

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."*

THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and Radio Engineers

VOLUME 35 No. 2

FEBRUARY 1968

Engineers in Hospitals

RECENT spectacular and highly-publicized operations in cardiac surgery carried out in South Africa and the United States have again demonstrated the importance of electronic and mechanical aids to the skills of the surgeon and physician. The advances made during the past decade in open-heart surgery, in kidney grafts and complicated brain operations, as well as in routine diagnosis and therapy, have only become possible through inter-disciplinary co-operation and the development of techniques of measurement and control of physiological functions in which the electronic engineer has played a prominent part. Some of the techniques developed by electronic engineers have been already the subject of meetings organized by the I.E.R.E., notably 'Electronic Instrumentation in Cardiac Surgery' (1961), a Symposium on 'The Brain' (1963) and 'Electronic Aids for the Handicapped' (1962 and 1966). A most topical and timely Collaborate Conference on 'Interdisciplinary Problems in Open-heart Surgery' will be held from 27th-29th March 1968 in Oxford.

Medical electronics is probably the branch of electronic engineering outside his own specialization in which the average engineer takes the greatest interest: perhaps because, potentially, we all have a rather personal interest! The recent decision of the Minister of Health to set up a Committee to consider the future of the hospital scientific and technical services of the British National Health Service is thus of concern to a much wider group than those relatively few electronic engineers who are professionally associated with the Health Service. The Committee, which is under the chairmanship of Sir Solly Zuckerman, invited the Institution, through its Medical and Biological Electronics Group, to give evidence and members of the Group Committee have now completed a detailed memorandum.

Nearly every field of endeavour which has evolved over a period of years provides room for criticism of its basic organization and the scientific and technical services in hospitals are no exception. Highly competent technical work is often hampered by lack of co-ordination and over-cumbersome administration and the memorandum makes concrete suggestions for increasing overall efficiency. For instance, in some hospitals, engineers and physicists who ought to be devoting their knowledge and experience to research and development in new and improved techniques become involved in maintenance and repair work. Clear division into two distinct departments is advocated at Regional Hospital level.

Again, research and development is at present carried out within hospitals by small and independent teams attached to different clinical departments, and a single 'Department of Physical Sciences' within each large hospital is regarded as being essential. The head of this Department should be of professional standing in physics or engineering and his department should be equal in status to the clinical departments. The memorandum discusses also the important matters of career structures within the hospital service, interchange of staff between departments of different hospitals, and educational programmes on bio-medical engineering in association with medical schools.

Although the Institution's opinion has been sought in this particular instance in connection with the British Hospital Service, many of the problems existing here are known to be as urgent elsewhere in the world. There has always been close international co-operation between workers in medical electronics and the International Federation for Medical Electronics, founded in 1958 by several national organizations including the I.E.R.E., was one of the first of its kind in any discipline. The wide dissemination of the information laid before the 'Zuckerman Committee', and particularly of its conclusions, will thus be awaited with considerable interest.

F. W. S.

CHANGE IN MEMBERSHIP DESIGNATIONS

Advice has now been received that on 20th December 1967 Her Majesty The Queen in Council allowed alterations to Articles 10 and 13 of the Institution's Charter. These Articles refer to the three Corporate Member grades which are now Honorary Fellow, Fellow, Member.

With immediate effect, all members who were originally elected to the grade of Member will be known as *Fellows* and must use the abbreviated designation 'F.I.E.R.E.'.

All members who were originally elected as Associate Members will be known as *Members* and must use the abbreviated designation 'M.I.E.R.E.'.

Alterations to the Institution's records and automatic addressing system will be effected as quickly as possible, but to avoid correspondence Corporate Members are assured that they are correct in using the above abbreviated designations, according to membership grade, for all appropriate professional references.

Approval has also been given to an alteration to the Bye-law of the Institution regarding election to the class of Associate so as to permit the election of senior technicians holding certain academic qualifications, e.g. Higher National Certificate, City and Guilds Certificates. An information pamphlet on the

requirements for election to Associate will be available from the Institution at the beginning of March.

Registration of Chartered Engineers

All Fellows and Members (formerly known as Members and Associate Members respectively) elected or transferred to Corporate Membership during the period from 4th August 1965 to 2nd November 1967 inclusive are advised that their names have now been recorded in the Registers of Chartered Engineers maintained by the Council of Engineering Institutions. All Corporate Members elected or transferred before 4th August 1965 will be aware that their names are already on the Registers.

Those elected or transferred to Corporate Membership after 2nd November 1967 will receive individual advice from the Institution when their names have been recorded. Inclusion in the Registers entitles Corporate Members to use the designation 'Chartered Engineer'. When used in addresses, etc., the abbreviated form 'C.Eng.' should follow honours and academic awards and precede Institution designation letters.

As announced in the January 1968 *Journal* (page 10), Certificates of Registration as a Chartered Engineer are now available for those who wish to have them. The cost is £3 and application forms may be obtained on request from the I.E.R.E., *not* from C.E.I.

INSTITUTION NOTICES

The Physics Exhibition 1968

The 52nd Physics Exhibition will be held from Monday, 11th to Thursday, 14th March, in the Great Hall of Alexandra Palace, Wood Green, London, N.22. Hours of opening are 10 a.m. to 6 p.m. (7.30 p.m. on Wednesday).

The Institute of Physics and The Physical Society have invited members of the I.E.R.E., as a sister institution, to attend the closed session (Monday morning, 11th March) and view the exhibits in greater comfort. Special admission tickets, which also admit to the open sessions, may be obtained on application to the Institution, 8-9 Bedford Square, London, W.C.1. (A stamped addressed envelope enclosed with applications will be appreciated.)

General tickets for the open sessions may be obtained on written request (preferably accompanied by a stamped addressed envelope) to the Exhibitions Officer of the I.P.P.S., 47 Belgrave Square, London, S.W.1.

Abstracting of Journal Papers

All papers and technical articles of full page length or over appearing in *The Radio and Electronic Engineer* are indexed in *British Technology Index* published by the Library Association, London. Major papers are listed or abstracted as appropriate in the following 'Science Abstracts' publications published by the Institution of Electrical Engineers: *Current Papers in Electrotechnology*, *Current Papers in Physics*, *Current Papers on Control*,

Physics Abstracts, *Electrical and Electronics Abstracts* and *Control Abstracts*.

Titles of all papers are also given in *Current Contents* (Institute for Scientific Information, Philadelphia) and abstracts of major papers are published in *Referativni Zhurnal* (Institute of Scientific Information, Academy of Sciences of the U.S.S.R.).

Abstracts also appear as appropriate in the following specialized abstracts journals: *Index Aeronauticus* (Journal of Aeronautical and Astronautical Abstracts, T.I.L.S., Ministry of Technology); *Computing Reviews* (Association for Computing Machinery, New York); *Acoustics Abstracts* (Multi-Science Publishing Co., Brentwood, Essex, England); *Solid State Abstracts Journal* (Cambridge Communications Corp., Cambridge, Mass.).

Index to Volume 34

The December issue completed Volume 34 of *The Radio and Electronic Engineer*. The Index for July to December 1967 is included with this issue of the *Journal*.

Correction

The following correction should be made to the Letter to the Editor by Dr. R. W. G. Haslett, printed in the December 1967 issue of *The Radio and Electronic Engineer*:

Page 384, col. 1, line 22, read Ref. 2 instead of Ref. 1.

System Engineering for Reliability and Ease of Maintenance

By

K. F. RANKIN†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.Prod.E.-I.E.E. Conference on the 'The Integration of Design and Production in the Electronics Industry' held in Nottingham on 10th-13th July 1967.

Summary: The paper discusses the lessons learned from several years of investigation into the performance of a large air traffic control data processing system. To achieve the twin requirements implicit in the title, the time between faults must be made as long as possible, whilst the time required to clear a fault must be reduced to the minimum. The practice of building up some degree of system redundancy with stand-by equipment is also one that can be employed to ensure that the occurrence of a fault does not cause an appreciable loss of system facilities.

The experimental air traffic control system, on which the original findings were based, illustrates the feasibility of designing a high standard of reliability into data processing equipment. Extensive data on the performance of the system have since been analysed and used to improve further the reliability factor.

1. Introduction

Designing for reliability and ease of maintenance covers the complete range of engineering from component selection to program writing.

The paper is based on work carried out during the design, development and production of an experimental system for data processing of air traffic control information. Feedback of information from the experimental system was used to improve the performance of the, later, fully operational system.

The vital features of a 'high-reliability' system are:

- (a) Low incidence of faults in the basic items liable to failure—soldered connections, electronic components etc.
- (b) Rapid detection, location and clearance of faults when they do occur.
- (c) Organization of the integral parts to give the desired level of availability in a system (by means of sectionalizing and the use of redundancy).
- (d) The feedback of fault data by a 'controlled maintenance operation'.

Each of these concepts will be considered in turn.

2. Component Selection and Tolerances

In the interests of low component fault rates the selection of quality components, although essential, is not in itself enough. Correct usage is needed which

† Systems Development Division, The Plessey Company Limited, Poole, Dorset.

necessitates a knowledge of the behaviour of the component as a function of the conditions of use. At the present time the designer draws on manufacturers' data, technical papers and laboratory tests for this information, and competent manufacturers have produced reliable electronic equipment on this basis, for some time past.

However, the acquisition of the data in this manner is not appropriate to a situation where reliability is moving out of the qualitative into the quantitative state and equipment manufacturers are being called upon to assure reliability, by test. A properly organized method of specification and quality assurance for professional components is required as an essential link in the reliability chain.

With regard to current equipment for data processing systems, the design is intended to be conducive to extremely low component failure rates, and various tests are carried out in order to assure the quality of the components. The need for very high orders of reliability has made the use of military quality-approved components, or those of similar standard, essential. Components are de-rated whenever such measures can be expected to improve the reliability. This de-rating effects a reduction of parameter drift with time, thereby decreasing the incidence of degradation failures. The effect of de-rating upon the incidence of catastrophic failure is less certain, since the nature of such failures is not well understood, but it is thought, in general, to reduce these also. Semiconductor devices are usually operated well within their maximum voltage and current ratings.

The design of the circuit elements is based on 'worst case' design philosophy and includes appropriate allowances for the component parametric changes which occur during life (e.g. 30% decrease in current gain is normally allowed for germanium alloy junction types of transistor) as well as taking into account the initial component tolerances.

In the interests of high element reliability, the temperature and relative humidity of the atmosphere in which the equipment will operate are controlled.

The design and manufacture of reliable electronic equipment does not necessarily call for components better than those made at the present time, although any improvement is of course of great value. As stated earlier, such design and manufacture can be accomplished, provided that usage appropriate to reliable operation is accorded to the basic components, both active and passive. For such usage to be generally possible, experience shows that even relatively enlightened documents, such as the government CV 7000 series specifications for semiconductor devices, are not adequate, although much good work has been made possible by their existence. The examination of possible means of improving the situation is therefore essential.

The advent of reliability as a quantitative design parameter emphasizes deficiencies not only in the life test data, but also in the specification of the initial parameters.

An essential point to appreciate is that the equipment manufacturer is likely to draw heavily on experience and prototype tests before accepting a contractual reliability index. The continuance of acceptable standards during the normal production output is therefore largely dependent on the basic components retaining their essential characteristics fairly closely. Improvement (e.g. stability) of life characteristics will of course be readily acceptable to the user, as will reduction in the standard deviation of initial parameters. Any appreciable change in the mean might present a definite hazard to the equipment designer, because certain secondary aspects of equipment performance may be adversely affected.

3. Modular Construction

Low component failure rates are not in themselves enough. For high availability, 'down time' must be kept short, and this requires rapid detection, location and rectification of faults.

To facilitate this a modular form of construction is used which permits the rapid replacement of faulty modules. Components, in various circuit configurations, are mounted on a printed wiring board (p.w.b.). These boards are gun-wrapped into plug-in units, 26 p.w.b.s per unit being the maximum number fitted.

Test points are provided on the front of units to facilitate testing and maintenance. Signal inputs and outputs enter the unit by 70-way plugs. The printed wiring boards can be quickly removed from the unit for inspection and repair by cutting the gun-wrapped connections between the board and unit field wiring. Incidentally, the number of different types of units employed has been limited, and whenever possible they contain a simple configuration of elements to facilitate fault finding. To reduce the number of spare units required, units of a similar type are designed to be completely interchangeable electrically and mechanically.

Units are grouped in cabinets according to circuit requirements, a maximum of 36 being mounted in any one cabinet. One of the problems in the design of large systems using transistors is the distribution of power at low voltages. Much work has been carried out with regard to this problem and the results of this are reflected in the low inductance distribution system which is incorporated into the cabinet. This distribution is a four-potential ring-main which supplies power to each plug-in unit on a cabinet shelf basis. The cabinet power circuits are kept to the front of the cabinet while the control signals are kept to the back. The whole power supply arrangement system is designed to reduce unwanted transients to a minimum and provide potentials which are within close tolerances at the elements. Power supplies from the central power distribution system enter the cabinet through a panel containing fuses and supervisory equipment.

The signal distribution methods have received similar detailed consideration and the methods employed here are such that relatively long interconnections between cabinets are possible. Strobing methods can be used to offer protection against spurious operation due to noise wherever this precaution is merited. All the input and output signals are connected via the links which are mounted in the bottom of the cabinet.

4. System Reliability

4.1. Minimal System

Consider a system comprised of n parts, each being essential to the useful functioning of the system. The arrangement is shown in Fig. 1.



Fig. 1. Minimal system.

This is a series probability situation and

$$Q = q_1 \times q_2 \times \dots \times q_r \dots \times q_n$$

where Q is the probability of success of the system and

q_1, q_2 , etc., are the probabilities of success of the various parts.

For the constant failure rate period the exponential reliability function can be used giving,

$$Q = \exp(-T/M)$$

where T is the period for which the system is required to be serviceable and M is the mean time between failures (m.t.b.f.) of the system. This is the exponential reliability equation commonly expressed as

$$R(t) = \exp(-t/m)$$

which defines the reliability of an item in which failures occur randomly (i.e. Poisson process) and is in fact a special case (zero failures) of the Poisson distribution.

Similarly, for the general system part r ,

$$q = \exp(-T/M_r)$$

where m_r is the m.t.b.f. of the r th part.

It is evident that

$$\frac{1}{M} = \frac{1}{m_1} + \frac{1}{m_2} + \dots + \frac{1}{m_r} \dots + \frac{1}{m_n}$$

For the special case where $q_r = q_1 = q_2$ etc. and $m_r = m_1 = m_2$ etc.,

$$M = \frac{m_r}{n}$$

For the situation where the failure rate is very low (i.e. $T \ll m$ and $T \ll m_r$) use can be made of an approximation to the value of $\exp(T/M)$ by the first two terms of the exponential series expansion

$$Q = \exp\left(-\frac{T}{M}\right) = 1 - \frac{T}{M} + \frac{1}{2!}\left(\frac{T}{M}\right)^2 - \frac{1}{3!}\left(\frac{T}{M}\right)^3 + \dots \approx 1 - \frac{T}{M}$$

Similarly,

$$q_r = 1 - \frac{T}{m_r}$$

Even with $t = 0.1m$, the error in making this approximation is less than 1%.

Now the sum of the probabilities of failure and success is always unity.

Therefore,

$$P = 1 - Q \approx \frac{T}{M}$$

and

$$p_r = 1 - q_r \approx \frac{T}{m_r}$$

where P is the probability of system failure in time

T and p_r is the probability of failure of the r th part in time T .

Hence

$$P = p_1 + p_2 + \dots + p_r + \dots + p_n$$

e.g.

$$m = \frac{1}{p}$$

In this minimal arrangement the percentage availability A of the system is given by

$$A = \frac{M-t}{M} \times 100$$

where M is the m.t.b.f. of the system and t is the time needed to restore the system to normal operation.

The acceptability of a minimal arrangement depends upon the particular situation in which it is used. Clearly it is not satisfactory where facilities are needed continuously for a long period of time (i.e. $T \ll m$) with a high probability of success. However, where a repair time of t or more can be allowed whenever necessary, perhaps subject to a reasonable value for A , then the arrangement will be acceptable.

For example, consider a system consisting of 4 parts having m.t.b.f. values of 500, 2000, 2000 and 500 hours, respectively, with a maximum repair time for any fault of 10 hours.

The overall m.t.b.f. is given by:

$$M = 1 / \left(\frac{1}{500} + \frac{1}{2000} + \frac{1}{2000} + \frac{1}{500} \right) = 200 \text{ hours}$$

and the availability is:

$$A = \frac{200-10}{200} \times 100\% = 95\%$$

Of course, if an availability of 95% was required, but a maximum out-of-service time for any particular fault is 2 hours, say, then this system would not be acceptable. However, if the parts are of modular design and spares can be held with reasonable economy the 2-hours limit might be met by replacing faulty modules by spares and repairing the faulty modules during system running time. Naturally, the provision of spares will not be effective in reducing fault clearance time unless the time taken to detect and locate a fault is sufficiently short.

The time needed to restore a faulty minimal system to normal working is seen to be sensitive to the 'maintenance philosophy' adopted, i.e. whether or not spare modules are used to clear a system fault, and the techniques used to detect and locate a fault.

Hence the total system repair time t is seen to be comprised of several parts, i.e.

$$t = t_D + t_L + t_R$$

where t_D is the time needed to detect that a fault has occurred, t_L is the time needed to locate the fault and t_C is the time needed to clear the fault from the system once it is found; t_C can be either an item repair time (i.e. no spare modules) or a faulty module replacement time.

4.2. Simple Redundancy

Consider a system comprised of n parts, any of which can satisfy the functional needs of the system, i.e. all must fail before the system fails and hence $n-1$ are redundant. No parts are repaired before system failure. The arrangement is shown in Fig. 2.

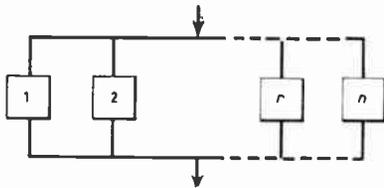


Fig. 2. Parallel reliability situation.

This is a parallel reliability situation and

$$P = p_1 \times p_2 \times \dots \times p_r \dots \times p_n$$

where P is the probability of failure of the system and p_1, p_2, \dots are the probabilities of failure of the various parts.

$$P = (1 - q_1) \times (1 - q_2) \times \dots \times (1 - q_r) \times \dots \times (1 - q_n)$$

where

$$q_r = \exp(T/M_r)$$

When all the parts are identical,

$$P = p_r^n$$

and, correspondingly,

$$P = (1 - q_r)^n$$

Hence,

$$Q = 1 - P = 1 - (1 - q_r)^n$$

where Q is the probability of success (i.e. the reliability).

For example, consider the simple case of triplication with no repairs before system failure. If $m = 50$

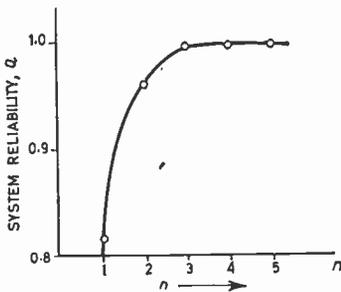


Fig. 3. Variation of system reliability against n , number of identical computers.

hours for each part and the mission time T is 10 hours,

$$Q = 1 - (1 - \exp(-10/50))^3 = 1 - 0.006 = 0.994$$

i.e. there is a 99.4% chance of completing 10 hours service time without failure of the system. The way in which the system reliability varies for values of n up to 5 is shown in Fig. 3. A law of diminishing returns is very evident for redundancy where each defective part is repaired at system failure only.

4.3. Redundancy with Repair

Although situations arise where redundancy without repair can happen, there are others—data processing complexes for ground use for example—where arrangements can be made to detect, locate and repair faults as they occur. This has important effects on the system reliability parameters.

An example of redundancy with repair is the ‘pool’ of similar computers which can be used in large data processing complexes. Consider an arrangement comprising n identical computers of which x are redundant and therefore $n-x$ computers are essential to provide full system facilities. The system interconnection arrangements are such that any $n-x$ computers will suffice in contrast to standard ‘stand-by’ provisions. Assuming that the maximum repair time for a faulty computer is t , the situation can be regarded, for simplicity, as a binomial distribution of failures occurring in a group of n computers in time t with expectation of failure p of any one computer in an infinite number of trials, and a corresponding expectation of success q .

The system will fail when at least $x+1$ computers fail within the maximum repair-time t , i.e. the system working computer complement falls below the $n-x$ computers necessary for full system operation, before a repair can be effected to restore the number available to $n-x$.

The expansion of $(q+p)^n$, taken term-by-term, expresses the probabilities of all the possible combinations of success and failure of the n computers in the ‘pool’ during a time interval t , e.g. the first term gives the probability of no computers failing, the second term gives the probability of one computer failing and the remainder succeeding, and so on.

Thus,

$$\begin{aligned} (q+p)^n &= q^n + nq^{n-1}p + \frac{n(n-1)}{2!}q^{n-2}p^2 + \dots + p^n \\ &= \sum_{r=0}^{r=n} \left[\frac{n!}{(n-r)! \cdot r!} \right] q^{n-r} p^r \\ &= \sum_{r=0}^{r=n} {}^n C_r q^{n-r} p^r. \end{aligned}$$

For system failure more than x computers must fail, therefore the probability of system failure is,

$$P_{(>x)} = 1 - (\text{probability of } x \text{ or less computers failing and } n-x \text{ or more succeeding})$$

$$= 1 - \sum_{r=0}^{r=x} {}^n C_r q^{n-r} p^r$$

Assuming that the times between failures of any one computer are exponentially distributed (i.e. the failure rate is constant), then the probability of success (i.e. of zero failures occurring) in a time interval t is,

$$q = \exp(-t/m)$$

where m is the m.t.b.f. of a single computer.

For the condition that $t \ll m$, the exponential series expansion can be used, giving

$$q = 1 - \frac{t}{m}$$

Hence, the probability p of failure in time interval t is given by

$$p = \frac{t}{m}$$

Substituting for q and p ,

$$P_{(>x)} = 1 - \sum_{r=0}^{r=x} {}^n C_r \left(1 - \frac{t}{m}\right)^{n-r} \left(\frac{t}{m}\right)^r$$

For unit time,

$$M = \frac{1}{P_{(>x)}}$$

Hence for time t ,

$$M = \frac{t}{P_{(>x)}}$$

For example, consider a 'pool' arrangement of four identical parts, one of which is redundant, each part having an m.t.b.f. of 100 hours and a maximum repair-time of 1 hour.

Hence $t \ll m$, therefore,

$$P_{(>x)} = 1 - [0.99^4 + 4(0.99)^3(0.01)]$$

$$= 1 - 0.9991 = 0.0009$$

From the previous equation,

$$M = \frac{10^4}{9} \approx 1100 \text{ hours}$$

4.3.1. Special cases

$x = 0$: For the special case of zero redundancy, ($x = 0$), the probability of the system failing is given by

$$P_{(>0)} = 1 - \left(1 - \frac{t}{m}\right)^n$$

and using the binomial series expansion with $t \ll m$,

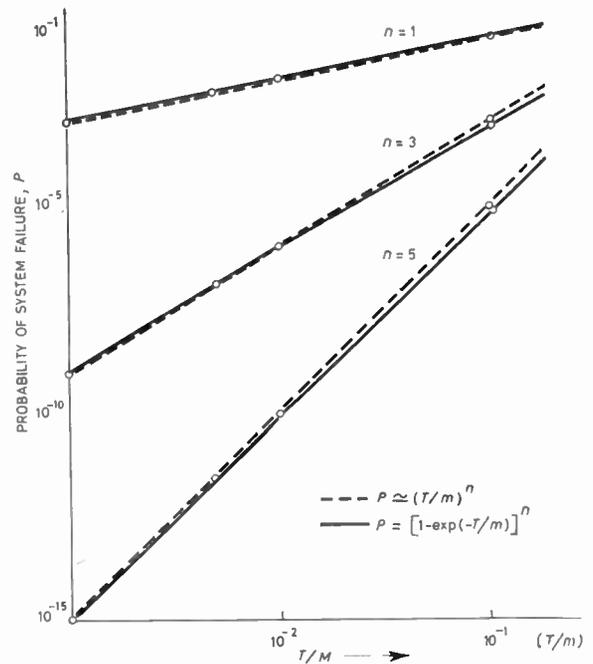


Fig. 4. Variation of the probability of failure with T/M .

$$P_{(>0)} = \frac{nt}{m}$$

and since

$$M = t/P_{(>0)}, \quad M = \frac{m}{n}$$

$x = n-1$: For this special case where all parts must fail in time t before the system fails, $P_{(>x)}$ is the last term of the binomial series.

$$P_{(>x)} = P_{(n)} = \sum_{r=n}^{r=n} {}^n C_r q^{n-r} p^r = p^n$$

Providing that $t \ll m$,

$$P_{(>x)} = p^n \approx \left(\frac{t}{m}\right)^n$$

and since

$$M = t/P_{(>0)}, \quad M = \frac{m^n}{t^{n-1}}$$

Figure 4 shows the variation of the probability of failure with values of T/M .

(Note that since $P = (T/M)^n$,

$$Q = 1 - P = 1 - (T/M)^n \approx \exp(T/M),$$

the failure pattern is approximated by the exponential reliability equation even in a case of simple redundancy when the above conditions apply.)

The 'pool' case is seen to contain the minimal series and simple parallel cases as special cases under certain conditions. It is seen to be very sensitive to the repair-time t .

In large systems where a multiplicity of equipments comprising standard modules are used, the repair-time t can be reduced very significantly by holding spare modules.

The t_c term is then the time needed to replace the faulty module with a spare module only and excludes the actual repair time of the module.

The system availability is given by

$$A = \frac{M-t}{M} \times 100\%$$

4.4. Off-line Data Processing

Consider a case where the overall commitment is heavy but no very precise terminal dates have to be met. Here high availability is needed but in the event of failure a fairly long repair time can be tolerated. A minimal system should be satisfactory here, no redundancy being provided. The essentials of such a system might be as shown in Fig. 5.

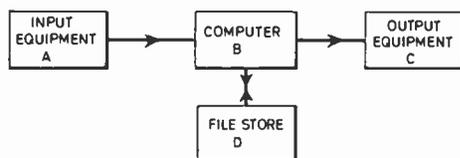


Fig. 5. A minimal system for off-line data processing.

This could be the example quoted in Section 3.1, with $M_A = 500$ hours, $M_B = 2000$ hours, $M_C = 500$ hours and $M_D = 2000$ hours, say, with a repair-time of 10 hours. The system m.t.b.f. is 200 hours and the availability is 95%.

Assuming that this availability figure is not high enough, it is clear that the situation can be improved either by reducing the repair-time or introducing redundancy at the input/output areas.

Suppose that equipments A and C are duplicated to give A_1, A_2 and C_1, C_2 . If an equipment fails it is of course repaired and therefore the risk concerned is that of both parts of a type failing in the repair-time. This can be treated as a case of simple redundancy with $T = 10$ hours, or as a special case ($x = n - 1$) of the 'pool' situation if t/m is sufficiently small. For the A and C equipments t/m is $10/500 = 0.02$, therefore the special case treatment can be used. Thus

$$m_A = \frac{m_{A_1}^n}{t^{n-1}} = \frac{500^2}{10} = 25\,000 \text{ hours}$$

Similarly, $M_D = 25\,000$ hours:

The new m.t.b.f. of the system is now

$$M = 1 / \left(\frac{2}{25\,000} + \frac{2}{2000} \right) = 925 \text{ hours}$$

The availability will now be

$$A = \frac{925-10}{925} \times 100\% = 99\%$$

Reducing the fault clearance time from 10 hours to 2 hours would achieve the same result and the provision of spare modules in a modular system should achieve this. The choice is a matter of economics.

4.5. On-line Control

An obvious example here is the automatic landing of an aeroplane. Here the period on risk is short and no repair is feasible. The situation in practice is one of simple redundancy. Since the risk-time is more or less constant the variables are the number of redundant parts and the basic reliability of each part.

Assume that the risk of failure must not be higher than 1 fault in 10^7 operations, the mission-time being 15 minutes, and that this will be achieved by the use of n identical equipments each having an m.t.b.f. of 100 hours,

$$\begin{aligned} n &= \log_{10} P / \log_{10} [1 - q] \\ &= \log P / \log [1 - \exp(-T/M)] \simeq \log P / \log \left(\frac{T}{M} \right) \\ &= \log 10^{-7} / \log 0.0025 = 2.69 \end{aligned}$$

Hence the commonly used triplication (i.e. $n = 3$) arrangement will be adequate.

4.6. Large Real-time Data Processing Systems

Consider a requirement for a large data processing system which must give continuous operation for long periods. Redundancy is provided and modular techniques are used to reduce the system repair-time.

A fundamental concept in the system with which the author is concerned is that of 'space-diversity'. By this is meant an arrangement in which, in general, main equipments are interconnected in such a way that failure of an equipment in a given functional block, such as the processing computers, will not render unusable any equipments in another related functional block, such as magnetic drum storage.

A frequently-used alternative system is that of the dual arrangement in which the basic system is virtually duplicated. If a failure occurs in a main equipment in either of the dual systems, the other equipments in the faulty system are not usable in the system complex. For similar amounts of system equipment, the dual system is always less reliable than the corresponding space-diversity system.

Figure 6 shows a hypothetical real-time data processing system of the space-diversity type designed to give a very high availability of a large proportion of the total facilities required.

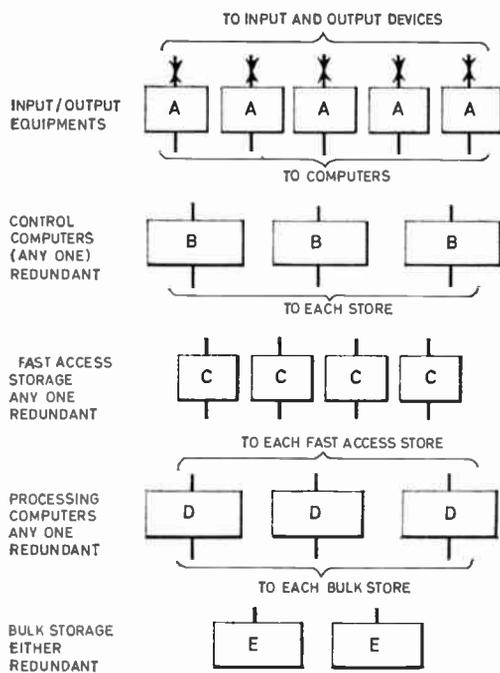


Fig. 6. A hypothetical real-time data processing system.

Each type A equipment comprises 3 sections giving a total of 15 sections in all and the situation is such that redundancy cannot be applied immediately in this area of the system. The failure of any of these blocks will therefore result in a reduction in the available facilities. Enough equipments are provided in each of the functions B, C, D, and E for the failure of one equipment per function to be acceptable without degradation of service. A loss of at least two equipments in a function is necessary for facilities to be lost at these points. File storage would be provided off-line.

Using the equations given earlier, the reliability of the system can be estimated from basic equipment data. The system repair-time t is far less than the equipment m.t.b.f.s.

The control computer, fast access store and process computer functions are all 'pool' type cases and the m.t.b.f.s M_B , M_C and M_D can be determined, redundancy x being 1 equipment in each case.

The m.t.b.f. M_E of the bulk storage function is also a 'pool' situation but best evaluated by the equations of Section 4.3.1 since the redundancy of one equipment makes this an $x = n - 1$ special case.

Regarding the loss of more than one section of the input-output equipments as a failure to provide a minimum acceptable facility level, the 'pool' case equations can be used with $n = 15$ $x = 1$. Basic equipment data are:

| | |
|--------------------------------------|------------------|
| Each A equipment, $M_A = 5000$ hours | } $t = 0.5$ hour |
| „ B „ , $M_B = 1100$ hours | |
| „ C „ , $M_C = 2500$ hours | |
| „ D „ , $M_D = 1100$ hours | |
| „ E „ , $M_E = 2000$ hours | |

Using these data, the mean time between instances when the facility provision falls below 93% is about 18.8 years while the availability of at least 93% of facilities is over 99.99%.

It is worth noting that where, as is frequently the case, the average data flow is less than the peak which the system is designed to accept, the loss of this proportion of facilities will usually cause no degradation of service in practice.

4.7. Main and Stand-by Equipment

In a large data handling system it is essential that the effect of an equipment fault should not result in an appreciable loss of system facilities. To avoid this, the designer splits up the equipment, particularly on a user basis, into convenient working units. For example, if there are 20 controllers in an air traffic control centre it would be desirable to divide the equipment into four groups. Each group would have its own storage and manipulation equipment and, if possible, its own power supplies. While there must be some equipment common to all four groups, this would be kept as simple as possible, even to the extent of somewhat elaborating the individual group equipment. It is often possible to have the common equipment in two parts, one of which could be a conventional stand-by equipment. Alternatively, each part can cater for half the groups under normal conditions with either capable of serving the whole centre if there is a fault on the other groups. This changeover can be carried out automatically as soon as a fault is detected. It is essential to use non-volatile stores, or where this is not possible, to duplicate equipment storing basic data which might be lost under fault conditions. Where information is regularly renewed, e.g. from a radar set, such duplication may not be necessary.

The system design should be such that it is possible for the duties of any common equipment to be taken over by another under fault conditions.

5. Maintainability

The measures taken to facilitate the maintenance of minimal and pool-type systems are extremely important since the reduction of system fault repair-time vitally affects the system reliability performance. The measures used to facilitate maintainability in the work with which the author is associated are briefly described here.

5.1. Fault Clearance

Fault clearance time, the term t_c , is reduced to a minimum by designing the individual equipments on a modular basis. The plug-in unit modules (called units) are designed so that those of a given type are mechanically and electrically interchangeable and require no adjustment after insertion into the equipment. As few different types of unit as possible are used so that the ratio of the total number to the number of different types in a system is kept as high as possible in the interests of an economic spares policy. Clearly the situation is at its best in large systems. In one large system, for instance, the total number of units is over 2500 while the number of different types is less than 70. (The best estimated m.t.b.f. of typical units is in the range 100 000 to 400 000 hours.)

However, before a spare module can be used to clear a fault from the system, the fault must be detected and located.

5.2. Fault Detection

Rapid fault detection is obviously needed if high system availability is to be realized. In order to achieve this automatic methods are essential. The design emphasis has therefore been to facilitate in-system testing using computer techniques.

'On-line' checking facilities are used extensively in all current system designs, and each equipment is carefully evaluated by the reliability group to ensure compatibility with the overall design policy, namely that no fault should arise and remain undetected by automatic means. The recommendations of this group are used by both system and equipment designers. Should an unsatisfactory report result from a maintainability assessment the design of the equipment is amended as far as is necessary to realize an acceptable standard.

'On-line' fault detection is achieved principally by:

- (a) routining,
- (b) internal check circuits.

Equipments are designed in such a manner that they can be checked under computer control using routiner detection techniques. This is much the preferred method of fault detection. Routiner-detected faults are those which are shown to exist by computer analysis of the result from a routining program. Computers are powerful aids to 'on-line' fault detection and every use is made of available programming capacity within them for system checking. This policy permits the simplification of the equipment which is in itself another factor in the attainment of high reliability.

Fault-detecting circuits are those auxiliary networks involved in the logical structure of the equipment to

detect failures of the main circuits. Parity-check circuits belong to this category.

With reference to system availability therefore, the adoption of on-line, in-system fault detection technique fulfills two important requirements:

- (1) The early warning of the presence of faults.
- (2) The indication of when the alternative measure built into the design of the system should be brought into use.

Wherever practicable, changeover to the alternative arrangement is automatic and computer controlled. Availability is thereby improved still further.

Any further interference to the system which may be caused by the faulty equipment, is prevented by isolating it prior to maintenance. Isolation is achieved at two levels: equipment and system.

For normal maintenance purposes, where a unit change clears the fault, isolation is effected at the equipment level. But where this action does not clear the fault, isolation at system level may be necessary. In this case a minor part of the system is isolated from the main equipment, thereby permitting the continuous routining of equipment under controlled conditions until the fault is precisely located, while the remainder of the system continues in normal operation.

In certain cases it may be necessary to use a 'test trolley' as an alternative to system isolation. Fault diagnostic routines are then injected from this source without involving the main system.

However, the facility of system isolation minimizes the necessity for special purpose test equipment, and promotes confidence in the fault identification because this is achieved under working-type conditions.

5.3. Fault Location

Fault location to an equipment level follows from the fault detection arrangements. The location of faults within equipment is also done by system computers which apply data patterns and test programs. Also used are the built-in fault circuits, the outputs of which are used by the routining computers in their analysis of the fault situation.

The fault is located to a 'fault area' which is less than five units in the majority of cases and is frequently one unit. In one arrangement which has been used the actual fault area is displayed as a light pattern (1 bulb per unit position) representing the equipment being tested. The maintenance aid used for this purpose is called a 'fault board'. More recently the fault areas have been printed-out after routining.

When the unit or units in the fault area are identified, they are replaced by spare units and the equip-

ment is routined to check that it is operating correctly before it is returned to the system for use.

When more than one unit comprises the fault area, the actual unit which is faulty is determined in the unit test area, a completely off-line function which also restores faulty units to proper working by identifying and replacing the faulty circuit. Since the printed boards which carry the circuits are gun-wrapped into the units, they are easy to replace.

5.4. System Control

The overall maintenance of a very large system is carried out by the 'system control' function. The object of this function is to keep the system at an optimum level of operation, and its organization is a subject in its own right, but nevertheless is a very legitimate part of system reliability philosophy.

The task of system control is aided by the provision of large mimic diagrams of the whole system, which enable the state of all the constituent equipment to be monitored.

6. Fault Recording and Data Analysis

The ultimate measure of reliability in terms of availability, maintainability and the mean time between failures, is in the field where the equipment is subjected to the actual conditions of use and operation for which it was designed.

Faults are cleared on faulty units by changing components rather than printed wiring boards. Complete boards are only replaced if they are damaged beyond repair.

These measures have made it possible for an accurate estimate to be made of the 'in service' time of each unit and consequently an accurate estimate of the 'component hours'. Environmental hazards associated with unit position are readily traced, and the history of any unit will be known accurately (Fig. 7).

The units are mounted in cabinets according to circuit requirements. An equipment is a number of cabinets connected together to form a functional entity.

6.1. Fault Recording System

The accuracy of fault data is often suspect. In this case the equipment is maintained by the manufacturers and rigorous control has been enforced in order to ensure correct records.

The key to the fault recording system is the 'incident sheet'. An incident sheet is initiated to cover any malfunction of the equipment, intentional or otherwise. For example, a power failure or a modification would be sufficient to call for an incident sheet.

Should a fault occur, the primary aim is to clear the system fault rapidly by use of automatically switched stand-by equipment, leaving the faulty equipment isolated from the system. The faulty equipment is subjected to diagnostic test routines which in conjunction with the built-in fault circuits locate the position of the fault. The units in the fault area are replaced with spares and the equipment retested to ensure that the fault has been cleared.

An incident sheet, with serial number, is completed to cover the fault. The serial number is used to identify the fault and the sheets are all pre-numbered

| Time & Date of Fault | | Time Fault Cleared in System | <u>ALL RELEVANT INFORMATION</u> |
|----------------------|----------|------------------------------|---------------------------------|
| Rack No. | Unit No. | Unit Designation | |
| Card Position | Card No. | Component | |

Fault or Symptom

Sig. of Engineer _____

Entered in Log by _____

Fig. 7. Typical history card for a unit.

| X TECHNICAL CONTROLLER | | | | | Q MAINTENANCE ENGINEER | | | | | |
|---------------------------------------------------------------------------------|-----------------------|----------------------|------------------------|----------------|------------------------|--------------------------------|---------------------------|-------------------|---------------------|--|
| INCIDENT TYPES: A Self-detected B Routiner C Others D Modifications | | | | | | | | | | |
| Equip. Locn. X | SFB No. X | Cabinet & Location X | SPARE UNITS Q | | | | | | | |
| | | | Locn. | Type | Ser. No. | MAINT. ENGINEER <u>RESULTS</u> | | | | |
| <u>SYMPTOMS X,Q</u> | | | | | | <u>RECORDS</u> | | | | |
| Incident No. X | Time/Date Initiated X | Type X | Technical Controller X | Time Started X | System Clear X | System X | Original Units Replaced Q | Incident Closed X | Unit Test Fault No. | |
| | | | | | | Usable Non-usable | | | | |

Fig. 8. Incident sheet.

to avoid the possibility of an incident number being used twice, and to ensure that the loss of an incident sheet would be noted.

The following details are recorded on the incident sheet (Fig. 8):

- (a) Date and time fault detected and date and time fault cleared from the system.
- (b) Equipment and cabinet location, unit(s) location, type and serial number and also serial numbers of spare units used.
- (c) Fault symptom—in system and equipment.
- (d) Type of fault (self-detected, routiner detected etc.).

The incident sheet is held in the data handling section until the units associated with the fault, tested and repaired where necessary, are replaced in

their original location. The equipment is then re-tested to ensure that the fault is still cleared. The incident sheet is then closed.

In order to provide fully comprehensive references concerning the system, and equipment affecting the system performance, the following documents are maintained:

- (a) History records for:
 - (i) Electronic equipments (identified by geographical location).
 - (ii) Spare units (identified by the serial number).
 - (iii) Power equipments (identified by geographical location).
 - (iv) Peripherals (identified by geographical location and/or manufacturer's serial number).

- (b) Modification records for:
 - (i) equipments, (ii) individual units.
- (c) Environmental log, in the form of temperature and humidity charts.

The history record for each electronic equipment is of particular importance. It contains details of all incidents affecting individual units and details of any occasions when a fault in the system failed to result in the location of a fault within an equipment. It also contains information such that the total number of hours run on an equipment at the time of an incident may be calculated.

All the information relevant to each fault occurrence is transferred to a punched card. This facilitates the use of automatic handling techniques and computer analysis. A punched card/computer system is also used to provide various other data including numbers of components in service, component hours and failure rates (Table 1).

The data are presented in various ways: failure rates, component histograms, unit p.w.b., equipment and system m.t.b.f.s etc. as required.

As a result of the analysis of data from the automatic routing equipment the presence of a number

of intermittent faults was deduced. Eventually the cause of these faults was found to be nine polystyrene film capacitors. Electrical measurements on the capacitors failed to reveal any defect, but physical examination, in this case by the manufacturer, showed that the weld between the lead-out wire and foil was absent in each capacitor. Further tests were carried out on samples obtained from stores and a number of potentially bad batches were found. This trouble was remedied by action at the manufacturing stage.

As a further example, some of the very first field returns for a germanium alloy junction transistor to a CV7000 series specification showed an unusually high fault rate. The failures were shown to be caused by excessive leakage current. The manufacturer's report showed that in each case the sealing was defective. Tests on unused devices were carried out on samples from batches with varying data codes. In some batches it was found that 5%–10% of the product was subject to this fault. Further investigations indicated that the fault was confined to components manufactured during a particular period when the sealing technique had been changed. Later devices were properly sealed, earlier difficulties with the process having been overcome.

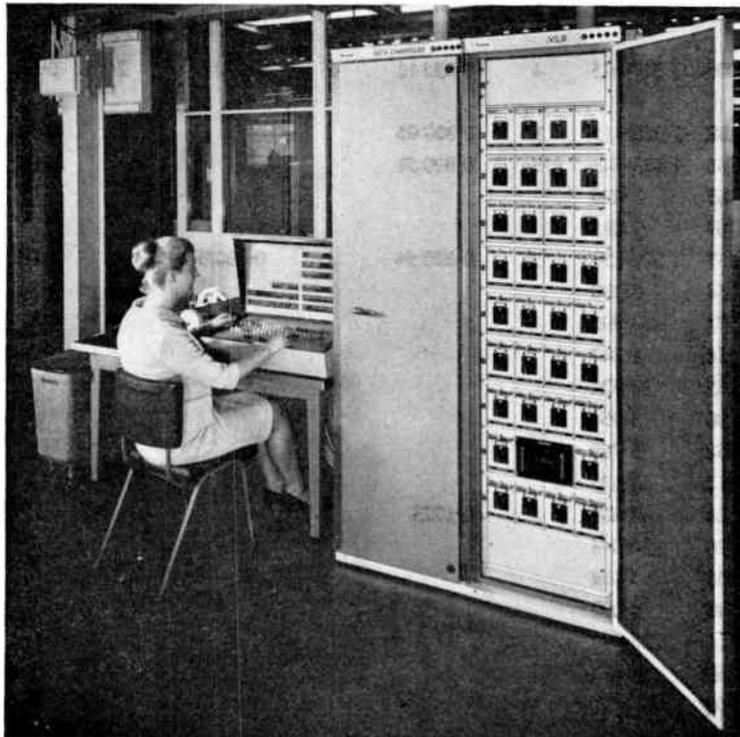


Fig. 9. The XL9 Computer system on test.

Table 1
Failures and failure rates

| Component Type | Component Totals | Component Hours ($\times 10^6$) | Catastrophic | | Degraded | | Accidental | | Other Failures | |
|--------------------|------------------|-----------------------------------|-----------------|-------------------------------|-----------------|-------------------------------|-----------------|-------------------------------|-----------------|-------------------------------|
| | | | No. of failures | Failure rate per 1000 hours % | No. of failures | Failure rate per 1000 hours % | No. of failures | Failure rate per 1000 hours % | No. of failures | Failure rate per 1000 hours % |
| TRANSISTORS | | | | | | | | | | |
| CV7042 | 59 017 | 1 320·164 | 15 | 0·0011 | 16 | 0·0012 | | | | |
| CV7007 | 4336 | 84·995 | 6 | 0·0071 | 1 | 0·0012 | | | | |
| CV7351 | 1354 | 30·006 | 2 | 0·0067 | | | | | | |
| CV7054 | 1270 | 26·920 | | | 2 | 0·0074 | | | | |
| CV7083 | 200 | 2·843 | | | | | | | 3 | 0·1055 |
| CV7087 | 28 442 | 644·240 | 8 | 0·0012 | 1 | 0·000 16 | 1 | 1 | 3 susp. sec. | 0·000 47 |
| CV7089 | 3120 | 67·624 | 2 | 0·0030 | 2 | 0·0030 | | | | |
| CV7184 | 1494 | 33·313 | | | | | | | | |
| CV7349 | 1144 | 24·979 | | | | | 2 | 0·0080 | | |
| CV7354 | 1690 | 34·548 | 1 | 0·0029 | | | | | | |
| CV7355 | 12 | 0·281 | | | | | | | | |
| CV7370 | 1092 | 20·062 | | | | | | | | |
| CV7348 | 64 | 1·183 | | | | | | | | |
| CV7350 | 310 | 5·040 | | | | | | | | |
| CV7376 | 240 | 3·150 | | | | | | | | |
| DIODES | | | | | | | | | | |
| CV 448 | 16 3064 | 3 610·095 | 4 | 0·000 11 | 2 | 0·000 06 | 2 | 0·000 06 | | |
| CV7130 | | | | | | | | | | |
| CV7047 | 88 982 | 1 980·915 | 1 | 0·000 05 | | | 5 | 0·000 25 | | |
| CV7127 | 54 765 | 1 173·643 | 6 | 0·000 51 | | | 1 | 0·000 09 | | |
| CV7102 | 65 | 1·436 | | | | | | | | |
| CV7104 | 64 | 1·411 | | | | | | | | |
| CV7049 | 58 837 | 1 304·859 | 7 | 0·000 54 | 1 | 0·000 08 | 3 | 0·000 23 | | |
| CV7138 | 154 | 3·388 | | | | | | | | |
| CV7106 | 2553 | 57·393 | | | | | | | | |
| CV7143 | 134 | 1·847 | | | | | | | | |
| CV7128 | 4552 | 83·524 | | | | | | | | |
| CV7171 | 180 | 3·954 | | | | | | | | |
| CV7174 | 252 | 5·740 | | | | | | | | |
| CV7177 | 212 | 4·305 | | | | | | | | |
| CV7169 | 26 | 0·564 | 1 | 0·1773 | | | | | | |
| HD1870 | 2720 | 50·044 | | | | | | | | |
| CV7172 | 248 | 4·032 | | | | | | | | |
| CV7173 | 248 | 4·032 | | | | | | | | |
| CV7364 | 1216 | 26·545 | | | | | | | | |
| CV7045 | 360 | 4·725 | | | | | | | | |
| CV7103 | 60 | 0·787 | | | | | | | | |
| CV7099 | 15 | 0·197 | | | | | | | | |
| CV7175 | 42 | 0·551 | | | | | | | | |
| CV7018 | 360 | 4·725 | 1 | 0·0212 | | | | | | |

Table 1 (contd.)

| Component Type | Component Totals | Component Hours ($\times 10^6$) | Catastrophic | | Degraded | | Accidental | | Other Failures | |
|-----------------------------|------------------|-----------------------------------|-----------------|-------------------------------|-----------------|-------------------------------|-----------------|-------------------------------|-----------------|-------------------------------|
| | | | No. of failures | Failure rate per 1000 hours % | No. of failures | Failure rate per 1000 hours % | No. of failures | Failure rate per 1000 hours % | No. of failures | Failure rate per 1000 hours % |
| POLYSTYRENE CAPACITORS | 10 9408 | 2 450·370 | | | | | | | | |
| STANTELAC CAPACITORS | 42 696 | 955·146 | 2 | 0·000 21 | | | | | | |
| METALLIZED PAPER CAPACITORS | 2832 | 62·370 | | | | | | | | |
| UNIT METAL PAPER CAPACITORS | 35 725 | 733·171 | | | | | | | | |
| B.M.E. CAPACITORS | 1608 | 32·645 | | | | | | | | |
| SILVERED MICA CAPACITORS | 2500 | 50·087 | | | | | | | | |
| GRADE I RESISTORS | 52 226711 | 600·882 | 2 | 0·000 02 | | | | | | |
| WIREWOUND RESISTORS | 6686 | 129·382 | 3 | 0·0023 | 1 | 0·000 77 | | | | |
| UNIT WIREWOUND RESISTORS | 7599 | 164·054 | 1 | 0·000 61 | | | | | | |
| METAL FILM RESISTORS | 7962 | 165·279 | | | | | | | | |
| METAL OXIDE RESISTORS | 8896 | 172·794 | | | | | | | | |
| POTENTIOMETERS | 463 | 8·232 | | | | | | | | |
| TRANSFORMERS | 7522 | 162·588 | | | | | | | | |
| CARDS | 42 453 | 934·902 | | | | | 5 | 0·00 53 | | |
| RELAYS | 4936 | 161·757 | 1 | 0·000 62 | | | 1 | 0·000 62 | | |
| PLUG AND SOCKET TAGS | 720 720 | 15 634·780 | | | | | 1 | 0·000 006 | | |
| PLUG AND SOCKET CONTACTS | 360 360 | 7 817·390 | 2 | 0·000 026 | | | | | | |
| SOLDERED CONNECTIONS | 3 079 690 | 66 608·358 | 9 | 0·000 014 | | | 4 | 0·000 006 | | |
| GUN-WRAPPED ENDS | 1 286 453 | 27 863·368 | | | | | | | | |

7. Conclusion

Although the paper has been based on larger systems, the knowledge and technique have been applied and used for small systems. Figure 9 shows an XL9 computer on test which with the system knowledge outlined has been used quite successfully in road traffic control, message switching, air traffic control and process control.

The value of accurate reliability data cannot be overestimated. Each failure in the system is a source of information. To be of value, however, the analysis must cover a period of several years during which the system has been functioning as a working entity.

Data properly recorded and used correctly indicate how closely equipment in use meets its reliability requirement in terms of availability and maintainability. This information is of considerable use to the design engineer working on new projects, and enables estimates to be made of the reliability of any new equipment at the design stage. It also makes possible the estimation of the maintenance procedure, personnel requirements and spare part requirements necessary to achieve any desired level of reliability.

8. Acknowledgments

The author wishes to thank the Management of the Plessey Company Limited for permission to present

this paper. He also wishes to acknowledge the help and co-operation of the Reliability Staff of Data Processing Division, Automation Group.

9. References and Bibliography

1. K. Blake and W. L. McLachlan, 'The Maintenance of Functional Availability in Large "On-line" Data Processing Systems', (Conference on the Impact of the Users' Needs on the Design of Data Processing Systems. Edinburgh, 1964).
2. P. Cox and V. J. McMullan, 'Reliability Feedback from Controlled Maintenance', (Conference on the Impact of the Users' Needs on the Design of Data Processing Systems. Edinburgh, 1964).
3. V. J. McMullan and P. Cox, 'The reliability of an experimental transistorized data handling system', *J. Brit. Instn Radio Engrs*, **22**, No. 1, pp. 17-29, July 1961.
4. P. Cox and V. J. McMullan, 'Reliability aspects of an experimental data processing system', *Electronics Reliability and Microminiaturization*, **2**, pp. 27-64, 1963. Also published, less appendix, in *A.T.E. J.*, **18**, No. 3, July 1962.)
5. 'Reliability of Military Electronic Equipment', Report by the Advisory Group on Reliability of Electronic Equipment, (AGREE). Office of U.S. Secretary of Defense, June 1957. (A condensed version of this report has been prepared by N. B. Griffin and is given in R.R.E. Memorandum No. 1934. 2nd issue, December 1962.)

Many text books and papers are available for study or reference among them being the following:

6. 'Reliability of Electronic Equipment', (Technical Committee Report), *J. Brit. Instn Radio Engrs*, **23**, No. 4, pp. 287-295, April 1962.
7. D. K. Lloyd and M. Lipow, 'Reliability, Management, Methods and Mathematics' (Prentice-Hall, London 1962).
8. S. R. Calabro, 'Reliability Principles and Practices' (McGraw-Hill, London 1962).
9. A. S. Goldman and T. B. Slattery, 'Maintainability' (John Wiley, London 1964).
10. 'Reliability Stress and Failure Rate Data for Electronic Equipment', MIL-HDBK-217A.

Manuscript received by the Institution on 28th September 1967. (Paper No. 1167/C102.)

© The Institution of Electronic and Radio Engineers, 1968

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10} , from nominal frequency for January 1968

| January 1968 | 24-hour mean centred on 0300 U.T. | | | January 1968 | 24-hour mean centred on 0300 U.T. | | |
|--------------|-----------------------------------|------------|-------------------|--------------|-----------------------------------|------------|-------------------|
| | GBR 16 kHz | MSF 60 kHz | Droitwich 200 kHz | | GBR 16 kHz | MSF 60 kHz | Droitwich 200 kHz |
| 1 | - 300.2 | 0 | 0 | 17 | - 299.9 | + 0.1 | 0 |
| 2 | - 300.0 | - 0.1 | + 0.1 | 18 | - 300.0 | - 0.2 | + 0.1 |
| 3 | - 300.0 | 0 | + 0.1 | 19 | - 300.0 | 0 | + 0.2 |
| 4 | - 300.2 | - 0.1 | + 0.1 | 20 | - 299.9 | 0 | + 0.2 |
| 5 | - 300.0 | - 0.1 | + 0.1 | 21 | - 300.1 | 0 | + 0.2 |
| 6 | - 300.1 | 0 | 0 | 22 | - 300.0 | 0 | + 0.2 |
| 7 | - 300.0 | 0 | + 0.1 | 23 | - 300.1 | 0 | + 0.2 |
| 8 | - 300.0 | 0 | + 0.1 | 24 | - 299.8 | 0 | + 0.2 |
| 9 | - 300.0 | 0 | 0 | 25 | - 300.1 | 0 | + 0.2 |
| 10 | - 300.0 | 0 | + 0.1 | 26 | - 300.1 | 0 | + 0.1 |
| 11 | - 300.1 | - 0.1 | 0 | 27 | - 300.1 | - 0.1 | + 0.2 |
| 12 | - 300.1 | 0 | + 0.1 | 28 | - 300.1 | - 0.1 | + 0.2 |
| 13 | - 300.2 | - 0.1 | + 0.1 | 29 | - 300.0 | 0 | + 0.2 |
| 14 | - 300.0 | 0 | 0 | 30 | - 300.0 | 0 | + 0.2 |
| 15 | - 300.2 | - 0.1 | 0 | 31 | - 300.0 | 0 | + 0.2 |
| 16 | - 300.1 | 0 | 0 | | | | |

Nominal frequency corresponds to a value of 9 192 631 770.0 Hz for the caesium F_m (4,0)-F_m (3,0) transition at zero field.

Notes: (1) All measurements were made in terms of H.P. Caesium Standard No. 134 which agrees with the NPL Caesium Standard to 1 part in 10^{11} .

(2) The offset value for 1968 will be -300 parts in 10^{10} from nominal frequency.

A High-speed TEM Junction Ferrite Modulator using a Wire Loop

By

J. HELSZAJN,

M.S.E.E., C.Eng., M.I.E.R.E.†

AND

M. E. HINES, B.S., M.S.‡

Summary: In this note a new TEM junction switch is described which is based on the three-port junction circulator. Microwave switching is achieved by applying a current waveform through a single wire loop inserted in each of the ferrite disks. In this way the demagnetizing fields of the ferrite shape and the eddy currents in the junction housing are eliminated. The magnetic energy is now determined by the inductance of a single wire turn and is extremely small. The switch can either be continuously modulated by a square current waveform or can be latched by short current pulses between the two remanence states of the closed ferrite shape.

In the three-port junction circulator the direction of circulation is dependent upon the sense of the applied transverse magnetic field.¹ This makes the junction useful as a prototype switching element. In the conventional approach the switch is operated by an electromagnet. The shortcomings of this technique are that the switching power is moderately large and that the switching speed is limited to about $10\ \mu\text{s}$ because of demagnetizing fields and eddy currents.² One way of removing these two limitations is to use a ferrite shape in the microwave circuit that can be biased with a single wire turn and which preferably has a square hysteresis loop.

The purpose of this note is to describe a new three-port TEM high-speed junction modulator which makes use of a single wire loop embedded in each of the two ferrite disks of the junction.§ Such an arrangement is shown in Fig. 1. The arrangement of the wire loop and the biasing magnetic field is shown in Fig. 2 for one of the ferrite disks. A waveguide junction latching switch has also recently been described.³

The behaviour of the junction can be understood with the help of Fig. 2. The perpendicular component of the magnetic field links downwards that part of the ferrite material which is outside the wire loop, and links upwards that part of the ferrite material which is inside the wire loop. Hence, the centre of the ferrite disk is biased in one sense and the outer part is biased in the opposite sense. A preferred direction of circulation results because the r.f. energy is not uniformly distributed over the whole disk cross-section. In fact, the r.f. magnetic field is circularly

polarized at the centre of the disk, and becomes more elliptically polarized as the radius increases. At the edge of the disk, it is linearly polarized.^{4,5}

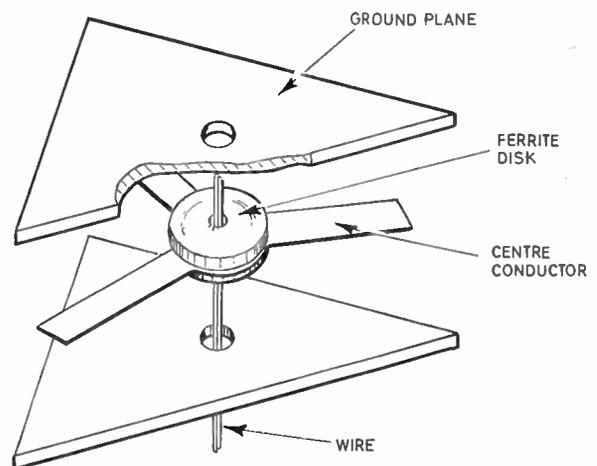


Fig. 1. High-speed TEM junction modulator.

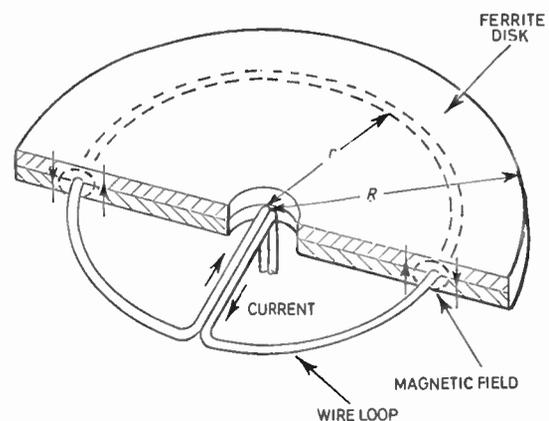


Fig. 2. Current and magnetic field of ferrite disk ($r/R = 0.65$).

† Department of Electrical and Electronic Engineering, The University of Leeds, Leeds 2; formerly with Microwave Associates Inc., Burlington, Mass., U.S.A.

‡ Microwave Associates Inc., Burlington, Mass., U.S.A.

§ U.S.A. patent filed.

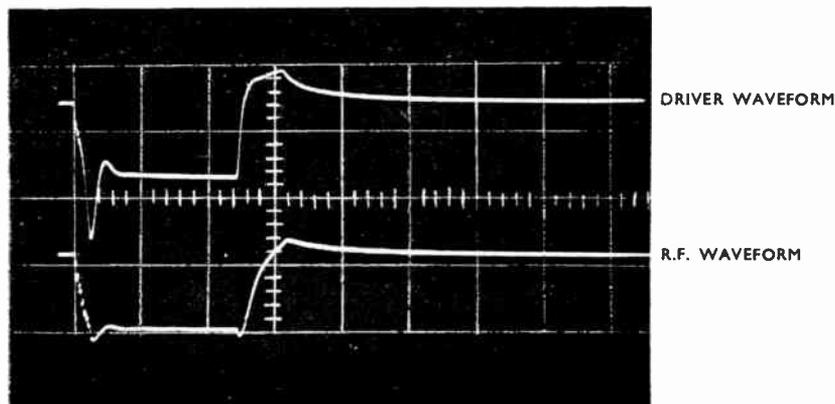


Fig. 3. Switching response of a 2450 MHz quarter-wave matched junction. (1 cm = 2.5 μ s)

The switch was operated in two ways. In the first mode of operation, the switch was continuously modulated by a square waveform. It is not necessary for the peak current to overcome the coercive force, and the current is therefore determined by the microwave circuit only, however a holding current is required. The switching response of a 2450 MHz quarter-wave matched junction is shown in Fig. 3. A rise-time of less than 500 ns was obtained using a 5 μ s pulse from a 4 watt driver. The peak current through the wires was 3 amperes. The calculated inductance of the wire loops agrees quite well with the measured one.

In the second mode of operation, the junction was biased between the two remanence states of the magnetic circuit by short positive and negative current pulses sufficiently large to overcome the coercive force. The advantage of this technique is of course that no holding current is required. In this mode, the peak current must correspond to 5 times the field of the coercive force. The switching energy is therefore about 10 times larger than for the first mode because the peak switching current in this instance must be at least 10 amperes.

The microwave adjustment of the switch was performed in the usual way, first by adjusting the ferrite diameter to satisfy the two degenerate electric field patterns of the junction. The resultant electric field pattern was then rotated through 30° with the

help of the current waveform.⁶ In both applications the isolation and insertion losses were typically 25 dB and $\frac{1}{2}$ dB respectively over a 4% band.

Acknowledgment

The authors wish to thank Microwave Associates, Inc., for permission to publish this paper.

References

1. U. Milano, J. H. Saunders and L. Davis, Jr., 'A Y-junction strip-line circulator', *Trans Inst. Radio Engrs on Microwave Theory and Techniques*, MTT-8, pp. 346-51, May 1960.
2. J. Helszajn, 'Switching criteria for waveguide ferrite devices', *The Radio and Electronic Engineer*, 30, pp. 289-96, November 1965.
3. P. C. Goodman, 'A latching ferrite junction circulator for phased array switching applications', Paper read at I.E.E.E. Professional Group on Microwave Theory and Techniques Symposium, 1965.
4. H. Bosma, 'On stripline Y-circulation at u.h.f.', *Trans. I.E.E.E.*, MTT-12, pp. 61-72, January 1964.
5. C. E. Fay and R. L. Comstock, 'Operation of the ferrite junction circulator', *Trans. I.E.E.E.*, MTT-13, pp. 15-27, January 1965.
6. J. Helszajn, 'A simplified theory of the three-port junction circulator', *The Radio and Electronic Engineer*, 33, pp. 283-8, May 1967.

Manuscript first received by the Institution on 2nd August 1966, in revised form on 7th January 1967 and in final form on 27th October 1967. (Contribution No. 100).

The Data Handling System for the *Ariel III* Satellite

By

A. LAWS†

Presented at a Symposium on 'The Ariel III Satellite' held in London on 13th October 1967.

Summary: The main features of the design and operation of the system are described. The techniques employed in the design and fabrication of the equipment are discussed in relation to the need for reliable operation in orbit, low power consumption and low weight. A brief account is given of the performance of prototype equipments during ground testing, and of the flight equipment since the launch of the satellite.

1. Introduction

The basic function of satellite data handling equipment is to commutate the various data outputs of the on-board sensors and to convert them to a form suitable for modulating the telemetry transmitter. The *Ariel III* telemetry system uses the same pulsed frequency modulation¹ (p.f.m.) technique as its forerunners, the *Ariel I* and *Ariel II* satellites, so that use can be made of existing ground-based receiving and data processing facilities. This paper discusses the essentials of p.f.m. telemetry, and its application in the *Ariel III* system. The techniques employed in the design and manufacture of the data handling equipment are also discussed, and a brief account is given of the performance of prototype equipments during ground testing, and of the flight equipment since the launch of the satellite.

2. P.F.M. Telemetry

Pulsed frequency modulation is a particular form of time division multiplexing in which the transmitter is modulated by a series of bursts of frequencies, the frequency in each burst representing a sample of the data at a particular sensor output. The commutation of these outputs is arranged in a repetitive sequence, some outputs being sampled more often than others, depending upon the bandwidths of the data at each particular output. Every sixteenth frequency burst contains alternately a synchronization or an identification frequency, so that when the data are processed, the position of each burst can be located in the sequence and hence related to a particular sensor output; the *Ariel III* high-speed telemetry format is given as an example in Fig. 1. To aid still further in the processing of the data, the blank period preceding each synchronization or identification burst is one-half the period of the blank preceding each data

burst; each synchronization and identification burst is made one-and-a-half times the period of a data burst, so that the same overall time is maintained for each blank-burst period. The frequency bursts are produced by gating on and off voltage-controlled oscillators, the inputs of which are either fixed voltage levels, chosen to produce the desired synchronization or identification frequencies, or analogue voltages from both the scientific experiments and the transducers monitoring the spacecraft performance. The voltage-controlled oscillators have a linear transfer characteristic producing output frequencies in the range 5 kHz to 15 kHz for input voltages in the range 0 to +5 V.

3. The *Ariel III* System

The system consists of a high-speed encoder, a low-speed encoder, a programmer and a tape recorder. The high-speed encoder commutates the data for real-time transmission at a rate of 55 samples per second, which is received on the ground while the satellite is making a pass over a ground station. There are however insufficient ground stations to provide world-wide coverage, and since some of the experimenters require continuous sampling of their sensor outputs throughout each orbit, their data are commutated by the low-speed encoder and stored on the tape recorder for replaying during a suitable pass. The programmer contains the circuits which control the flow of data from the encoders to the tape recorder and the telemetry transmitter; a simplified block diagram of the system is shown in Fig. 2. The tape recorder has a tape speed of $\frac{1}{4}$ inch per second (0.625 cm/s) in the record mode, and 12 inches per second (30.45 cm/s) in the replay mode, so that all the information stored during a 96-minute orbit can be transmitted in 2 minutes. Replay is initiated by command from the receiving ground station, while the replay period is controlled by timing circuits in the programmer. At the end of the replay period, the

† Ministry of Technology, Space Department, Royal Aircraft Establishment, Farnborough, Hampshire.

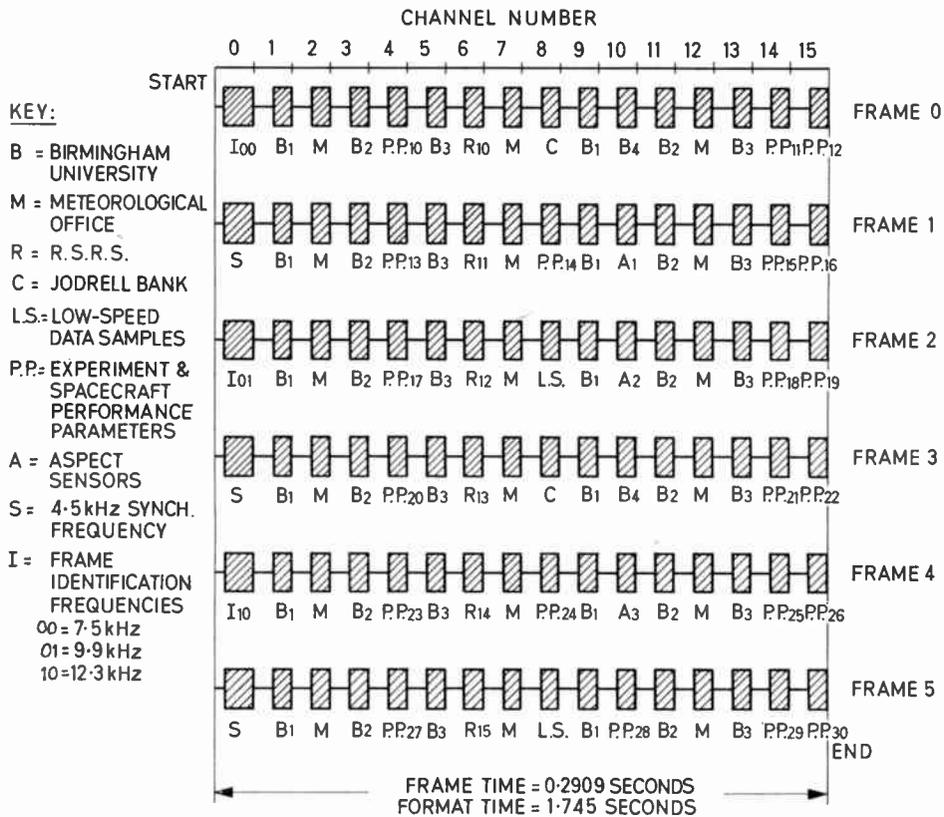


Fig. 1. Ariel III high-speed telemetry format.

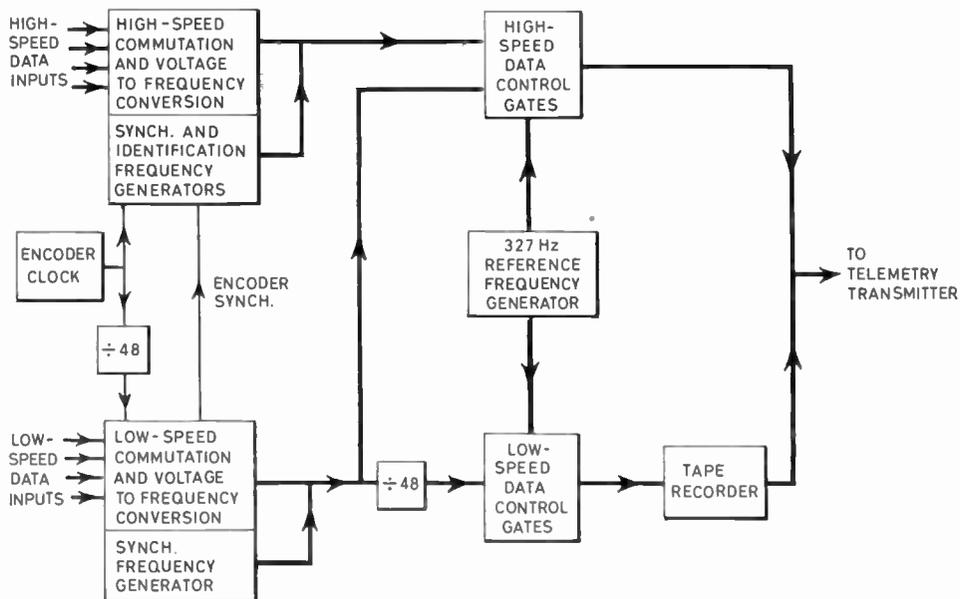


Fig. 2. Simplified block diagram of the data handling system.

recorder is switched back to the record mode and real-time data transmission is resumed. The data to be recorded are commutated by the low-speed encoder at $\frac{1}{48}$ the high-speed rate and the sub-carrier frequencies are divided by 48 before being recorded, so that during replay the information has the same commutation rate and bandwidth as that of the high-speed transmission. The 'blanks' in the low-speed transmission are filled in with a 15.7 kHz reference frequency, which is derived from a tuning fork oscillator included in the programmer. This reference frequency provides not only a means of identifying the low-speed transmission, but also a means of compensating for the errors introduced by variations in speed of the tape recorder. In response to a replay command, this reference frequency (divided by 48) is simultaneously transmitted to the ground station and recorded on the tape recorder for a period of 3.5 seconds. The low-frequency (327 Hz) transmission gives the receiving station operator an indication that the data system has been activated by the command, while the extended period of the recorded reference frequency provides a time marker to which the previously recorded data can be referred.

Four of the experiments require data storage, but since the total capacity of the tape recorder is needed to meet the data bandwidth requirements of the Jodrell Bank galactic noise experiment, the low-speed system has been arranged to have two modes of operation. In one mode, the tape recorder capacity is shared by three of the experiments and the Jodrell Bank experiment is switched off to minimize the hazard of r.f. interference. Every fifth orbit the system is switched to the other mode, when the Jodrell Bank experiment is switched on and only the data from this experiment are stored; at the end of this orbit the system automatically reverts to the first mode. Switching from one mode of operation to the other is effected by control circuits in the programmer, which count the number of commands received, since in general the tape recorder is commanded to replay only once per orbit.

The life of the tape recorder is governed largely by wear-out of the tape, and may not be much more than about 3 months.† A high-speed back-up of the low-speed data has therefore been incorporated into the system, which, if the tape recorder does fail, will provide a useful, if somewhat limited, supply of data

† 5.5.67. Satellite launched.

28.11.67. Tape recorder not working in record mode.

7.12.67. Tape recorder again functioning correctly.

22.12.67. Tape recorder not working in either record or playback mode.

6.1.68. Tape recorder again functioning correctly.

At present (end of January) still functioning correctly. The cause of the erratic behaviour has not yet been positively established.

to the experimenters. This facility provides for the insertion of frequency bursts corresponding to successive samples of the low-speed data in the high-speed sequence. This is arranged by synchronizing the high-speed and low-speed encoders, and gating the output of the low-speed sub-carrier oscillators (before frequency division) so as to provide every 48th burst of the high-speed sequence. This back-up facility has proved invaluable during satellite integration and check-out, since the low-speed data can be monitored by direct telemetry, thereby overcoming the delay which would otherwise be incurred by the necessity to record and replay the data.

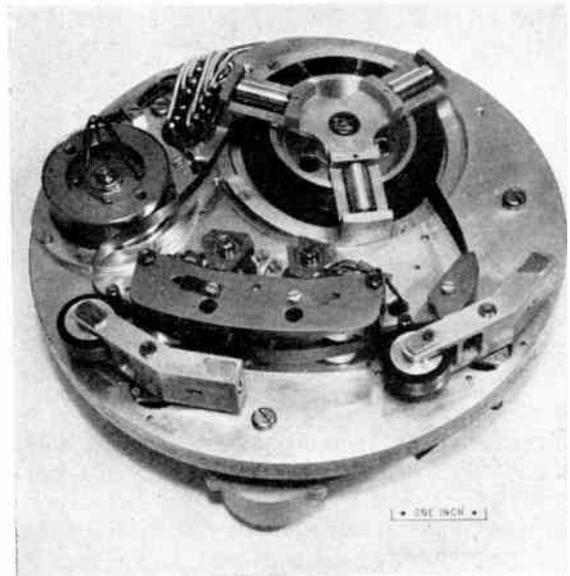


Fig. 3. The tape transport mechanism of the tape recorder.

The tape recorder uses an endless loop of standard $\frac{1}{4}$ -inch instrumentation tape which is drawn from the inside of the tape spool, passes across the replay, erase and record heads before being fed back on to the outside of the spool (Fig. 3). This mechanism involves a slipping action between adjacent layers of the tape on the spool, and therefore the tape is lubricated with graphite. For this to be effective, the mechanism must run in an atmosphere containing some moisture, and for this reason, the tape recorder is housed in an hermetically-sealed pressure vessel. The tension of the tape across the heads is maintained by having a 0.2% speed differential between the entry and exit capstans. These capstans are driven through a series of polyester film belts and pulleys, from a two-phase hysteresis motor. The change in tape-speed is effected by reversing the phases of the motor drive, the resulting change in direction of shaft rotation causing a clutch mechanism to operate, which in turn causes the drive to be transmitted through a different set of pulleys.

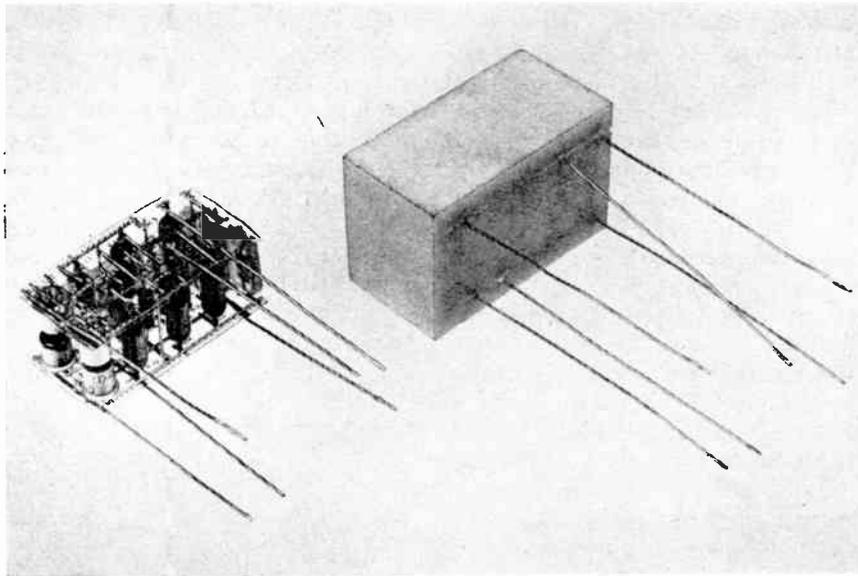


Fig. 4. Examples of electronic sub-modules, before and after encapsulation.

4. Design, Manufacture and Test of the Equipment

Most of the circuits in the data handling system perform switching functions and are therefore digital. In the interests of reliability, the number of types of logical element used in the system has been kept as small as is commensurate with other design constraints. All of these circuits have been designed so that they will still operate satisfactorily under the most pessimistic combination of component parameter variations caused by manufacturing tolerances, change over a year's life, temperature of the environment and level of the voltage rails. As a preliminary to the circuit design an evaluation was carried out of available components to select those types which were most likely to give reliable operation in circuits operating at very low current levels.² The number of different types of component has been minimized to reduce cost and to obtain better performance data. Since it was not possible to obtain the quantitative component failure rate data necessary to make a confident prediction of the reliability of a given system design, the criterion that a single failure would not cause a complete loss of all data has been used to set the minimum level of redundancy throughout the system. The circuit techniques employed have been chosen to minimize power dissipation, and it has been possible to design a system which uses 2771 electronic components, yet dissipates only 265 mW of power.

In order to minimize the weight of the electronic equipment, particularly in the amount of encapsulant,

the mechanical design is arranged to produce as high a component packing density as is consistent with ease of manufacture (approximately 10 000 components/ft³; 35 000/m³). To achieve this high packing density, the basic digital elements are constructed into sub-modules of the type shown in Fig. 4. The encapsulant used is a polyurethane foam, having a density of approximately 160 kg/m³ (10 lb/ft³). Resistance welding³ has been used as the method of interconnection throughout the equipment, and in the sub-module shown, the component leads are welded to nickel wire interconnection matrices which are secured to 0.010 mm (0.004 in) thick polyester films. Since the repair of these compact units is considered to be incompatible with the reliability requirements, the number of components in each is restricted. All of the components are aged for a minimum period of 200 hours before use to minimize the risk of having to carry out repairs. Each sub-module is arranged to form a basic logical element to facilitate electrical testing. Functional tests are carried out at room ambient temperatures prior to encapsulation to confirm correct assembly. After encapsulation, each unit is again tested at the specified temperature extremes, i.e. -15°C and +60°C. Groups of sub-modules are inter-connected by further nickel wire polyester film matrices to form modules, the grouping of the sub-modules being arranged so that each module can be separately electrically tested. The sub-modules are welded to sprigs, which in turn are welded to the module interconnection matrix, so that a sub-module can be removed and replaced without degrading the

potential reliability of the equipment. Each sub-system (high-speed encoder, low-speed encoder or programmer) consists of seven modules, which are interconnected by a separately encapsulated multi-layer nickel wire polyester film interconnection matrix, which also contains the plugs and sockets for external connections. Again, the sprigging technique is employed, so that the module can be removed and replaced should repair action be necessary. The performance of the complete system is fully tested at room ambient and at the temperature extremes, so that calibration data can be obtained. The completed units are secured to aluminium base plates (which include mounting lugs) before being finally encapsulated and having their protective aluminium covers fitted. After certain environmental testing and further electrical checks, each equipment is released for incorporation into a satellite.

During the project, six equipments have been manufactured for prototype testing and flight units. As soon as the prototype units had served their initial purposes, they were operated continuously for as long as available staff and facilities would allow. Although most of this operation was carried out at room ambient, one equipment was continuously thermally cycled for periods of 48 hours at either -15°C or $+60^{\circ}\text{C}$ for a total period of 8040 hours. Prior to launch of the flight unit, the six equipments had operated for a combined period of nearly 15 000 hours. Although component, manufacturing and design defects were detected during the production of the equipments, only one defective component has been detected during equipment operation.⁴ At the time of writing, the *Ariel III* satellite has been in orbit for more than six months, and apart from six occasions when short tape recorder replay periods have been reported, the system has performed very well. These shortened periods are believed to be due to the spurious triggering of bi-stables in the 2-minute timer caused by transients resulting from the phase switching of the tape recorder motor. It should be pointed out however, that well over 2000 successful replays have been obtained. The original design aim of three months'

operation for the low-speed system has already been surpassed, and in view of the history of successful operation of the equipments, both during ground testing and in orbit, there is every reason to believe that the remaining design aim, that is successful operation of the high-speed system for one year, will be achieved.

5. Conclusions

The main features of the design and manufacture of the data handling equipment for the *Ariel III* satellite have been described. The successful operation of prototype equipments during extended testing, and of the equipment in the flight satellite since launch, gives reason to believe that the remaining design aim of direct transmission for one year will be achieved.

6. Acknowledgments

The author wishes to acknowledge the work of his colleagues in the Data Handling Section of Space Department in the design and development of the system, the staff of the G.E.C. Applied Electronics Laboratories at Portsmouth who were responsible for the manufacture and test of the equipment, and the staff of A.W.R.E., Aldermaston, who were responsible for the manufacture and test of the tape recorder.

7. References

1. R. W. Rochelle, 'Pulse Frequency Modulation', NASA Technical Note D1421, Goddard Space Flight Center, October 1962.
2. A. Laws and J. K. Williams, 'The Evaluation, Testing and Ageing of some Electronic Components for the *UK-3* Satellite', R.A.E. Farnborough, Technical Report 66310, October 1966.
3. D. W. Allen and H. D. Fisher, 'Welded Components for Electronic Packages', R.A.E. Farnborough, Technical Note Space 70, August 1964.
4. A. Laws, D. Verrian and H. D. Fisher, 'The Data Handling Electronics for the *UK-3* Satellite', R.A.E. Farnborough, Technical Report. (To be issued.)

Manuscript first received by the Institution on 24th August 1967 and in final form on 23rd November 1967. (Paper No. 1168/AMMS 4.)

© The Institution of Electronic and Radio Engineers, 1968

The British Nuclear Energy Society

The British Nuclear Energy Society was established in 1962 in succession to the British Nuclear Energy Conference; its membership comprises eleven engineering and scientific institutions who have interests in some or all aspects of nuclear energy. The I.E.R.E. has been a member since shortly after the formation of the Society.

One of the most important roles of the Society is to provide a channel for the exchange of experience and the cross-fertilization of ideas between all engaged in this field. To achieve this aim, the Society has gradually expanded its programme from mainly evening meetings to several international conferences and symposia each year. During 1967 these have included meetings sponsored by the Society itself or through one or more of the constituent bodies on, for instance, 'The physics problems in thermal reactor design', 'The physics problems of reactor shielding', 'The applications of ionizing radiations in the chemical and allied industries', and 'Nuclear desalination'. In 1968, the programme will include a three-day conference in May on 'Steam generating and other heavy-water reactors', and on June 13th Lord Penney, F.R.S., will deliver the first of the Cockcroft Lectures which have been established as a tribute to a man who played such a major part in the development of atomic energy.

Lord Penney has taken as the title of his lecture 'Cockcroft and Atomic Energy'.

The Society publishes a quarterly Journal which contains original papers and discussions on a wide variety of subjects, and information and authoritative scientific and technological comments on developments in nuclear energy throughout the world. Bibliographies, lists of references and background information on selected topics, etc., and reports on the proceedings of the Society are also published.

Membership of the Society is open on application to members of all constituent bodies, and to all who satisfy the Board that they are 'actively engaged in the professional, scientific or technical aspects of the application of nuclear energy and ancillary subjects'. Membership is not confined to those of British nationality. The membership subscription, which includes copies of the quarterly journal and reduced registration fees for conferences, etc., is £4 4s. per annum for those over twenty-seven years of age. For those under twenty-seven (and for members who have retired) the subscription, as from 1st January 1968, is £1 10s. per annum.

The affairs of the Society are managed by a Board made up of representatives of each constituent body and of members elected by and from the general membership of the Society. The I.E.R.E. representative is Mr. R. J. Cox, B.Sc. (Member).

A Joint Technical Seminar in India

A Seminar on 'Electronics Industry in India' was held in Bangalore on 15th to 17th December 1967, under the joint auspices of the Institution of Engineers (India), The Institution of Telecommunication Engineers (India), I.E.R.E. (Bangalore Zone of the Indian Division) and I.E.E. (Bangalore Overseas Branch).

The Seminar, organized in the memory of Sir Mokshagundam Visvesvaraya, the eminent hydro-electric engineer and a former Dewan of Mysore State, was inaugurated by the Hon. Shri S. Nijalingappa, Chief Minister of Mysore.

The programme for the Seminar was divided into eight groups and in all some 55 papers were presented. Among the topics discussed were:

- (i) Design and Development of Electronic Equipment, including Reliability and Economic Aspects;
- (ii) Import Substitution in Electronics Industry;

- (iii) Scope for Development of Materials, Electron Devices, Components, etc.;
- (iv) Production Problems in Electronics Industry; Quality Control and Value Analysis;
- (v) Manpower: Education and Training;
- (vi) Research and Development;
- (vii) Evaluation, Testing and Test Facilities;
- (viii) Electronic Measurements and Measuring Instruments.

The authors of the papers presented were representatives of organizations such as Bharat Electronics Ltd., Indian Telephone Industries Ltd., and also the Indian Institute of Science. Several members of the I.E.R.E. presented papers at the Seminar and it is hoped that some of the contributions will, in due course, be submitted for publication in the *Proceedings* of the Indian Division of the I.E.R.E. or in *The Radio and Electronic Engineer*.

The Signal/Noise Ratio of Quantum Detectors in Coherent Light Systems

By

R. E. SCHEMEL,

B.Sc.(Eng.)†

Summary: The signal/noise ratio of a quantum detector is considered under conditions where background illumination is the dominant noise component. It is shown that this is a function of the optical (pre-detector) and post-detector bandwidths. To obtain as general a case as possible, the modulation (information) signal is assumed to be white-noise of a defined bandwidth.

Two separate methods of analysis are used. The first method assumes that the received light may be described in terms of its electromagnetic field while the other assumes that simple quantum theory is more appropriate. The first approach is valid for high levels of background radiation, the second for low levels. The overlap is shown to be in the region where the average quantum rate is just sufficient to define the pre-detector bandwidth—a result which is in agreement with sampling theory.

The results are shown to be applicable to superheterodyne and homodyne (synchronous detection) detectors by suitable parameter adjustment.

List of Principal Symbols

| | |
|------------|-------------------------------------------------------|
| P_n | noise power per unit area of photodetector surface |
| P_s | signal power per unit area of photodetector surface |
| Z_0 | free-space impedance |
| B | pre-detection (optical) bandwidth |
| b | post-detection bandwidth |
| d | small noise bandwidth of noise in B |
| E_n | equivalent peak electric field of noise bandwidth d |
| E_s | equivalent peak field of carrier |
| ω_n | lowest angular frequency of noise band |
| Ω | angular signal frequency ($= 2\pi f$) |
| $m(t)$ | normalized modulation function |
| Q | normalized power of $m(t)$ |
| c | bandwidth of $m(t)$ |
| x | small bandwidth of noise in c |
| I | detector output current |
| e | electronic charge |
| h | Planck's constant |
| G | image area |
| A | conversion constant of photodetector surface |
| α | quantum to electron conversion efficiency |

1. Introduction

The invention of the laser has suggested many new possibilities for communication and ranging. Current interest stems from the ease with which highly collimated spectrally pure beams of light may be generated. For the first time modulated optical signals may be produced with bandwidths approaching those used at radio frequencies, a fact which will enable optical systems with an improved signal/noise ratio to be obtained. To make use of such bandwidths effectively, superheterodyne or homodyne (synchronous detector) techniques will have to be employed,^{1,2} and whilst a quantum detector is capable of performing this—the subject is dealt with briefly at a later point—the most common method of use is as a simple energy detector. Most present day ranging and communication systems employ this method of reception almost exclusively.

A feature of a correlation type of detector such as the homodyne is that bandwidth is exchangeable between input and output. This means that the point at which selectivity is built into the system is unimportant. Quantum detectors, which are essentially square-law detectors, do not possess such advantages. At optical frequencies the best available filters might have bandwidths in excess of 300 GHz, and the external unwanted light sources falling in such a wide bandwidth could easily be much greater than the wanted signal. Nevertheless, as is shown in the analysis, the post-detection signal/noise ratio can still be positive under such conditions.

Signal/noise ratio is a quantity used widely at radio frequencies to define system quality. In the context of

† Formerly with the Future Projects Department of Hawker Siddeley Dynamics Ltd.; now with the European Space Research Organisation, Noordwijk, Holland.

communication it is often the modulation, rather than the actual carrier signal, in which the interest lies, and for this reason this paper derives signal/noise ratio in terms of the modulation. White noise has been chosen as the modulating signal because it is one of the most promising means of representing information. For example, multichannel telephony, or even several simultaneous modulating frequencies, are both well represented by white noise. For radar-like systems, where the modulation is not of such random form, the results are equally valid, and only need simple interpretation.

The analysis of signal/noise ratio has been presented in terms of two different concepts, wave theory and quantum theory. The first will be familiar to radio engineers, and is valid when signal levels are such that the quantum rate is high. The second approach is strictly valid at all signal levels. It is shown in the analysis that the two methods yield similar signal/noise ratios when the average quantum rate is equal to the optical bandwidth.

The situation where the quantum rate is sufficiently high for the light to be considered as having wave-like properties is discussed in Section 3. In Section 4 the case where the quantum rate is so low that the probability of arrival of an individual photon is independent of its neighbours is considered. This case has been considered in detail elsewhere.^{1,2}

2. Sources of Noise

Noise falling within an optical system may be divided into two distinct categories: (a) internal and (b) external. Noise in category (a) is partially under the control of the system designer, and obviously is minimized as far as possible at the design stage. Two such examples are thermal noise generated in the detector load resistance or in the following amplifiers, and noise due to detector dark current. The first type of noise may be reduced to negligible proportions by the use of photomultiplier types of detector (but at the expense of the reduced quantum efficiency of photo-emissive surfaces), the second type can be reduced by careful design of the detector, and finally by cooling. It is worth while pointing out that both types of noise can be made negligible by the use of homodyne or synchrodyne reception.^{1,2}

Category (b) contains those sources not immediately under the designer's control, and of these quantum noise^{1,2,6} forms the irreducible limit in an analogous manner to Johnson noise at radio frequencies.

The other important contribution, with which the present paper is concerned, is the background noise. This occurs by illumination from the Sun and stars, by scattering, and by reflection. Generally speaking it is much larger than quantum noise, and in modern

detectors is usually greater than the noise from sources in category (a). Present-day detectors have a much wider acceptance angle than is necessary (strictly, most light signals behave as a point source), and this again tends to ensure that systems are background noise limited.

3. Analysis of the Signal/Noise Ratio at High Quantum Rates

In this section the background light is assumed to possess the properties normally associated with radio waves, that is, its phase can be described accurately, and the power flux may be thought of directly in terms of an electric field related to it by the free space impedance. The background light is in itself treated as a band-limited (by the pre-detector filter) Gaussian process. The wanted signal is assumed to be monochromatic in character.

To calculate the detector output there are separate methods of approach. The background light is essentially a statistical process whose mean square value and spectral density is known. Application of the Wiener-Khintchine theorem shows that the Fourier transform of the spectral density yields the auto-correlation of the input random process. The input statistics are then known with enough details to specify the autocorrelation of the detector output. Inverse Fourier transformation will yield the wanted quantity, namely, the output spectral density. The method is elegant, but the fact that use is made of the spectral properties of the noise makes the second alternative attractive.

The second approach, the one which will be used in this section, is more intuitive. It is well known that the noise existing in a narrow bandwidth has many of the characteristics of a sine-wave at the band centre frequency. It is therefore allowable to represent a noise spectrum as the summation of a large number of discrete, but non-coherent, sine-waves, each spaced a small increment in frequency apart. The statistical justification of such a representation is given, for example, by Bennett.⁴

The background light (from now on this will be referred to simply as the noise) falling in the pre-detector bandwidth or the optical bandwidth B , is supposed to consist of a number of discrete sine-waves occupying a small band d . If P_n is the noise power density (the noise power per unit area present at the photodetector surface), then the noise in a small band d is $P_n d/B$ and the peak electric field will be:

$$E_n = (P_n Z_0 d/B)^{\frac{1}{2}}$$

The complete noise spectrum may now be treated as an assembly of sine-waves all having a peak amplitude E_n and may be written as:

$$I_p \sum_{k=0}^{B/d} \left(2P_n Z_0 \frac{d}{B} \right)^{\frac{1}{2}} \exp [j(\omega_a t + k dt + \phi_k)] \dots\dots(1)$$

where I_p denotes the imaginary part. The arbitrary phase constant is inserted to ensure incoherence between individual components. In a similar manner the signal power density may be interpreted as a peak electric field of:

$$E_s = (2P_s Z_0)^{\frac{1}{2}}$$

For the signal to transmit information in a form which the detector can make use of, this field must vary in amplitude, and this applies equally to cases where the carrier might be frequency modulated; in this case the variation would be brought about by a discriminator. We will assume that the amplitude of the carrier is modulated so that:

$$E_s = (2P_s Z_0)^{\frac{1}{2}} [1 + m(t)]$$

Provided that the signal and noise are both equally polarized, and this condition may be realized by a suitable filter, E_s and E_n may be added vectorially. Thus the resultant field is:

$$E = (2P_s Z_0)^{\frac{1}{2}} [1 + m(t)] \exp [j\Omega t] + \left(2P_n Z_0 \frac{d}{B} \right)^{\frac{1}{2}} \sum \exp [j(\omega_a t + \phi_k + k dt)] = (2P_s Z_0)^{\frac{1}{2}} \exp [j\Omega t] \left\{ 1 + m(t) + \left(\frac{P_n d}{P_s B} \right)^{\frac{1}{2}} \sum \exp j[(\omega_a - \Omega)t + k dt + \phi_k] \right\} \dots\dots(2)$$

The expression represents a constant carrier $\exp [j\Omega t]$ modulated in amplitude and phase by the curly bracketed term. However, the photo-sensitive surface is responsive only to fluctuations in the received power level, which is simply proportional to the square of the modulus of the bracketed term; thus the detector output current will be:

$$I = AP_s \left\{ 1 + 2m(t) + m^2(t) + 2 \sum_1 + 2m(t) \cdot \sum_1 + \left[\sum_1 \right]^2 + \left[\sum_2 \right]^2 \right\} \dots\dots(3)$$

where the summation signs have the following meaning:

$$\sum_1 = \sum_{k=0}^{B/d} \left[\frac{P_n d}{P_s B} \right]^{\frac{1}{2}} \cos (\omega_a - \Omega + kd)t + \phi_k \dots\dots(4)$$

$$\sum_2 = \sum_{k=0}^{B/d} \left[\frac{P_n d}{P_s B} \right]^{\frac{1}{2}} \sin (\omega_a - \Omega + kd)t + \phi_k \dots\dots(5)$$

Of the terms in eqn. (3), $2m(t)$ represents the wanted signal, $m^2(t)$ is a distorted term associated with square-law detection, and the next term represents noise

which is present because of a true product demodulation. The fifth term is present because of intermodulation between the modulation and noise, and gives rise to the characteristic increase in noise associated with the presence of a signal. This is sometimes referred to as 'noise in signal'. The last two terms represent non-linear actions on an assembly of sine-waves, and will generate difference tones of the constituent frequencies. Expansion and addition of these two summations gives:

$$\left[\sum_1 \right]^2 + \left[\sum_2 \right]^2 = \frac{P_n dB}{P_s B d} + \sum_{q=1}^{B/d} \sum_{k=2}^{B/d} \frac{2P_n d}{P_s B} \times \cos (q-k)dt + \phi \dots\dots(6)$$

where the phase term ϕ is retained to ensure non-coherence and where the double summation means taking every available value of k for each value of q (k to be greater than q).

The first term of this expression is a steady or d.c. value, which is made up of a total of B/d separate contributions, and when summed, the result is simply P_n/P_s , or a d.c. term proportional to noise input power.

Selecting only those components which contain the modulation or the noise and neglecting all other terms, the detector output current is now given by

$$I = 2AP_s m(t) + 2 \sum_{k=0}^{B/d} \left(\frac{P_n d}{P_s B} \right)^{\frac{1}{2}} AP_s \cos (\omega_a - \Omega + kd)t + \phi_k + 2m(t) \sum_{k=0}^{B/d} \left(\frac{P_n d}{P_s B} \right)^{\frac{1}{2}} AP_s \cos (\omega_a - \Omega + kd)t + \phi_k + \sum_{q=1}^{B/d} \sum_{k=2}^{B/d} 2AP_n \frac{d}{B} \cos (q-k)dt + \phi \dots\dots(7)$$

Taking the first noise term, (one which is identical with product demodulation), it can be seen that the original noise spectrum has simply become translated in frequency so as to extend over a frequency range of $-B/2$ to $+B/2$, in physical terms the negative sign meaning a 180° phase shift. This is shown more clearly in Figs. 1 and 2. The noise output power will be proportional to the square of r.m.s. current, so that the power per unit of d in a 1-ohm resistance will be:

$$\left[\frac{2}{\sqrt{2}} \left(\frac{2P_n d}{P_s B} \right)^{\frac{1}{2}} AP_s \right]^2 \dots\dots(8)$$

In all there will be $(2b/d)$ such contributions, the factor 2 arising because of the negative frequency terms; thus the total product demodulated noise is equal to

$$4P_n P_s \frac{b}{B} A^2, \dots\dots(9)$$

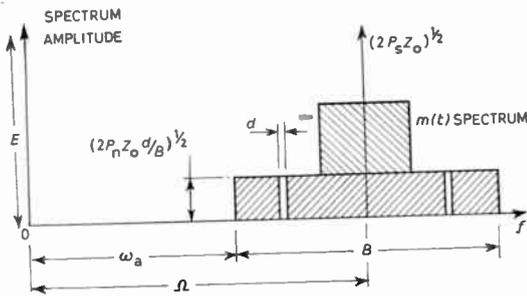


Fig. 1. Input spectrum of detector.

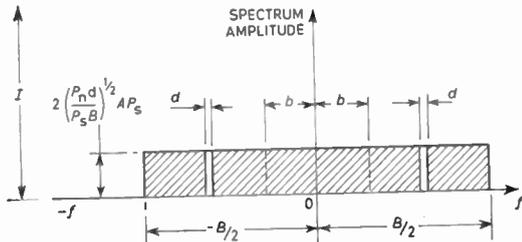


Fig. 2. Derivation of product demodulation term.

it being assumed that the video bandwidth b is less than $B/2$.

Although this derivation assumed a constant spectral density of the original noise signal over the entire range from ω_a to $\omega_a + B$, this need not be so. Product demodulation merely implies a frequency translation with all relative amplitudes preserved, the video spectral density therefore remaining unaltered from the corresponding optical spectrum. In all practical systems, however, the proportions of b and B are such that the video density is constant.

The next term to be dealt with is the double summation. Holding the value of q constant at 1, and allowing k to assume all values between 2 and B/d , will lay down a baseband spectrum extending from an infinitesimal step d , to $B-d$, and of component value $2AP_n d/B$ as shown in Fig. 3. q is now advanced to 2, and k takes on all values above 3. This action lays down a spectrum from d to $B-2d$. Advancing the value of q to 2 will lay down a spectrum from 0 to $B-3d$, and so on. Taking an interval d in any of the spectra, this has an r.m.s. current of $\sqrt{2}AP_n d/B$. At the lowest frequency, there is a total of B/d such currents, all randomly phased with respect to one another, and the number falls linearly to reach zero at a frequency B . Over a small interval $b, b \ll B$, the number of spectra are sensibly constant and so the total noise power will be:

$$2A^2 P_n^2 \frac{d^2 B b}{B^2 d d} = 2A^2 P_n^2 \frac{b}{B} \dots\dots(10)$$

The question of constant spectral density must again be considered. Clearly all parts of the optical spectrum contribute to the video noise, and in the derivation it was assumed that all such parts were equal in amplitude. Such an approximation is valid because every part is multiplied by every other, and as some parts are higher and the others lower than the mean value, a good deal of averaging will occur. Brockbank and Wass⁵ give a demonstration of this.

Depending on the type of modulation used, the modulation dependent term of eqn. (7) will vary. To obtain as general a picture as possible of this noise mechanism, $m(t)$ is assumed to possess band-limited white noise properties with the following characteristics:

$$m(t) = I_p \sum_{r=0}^{r=c/x} \left(2Q \frac{x}{c} \right)^{\frac{1}{2}} \exp j(rxt + \phi_r) \dots\dots(11)$$

where I_p again denotes that this is the imaginary part.

Figure 4 shows the noise and the signal $[2m(t)]$ spectra, with the noise spectrum folded over so that

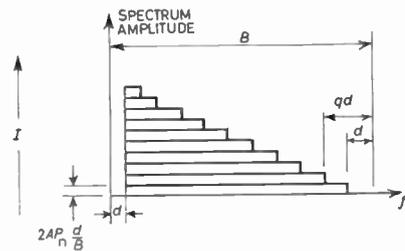


Fig. 3. Derivation of intermodulation term.

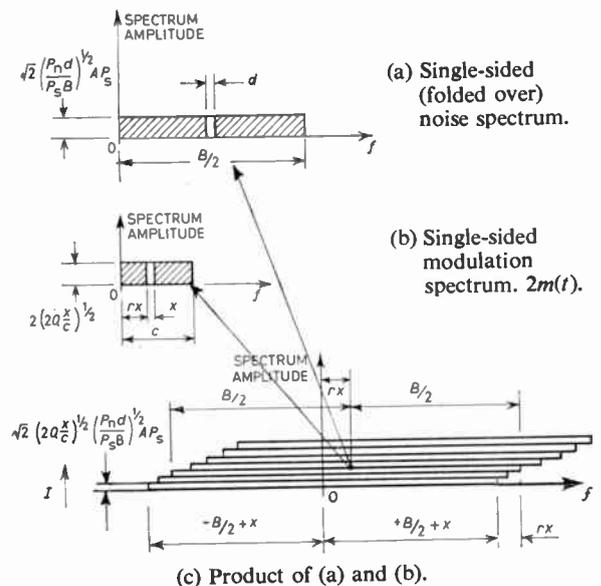


Fig. 4. Derivation of modulation-dependent term.

it lies entirely in the real frequency part of the graph. The spectrum is uncorrelated, and since the currents add power-wise, the effect of folding is to increase the peak current by an average amount $\sqrt{2}$. We now select a frequency from $m(t)$ and assume that this becomes a carrier for suppressed carrier modulation by the noise spectrum, for this is the physical representation of multiplication. Starting with $r = 1$ will lay down a spectrum extending from $(-B/2+x)$ to $(B/2+x)$, and each interval d of this will contain a peak noise current of

$$\sqrt{2} \left(\frac{P_n d}{P_s B} \right)^{\frac{1}{2}} A P_s \left(2Q \frac{x}{c} \right)^{\frac{1}{2}} \dots\dots(12)$$

Advancing r to 2 lays down a similar spectrum, and this may be repeated a total of c/x times. Thus in a band d the summated noise power becomes:

$$2 \frac{P_n d}{P_s B} A^2 P_s^2 2Q \frac{x}{c} \frac{c}{x} \dots\dots(13)$$

the factor 2 arising because of the negative frequency contributions. The effect of folding over negative frequencies makes the summated spectral density constant up to $B/2+c$, provided that $c < B/2$. It may be seen that only a small part of the optical spectrum, that extending to frequencies $b+c$ on either side of the carrier, contributes to the noise falling in the video band, and as the band is very small relatively, its spectral density will be constant. The total noise falling within the video band b will be b/d times that within the width d , giving:

$$4Q \frac{P_n}{P_s} A^2 P_s^2 \frac{b}{B} \dots\dots(14)$$

which, it will be noticed, is independent of the chosen bandwidth c . The form of $m(t)$ could therefore be sinusoidal and the generated noise term would be similar.

The remaining noise component to discuss is the term $m^2(t)$ in eqn. (3). Strictly speaking, this term does not fall within the context of background noise, and other forms of modulation, such as a sine-wave or a pulse train, will not generate it. However, when $m(t)$ is a white-noise process, the non-linearity generates an un-correlated noise spectrum whose density is $2Q^2 A^2 P_s^2 / c$ at low frequencies, falling linearly to zero at twice the highest modulation frequency c . Thus the total noise falling within a band equal to the original modulation signal is $3Q^2 A^2 P_s^2 / 2$. Because it is unrelated to background noise the derivation has not been given, but this follows the same method as is used for the noise in signal component in eqn. (7).

The wanted modulation current is $2AP_s m(t)$, and adopting a white-noise definition for $m(t)$, the total power in a 1 ohm resistance will be $4QA^2 P_s^2$.

Adding the contributions of eqns. (9), (10) and (14), and also including the intermodulation term due to $m^2(t)$, the signal/noise ratio becomes:

$$S/N = \frac{Q}{\frac{b}{B} \left[\frac{P_n}{P_s} + \frac{P_n}{P_s} Q + \frac{1}{2} \left(\frac{P_n}{P_s} \right)^2 \right] + \frac{3Q^2}{8}} \dots\dots(15a)$$

The case for the homodyne can now be considered. In homodyne detection a large, ideally infinite, carrier of identical phase to the signal is added at a point prior to detection. In expression (15a), this is equivalent to increasing the value of P_s without limit and reducing Q by an equal factor, the useful modulation signal remaining unaltered. Equation (15a) then reduces to:

$$S/N = \frac{BP_s}{bP_n} Q \dots\dots(15b)$$

It will be noted that all the non-linear contributions have disappeared, leaving only that part of the noise due to product demodulation. The homodyne is a special case of product detection, and as this can be regarded as a superheterodyne (with zero i.f.), eqn. (15b) is valid for this method of reception as well.

4. Quantum Effects

At low light-levels the form of analysis adopted in Section 3 is no longer valid, because the average quantum rate will be lower than the frequency, and the phase of a signal then becomes difficult to define. In this Section it is assumed that the arrival of a quantum is a random process which generates a noise spectrum at the detector output.

Each quantum on arrival at the photo-detector surface has a definite probability of releasing an electron; this is the quantum to electron efficiency. At low light-levels, the arrival of a quantum is supposed to be a random event whose statistics are time stationary. The release of electrons generates a mean current I at the detector output, associated with which is an r.m.s. noise current given by the well known shot noise formula;

$$i_n^2 = 2elb \dots\dots(16)$$

At high light-levels the quantum arrival time is no longer such an easily defined random process, but is one which is a function of the beating between the component frequencies of the applied spectrum. It is not time-stationary and eqn. (16) needs reinterpretation. However, this is the region already analysed in Section 3.

The steady current, or mean direct current, may be found from:

$$I = AP_s [1 + m(t)]^2 + AP_n$$

from which it will be noted that I is a function of the

modulation, and hence of time. It is not permissible to substitute this into eqn. (16) without further qualification, but time effects can be ignored if P_n is dominant, or if $m(t)$ is small, as it would be with a small modulation index.

The factor A may be expressed more fundamentally as:

$$A = \frac{\alpha e G}{hf}$$

where α is the quantum to electron efficiency and f is the light frequency. Over the range of bandwidths being considered, f will be a constant.

Accepting the same definition of the modulation as is used in Section 3, that $m(t)$ is a white-noise process, the wanted r.m.s. signal current will be $AP_s\sqrt{(4Q)}$, and on substitution for I in eqn. (16), this gives the signal/noise ratio as:

$$S/N = \frac{i_s^2}{i_n^2} = \frac{2\alpha GP_s^2 Q}{hfb(P_n + P_s)} \quad \dots\dots(17a)$$

where GP_s is the total power incident on the photo-detector surface. When the background noise is zero, this reduces to the quantum limit discussed in rather more detail by, for example, Oliver.⁶ The apparent difference between eqn. (17(a)) and Oliver's result is due to the definition of the modulation.

Homodyne reception is analysed, as before, by allowing P_s to assume arbitrarily large values, with a corresponding decrease in the value of the modulation power Q . This forces I to be independent of both $m(t)$ and P_n , and the signal/noise ratio becomes:

$$S/N = \frac{2\alpha QGP_s}{hfb} \quad \dots\dots(17b)$$

that is, the quantum noise approaches the fundamental limit, irrespective of the background power P_n .

Quantum noise will not be noticeable when the current given by eqn. (16) is less than the contributions described in Section 3. To define this region the ratio of the sum of eqns. (9) and (10) to (16) is taken, and the result gives the inequality

$$\frac{P_n \alpha G}{hfb} \left[1 + \frac{P_s}{P_n + P_s} \right] > 1 \quad \dots\dots(18)$$

Since $m(t)$ is assumed small in the preceding analysis, the contributions of the noise in signal term of eqn. (14), and the noise spectrum due to $m^2(t)$, have both been neglected.

In the case where the background power is dominant, the inequality gives a result which may be interpreted as a form of the sampling theorem. The effective average quantum rate is $\alpha GP_n/hf$, and to satisfy the inequality, this must be greater than the pre-detector bandwidth B . Now the background power may be considered to be a form of modulation of bandwidth

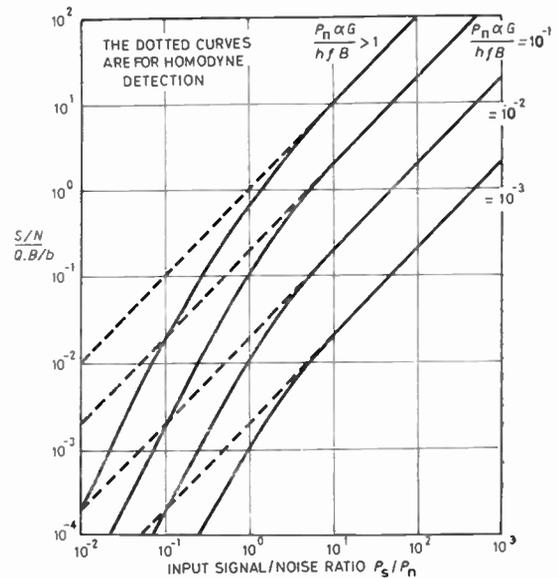


Fig. 5. Normalized signal/noise ratio at detector output as a function of P_s/P_n , with P_n as a parameter.

$\pm B/2$, centered on the carrier; the envelope of such a signal may be completely defined in terms of a unit impulse function which samples the waveform at a rate of at least twice the highest modulation frequency. In the context of the present analysis, the unit impulse function is represented by the arrival of any quantum which is usefully converted into an electron, and it is of interest to note that the sampling rate need only be sufficient to define the pre-detector band-width, rather than the signal itself.

When the signal is greater than the background noise, the detector virtually behaves as a homodyne, and in this case only the product demodulation term and shot noise due to the carrier are of significance. In this case, eqn. (18) shows that the background noise power spectral density, referred to a post-detection point, should be greater than hf , the fundamental quantum noise limit.

Shot noise is present at all input powers, and will be uncorrelated with the parts of the noise spectrum derived in Section 3. It will, therefore, add an extra term into the denominator of eqns. (15a) or (15b).

Figure 5 shows post-detection signal/noise ratio normalized with respect to Q and B/b . The background noise spectral density, $\alpha GP_n/B$, after allowing for conversion efficiency, is chosen as a parameter, and is normalized with respect to the fundamental quantum noise limit hf . The intermodulation limited region, as defined in Section 3, is represented by a single curve, whilst the shot-noise region is shown for several values of the parameter. Homodyne detection is shown by the dotted curves.

5. Discussion

Figure 5 shows that a photo-detector has a threshold region, defined roughly as the region where the signal and background noise powers are equal. Below this, the post-detection signal/noise ratio decreases as the square of the input, while above it this ratio varies linearly. Two distinct regions of operation are shown. One of these is represented by the parameter $P_n \alpha G / hf B$ being greater than unity. This is the region where shot noise is insignificant in comparison with the non-linear noise mechanisms given in Section 3. The second region is where shot noise is dominant and all other non-linear effects may be neglected.

Below threshold, the simple quantum detector has a considerable disadvantage as compared to the homodyne, and it is of interest to note that with the wide bandwidths of available filters, many optical systems will be operating in this region. However, above threshold, the homodyne has no great advantage, except that all non-linear mechanisms such as the noise in signal term of eqn. (14), and the extra noise due to the non-linear modulation term $m^2(t)$, are suppressed.

The threshold-free characteristics of the homodyne apply equally to the superheterodyne, where the dotted curves of Fig. 5 are equally valid if the final detection system is a linear one.

Inspection of Fig. 5 will show that the ratio of signal/background noise power is not the only parameter which determines the post detection signal/noise ratio; the signal can be completely obscured and yet be successfully received. In fact, the ratio of the pre-detection and post-detection bandwidth is equally important; if the ordinate and the abscissa of Fig. 5 are multiplied by B , it is seen that the essential parameters are really the pre-detection and post-detection noise spectral densities.

6. Conclusions

An optical detector working under background-noise-limited conditions has the post-detection signal/noise ratio determined not only by the pre-detection ratio of signal to background noise power, but by the ratio of the bandwidths at these two points as well. There are two distinct regions of operation; one region is where the background noise quantum rate is sufficiently high to define the modulation on the carrier that it may be assumed to cause and the second region is where such an assumption is not valid. In the first region the predominant contribution to the noise is made by demodulation of the noise envelope of the

carrier, whilst in the second region such effects are insignificant compared with shot noise. A simple quantum detector has a definite threshold below which there is a square-law relation between the pre-detection and post-detection signal/noise ratios. This is true in either of the two regions of operation just discussed.

Inspection of eqn. (15a), and the method used to derive eqn. (15b) for the homodyne, shows that this is essentially a process where the output signal is amplified without limit as the added carrier is increased. In this case, homodyne detection can be made to override all subsequent noise sources, with the result that only the background power spectral density is of importance, or with zero background power, the quantum limit is reached.¹ This fact, related with threshold free characteristics, makes the homodyne extremely attractive. Unfortunately, the optical homodyne is in its infancy, and the only practical detector at present is the simple quantum device.

Since completion of the work described in this paper, several papers dealing with various aspects of detection at radio frequencies have appeared. In particular, W. J. Lucas⁷ has anticipated part of the analysis of Section 3. However, it is believed that application to an *optical* detector is original, especially in the use of white-noise as an analytical representation of the modulation.

7. References

1. F. L. Warner and M. P. Warner, 'Superheterodyne reception at optical frequencies', 'Lasers and their Applications'. Joint I.E.E.-I.E.R.E.-I.E.E.E. Conference, 1964. (Paper No. 24).
2. S. Jacobs, 'The optical heterodyne: key to advanced signalling', *Electronics*, 36, No. 28, pp. 29-31, 12th July 1963.
3. J. C. Moden, 'Considerations affecting the design of short range single pulse ruby laser rangefinders', 'Lasers and their Applications' 1964, *ibid.* Paper No. 16.
4. W. R. Bennett, 'Methods of solving noise problems', *Proc. Inst. Radio Engrs*, 44, pp. 609-38, May 1956.
5. R. A. Brockbank and C. A. A. Wass, 'Non-linear distortion in transmission systems', *J. Instn Elect. Engrs*, Part III, 92, pp. 45-56, March 1945.
6. B. M. Oliver, 'Thermal and quantum noise', *Proc. Inst. Elect. Electronics Engrs*, 53, No. 5, pp. 436-54, May 1965.
7. W. J. Lucas, 'Tangential sensitivity of a detector video system with r.f. preamplification', *Proc. I.E.E.*, 113, No. 8, pp. 1321-30, August 1966.

Manuscript first received by the Institution on 31st March 1967 and in final form on 6th November 1967.

(Paper No. 1169/AMMS 5.)

© The Institution of Electronic and Radio Engineers 1968

Joint I.E.R.E.–I.E.E. Conference on 'THICK FILM TECHNOLOGY'

With the association of the International Society for Hybrid Microelectronics

Imperial College, London, Monday, 8th April and Tuesday, 9th April, 1968

OUTLINE PROGRAMME

Papers on the following subjects will be given:

Session 1—SUBSTRATES

Ceramic Considerations Controlling the Quality of Alumina Substrates.
Processing Ceramics to give Suitable Substrate Characteristics.
Glazed Ceramic Substrates of High Thermal Conductivity.

Session 2—MATERIALS, INKS AND DEPOSITION TECHNIQUES

Thick Film Tin Oxide Glaze for High Value, High Voltage Resistors.
Ruthenium Resistor Glazes for Thick Film Circuits.
Thallium Oxide Resistive Glazes.
Reliability Characteristics of Palladium-Silver Thick Film Resistors.
Capacitors Compatible with Thick Film Circuit Technology.
Glaze Capacitors for use in Thick-Film Circuits.
Performance Data on Thick Film Crossovers.
Thick Film Transistors.

Session 3—ASSEMBLY AND PACKAGING

Ultrasonic Bonding of Silicon Integrated Circuit Chips.
Active and Passive Components for Attachment to Thick Film Circuits.
Multilayer Thick Film Interconnection System.
High Density Soldered Interconnections.
Drag and Reflow Soldering Techniques for Mounting Flat-Pack Integrated Circuits on Thick Film Substrates.
Thick Film Interconnection Pattern for the Assembly of Silicon Integrated Circuits.
Packing Thick Film Circuits and Large-scale Hybrid Integration.

Session 4—APPLICATIONS

A Cermet Trimming Potentiometer.
Microelectronics in Automotive Equipment.

A full programme and synopses of the papers will be published in the March issue of *The Radio and Electronic Engineer*.
An application form for registration details, which are now available, is included in the back pages of this issue.

The Meteorological Office Experiment in *Ariel III*

By

P. J. L. WILDMAN, Ph.D.,†

G. P. CARRUTHERS†

AND

C. V. ELSE†

Presented at the Symposium on 'The Ariel III Satellite' held in London on 13th October 1967.

Summary: The experiment is designed to measure O₂ concentration from 150 km upwards with wide geographical coverage. The attenuation of solar ultra-violet light of 1425 Å to 1490 Å is observed as the satellite enters and leaves the Earth's shadow. The ultra-violet sensors used in the experiment are four gas-filled ionization chambers each of which produces about 10⁻¹¹ amperes in full sunlight above the molecular oxygen layer. These four chambers all feed into a single electrometer valve amplifier whose circuit and performance are outlined. The constraints on the layout and mechanical structure of the experiment are discussed and the final experiment package is described. An outline is given of the performance of the equipment in orbit.

1. Introduction

The aim of the Meteorological Office experiment in *Ariel III* is to determine the concentration of molecular oxygen (O₂) in the Earth's atmosphere at heights from 150 km to about 300 km. At these heights ultra-violet radiation from the Sun splits up the O₂ to form atomic oxygen as well as causing the ionization which is the basis of the ionosphere. These processes are of interest for their effects on the heat balance of the atmosphere both here and at lower levels, as well as for their importance to ionospheric physics.

A means of detecting an atmospheric constituent such as O₂ is to observe how it affects the penetration of light from the Sun into the atmosphere. In this case a waveband was chosen which is absorbed only by O₂, that is 1425 Å–1490 Å, and a detector, which is sensitive to this wavelength only, is mounted on the satellite. Figure 1 shows the principle of the *Ariel III* experiment, with the satellite just emerging into sunlight from the Earth's shadow. At this time the light from the Sun passes through the Earth's atmosphere, scanning through it as the satellite 'rises' into the sunlight. As the distance h increases, the quantity of O₂ between the satellite and the Sun is reduced until the Sun's ultra-violet radiation is unattenuated and 'full sun signal' is obtained from the detector. Thus a series of values is obtained from the detector for a series of values of h , and if a single wavelength is being dealt with then the following expression holds.

$$\frac{I_h}{I_0} = \exp(-\sigma_\lambda N_h) \quad \dots\dots(1)$$

where the symbols have the following meaning:

I_0 'full sun signal',

I_h signal corresponding to h ,

σ_λ absorption cross-section of O₂ at wavelength λ and

N_h total number of O₂ molecules between the satellite and the Sun.

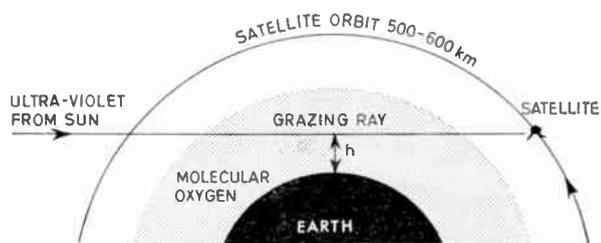


Fig. 1. Principle of the *Ariel III* Meteorological Office experiment.

By knowing the position of the satellite when the successive values of N_h are determined, the O₂ concentration per unit volume at the height h may be determined. Since at 1450 Å the absorption cross-section (σ_λ) of O₂ is at its maximum value, the experiment allows us to measure the smallest possible concentration of O₂ at very high altitudes. To detect the solar u.v. of 1450 Å, and to reject all other wavelengths, an ionization chamber is used into which the u.v. light enters through a sapphire window, and which is filled with an organic gas, para-xylene (C₆H₄(CH₃)₂). The bandwidth of this detector is set by the facts that sapphire does not transmit below 1425 Å, and that para-xylene is not ionized by light beyond 1490 Å. Since the value of σ_λ is almost constant between these wavelengths from the point

† High Atmosphere Research Unit, Meteorological Office, Bracknell, Berkshire.

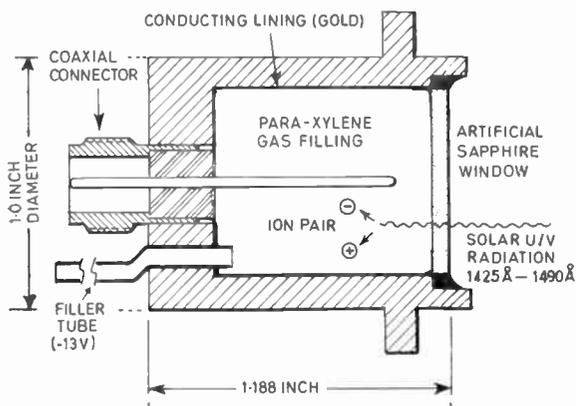


Fig. 2. Basic structure of a u.v. ionization chamber.

of view of eqn. (1), we are dealing with monochromatic light. Four such ionization chambers (Fig. 2), feeding a single d.c. amplifier, were mounted in the satellite unit to be described later. Unfortunately the para-xylene filling of these chambers is dissociated by u.v. from the Sun of about 1900 \AA so that the chamber sensitivity falls steadily with use. In *Ariel III* the 'full sun signals' had fallen to about 30% of their initial value after one month in orbit. The orbit requirements of other experiments on *Ariel III* meant that one week after launch the satellite orbit went into an 'all sun' condition where no useful results are obtained by the Meteorological Office experiment since the satellite never enters the Earth's shadow. At the end of this 'all sun' period the 'full sun signal' had reached the 30% value quoted, and has continued to fall since then.

2. The Experiment Package

The mechanical layout of the experiment package was influenced by the following points:

- (1) The necessity to expose the four ion chambers with an equatorial field of view of 90° each, to give an all-round view. The intersections of fields of view were arranged to be in line with the telemetry aerials.
- (2) The need to ensure that the $\pm 45^\circ$ longitudinal field of view did not include the top cone of the satellite.
- (3) The need to keep the centre of gravity low, to assist the spacecraft dynamics.

The package was to be mounted at the top of the conical section of the satellite for the best field of view, and the general layout is shown in Fig. 3, where the top cover of the package has been removed to show the four ionization chambers with the spaces between being filled with, clockwise from the top,

the amplifier, the polarizing batteries for the ion chambers, a junction box and the calibration resistor referred to later. The required fields of view were obtained using baffles (not shown) mounted on the outside of the package over each ionization chamber window.

It was decided to machine the package shell from solid aluminium bar, so that consistent and repeatable structures could be made and any changes that might be needed could be readily machined on later models. It also meant that the mechanical work could all be done at the Meteorological Office experimental workshops. Vibration and the thermal vacuum testing showed no need to change the initial design. The surface finish was matt gold to comply with spacecraft heat balance and electrical requirements. To ensure that a low resistance thermal path existed between the experiment and the satellite the wall thickness of the package shell was not less than 0.1 inches, and this also ensured that no large low-frequency resonances were introduced by the shell. The ion chambers, the amplifier, ion chamber polarizing batteries, and a junction box were all attached to this cylindrical wall, which was the stiffest part of the structure.

From a vibrational point of view the amplifier electrometer valve was the weakest component and it was mounted in an open cell polythene foam, the outer surfaces of which were sealed with silicone rubber. Many valves were vibrated in this form, those showing the least electrical change being selected as

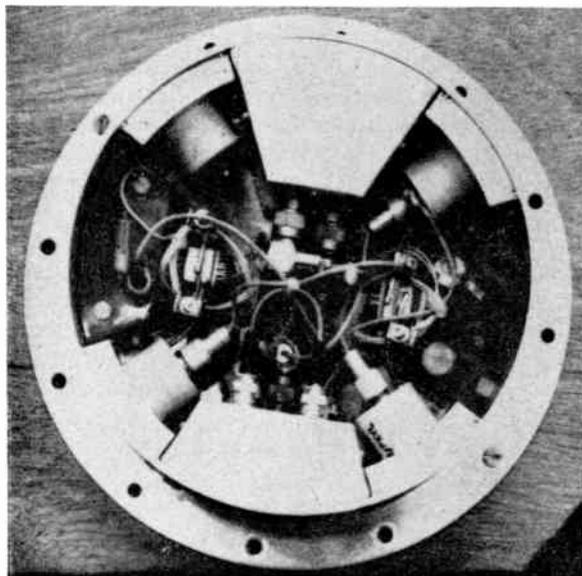


Fig. 3. The Meteorological Office package with top cover removed.

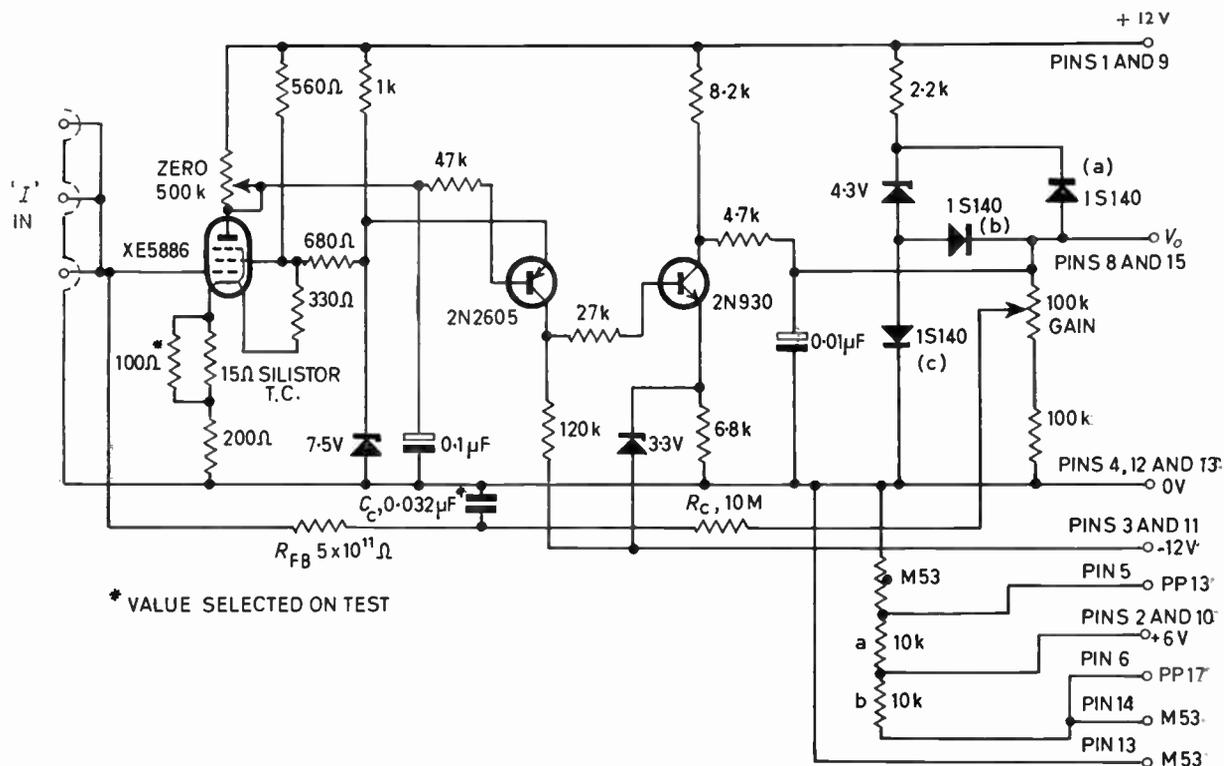


Fig. 4. Low-current electrometer amplifier.

potential flight valves. It was found that a coating of silicone rubber over all electrical components was sufficient to stop excessive movement. The two ion-chamber polarizing batteries were potted together under low vacuum in a resin of low exothermal temperature rise to reduce risk of damage to the mercury zinc cells used.

The wiring loom was made using p.t.f.e. insulated 7/0-0076 wire, and a system of redundancy was used such that experiment parameters were carried by at least two wires or pins within the package and by three pins at external connections. Power supplies were always carried by two wires or pins. The ion chamber polarizing batteries were each used with a 1 MΩ current limiting resistor, to guard against accidental short-circuits, and a diode, so that should the voltage of one battery become low the diode would cease to conduct thereby allowing the other battery to continue.

In addition to the ionization chamber amplifier outputs, temperatures at three points within the experiment package are telemetered back to the ground. These points are within the amplifier itself, on the inner surface of the cylindrical shell of the package, and on a wall of an ionization chamber. The temperatures are measured by a simple potential divider across the satellite +6 V supply, consisting of a 10 kΩ 1% metal oxide resistor of low temperature

coefficient and a thermistor. The thermistors were calibrated in the laboratory at several temperatures and the voltage output of the divider calculated including self-heating effects. The results were plotted to produce a smooth curve which was checked against the actual voltages produced at different temperatures by the dividers during the environmental testing of the satellite.

3. The Amplifier

The amplifier (Fig. 4) has three stages, d.c.-coupled with output limiting facilities and provision for powering the thermistor potential dividers. The amplifier stages comprise an electrometer valve input followed by two complementary transistors, the overall gain of the arrangement being controlled by applying negative feedback from output to input. The amount of feedback is variable from 100% to approximately 50% allowing the overall gain of the amplifier to be varied between $\times 1$ and $\times 2$ without upsetting the d.c. level balance of the amplifier as a whole. Phase compensation of the stray capacitance of the feedback resistor is provided to allow optimum response to a step input. Without feedback the input impedance of the amplifier is of the order of $10^{14} \Omega$, and the output impedance is about 12 kΩ. With feedback the input impedance is of the order of $10^8 \Omega$ and the output impedance about 1.2 Ω. The output capability is 0.6 mA into a 10 kΩ load.

The first stage uses an XE 5886 electrometer pentode operated directly from the 12 V line. Overall d.c. level adjustment of the amplifier is obtained by feeding the signal to the next stage from a 500 k Ω variable resistor as the anode load. This means that the overall loop gain of the amplifier changes as the d.c. level is changed, but the effect is not serious. The filament current for the valve, and the screen grid supply are derived from a resistor chain connected across the +12 V supply, and the filament is thereby run under virtually constant current conditions giving long filament emission life which is further extended by running the filament current at approximately 90% of the manufacturer's recommended rating. In order to provide stabilization against changes of d.c. output level if the +12 V line should change, the upper end of the screen and filament supply chain is fed partly from a resistive source and partly from a Zener stabilized source. Proper proportioning of the supply current from each of these sources provides screen voltage changes which cause compensation of anode voltage changes due to changes of the supply line voltage. Overall temperature compensation of changes of d.c. output level is also provided in this stage by the low-value positive temperature coefficient silistor in series with the earth return connection of the filament supply. The temperature variations in the resistive value of this component vary the effective grid bias on the valve and cause compensation of the d.c. levels. The required temperature coefficient is lower than that of the silistor itself, so the latter's effect is reduced by connecting a resistor in parallel with it. Exact compensation may then be achieved by varying the value of this resistor.

The second stage uses a 2N2605 p-n-p silicon planar transistor connected in a straightforward common-emitter configuration. In order to keep the d.c. levels through the amplifier correct the emitter is supplied from the Zener reference diode which also supplies the stabilized portion of the XE5886 valve filament current. The collector is returned to the -12 V line through a high-value load resistor, and the signal output from the first stage is applied to the base through a safety resistor of 47 k Ω which, in conjunction with the 0.1 μ F tantalum electrolytic capacitor and the anode load of the valve, provides a high-frequency roll-off characteristic, to ensure amplifier stability under 100% feedback conditions.

The third stage uses a 2N930 n-p-n silicon planar transistor which is approximately complementary in characteristics to the 2N2605 used in the second stage. Again the transistor is used in the common-emitter connection and correct d.c. levels result from the emitter being fed from a Zener stabilizer connected to the -12 V supply. A relatively low collector load is used to provide adequate power to the

specified external load of 10 k Ω and again the input signal from the collector of the 2N2605 is fed to the base of the 2N930 via a 27 k Ω safety resistor. This resistor and the 47 k Ω safety resistor in the second stage prevent excessive base current and transistor damage in the event of a shorted output or certain component failures.

From the 2N930 collector the output is taken via a 4.7 k Ω limiting resistor to the limiting and feedback networks. Positive limiting occurs at approximately 5.4 V due to the voltage across the 4.3 V Zener diode and the forward drops (approximately 0.6 V each) across the two 1S140 diodes (a and c), whilst negative limiting occurs at almost exactly zero volts, the forward drop across the 1S140 diode (b) being cancelled by the forward drop across the 1S140 diode (c). From this output point the external load is fed and also the feedback connection made so that the whole of the amplifier and its limiters are inside the feedback loop. This ensures a very low output impedance within the dynamic range of the amplifier, and a sharp limiting action. The feedback connection is taken from a 100 k Ω potentiometer in series with a 100 k Ω resistor connected between output and earth, allowing 100% feedback with an overall gain of 1 when the potentiometer slider is at the end connected to the output terminal, and 50% feedback with an overall gain of 2 when the slider is at the other end of its travel.

The conversion gain of the amplifier, in terms of volts out per unit current in, is set by the physical value of the feedback resistor, which in this case is two 10^{12} Ω resistors in parallel giving a sensitivity of 2×10^{-12} A in/V out and a full-scale output of 5 V for 1×10^{-11} A in. Due to the very high value of the feedback resistor, the effective value is reduced at a.c. due to the value of the impedance of the parallel stray capacitance of the feedback resistor. This reduction in the effective value of the feedback impedance progressively reduces the sensitivity of the amplifier at frequencies above zero and drastically increases the response-time to a step function. This situation is corrected by progressively reducing the amount of signal applied to the feedback resistor at frequencies above zero with a time-constant $R_c C_c$ equal to the time constant of the feedback resistor in parallel with its stray capacitance.† In practice this is adjusted by varying C_c for optimum response after construction and encapsulation are complete, whilst applying a repetitive current step to the input. By this means an initial response time of about 1 second can be reduced to a few tens of milliseconds, and the

† J. Praglin and W. A. Nichols, 'High-speed electrometers for rocket and satellite experiments', *Proc. Inst. Radio Engrs*, 48, No. 4, pp. 771-9, April 1960.

amplifier flown in *Ariel III* had a response-time of 50 ms with 5% overshoot.

Tests carried out on about 20 amplifiers show that without excessive adjustment one may expect a zero drift not exceeding ± 5 mV between -20°C and $+60^{\circ}\text{C}$, and a long-term zero drift, mainly due to decreasing filament emission, not exceeding 1.5 mV per week. Since the amplifier operates with approximately 80 dB of negative feedback, the linearity, slope and constancy of the input-output characteristic depend solely on the stability of the feedback resistors.

The power consumption of the amplifier is approximately 204 mW from the +12 V line, 36 mW from the -12 V line and 2 mW from the +6 V line. Of the power consumed from the +12 V line, about 180 mW is accounted for by the power needed to supply the valve filament and the Zener circuit used to stabilize the output against line voltage changes.

The amplifiers are constructed on printed circuit boards made from glass epoxy laminate board, using conventional soldering techniques. The valve is encapsulated in plastic foam to protect it against vibration, and for the same reason a thick layer of silicone rubber encapsulation is put over all components except the live grid, feedback and input connections to the valve, as a final protection. The input and feedback connections must be kept well clear of encapsulation since its bulk resistance falls sufficiently at temperatures above $+35^{\circ}\text{C}$ to affect seriously the d.c. levels and performance of the amplifier.

To enable pre-flight calibrations of the amplifier to be carried out when the complete experiment package was installed in the satellite, a high-value resistor, mounted in epoxy, was included in the package so that a known voltage fed to the amplifier through the resistor, would allow a calibration to be obtained in terms of input current against output volts. This resistor was normally connected in the place of one of the ionization chambers and the connections removed shortly before launching, when the fourth ionization chamber was reconnected.

4. Preliminary Experimental Results

The outputs from all four ion chambers were close to the forecast values, and during the first week of

life-time, before the 'all-sun' period a large number of satellite 'sunrises' and 'sunsets' were obtained, which will allow the determination of O_2 vertical concentration profiles over a large portion of the Earth. Results are obtained only when the satellite is within range of a telemetry receiving station so that only the high-speed 'real time' telemetry system can be used because the tape recorder low-speed system does not give a sufficiently high data rate for a satisfactory O_2 concentration profile to be obtained.

A problem has arisen in the form of a telemetry signal obtained when the satellite is well into the Earth's shadow. This has four peaks similar to the true ultra-violet output from the chambers, and can be as much as 10% of the initial 'full-sun signal'. It is believed that this is due to the electrostatic effects of the ions in the path of the satellite as they are 'swept up' along the orbital path. The value of this spurious 'dark signal', which is also apparently present when the experiment is operating correctly in solar u.v. light, will have to be determined at all times before values of O_2 density to a degree of accuracy better than approximately $\pm 40\%$ can be obtained.

So far only data from a few passes have been examined by hand computation. These have been on occasions in the Southern Hemisphere when the spurious signal mentioned above has been at its lowest, allowing the quoted accuracy of $\pm 40\%$ to be obtained. For example, at 60° South near the Falkland Islands tracking station a density of $(5 \pm 2) 10^8$ molecules/cm³ of O_2 has been measured at 175 km, this being significantly less than theoretically derived model atmospheres for this period (mid-1967) of comparatively high solar activity.†

5. Acknowledgments

This paper is published with the permission of the Director-General of the Meteorological Office. The authors gratefully acknowledge the work done on this project by other members of the High Atmosphere Research Unit, and by the Meteorological Office workshops.

† 'Cospar International Reference Atmosphere (CIRA)', (North Holland Publishing Co., Amsterdam, 1965).

*Manuscript received by the Institution on 7th November 1967.
(Paper No. 1170/AMMS 6.)*

© The Institution of Electronic and Radio Engineers, 1968

Forthcoming Conferences

Conference on Electronics Design

Electronics design is the subject of a conference to be held at the University of Cambridge from 23rd to 27th September 1968. Sponsored by the I.E.E. Electronics Division, the Institution of Electronic and Radio Engineers, and the Institute of Electrical and Electronics Engineers (U.K. and Republic of Ireland Section), this Conference was conceived partly as a result of the Joint I.E.R.E.-I.E.E.-I.Prod. E. Conference on 'The Integration of Design and Production in the Electronics Industry' held in Nottingham last July. The main theme of this Conference, however, will be confined to the management of design in all branches of the electronics industry, and topics will include:

- planning of design projects;
- formation and behaviour of design teams;
- human factors in the management of design teams;
- time, cost and technical control in design projects;
- measurement of the efficiency of design activity;
- value engineering.

The Organizing Committee are now considering contributions for inclusion in the conference programme, but further offers of papers may be accepted if sent without delay and accompanied by synopses (150 to 250 words) to the Conference Department, I.E.E., Savoy Place, London, W.C.2.

Nucleonic Instrumentation Conference and Exhibition

A three-day Conference and associated exhibition on Nucleonics Instrumentation is being sponsored by the Institution of Electrical Engineers, the Institution of Electronic and Radio Engineers, the Institute of Physics and The Physical Society, the British Nuclear Energy Society and the Scientific Instrument Manufacturers' Association. It will be held in Oxford from 24th to 26th September 1968. Lectures will be held at the Clarendon Laboratory.

It is intended that the Conference should deal with recent developments in nucleonic instrumentation with particular emphasis on research and industrial use of complete radiation detection systems including automation and on-line computer aspects of control.

The scope of the Conference has therefore been divided into five main areas: reactor instrumentation; health physics and medicine; industrial; physics research; on-line computers.

The Organizing Committee invite offers of contributions for inclusion in the Conference programme. Synopses (200 words) should be submitted to the Conference Department of the I.E.E. *on or before 1st March 1968*. Manuscripts, not exceeding 2500 words, will be required for assessment on or before 19th June 1968.

Requests for registration forms and other details should be sent to the Conference Department, I.E.E., Savoy Place, London, W.C.2.

The associated scientific exhibition is being organized by the Scientific Instrument Manufacturers' Association, and further details may be obtained from the Secretary, S.I.M.A., 20 Peel Street, London, W.8.

Advances in Control of Systems

The Third Convention of the United Kingdom Automation Council will be held at the University of Leicester, from 2nd to 4th April, 1968. The Council's first Convention was on 'Advances in Automatic Control', held at Nottingham University in 1965, and the second on 'Advances in Computer Control', at Bristol University last year.

The Convention has been organized on behalf of the U.K.A.C. by the Institute of Measurement and Control. Serving on the Organizing Committee are members of the interested Member Societies of the U.K.A.C., namely, the British Computer Society, the Institution of Chemical Engineers, the Institution of Electrical Engineers, the Institution of Electronic and Radio Engineers, the Institution of Mechanical Engineers, the Institution of Production Engineers and the Royal Aeronautical Society. The I.E.R.E. representative is Mr. W. E. Willison.

Compared to the previous Conventions, this time the scope has been broadened, and the application of control principles in a wide variety of situations will be discussed, at sessions opened by the following:

- Systems Theory—Professor G. M. Jenkins
- Management Systems—Mr. Stafford Beer
- Process Systems—Professor J. F. Coales, O.B.E.

Papers already accepted by the Committee include:

- 'Insensitivity of Control Systems'—S. Barnett.
- 'Simulation and Dimensional Analysis of a Stepping Extremum Control System'—O. Jacobs.
- 'Symbolic Array Programming for Control System Analysis by Digital Computer'—F. L. N-Nagy and O. Bar.
- 'Comparison of Numerical Methods in Non-Linear Stability Analysis'—J. Hewitt and C. Storey.
- 'On-Line Optimization of a Non-Linear Control System using Correlation Methods'—D. Bell and R. E. Selway.
- 'Equivalent Gain Matrix of a Multivariable Non-Linear Function of Stochastic Processes'—J. West.
- 'Management—A Control System'—F. L. Pitt.
- 'An Introduction to the Analysis and Control of Dynamic Models in Economics'—F. G. Pyatt and P. C. Parks.
- 'Centralization of Road Traffic Control in London'—K. W. Huddart.
- 'Network Theory Applied to Railway Scheduling'—J. Heaton.
- 'Dynamic Automated Design'—P. T. S. Buckerfield.
- 'A Fast Information Retrieval System based on Statistical Indexing'—H. J. Gawlik.
- 'The Digital Control of a Fourdrinier Paper Machine'—R. E. Jones and W. T. Wright.

On Monday evening, 1st April, Professor C. Storey will give a tutorial on Mathematical Programming, and on the Tuesday evening, 2nd April, Dr. Harold Chestnut will give an Address in commemoration of the centenary of the publication of J. C. Maxwell's paper, 'On Governors'.

Registration fees for the Convention are: members of bodies affiliated to the U.K.A.C.—£12; all others—£15. Registration forms from The Secretary, The Institute of Measurement and Control, 20 Peel Street, London, W.8.

V.L.F. Observations on *Ariel III*: A Preliminary Report

By

K. BULLOUGH,
D. de l'U. (Paris)†,

A. R. W. HUGHES, M.Sc.†

AND

Professor

T. R. KAISER, D.Phil., D.Sc.†

Presented at a Symposium on 'The Ariel III Satellite' held in London on 13th October 1967.

Summary: The very-low-frequency receiver carried by *Ariel III* makes synoptic studies at 3.2, 9.6 and 16 kHz of the world-wide distribution and occurrence of natural v.l.f. phenomena such as whistlers, chorus and hiss and also determines the configuration of the wave-field above the ionosphere due to the GBR (Rugby) transmission at 16 kHz. At these three frequencies the peak, mean and minimum signals are measured in each 28 second period ($\sim 2^\circ$) around the orbit.

This report is based on a preliminary analysis of 36 orbits (quick-look data). Zones of emission, mainly hiss, are present on every orbit. They occur at both high and medium latitudes and may often extend over 20° in invariant latitude and down to within 20° of the geomagnetic equator.

Whistlers, at medium latitudes, are usually easily identified by the large peak, very low mean and zero minimum signal levels. The signal amplitudes often vary rather smoothly along the orbit and the pattern of signal amplitude may be repeated on different orbits following the same sub-satellite path.

The geomagnetic field and the D-region are the main factors determining the configuration and intensity of the GBR (Rugby) 16 kHz wave-field above the ionosphere. Signals were identified at locations as remote as the Antipodes and the North Pacific.

1. Introduction

Naturally occurring v.l.f. phenomena (< 30 kHz) such as whistlers, chorus and hiss owe their name to the characteristic sound heard when such a signal is amplified and fed to a loudspeaker. Thus a whistler is heard as a falling tone while chorus, a mixture of falling and rising tones, is similar to the 'dawn chorus' of birds; hiss is a slowly varying noise-like signal.

The study of such phenomena¹ received a great impetus following the pioneering work of Storey² at Cambridge in 1953. He was able to show that, at v.l.f., a small fraction of the large pulse of radio-energy liberated during a lightning discharge may penetrate the lower ionosphere boundary and then propagate along the geomagnetic field line to the conjugate point in the opposite hemisphere. The path along which the pulse travels may extend out into the exosphere to several Earth-radii at the geomagnetic equator. The medium is dispersive so that the higher frequencies travel faster; hence if the signal at the conjugate point is amplified and fed to a loudspeaker a falling tone or whistler is heard. Measurements of whistler dispersion at many ground-based stations

have yielded information on the distribution of electrons in the exosphere.

At high latitudes, natural v.l.f. emissions such as hiss and chorus are closely associated with the aurora. They are mainly due to energetic electrons (< 20 keV) flowing along the field lines into the auroral zones. Such particles radiate most strongly either when their velocity equals the wave-phase velocity in the ionosphere/exosphere (e.g. Cerenkov radiation) or when the particle gyro-rotation is in phase with the rotating E-vector of the circularly polarized v.l.f. wave (cyclotron resonance).³

The different emissions are shown schematically in Fig. 1. Ground-based observations of these phenomena are severely limited. Most of the signals do not penetrate the ionosphere/atmosphere boundary either because their wave-normals are not sufficiently well-aligned with the local magnetic field, or because of absorption in the D-region which, during the day, may reduce the signal level by as much as 60 dB. Also, the source of emission may be difficult to locate since the signal, having penetrated the lower ionosphere boundary into the Earth/ionosphere duct, may then propagate to considerable distances. Finally, the satellite is immune to industrial interference and radiation from 50 Hz power lines, though this

† Department of Physics, University of Sheffield.

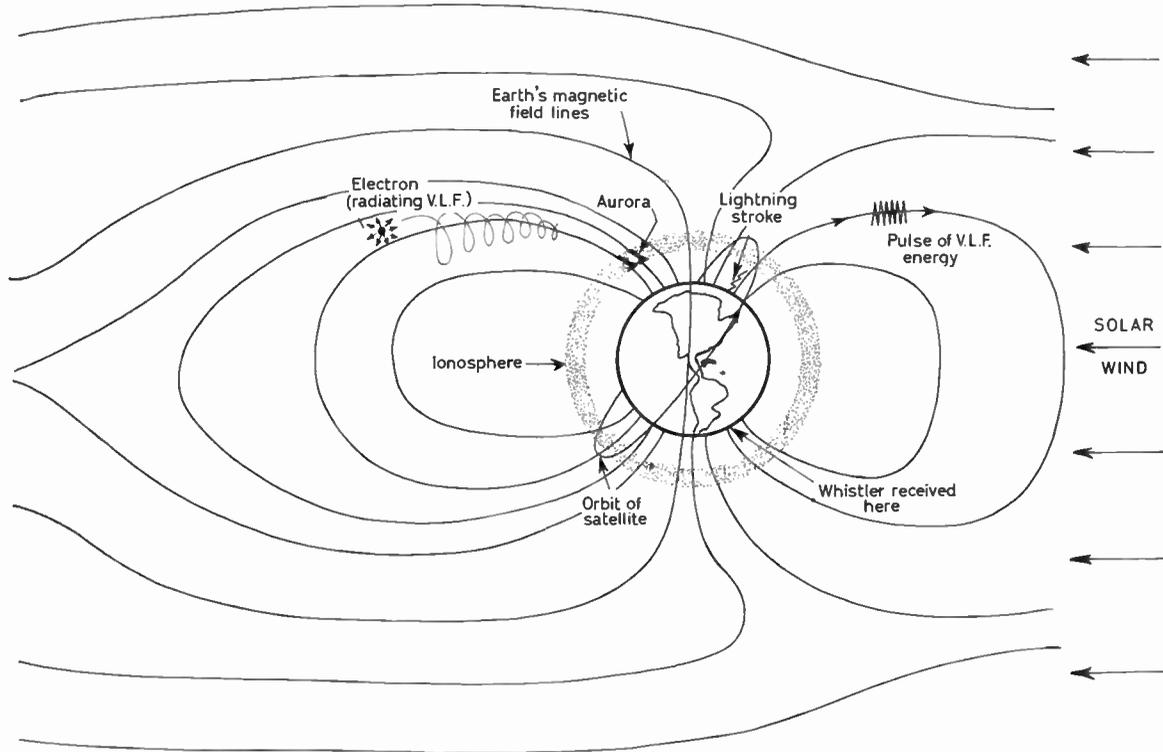


Fig. 1. Natural v.l.f. emissions (schematic).

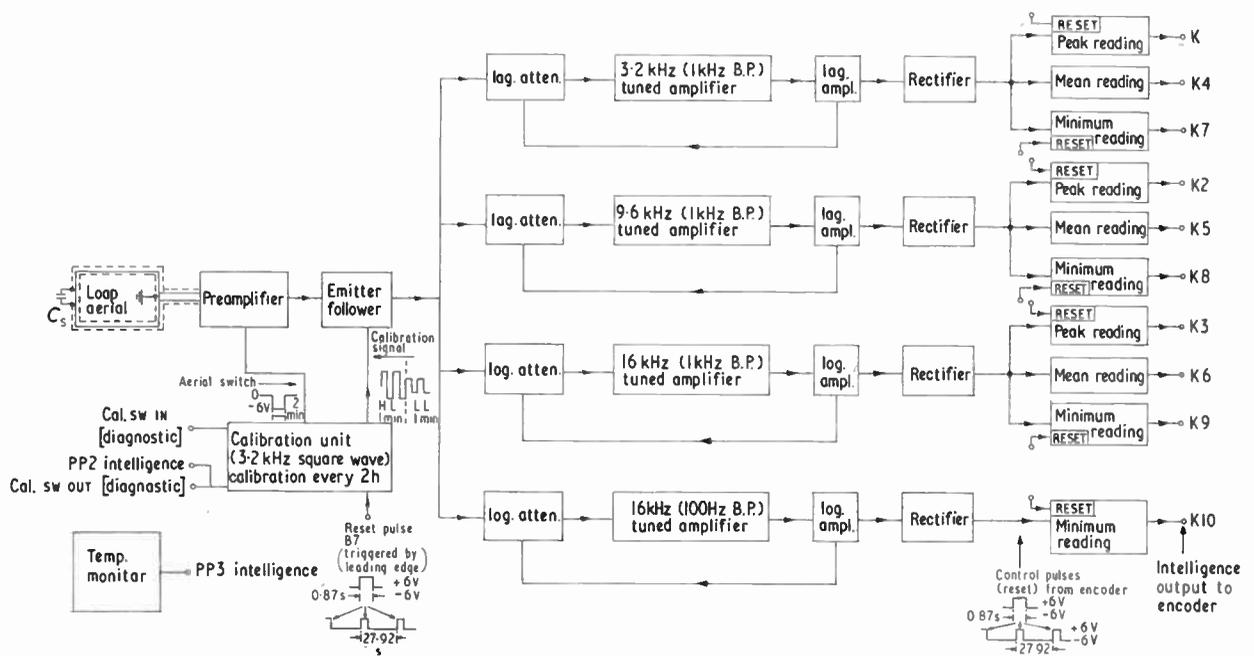


Fig. 2. Block diagram of the Ariel III v.l.f. receiver. (Courtesy J. Sci. Instrum.)

advantage may, of course, be offset by severe compatibility problems on the spacecraft itself.

2. The Equipment

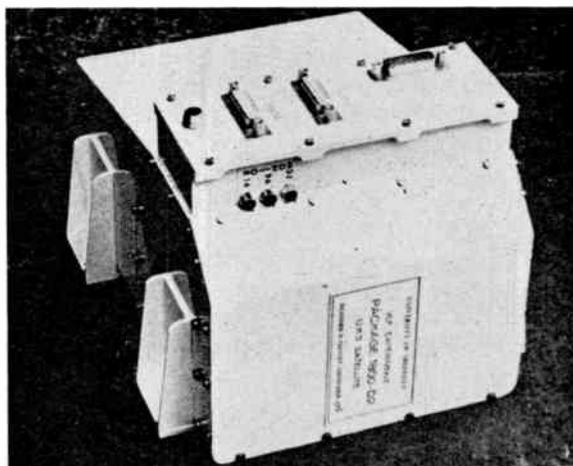
2.1. Choice of Parameters

The choice of satellite orbital inclination was governed by two considerations; firstly, observations over the whole range of geomagnetic latitudes meant that the orbital inclination had to be at least 78° since the geomagnetic poles are located at 68°W , 78°N , 112°E , 78°S ; secondly, the orbital precession had to be such that complete local time coverage was obtained for each season. This meant that the inclination should not exceed 83° .

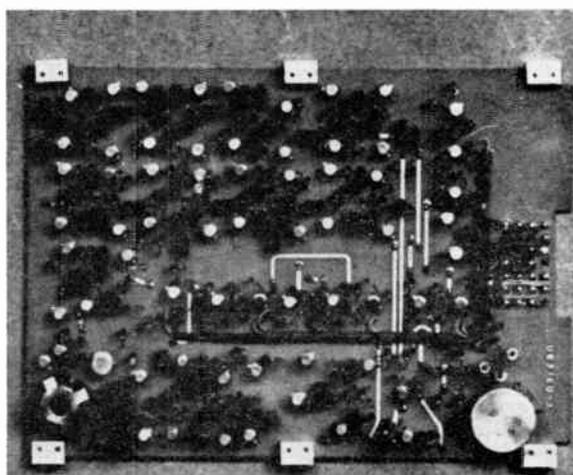
The main design aim was to obtain useful synoptic data within the limitations of the available data rate (~ 3 bits/s). To attain this end there is 'on-board' processing of the data and key parameters of the signals are stored on the satellite tape-recorder. The latter is subsequently read out to a conveniently situated STADAN station. In this respect, the experiment differs from those carried on the *Alouette* and *Injun III* satellites^{4,5} where detailed studies of the spectral and dispersive characteristics of the signals have necessitated relaying the output of a broad band v.l.f. receiver to ground. The latter are, so far, 'real-time' studies limited to the vicinity of STADAN stations.

A block diagram of the *Ariel III* v.l.f. receiver and photographs of the receiver and circuit boards are shown in Figs. 2 and 3. A detailed description of the equipment is given elsewhere.⁶ Observations are made at three representative frequencies: 3.2, 9.6 and 16 kHz, each channel having a bandwidth of 1 kHz. These frequencies are harmonically related for ease in calibration. The lower frequency lies in the middle of the range 1 to 5 kHz where chorus usually occurs; the centre frequency lies in the auroral hiss band. The reason for choosing 16 kHz was that the configuration of the wave-field due to the GBR (Rugby) transmitter above the ionosphere could be investigated. At 16 kHz an additional narrow band (100 Hz) channel was used in order to discriminate between the c.w. transmission and hiss (Fig. 4(b)). The Admiralty kindly agreed to code the GBR signal during 'standby periods' (90 seconds 'on', 30 seconds 'off'). This greatly facilitated signal recognition at distant locations.

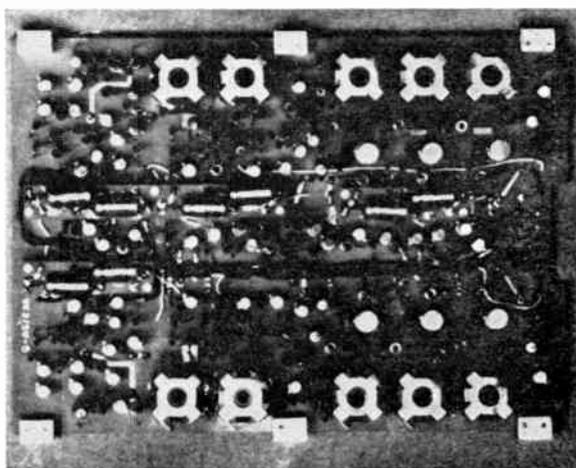
To distinguish between the different classes of signal, the peak, mean and minimum amplitude at each frequency was measured in each observing interval of 28 seconds ($\sim 2^\circ$) around the orbit. Thus whistler activity is characterized by large peak, very low mean and zero minimum readings; chorus, observed mainly at 3.2 kHz, by minimum readings perhaps 10 dB lower than the mean readings and hiss by mean and minimum readings separated by only



(a) The complete unit.



Pre-amplifier and calibration unit.



3.2 and 9.6 kHz channels.
(b) Two circuit boards.

Fig. 3. The v.l.f. receiver (weight 3.6 kg (~ 8 lb); power consumption $\sim \frac{1}{4}$ W). (Courtesy *J. Sci. Instrum.*)

a few decibels (Fig. 4(a)). The time-constants of the peak, mean and minimum reading circuits, 0.01, 30 and 0.1 second respectively, are such that a white-noise signal would give virtually identical outputs. The narrow band channel has a minimum reading output only, with a time-constant of 1 second. Whistlers are rejected by the minimum reading circuits and at 16 kHz the special coded transmission produces a characteristic pattern of two up, two down, . . . etc. Peak and minimum reading circuits are reset immediately after being read to the tape-recorder whereas the mean circuit records a 'running mean'.

2.2. General Description

The untuned 14-turn screened loop aerial of 3 m² area is mounted near the ends of the four booms, concentric with the Jodrell Bank loop. The gap in the loop-screen is by-passed by a capacitor (0.01 μ F) so that the loop approximates to a shorted turn in the Jodrell Bank experiment's range of frequencies (2 to 4.3 MHz). Detailed calculations and measurements⁶ showed that coupling between the two loops resulted in a reduction of less than 1 dB in the Jodrell signal. The Jodrell loop appears open-circuited at v.l.f. and has negligible effect on the Sheffield University experiment's sensitivity. Another reason for the capacitor is to reduce the possibility of cross-modulation in the pre-amplifier due to powerful short-wave transmissions. Additional r.f. filtering at the loop terminals removed any residual telemetry r.f. (137 MHz) signal. The screening was essential to stop electrostatic interference from the 6.5 kHz waveform applied to Birmingham University's electron density probe.

Optimum signal/noise ratio was obtained by employing transformer coupling between the loop and the pre-amplifier so that the transformed inductive reactance at 10 kHz, presented to the input of the pre-amplifier transistor (2N930), was equal to the optimum generator impedance specified in the data sheet of the transistor. The pre-amplifier comprised a low-noise, common-emitter stage having a limiting sensitivity equal to $6 \times 10^{-6} \gamma$ on the 16 kHz narrow band channel.† The amplified broad band signal is fed to the four tuned amplifiers. Each of the selective amplifiers has a logarithmic response (~ 75 dB) and was carefully temperature-compensated so that it was unnecessary to apply any temperature correction to the data for the range of receiver temperatures (10 to 35°C) so far encountered. The sensitivity ($\sim 10^{-5} \gamma$) and the logarithmic response were such that zones of emission are easily identified on every orbit while, at the same time the upper limit (0.1 γ) to the dynamic

range was only very rarely exceeded. The peak and minimum reading circuits were improved versions of a circuit originally due to Wager.⁷

During a 112-second period every two hours a calibration signal (3.2 kHz square-wave) is fed into the receiver at a point immediately after the pre-amplifier. The latter is biased beyond cut-off during this period. The calibration consists of two intervals (56 s) of high-level signal followed by two intervals where the signal is reduced by 30 dB.

3. Emissions: General Characteristics

'Quick-look' data for 36 orbits were available to the authors at the time of writing and zones of emission are present on every orbit. The equipment is some 30 dB more sensitive than the v.l.f. experiment on *Injun III*⁸ where such emissions were identified on 171 orbits out of 4000.

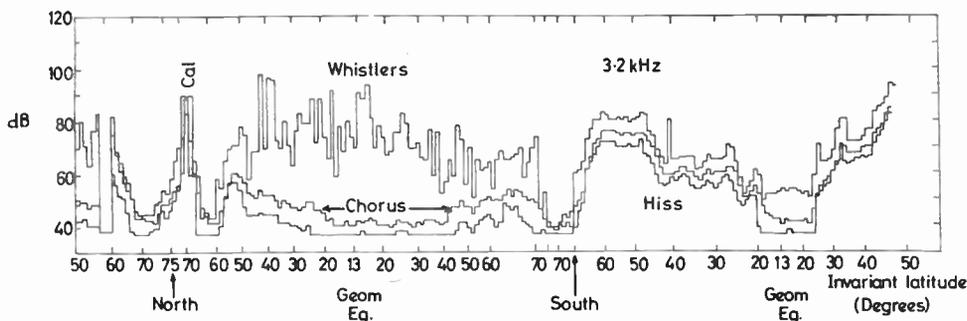
The distribution of the zones is seen most clearly on N \rightarrow S passes at Winkfield where the satellite orbit includes the magnetic poles and lies close to a magnetic meridian. The high-latitude hiss zone studied by Gurnett⁸ is often present. At lower latitudes hiss at all three frequencies and chorus at 3.2 kHz have often been observed (Fig. 4(a)). In a typical 'hiss' event the minimum signal is generally only 1 or 2 dB below the mean signal level with the peak reading about 15 dB higher. The signal amplitude varies smoothly over the hiss zone which may often extend over 20° in latitude and occasionally, down to within 20° of the geomagnetic equator. On many orbits the zonal maxima for the three frequencies are far from coincident, also there is often marked asymmetry about the geomagnetic equator. The mean levels for the most intense hiss recorded were $5 \times 10^{-2} \gamma$, $10^{-2} \gamma$ and $6 \times 10^{-3} \gamma$ on 3.2, 9.6 and 16 kHz respectively. The maximum hiss observed on individual orbits ranged over 40 dB and showed an average spectral decrease towards the higher frequencies of 7 dB/octave with a wide scatter.

The very limited data available for the magnetically disturbed period which started on 25th May suggested that the principal effect might be a broadening of the zones, which then extend to low latitudes, rather than an increase in the maximum observed signal.

4. Whistlers

At low and medium latitudes whistler signals give rise to large peak readings, which occasionally fill the dynamic range, very low mean, and zero minimum readings. An unexpected feature of the peak readings was that they frequently vary rather smoothly along the orbit (Fig. 4(b)). In such instances small periodic fluctuations in the signal strength may reflect latitudinal variations in the whistler path possibly due to travelling ionospheric disturbances.

† $1 \gamma = 10^{-9} \text{ Wbm}^{-2}$.

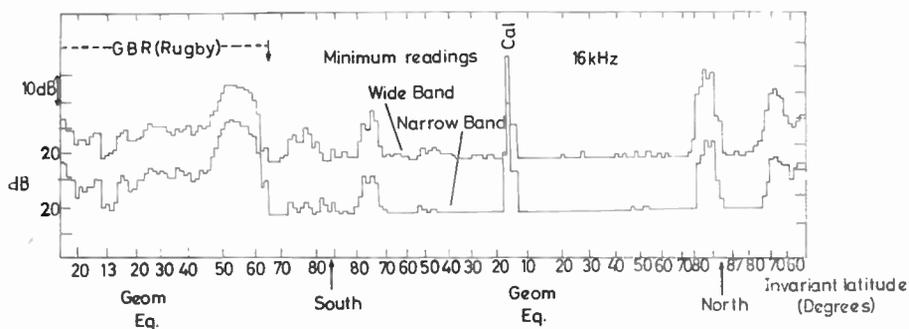


(a) 3.2 kHz observations. A S → N pass, read-out at Winkfield: 1855 U.T., 26th May 1967.

Record begins at 1719 U.T., sub-satellite point: 11.5°E, 49.2°N.

Record ends at 1855 U.T., sub-satellite point: 12.2°W, 50.3°N.

On this orbit chorus was observed at medium latitudes on the morning side of the earth, hiss on the evening side. The calibration signal is superimposed on the high-latitude hiss zone. Whistlers were present mainly on the morning side of the Earth.



(b) 16 kHz minimum readings.

A N → S pass, read-out at Winkfield: 2159 U.T., 10th July 1967.

Record begins at 2030 U.T., sub-satellite point: 23.9°E, 33.5°N.

Record ends at 2159 U.T., sub-satellite point: 6.5°W, 55.3°N.

On this orbit the GBR transmission was c.w. only. The signal was present from the beginning of the record south of Rugby down to 65°S invariant latitude. The whistler-mode signal is particularly well-defined in the conjugate zone to Rugby (50 to 60° invariant latitude). Note also the high-latitude hiss zone during this particularly quiet (magnetically) period.

Fig. 4. Typical records (abscissa: invariant latitude (degrees); ordinate: dB above $1 \mu\gamma$ ($10^{-15}\gamma^2/\text{Hz}$)).

When the pattern of peak signals on the three channels is similar it is probable that one whistler is responsible for all three in a given interval. In this case it will be possible to determine relative whistler amplitudes at the three frequencies as a function of invariant latitude.

Another unexpected feature of the peak readings was the discovery of a discontinuity in the vicinity of the South Atlantic anomaly. This feature, which was present on five orbits, was especially pronounced on the 16 kHz peak reading where it appeared as a sudden rise, in less than one observing interval, of about 30 dB in signal amplitude to the south (S → N pass) of the geomagnetic equator at an invariant latitude of approximately 24°.

5. Intensity and Occurrence of the GBR 16 kHz Signal Above the Ionosphere

GBR signals have been identified on 26 of the 36 orbits. They are present on the 7 N → S passes at Winkfield near to Rugby† and/or at the conjugate point‡ where the signals travel in the whistler mode (Fig. 4(b)). Signals have also been received at remote locations such as the tip of South America and the North Pacific (five orbits). In the latter case reception is probably due to transmission via the Earth/ionosphere waveguide to a location where the orientation of the geomagnetic field is favourable for transmission into the ionosphere. Signals may be received in the vicinity of the geomagnetic equator at night when D-region absorption is low (Fig. 4(b)). Attenuation of the signal in the D-region is greater when the latter is sunlit. Reception at remote locations occurs usually when the D-region is in shadow.

On several occasions, during periods when clear, specially coded or c.w. signals were being received, it was found that the narrow band minimum reading was consistently a few decibels higher than the wide band minimum reading. This is probably due to fading of the signals such that the fluctuation period is short relative to the time-constant (1 s) of the narrow band channel. This fading is evident in Fig. 4(b) in the conjugate region and is consistent with a focusing effect predicted by Yabroff,⁹ for whistler-mode paths. The fading takes place between signals which have travelled along different paths in the magnetosphere.

Previous observations¹⁰ of man-made transmissions made in the *Lofti I* experiment in 1961 were few in

† Peak signal levels greater than 50 dB above 1 $\mu\gamma$.

number (36 days), they were also real-time and therefore limited to the vicinity of STADAN stations and limited to latitudes lying between 28°N and 28°S. These observations detected signals from NBA (18 kHz) in North America as far away as Australia. The attenuation of magnetic field intensity, measured near extreme line-of-sight distances to the north of the station NBA, was less than 13 dB for half the time at night and less than 38 dB for half the time during daylight hours.

6. Acknowledgments

The authors are indebted to many people of the Royal Aircraft Establishment, British Aircraft Corporation, General Electric Company, National Aeronautics and Space Administration, the Admiralty, and the Space Research Management Unit, Science Research Council for the success of this enterprise.

7. References

1. R. A. Helliwell, 'Whistlers and Related Ionospheric Phenomena' (Stanford University Press, California, 1965).
2. L. R. O. Storey, 'An investigation of whistling atmospherics', *Phil. Trans. Roy. Soc., A*, **246**, No. 908, pp. 113–41, 1953.
3. N. Brice, 'Fundamentals of very low frequency emission generation mechanisms', *J. Geophys. Res.*, **69**, No. 21, pp. 4515–22, 1st November 1964.
4. R. E. Barrington, J. S. Belrose and D. A. Keeley, 'Very low frequency noise bands observed by the *Alouette I* satellite', *J. Geophys. Res.*, **68**, No. 24, pp. 6539–41, 15th December 1963.
5. D. A. Gurnett and B. J. O'Brien, 'High latitude geophysical studies with satellite *Injun 3*, Part 5: Very-low-frequency electromagnetic radiation', *J. Geophys. Res.*, **69**, No. 1, pp. 65–89, 1st January 1964.
6. K. Bullough *et al.*, 'The Sheffield University very low frequency Experiment on *Ariel III*', *J. Sci. Instrum. (J. Phys.-E.)* Series 2, **1**, pp. 77–85, February 1968.
7. J. H. Wager, 'A short term analogue data storage unit', *Electronic Engng*, **35**, pp. 254–6, April 1963.
8. D. A. Gurnett, 'A satellite study of v.l.f. hiss', *J. Geophys. Res.*, **71**, No. 23, pp. 5599–5615, 1st December 1966.
9. I. Yabroff, 'Computation of whistler ray paths', *J. Res. Nat. Bur. Stands*, **65D**, No. 5, pp. 485–505, May 1961.
10. J. P. Leiphart *et al.*, 'Penetration of the ionosphere by very-low-frequency radio signals—interim results of the *Lofti I* experiment', *Proc. Inst. Radio Engrs*, **50**, No. 1, pp. 6–17, January 1962.

Manuscript first received by the Institution on 6th September 1967 and in final form on 8th December 1967.

(Paper No. 1171/AMMS 8.)

© The Institution of Electronic and Radio Engineers, 1968

Performance Factor of Linear Two-port Active Networks

By

Professor

S. VENKATESWARAN,

B.Sc., M.A., Ph.D., D.I.C., C.Eng.,
M.I.E.E.†

Summary: Callendar's 'curve distortion criterion', S_d , and a factor, K_r , in his power gain expressions are simply related to the 'performance factor', n and 'inherent performance factor', n_i , of the linear two-port active network. A lower limit of 10, proposed earlier for the 'invariant stability factor', s , is shown as equivalent to a lower limit of 4 for S_d . The role of the performance factor in the maximum power gains of mismatched and optimum unilateralized single active stages is emphasized. Finally, variations in the total port immittance due to changes in terminating immittance at the opposite port are expressed in terms of the performance factor.

1. Introduction

A factor, S_d , called the 'curve distortion criterion' has been proposed by Callendar¹ for tuned transistorized radio-frequency and intermediate-frequency amplifiers. It is in terms of admittance parameters of the network. In particular, if S_d is equal to 4, the asymmetry in the power gain-frequency response curve is not excessive. However, if Stern's stability factor,² k , is equal to 4, the asymmetry in the response curve is excessive.

Characterize the linear two-port active network with passive source immittance, p_s and passive load immittance, p_L , by its p -matrix,

$$[p] = \begin{bmatrix} p_{11} + p_s & p_{12} \\ p_{21} & p_{22} + p_L \end{bmatrix} = \begin{bmatrix} p_1 & p_{12} \\ p_{21} & p_2 \end{bmatrix} \dots\dots(1)$$

where p_1 , etc., may be sets of h -, z -, y - or g - matrix parameters. This p - matrix can be in terms of the general complex variable, λ , or in terms of real frequencies, i.e. with λ equal to $j\omega$.

The two 'total driving point immittances', p_{p1} and p_{p2} include the source terminating immittance, p_s , at the input or port 1 and the load terminating immittance, p_L , at the output or port 2. They are given by

$$p_{p1} = p_1 \left\{ 1 - \frac{p_{12} p_{21}}{p_1 p_2} \right\} \dots\dots(2)$$

$$p_{p2} = p_2 \left\{ 1 - \frac{p_{12} p_{21}}{p_1 p_2} \right\} \dots\dots(3)$$

The forward transfer function, γ_{21} , of the network with terminations, p_s , p_L is a ratio of the output voltage or current to the input voltage or current. It is given by

$$\gamma_{21} = \frac{\gamma'_{21}}{1 - \frac{p_{12} p_{21}}{p_1 p_2}} \dots\dots(4)$$

where γ'_{21} equals $\frac{p_{21}}{p_1 p_2}$ for h - or g - set of matrix parameters and equals $\frac{p_{21} p_L}{p_1 p_2}$ for z - or y - set of matrix parameters. Thus γ'_{21} is the forward transfer function, when the reverse transfer parameter, p_{12} vanishes.

Similarly the reverse transfer function, γ_{12} , of the network is given by

$$\gamma_{12} = \frac{\gamma'_{12}}{1 - \frac{p_{12} p_{21}}{p_1 p_2}} \dots\dots(5)$$

where γ'_{12} equals $\frac{p_{12}}{p_1 p_2}$ for h - or g - set of matrix parameters and equals $\frac{p_{12} p_s}{p_1 p_2}$ for z - or y - set of matrix parameters. Thus γ'_{12} is the reverse transfer function, when the forward transfer parameter, p_{21} , vanishes.

The reciprocal term, $\frac{p_{12} p_{21}}{p_1 p_2}$, in eqns. (2) to (5) is a function of all the four parameters, p_1 , p_{12} , p_{21} and p_2 . It may be defined as the 'internal loop gain', g_l , of the two-port network in its p - matrix.

Assume that p_{11} , p_{12} , p_{21} and p_{22} have no poles or zeros on the imaginary axis or on the finite right half complex plane whether p equals h , z , y or g set of matrix parameters. This ensures that the network is stable whether each of its ports be open or short circuited, a stipulation satisfied by a wide range of active networks, including valve and transistor amplifier stages. Networks violating one or more of the above conditions can be made to satisfy them by the

† Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India.

addition of resistances at the ports and/or external passive feedback.

The poles of $\frac{p_{12}p_{21}}{p_1p_2}$ and $\left\{1 - \frac{p_{12}p_{21}}{p_1p_2}\right\}$ are identical but not their zeros. p_1 is the sum of p_{11} and the passive source immittance p_s ; similarly p_2 is the sum of p_{22} and the passive load immittance p_L . As a result of our assumptions in the previous paragraph, $\frac{p_{12}p_{21}}{p_1p_2}$ and hence $\left\{1 - \frac{p_{12}p_{21}}{p_1p_2}\right\}$ cannot have poles on the finite right half complex plane including the imaginary axis. Equations (2) to (5) suggest that for stability, no zero of $\left\{1 - \frac{p_{12}p_{21}}{p_1p_2}\right\}$ should lie on the finite right half complex plane including the imaginary axis.

Call $\frac{p_{12}p_{21}}{p_1p_2}$, the loop gain, g_l ; note that in practice g_l tends to zero at infinite frequency so that $(1-g_l)$ tends to unity as frequency tends to infinity. $\{1-g_l(\lambda)\}$ is analytic within and on a contour enclosing the imaginary axis and the right half complex plane. Nyquist's steps therefore lead to the familiar stability criterion,³ viz.

$$g_l(j\omega) \neq 1 \quad \text{for } 0 < \omega < \infty \quad \dots\dots(6)$$

because of the symmetry of the poles and zeros about the real axis.

Thus, the two-port network is stable, if

$$g_{lr} < 1 \quad \dots\dots(7)$$

where g_{lr} refers to the pure real loop gain; loop gain is real when $\arg g_l(j\omega)$ equals $2r\pi$ radians, r being an integer.

2. Stern's Stability Factor

For tuned amplifier stages, the power gain frequency response curve exhibits the most significant peak⁴ at the 'tuned' or 'central' frequency. It is therefore sufficient to apply eqn. (7) at this central frequency taking into account the variations in reactances (or susceptances) at the ports; such variations occur when tuning the stage.

A given two-port network with fixed real parts of terminations, ρ_s, ρ_L , is said to be 'absolutely stable', if it is stable for all values of imaginary parts of terminations σ_s, σ_L , i.e. the network is stable for all reactive terminations. This requirement is more stringent than the one associated with fixed terminations. If the network is not absolutely stable, it is said to be 'potentially unstable'.

By suitable choice of σ_s, σ_L the imaginary parts of terminating immittances, p_s, p_L , the internal loop gain, g_l , can be made purely real and maximized.^{5,6}

This maximum real loop-gain,

$$g_{lr \max} = \frac{|p_{12}p_{21}| + \text{Re}(p_{12}p_{21})}{2 \text{Re}(p_1) \text{Re}(p_2)} = \frac{L+M}{2\rho_1\rho_2} = \frac{1}{k} \quad \dots\dots(8)$$

where

$$p_1 = \rho_1 + j\sigma_1; \quad p_2 = \rho_2 + j\sigma_2 \quad \dots\dots(9)$$

$$p_{12}p_{21} = M + jN; \quad |p_{12}p_{21}| = L \quad \dots\dots(10)$$

Maximum real loop gain occurs when the imaginary parts of the source and the load terminations are such as to satisfy

$$\begin{aligned} \frac{\sigma_{11} + \sigma_s}{\rho_{11} + \rho_s} &= \frac{\sigma_{22} + \sigma_L}{\rho_{22} + \rho_L} \\ &= \frac{|p_{12}p_{21}| - \text{Re}(p_{12}p_{21})}{\text{Im}(p_{12}p_{21})} \\ &= \frac{L-M}{N} \quad \dots\dots(11) \end{aligned}$$

For 'absolute stability', according to eqns. (7) and (8), $g_{lr \max} < 1$ or Stern's stability factor, k must be greater than unity.

When the values of σ_s, σ_L are adjusted to those given by eqn. (11) for the central frequency, f_0 , a polar plot of the loop gain, g_l , for frequencies ranging from 0 to ∞ gives g_l real and equal to $g_{lr \max}$ or $1/k$ at the central frequency, f_0 . So the 'gain margin' now corresponds to k , if $k > 1$. This is a further interpretation for k .

3. Performance Factor and Inherent Performance Factor

By varying σ_s, σ_L , the modulus of the loop gain can be changed. It is a maximum⁶ when

$$\frac{\sigma_{11} + \sigma_s}{\rho_{11} + \rho_s} = \frac{\sigma_{22} + \sigma_L}{\rho_{22} + \rho_L} = 0 \quad \dots\dots(12)$$

and is given by

$$|g_l|_{\max} = \frac{|p_{12}p_{21}|}{\text{Re}(p_1) \text{Re}(p_2)} = \frac{L}{\rho_1\rho_2} = \frac{1}{n} \quad \dots\dots(13)$$

where n is the 'performance factor of the network'.⁶⁻⁸

From eqns. (8) and (13),

$$n = k \left\{ \frac{|p_{12}p_{21}| + \text{Re}(p_{12}p_{21})}{2|p_{12}p_{21}|} \right\} = \left(\frac{L+M}{2L} \right) k \quad \dots\dots(14)$$

If n equals 1, k is greater than 1, according to eqn. (14), except for the special case where M equals L . Thus if $n > 1$ or $|g_l|_{\max} < 1$, we are assured more than the minimum requirement for 'absolute stability'.

By making ρ_s and ρ_L both equal to zero, a particular

performance factor may be obtained. It is the 'inherent performance factor', n_i , of the network.

$$|g_{ii}|_{\max} = \frac{|p_{12} p_{21}|}{\text{Re}(p_{11}) \text{Re}(p_{22})} = \frac{L}{\rho_{11} \rho_{22}} = \frac{1}{n_i} \quad \dots\dots(15)$$

where $|g_{ii}|_{\max}$ refers to the maximum modulus of the 'inherent loop gain' of the network, i.e. with $\rho_s = \rho_L = 0$. It occurs when

$$\frac{\sigma_{11} + \sigma_s}{\rho_{11}} = \frac{\sigma_{22} + \sigma_L}{\rho_{22}} = 0 \quad \dots\dots(16)$$

Callendar's curve distortion factor,¹ S_d , extended to the general ρ -matrix is $\frac{2\rho_1 \rho_2}{L}$ and is therefore equal to $2n$, according to eqn. (13). Another factor¹ that appears often in his power gain expressions is K_r , which in the general ρ -matrix is equal to $\left\{ \frac{L}{\rho_{11} \rho_{22}} \right\}^{\frac{1}{2}}$; according to eqn. (15), K_r equals $\left\{ \frac{1}{n_i} \right\}^{\frac{1}{2}}$. Thus, Callendar's factors, S_d and K_r are simply related to the 'performance factor', n , and the 'inherent performance factor', n_i of the two-port network.

For a given performance factor, n , the operating power gain of the two-port network can be maximized. For a large value of this performance factor, this maximum power gain is closely given by⁷

$$g_{\max n} \approx \frac{4}{n} \left| \frac{p_{21}}{p_{12}} \right| \left\{ 1 - \sqrt{\frac{n_i}{n}} \right\}^2 \quad \dots\dots(17)$$

Again, with optimum unilateralization, the maximum available power gain is given by⁷

$$g_{\max c} \approx \frac{1}{4n_i} \left\{ \frac{|p_{21} \pm p_{12}|^2}{|p_{12} p_{21}|} \right\} \quad \dots\dots(18)$$

with plus sign for h - or g - set of matrix parameters and minus sign for z - or y -set of matrix parameters. In eqn. (18), n_i refers to the inherent performance factor of the two-port network before the application of unilateralizing feedback.

It is therefore possible to relate the power gains obtained with and without unilateralizing feedback, if the performance factor in the 'mismatch' case, is large compared to unity. From eqns. (17) and (18),

$$g_{\max n} \approx 16 g_{\max c} \frac{n_i}{n} \quad \dots\dots(19)$$

It is clear that the numerical constant in this relationship ensures that for a range of performance factor values the simple mismatched two-port network can develop a larger power gain than the same network when optimum unilateralized. This corresponds to a range of fractional bandwidths.⁷

4. Invariant Inherent Stability Factor and Invariant Stability Factor

An 'invariant inherent stability factor', s_i , and an 'invariant stability factor', s , have also previously been proposed.^{6, 8} If a 'device network' is absolutely stable, its s_i is ≥ 1 . Its maximum available power gain is obtained with conjugate matched terminations at its ports and is given by

$$g_{\max s_i} \quad s_i = \left| \frac{p_{21}}{p_{12}} \right| \quad \dots\dots(20)$$

If $s_i < 1$ (or otherwise), the given 'device network' can be padded by resistances at the input and output ports such that the resultant 'modified network' has its s value ≥ 1 . The maximum power gain⁹ of the device network for a given s is then given by

$$g_{\max s} \quad s = \left| \frac{p_{21}}{p_{12}} \right| \quad \dots\dots(21)$$

Maximum power gain of eqn. (21) is obtained with conjugate match for the 'modified network' but indirect mismatch⁹ for the 'device network'.

If $s_i \geq 1$, it is the inverse of the modulus of the internal loop gain of the network with conjugate matching terminations. However, if $s \geq 1$, it is the inverse of the modulus of the internal loop gain of the 'device network' with its mismatched terminations that yield the maximum power gain of eqn. (21).

'Invariant inherent stability factor', s_i , is defined as

$$s_i = \eta_i + \sqrt{\eta_i^2 - 1} \quad \dots\dots(22)$$

where the 'invariant inherent factor'

$$\eta_i = \frac{2\rho_{11} \rho_{22} - M}{L} \quad \dots\dots(23)$$

Similarly, the 'invariant stability factor', s , is defined as

$$s = \eta + \sqrt{\eta^2 - 1} \quad \dots\dots(24)$$

where the 'invariant factor'

$$\eta = \frac{2\rho_1 \rho_2 - M}{L} \quad \dots\dots(25)$$

η is also the 'overall invariant stability factor' of Rollett.¹⁰

In earlier papers^{11, 12} on synchronously tuned, cascaded, linear two-port networks, alignability and skew in power gain-frequency response are considered. A lower bound of 10 is imposed on s to avoid excessive asymmetry in the response curve and make amplifiers alignable. From eqns. (24) and (25)

$$\eta = \frac{s^2 + 1}{2s} \quad \dots\dots(26)$$

and

$$n = \frac{\rho_1 \rho_2}{L} = \frac{1}{2} \left\{ \eta + \frac{M}{L} \right\} \quad \dots\dots(27)$$

The performance factor of eqn. (27) is the inherent performance factor of the 'modified network' specified by eqn. (1). It does not include further source and load conjugate matched terminations of 'indirect mismatch' that yields the maximum power gain of eqn. (21).

Equation (26) gives a lower bound of 5 for η , if s has a lower bound of 10. M is the real part of $p_{12}p_{21}$, while L is the modulus of same. Hence eqn. (27) gives a lower bound of 2 for n . S_d , which is $2n$, must therefore have a lower bound of 4, a conclusion in agreement with that of Callendar.¹

5. Performance Factor in Interactions between Ports

The input and output ports interact with each other due to the internal feedback associated with the loop gain. The total input port immittance (with source immittance), p_{p1} is given by eqn. (2) as

$$p_{p1} = p_1 \left\{ 1 - \frac{p_{12}p_{21}}{p_1 p_2} \right\} \quad \dots\dots(28)$$

Since the loop gain has a maximum modulus of $1/n$, according to eqn. (13)

$$\left| \frac{p_{p1} - p_1}{p_1} \right| \leq \frac{1}{n} \quad \dots\dots(29)$$

Similarly the total output port immittance, p_{p2} (with load immittance), differs from the total output self immittance, p_2 , by a normalized amount given by

$$\left| \frac{p_{p2} - p_2}{p_2} \right| \leq \frac{1}{n} \quad \dots\dots(30)$$

Let the total input port immittance, p_{p1} , change by a small amount, Δp_{p1} , as a result of a small change Δp_L , in the load immittance, p_L . Differentiate eqn. (28) with respect to p_L to obtain

$$\Delta p_{p1} = \frac{p_{12}p_{21}}{p_2^2} \Delta p_L \quad \dots\dots(31)$$

which can be rearranged as

$$\frac{\Delta p_{p1}}{p_1} = \frac{p_{12}p_{21}}{p_1 p_2} \frac{\Delta p_L}{p_2} \quad \dots\dots(32)$$

Take modulus on each side of eqn. (32) to get

$$\left| \frac{\Delta p_{p1}}{p_1} \right| \leq \left| \frac{p_{12}p_{21}}{p_1 p_2} \right|_{\max} \left| \frac{\Delta p_L}{p_2} \right| \quad \dots\dots(33)$$

Equations (13) and (33) now yield

$$\left| \frac{\Delta p_{p1}}{p_1} \right| \leq \frac{1}{n} \left| \frac{\Delta p_L}{p_2} \right| \quad \dots\dots(34)$$

Similarly, if the source immittance changes by a small amount, Δp_s , the small change, Δp_{p2} , in the total output port immittance is given by

$$\frac{\Delta p_{p2}}{p_2} = \frac{p_{12}p_{21}}{p_1 p_2} \frac{\Delta p_s}{p_1} \quad \dots\dots(35)$$

and

$$\left| \frac{\Delta p_{p2}}{p_2} \right| \leq \frac{1}{n} \left| \frac{\Delta p_s}{p_1} \right| \quad \dots\dots(36)$$

If both the ports are tuned in the absence of feedback, both σ_1 and σ_2 will vanish. Eqns. (34) and (36) then become

$$\left| \frac{\Delta p_{p1}}{\rho_1} \right| = \frac{1}{n} \left| \frac{\Delta p_L}{\rho_2} \right| \quad \dots\dots(37)$$

and

$$\left| \frac{\Delta p_{p2}}{\rho_2} \right| = \frac{1}{n} \left| \frac{\Delta p_s}{\rho_1} \right| \quad \dots\dots(38)$$

Equations (29), (30), (34), (36) to (38) emphasize the role of the performance factor in assessing interactions between input and output ports of linear active and passive networks. Ease of tuning the ports requires this interaction to be small, which is achieved by a moderate value for the performance factor. A minimum value of between 2 and 3 is suggested for the performance factor to ensure alignability and negligible skew due to feedback. This is equivalent to a minimum value of 10 for the invariant stability factor.

6. Acknowledgment

The author is grateful to Dr P. K. Kelkar, Director, Indian Institute of Technology, Kanpur, for providing facilities.

7. References

1. M. V. Callendar, 'Gain and stability of tuned transistor r.f. and i.f. amplifiers', *The Radio and Electronic Engineer*, 32, pp. 233-242, October 1966.
2. A. P. Stern, 'Considerations on the stability of active elements and applications to transistors', *Inst. Radio Engrs Conv. Rec.*, Part 2, pp. 46-52, 1956.
3. H. Nyquist, 'Regeneration theory', *Bell Syst. Tech. J.*, 11, pp. 126-147, 1932.
4. D. E. Thomas, 'Some design considerations for high frequency transistor amplifiers', *Bell Syst. Tech. J.*, 38, pp. 1551-1580, 1959.
5. J. B. Rodgers, 'The Stability and Power Gain of a Tuned Grounded Emitter Transistor Amplifier Stage', Mullard Research Laboratories Report No. 333, 1960.
6. S. Venkateswaran, 'An invariant stability factor and its physical significance', *Proc. Instn Elect. Engrs*, 109C, pp. 98-102, 1962. (I.E.E. Monograph No. 468E, September 1961.)
7. S. Venkateswaran and A. R. Boothroyd, 'Power gain and bandwidth of tuned transistor amplifier stages', *Proc. Instn Elect. Engrs*, 106B, Suppl. 15, pp. 518-529, January 1960.
8. A. R. Boothroyd, 'The transistor as an active two-port network', *Scientia Electrica*, 7, pp. 3-15, March 1961.

9. S. Venkateswaran, 'Stability and power gain of linear two-port active networks', *The Radio and Electronic Engineer*, 25, p. 276, March 1963.
10. J. M. Rollett, 'Stability and power gain invariants of linear two ports', *Trans. I.R.E. on Circuit Theory*, CT-9, pp. 29-32, March 1962.
11. S. Venkateswaran, 'Stability, power gain and bandwidth of synchronously tuned cascaded linear active two-port networks', *I.R.E. Int. Conv. Rec.*, Part 2, pp. 49-60, 1962.
12. S. Venkateswaran, 'A design basis for synchronously tuned multistage linear amplifiers taking into account spread in device parameters', *The Radio and Electronic Engineer*, 33, pp. 55-64, January 1967.

Manuscript first received by the Institution on 18th January 1967 and in final form on 4th April 1967. (Paper No. 1172.)

© The Institution of Electronic and Radio Engineers, 1968

Discussion on 'Gain and Stability of Transistor Amplifiers'

Mr. M. V. Callendar: The paper by Professor Venkateswaran published in the January 1967 issue of *The Radio and Electronic Engineer*¹² and his later paper which is published in the current issue, covered broadly the same subject—gain and stability of transistor amplifiers—as my own paper on that topic which was published in October 1966.^{1†} But several points of disagreement are evident when comparing the results obtained.

The most noticeable discrepancy stems from the fact that Venkateswaran's design theory, as set out in his first paper, explicitly requires the use of conjugate matching for the 'device' (or 'modified device') under all conditions. It is true that conjugate matching gives the 'maximum available power gain' in the simple case of an unconditionally stable device. But where, as with transistors under most conditions, the device is not unconditionally stable, this simple scheme is no longer directly applicable, and his expedient of adding damping resistors to form a 'modified device' while retaining conjugate matching, does not, in general, give the maximum insertion gain for a single stage for a given stability factor. As a simple example, we may take the final output circuit in the amplifier of his first paper (Fig. 6): the 'damping' conductance ${}^d\rho_L$ is much larger than ρ_{22} and so is roughly equal to the true load ρ_{Lo}/T^2 : if we remove ${}^d\rho_L$ and alter T from 10 to 7, we shall still have the same effective stability, but the power gain is increased by about 3 dB. A corresponding loss exists in the input circuit, making the stage gain in this case about 6 dB less than that with the correct mismatch: in other cases the difference could be greater or less.

In the case of an 'iterative' stage (in a multistage amplifier) mismatching does not give more gain than damping, and so my approach gives, in principle, the same results as his.

There is also a difficulty in comparing results which arises from an apparent inconsistency in the use of symbols in his two papers. From Fig. 3 and eqns. (10) and (11) in his first paper, it is clear that the input and output resistive immittance symbols ${}^d\rho_1$ and ${}^d\rho_2$ do not include the source (ρ_{s0}) and load (ρ_{Lo}), and his s and η are given as functions of ${}^d\rho_1$ and ${}^d\rho_2$. In his second paper, however, he gives eqns. (24) and (25) which are identical to eqns. (10) and (11) in his first paper, except for the substitution of ρ_1

and ρ_2 (defined as including source ρ_s and load ρ_L) for ${}^d\rho_1$ and ${}^d\rho_2$. But he also gives, as eqn. (21),

$$g_{\max} = \frac{1}{s} \cdot \left| \frac{\rho_{21}}{\rho_{12}} \right|$$

which is correct only for the previous definition of s which refers to the 'device' alone ('modified' if necessary) and not to the complete circuit including source and load. (N.B. His g_{\max} does not, in any case, agree with my A_p for reasons given in my second paragraph.) The ambiguity underlying this and other passages radically affects the relations between his symbols and mine: his s , for example, comes out (assuming a fairly high level of stability) roughly equal to half my s_d on his first definition, but about double my s_d on the second definition.

Professor S. Venkateswaran (in reply): A potentially unstable two-port network can be made absolutely stable (a) by unilateralization (also part-unilateralization), or (b) by direct mismatch, or (c) by indirect mismatch. Unilateralization is well known. Direct mismatch is dealt with in detail by Venkateswaran and Boothroyd.⁷ Indirect mismatch is covered in two of my earlier papers.^{11,12} Combinations of part-unilateralization and direct or indirect mismatch have also been used in i.f. amplifiers.

The maximum power gain that is obtainable with direct mismatch is given by eqn. (17) when $n \gg 1$.

If also $n \gg n_1$,

$$g_{\max(n)} \simeq \frac{4}{n} \left| \frac{\rho_{21}}{\rho_{12}} \right|$$

The maximum power gain that is realizable with indirect mismatch is given by eqn. (21). When $s \gg 1$ (say 10) and equal to n , the maximum power gain with direct mismatch is 6 dB higher than that with indirect mismatch for the same stability factor (n or s associated with loop gain). This explains Mr. Callendar's statement on power gain (para. 2).

The measure of non-reciprocity, $\left| \frac{\rho_{21}}{\rho_{12}} \right|$

may be altered by part-unilateralization and/or by bias change of the transistor. Therefore, it is possible to change the maximum power gain for a given stability factor.

† The references cited in this discussion are those given in Professor Venkateswaran's paper in this issue.

However, this gain cannot exceed the unilateral power gain

$$\frac{|p_{21}|^2}{4\rho_{11}\rho_{22}};$$

here the parameters p_{21} etc., refer to the initial device network.

Thus, there is no unique power gain for a given stability factor. This is illustrated by an example of an AF117 transistor amplifier stage in its common-emitter configuration at 455 kHz. At $I_e = 2$ mA and $V_{ce} = -6$ V, device admittance parameters are

$$y_{11} = (510 + j410) \mu\text{mho};$$

$$y_{12} = -j8 \mu\text{mho};$$

$$y_{21} = 70 \mu\text{mho}$$

$$\text{and } y_{22} = (2.5 + j14) \mu\text{mho}.$$

Its $k_1 = 2n_1 = 0.00455$ and $s_1 \approx \pm j1$, so that it is potentially unstable. If this device network is unilateralized, it can realize a maximum power gain of 59.8 dB.

Equation (17) gives a maximum power gain of 35.4 dB for direct mismatch corresponding to $n = 10$. Equation (21) gives a maximum power gain of 29.4 dB for indirect mismatch corresponding to $s = 10$. The corresponding maximum power gains with 50% unilateralization for direct and indirect mismatches (with $n = s = 10$) are 38.4 dB and 32.4 dB respectively. If the emitter current is doubled to 4 mA, the maximum power gains for the above four cases are 38.4 dB, 32.4 dB, 41.4 dB and 35.4 dB respectively. These results were computed and experimentally verified by four post-graduate students of this Department, J. S. Agarwal, P. S. Sarma, K. Kishan Rao and K. Radhakrishna Rao.

In this simple example, power gain has varied from 29.4 dB to 41.4 dB or by 12 dB for the same stability factor. This confirms that there is no unique power gain for a given stability factor.

The definition of s in the present paper is identical with that in my previous paper.¹² ${}^d\rho_1$ of previous paper is ρ_1

here. Similarly ${}^d\rho_2$ of previous paper is ρ_2 here. As a stability factor, s is purely a characteristic of the modified network. However, the maximum power gain as given by eqn. (21) implies conjugate matching of the modified network but resistive mismatching (reactive tuning) of the original device network. Then, s equals the inverse of the modulus of the loop gain of the modified network with its conjugate matching terminations.

The modulus of loop gain of the modified network with its conjugate matching terminations equals 0.1, if $s = 10$. The maximum modulus of loop gain with these terminations equals approximately 0.1 and hence $n \approx 10$. By definition, Callendar's $S_d = 2n$ and hence it is approximately equal to 20 in this situation. I agree with Mr. Callendar on this point.

The maximum interaction between the two ports is expressed in terms of the loop gain's maximum modulus and hence in terms of the performance factor, n , in Section 5 of the present paper.

Three major requirements will play a decisive role in future designs of synchronously tuned amplifiers, including i.f. amplifiers. These are as follows:

(a) Ease in design and fabrication with solid-state devices and components. Identical stages are preferable to non-identical stages all through the chain of stages.

(b) Sensitivity of amplifier performance (gain, bandwidth etc.) to spread in device and circuit parameters must be low. Designs based on average parameters that take into account spread are preferable.

(c) Reliability of the finished product under varying environmental conditions must be high.

Designs that minimize the number of stages or circuit components, or both, are most welcome, provided they also take into account the above three requirements.

I thank Mr. Callendar for initiating a stimulating discussion.

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.

PLASMA AMPLIFIER

Studies of microwave amplification by the interaction of an electron beam and a plasma are described in a Japanese paper. In this programme of work the electron beam was passed through a caesium plasma generated by tungsten heaters in a caesium atmosphere. The frequency of maximum gain was found to increase with the heater temperature. A gain of 24 dB was obtained at 4000 MHz when the temperature of the tube wall was 67°C (vapour pressure of caesium 8×10^{-6} mm Hg).

The plasma density was measured by the double-probe technique, and it was shown that the frequency of maximum gain corresponded to the plasma frequency. The noise figure of the amplifier was measured and found to be in the range of 25 to 35 dB, and was found to be dependent on the pressure of the tube and was little affected by the presence of the plasma.

'Amplification of microwaves by the interaction of an electron beam with a caesium plasma', H. Katoh, K. Ayaki, M. Ozawa, and Y. Asami, *Electronics and Communications in Japan* (English language edition of *Denki Tsushin Gakkai Zasshi*), 49, No. 8, pp. 84-90, August 1966.

RADIO ASTRONOMICAL AND RADAR INVESTIGATIONS OF VENUS

A number of hypotheses have been put forward for the purpose of explaining the high effective brightness temperature of Venus in the centimetre wavelength region. The most important of these hypotheses are the greenhouse hypothesis and the ionospheric hypothesis. According to the former the high temperature of the planet in this wavelength region is due to emission from a heated surface, while the latter is based on the assumption that Venus is enclosed by a dense and hot ionosphere whose opacity in the centimetre region is fairly high as a result of free 'free-electron-transitions'.

In a Soviet paper the ionosphere hypothesis is considered and the radio brightness temperature distribution over the Venus disk at a wavelength of 1.9 cm has been determined for various ratios of the horizontal dimension of ionospheric inhomogeneities to the ionosphere thickness. The results of calculations, performed with allowance for the curvature of the ionosphere, or, in other words, with allowance for the gradual increase in the opacity of the ionospheric 'holes' with increasing distance from the centre of the visible disk of the planet, are shown to contradict the results obtained by scanning the Venus disk at $\lambda = 1.9$ cm by the *Mariner-2* spacecraft.

The correlation between the dense model of the ionosphere and the available radar data is examined. It is

concluded that the dense model cannot explain the presence of radar reflections from the edges of the visible disk of the planet, nor does it explain the existence of strong radar reflection from the point of the planet nearest to Earth. Some views are expressed about the present-day hypothesis of a high electron density in the 'night' ionosphere of Venus.

'The dense model of the Venus ionosphere', G. M. Strelkov, *Radio Engineering and Electronic Physics* (English language edition of *Radiotekhnika i Elektronika*), pp. 721-7, No. 5, May 1967.

GROUP DELAY MEASUREMENTS IN TELEVISION RECEIVERS

In television receivers an important requirement for good fidelity is greater immunity to adjacent channel television signals and/or other strong signals which may lie just outside the wanted channel spectrum. If only minimum phase-shift networks are used, this increased immunity is likely to result in some set-back to picture fidelity unless the special precaution of phase equalization is applied. Without equalization, however, the design of the circuitry of a television receiver to meet a satisfactory standard must involve compromises. The governing factor is that the rate of change of slope of the attenuation/frequency characteristic of the receiver is a measure of the 'group delay' errors introduced.

In practice, the circuit designer has to seek a compromise between a linear phase response and good selectivity. It is therefore desirable that a quick assessment of these two parameters is made.

An Australian paper describes a technique by means of which it is possible to display in a swept manner on an oscilloscope either (a) the total group delay characteristic of a television receiver, or (b) the partial group delay, or phase characteristics of the individual sections which are cascaded in forming the complete r.f., i.f. and video complex of the receiver. This means that the group delay or phase response of the video amplifier along with its amplitude response may be displayed in a swept manner. Similarly, the group delay response of the i.f. amplifier or the r.f. amplifier or both may be displayed conveniently.

The paper also describes a laboratory instrument which has been constructed to achieve these results, and certain design requirements of the various sections of this instrument are indicated. Finally, some practical results obtained with the use of this instrument are given.

'A sweep technique for group delay measurements on TV receivers', T.A. Pascoe, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, 28, No. 10, pp. 398-411, October 1967.

METEOR FREQUENCY TELEGRAPHY CHANNEL

The noise immunity of a meteor communication channel is determined not only by noise itself but also by multipath propagation, which is mainly due to the following factors: the presence of an ionospheric scatter signal, reflection from different parts of the same meteor trail or from several meteor trails occurring simultaneously, and reflection from sporadic layers. The last two types of noise seem to be the most important, in spite of the fact that only part of the meteor burst is subject to their influence. This type of multi-path structure, which can increase the error probability by one order or more, can be combated, according to a Soviet paper, by using circuits that analyse and reject signals formed by several beams. As regards noise and ionospheric scatter, these are always present, and the degree to which they can be suppressed is limited.

The paper assesses the potentialities of a meteor channel by determining its noise immunity and carrying capacity in the presence of noise and ionospheric scatter.

The error probability in a frequency telegraphy meteor channel in the presence of Gaussian noise and Rayleigh-distributed ionospheric scatter signals is determined. The dependence of the channel carrying capacity on the system parameters is also discussed.

'The carrying capacity of a meteor frequency telegraphy channel in the presence of noise and ionospheric scatter signals', Y. F. Korobov, *Telecommunications and Radio Engineering* (English language edition of *Elektrosvyaz and Radiotekhnika*), pp. 24-8, No. 1, January 1967.

PREAMPLIFIERS FOR OPTICAL-BAND RECEIVERS

Effects of a quantum preamplifier in optical (coherent-light) receivers on the following detector or mixer stage are investigated in a Japanese paper. The overall signal/noise ratio is calculated to permit evaluation of the effectiveness of the preamplifier. Calculations employ the photon analysis method, which is based on the fundamental equations for a general quantum-transition system derived by Shimoda. Characteristics of the amplifier, free space, the optical detector and the heterodyne mixer are investigated individually and various combinations of these are considered. A single-mode continuous wave is assumed throughout the paper.

Effects on the preamplifier of the 'dark current' in detectors and mixers, the input signal level, the local-oscillator power, the quantum efficiency, the gain, and the internal noise are calculated quantitatively. It is shown that for mixers there is an optimum amplifier gain at which the signal/noise ratio is maximum.

'Effects of quantum preamplifier in optical-band receivers', N. Hirano, *Electronics and Communications in Japan* (English language edition of *Denki Tsushin Gakkai Zasshi*), 49, No. 10, pp. 63-71, October 1966.

REFERENCE TELEVISION ON TAPE

The principles adopted for defining the recording characteristics of the vision signal in the television tape-recording standards formulated by the European Broadcasting Union are described in a paper (in English) by an engineer with the Italian broadcasting organization; the work was carried out for the Working Party of the E.B.U.

The function of the reference tape and the several methods of specifying such tape for sound recording, where it is possible to measure the surface inductance with precision, are described. This measurement cannot, unfortunately, be made on a television tape-recording because of the arrangement of the tracks, and the only practicable solution is to specify a reference recording and playback chain. The author draws attention to the difficulties encountered in applying this solution, as a result of the design of the modulators and demodulators and of the r.f. sections of the recording and playback chains; and in conclusion it is stated that the standardization of a recording chain is subject to less error than that of a playback chain.

A survey of the results of experiments, indicating the dispersion to be expected (at the higher video frequencies) when using reference tapes recorded under different conditions is given.

'Specification for recording a reference television tape', P. Zaccarian, *E.B.U. Review*, No. 106-A, pp. 234-40, December 1967.

EFFECT OF ATMOSPHERIC REFRACTIVE INDEX ON MICROWAVE FADING

Ray-optical analyses were made by Japanese engineers on the occurrence mechanisms of microwave fading due to duct formation. In a paper describing this work it is shown that attenuation and interference regions are caused by the M-profile curvature, which may produce attenuation- and interference-type fading as the atmosphere varies with time.

The effect of divergence and convergence of waves due to the M-profile curvature was also calculated. Various characteristics of fading can be interpreted by optical analysis of rays. The results of the analyses are compared and are found to be in fairly good agreement with measurement, and have presented effective information for the analysis of multi-path distortion.

Various characteristics of fading can be better understood with such a ray-optical interpretation, and the physical nature of fading phenomenon can be clarified.

'Analyses of microwave fading due to laminar structure of the atmospheric refractive index', F. Ikegami, *Review of the Electrical Communication Laboratory NTT*, 15, Nos. 7-8, pp. 483-506, July-August 1967.

SCATTERING PROPERTIES OF RADAR REFLECTORS

The basic relationship between scattering properties and size of a radar reflector can be used to show that, for technically suitable design forms, the aim should lie on p.p.i. properties. From the principle of physical optics an approximate calculation of the radar cross-section of reflectors is developed in a German paper, and is applied to a triangular triplet reflector and a Luneberg reflector with a special delay. Furthermore, back-scatter measuring equipment designed for the 3 cm band is described and a good agreement between calculation and measurement is proved in the case of several reflector designs.

'A standardized method of investigating the scattering properties of known and new radar reflectors', K. Hoffmann, *Nachrichtentechnische Zeitschrift*, 29, No. 10, pp. 610-15, October 1967.