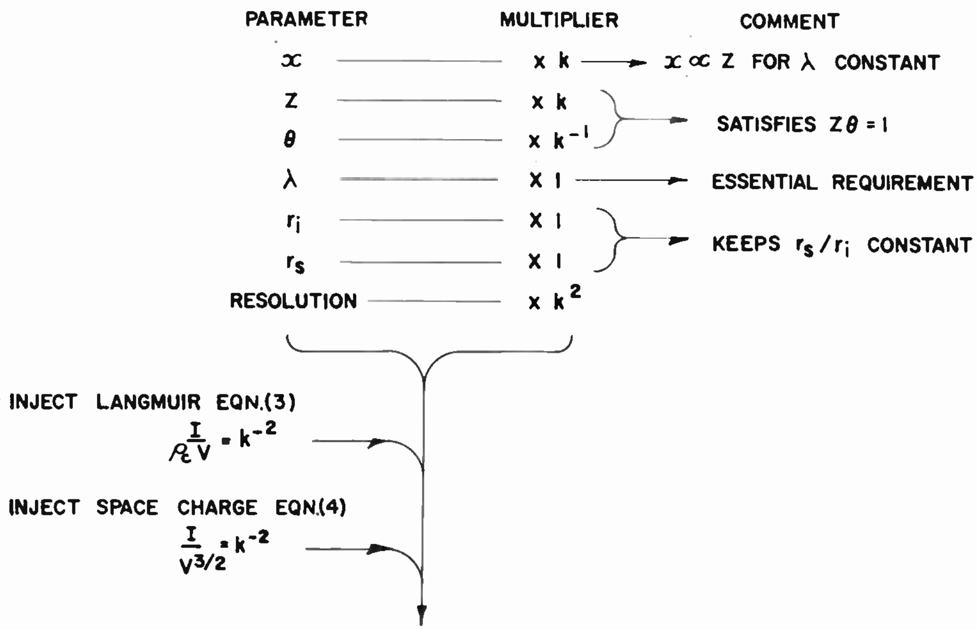


Fig. 1

deflecting system, either electromagnetic or electrostatic, turns the beam through an angle λ from the tube axis. We ignore the curvature of the electron beam in the deflecting region, and postulate a centre of deflection, which is supposed fixed. We suppose that the beam has a radius r_i at the centre of deflection. At the screen, distant z from the centre of deflection, the spot radius is r_s. The semi-convergence angle of

Table 2 Details of Stages 2 and 3 for the Case where r_s is Fixed



TRANSFORMATION TABLE

DEFINED OPERATION	UNIQUE CONSEQUENCES		
$V \times k^{n}$	$I \times k^{\frac{3}{2}n-2}$	$\mathcal{R}_c \times k^{\frac{1}{2}n}$	$\mathcal{R}_o \times k^{\frac{3}{2}n-2}$
$I \times k^{n}$	$V \times k^{\frac{2}{3}n+\frac{4}{3}}$	$\mathcal{R}_c \times k^{\frac{1}{3}n+\frac{2}{3}}$	$\mathcal{R}_o \times k^{n}$
$\mathcal{R}_c \times k^{n}$	$I \times k^{3n-2}$	$V \times k^{2n}$	$\mathcal{R}_o \times k^{3n-2}$

the beam is denoted by θ and since θ is very small, we may write without sensible error $r_i = z\theta$. The absolute deflection on the screen from the screen centre will be denoted by x.

In Appendix 1 we study beam deflection by sharply terminated electrostatic or magnetic fields, showing that with reasonable approximations, the linear increase in spot size on deflection is proportional to $\lambda^2.z\theta$. The most dubious approximation here resides in the presence of λ^2 but in the analysis which follows (Part 1 of the whole paper), we do not need to be so explicit and it suffices to write the much more general expression—deflection defocusing is proportional to $f(\lambda)z\theta$ where $f(\lambda)$ is entirely arbitrary. This is certainly a satisfactory engineering approximation.

From our basic postulate, we require the deflection defocusing to be proportional to the spot size. Thus on the derived tube,

$$f(\lambda_2)z_2\,\theta_2 \propto r_{s2} \qquad \ldots\ldots(1)$$

Similarly for the prototype tube,

$$f(\lambda_1)z_1\,\theta_1 \propto r_{s1} \qquad \ldots\ldots(2)$$

To make the transformation conformal with respect to space charge at the screen, eqn. (4) requires that $r_{s2}/r_{s1} = r_{i2}/r_{i1} = z_2\theta_2/z_1\theta_1$ so that dividing (1) by (2) proves that λ *must be constant*.† This is an interesting conclusion due essentially to the necessity

† More precisely it proves that $f(\lambda_2) = f(\lambda_1)$. This however necessitates $\lambda_2 = \lambda_1$, since it is assumed that the same field shape is being used in both prototype and derived tubes.

of making the ratio r_s/r_i invariant during the transformation. This last condition is necessitated by the form of the normalized space charge equation (4) which would become nonanalytic if r_s/r_i varied.

This completes Stage 1 of the reasoning.

2.2. *Stage 2*

We now recognize that two situations can arise. In the first place, the central spot size can remain fixed. In the second place, it can vary. These two possibilities are shown in the bifurcation in Table 1. Since r_s must be proportional to r_i which in turn is equal to $z\theta$, it follows that our two possibilities are either the system $z\theta = 1$ or the system $z\theta = k$.

Figure 2 shows a diagram illustrating these two transformation systems. The line SPT is the rectangular hyperbola $z\theta = 1$. For a particular tube let P be the operating point corresponding to specific values of z and θ. The transformation $z\theta = 1$ simply means that P wanders along this hyperbola. Numerically, this transformation may be represented by the symbolism $[z \times k,\ \theta \times k^{-1}]$ where k is the scaling factor and where $k \geq 0$. The second situation, where r_s varies, corresponds to the condition $z\theta = k$. This

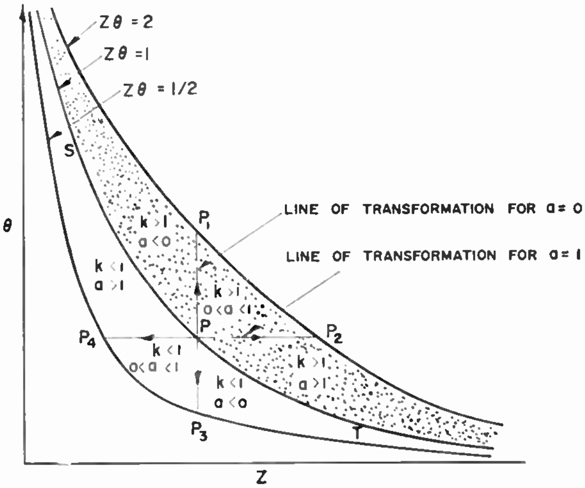


Fig. 2. The regions of the transformation. $[z \times k^a,\ \theta \times k^{1-a}]$ for $-\infty < a < +\infty$

Table 3 Details of Stages 2 and 3 for the Case where r_s is Variable

PARAMETER	MULTIPLIER	COMMENT
x	$\times k^{a}$	$x \propto Z$ FOR λ CONSTANT
Z	$\times k^{a}$	SATISFIES $Z\theta = k$
θ	$\times k^{1-a}$	
λ	$\times 1$	ESSENTIAL REQUIREMENT
r_i	$\times k$	KEEPS r_s/r_i CONSTANT
r_s	$\times k$	
RESOLUTION	$\times k^{2(a-1)}$	

INJECT LANGMUIR EQN.(3)

$$\frac{I}{\rho_c V} = k^{4-2a}$$

INJECT SPACE CHARGE EQN.(4)

$$\frac{I}{V^{3/2}} = k^{2(1-a)}$$

TRANSFORMATION TABLE

DEFINED OPERATION	UNIQUE CONSEQUENCES		
$V \times k^{n}$	$I \times k^{\frac{3}{2}n + 2(1-a)}$	$\rho_c \times k^{\frac{1}{2}n - 2}$	$\rho_o \times k^{\frac{3}{2}n - 2a}$
$I \times k^{n}$	$V \times k^{\frac{2}{3}n + \frac{4}{3}(a-1)}$	$\rho_c \times k^{\frac{1}{3}n + \frac{2}{3}a - \frac{8}{3}}$	$\rho_o \times k^{n-2}$
$\rho_c \times k^{n}$	$I \times k^{3n - 2a + 8}$	$V \times k^{2(n+2)}$	$\rho_o \times k^{3n - 2a + 6}$

is a little more complicated. To illustrate it we have drawn two additional rectangular hyperbolae (in Fig. 2), namely the curve $z\theta = 2$ and the curve $z\theta = \frac{1}{2}$. These correspond to $k = 2$ and $k = \frac{1}{2}$ respectively. Starting from P, *all* transformations for $k = 2$ must end somewhere on the upper hyperbola and for $k = \frac{1}{2}$ must end somewhere on the lower hyperbola. Since k can have any value from 0 to infinity, it is clear from Fig. 2, that P can transform into any point in the first quadrant of the z, θ axes. In terms of a scaling factor k, this transformation may be rather generally written in the form $[z \times k^a,\ \theta \times k^{1-a}]$, $-\infty < a < +\infty$. This satisfies the requirement $z\theta = k$ for all values of a.

The choice of the index a defines the path of transformation. For example, the transformation $P_1 \leftarrow P \rightarrow P_3$ represents the condition in which z is unchanged so that $a = 0$. Similarly the transformation $P_4 \leftarrow P \rightarrow P_2$ represents the situation in which θ is unchanged and $a = 1$. For illustrative purposes, the curves of Fig. 2 are divided into areas in each of which are inserted the relevant values of k and a. For the dotted area $k > 1$. The transformation paths traverse these areas. This diagram vividly portrays the significance of a. If it lies near 0, the transformation represents the case where z changes only slightly. If it lies near unity, it represents the case where θ changes only slightly. All possible transformations are contained in the first quadrant of the z, θ axes since the $-z$, $-\theta$ quadrant is excluded by the physical requirement that z and θ are inherently positive.

For each of the transformations, r_s fixed and r_s variable, the geometry of the deflecting region is now uniquely defined. This is indicated at the top of Tables 2 and 3 respectively.

At this stage, we introduce the two basic equations of the beam physics. The first is Langmuir's equation

$$\rho_0 = \rho_c \left\{ \frac{eV}{kT} + 1 \right\} \sin^2 \theta \qquad \text{......(3)}$$

The second is the normalized space charge equation[6, 7]

$$\frac{I^{1/2}}{V^{3/4}} \cdot \frac{z}{r_i} = f\left(\frac{r_s}{r_i}\right) \qquad \text{......(4)}$$

Recollecting that $\rho_0 \propto I/r_s^2$, these equations lead to the results:

From Table 2		*From Table 3*	
$\dfrac{I}{\rho_c V} = k^{-2}$	...(3a)	$\dfrac{I}{\rho_c V} = k^{4-2a}$	...(3b)
$\dfrac{I}{V^{3/2}} = k^{-2}$	...(4a)	$\dfrac{I}{V^{3/2}} = k^{2(1-a)}$	...(4b)

where for all cases $\dfrac{eV}{kT} \gg 1$

These results are entered on the left-hand side of Tables 2 and 3. This concludes Stage 2 in the reasoning.

2.3. *Stage 3*

Since only two equations are defining three variables, an infinity of solutions exist, and at this stage we have to arbitrarily define the locus of any one of these variables so that the remainder become determined. At the bottom of Tables 2 and 3 are shown the inevitable and unique consequences of making three separate choices as to how each of these variables shall change. The transformation tables in Tables 2 and 3 contain all the consequences of the analysis so far, in a most succinct form. They are exact within their postulate framework.

2.4. *Stage 4—Physical Realizability*

So far our scaling transformations have used only rather basic geometrical ideas in association with two equations concerning beam physics. Now it is necessary to consider how a practical electron gun operates, and especially to study how it may be scaled to achieve the conditions defined by Tables 2 or 3.

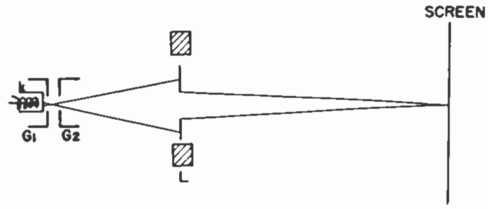


Fig. 3 (a)

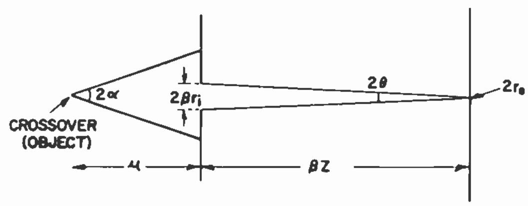


Fig. 3 (b)

To do this it is necessary to establish a simple "model" of an electron gun. Such a model is indicated in Figs. 3 (a) and 3 (b). Figure 3 (a) represents a very simple type of gun. The conventional triode formed by cathode K, modulator G_1 and first accelerator G_2 serves to produce a crossover which is imaged by focusing lens L to form a spot on the fluorescent screen. The lens L is assumed to be "thin". For simplicity the system of Fig. 3 (a) uses a magnetic lens but this is in no way essential to the argument.

A beam trimming aperture is included in the plane of the lens L. Figure 3 (b) shows the essential geometry of this model. The focusing lens to screen distance is βz where β is a positive constant equal to about 1·2 in typical tubes. This allows for space in which to mount the deflection system. The lens to centre of deflection separation is now $(\beta-1)z$. The beam radius at the lens plane becomes βr_i. Whenever, in the subsequent working, we scale z, it is assumed that βz is scaled in exactly the same proportion.

2.4.1. The semi-empirical equations of the electron gun

To proceed with the reasoning it is now necessary to introduce some semi-empirical equations which describe the performance of an electron gun. The derivation of these results has been given elsewhere.[8, 9] They are dimensionally correct. Since in this paper our concern is wholly with *relative* values, any uncertainties in numerical multipliers have no effect. This entitles us to reasonable confidence in the analysis. The following results are used:

The cathode loading ρ_c is proportional to

$$V_d^{3/2}/D^2 \qquad \ldots\ldots(5)$$

The total cathode current I_c for *constant triode proportions* is proportional to

$$V_d^{7/2}\,.\,V_c^{-2} \qquad \ldots\ldots(6)$$

The final spot radius r_s is proportional to the crossover size, which is proportional to the triode size, to the geometric magnification of the final focusing lens $\beta z/u$, and inversely proportional to the square root of the beam voltage V ......(7)

2.4.2. Space charge conformality at the crossover

Use of eqn. (4) has ensured that our transformations are conformal with respect to space charge at the *screen*. Similarly in order to ensure that any scaling transformation satisfies the space charge restriction at the *crossover* it is necessary and sufficient that the current in the crossover (i.e. the *total* cathode current I_c) is proportional to $V^{3/2}$ where V is the potential on the crossover forming anode.[2]

By result (6)

$$I_c \propto V_d^{7/2}\,.\,V_c^{-2} \qquad \ldots\ldots(6)$$

Now suppose that *in any changes to be made in the triode all proportions are maintained constant*. This will henceforth apply *always*. Then the cut-off voltage V_c is proportional to anode potential V.† Hence eqn. (6) may be re-written

$$I_c \propto V_d^{7/2} V^{-2} \qquad \ldots\ldots(6a)$$

Entering the requirement that $I_c \propto V^{3/2}$ in (6 a) gives

$$I_c \propto V^{3/2} \propto V_d^{7/2} V^{-2}$$

so that *inevitably*

$$V_d \propto V \qquad \ldots\ldots(8)$$

2.4.3. Derivation of gun parameters from the semi-empirical defining equations

As usual the suffix 2 relates a quantity to the derived tube and suffix 1 to the prototype.

Let

$$\frac{\rho_{c2}}{\rho_{c1}} \equiv k^m \qquad \ldots\ldots(9)$$

where k^m is the cathode loading multiplier obtained from the relevant section of Tables 2 or 3.

From result (5) it now follows that

$$\frac{(V_{d2})^{3/2}}{D_2^2} = k^m\,.\,\frac{(V_{d1})^{3/2}}{D_1^{\,2}} \qquad \ldots\ldots(10)$$

From result (7) it follows that

$$\frac{D_2}{\sqrt{V_2}}\,.\,\beta z = k^p\,.\,\frac{D_1}{\sqrt{V_1}}\,.\,u \qquad \ldots\ldots(11)$$

where $p = 0$ for transformation of Table 2

$p = 1$ for transformation of Table 3.

Eqn. (11) takes note of the fact that we have postulated maintenance of constant triode proportions, so that D_2/D_1 is also the ratio of triode and crossover sizes.

Entering result (8) in eqn. (10) gives

$$\frac{V_2^{3/2}}{D_2^2} = k^m\,.\,\frac{V_1^{3/2}}{D_1^2} \qquad \ldots\ldots(12)$$

For any particular transformation there will be a value of V_2/V_1 found from either Table 2 or 3. Let

$$\frac{V_2}{V_1} \equiv k^g \qquad \ldots\ldots(13)$$

The solution of eqns. (11), (12) and (13) is

$$\frac{D_2}{D_1} = k^{\frac{3}{4}g-\frac{m}{2}} \qquad \ldots\ldots(14)$$

$$\frac{u_2}{u_1} = k^{\frac{g}{4}-\frac{m}{2}-p+a} \qquad \ldots\ldots(15)$$

[Enter $a = 1$ for all transformations of Table 2]

Remembering that $V_d \propto V$, $I_c \propto V^{3/2}$ and that α is proportional to V_d/V_c for constant triode proportions, gives in addition

$$\frac{V_{d2}}{V_{d1}} = k^g \qquad \ldots\ldots(16)$$

$$\frac{V_{c2}}{V_{c1}} = k^g \qquad \ldots\ldots(17)$$

$$\frac{\alpha_2}{\alpha_1} = 1 \qquad \ldots\ldots(18)$$

$$\frac{I_{c2}}{I_{c1}} = k^{\frac{3}{2}g} \qquad \ldots\ldots(19)$$

† Neglecting the second-order effects of thermionic emission velocity and contact potential.

Equations (14), (15), (16), (17), (18) and (19) completely define the new gun conditions, after we enter the relevant values of m, g, a and p from Tables 2 or 3. When satisfied they are necessary and sufficient to ensure that the derived gun satisfies any transformation described in these two tables. Again note that D_2/D_1 in eqn. (14) is also the triode size ratio.

2.4.4. Invariance of the magnification M

With reference to Fig. 3 the magnification M is defined simply as $\beta z/u$. We now prove that for all these transformations this quantity is unchanged.

Case 1—Transformations at variable spot size—Table 3

From eqn. (15) the general expression for the u ratio is

$$u \times k^{\frac{g}{4}-\frac{m}{2}-p+a} \qquad \ldots\ldots(15a)$$

But the βz ratio is (from Table 3)

$$(\beta z) \times k^{a} \qquad \ldots\ldots(20)$$

Thus the M ratio is

$$M \times k^{p+\frac{m}{2}-\frac{g}{4}} \qquad \ldots\ldots(21)$$

and is independent of a.

Eliminating I between the basic defining eqns. (3 b) and (4 b) gives

$$\frac{V}{\rho_c^2} = k^4 \qquad \ldots\ldots(22)$$

Entering in (22) the V and ρ_c ratios *defined* by (13) and (9) gives

$$\frac{k^g}{k^{2m}} = k^4 \qquad \ldots\ldots(22a)$$

whence

$$\frac{g}{4} = 1 + \frac{m}{2} \qquad \ldots\ldots(22b)$$

Insertion of (22 b) in (21) shows immediately that M is always constant under this transformation where r_s is varied, since here $p = 1$.

Case 2—Transformations at fixed spot size—Table 2

The argument is similar. Eliminating I between the defining eqns. (3 a) and (4 a) gives

$$\frac{V}{\rho_c^2} = 1 \qquad \ldots\ldots(23)$$

Entering in (23) the V and ρ_c ratios defined by (13) and (9) gives

$$\frac{k^g}{k^{2m}} = 1 \qquad \ldots\ldots(23a)$$

so that now

$$\frac{g}{4} = \frac{m}{2} \qquad \ldots\ldots(23b)$$

Insertion of (23 b) in (21), recollecting that now $p = 0$, immediately shows again that M is constant.

The main interest in this result is that since z, βz and u change in the same ratio, it follows that *the ratio change in any one of these quantities is also the ratio change in the total tube length* (ignoring the small fixed base length).

2.4.5. Proof that the beam current I scales as required

It remains to prove that all the transformations indicated in Tables 2 and 3, followed by the defining eqns. (14) to (19), when applied to the model of Fig. 3 automatically yield the necessary beam current.

This is very easy. With reference to Fig. 3, since the emergent beam angle α is always constant in all these transformations (eqn. (18)), then it is necessary and sufficient that

$$I \propto \frac{(\beta r_i)^2}{u^2} \,.\, I_c \qquad \ldots\ldots(24)$$

Case 1—Transformations at variable spot size

Here the βr_i ratio is

$$(\beta r_i) \times k \qquad \ldots\ldots(25)$$

and the u ratio is

$$u \times k^{\frac{g}{4}-\frac{m}{2}-p+a} \qquad \ldots\ldots(15a)$$

while the I_c ratio is

$$I_c \times V^{3/2} \qquad \ldots\ldots(26)$$

Entering (25), (15 a), and (26) in (24) gives

$$\frac{I}{V^{3/2}} = k^{m-\frac{g}{2}+4-2a}$$

$$= k^{2(1-a)}$$

$$\text{since from (22b)} \qquad m - \frac{g}{2} + 2 = 0$$

This last result is eqn. (4b)—one of the basic defining equations—and is automatically satisfied.

Case 2—Transformations at constant spot size

Here the βr_i ratio is

$$(\beta r_i) \times 1 \qquad \ldots\ldots(27$$

and the u ratio is

$$u \times k^{\frac{g}{4}-\frac{m}{2}+1} \qquad (15b)$$

Entering (27), (15b), and (26) in (24) gives

$$\frac{I}{V^{3/2}} = k^{m-\frac{g}{2}-2}$$

$$= k^{-2}$$

$$\text{since from (23b)} \qquad m = g/2$$

This last result is eqn. (4a)—one of the basic defining equations—and is automatically satisfied.

These proofs depend on the correctness of the form (24) to represent the beam current. Except when $\beta r_i/u$ is constant, eqn. (24) implicitly assumes that the current density distribution in the ray cone from the triode is uniform. In practice this distribution is bell-shaped, and (24) applies only for limited variations in $\beta r_i/u$. Discussion on this matter is given later.

2.5. *Relationship Between Transformations of Table 2 (r_s fixed) and Table 3 (r_s variable)*

We may write:

transformation of Table 3

$$\equiv \text{transformation } [z \times k^a,\ \theta \times k^{1-a}]$$
$$\rightarrow \text{transformation } [z \times k^b,\ \theta \times k^{-b}]$$
$$\text{as } |a| \rightarrow |b| \text{ where } b \rightarrow \infty$$

Writing $k^b \equiv k_1$ gives $k^{-b} \equiv k_1^{-1}$ whereupon:

$$\text{transformation } [z \times k^b,\ \theta \times k^{-b}]$$
$$\equiv \text{transformation } [z \times k_1,\ \theta \times k_1^{-1}]$$
$$\equiv \text{transformation of Table 2}$$

Apparently then no sharp distinction exists, and we can regard the transformation at constant spot size as a limiting case of the more general transformation where the spot size is variable and $|a| \rightarrow \infty$ while $k \rightarrow 1$.

2.6. *Extensions of the Results to Any Form of Electron Gun With or Without Aberrations*

So far, the physical realizability of the transformations of Tables 2 and 3 has been proven only for the very simple type of electron gun shown in Fig. 3. In addition aberrations have been neglected. Now we shall consider to what extent these transformations can be realized for any type of gun system, and when aberrations may be present.

Figure 4(a) illustrates the essentials of a c.r.t. embodying a more complex form of gun. This can have any shape whatever and can use any number of anodes. It is necessary *only* that it can be divided into a "generator" (usually a triode), producing the cross-over (object), and into a subsequent system of electrodes which focus the beam. These two essential elements of any gun will be termed "the triode" and "the lens system" respectively.

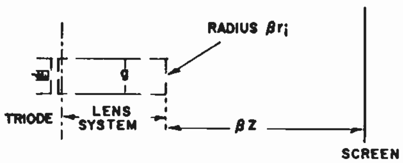


Fig. 4 (a)

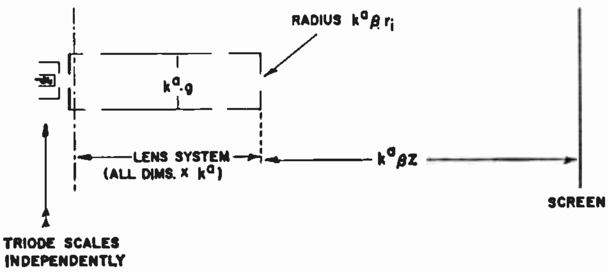


Fig. 4 (b)

2.6.1. Assumption 1—Aberrations negligible in both derived and prototype tubes

This restriction will be removed later, but is introduced to simplify the arguments.

Suppose that any transformation from Table 3 is applied to the c.r.t. of Fig. 4 (a). It has been proved that any transformation maintains constant geometric magnification. This requirement is satisfied by uniformly changing *all* dimensions of the prototype tube except for triode size which can change in any way, or may remain fixed.

Such a transformation is shown in Fig. 4 (b). Table 3 shows the scaling factor for z as k^a, which also therefore becomes the scaling factor for the rest of the tube dimensions except for the triode.

Figure 4 (b) represents an intermediate stage in developing the derived tube. Now consider its characteristics. Let all its applied voltages be in the same proportion (k^g – eqn. (13)) as those at corresponding electrodes in the prototype. Then by the principle of voltage similitude, if these voltages focus the prototype tube, so also will they focus the tube of Fig. 4 (b).

To complete the transformation, it is only necessary to scale the triode of Fig. 4 (a) according to eqn. (14). This forces the spot size to be multiplied by k as needed. Such a change has no effect on the focus condition, as it does not involve the lens system.

Finally as a last step, the beam trimming stop is made k times as large as in the prototype. This may be achieved by multiplying the stop diameter in Fig. 4 (b) by k^{1-a}. Since this aperture will be in field-free space, this final step causes no change in the focus condition.

This proves the following result. If aberrations are neglected then *any* transformation in Table 3 may be physically realized by the change from Fig. 4 (a) to Fig. 4 (b), followed by adjustment of the triode size and adjustment of the beam trimming stop size. Obviously these processes are *always* possible, however complex the electron gun may be, provided *only* the "object" generator and the focusing system are separable.

Exactly similar arguments show that *any* transformation from Table 2 can also be physically realized for *any* electron gun. Here however the dimensions of the prototype are multiplied by k, instead of by k^a.

3.2. *Transformations Yielding Constant Spot Writing Speed at Constant Screen Size*

What transformations maintain constant spot writing speed (I/r_s constant†) and also maintain constant screen size?

Since the screen size is to be constant it follows from Table 3 that $a = 0$ (since x is $\times k^a$). From row 2, the I/r_s constant condition requires k^{n-1} constant, i.e. $n = 1$. Accordingly, we have $z \times 1$, $\theta \times k$, $r_i \times k$, $r_s \times k$, resolution $\times k^{-2}$, $V \times k^{-2/3}$, $\rho_c \times k^{-7/3}$, $I \times k$, $\rho_0 \times k^{-1}$, $D \times k^{2/3}$, $u \times 1$, $V_d \times k^{-2/3}$, $V_c \times k^{-2/3}$, $\alpha \times 1$, $I_c \times k^{-1}$. This transformation is mono-directional and applies safely for $k < 1$.

Let us investigate the application of row 1 on Table 3 to this same problem. Now from column 2 we must have $\frac{3}{2}n + 2 - 2a - 1 = 0$. But $a = 0$ as before so that $n = -2/3$. Thus ρ_c is $\times k^{-7/3}$, $V \times k^{-2/3}$ and the previous transformation is repeated. Use of row 3 leads to the same results. Again only one transformation satisfies the postulates. Note that the tube length is unchanged—a fact which would be expected since λ is always constant and we have postulated constant screen size, so that z must be constant. Practical limits to the increased resolution are set by the cathode loading increase and by the mechanical problems of reducing the triode size beyond a certain point.

3.3. *Equi-Angular Transformations*

As we have seen, these are of special interest since alone they are exact and bi-directional. The necessary and sufficient condition is $a = 1$. This imposes immediately the requirement that the screen diameter is multiplied by k on any equi-angular transformation.

The whole scheme is now Table 3 with $a = 1$. As an example let us inquire what happens if we again require constant writing speed. Row 1, column 2 now gives $\frac{3}{2}n - 1 = 0$, i.e. $n = 2/3$ so that $V \times k^{2/3}$, $\rho_c \times k^{-5/3}$, $\rho_0 \times k^{-1}$, $z \times k$, $\theta \times 1$, $r_i \times k$, $r_s \times k$, $I \times k$, resolution $\times 1$. Eqns. (14) to (19) complete the transformation by giving $D \times k^{4/3}$, $u \times k$, $V_d \times k^{2/3}$, $V_c \times k^{2/3}$, $\alpha \times 1$, $I_c \times k$.

As a final illustration let us inquire into an equi-angular transformation giving constant spot density. Row 2, column 4 gives $n - 2 = 0$. This defines $I \times k^2$, $V \times k^{4/3}$, $\rho_c \times k^{-4/3}$, $z \times k$, $\theta \times 1$, $r_i \times k$, $r_s \times k$, resolution $\times 1$. Eqns. (14) to (19) now give $D \times k^{5/3}$, $u \times k$, $V_d \times k^{4/3}$, $V_c \times k^{4/3}$, $\alpha \times 1$, $I_c \times k^2$.

Notice that Table 3 shows the impossibility of an equi-angular transformation maintaining constant spot size. It also shows that *no equi-angular transformation can give a gain in resolution.*

† This statement of the constant writing speed condition ignores variation of phosphor efficiency with voltage.

3.4. *Transformations Yielding Increased Resolutions*

Here we may apparently use either Table 2, or Table 3. It is of interest to explore both approaches. Let us require concurrently that the writing speed be constant.

(a) *Use of Table 2*

Here I must be constant so that row 2 requires $n = 0$ which gives $V \times k^{4/3}$, $\rho_c \times k^{2/3}$, $\rho_0 \times 1$. The rest of the transformation is $z \times k$, $\theta \times k^{-1}$, $r_i \times 1$, $r_s \times 1$, resolution $\times k^2$, $D \times k^{2/3}$, $u \times k$, $V_d \times k^{4/3}$, $V_c \times k^{4/3}$, $\alpha \times 1$, $I_c \times k^2$. This is safe for $k > 1$.

(b) *Use of Table 3*

Row 2 now requires $n = 1$ to keep I/r_s constant. This gives $V \times k^{4a/3 - 2/3}$, $\rho_c \times k^{2a/3 - 7/3}$, $z \times k^a$, $\theta \times k^{1-a}$, $r_i \times k$, $r_s \times k$, resolution $\times k^{2(a-1)}$. Eqns. (14) to (19) give $D \times k^{2a/3 + 2/3}$, $u \times k^a$, $V_d \times k^{4a/3 - 2/3}$, $V_c \times k^{4a/3 - 2/3}$, $\alpha \times 1$, $I_c \times k^{2a-1}$.

If in this last set of results we set $a = 2$ then the resolution becomes $\times k^2$ and exactly equals the figure from the transformation using Table 2. The corresponding results are $V \times k^2$, $\rho_c \times k^{-1}$, $\rho_0 \times k^{-1}$, $z \times k^2$, $\theta \times k^{-1}$, $r_i \times k$, $r_s \times k$, resolution $\times k^2$, $D \times k^2$, $u \times k^2$, $V_d \times k^2$, $V_c \times k^2$, $\alpha \times 1$, $I_c \times k^3$. This is safe for $k > 1$.

The transformations of Table 2 and 3 may be directly compared. Taking for instance $k = 2$ shows that for Table 2 the screen diameter has been multiplied by 2 and the spot size unchanged. For Table 3 the spot size has doubled and the screen diameter has quadrupled so that resolution gain is 4 in both cases. The transformation of Table 3 gives a tube twice as long as that from Table 2. It requires four times the prototype voltage—that from Table 2 about 2½ times the prototype voltage.

4. Discussion on Part 1

All these transformations in Part 1 require that the scan angle λ is constant. It will have been noted that for exactitude we also require θ constant. The question now arises—what significance, if any, attaches to this necessity for angle constancy?

Fundamentally this invariance is deeply rooted in scaling theory. Whenever assertions are made about scaling of geometries it is apparent that angle constancy is an essential prerequisite. If, for example, we state that for a constant material density the masses of two objects of linear size ratio k, vary as k^3, it is of course implied that the *shape* of the objects is identical. This in turn implies angle constancy at corresponding points.

From a dimensional viewpoint, angles are truly dimensionless ratios. It follows that they must be preserved in any *exact* scaling transformation.

This explains the necessity for keeping λ constant, and why θ also must be constant for exact deductions.

By an ingenious argument involving the concept of a "safe" (i.e. conservative) transformation, it has been possible to permit *unidirectional* θ variations and to extract useful engineering data. However, this has nothing to do with scaling theory limitations as such. For exactitude, θ as well as λ, must be constant.

The question now arises—would it be possible to allow λ as well as θ to vary and establish a broader theory? Such transformations would certainly be inexact, but presumably could be constrained by the condition $\lambda_2/\lambda_1 < 1$ to yield "safe" answers. This is now under investigation.

Although some such technique is certainly possible it may not prove worthwhile from a practical standpoint. Perhaps it is preferable to accept the limitations of the conformal transformations herein discussed as yielding a homologous family of tubes all with constant λ parentage, and not to attempt to pass from one family to another *except by direct measurement.* Our scaling theory relates members of each family but has nothing to say about their relative family attributes.

5. Acknowledgment

This work was performed under the sponsorship of the Air Research and Development Command, United States Air Force, Contract 33(616)–6219.

I am indebted to Dr. I. Brodie of these laboratories for interesting discussions on this paper.

6. References

1. H. Moss, "The optimum design of electrostatically-deflected cathode-ray tubes", *J. Brit.I.R.E.*, **18**, pp. 485–91, August 1958.
2. B. Meltzer, "Electron flow in curved paths under space-charge conditions", *Proc. Phys. Soc. (Lond.)*, **62B**, pp. 813–7, December 1949.
3. H. Moss, "Simplification of cathode-ray tube design by the application of the theory of similitude", *J. Television Soc.*, **4**, pp. 209–19, March 1946.
4. D. B. Langmuir, "Theoretical limitations of cathode-ray tubes", *Proc. Inst. Radio Engrs*, **25**, pp. 977–91, August 1937.
5. B. J. Thompson and L. B. Headrick, "Space-charge limitations on the focus of electron beams", *Proc. Inst. Radio Engrs*, **28**, pp. 318–24, July 1940.
6. J. W. Schwartz, "Space-charge limitation on the focus of electron beams", *R.C.A. Rev.*, **18**, No. 1, pp. 3–11, March 1957.
7. H. Moss, "The cathode loading limit in circular beam electron devices", *J. Brit.I.R.E.*, **21**, pp. 35–9, January 1961.
8. H. Moss, "The electron gun of the cathode ray tube—part 1", *J. Brit.I.R.E.*, **5**, pp. 10–25, January 1945.
9. H. Moss, "The electron gun of the cathode ray tube—part 2", *J. Brit.I.R.E.*, **6**, pp. 99–128, June 1946.

7. Appendix 1

On the Theory of Deflection Defocusing

At the basis of the paper has been the following theorem:

"If an electron beam is deflected through an angle λ, the linear increase in spot size can be expressed in the form $f(\lambda)z\,.\,\theta$ where $f(\lambda)$ is an arbitrary function of λ."

This statement will now be justified, both for magnetic and electric fields, provided only that both are uniform, wholly normal to the axis, and sharply bounded. Many practical fields—especially those designed for high performance instrument tube types—approximate these conditions, so that the result has more than academic significance.

Although this same expression can apply equally to either electrostatic or magnetic deflection, yet there is an important difference in the two cases. For magnetic deflection the result is true regardless of the physical size of the field. With electrostatic deflection, on the other hand, the physical size of the electric field is an essential parameter.

7.1. *Case (1)—Magnetic Deflection*

Figure 5 shows the tube axis ZZ′ and the region of deflection. The postulated uniform magnetic field is sharply bounded by the planes X_1Y_1, X_2Y_2. On account of the initial beam convergence (angle 2θ) the extreme electrons on the upper and lower edges

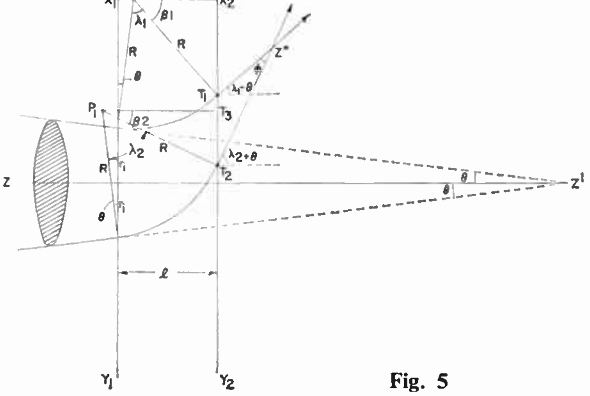


Fig. 5

of the beam are deflected through differing angles, λ_1, and λ_2 respectively and the new point of coincidence becomes Z'' due to a new convergence angle ϕ where $\phi > 2\theta$. Under the given assumptions the track of either top or bottom edge electron is *exactly* circular and of the same radius R.

P_1 and P_2 are the centres of the circles of deflection for electrons at the bottom and top of the beam respectively.

From Fig. 5 it is seen that the extent of the deflection defocusing will be a function of the change in the convergence angle, i.e. a function of the quantity $(\phi - 2\theta)$. But

$$\phi = \beta_1 - \beta_2 \qquad \ldots\ldots(33)$$

From (33) we may evaluate ϕ.

Now

$$\phi \sim \sin\phi = \sin(\beta_1 - \beta_2) = \sin\beta_1\cos\beta_2 - \cos\beta_1\sin\beta_2 \qquad \ldots\ldots(34)$$

Furthermore from the triangles $P_2X_2T_1$, $P_1T_3T_2$ we find:

$$\left.\begin{aligned} \cos\beta_2 &= \frac{l + R\sin\theta}{R} \\ \cos\beta_1 &= \frac{l - R\sin\theta}{R} \\ \sin\beta_2 &= \frac{\sqrt{R^2 - (l + R\sin\theta)^2}}{R} \\ \sin\beta_1 &= \frac{\sqrt{R^2 - (l - R\sin\theta)^2}}{R} \end{aligned}\right\} \qquad \ldots\ldots(35)$$

Using the approximation $(1-x)^{1/2} \sim 1 - x/2$ provided $x \ll 1$ and substituting from eqns. (35) into (34) gives

$$\sin\phi = \frac{\sqrt{R^2 - l^2}}{R^2}\left[\left\{1 + \frac{lR\sin\theta}{R^2 - l^2}\right\}(l + R\sin\theta) - \right.$$
$$\left. - \left\{1 - \frac{lR\sin\theta}{R^2 - l^2}\right\}(l - R\sin\theta)\right]$$

Whence after some reduction

$$\sin\phi = \frac{R}{\sqrt{R^2 - l^2}} \cdot 2\sin\theta \qquad \ldots\ldots(36)$$

From Fig. 5 it is seen that $\dfrac{R}{\sqrt{R^2 - l^2}} = \sec\lambda$ where λ is the deflection angle of the *centre* ray and equals $(\lambda_1 + \lambda_2)/2$.

Thus finally $\sin\phi \sim \phi \sim 2\theta\sec\lambda$ since θ is small.

The *increase* in spot diameter on deflection is approximately $(\phi - 2\theta)z$ where z is coil to screen distance.

Thus the deflection defocusing is

$$\begin{aligned} \delta(r_s) &= (\sec\lambda - 1).z\theta \\ &\simeq \lambda^2 z\theta \\ &= f(\lambda)z\theta \end{aligned} \qquad \ldots\ldots(37)$$

Equation (37) proves that the deflection defocusing is proportional to the beam width, all else being constant.

7.2. *Case (2)—Electrostatic Deflection*

Figure 6 shows the essential geometry. Again the field is supposed uniform, wholly normal to the axis, and sharply terminated at the sections X_1Y_1 and X_2Y_2, which are the ends of the parallel deflection plate system.

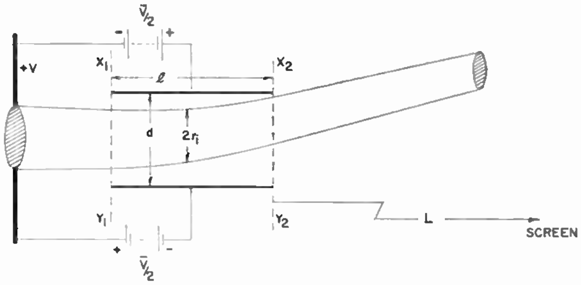


Fig. 6

In the case of electrostatic deflection, the primary cause of the defocusing resides in the differing energy changes among the component electron rays comprising the beam. Although there will be *some* purely *geometrical* defocusing due to the effects of beam convergence (this was the *only* cause of the defocusing for the magnetic field case), yet this effect is far outweighed by the energy change considerations.† Accordingly it is permissible to neglect the slight beam convergence when treating the electrostatic deflection problem.

Accordingly with these assumptions, an electron moving along the *top* edge of the beam, on passing the plane X_1Y_1 *suddenly* jumps in axial velocity from v_z to (say) v_1 where

$$\tfrac{1}{2}mv_1^2 = e\left[V + \bar{V}.\frac{r_i}{d}\right] \qquad \ldots\ldots(38)$$

Similarly an electron moving along the bottom edge of the beam, on passing the plane X_1Y_1 suddenly falls in axial velocity from v_z to (say) v_2 where

$$\tfrac{1}{2}mv_2^2 = e\left[V - \bar{V}.\frac{r_i}{d}\right] \qquad \ldots\ldots(39)$$

Now consider an electron following the *top* edge of

† The truth of this assertion is indicated by observation showing defocusing to be severe at angles so small (~ 10°) that it could not be explained on geometrical grounds only.

the beam. Its transit time in the plate system during its flight from X_1Y_1 to X_2Y_2 is l/v_1. Its lateral acceleration is uniform and equal to $\frac{e}{m}\nabla\bar{V}$. Therefore its lateral velocity on reaching the plane X_2Y_2 is $\frac{e}{m}.\nabla\bar{V}.\frac{l}{v_1}$ and its additional displacement from the axis is $\frac{e}{2m}.\nabla\bar{V}.\left(\frac{l}{v_1}\right)^2$.

On passing the plane X_2Y_2 the axial speed of the electron *suddenly* reverts to v_z. There is no change in the radial component of velocity under the assumed postulates. From this it follows that the exit angle of the top edge electron is arc tan $\frac{e}{m}\nabla\bar{V}\left(\frac{l}{v_1}\right)\frac{1}{v_z}$. Hence it follows that at the screen, distance L from the exit edge X_2Y_2 of the plate system, the *total* deflection for a top edge electron is exactly

$$x_t = \frac{e}{m}.\nabla\bar{V}l\left[\frac{l}{2v_1^2} + \frac{L}{v_z v_1}\right] \qquad \ldots\ldots(40)$$

Exactly similar reasoning shows that for a *bottom* edge electron the corresponding total deflection is

$$x_b = \frac{e}{m}\nabla\bar{V}l\left[\frac{l}{2v_2^2} + \frac{L}{v_z v_2}\right] \qquad \ldots\ldots(41)$$

Hence the deflection defocusing $\delta(r_s)$ is given by half the difference of (40) and (41) so that

$$2\delta(r_s) = \frac{e}{m}\nabla\bar{V}l\left[\frac{l}{2}\left(\frac{1}{v_2^2} - \frac{1}{v_1^2}\right) + \frac{L}{v_z}\left(\frac{1}{v_2} - \frac{1}{v_1}\right)\right] \ldots(42)$$

Using eqns. (38) and (39) to eliminate v_1 and v_2 from eqn. (42), writing $v_z = \sqrt{\frac{2eV}{m}}$, neglecting $\left(\bar{V}.\frac{r_i}{d}\right)^2$ in relation to V^2 and writing $\left[1 \pm \frac{\bar{V}}{V}.\frac{r_i}{d}\right]^{\frac{1}{2}} \simeq 1 \pm \frac{\bar{V}}{2V}.\frac{r_i}{d}$ gives after some reduction:

$$2\delta(r_s) \simeq \frac{1}{2}\frac{\bar{V}^2}{V^2}\cdot\frac{1}{d^2}\cdot l(L+l)r_i \qquad \ldots\ldots(43)$$

Similarly it is readily shown that the deflection of the principal ray of the beam cone at the screen is given by

$$x = \frac{\bar{V}.l}{2dV}\left(L + \frac{l}{2}\right) \qquad \ldots\ldots(44)$$

Since $L \gg l$ we may write (44) in the near approximation

$$x \simeq \frac{\bar{V}l}{2dV}(L+l) \qquad \ldots\ldots(44a)$$

and substituting from (44a) in (43) then gives

$$\delta(r_s) = \frac{x^2}{l(L+l)}\cdot r_i \qquad \ldots\ldots(45)$$

Equation (45) establishes that, all else being constant, the deflection defocusing is again proportional to the beam width, and to λ^2 for small angles of deflection.

7.3. *Relationship between Results for Magnetic and Electric Fields—The* f(λ)z . θ *Law*

For magnetic deflection the analysis immediately gave the desired result in eqn. (37). Now consider eqn. (45). Substituting $x = z\tan\lambda$ gives

$$\delta(r_s) = \frac{\tan^2\lambda}{l(L+l)}\cdot z^2 r_i$$

$$= \frac{f(\lambda)}{l(L+l)}z^3.\theta \qquad \ldots\ldots(45a)$$

At first sight this looks quite different from (37). However proper interpretation shows that they can both be written in the same way. For if in (45a) we always scale the electrostatic deflections *in the same proportion* as z itself, then l and L both are proportional to z so that again $\delta(r_s) \propto f(\lambda)z\theta$.

All this has the following meaning when applied to the analysis in the main body of the paper. In the scaling operation of Fig. 4 if the deflection is *electrostatic* then all the deflectors *must* be scaled in the same proportion as z (i.e. by k^a, or k, depending on whether Table 3 or Table 2 is involved). On the other hand if the deflection is magnetic it is not *essential* to scale up the coil in proportion. The analysis leading to eqn. (37) shows that the defocusing depends only on the angle of deflection and not on the thickness of the field. A weak thick magnetic field turning the beam through angle λ causes just as much defocusing as a strong thin magnetic field also deflecting through the same angle λ.

In contradistinction the defocusing in an electric field involves the field gradient as shown by eqn. (43). This means that the physical size of the field is involved. Thus any scaling of the deflecting region requires proportional scaling of the deflection plate system. Basically the two mechanisms of deflection and defocusing are quite dissimilar. In magnetic deflection no beam energy change occurs. Electrostatic deflection involves alteration of beam energy.

Manuscript received by the Institution on 7th June 1961 (Paper No. 674).

# of current interest . . .

The 1961 Medical Electronics Conference

Several members of the Institution were among British delegates to the 4th International Conference on Medical Electronics which was held in New York during July. Four members of the Brit.I.R.E. Medical and Biological Electronics Group Committee were in the British delegation: Messrs. R. Brennand, K. Copeland, W. J. Perkins and Dr. C. A. F. Joslin; Dr. A. Nemet and Mr. H. W. Shipton (now of Iowa State University) were also present.

Papers were read by Dr. Joslin and Messrs. Copeland and Perkins. Mr. Perkins, who was re-elected to the Executive Committee of the International Federation of Medical Electronics as Membership Secretary, took the chair at a session on Instrumentation. A short report on the Conference is given on page 304 of this *Journal*.

Conference of European Postal and Telecommunications Administrations

Since the war there has been a growing demand for closer co-operation within Europe in the field of Posts and Telecommunications. The idea gained momentum from 1955 onwards and in June 1959, the Conference of European Postal and Telecommunications Administrations (C.E.P.T.) was established. It was decided that member countries should meet in full session once a year: the first meeting was held in Paris in October 1960.

Representatives of 19 European countries met at Torquay, Devon, between September 11th and 22nd for the 1961 meeting. The aims of the Conference, which was opened by the Postmaster-General, the Rt. Hon. Reginald Bevins, M.P., are the development of co-operation between its members and the practical betterment of their administrative and technical services by the exchange of information and joint study of common problems.

In the telecommunications field the Conference concerned itself with a wide range of operational and financial questions concerning telecommunication services between member countries of C.E.P.T. and also between member countries of C.E.P.T. and the rest of the world. A study group was set up to consider from the viewpoint of the member countries of C.E.P.T. the problems which will arise as developments proceed in the new field of radio communication via artificial earth satellites. A number of C.E.P.T. countries are already building ground stations to participate in tests with American launched telecommunication satellites in 1962. (See *J. Brit.I.R.E.* May 1961, page 408 and September 1961, page 239.)

Membership is restricted to administrations of European member countries of the Universal Postal Union and International Telecommunication Union. At present the following 19 administrations are members, namely, Austria, Belgium, Denmark, Finland, France, Germany (Western), Greece, Iceland, Ireland, Italy, Luxembourg, Norway, Netherlands, Portugal, Spain, Sweden, Switzerland, Turkey, and the United Kingdom.

The Conference which is independent of any political or economic body, operates within the spirit of the Conventions of the Universal Postal Union and International Telecommunication Union. It is essentially a consultative body—its conclusions take the form of recommendations to the member countries.

The Association of Retired Engineers

Retired and semi-retired members living in the South of England in and around the County of Sussex, will no doubt be interested to learn of the "Association of Retired Engineers". The Association was formed in 1951 and is for retired engineers who are, or have been, corporate members of a Chartered Engineering Institution.

The object of the Association is to help retired engineers to associate with others of similar interests and with this in mind meetings are held in Worthing during alternate months in the Winter on engineering and allied subjects, and visits are organized to local engineering works. There is also a good social programme and at many of these functions lady guests are invited. Any member who may be interested should get in touch with the Honorary Secretary, Mr. L. L. Ruderman, B.Sc., A.C.G.I., M.I.E.E., 37, Hayling Rise, Worthing, Sussex.

There are many good reasons for hoping that there will be a national extension of this very excellent Association.

Conference on Semi-conductors

The Institute of Physics and the Physical Society, on behalf of the International Union of Pure and Applied Physics and the British National Committee for Physics, is arranging an International Conference on The Physics of Semi-conductors, which will be held at the University of Exeter from 16th–20th July 1962. The Conference is planned to follow the previous sequence of Conferences on the physics of semi-conductors, which were held in Reading in 1950, Amsterdam in 1954, Garmisch in 1956, Rochester in 1958 and Prague in 1960.

Accommodation will be provided in Halls of Residence at the University. Further information regarding the Conference may be obtained from the Administration Assistant, The Institute of Physics and The Physical Society, 47, Belgrave Square, London, S.W.1.

U.D.C. 621.396.946: 629.19

Communications at Megamile Ranges

By

R. C. HANSEN, Ph.D.†

AND

R. G. STEPHENSON, B.Sc.†

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

Summary: Megamile communications call for the fullest use of advanced modern technology. This paper surveys the design of a deep-space communication system. The basic constraints and capabilities of a low-noise ground receiving system are covered in detail, including antenna temperature contributions from galactic noise, tropospheric and ionospheric attenuation, and earth radiation through antenna side lobes. Low-noise preamplifiers, large high-gain antennas, and narrow-band phase-lock receivers are also considered. Digital biphase modulation, used in *Pioneer V*, is compared with more sophisticated bi-orthogonal coding systems. Both types of coding systems are compared with error rate as a parameter, with the ideal error-free Shannon limit.

The design of *Pioneer V* and its record of achievement are described. Included are explanations of the coherent transponder within the spacecraft, and of the command transmitter and telemetry reception equipment at the ground stations.

1. Introduction

This paper reviews the major compromises that arise in the design of a deep space communications system. Section 2 covers the factors in ground receiving station performance. Graphs are presented for the noise temperature contribution of several sources including the galaxy, radio stars, atmospheric absorption, ionospheric refraction and Faraday rotation, and earth radiation through the antenna sidelobes. Transmission line losses and high-gain antennas are also covered. Section 3 discusses system calculations and digital modulation techniques. Biphase systems are compared with more sophisticated orthogonal systems. Also covered are the principles of a narrow-band phase-lock receiver and of a coherent demodulation system for binary bits. Section 4 describes the *Pioneer V* communication system.

2. Low-Noise, High-Gain Ground Stations

It is convenient to discuss system performance in terms of a system noise temperature rather than the easily misinterpreted noise figure or noise factor. The system noise temperature is of course the temperature of the input resistive termination of a hypothetical noiseless system which provides the same output noise power as the actual system. Below about 50 Mc/s atmospheric noise is usually the limiting factor. At lower frequencies, the effective antenna temperature (defined below) due to atmospheric noise can easily be 10^3 deg K or higher. Above this frequency, atmospheric noise is negligible and other noise sources become important. The noise appearing at the antenna terminals originates from a multitude of sources. We will attempt to categorize the most important sources. These are galactic noise, star and planetary noise, atmospheric absorption noise, and earth radiation. Man-made noise is not included here. All of these may contribute noise through the antenna main beam or through the sidelobes. In addition, resistive loss in the antenna and transmission line structure introduces noise.

At frequencies above 1 Gc/s (and to a lesser degree below 1 Gc/s) the spatial distribution of noise is not isotropic and the antenna pattern is typically highly directional. In addition, the multitude of sources poses a problem. For example, the antenna may see one noise source partially reflected in the surface of a second noise source, e.g. land or sea. Reciprocity allows us to say that the noise received by the antenna from each direction (through a small solid angle) is proportional to the power absorbed in that direction when the antenna is transmitting, multiplied by the ambient temperature of the absorbing body. This is given by eqn. (1), where the effective antenna temperature T_a due to a source of temperature T_s, subtending small solid angle θ_s^2 is given. G is the antenna gain averaged over the solid angle θ_s^2.

$$T_a = \frac{\theta_s^2 T_s G}{4\pi} \qquad \text{......(1)}$$

Typically the antenna main beam and sidelobes are divided up into a number of small solid angle regions, according to the distribution of noise sources. Formula (1) is then used to find the contribution to antenna temperature from each, with gain for each solid angle used in (1). For a noise source subtending angle θ_s, with $\theta_s \geq \theta_a$ where θ_a is the antenna beamwidth, the gain in (1) is now the maximum gain or the

† Aerospace Corporation, Los Angeles; formerly at Space Technology Laboratories, Inc.

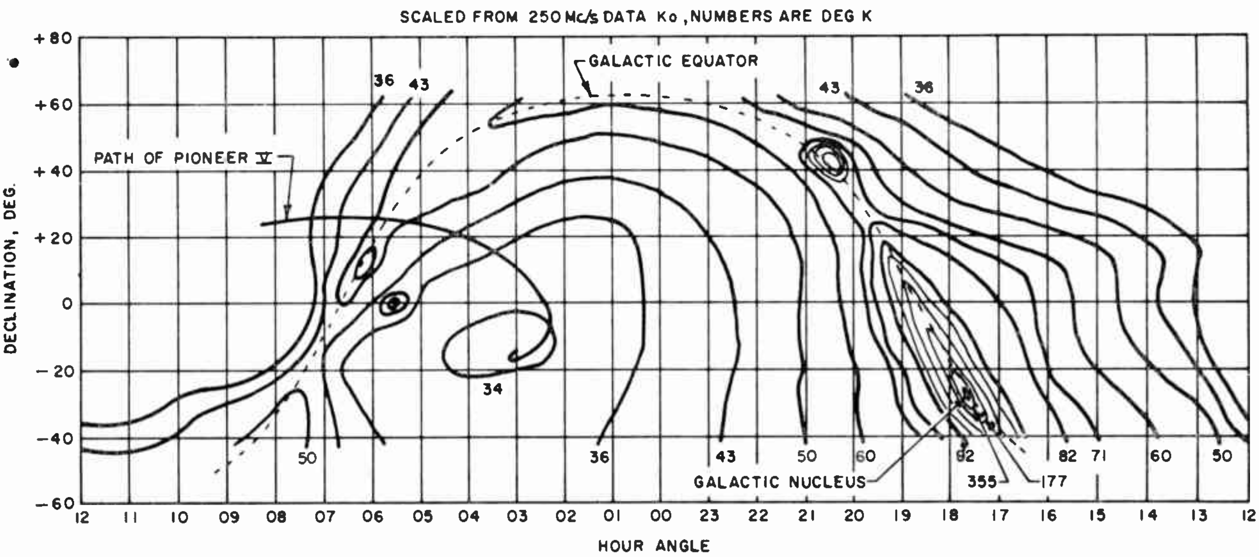


Fig. 1. Cosmic background radiation at 378 Mc/s.

directivity, given for an ideal antenna as

$$G = \frac{4\pi}{\theta_a^2} \qquad \ldots\ldots(2)$$

This equation coupled with (1) gives a main beam noise temperature contribution from a source $\theta_s \geq \theta_a$ of

$$T_a = T_s.$$

To this must be added the side lobe contribution if $\theta_s > \theta_a$. For the case where $\theta_s < \theta_a$ and the beam maximum is pointed at the noise source, (2) holds, giving with (1)

$$T_a = T_s \left(\frac{\theta_s}{\theta_a}\right)^2 \qquad \ldots\ldots(3)$$

The details of how to calculate antenna temperatures for a narrow-beam antenna looking at a multitude of sources have been given by one of the authors.[1] It will suffice here to detail the magnitude and variation of the important constituents so that the boundary of the problem may be delineated. This is of special interest in the selection of an optimum frequency which will be discussed later.

2.1. *Galactic Noise*

Information on cosmic or galactic noise has appeared from many sources. The most extensive are the charts of Menzel[2] which cover the heavens at 200 and 600 Mc/s and the paper of Ko[3] which gives charts from 64 Mc/s to 910 Mc/s. Menzel's data were measured with a 10 deg pencil beam, producing a somewhat coarse mapping; Ko summarizes data from a variety of sources. Other sources give data for a 20 deg beam.[4] Figure 1 shows a typical distribution of galactic noise temperature (in degrees Kelvin) for 378 Mc/s; this was the frequency of the *Pioneer V* telemetry link from vehicle to earth. These data have been obtained by using the 250 Mc/s contours of Ko, scaling the intensities by

$$T \propto f^{-2\cdot5}$$

as determined empirically by Hanbury Brown and Hazard.[5] The co-ordinate system is one commonly used by astronomers, with right ascension (hour angle) for the abscissa and declination as the ordinate.

In general, our sun is a small star in one arm of a spiral "pancake" galaxy. Antennas pointed in the direction of the galactic centre see many sources, hence have a high antenna temperature; another peak not quite so large occurs when the antenna looks along the galactic arm. When, however, the antenna looks out of the galactic plane, the noise temperature is very low. The galactic centre and arm can be observed in Fig. 1, along with the low values out of

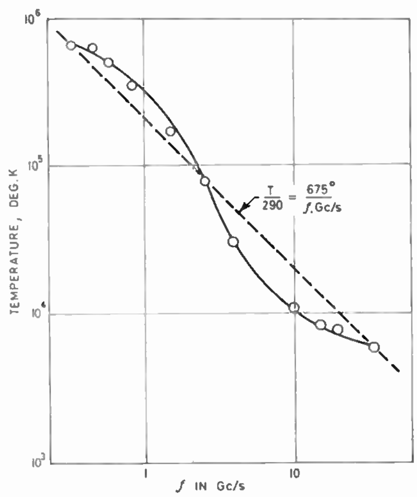


Fig. 2. Sun noise temperature.

the plane. Examination of this figure and of Ko's plot at 910 Mc/s indicates that only a few localized regions contribute appreciable antenna temperature. At higher frequencies this is even more true. For the galactic centre, Hanbury Brown and Hazard[5] quote 17° K for 1·2 Gc/s with a 2·8 deg beam, and 2·6° K for 3 Gc/s with a 3·4 deg beam. Thus above 1 Gc/s the continuous background is very low; a list of discrete sources is valuable.

2.2. *Discrete Sources*

Of the discrete body sources, the sun is certainly the "shining" example. Hogg and Mumford[6] have given the noise temperature contribution of the sun for antennas whose beamwidth is not greater than the 0·5 deg plane angle subtended by the sun. Their data are shown in Fig. 2. For antennas of broader beamwidth, the antenna temperature T_a is decreased by eqn. (3) where $\theta_s = 0{\cdot}5°$. For radio stars, the flux density S is given by

$$S = \frac{2kT_s\theta^2}{\lambda^2} \quad \text{Wm}^{-2}/\text{cycles . s}^{-1} \qquad \ldots\ldots(4)$$

where T_s and θ_s are the star temperature and subtended angle, and k is Boltzmann's constant. For almost all stars, $\theta_a \geq \theta_s$; combining (3) with (4) one obtains antenna temperature contribution as a function of antenna effective aperture A_e and flux density:

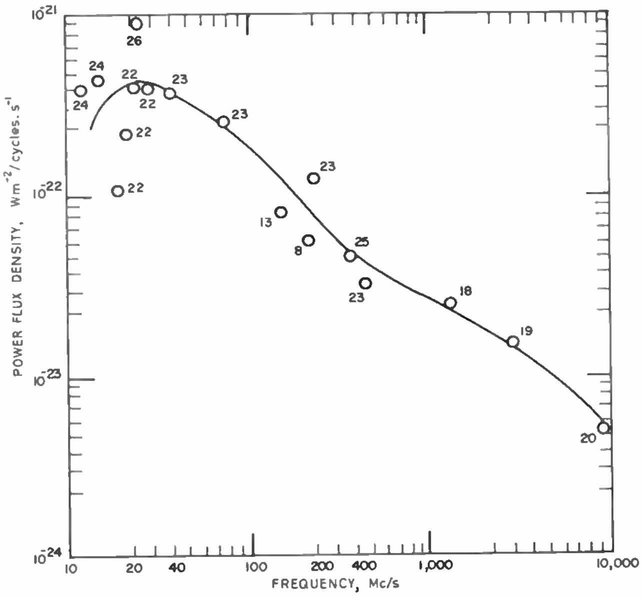


Fig. 3. Radio spectrum of Cassiopeia A (IAU23N5A). The numbers indicate observers.

$$T_a = \frac{SA_e}{2k} \qquad \ldots\ldots(5)$$

Ko[7] gives a list of 57 "reliable and intense" radio stars; he also gives flux density vs. frequency data on

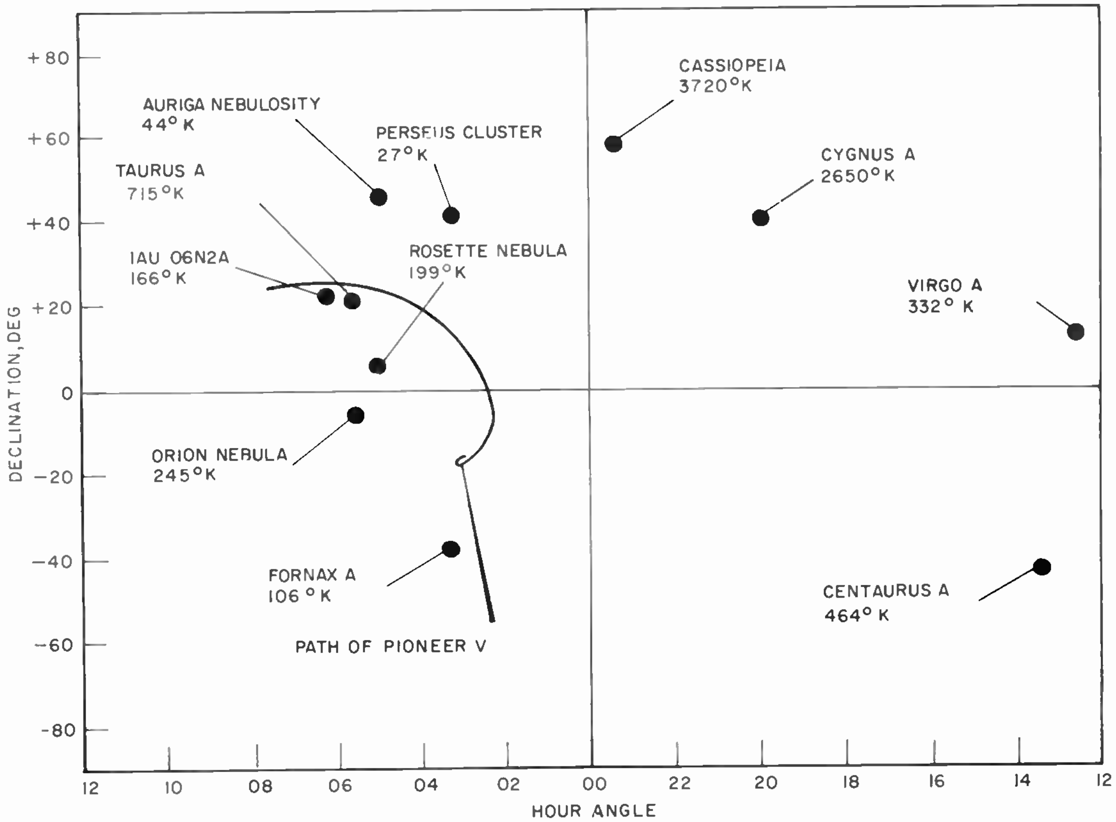


Fig. 4. Stellar radio sources at 378 Mc/s for Manchester antenna.

eleven of the more prominent sources. For example, the data on Cassiopeia A are shown in Fig. 3. From Ko's data we have prepared Fig. 4, a plot of the more prominent radio stars, again on a right ascension-declination co-ordinate system. The temperatures shown have been computed as the effective noise temperature when observed by the 250-ft parabolic antenna at Jodrell Bank; a 40% aperture gain factor at 400 Mc/s has been estimated (since the *Pioneer V* measurements, the instrument has been improved).

For very low noise systems, weaker radio stars as well as planets must be considered. Table 1 shows approximate temperatures for the most important planets.[8-11] Equation (1) is also used to calculate planetary contributions to antenna temperature.

Table 1

Planetary Radiation Temperatures

	Frequency Gc/s	Surface temperature deg K
Venus	35	410
	9·5	590
	3	580
Jupiter	9·5	145
	2·9	640
	1·4	3000
	0·97	5500
	0·44	50 000
Mars	9·5	218

2.3. *Atmospheric Absorption*

In addition to noise from extra-terrestrial sources, noise can be introduced into a receiving antenna system because of absorption of radio signals by oxygen,

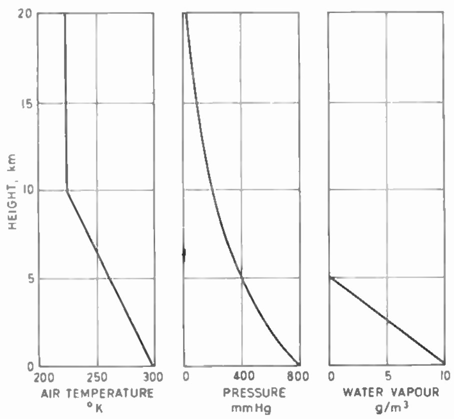


Fig. 5. International standard atmosphere.

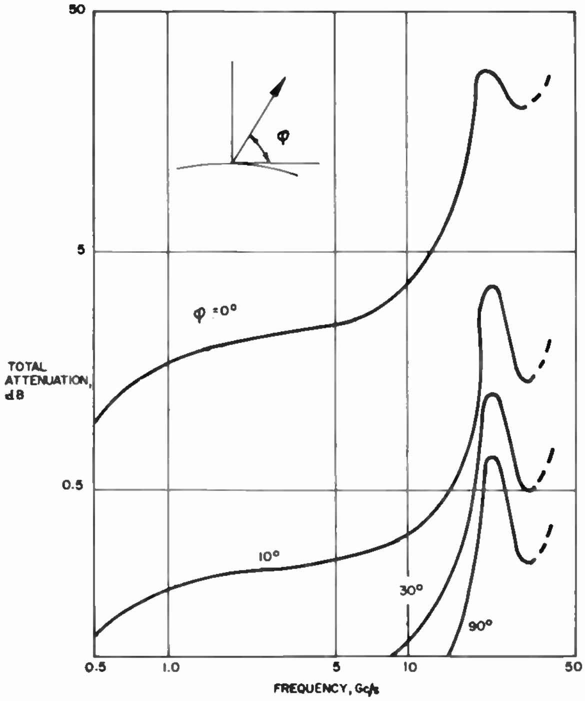


Fig. 6. Absorption through troposphere.

water vapour, and rain. This noise, due to the attenuating qualities of the atmosphere, should not be confused with the "atmospheric noise"; this latter term is commonly used to denote low-frequency noise mainly due to lightning. The effects of rain have been indicated by Hogg and Semplak,[12] who observed increases of sky temperature from 3° K to 70° K (120° K during storms) at 6 Gc/s. Atmospheric absorption is computed for the international standard atmosphere, unless more specific data are available; the summer standard atmosphere is shown in Fig. 5. The actual noise contribution depends upon the absorption coefficient of the atmosphere, upon the actual temperature of the atmosphere, and upon the antenna elevation angle. Data on atmospheric absorption in the lower portion of the microwave region are given by Hogg and Mumford,[6, 13] (see Fig. 6), while recent results on millimetric absorption have been given by Tolbert and Straiton[14] and by Rosenblum.[15] Translated for a low side lobe hoghorn antenna, the attenuation results in the temperature versus frequency curve of Fig. 7 where elevation angle is a parameter. To this effect should be added the ionospheric attenuation,[16] which is shown in Fig. 8. It will be noted that attenuation in the ionosphere is negligible above about 300 Mc/s, while the tropospheric curves, given earlier, are smaller below that frequency. The reader is cautioned that these data are "typical" and cannot be relied upon for exact values on a specific experiment.

All of these sources contribute principally through the antenna main beam. The side lobes are an important noise source, however, in that side lobes look at such hot sources as the sun and the earth (the earth may not be as hot, but it looks much bigger). Noise temperature is calculated[1] by multiplying the fraction of power contained in each side lobe by the average temperature that the side lobe sees. These temperatures are then added to the main beam contributions. When a side lobe looks at the sun, formula (2) is used.

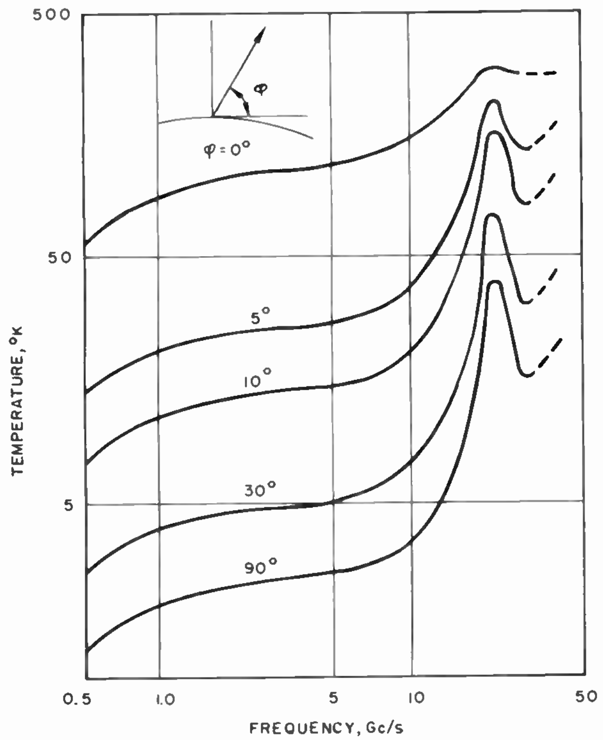


Fig. 7. Antenna temperature (after Hogg and Mumford[6]).

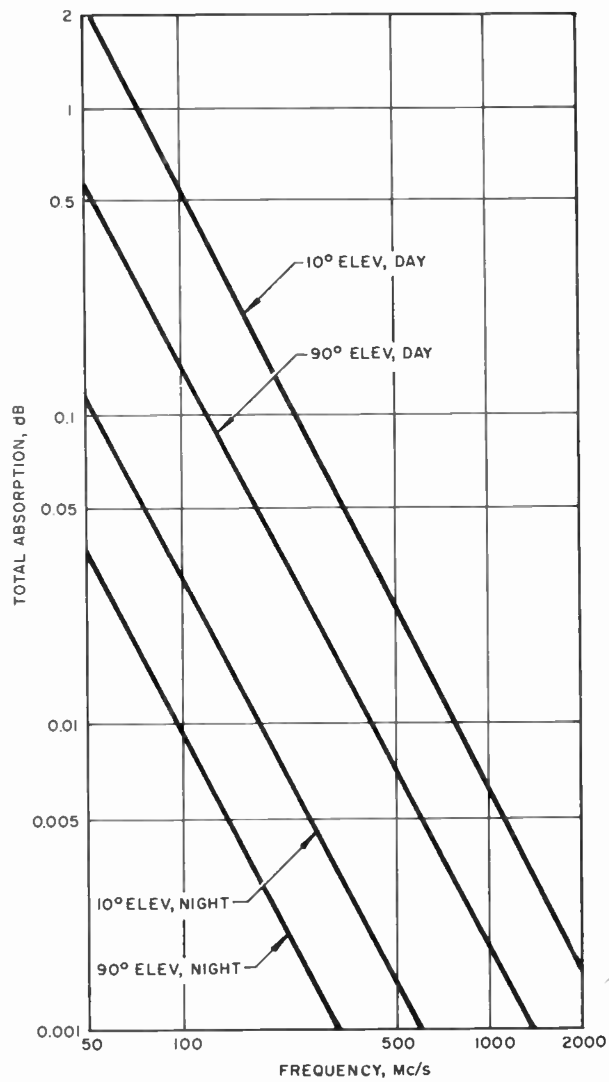


Fig. 8. Absorption through ionosphere for source at 1000 km height.

2.4. *Transmission Line Loss*

Between the incident electromagnetic wave and the receiver terminal there appear a number of metal structures including the antenna, transmission lines, duplexers, etc. Resistive losses in all of these introduce noise as well as signal attenuation. Very small losses in structures at room temperature can best be handled by adding 7° K for every 0·1 dB; losses higher than a few tenths of a decibel should be treated as a four terminal network. When the attenuation ratio (input to output) is called L, the noise temperature at the output of the transmission line is given by the following formula, where T_a is the antenna temperature and T_L is the ambient temperature.

$$T_{\text{out}} = \frac{L-1}{L} T_L + \frac{1}{L} T_a \qquad \ldots\ldots(6)$$

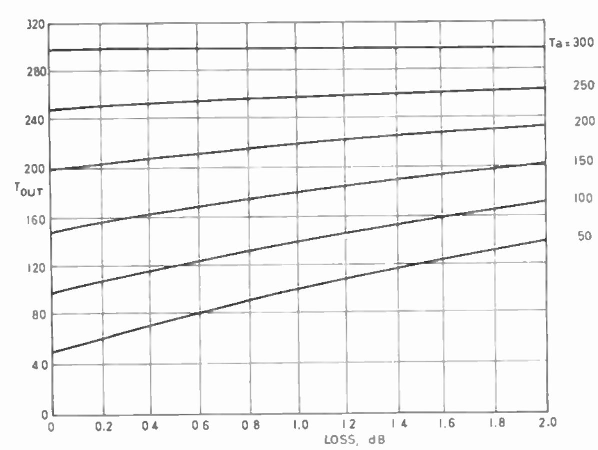


Fig. 9. Output temperature vs. line loss.

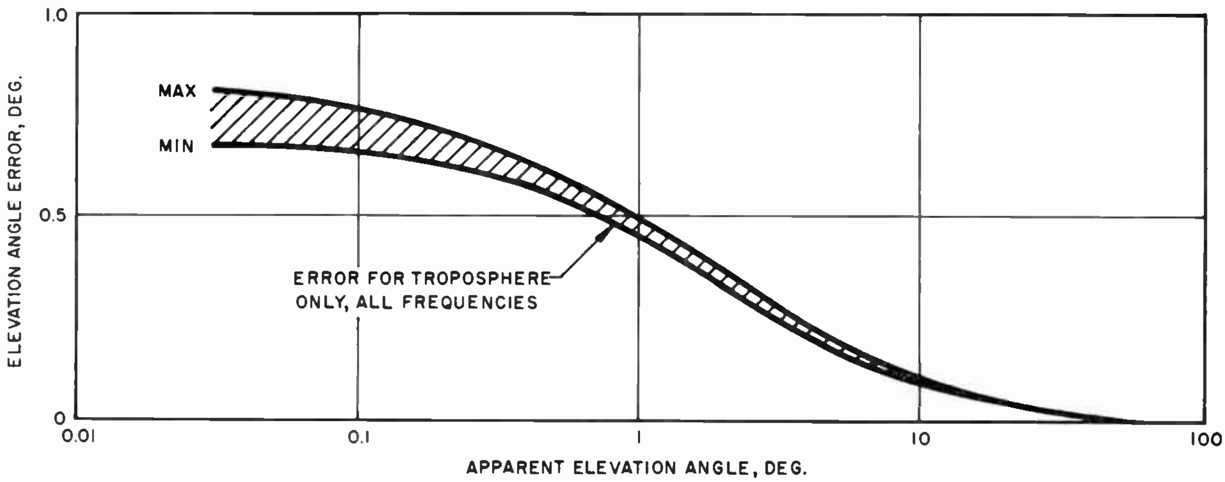


Fig. 10. Atmospheric refraction errors.

Figure 9 depicts the variation for an ambient of 300° K. For an antenna temperature of 50° K, for example, 1 dB of loss raises the input temperature to 100° K, resulting in an appreciable degradation in system sensitivity. In addition, the signal itself is attenuated by 1 dB.

2.5. *Refraction*

The effects of tropospheric refraction, which are independent of frequency, are plotted in Fig. 10. The shaded area shows the variation of the ray-bending by the troposphere under widely varying conditions. In this figure one sees that errors in elevation angle are appreciable only if the elevation angle is very low. The data in this figure summarizes a number of specific calculations on tropospheric refraction. However, the bulk of the data were obtained from Weisbrod and Anderson.[17]

Ionospheric refraction is very small for u.h.f. signal rays from ground to outer space, and can be neglected in comparison to the tropospheric effect for frequencies above 100 Mc/s. (However, the refraction cannot be neglected for sources within the ionosphere.)

2.6. *Faraday Rotation*

Faraday rotation of the plane of polarization of radio waves must also be recognized during the design of a space communication system. This rotation phenomenon is caused by an interaction of the earth's magnetic field and free electrons in the ionosphere. Figure 11 shows the number of radians of rotation of incoming radio signals as a function of frequency and of elevation angle, as calculated by Millman.[16] From that figure, we see that rotation experienced is markedly different according to the direction of the magnetic field. Propagation along a magnetic field line of force can cause a rotation of about 15 radians for 378-Mc/s signals at low-elevation angles during the daytime; propagation in a direction perpendicular to magnetic lines of force will cause a rotation of only about 1·0 radian. Night-time effects are about one third of those shown.

The Faraday rotation is important to consider if linearly-polarized antennas are to be used for both transmission and reception of signals from space probes. As the transmission ray moves across the sky, and as the condition of the ionosphere changes, the magnitude of the Faraday effect will change; thus, the receiving signal plane of polarization will change with time. If a linearly-polarized receiving antenna is used for such signals, the receiving antenna plane of polarization must be rotated continually to maintain a strong receiving signal; if this is not done, the signal will fade severely as its plane of polarization becomes perpendicular to that of the receiving antenna. Because continual rotation of a dipole feed on a parabolic

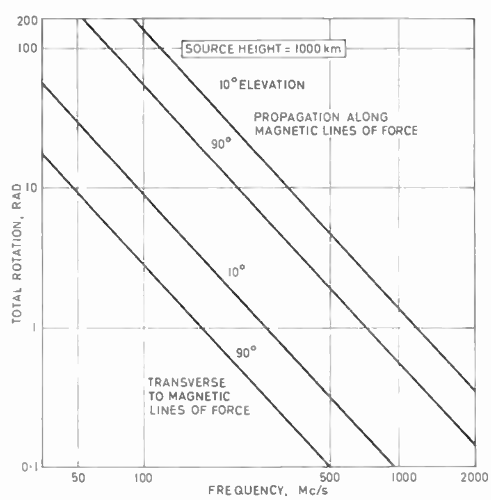


Fig. 11. Daytime Faraday rotation.

antenna is an unattractive requirement, it has become common to use circularly-polarized receiving antenna feeds. Such a feed will have 3 dB less gain than a linear feed for linearly polarized incoming signals; however, the signal amplitude from such a feed will be steady and signal dropouts can be avoided. This effect decreases rapidly with increasing frequency; above 2 Gc/s antennas of the same polarization can be used (if vehicle orientation permits).

2.7. *Antennas*

In considering various components of the system, the interesting question arises of how high a gain is practical in parabolic antennas. Data on the performance of a number of large parabolic antennas were gathered with the interesting result that none of the large group of antenna systems which were investigated has achieved gains above 65 dB. This apparent present limitation on gain appears to be caused by mechanical imperfections in the surface of the parabolic reflector. Some large antennas may have gain restrictions substantially below this limit of 65 dB where mesh rather than solid-surface parabolas are used, or where the structural rigidity is insufficient to preclude elastic deformation as the dish is tilted. A gain limitation due to tropospheric irregularities or "blobs" may be introduced at higher gain levels.[18]

Typical aperture gain factors are 50% for parabolic dish tracking antennas. The directivity G is given by

$$G = D^2 f^2 (7{\cdot}2 \text{ dB}) \qquad \ldots\ldots(7)$$

with D in ft, f in Gc/s, where the constant in parenthesis is to be added to the other quantities expressed in dB. (This shorthand allows visualization of the compact equation, while allowing the constant to be expressed in dB.) Many other types of antennas, e.g. helices and helical arrays, have been used as well as parabolic dishes. The most vital parameter is antenna directivity; as the next section will show, the directivities must be as large as possible. For the critical space-to-earth link, there is a limit on system noise temperature. Greater ranges/data rates require higher vehicle power, higher vehicle antenna effective area, or higher ground antenna effective area. Thus larger and larger ground antennas or arrays of large antennas are to be expected.

Another very practical consideration in the use of high-gain antennas is that of signal acquisition with very fine pencil beam. A parabolic antenna whose gain is 60 dB will have a total beamwidth of 0·14 deg and thus require steering information accurate to 0·07 deg, or roughly 1 mil. Prediction of signal azimuth and elevation to such accuracy is very difficult for most space probes because of uncertainties in trajectory data, and also because the atmospheric refraction may be large compared to such a tolerance. Thus use of very narrow beam antennas poses some difficult operational problems.

3. System Calculations

3.1. *Range Equation*

The one-way range equation is given by

$$P_T = \frac{16\pi^2 R^2 P_R}{G_1 G_2 \lambda^2} = \frac{16\pi^2 R^2 kTB\, S/N}{G_1 G_2 \lambda^2} \qquad \ldots\ldots(8)$$

This formula can be rewritten in a more usable form as

$$P_t = \frac{R^2 f^2 TB\, S/N}{G_1 G_2 (130{\cdot}8 \text{ dB})} \qquad \ldots\ldots(9)$$

Units used are the following:

- P_t transmitted power in watts
- G_1 ground antenna directivity
- G_2 vehicle antenna directivity
- B bandwidth in cycles/second
- T overall system noise temperature,† degrees K
- f frequency, gigacycles/second
- R range, nautical miles
- S/N signal/noise ratio

Path loss is a useful concept, and it consists of part of eqn. (8) above. It is

$$\text{path loss} = \frac{16\pi^2 R^2}{\lambda^2} = R^2 f^2 (97{\cdot}8 \text{ dB}) \qquad \ldots\ldots(10)$$

using the units above.

Figure 12 has been drawn to show the total path loss, including two antennas, over a distance of one million nautical miles for three types of antenna systems. It can be seen that the loss between two dipoles increases with the square of frequency. The dish antennas have an assumed aperture gain factor of 50% for frequencies up to 3·1 Gc/s where the hypothetical 250 ft antenna has 65 dB directivity. Above this frequency the 250 ft antenna is assumed to have an arbitrary maximum directivity of 65 dB; this accounts for the "corners" in the ideal curves of Fig. 12. These data emphasize the need for high antenna directivities (large antennas).

† In many formulas for both one-way and two-way range equations, one often sees the noise power kTB replaced by the approximate expression kT_0BF_N where T_0 is 290° K and F_N is the system noise figure defined as

$$F_N = 1 + \frac{T}{T_0}.$$

The approximation is good only for high-noise systems and becomes very misleading for low-noise systems. The correct result for noise power, of course, is $N = kT_0B(F_N - 1) = kTB$. However, it is much simpler to use noise temperature as the antenna and loss contributions are readily available in that form, as are the parameters for low-noise preamplifiers.

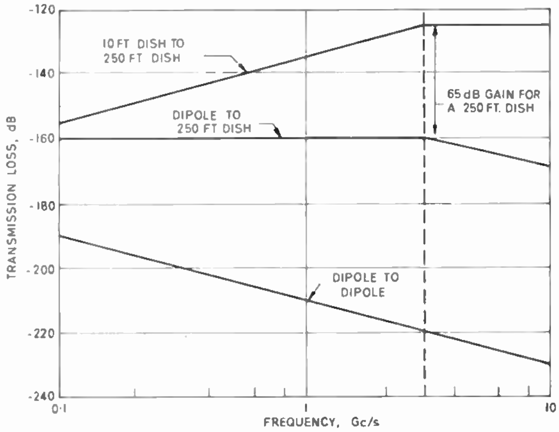


Fig. 12. Path loss and antenna gain for one nautical megamile.

The antenna gains and noise temperatures have already been discussed. The remaining two items of interest are the choice of frequency and the signal/noise behaviour. Although the *Pioneer V* frequencies were chosen on the basis of compatibility with existing systems, it is not inappropriate to discuss the behaviour of system performance with frequency. There are generally three prime considerations. These are the sharp rise in galactic noise below 1 Gc/s (and unacceptable incidence of atmospheric noise below 100 Mc/s), severe atmospheric absorption above 15 Gc/s, and the declining efficiency of both tube and solid-state transmitters as frequency goes towards 10 Gc/s. Beamwidth of large aperture antennas may be a problem above 2 Gc/s. Thus one may generally state that optimum frequencies for deep space communications lie in the range 1 to 10 Gc/s. Within this range one may derive an optimum region by assuming cost functions for vehicle antenna directivity, vehicle transmitter power, etc. A number of papers have appeared in the literature essentially doing this. Most of these take into account the variation of antenna noise temperature with frequency, and make assumptions about vehicle antenna, vehicle transmitter, ground antenna, etc.[19-22] Most of these studies arrive at an optimum somewhere in the middle of the range mentioned above, but it should be emphasized that the set of assumptions will be different for each case, hence the optimum frequency will also vary. Davies and Weaver[23] have derived optimum performance for a fixed vehicle weight as a function of range and bandwidth. Others have assumed an upper limit to ground antenna gain.[24-26] Raisbeck[27] has considered the case where the system attenuation (noise temperature) is a function of frequency. Since the actual choice of frequency is usually dictated by a host of quasi-technical factors such as frequency allocations, many of the above references are of academic interest only.

Another item that is sometimes of importance is the splitting of spectral lines due to satellite spin. An excellent paper by Bolljahn[28] contains the necessary formulas for calculating these effects. This effect, however, does not affect phase coherent two-way systems.

Instead of using the vehicle antenna as a two-port transducer, an attractive scheme uses an active automatic angle return array. In this scheme the Van Atta array, which is an array of cross-connected elements with the interesting property that energy is radiated in the direction of the incident energy, is made active by use of solid state printed circuit amplifiers. The advantages of the active Van Atta system are the realization of two-way antenna gain over $\pm$ 45 deg without vehicle stabilization, and intrinsic high reliability. The interested reader is referred to the literature for further details.[29]

3.2. *Modulation Techniques*

Since the limited available vehicle transmitter power must be used in the most efficient manner, a digital modulation system such as p.c.m. is indicated.[30] One system which offers a good compromise between system complexity and high coding efficiency is biphase p.c.m. wherein the binary coded data (in a non-return-to-zero form) are used to alternate the phase of a carrier. Cahn[31] and others[32] have derived error rates for biphase p.c.m. as a function of signal-to-

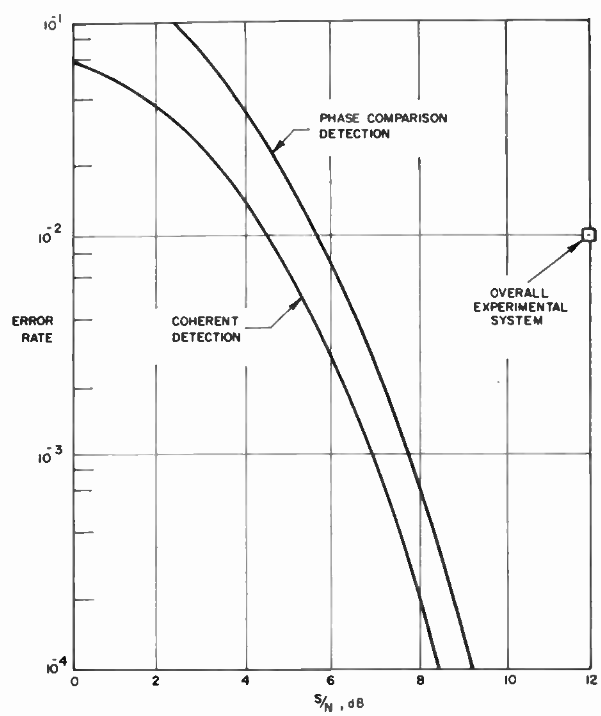


Fig. 13. Ideal biphase error rates.

noise ratio. Figure 13, from Cahn's paper, shows that for ideal coherent detection an error rate of 10^{-2} requires a signal/noise ratio of 4·3 dB, error rate of 10^{-3} requires 6·8 dB, and error rate of 10^{-4} requires 8·4 dB. The ideal coherent detection is approached by a narrow-band phase-lock receiver, suitable for the low bit rate capabilities at megamile ranges. Details of such a system will be described in the next section along with comparison of actual performance with the ideal.†

Discussions of the relative merits of various comparisons have been made by a number of authors including Sanders,[33] Balakrishnan,[34] and Helstrom.[35] For narrow-band systems, a useful comparative parameter is signal energy per bit divided by noise spectral density, or $ST/(N/B)$. Two useful ways of utilizing this parameter are to plot it against error probability, and against the bandwidth/information rate ratio, B/H. The latter is usually used for the Shannon ideal capability. It is instructive to compare digital biphase coding and orthogonal coding[36] with the ideal performance of Shannon. Figure 14 contains these data plotted in terms of $ST/(N/B)$ vs. B/H. The Shannon curve which can be obtained only with ideal coding is given by the well-known equation[33]

$$B/H = \frac{ST}{N/B}\left[2^{\frac{N/B}{ST}} - 1\right] \qquad \text{......(11)}$$

For low error rates, biphase coding lies on a straight line (see Fig. 14) and, as might be expected, lower error rates require more signal energy. Bi-orthogonal systems are also shown in Fig. 14; for two cases of 5 and 10 bits per character, these curves were obtained from data calculated by Viterbi.[37] The higher efficiency of bi-orthogonal systems may be observed from this figure.

A different comparison, also obtained from Viterbi, is shown in Fig. 15 for $n = 5$, and Fig. 16 for $n = 10$. These graphs give word error probability against $ST/(N/B)$. It is not possible to make a direct comparison between orthogonal coding and biphase coding as the individual bit errors are "clustered" in the former system, i.e. the errors in a word are not randomly distributed. From Fig. 15 it may be seen that for $n = 5$ the theoretical advantage of a bi-orthogonal system is about 3 dB; from the next figure for $n = 10$, it is about 4 or 5 dB. Thus the orthogonal systems, which require vastly more complicated equipment—both airborne and ground—offer only 3 to 5 dB advantage over coherent biphase coding.

The design of the ground-to-vehicle command link is somewhat less critical in that sufficiently higher ground transmitted power can be used to more than offset the high vehicle receiver noise temperature. If the two links can be made phase coherent (and this is

† It is interesting to note that for binary systems which detect bit-by-bit, biphase with coherent detection achieves the optimum capability. This is because one binary datum has a correlation exactly equal to the negative of the correlation of the other binary datum. To achieve higher efficiency, sequences of bits must be detected together.

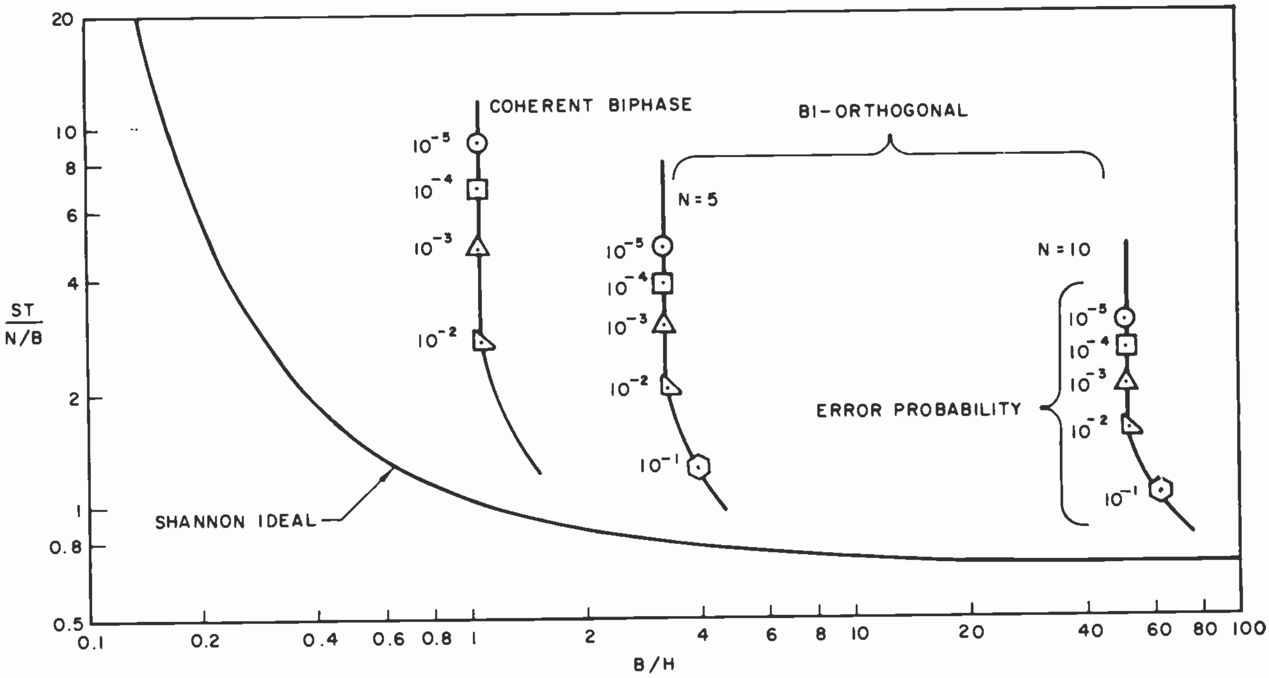


Fig. 14. Comparison of digital coding with Shannon ideal.

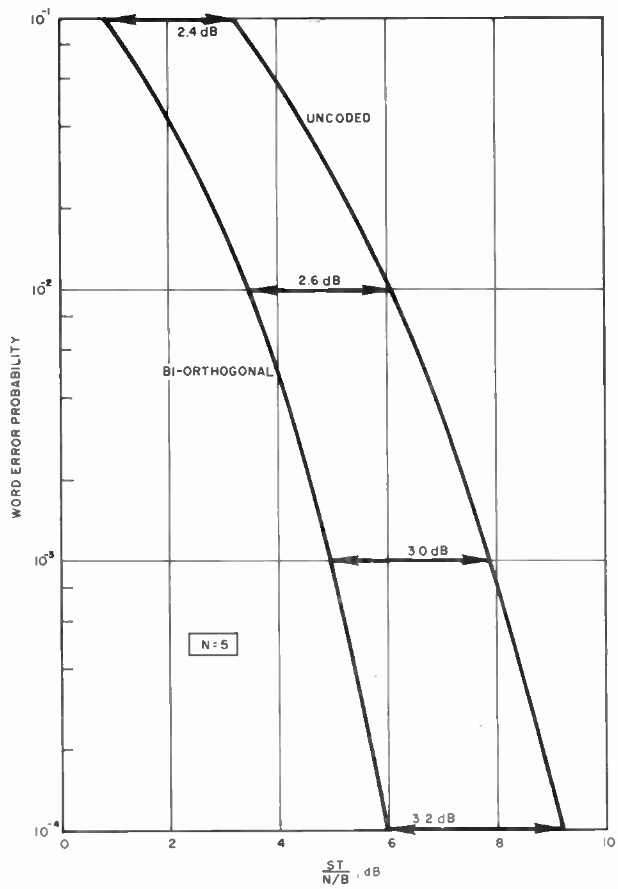


Fig. 15. Comparison of coded and uncoded systems. $N = 5$

readily done with digital phase modulation systems), the system can be used for the extraction of two-way Doppler data. For deep space probes these data allow exceedingly accurate determination of the vehicle ephemeris. If direct biphase modulation is used, the signal contains no discrete carrier and the phase-lock receiver must create a virtual carrier, usually by frequency doubling both the incoming signal and the phase-lock oscillator. However, the problems that arise when the bit rate is within the frequency pass-band of the phase-lock servo have not been completely solved. For this reason the current practice is to modulate the digital data in a biphase fashion on a sub-carrier, with the sub-carrier in turn phase-modulating the carrier. This leaves a discrete carrier which allows the carrier phase-lock loop to be simpler. After demodulation of the sub-carrier, a second phase-lock loop is used for demodulation of the digital data. The overall system is more reliable at present, but the price paid is the division of energy between carrier and sub-carrier. This has been investigated for several systems by Costas.[38] For a sub-carrier phase modulated on a carrier, the spectrum consists of a series of frequencies separated from the carrier frequency by multiples of the sub-carrier frequency. The relative power in carrier and in each of the sidebands is given by the square of appropriate Bessel functions and is shown in Fig. 17; see Sturley.[39] J_0^2 represents power in the carrier; $2J_1^2$ represents power in the first sideband, etc. In order to utilize the full sub-carrier power, a comb filter is needed to include all the sidebands. For systems utilizing only the first sideband, $2J_1^2$ gives the relative power level. The choice of modulation index, which governs the distribution of power among the Bessel components, should be such that the carrier phase-lock loop break point is the same as the sub-carrier phase-lock loop break point. This is a function of the desired Doppler tracking speed, data rate, etc. Techniques for implementing a two-way phase coherent system will be described next.

4. *Pioneer V* Communication System

The *Pioneer V* interplanetary space probe was launched on 11th March 1960 from Cape Canaveral, Florida, by a *Thor-Able* three-stage missile. The 95-pound payload was designed to telemeter back to the earth data on cosmic rays of high and low energy, of magnetic fields (evidently caused by high magneto-hydrodynamic currents flowing through space) and

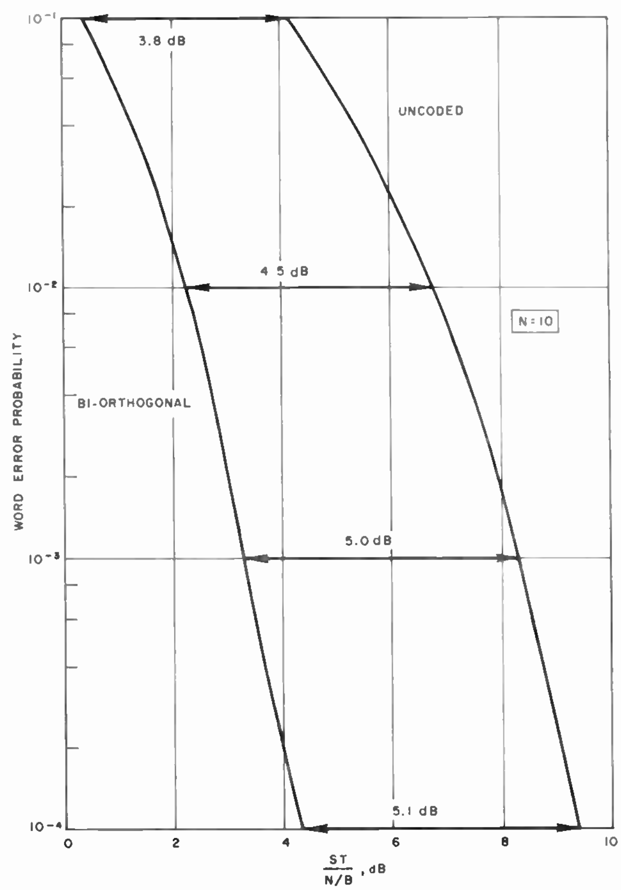


Fig. 16. Comparison of coded and uncoded systems. $N = 10$

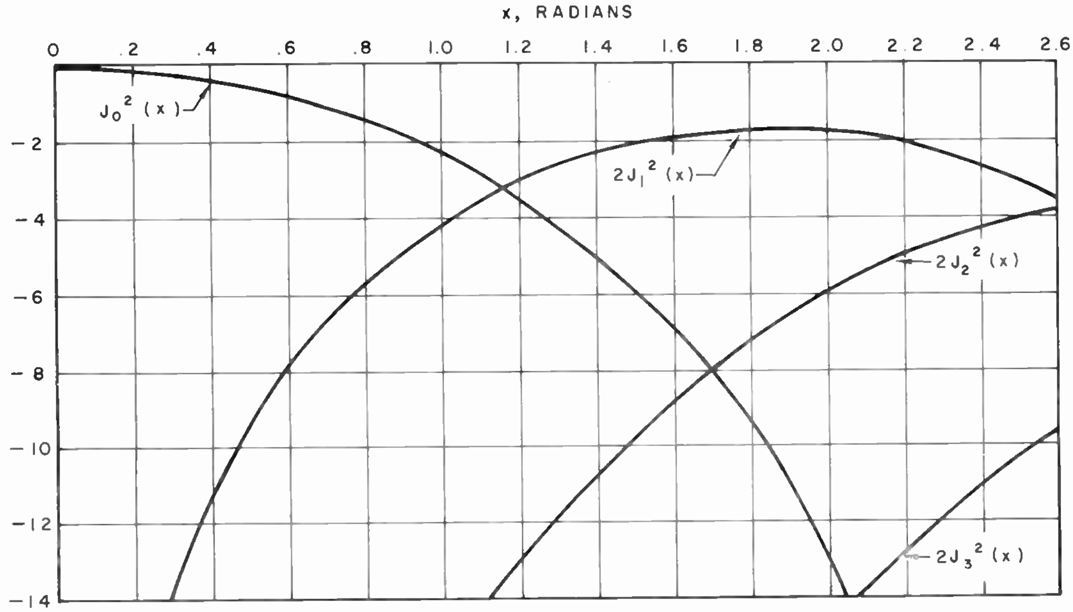


Fig. 17. Phase modulation power levels.

on micrometeorites. Details of the scientific objectives of the *Able-5* (*Pioneer V*) programme have been given by Altshuler *et al.*[40] The probe was equipped with solar-cell paddles which supplied about 25 watts of power continuously. The communication system of *Pioneer V* was required to receive commands from the earth, to transmit scientific and vehicle status data back to the earth, and to provide a coherent frequency shift between the received command carrier and the transmitted telemetry carrier. This coherent frequency shift provided very accurate Doppler information by comparison of the ground transmitted and ground received frequencies.

Since the transponder within the payload is basic to an understanding of the system, an illustrative block diagram of the transponder is presented in Fig. 18. The actual transponder is more complicated, to reduce stray coupling from the voltage-controlled-oscillator to the i.f. strip. In this figure, the received signal from a quarter-wave stub antenna passes through a diplexer which permits the use of a single antenna for transmission and reception. The received signal is then mixed with a reference frequency whose frequency is $\frac{16}{17}$ of that of the received signal. The difference frequency, $\frac{1}{17}$ F_{rec}, is amplified and filtered. The output of the i.f. amplifier at $\frac{1}{17}$ R_{rec} is compared in the phase detector with the output of a voltage-controlled-oscillator at the same frequency. The error signal from the phase detector controls the frequency of this voltage-controlled-oscillator, so that once the oscillator frequency has been locked to the output signal of the i.f. amplifier, this lock will be maintained indefinitely. The effective bandwidth of this phase-locked loop is adjusted to be about 200 c/s. The receiver will reliably lock on to incoming signals at about -140 dBm. The voltage-controlled-oscillator frequency is multiplied by 16 to obtain the $\frac{16}{17}$ F_{rec} frequency which was used in the first mixer of the receiver. With a little thought, it is apparent that the voltage-controlled-oscillator frequency will always be precisely $\frac{1}{17}$ F_{rec} unless the signal level becomes so weak that the phase-lock-loop "slips". The $\frac{16}{17}$ F_{rec} carrier, which is now phase-locked to the received carrier, is phase-modulated by the audio sub-carrier. This 1024 c/s sub-carrier contains the telemetry data through biphase modulation. The carrier modulated with sub-carrier is then amplified in the transmitter and radiated from the vehicle antenna.

Details on the data system have been given by Greenstadt[41]; it is similar to that of *Explorer VI.*[25] In this system, called Telebit, digits were transmitted at a steady rate of 64-, 8-, or 1-bit/second. At the beginning of each bit interval, the phase of the 1024-c/s sub-carrier was inverted if a "1" was to be transmitted. If a "0" was to be transmitted, the phase of the audio sub-carrier remains unchanged. For *Pioneer V*, the total message length of eight "words" was used. Each word transmitted 10 binary bits of information, but had two additional bits to provide word synchronization pulses for the ground equipment. In addition, one of the eight words always consisted of 12 "zeros" to provide a frame synchronizing signal for the ground equipment. Telebit is a digital telemetry storage and programming unit to prepare data in the format described above. This unit accumulates digital data such as cosmic ray or micrometeorite counts, converts

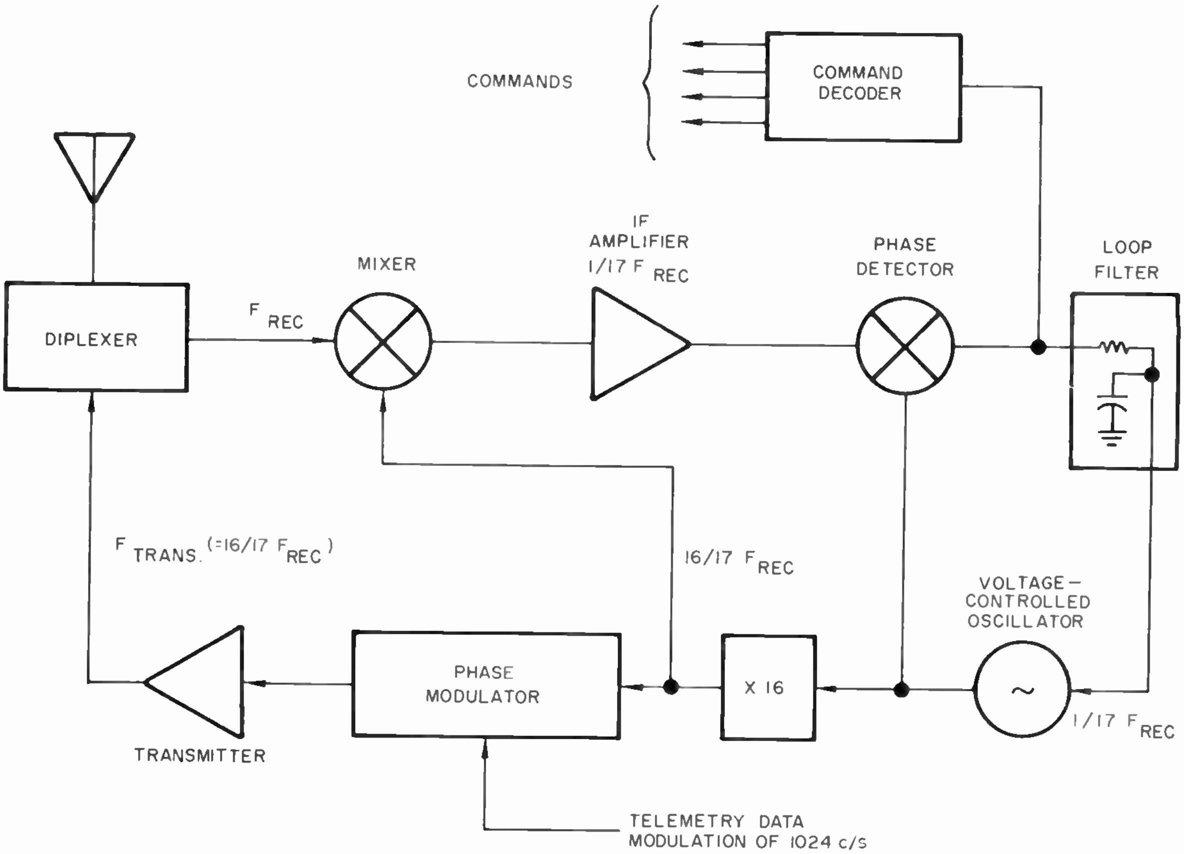


Fig. 18. Schematic of the coherent transponder.

analogue data such as magnetic field strength or temperature to digital form, and then sequentially modulates the transmitter with these data whenever the transmitter is on.

Figure 19 shows a photograph of the interior of the payload with the principal components identified.

There were two primary tracking stations on the earth which were in regular daily contact with the *Pioneer V* space probe. These were the Jodrell Bank radio telescope of the University of Manchester in England, and the Space Technology Laboratories' ground tracking station at South Point, Hawaii. Jodrell Bank is of course equipped with the world's largest (250-ft) radio telescope, and offered the best capability for communicating with the space probe at great distances. The Hawaii station was equipped with a 60-ft parabolic antenna. Except for antenna size, the two stations had identical electronic equipment, described in the following paragraphs.

The command transmitters used were capable of transmitting the 400-Mc/s signal at a 10 kW level. The transmitters employed conventional air-cooled tetrode amplifiers with coaxial-cavity resonators in order that the command signals could have extreme phase stability. Commands were transmitted at a low audio frequency using a peak phase modulation of 0·8 radian. Signals were received on the ground at $\frac{16}{17}$ of the 402-Mc/s command frequency or at about 378 Mc/s. The actual transponder which was used in *Pioneer V* swept its voltage-controlled-oscillator frequency back and forth across a 30-kc/s band continuously, seeking a command signal from the ground stations. When a command signal was received, the saw-tooth frequency search stopped and the receiver locked on to the carrier signal. The ground station had the option of commanding vehicle telemetry rates of 64, 8, or 1 bit/s, at the 5-watt (nominal) level. As soon as the 5-watt transmission had been initiated, the ground station could, at its discretion, turn on the filaments for the 150-watt transmitter in the payload, and then, after a short wait, turn on the anodes of the 150-watt transmitter. The payload could transmit for over an hour at the 5-watt level before the batteries were exhausted and could transmit for about five minutes at the 150-watt level. The payload was normally commanded off at the end of the desired transmission time. However, a low-voltage protective relay was included in the space probe, which turned off the transmitters automatically in any case when the battery voltage level reached its lower limit. (On at least one occasion this relay shut off the transmitter.)

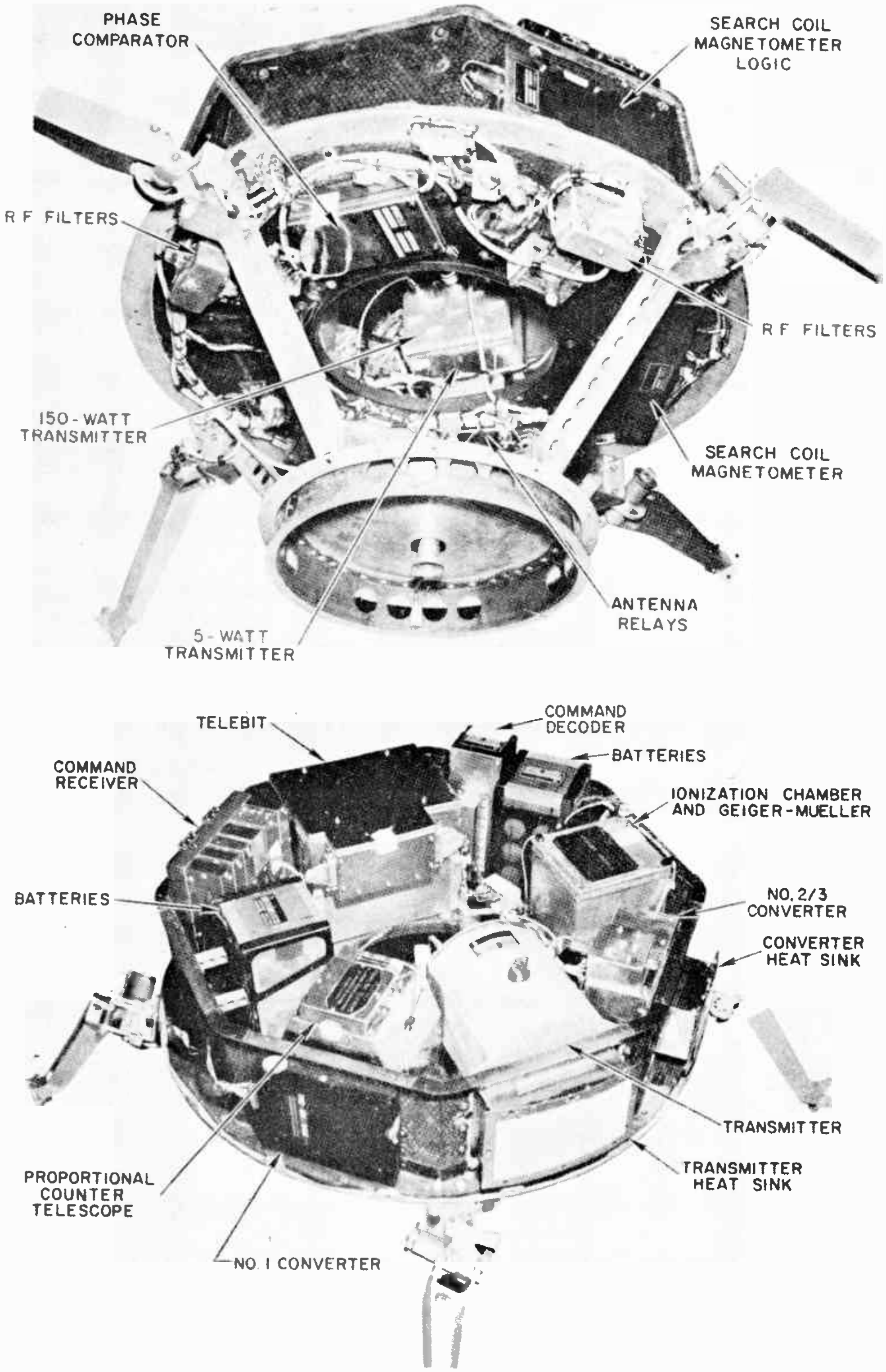


Fig. 19. *Pioneer V* payload.

Since the system operation required transmission and reception simultaneously (until the payload range had increased to the point that round-trip travel time exceeded transmitter on-time of five minutes at about 27 million nautical miles), a diplexer was provided at each ground station. The received signal was attenuated less than 0·2 dB by this diplexer and was then received by a phase-lock receiver employing a parametric preamplifier whose noise figure was 1·5 dB. The ground receivers, whose design was patterned closely after that of the microlock receivers developed by the Jet Propulsion Laboratory,[42, 43] were able to lock on to the signals at a received carrier level of −160 dBm, despite transmitter power on the same antenna feed of +70 dBm. The *Able* ground receivers contained an interesting device for coherent detection of data bits; a coherently sampled bit filter. This bit filter was gated by a 64, 8, or 1 c/s square wave signal obtained by counting down from an oscillator phase locked with the 1024 c/s sub-carrier. Synchronization of the countdown circuit was performed by adding or dropping cycles of the sub-carrier until the square wave synchronizing signal coincided with the data pulses. This circuit, together with the carrier and sub-carrier phase-lock loops and demodulators, performed as indicated in Fig. 20. A single point representing over-all performance is also shown in Fig. 13. It is to be expected that significant improvement will be achieved with subsequent equipment; experience gained with the *Pioneer V* equipment has been invaluable.

The telemetry information at 1024 c/s was automatically punched on standard five-hole teletype tape. Time information was automatically interlaced with the data on the same tape. At the end of each tracking period, the punched paper tape was fed into a tape reader for automatic transmission of the data back to the Space Navigation Center at Los Angeles.

Within a few hours after launch, the Center was able to predict the azimuth and elevation of the interplanetary space probe much more accurately than these quantities could be measured at the ground stations. These predictions were based on accurate measurements of position and velocity of the space probe at the burnout of the launching vehicle. Each of the stations always positioned its antennas according to the predictions supplied from the Space Navigation Center over leased teletype lines and did not attempt to refine the angular information by angular position measurements. Each station was, however, able to make extremely accurate Doppler measurement because of the coherent frequency translation which was performed within the space probe as described previously. The transmitted and received signal frequencies at a single ground station were compared, after multiplying the transmitted frequency by $\frac{16}{17}$ at the ground station, to develop a Doppler signal. A number of observations were made of the Doppler frequency, during each data reception period, by counting the Doppler frequency for 10 seconds at each observation. The stations reported back to Los Angeles the results of each tracking operation, and transmitted all telemetry data automatically to the Space Navigation Center by the use of the same teletype circuits. Corrections to the predicted ephemeris were made from the range rate data. Additional details of the Space Navigation Network have been given by Hansen and Spangler.[44]

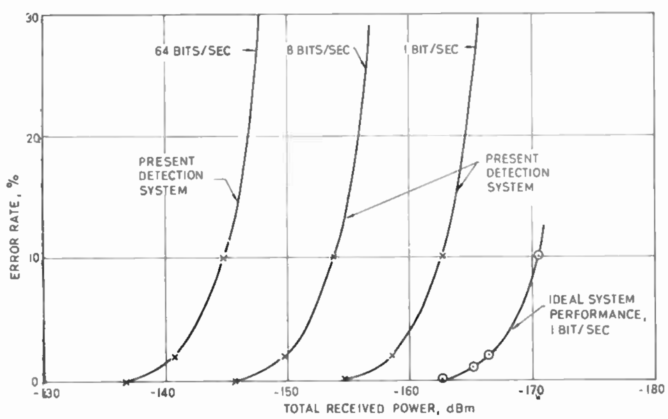


Fig. 20. Errors for biphase telemetry reception.

Table 2

Able System Calculations, from equation (8)

Vehicle transmitter—3 W actual	4·8 dBW
Vehicle diplexer and plumbing loss	−2·0 dB
Vehicle antenna directivity	0 dB
R^2 (1 nautical megamile)	−120·0 dB
f^2 (0·401 Gc/s)	7·9 dB
Ground antenna directivity (Manchester)	46·2 dB
Ground diplexer and plumbing loss	−0·7 dB
Polarization loss (linear-to-circular)	−3·0 dB
Constants	130·8 dB
System noise temperature (250° K)	−24·0 dB
Ideal S/N for 1% error rate	−4·3 dB
Subcarrier level (0·8 radian modulation)	−5·7 dB
Bandwidth—1 c/s	0 dB
Receiver and demodulation inefficiency	−7·8 dB
Excess signal	22·2 dB

Calculations of the expected performance of *Pioneer V* and the comparison with actual performance will be of interest. The general system considerations and the appropriate range equation were given in an earlier section; Table 2 summarizes the computations. Additional details have been given by Stephenson.[45]

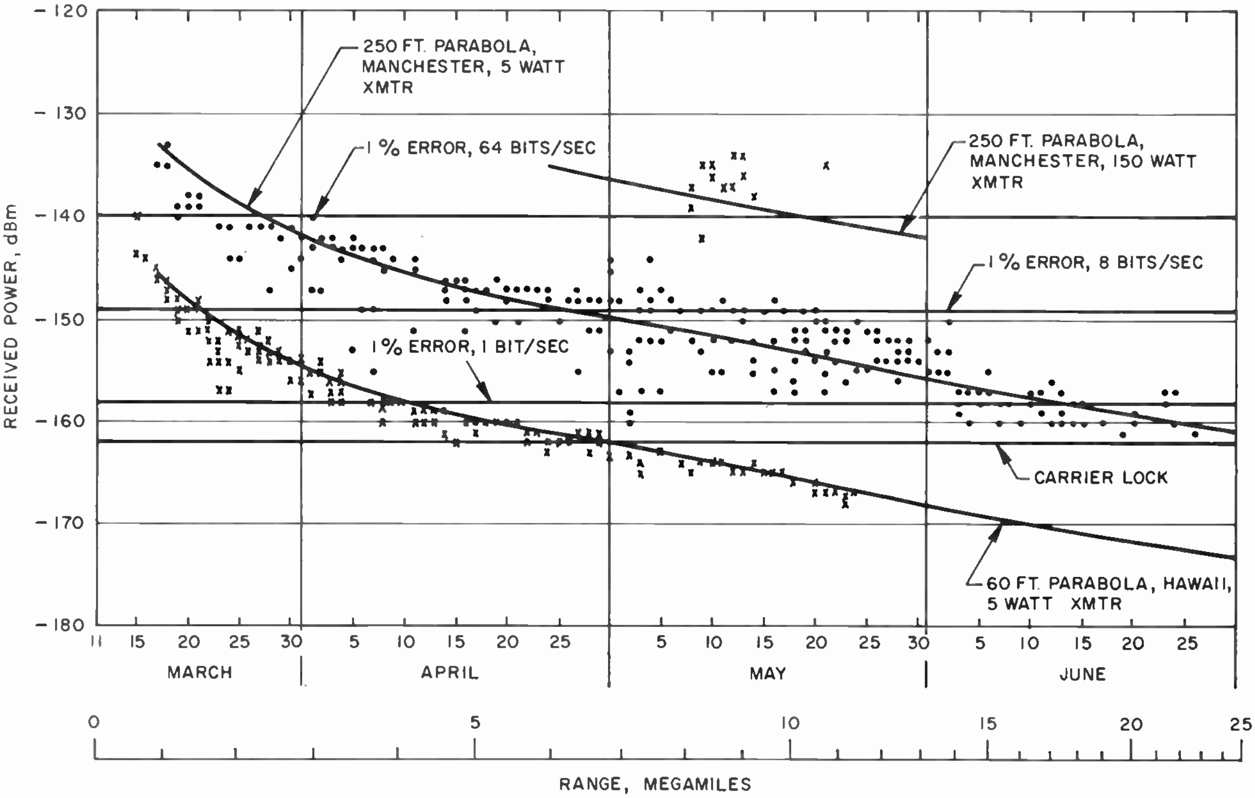


Fig. 21. *Pioneer V* signal levels.

These calculations are based on a range of 1 nautical megamile and on a data rate of 1 bit/s. Note that the over-all receiver and coding inefficiency of 7·8 dB is that obtained from Fig. 20. The system noise temperature constituents are detailed in Table 3.

Table 3

System Noise Temperature

Cosmic noise	45° K
Side-lobe noise	30°
Atmospheric absorption	10°
Plumbing loss	45°
Receiver and preamplifier	120°
	250° K

The calculations of Table 2 result in a predicted excess signal of 22 dB at 1 megamile. Actual signal strength readings fluctuated considerably, due at least partially to minor equipment performance variations. Figure 21 depicts the predicted and measured performance. Each dot and cross represents actual recorded ground station data. The predicted margin at 1 megamile is close to that observed in Fig. 21 for the Manchester parabolic antenna. In general, signal strengths appeared to be within 2 to 3 dB of predicted values at both Hawaii and Manchester.

The computations were made for the 5-watt (nominal) payload transmitter. The 150-watt payload transmitter was expected to allow communications to 90 megamiles. However, battery degradation allowed this high-power transmitter to be utilized only for very short periods and the communications record was, in fact, set by the 5-watt transmitter. Two-way communications were maintained to a distance of 23 nautical megamiles.

5. Acknowledgment

Gratitude is due to many colleagues at STL who participated in the *Able-5* programme, particularly John Taber and Jean Develet, to Dr. Cahn of the Bissett-Berman Corporation for information on coding, and to Mary Lee Buschkotter for invaluable aid in preparing the manuscript.

6. References

1. R. C. Hansen, "Low noise antennas", *Microwave J.*, **2**, pp. 19–24, June 1959.
2. D. H. Menzel, "The Radio Noise Spectrum", pp. 151–76 (Harvard Univ. Press, 1960).
3. H. C. Ko, "The distribution of cosmic radio background radiation", *Proc. Inst. Radio Engrs*, **46**, pp. 208–15, January 1958.

4. F. Horner, "Extra-terrestrial radio noise as a source of interference in the frequency range 30–1000 Mc/s", *Proc. Instn Elect. Engrs*, **107B**, pp. 373–6, July 1960.

5. R. Hanbury Brown and C. Hazard, "A model of the radio-frequency radiation from the galaxy", *Phil. Mag.*, Ser. 7, **44**, pp. 939–63, September 1953.

6. D. C. Hogg and W. W. Mumford, "The effective noise temperature of the sky", *Microwave J.*, **3**, pp. 80–4, March 1960.

7. J. D. Kraus and H. C. Ko, "Celestial Radio Radiation". Report AFCRC-TN-57-557, Radio Observatory, Ohio State Univ., May 1957.

8. C. H. Mayer, "Planetary radiation at centimeter wave lengths", *Astronom. J.*, **64**, pp. 43–5, March 1959.

9. C. H. Mayer, "Radiation from the planets", *IRE Stud. Quart.*, Sept. 1960.

10. Alex G. Smith, "Extraterrestrial noise as a factor in space communications", *Proc. Inst. Radio Engrs*, **48**, pp. 593–9, April 1960.

11. C. H. Mayer *et al.*, "Measurement of planetary radiation at centimeter wavelengths", *Proc. Inst. Radio Engrs*, **46**, pp. 260–6.

12. D. C. Hogg and R. A. Semplak, "The effect of rain on the noise level of a microwave receiving system", *Proc. Inst. Radio Engrs*, **48**, pp. 2024–5, December 1960 (Letter).

13. D. C. Hogg, "Effective antenna temperatures due to oxygen and water vapor in the atmosphere", *J. Appl. Phys.*, **30**, pp. 1417–9, September 1959.

14. C. W. Tolbert and A. W. Straiton, "An analysis of recent measurements of the atmospheric absorption of millimetric radio waves", *Proc. Inst. Radio Engrs.*, **49**, pp. 649–50, March 1961 (Letter).

15. E. S. Rosenblum, "Atmospheric absorption of 10–400 Kmcps radiation", *Microwave J.*, **4**, pp. 91–6, March 1961.

16. G. H. Millman, "Atmospheric effects on vhf and uhf propagation", *Proc. Inst. Radio Engrs*, **46**, pp. 1492–501, August 1958.

17. S. Weisbrod and L. J. Anderson, "Simple methods for computing tropospheric and ionospheric refractive effects on radio waves", *Proc. Inst. Radio Engrs*, **47**, pp. 1770–7, October 1959.

18. C. M. Angulo and J. P. Ruina, "Antenna Resolution as Limited by Atmospheric Turbulence". Report R-96, Control Sys. Lab., Univ. of Ill., July 1957.

19. G. W. Haydon, "Optimum frequencies for outer space communication", *J. Res. Nat. Bur. Stds.*, **64D**, pp. 105–9, March-April 1960.

20. S. Perlman *et al.*, "Concerning optimum frequencies for space vehicle communication", *IRE Record of Nat. Symp. on Extended Range & Space Communications, Geo. Wash. Univ.*, pp. 24–31, 1958.

21. J. H. Vogelman, "Propagation and communications problems in space", *Proc. Inst. Radio Engrs*, **48**, pp. 567–9, April 1960.

22. H. J. Pratt, "Propagation, noise, and general systems considerations in earth-space communications", *Trans. I.R.E. (Communications Systems)*, **CS-8**, pp. 214–21, December 1960.

23. R. S. Davies and C. S. Weaver, "Minimum transmitter weight for space communications", *Proc. Inst. Radio Engrs*, **47**, pp. 1151–2, June 1959 (Letter).

24. G. E. Mueller, "A pragmatic approach to space communication", *Proc. Inst. Radio Engrs*, **48**, pp. 557–66, April 1960.

25. G. E. Mueller and J. E. Taber, "An interplanetary communication system", *I.R.E.-WESCON Conv. Rec.*, Pt. 5, pp. 68–80, 1959.

26. M. E. Breese and P. J. Sferrazza, "Optimization of system parameters for deep space communication systems", *I.R.E. Globecom Conv. Rec., Chicago, Ill.*, pp. 180-3, 1961.

27. Gordon Raisbeck, "The optimal distribution of signal power in a transmission link whose attenuation is a function of frequency", *Trans. I.R.E. (Information Theory)*, **IT-4**, pp. 129–30, September 1958 (Letter).

28. J. T. Bolljahn, "Effects of satellite spin on ground-received signal", *Trans. I.R.E. (Antennas and Propagation)*, **AP-6**, pp. 260–7, July 1958.

29. R. C. Hansen, "Communications satellites using arrays", *Proc. Inst. Radio Engrs*, **49**, pp. 1066–74, June 1961.

30. B. M. Oliver *et al.*, "The philosophy of pcm", *Proc. Inst. Radio Engrs*, **36**, pp. 1324–31, November 1948.

31. C. R. Cahn, " Performance of digital phase-modulation communication systems", *Trans. I.R.E. (Communication Systems)*, **CS-7**, pp. 3–6, May 1959.

32. J. W. R. Griffiths, "Signal/noise ratio in pulse-code modulation systems—use of the 'ideal observer' criterion", *J. Brit. I.R.E.*, **19**, pp. 183–6, March 1959.

33. R. W. Sanders, "Communication efficiency comparison of several communication systems", *Proc. Inst. Radio Engrs*, **48**, pp. 575–88, April 1960.

34. A. V. Balakrishnan and I. J. Abrams, "Detection levels and error rates in pcm telemetry systems", *I.R.E. Conv. Rec.*, Part 5, pp. 37–55, 1960.

35. C. W. Helstrom, "The comparison of digital communication systems", *Trans. I.R.E. (Communication Systems)*, **CS-8**, pp. 141–50, September 1960.

36. H. F. Harmuth, "Orthogonal codes", *Proc. Instn Elect. Engrs*, **107C**, pp. 242–8, September 1960 (I.E.E. Monograph No. 369E).

37. A. J. Viterbi, "On coded phase-coherent communications", *Trans. I.R.E. (Space Electronics and Telemetry)*, **SET-7**, pp. 3–14, March 1961.

38. J. P. Costas, "Some notes on space communications", *Proc. Inst. Radio Engrs*, **47**, pp. 1383–4, August 1959 (Letter).

39. K. R. Sturley, "Frequency modulation", *Proc. Instn Elect. Engrs*, **92**, Pt. 3, pp. 197–207, September 1940.

40. S. Altshuler *et al.*, "The scientific objectives of the *Able-5* program", *Physics Today*, **14**, pp. 20–7, January 1961.

41. E. W. Greenstadt, "The data systems for *Explorer VI* and *Pioneer V*", *Trans. I.R.E., (Space Electronics and Telemetry)* **SET-6**, pp. 122–9, September/December 1960.

42. C. E. Gilchriest, "Application of the phase-locked loop to telemetry as a discriminator or tracking filter", *Trans. I.R.E. (Telemetry and Remote Control)*, **TRC-4**, pp. 20–35, June 1958.

43. M. H. Brockman *et al.*, "Extra-terrestrial radio tracking and communication", *Proc. Inst. Radio Engrs*, **48**, pp. 643–54, April 1960.

44. R. C. Hansen and E. R. Spangler, "Providing communication and navigation for space probes", *Electronics*, 8th July, 1960.

45. R. G. Stephenson, "Interplanetary Communication System Development". AIEE Special Publication T-126, Conference on Space Technology, 1960.

Manuscript received by the Institution on 6th July 1961 (Paper No. 675.)

POINTS FROM THE DISCUSSION

Mr. C. Williams: Mr. Stephenson displayed an interesting slide on the sun noise temperature, being of the order of 10^5°K. If we take a typical large aperture receiving aerial with a typical gain of 50 dB and sidelobe levels say 30 bB down on the main lobe, then for any orientation of the receiving aerial, noise power from the sun will enter the receiver through the side lobes at a level which is 10^5°K less 30 dB and is of the order of 100°K. Will Mr. Stephenson please comment on the apparent situation that sun noise is greater than source noise in a typical maser receiver.

Dr. R. Hansen (*in reply*): Mr. Williams is certainly correct in saying that the first few sidelobes of the antenna would see a noise temperature of 100°K, and this noise would probably exceed that from a maser receiver. However, for an antenna gain of 50 dB, the average weighted sidelobe level must be —50 dB and thus the sidelobe level decreases very quickly. Care must be exercised not to look at the sun within a few sidelobes of the main beam, and if antennas of higher gain are used, the safe zone will start farther from the main beam.

Dr. R. Jennison: There is much to recommend short focus paraboloids for ground systems as the reduction in side lobes and overspill to the ground reduce the extraneous contributions from the sun and earth to system temperatures.

Dr. Hansen (*in reply*): Short focus paraboloids are good; Cassegrain designs reduce spillover even farther. For narrow beamwidths of a few degrees or less, Cassegrain designs can produce very low aperture blockage and spillover. These probably represent one of the best low-noise antenna designs.

Mr. S. D. Abercrombie: The present electronic reliability figures, the short life of the satellite power supplies (solar or chemical), and the limitation in weight of the satellite using an "active", i.e. receiver and transmitter, system, all produce a very limited probability of a successful satellite system that will work for long periods (i.e. more than a year).

What is the feasibility, and cost, of a completely passive satellite system, using just a simple reflector on the satellite, say multiple corner reflectors, or parabolic system, or even a sphere? Is it possible to track accurately a satellite which does not radiate actively? Would it not be better to put the extra complexity of equipment at the ground stations where they can be maintained?

Dr. Hansen (*in reply*): Reliability of electronic subsystems is certainly a serious limitation at present and, in fact, limited the performance of *Pioneer V* to 22 megamiles. However, the extra path loss of a passive system for great distances is so great that it cannot be made up by ground station complexity at present. In reference 29 of the paper it is shown that for equal distances and identical ground stations, a passive satellite will have identical performance to an active satellite if the following conditions are met—

$$\sigma = G_{amp} G_{ant}^2 \lambda^2 / 4\pi$$

where G_{amp} is the active amplifier gain and G_{ant} is the active antenna gain and σ is the passive cross-section. To consider the *Pioneer V*/Jodrell Bank example, $G_{amp} = 170$ dB and $G_{ant} = 0$ dB. Thus for the same performance at ground station, a cross-section of 167 dB above 1 squared foot is required. This requirement could be reduced by increasing ground power somewhat and by decreasing system noise temperature. The ground station used a 250-ft antenna and 10 kW transmitter with about 300°K system temperature, so that about 20 dB improvement is the limit here. Additional gain could be realized by the use of an unfurlable passive modulated Van Atta array which would not require stabilization. However, even this does not come close to the 167 dB required. It appears that the use of passive systems must wait until technology allows use of unfurlable or inflatable reflectors many hundreds of feet in size.

In reply to the question regarding tracking of non-radiating satellites, even at lunar distances this is probably exceedingly difficult.

U.D.C. 522.14: 535.61: 629.19

Ultra-Violet Astronomy from Rockets and Satellites

By

D. W. O. HEDDLE, PH.D.†

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

Summary: The astronomical observations which can be made from the surface of the earth are severely limited by the presence of the earth's atmosphere. These limitations can be overcome by conducting observations from a vehicle at a sufficient height above the earth. The inaccuracies introduced by atmospheric refraction (scintillation and apparent displacements) are extremely small. The vehicle stability required to offer significant advantages for astrometric observations is perhaps too great for serious consideration at the present time, but ultimately a system with a short-term drift of less than 0·01 sec of arc per minute will require study.

At the present time, the most significant advantage to be obtained from the use of rockets lies in the extension of the observable spectrum to wavelengths less than 3000 Å. The experiments so far carried out have involved the installation in rockets of detectors sensitive to wavelengths within a band of a few hundred angstroms centred in the 1000 Å to 2000 Å region. The limitations of this technique lie in the short observing times dictated by the roll rate of the vehicle, and a vehicle with a certain degree of stability is urgently required. With the "pre-determined" vehicle it will be possible to obtain crude spectra of selected bright stars in the short time offered by a sounding rocket flight. The more stable vehicle would enable spectra of good resolution to be obtained from a satellite, or detailed broadband photometry to be carried out from a rocket.

The development of experimental astronomy has been limited in many ways by the earth's atmosphere. The inhomogeneities and movements of the atmosphere prevent the theoretical resolving power of large telescopes from being attained and so make it impossible to observe stellar discs. The random motion of a stellar diffraction image produces a considerable light loss in high resolution spectroscopy of the stars and consequently reduces the accuracy of intensity measurements. Any observation of a star is made against a background (or more correctly, through a foreground) of light from the sky, and for this reason it is not possible to observe stars fainter than a certain magnitude. A large proportion of the light from the sky originates in the atmosphere, either as light scattered from terrestrial sources or as the result of atomic and molecular excitation which produces the airglow. The third and, at the present time, most significant effect of the atmosphere is the restriction of the observable spectrum regions to wavelengths between 10 000 Å and 3000 Å, between 15 metres and 3 millimetres, and to a very few narrow bands in the infra-red.

† University College, London.

The development of vehicles which can carry scientific instruments beyond the atmosphere has enabled astronomers to consider experiments which could not be carried out from the surface of the earth. The diffraction image produced by a 40-in. telescope is approximately 0·1 seconds of arc in radius for visible wavelengths and the vehicle should be stable to this degree if the full advantage of avoiding atmospheric effects is to be obtained. If observations are to be made in the far ultra-violet the requirements for vehicle stability are increased. Stabilities of this order are not likely to be achieved for some years and the experiments which will be carried out from rockets and satellites will be concerned with exploring the previously unobserved regions of the spectrum. Radiation in the infra-red region penetrates quite deeply into the atmosphere and can to a large degree be studied from high altitude balloons, but in the ultra-violet, particularly at wavelengths less than 2000 Å, observations must be made from altitudes of 100 km or more and the rocket or satellite are the only suitable vehicles.

It would be as well to remind ourselves of the chief features of stellar spectra. We can regard the spectrum of a star as a continuum crossed by discrete spectrum

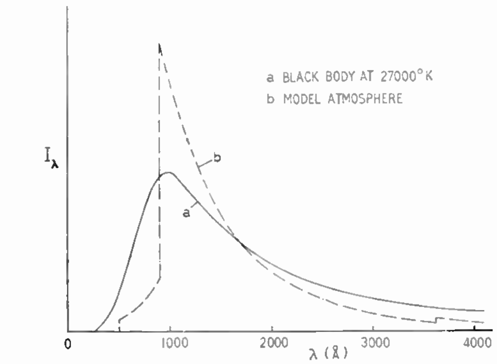
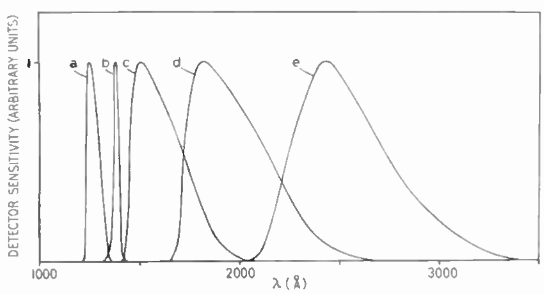


Fig. 1. The continuous spectrum of a star of spectrum class B I.



	WINDOW	SENSITIVE AGENT
a	Ca F_2	Nitric Oxide (Gas)
b	Ba F_2	Iodine (Vapour)
c	Sapphire	Platinum (Solid)
d	Silica (High Purity)	Gold (Semi-transparent)
e	Silica (Low Purity)	Cadmium (Semi-transparent)

Fig. 2. The spectrum sensitivities of representative selective ultra-violet detectors.

lines. In order to study the line spectrum we require an instrument having a resolution of ~ 1 Å and if we wish to study the absorption lines which appear in stellar spectra, because they originate in the inter-stellar medium, a resolution of $\sim 0{\cdot}01$ Å is desirable. The continuum presents a rather simpler problem, a resolution of tens, or even hundreds of angstroms being adequate. The brightest stars are, of course, the easiest to study, but the brightest stars in the ultra-violet will not be those stars which appear brightest to the eye. We expect that the very hot, early type stars will be the most conspicuous in the ultra-violet, but it is not easy to predict the relevant intensities. The continuous spectrum will almost certainly differ from that of a black body; perhaps in the manner indicated in Fig. 1. Different assumptions about the stellar model atmospheres will lead to different predictions of the ultra-violet intensities and experimental measurements are necessary.

Studies with low spectrum resolution can be carried out by using detectors sensitive to a limited band of wavelengths and do not require the use of conventional dispersing elements. There are no filters, in the usual sense of the word, for wavelengths less than about 2500 Å, but one can make selective detectors by a suitable combination of photo-electric sensor and window material. Some examples are shown in Fig. 2 in which all the maximum sensitivities are normalized to unity. All the experimental work so far reported has made use of detectors of this general type mounted on an unstabilized rocket and provided with some collimation device that allows the detector to receive light from a limited region of the sky.† The roll and precession of the vehicle cause the field of view to sweep across the sky and, provided we know the vehicle attitude at all times, we can interpret the telemetered data in terms of stellar brightness.

One example of this type of instrumentation is shown in Fig. 3. The detectors were photomultiplier tubes with semi-transparent gold cathodes deposited on silica windows (response (*d*) of Fig. 2). The fields of view of five of the six detectors were limited by lengths of aluminium honeycomb to a circle of some

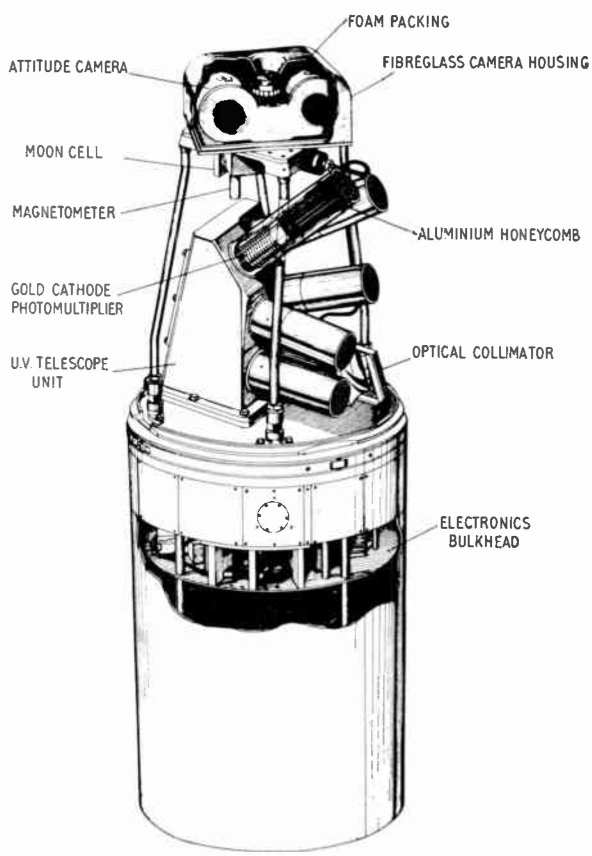


Fig. 3. The photometric installation in *Skylark 43*.

† A. Boggess III, *Mem. Soc. R. Sc. Liège*, 5th Series, **4**, p. 459, 1960; E. T. Byram, T. A. Chubb, H. Friedman, *Mem. Soc. R. Sc. Liège*, 5th Series, **4**, p. 469, 1960.

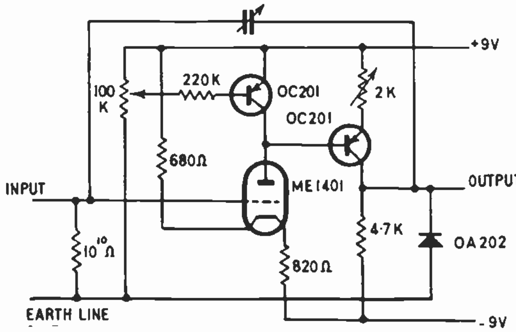


Fig. 4. The logarithmic amplifier.

4½ deg in diameter. A cylindrical Cassegrain mirror system was used to give the sixth detector a field of view less than 1 deg wide extending some 50 deg in a plane containing the rocket axis. The outputs of the photomultipliers were fed into logarithmic amplifiers (Fig. 4) which gave outputs suitable for the rocket telemetry system. The attitude of the vehicle was measured by magnetometers and by photocells which detected the passage of the moon across a slit system. This vehicle, a *Skylark*, was successfully launched from Woomera on 1st May of this year. The data is still being analysed but ultra-violet emission in the 2000 Å region was certainly observed.

This type of experiment, while making little demand on vehicle performance, can make comparatively heavy demands on telemetry. It would be unwise to anticipate a roll rate of less than 100 deg/second from an unstabilized vehicle, and if we wish to be able to detect (in round figures) 1000 stars, we must limit the field of view of our detector to about 1 deg square in order that we may be reasonably sure of having only one star in the field at a time. To determine the intensity and position of an ultra-violet source adequately we require a telemetry bandwidth of about 1 kc/s for each of our detectors. In principle we could increase the sensitivity of the system, and be forced to reduce the field of view—with the consequence that we would require greater telemetry bandwidth—but we are limited, with presently available detectors, by considerations of signal statistics. The signal from a fairly bright star, as detected by a gold photo-cathode, is some 10^3 photo-electrons/cm^2/s. We can justify a 1 kc/s bandwidth only if we receive about 5 photo-electrons/millisecond from a star, so we require a collecting area of 5 cm^2 to be confident of detecting such a star with a field 1 deg square. Our choice of a field of view of this size was made on the basis of 1000 detectable stars and, as the number of stars brighter than a given magnitude, m_c, varies approximately as 3^{m_c}, our system must be capable of detecting stars some five magnitudes (a factor of 100) fainter than our fairly bright star. The collecting area required to make this possible is 500 cm^2, a mirror of 10 in. diameter. This is about as large a mirror as can be fitted into *Skylark*. An extension of photometric experiments to still fainter stars will require a larger unstabilized vehicle or a *Skylark* with a certain degree of stabilization, particularly in respect of roll rate. The observation time available during the flight of a vertical sounding rocket is only a few minutes, and it would be advantageous to increase this by firing rockets to greater heights. The *Skylark* of 1st May reached an apogee of about 150 km, and spent some 200 seconds above 100 km. In order to double this observation time an apogee of 300 km would be necessary and the telemetry power required would be increased four times.

With the development of stabilized satellites it will become possible to carry out experiments of increasingly higher spectrum resolution. An approximate figure to be kept in mind is that a spectrum resolution of 1 Å requires a vehicle stability (more correctly, a stellar image stability) of approximately 1 minute of arc. The rate at which the spectrum can be scanned will be governed more by the limitations of the small signals than by telemetry bandwidth considerations, though the need to record data throughout the orbit and transmit it for the 5 minutes or so that the satellite is within range of a ground station, will need to be kept in mind. One can telemeter continuously from a satellite in an orbit of 24 hour period, but the transmitter power required is comparatively large.

The broadband photometric measurements already carried out and currently planned serve two main purposes. The information they yield will enable the later, high resolution experiments to be designed in an optimum fashion and, because of the comparative simplicity of broadband photometry, the situation will be that results of this type will be available for many more stars than one can hope to study in detail. Photometers mounted on stabilized satellites will be capable of detecting stars at greater distances than will similarly mounted high resolution instruments, and may well play a significant part in the elucidation of the problems of the evolution of the Universe.

Manuscript received by the Institution on 5th June 1961. (Contribution No. 39.)

U.D.C. 621.396.946 : 621.397.7

Television Communications using Earth Satellite Vehicles

By

L. F. MATHEWS,
(*Member*)†

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

Summary: This paper describes the type of orbits, and operational frequencies that could be used for a television relay satellite system, including the choice of modulation method and repeater equipment. This is followed by a consideration of the requirements to be met in international television exchanges, including waveform specifications. Television observation satellites (e.g. Tiros) are also described. The paper concludes with a speculation on the development to be expected with satellite television communication systems, and the problems to be encountered.

1. Introduction

A great deal has been written about, and a number of international research units have been studying, the possibilities and problems of radio and television transmissions, either from—or via—earth satellite vehicles. A very positive contribution to this study was made at the Thirteenth General Assembly of the International Scientific Radio Union held in London in September 1960, when special sessions on space radio research were held, and two committees were formed, one dealing with Space Radio Relays, and the other Radio Research.

Although a lot of work has still to be done before public service transmissions can be expected, many experiments are now proceeding, particularly in the United States, where the Bell Telephone Laboratories, R.C.A., and A.T.T., amongst others, are making rapid progress.

This Paper will firstly discuss some of the possible orbits, and the signal frequencies that could be used for radio and television relay transmissions, and also the type of equipment most likely to be carried in the satellite.

Following this, the author will consider the progress being made towards television communications via satellites, as a result of experience with vehicles at present in orbit, and the paper will conclude with a speculation on further developments.

2. Type of Orbit

The selection of an orbit in which to operate a communications satellite presents a number of problems, on account of the advantages and disadvantages of any particular orbit. The final decision, therefore, must be a compromise, which will take into account a number of factors, including the following:

(a) The economics involved in the launching (size of launching rocket to be used).

(b) Area of land to be covered at any one time.

(c) Size and complexity of ground equipment.

(d) Size and complexity of satellite equipment.

It might appear from a first consideration that a satellite system covering practically the whole surface of the earth, for twenty-four hours a day, must be aimed for, such a communications system, including radio and television, could then provide a service to any desired point on the globe. For such a system, a minimum of three satellites in a circular equatorial orbit would be required, orbiting at 22 300 miles above the earth. At this height they would effectively remain stationary with respect to the earth, as their rotation would be equal to the earth's period, i.e. a synchronous orbit, and correctly aligned transmitters and receivers on the earth would allow of continuous transmissions. Unfortunately this system presents a number of important problems:

(1) Projecting the satellite to this altitude requires an extremely costly assembly of rockets.

(2) Having placed the satellite in the required orbit, it requires the addition of complex controlling equipment built into the satellite to maintain its period and altitude with respect to the earth correctly.

(3) At 22 300 miles above the earth, signals between the two locations will suffer high signal attenuation, which in order to offset, will require very high-power base transmitters, and super-sensitive receiver equipment; whilst in the satellite the equipment will likewise have to be capable of higher sensitivity and power output. In addition, the use of passive repeaters can—for all practical purposes—be ruled out.

† Associated Television, Ltd., London, W.1.

(4) Signal/noise ratio will suffer, due to increased interference, resulting from interstellar and ionic plasmas, etc., which again introduces further complications in the design of receiving equipment.

At lower altitudes, within the range of 500 to 4000 miles, many of the problems just described can be reduced appreciably, but in their turn introduce other difficulties. For instance, a satellite orbiting at a height of 500 miles, will circle the earth in a period of $1\frac{3}{4}$ hours, and at a height of 4000 miles, in four hours, and as a result of the high speeds of the satellite, relative to the earth's rotation, a severe limitation is placed on time availability over any particular area. In addition, complicated tracking equipment must be coupled with the transmitter and receiver aerial systems of the ground stations, in order to maintain contact with the satellite. If a continuous service at any one point is required, then a large number of satellites located along the orbit will be required to provide an almost continual service.

A compromise solution to the above problem was, however, put forward by Hilton and Dauncey in a paper to the International Astronautics Federation at Stockholm in August 1960,[1] proposing a different approach to this problem. This proposal was to launch the communications satellite into an elliptical orbit, inclined at approximately 63 deg to the equatorial plane. It is claimed that such an orbit would satisfy—to a very large degree—the communications requirements, and at the same time reduce many of the disadvantages of the synchronous and the low altitude orbit.

In the proposed 63 deg orbit, the apogee would occur at latitude 63 deg North, and be at a height of approximately 12 500 miles above the earth's surface. The perigee would occur in the Southern hemisphere, at a height of approximately 300 miles. This orbit would provide a satellite availability for communication over a large portion of the complete orbital time, i.e. approximately nine hours per day. By correctly disposing three such satellites, in complementary orbits, almost 24-hour coverage could be achieved. The case made for this system is based upon the following:

(1) The apogee remains fixed at 63 deg latitude, and therefore at this point does not require complex tracking.

(2) Rockets at present available could satisfactorily launch satellites into this orbit at its perigee point of 300 miles.

Although the satellite is not available for continuous communication, this is no great problem, as communication requirements for a particular area are small during the night time period. For instance, radio and television services would be closed down, or special programmes could have been recorded during satellite availability time. In the case of telephony, reduced traffic during night time can be accommodated by international cable and conventional radio circuits.

Hilton and Dauncey have calculated practical service areas that could be accommodated by a satellite system in the 63 deg slot, and have shown that service between points reasonably high in the Northern hemisphere could be good. For instance, between London and New York, communication time, using one satellite, could be for about nine hours a day, and using two satellites 180 deg out of phase in orbits 180 deg apart in longitude, a communications period of eighteen hours per day could be provided between London and New York; while the use of three satellites 120 deg out of phase, with orbits 120 deg apart gives virtually a 24-hour coverage between all points approximately 35 deg or 40 deg North.

3. Choice of Operational Frequencies

Irrespective of which of the foregoing orbits is finally adopted, and whether passive or active repeaters are used, the choice of operational frequency is broadly limited to the spectrum between 100 and 10 000 Mc/s.

Atmospheric and man-made noise, which decreases inversely with the square of frequency, is at such a value below 100 Mc/s as to impose virtually impossible receiver design requirements, and this is, therefore, the lowest usable frequency when considerations of high gain and unwanted noise rejection are taken into consideration. In fact a more likely minimum value is of the order of 1000 Mc/s, which not only allows for sufficient signal and noise separation, but will also provide a sufficiently broad bandwidth for an adequate signal/noise ratio when operating with the prescribed modulation techniques.

Attenuation and absorption of the transmitted signal, due to water, vapour and oxygen, increases with the cube of the frequency, and as a result, puts a practical upper limit at approximately 10 000 Mc/s. Two further factors in determining ideal operating conditions are, however, conflicting, for although the use of higher frequencies means a reduction in size of some satellite components, with the resultant reduction in weight, at the same time power efficiency is reduced. This latter fact is important when d.c. input power must be developed by the use of devices forming part of the satellite.

Taking all the conflicting factors into consideration, the resultant frequency band which appears best suited for satellite communication systems is between 2000 and 4000 Mc/s. The decision to use frequencies within the ideal band, however, introduces many possibilities of interference from and to other services, which already extensively use frequencies within this part of the spectrum. This situation would be aggravated, due to the high sensitivity required by the satellite ground receiving station, and also the necessary high powers for ground transmitters. To alleviate this situation, geographical guard areas have been proposed, where within a prescribed area around transmitter and receiver, equipment using similar frequencies will not be allowed to operate. This would probably be a quite impracticable solution for a small island such as Great Britain, with a large concentration of microwave operations. From calculations it has been suggested that interference could permeate to a distance of 38 miles at frequencies of 2000 Mc/s, and to a distance of 55 miles at 6500 Mc/s.

Many recommendations have been made to the F.C.C. in the United States for combating this problem, among which have been the proposal that all present allocations be rescinded, and a reallocation can be carried out, reserving the frequency band 1000 to 10 000 Mc/s for space communications. As communications via satellites will have international implications, irrespective of orbit to be operated, it is plain that any plan will have a wider effect than that area governed by the F.C.C., and as a result, frequency allocations will need to be discussed and agreed internationally.

4. Choice of Repeater and Modulation Methods

A communications satellite may be designed to operate either as an active or as a passive repeater. With the former, the satellite contains a transceiver, which is capable of receiving a signal from a transmitter on the ground, and re-transmitting the signal back to a ground receiving system after frequency change and amplification. In the case of a passive repeater, however, the satellite provides a reflecting surface, which is for returning the ground transmitted signal back to the ground receiver. The active satellite method requires smaller transmitter power, and lower gain ground aerial systems than that for passive satellites, and also less sensitive receivers, allowing better signal/noise ratios to be achieved. Against the advantages of the active satellite, the passive counterpart contains no electronics, and so has none of the problems associated with space borne equipment. As a result, any improved equipment techniques and designs can be embodied, whereas this is at the present time quite impossible with satellite repeater equipment. Additionally, at the present stage of the art, once an active repeater has been placed into orbit, it can only accommodate the services for which it was originally designed, whereas it is possible for additional services to be established via the passive repeater.

Some of the disadvantages associated with either the passive or active repeater system are being off-set by new developments in equipment design, and the various problems are being investigated in practice by the experience being gained with passive satellites, such as the American *Echo I* and active satellites such as *Courier*.

In addition to the problems concerned with choice of frequency for satellite communication systems, propagation of the radio signal can be affected by the formation of an ion sheath around high altitude vehicles. Such a sheath has been predicted to exist at altitudes as low as 1 000 000 ft (305 km), and such a sheath could cause severe attenuation to the radio signals passing through it. This has not been completely established by any recent observations, although the existence of such an ion sheath around a vehicle will exaggerate the actual size of the vehicle. Problems concerned with such an obstruction consist in determining more accurately its area, electron density, and also its effect on propagation of electromagnetic waves. This in its basic form is a propagation problem, which is similar to signal transmission in the ionosphere, as the sheath will have a normal incident critical frequency and a propagation constant, affecting the index of refraction and the absorption per unit thickness.

A further problem can arise due to extra-terrestrial noise, which is predominantly cosmic noise, and needs to be taken into account in the overall system design, especially as it influences aerial configurations. A communication mode at present under investigation is one utilizing the magneto-ionic duct, in which a signal entering the ionosphere encounters several modes of propagation, and many of these modes at frequencies below the plasma frequency are reflected back to earth. One mode—the extraordinary, quasi-longitudinal—is not reflected, but continues to be propagated with low loss, through the ionosphere to the exosphere, and back to earth. This mode has possibilities for the communications from ground to ground, and also between ground and space vehicle.

Although much has still to be learned of propagation effects at high altitudes, it is unlikely that future investigations will affect appreciably the choice of optimum frequencies for transmission of signals between earth and satellite. Carrier modulation, however, for active or passive television relay satellites can be achieved by either amplitude or frequency modulation. The former has the advantage of smaller bandwidth requirements, but the latter allows for a much simpler design of the modulation equipment. In addition, the f.m. system, having a triangular noise distribution provides a better signal/noise ratio than the flat characteristic of the a.m. system. As a result of these factors, frequency modulation is likely to be generally adopted. However, with an f.m. system the receiver bandwidth must normally be large, in order to accommodate the frequency deviation of the transmitted signal, and in fact for television transmission requires from 25 to 50 Mc/s. With such wide bandwidths, the necessary highly sensitive receiver would receive so much noise in addition to the wanted signal as to be completely unusable, and so for all practical purposes it is necessary to use an f.m. system, employing a negative feedback arrangement such as was first suggested by Chaffee.[2] This arrangement requires only a narrow bandwidth for the receiver input stages in the region of one-tenth of the transmitter frequency deviation. The receiver tuning is so controlled that it automatically follows the transmitter carrier, but at the same time the receiver only presents the restricted bandwidth input to its aerial system.

5. Equipment for Satellite Communications

In the case of the passive repeater, there is generally no equipment requirement within the satellite, for the purpose of returning the transmitted signals to earth, as this is achieved by reflection from the surface of the vehicle. As has been stated, however, very high power transmitters and extremely sensitive receivers are required, both utilizing highly directional aerial systems.

The requirements of the receiver can largely be met by utilizing the advantages offered by parametric amplifiers and masers, and receiver tuning tracking systems, adopting the Chaffee principles. With conventional receiver design it would be difficult to produce equipment in which the inherent circuit noise would have a lower value than required signal, but the principles of operation for both the parametric amplifier and maser reduce the inherent noise to a practical level. Although both will operate satisfactorily at microwave frequencies, to achieve the low noise value demands special requirements.

The parametric amplifier derives its principal source of noise from the idler circuit, and to reduce this it is necessary to make the idler frequency as high as possible compared with the signal frequency, other forms of noise such as "Johnson noise" being reduced by cooling the idler circuit. In the case of the maser, noise is kept to a low value by operating the device at a temperature of approximately 4° K. The simplest type of parametric amplifier employs a semi-conductor diode as the non-linear reactor, but a parametric amplifier using an electron beam as the non-linear reactor—such as devised by Adler[3]—has been successfully produced, giving an extremely low noise performance, and wider bandwidth at microwave frequencies.

The foregoing amplifiers in their normal form, however, have certain disadvantages, among which are:

(a) In order to achieve the very low noise figure, the product of the gain bandwidth, is rather small.

(b) The amplifier can easily drift into instability.

(c) The input and output circuits are not normally isolated from each other, and noise amplified in one stage can be returned to the previous stage.

(The Adler tube previously mentioned, however, does not suffer from the basic defects, except that the bandwidth is rather restricted.)

In order to overcome the foregoing disadvantages, the travelling-wave tube principle has been embodied in the amplifiers to produce travelling-wave masers, and travelling-wave parametric amplifiers.

It has been previously mentioned that in order to reduce appreciably the effects of received noise, a feedback system is introduced in the receiver to provide a narrow bandwidth in the tuning stage, which will follow the carrier deviation from the transmitter. There are, of course, a number of problems associated with the use of this arrangement, as it must be adjusted to follow the transient type of required incoming signal and yet reject the effect of transients resulting from unwanted noise.

Turning now to the relay system requiring the use of an active repeater the major problem becomes one of designing transceiver equipment sufficiently small and reliable enough for installation within the satellite, and also provision of equipment capable of providing the necessary power for the operation of the repeater equipment. Masers or parametric amplifiers are—at the present stage of the art—unsuitable for installation and operation in a satellite, and therefore repeaters employed over the near future are most likely to continue with more conventional travelling-wave tube arrangements. Many new microwave tubes have been developed for this purpose, amongst which are multiple beam klystrons, combining a number of beams within a single vacuum envelope and the orthitron cross-field tube, embodying travelling waves.

The latter, operating in tandem connections, can provide saturated gain in excess of 60 dB, this level of gain assuring a minimum power output of 1 W from an input drive of 1 μW approximately, over a frequency band of 2000 to 4000 Mc/s. Circuits using these tubes can be reduced in size and weight by the use of miniaturized periodic focusing systems, made possible by the small-diameter helical slow-wave circuits and the small external r.f. helical transducers. The complete amplifier chain can be made to consume less than 25 W (exclusive of heater power) with an applied maximum potential of 1100 V, whereas a similar system employing conventional travelling-wave tubes would consume ten times this power, purely for the focusing requirements. The use of a number of microwave tubes in tandem, can be overcome by the use of a reflex arrangement, a technique where the input signal is passed through a travelling wave at a frequency F_1, the output being then reintroduced into the modulator by the use of suitable filters, and an output being derived at a frequency F_2. This signal is then sent back through the gain channel, where the process is repeated, after which further frequency conversion is then carried out prior to the signal transmission. This process cannot, however, be achieved all in one tube, due to the inherent low power capabilities of a single tube, and so at least two tubes are required to complete the conversion with an adequate gain.

The ancillary equipment required for the operation of the travelling wave tube circuitry, can to a large extent consist of transistors, as the latter are less susceptible to vibration and shock effects, the reduced power requirements and light-weight character of transistors being additional reasons for such a choice.

For the operation of satellite equipment, conventional batteries are quite unsuitable, owing to the weight and space required, and the most practical method of providing the electrical power is by the use of solar cells arranged in banks, which can operate directly when facing the sun, and at the same time charge light-weight storage batteries, providing alternative power when the cells are not within the sun's rays. The solar cell output, directly, or via the storage batteries, can be used to operate equipment directly, or after voltage regulation; it can also be converted to a.c. where required. Many solar cell systems consist of series connected silicon photo-voltaic converters mounted on an aluminium sheet metal structure by a ceramic cement, the cells themselves often being protected from micro-meteorite damage by placing behind fused silica windows; the output can then be connected in parallel with similar banks, and via silicon diodes to the equipment or batteries.

In order to achieve a good operational signal path between satellite and ground equipment, it is necessary that space-borne aerial systems be used, capable of high gain and directivity, and this presents a number of problems in the design of the satellite. In addition, the satellite needs to be stabilized for spin and any attitude variation, if high directivity aerials are not to move out of focus with their counterparts on earth. Although very little attempt has been made to date to provide such highly directional aerial systems, this requirement will become very much more important for relaying radio and television signals, if reliable quality is to be achieved.

The above requirement will also apply to ground aerial systems, and at the present time large horn aerials have been amongst those adopted for reception of space signals. An interesting satellite design has been suggested by Hawker Siddeley Aviation[4] in which, to maintain bearing, the satellite attitude stabilization is accomplished by means of jets, the horizon sensing unit protruding through the light alloy skin enclosing the main body.

6. Experience so Far

At the time of writing, it is believed that seventeen satellites are in orbit around the earth, but the event has now become so commonplace that this number can be greatly in error in even a very short time. Initially, one reason for launching a satellite was to determine possible orbits, and to investigate the effects on a vehicle in differing orbits; this was followed by installing equipment within the satellite to provide information on weather conditions and structure of the upper atmosphere, so giving more detailed information on the earth's shape, and many other imponderables.

In satisfying the foregoing requirements, these satellites have provided information which is important for the design of equipment, and systems necessary for achieving radio and television programme relays. Latterly, satellites designed specifically to investigate space communications have been placed into orbit, and the results so far have been most encouragingly satisfactory. What are considered the most important advances in the history of satellites are catalogued in Appendix 2.[5]

Three satellites which are worthy of particular note, however, are *Tiros I*, *Echo*, and *Courier*, as they have in their instrumentation satisfied many of the requirements of a television satellite.

6.1. *Tiros I*

This American satellite, whose title is an abbreviation for Television Infra-red Observation Satellite, was designed to take television pictures of interest to the meteorologist, to record and to transmit them to the ground on the receipt of the command signal. For this purpose two half-inch vidicon cameras are mounted with their lens assemblies projecting through the satellite bottom surface, and a miniaturized magnetic tape recorder is associated with each camera (Figs. 1 and 2). The tape capacity of each machine is 400 ft of a special half-inch mylar base tape moving at a rate of 50 in/s during recording and playback; as the tape is played back to the ground, it is erased in readiness for further recording. The top and sides of the satellite are covered with solar cells, in order to provide the 18 W of power (Fig. 3). As it is important for the television cameras to be maintained in a to provide a solid unit, and resistors, capacitors, inductors, transistors, diodes, and crystals can be formed into such micro-elements, resulting in a ten-to-one reduction in size over the standard type of miniaturized component. The minimum overall size, however, is determined by the soldered connections between the wafers, this point being also the most likely source of component breakdown. The two-dimensional or printed circuit, on the other hand, also produces large reductions in weight and size, but the reliability factor at the present time is apparently not sufficient to be acceptable for use in the satellite, where maintenance and changing of circuit cards cannot be undertaken.

The rapidly increasing knowledge arising from experiments connected with the introduction of impurities into materials, in controlled amounts, is, however, opening up a comparatively new field for solid circuit design. By these methods, based on a knowledge of semi-conductor materials, it is possible to produce microscopical blocks, which can operate as a complete unit, i.e. oscillator or amplifier, etc. The various blocks can then be clamped together, and with only input/output connections, provide a complete and complex circuit. Although the method of miniaturization can produce extremely small circuit configurations, it is necessary that the circuits operate with a high degree of reliability for indefinitely long periods under conditions of stress and varying temperatures. Also during manufacture close tolerances will have to be achieved, as circuit adjustment will not be possible after construction. However, as the problem of reducing size is an extremely important one, it will be necessary for experiments to continue in the whole field of miniaturization. At the same time, development with the miniaturization of television camera systems has been proceeding, and a number of manufacturers, notably R.C.A., have designed 1 in. vidicon cameras 5 in. long and 7½ in. diameter, weighing approximately 4 lb, and producing a video bandwidth of 62·5 kc/s, capable of giving a line resolution in the order of 800–1000 lines.[6]

8. Orbital Scatter Communications

A communications system has been proposed primarily by W. E. Morrow[7] of the Massachusetts Institute of Technology, in which a belt of dipoles placed around the earth at altitudes of 3000 to 6000 miles, provide a reflecting medium for a radio signal at microwave frequencies, the most favoured orbits being the polar or equatorial orbit. Using high powered transmitters and sensitive receivers, the dipole layer would then act as a passive reflector, and transmissions over extremely long distances could be achieved. Due however to the very serious doppler effects that could occur, the received signal would undergo rapid phase changes and variations of the transmitted bandwidth. The system therefore introduces so many distortions, that it is unlikely to be satisfactory for conventional telephone systems, or radio and television transmissions, and as a result its main use appears to be confined to the transmission of digital information, where changes of the phase characteristic does not have such disastrous effects.

9. Television Services

Many of the problems and solutions that have so far been dealt with in this paper are applicable, whether the transmission is connected with multi-channel telephony, radio, or television. Propagation at microwave frequencies is mainly unaffected by the spectrum requirements with either service, particularly as the bandwidth for multi-channel telephony—capable of accommodating sufficient channels—is comparable with that required for a Television Service.

In the overall system, however, a number of different requirements exist, and must be accommodated if the required intelligibility is to be preserved. Contrary to popular belief, the linearity requirement is of greater importance in the case of wideband multi-channel telephony, than that for the transmission of television signals. With the latter, unless non-linearity is severe, the noticeable change in the relative gradation of tones between black and white is small and acceptable; in addition, the picture/synchronizing ratio can be corrected by use of clamp circuit equipment. In the multi-channel telephony case, non-linearity can produce both intelligible and unintelligible cross-talk between the various channels. An important consideration for television, however, relates to the phase/frequency characteristic, which although relatively unimportant in the multi-channel telephony link, requires a group delay substantially flat from d.c. to 3 Mc/s (for the British system) in the case of television.

As a television link, via a satellite, will require extension over a conventional ground network at either terminal, to preserve reasonable quality over the complete connection, it will be necessary that the satellite link provides a high quality transmission path. The various ground networks are designed and maintained to operate within certain prescribed limits, and this condition will have to apply also to the satellite link. A suggested specification is shown in Appendix 1.

The bulk of the world's television transmitters and networks are confined to the North American continent, and to Europe, including Great Britain, and in both of these areas a large ground network is employed to provide the connection between national networks and programme operators. The television satellite will achieve connection between these two large networks, so that programme material may be exchanged, and will require in its simplest form, one satellite ground station associated with each network. This basic requirement can be elaborated by the use of other strategically disposed ground stations, connected into other points on the network, so providing standby and re-routing facilities, in the event of ground station or ground network failures.

Depending upon orbits selected, a number of satellites in operation, service can also be provided to other parts of the world, including the British Commonwealth, Japan, and other widespread television networks. A scheduled service could then be introduced, covering that part of the day most suited between two or more networks. A direct mobile service can also be envisaged, whereby mobile camera equipment and ground station transmitters, using collapsible 50 ft dishes, could then be sent to areas not normally connected to a network, and from this point carry out a transmission directly via the appropriate satellite. As a result, therefore, a "world wide" coverage of television could be achieved, which apart from connecting the various permanent networks, for exchange of programmes, would enable outside broadcasts from a great number of locations to be linked with the international network.

There are, unfortunately, a number of operational problems with such an international network as just described.

9.1. *Incompatibility of Line and Frame Scanning Standards*

The Table in Appendix 3 shows the main scanning standards used throughout the world, from which it is seen that any direct exchange between major operating countries requires the adoption of standards conversion procedures. At the present time, such an operation inevitably introduces a number of distortions, which although acceptable for news and some actuality events, does not provide picture translation of sufficiently high quality for normal entertainment programmes. Although C.C.I.R. recommendations advocate 625 lines as a standard, it is unlikely that the 525-line system will now ever be disbanded throughout the North American Continent, and so the only solution to this particular problem is the development of an extremely high quality standards converter of the completely electronic type. Exchanges of television programmes recorded on film, and then flown to other parts of the world, do not suffer from this problem, and can be used directly at any location equipped with the requisite telecine equipment. Video tape recording does however suffer from the need to convert standards, in the same way as the "live" transmission via a satellite.

9.2. *Language Barriers*

The number of different languages employed in the major television areas places a serious restriction on the use of international television exchanges. Even at the present time, this factor results in a curtailment of the number of programmes which can be satisfactorily exchanged, and apart from programmes of international flavour, such as music and ballet, other programmes—even of the simple commentary type—require the use of a multiplicity of music lines, in order that separate language commentaries can be undertaken. The use of an electronic translator is currently being investigated in a number of countries, but it is unlikely that anything but a simple phonetic translation will be achieved in the foreseeable future. It is therefore envisaged that the language barrier will restrict appreciably "live" television exchanges for a very long time.

9.3. *Economics*

The cost of providing and maintaining a satellite system for television exchanges, would—if borne by the television organizations alone—be prohibitive, and could only be considered economic if provided from a satellite system whose prime function was connected with international telephony. Such a system could easily cost in the region of £40,000,000 to establish, and £20,000,000 per annum for maintenance and replacement, exclusive of research and development. It has been stated, by a number of well informed sources, that costs of this order could be incurred and still show a handsome profit for telephony traffic alone, the television transmissions being undertaken during the telephony "off-peak" periods. This would, however, mean that the television relays would of necessity have to take place during the night time period of one of the terminals. Such a system, therefore, would most probably require recording facilities at one end or the other, resulting in an additional cost in the television exchange. Therefore, for a television system to become a practical proposition—from an economic point of view—it will be necessary for a relatively low tariff to be set if the system is to compete with present-day methods.

9.4. *Geographical Time Differences*

As the bulk of the world's television coverage is widely displaced in longitude, there are wide time differences to consider, and these are serious obstacles for "live" television transmissions, as the peak programme period within one area can coincide with a night-time period within the network to be connected. In this situation, the video tape recording can compete satisfactorily with the satellite network, as such a recording can be flown to the required area, for re-transmission at a more suitable time, whereas the "live" programme via satellite, must again be recorded before final transmission, in which event the immediate nature of the system is of no advantage. To illustrate this point, a "live" transmission in New York at local peak viewing time of 8 p.m. would, if relayed "live" to London, coincide with 1 a.m. G.M.T.—hardly a peak viewing time! In reverse, of course, a programme at British peak viewing time of 8 p.m. would coincide with 3 p.m. New York time, where it would probably be recorded for showing in the peak viewing evening period.

A similar situation can exist even with countries situated closer together, and in television programme presentation departments, it is well known that changing time tables to satisfy other organizations is an extremely difficult business; in fact this problem has been experienced on a number of occasions, where the time difference has been as little as two hours.

10. Conclusion

An attempt has been shown in this paper to outline the problems associated with setting up a satellite system for television, and the most likely solution appears to rest with the Common Carrier telephony system, with the television requirements being satisfied during telephone traffic off-peak periods. It is not suggested that this facility will result in a large number of entertainment programme exchanges, but provided the rental per minute is reasonably small, there would undoubtedly be an increasing amount of traffic as the system becomes established. News exchanges and special events will obviously be routed via satellite system, but this form of business does not represent a very large use of the service.

The greatest opposition to the development of television relays—via satellite—is the existence of video tape recording, which allows of immediate replay, with excellent quality, after recording process. By this method recordings can be flown to most capital cities of the world, and be transmitted locally within a few hours of the original performance, if necessary. In many cases a few hours delay is unimportant, and in some cases desirable, for instance, recording a programme in London at 8 p.m., and flying by modern jet to New York, can result in the programme arriving, and being transmitted by 10 p.m. local time and, with the advent of the supersonic aircraft taking some three hours on the Transatlantic crossing, this situation will be even further improved.

11. Acknowledgment

The Author wishes to thank the Managing Director and Technical Controller of Associated Television Limited for permission to publish this Paper.

12. References

1. W. F. Hilton and S. R. Dauncey, "Communications Satellite Orbits". Paper presented at 11th Congress of International Astronautical Federation, Stockholm, August, 1960.
2. J. G. Chaffee, "Application of negative feedback to frequency modulation systems", *Proc. Inst. Radio Engrs*, **27**, p. 317–31, May, 1939.
3. R. Adler, "Parametric amplification of the fast electron wave", *Proc. Inst. Radio Engrs*, **46**, p. 1300, 1958.
4. "Hawker Siddeley's communications satellite", *Aeroplane & Astronautics*, 10th March, 1961.
5. D. G. King-Hele, "Satellites and Scientific Research". (Routledge & Kegan Paul, London 1960.) See also *R.C.A. Engineer*, **5**, No. 6, April-May, 1960.
6. "Television Camera Systems for Space Applications". R.C.A. Princeton, New Jersey.
7. W. E. Morrow, Jr., "Orbital Scatter Communications". Paper presented at 13th General Assembly of International Scientific Radio Union, London, 5th–15th September, 1960.

13. Appendix 1: Satellite Link Specification

VIDEO CHANNEL

Applied signal	1 V d.a.p.
Output signal	1 V d.a.p.

1. *Waveform distortion*
   Using sine-squared pulse and bar generator:
   K rating not exceeding 5% using 2T pulse
2. *Non-linear distortion*
   (a) Synchronizing pulse crushing
   Any level between zero and 0·7 V: Synch. pulse within +10% and −30% of nominal
   (b) Not in excess of 15% (picture non-linearity)
3. *Signal/noise ratio*
   Random noise 26 dB
   Periodic noise 36 dB (excluding hum)
   Hum 30 dB
4. *Gain stability*
   10 kc/s signal not varying greater than ±1 dB per hour
5. *Group delay*
   Less than ±0·05 μs between 200 kc/s and 3 Mc/s
   Phase-frequency not in excess of 2 deg between 50 c/s and 200 kc/s

AUDIO CHANNEL

Applied signal	1 mW
Output signal	1 mW

1. *Frequency/attenuation characteristic*
   ±2 dB 50 c/s to 8 kc/s
2. *Harmonic distortion*
   1000 c/s +10 dB (relative to zero level) 40 dB sep.
   100 c/s +10 dB (relative to zero level) 26 dB sep.
3. *Signal/noise ratio*
   40 dB
4. *Gain stability*
   1 kc/s signal not varying greater than ±1 dB during any one operational period

14. Appendix 2: Satellite Information

Launch date	Satellite	Orbit relative to Equator	Apogee and perigee (miles)	Satellite size	Satellite weight (lb)	Radio equipment	Transmitter frequencies (approx.)	Remarks
4 Oct. 1957	*Sputnik 1*	65°	360	Sphere 23″ dia.	184	2 transmitters	20 Mc/s 40 Mc/s	Equipment powered by chemical battery which discharged in 3 weeks
1 Feb. 1958	*Explorer 1*	33°	1593 222	Cylinder 80″ long 6″ dia.	31	2 transmitters	108 Mc/s	Designed to measure intensity of cosmic rays and measure internal temperature
17 March 1958	*Vanguard 1*	34·3°	2468 410	6·4″ dia.	3¼	2 transmitters	108 Mc/s	Power provided by solar cells
15 May 1958	*Sputnik 3*	65°	1164 136	Cone 141″ long 68″ dia.	2925	1 transmitter	20 Mc/s	Contained instruments to measure pressure and composition of the atmosphere
26 July 1958	*Explorer 4*	50°	1377 167	Cylinder 80″ long 6½″ dia.	38½	2 transmitters	108 Mc/s	Designed to measure corpuscular radiations
18 Dec. 1958	*Atlas-Score*	32·4°	925 119	1020″ long 120″ dia.	8700	4 transmitters	2–108 Mc/s 132 Mc/s 140 Mc/s	Intended to investigate the feasibility of radio relay space stations
17 Feb. 1959	*Vanguard 2*	33°	2065 350	Sphere 20″ dia.	21	2 transmitters	108 Mc/s	Carries banks of photoelectric cells to measure cloud cover
1 April 1960	*Tiros 1*	48·4°	450	Pill-box 42″ dia. 19″ high	270	2 transmitters	108 Mc/s 235 Mc/s	Includes television cameras and tape recorders
12 Aug. 1960	*Echo 1*	50°	1000	Sphere 100′ dia.	132	2 beacon transmitters	108 Mc/s	Plastic balloon coated with aluminium for reflecting radio signals
4 Oct. 1960	*Courier 1B*	28·3°	752 602	Sphere 51″ dia.	500	Includes 6 transmitters	108 Mc/s 1700 Mc/s 2300 Mc/s	Designed as communications relay with delayed re-transmission

15. Appendix 3: Major Standards

Country	Line standard	Field standard (fields/second)	Video bandwidth requirement (Mc/s)
U.K.	405	50	3 Mc/s
Continental (except France, etc.)	625	50	5 Mc/s
France	819	50	12 Mc/s
U.S.A./Canada	525	60	4·2 Mc/s
C.C.I.R. recommendation	625	50	6 Mc/s

Manuscript first received by the Institution on 18th March 1961 and in revised form on 1st June 1961 (Paper No. 676).

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".

PHASE AND FREQUENCY COMPARISON CIRCUIT

A new method is discussed in a recent German paper for producing a rectified voltage that is a function of the phase and frequency difference of two alternating voltages. It is intended in particular for the comparison of pulsed alternating voltages and is outstanding by very moderate complexity. This method uses phase comparison circuits and storage properties that satisfy the highest demands on quality, despite moderate complexity. The application in follow-up synchronization circuits of phase and frequency comparison circuits for the synchronization of an oscillator (e.g. a horizontal deflection oscillator in television receivers) offers the considerable advantage that a wide pull-in range is attained as well as excellent freedom from interference. With the adoption of conventional phase comparison circuits without frequency comparison these two properties cannot be reconciled so that the demand for a wide pull-in range must be sacrificed and control elements provided for manual tuning corrections. The phase and frequency comparison circuits, however, make possible a fully automatic synchronization with immunity to interference. The rectified voltage produced by the frequency comparison adjusts the tuning condition of the oscillator in a way that the difference between the oscillator and the synchronizing frequency becomes very low and capable of synchronizing the phase-dependent rectified voltage.

"A new phase and frequency comparison circuit", G.-G. Gassmann. *Archiv der Elektrischen Ubertragung*, **15**, No. 8, pp. 359–76, August 1961.

TIME-AMPLITUDE CONVERTER

To determine the mass of a particle, the usual practice in high energy physics is to measure its transit time. This measurement is not very simple, for it is easier to analyse an amplitude than a time. Thus, a time-amplitude converter has been designed at the Centre of Nuclear Studies Saclay. The converter uses a double control grid tube with a resolving time of 5×10^{-11} seconds, as measured with a pulse generator. Analysis of the response of a particle detector consisting of a scintillator and a photomultiplier, shows that a resolving time of 5×10^{-11} seconds can be attained. A time of this magnitude has in fact been experimentally achieved. The converter described has been used to study the transit time of particles in a secondary beam in the accelerator *Saturne*. The energy spectra of π mesons, protons, deutons, emitted by a polyethylene target under 1·4 and 2 GeV proton bombardment were measured by this method.

"Time-amplitude converter", M. Banner. *Acta Electronica*, **5**, No. 1, pp. 73–88, January 1961.

DISTRIBUTED AMPLIFIER TUBE

A wide-band distributed-amplifier tube in which the anode and grid form a double helix has been developed by a Japanese engineer. Part of the amplified signal on the anode line is fed back to the grid line loss in the high frequency. The bandwidth of amplification obtained with the experimental tube was from zero to 300 Mc/s. The theory of the design of the tube is described in this paper.

"Feedback type distributed amplifier tube", T. Kojima. *Review of the Electrical Communication Laboratory, NTT, (Tokyo)*, **9**, pp. 150–64, March-April 1961. (In English.)

SIMULATION OF PROPAGATION VARIATION

Reflection from neighbouring mountains may cause multiple f.m. transmission paths of different lengths between transmitting and receiving aerial and can give rise to troublesome distortion. An analysis of this phenomenon is given in a recent Dutch paper. Means of improving reception are the use of sharply directional aerial, increasing the bandwidth of limiter and discriminator, more rigorous limiting, and the use of a discriminator capable of handling signal ratios close to unity. An apparatus is described for simulating multipath effects in the laboratory, irrespective of terrain or reception conditions. It delivers two r.f. signals, one delayed and the other not, which are modulated in frequency by an audio signal (sine wave; music or speech). The time delay is continuously variable, and simulates a path difference up to 300 km.

"Multipath transmission effects in f.m. reception and their simulation in the laboratory", J. Foster. *Philips Technical Review*, **22**, pp. 393–402, No. 12, 1960–61. (In English.)

ELECTRONICS AND MAP-MAKING

Electronic techniques have contributed in recent years to the advances in map-making which have taken place in Canada. Successful distance measurements have been achieved by the use of shoran, tellurometer and geodimeter and other electronic techniques have been employed to assist the hydrographer and photogrammetrist. An account of this work is given in a paper from the National Research Council of Canada. Among other radio methods employed has been an airborne profile recorder operating on the same principles is the radar altimeter, the two range Decca navigator and a microwave position fixing system developed by the N.R.C. The two latter systems have been used primarily for hydrographic survey.

"Electronic aids in Canadian mapping", J. T. Henderson, *The Engineering Journal of Canada*, **44**, No. 8, pp. 58–65, August 1961.