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 34 No. 3 SEPTEMBER 1967

R.F. Measurements and Standards

KELVIN'S dictum that knowledge is meagre and unsatisfactory unless supported by measurement must surely be appreciated by all engineers: accurate measurement forms the foundation stone of all research, development and production. Like most foundations, however, it is frequently taken for granted, and it is therefore all the more desirable from time to time to reappraise existing methods and techniques of measurement and to look carefully at new proposals for their improvement.

The fundamental importance of measurements to the radio and electronic engineer was one of the reasons which lead the Institution's Programme and Papers Committee to recommend that a three-day Conference on 'R.F. Measurements and Standards' should be held on 14th to 16th November next at the National Physical Laboratory; the Electronics Division of the Institution of Electrical Engineers was invited to be a co-sponsor of the Conference and to appoint members to the Joint Organizing Committee. An outline programme and synopses of some of the papers are printed on pages 131–136 of this *Journal.*

The timeliness of the Conference may be considered to spring from two facts. Firstly, the whole of British industry is regarding the twelve-month period from October 1966 to October 1967 as 'Quality and Reliability Year' and this Conference, coming just after the conclusion of this year of especial emphasis, is a relevant and important contribution on the part of the Institution to the aims of 'QRY'. QRY has been sponsored by the National Council for Quality and Reliability (N.C.Q.R.), on which the Institution has been represented since its formation in 1962. Secondly, there has recently been established a British Calibration Service (B.C.S.), under the Ministry of Technology, which has among its objects the speeding-up of technological advances and the inauguration of a new and controlled measurements policy. On a number of occasions in the past† the Institution has advocated the setting up of a co-ordinated standards organization for electrical quantities at radio frequencies and this new body, with the associated Advisory Council on Calibration and Measurement and its Panel dealing with H.F. Measurements, goes far to meeting a long-standing need. The co-operation of B.C.S. in the organization and programme of the Conference is expecially welcome.

In drawing up the terms of reference of this Conference, the Joint Organizing Committee has laid down two broad conditions. With the exception of one invited paper, contributions on frequency (or time) measurement have been purposely omitted. This particular field of measurement has, it is felt, been covered adequately already through existing national and international organizations and meetings. Papers to be presented in the main cover measurements on attenuation, impedance, power, voltage and noise. Secondly, it has been decided to confine the frequency range of these measurements and standards to the r.f. band 100 kHz to 3 GHz, i.e. the band in which discrete components and coaxial techniques are used.

The significance of the subject of this Conference thus lies both in its inherent importance as a fundamental subdivision within radio and electronic engineering, and in its essential timeliness. The papers and discussions, both formal and informal, should be of considerable value and interest to all engineers concerned with extending quantitative knowledge in radio and electronics.

A. G. WRAY

† See, for instance, 'Radio Frequency Standards', *J. Brit. Instn Radio Engrs*, 21, p. 105, February 1961.

Joint I.E.R.E.–I.E.E. Conference on

'R.F. Measurements and Standards'

National Physical Laboratory, Teddington, Middlesex, 14th to 16th November 1967

PROVISIONAL PROGRAMME AND SYNOPSES OF PAPERS

Final details, which may indicate further papers and/or alterations to the provisional programme shown, will be sent to those registering to attend the Conference.

Tuesday, 14th November (10.30–12.45; 2.30–5.15)

9.00—10.30 a.m. REGISTRATION.

10.30 a.m. **Formal Opening of Conference**

Opening Address by Mr. A. H. A. Wynn, M.A., *Adviser on Engineering Standards, Ministry of Technology.*

Session 1

'A Comparative Survey of U.K. Measurement Capability at Radio Frequencies'—D. P. THURNELL.

'Frequency Standards and Measurements'—DR. L. ESSEN, O.B.E., F.R.S.

'Frequency Deviation Measurements'—P. BRODERICK.

'Standards for Electrical Circuit Properties at Radio Frequencies'—I. A. HARRIS.

Discussion on 'Preferred Frequencies for Measurements'—opened by G. J. HALFORD.

Wednesday, 15th November (9.30–12.45; 2.30–5.15)

Session 2: IMPEDANCE, POWER AND ATTENUATION MEASUREMENTS

'Admittance Measurements on Thin Film Components from 50–1500 MHz'—G. G. BLOODWORTH and A. M. NEGANDHI.

'Admittance Measurements on Monolithic Microcircuits in the range 50–1500 MHz'—H. A. KEMHADJIAN and B. J. LEWIS.

'The Measurement of Lumped Immittance at R.F.'—T. MCCARTNEY.

'A New R.F. Power Meter'—A. A. LUSKOW.

'The Measurement of Attenuation at Radio Frequencies'—M. MCHATTIE.

'A C.W. Comparator for Precision Attenuators'—R. W. A. SIDDLE and I. A. HARRIS.

'An Equipment for the Measurement of Insertion Loss, Gain and Delay over a Wide Range of Amplitude and Frequency'—G. J. CRANK.

'U.H.F. Transmission Loss Measurements'—N. L. MERRITT.

5.45—7.00 p.m. Conference Reception

Thursday, 16th November (9.30–12.45; 2.30–5.15)

Session 3: NOISE AND FIELD STRENGTH MEASUREMENTS

'The Use of Noise Measurements in Radar Receiver Analysis'—E. W. HOUGHTON, R. S. PETERS and M. W. SINCLAIR.

'The Ion Pair Method of Specifying Noise in a Nuclear Pulse Amplifier, in Relation to the Noise Figure and Noise Temperature'—M. O. DEIGHTON.

'Noise Source Calibration in the Decimetre Band'—G. J. HALFORD and E. G. ROBUS.

'A Coaxial Primary Standard for Noise Source Calibration'—R. W. MURRAY.

'Shielded Enclosures'—A. P. HALE.

'Standardization of Radio Interference Measuring Equipment and Techniques'—G. A. JACKSON.

Closing Address by Professor H. M. Barlow, D.Sc., F.R.S.

Synopses of some of the Papers to be presented at the Conference

A Comparative Survey of U.K. Radio Frequency Measurement Capability

D. P. THURNELL, B.Sc.(Eng.), C.Eng., M.I.E.E. (*Deputy Director, British Calibration Service, Ministry of Technology, London.*)

A quantitative assessment of the range and uncertainty of measurement available in United Kingdom R.F. Standards Laboratories. The progress made in the establishment of national standards will be reviewed and the relationship between U.K. achievements and those of other countries, especially in the National Bureau of Standards in the U.S.A., will be discussed.

Future needs will be high-lighted and particular reference will be made to the provision of facilities under the aegis of the British Calibration Service. The planned programme of support at national level by the establishment of improved standards will be outlined.

The data presented will be summarized pictorially. Particular measurement techniques and individual practical standards will not be discussed in detail, but appropriate reference will be made to other papers presented at the Conference or published elsewhere.

Shielded Enclosures

A. P. HALE, C.Eng., M.I.E.R.E. (*Belling & Lee Limited, Enfield, Middlesex.*)

The paper will be non-mathematical and discuss the need for shielded enclosures, with particular emphasis on the measurements field. The basic principles of shielding will be dealt with, followed by a brief review of enclosures commercially available over the past twenty-five years. A more detailed examination of shielding facilities now available will be made considering such things as insertion loss, resonances, ventilation, door design and signal and power line filters. A mention will be made of anechoic chambers.
Finally, the insertion loss testing of shielded enclosures will be described.

Frequency Deviation Measurements

P. BRODERICK. (*Marconi Instruments Ltd., St. Albans, Hertfordshire.*)

The technique used for a frequency modulation measurement depends upon various factors such as the value of modulation index and frequency deviation. Even with the same set of variables more than one method of measurement may be available and the choice of method is decided as much by its ease of application as by its inherent accuracy. Some of the more commonly used methods of frequency modulation measurement are discussed and the conditions under which each is used examined. The errors arising in the use of these methods and the precautions which must be taken to minimize these errors are also examined. Finally the application of these methods in the setting up of standard frequency deviation for the calibration of an f.m. meter is discussed and the sources of error within the meter identified.

Preferred Frequencies for Calibration and Measurement

G. J. HALFORD, B.Sc. (*Services Valve Test Laboratory, Haslemere, Surrey.*)

The choice of preferred frequencies for electrical calibration and measurement is discussed in relation to present practice. A new system of calibration frequencies is proposed based on a series of preferred numbers. It is shown to be applicable to all coaxial and waveguide frequency bands in common use.

A New R.F. Power Meter

A. A. LUSKOW, B.Sc., C.Eng., A.M.I.E.R.E. (*Marconi Instruments Ltd., St. Albans, Hertfordshire.*)

A new range of wideband r.f. wattmeters is discussed. Based on a coaxial thin film thermocouple, the instrument measures incident power as a function of the current passing through the thermocouple and into its internal terminating load. The thermoelectric e.m.f. produced by this square law device is directly proportional to true mean power and therefore independent of waveform distortion. By applying thin film techniques this measurement method can be extended from d.c. into the microwave region for power ranges from 100 mW to 1 kW. The limitations of thermocouple devices and methods of overcoming them are dealt with.

Standards for Electrical Circuit Properties at Radio Frequencies (A Survey of some U.K. Developments)

I. A. HARRIS, C.Eng., M.I.E.E., A.M.I.E.R.E. (*Electrical Inspection Directorate, Ministry of Technology, Harefield, Middlesex.*)

In an introduction, the choice of properties for basic standards and their precise definitions are discussed. The differences in approach with lumped-element circuit language on the one hand, and wave transmission language on the other, are pointed out together with the connection between them. This has a bearing on the choice of measurements and their meaning, especially at medium radio frequencies. The importance of using precision coaxial connectors to realize well-defined cross-sections in a circuit is stressed. The basic properties chosen are impedance, attenuation and power. Frequency and noise level are dealt with elsewhere and are not considered here.

1. *Impedance*

   Bridge methods up to 250 MHz;

   Special standing-wave methods for 0·4 to 3 GHz;

   Precision reflectometer methods.

2. *Attenuation*

   Definitions;

   Standards;

   Methods for comparison;

   Special methods for small attenuations.

3. *Power*

   Definitions;

   Means of referring r.f. power to d.c. power;

   Methods for comparing wattmeters accurately.

The Measurement of Attenuation at Radio Frequencies

J. MCHATTIE, B.Sc. (*Marconi Instruments Limited, St. Albans, Hertfordshire.*)

Various methods of measuring attenuation at radio frequencies are in common use. The capabilities of some of the methods are discussed, together with their advantages, disadvantages and practical precautions necessary to achieve the maximum accuracy, reliability and repeatability.

Most of the techniques discussed have been used by the author for measurements on high frequency attenuators and for the calibration of signal generator output levels. The considerations which led to the adoption of the techniques used are outlined and practical details are given for the test gear layout and methods employed.

An Equipment for the Measurement of Insertion Loss, Gain and Delay over a Wide Range of Amplitude and Frequency

G. J. CRANK, B.Sc.(Eng.), C.Eng., M.I.E.E. (*Post Office Research Station, London.*)

In the frequency range between 1 and 30 MHz there is very little commercial equipment available for making precision loss or gain measurements. This paper discusses the limitations of currently employed measuring techniques in this frequency band.

A general purpose equipment, currently under development, intended for relatively precise measurements over the range 0–30 MHz and for less accurate work at much higher frequencies, is described.

A direct comparison type of measurement is employed using either series or parallel substitution and the relative advantages of these two methods are discussed.

One of the main areas of difficulty has been the realization of good return loss throughout the comparison panel. Particular attention has also been given to minimizing cross-talk between individual units within the panel. Novel features include the design of the fractional dB incremental attenuators and the method for correcting basic attenuator errors.

U.H.F. Transmission Loss Measurements

N. L. MERRITT. (*Ministry of Technology, Signals Research and Development Establishment, Christchurch, Hampshire.*)

Reliable and accurate transmission loss measurements were needed in order to establish the validity of theoretical path loss prediction methods proposed for u.h.f. communication links (with losses up to 180dB). Fluctuations due to tropospheric scatter and reflections from elevated inversions are found on these high loss paths, seasonal variations are also possible.

To save setting up a number of permanent links mobile equipment which could have its sensitivity standardized was necessary. The bandwidth of the system was 3 kHz and at u.h.f. this calls for the use of stable crystal oscillators. Height-loss patterns were required and to facilitate their measurement pneumatic masts were used to give continuously variable height adjustment.

The use of mobile equipment could only be considered provided the risks of inaccuracies due to poor connections and variable equipment performance were minimized. This was achieved by observing extreme care in the handling of aerials and connectors, and by making the electronic equipment virtually a fixed installation shock mounted within a completely screened metal vehicle; power supplies were stabilized and only experienced staff were used to man the terminals. Consistency of calibration improved with time and practice. The local field patterns were inspected before a measurement site was decided on and wherever possible sites where large or complicated electromagnetic field gradients existed were avoided.

Standardization of Power Output: The measurement of transmitter power appeared to be straightforward with standard broadband wattmeters and attenuators but the transistor-varactor amplifier-multiplier proved to be capable of producing the majority of its power over a 500 kHz bandwidth which gave rise to an apparent loss of 20 dB on the narrow band receiver. This necessitated inspection of the waveform as well as r.m.s. power.

Standardization of Sensitivity: Because of the narrow band of the receiver used (3 kHz) the only signal generator that one could use was a crystal-controlled oscillator with the same specification as the transmitter and the oscillator was used before and after each day's measurements, taking care to eliminate inaccuracies due to leakage. At each site the receiver was calibrated with a noise generator (assuming the bandwidth to remain constant). The complete system including the aerials was regularly calibrated on a flat airfield runway well away from any potential source of field pattern distortion or interference.

The results obtained were standardized to within ½ dB and established the prediction techniques to be better than 5 dB on the majority of profiles measured.

A C.W. Comparator for Precision Attenuators

R. W. A. SIDDLE, C.Eng., A.M.I.E.R.E. and I.A. HARRIS, C.Eng., M.I.E.E., A.M.I.E.R.E. (*Electrical Inspection Directorate, Ministry of Technology, Harefield, Middlesex.*)

The most accurate piston attenuators work at frequencies below 100 MHz and standard piston attenuators are usually operated at a fixed frequency such as 30 to 60 MHz. In designing a comparator in which attenuators operating at any frequency up to, say, 3 GHz can be compared with a standard over a wide range of attenuation up to 100 dB, a number of problems have to be solved. The paper describes how these problems were taken into account in designing a comparator which would function with an unmodulated c.w. r.f. source and would cover up to 100 dB attenuation without producing an uncertainty of comparison worse than 0·01 dB.

The problems are:

(a) Choice of 'series' or 'parallel' method: the parallel method is adopted.

(b) Choice of square-wave modulation or c.w. with manual switching between 'unknown' and 'standard' circuit channels: reasons are given why the c.w. method is adopted, and the resulting requirements on stability and the means adopted to achieve them are stated.

(c) The design of frequency changers is conditioned by a stringent requirement of linearity over the range —7 dBm to —107 dBm and the means adopted to achieve this are described.

(d) The noise level must be sufficiently below the c.w. level at —107 dBm. The manner in which this is achieved is described.

The overall performance of the comparator is such that the uncertainty in comparison is well within ±0·01 dB over the whole range of attenuation, for frequencies up to 1 GHz.

Admittance Measurements on Monolithic Microcircuits in the range 50–1500 MHz

H. A. KEMHADJIAN, M.Sc. and B. J. LEWIS, B.Sc. (*University of Southampton.*)

Measurements on integrated circuits involve special problems because of the very small size of the components. In order to make accurate measurements at high frequencies it is essential that the connecting leads introduce negligible or at least predictable effects. This short contribution describes a stripline jig used to connect silicon monolithic circuitry to a GR transfer bridge and the results of some preliminary measurements on resistors with an estimate of the errors.

Admittance Measurements on Thin Film Components from 50–1500 MHz

G. G. Bloodworth, M.A., D.U.S., C.Eng., A.M.I.E.R.E. and A. M. Negandhi, M.Sc., D.U.S. (*University of Southampton.*)

The admittance of typical thin-film capacitors in the frequency range 50–1500 MHz is strongly dependent on the shape and thickness of the electrodes. Usually the loss factor is mainly determined by the resistance of the electrodes, which is not increased significantly by the skin effect. The electrode inductance resonates with the capacitance at a frequency which is often in this range. The paper discusses the measurement of admittance by connecting capacitors to an admittance bridge by wires and by strip-lines. The losses due to the thin-film dielectric can be measured using a direct connection to the coaxial bridge of capacitors fabricated with circular symmetry, thus minimizing the electrode effects. The loss factor for silicon oxide films formed by evaporation in vacuo is being investigated as a function of physical composition. A similar arrangement is used to measure the impedance of thin-film resistors, to eliminate the inductance of the connections to the bridge.

The Measurement of Lumped Immittance at R.F.

T. McCartney, C.Eng., M.I.E.E., A.M.I.E.R.E. (*The Wayne Kerr Company Limited, Bognor Regis, Sussex.*)

The paper covers the general approach to the measurement of lumped elements at r.f. The term 'lumped' implies that the elements have electromagnetic field and lead configurations such that the immittance is definite, independent of the surrounding environment and dimensionally small compared with the wavelength.

For a working device, a complete current circuit is essential and measurement of the elements should be made without disturbing the circuit. In practise the isolation of what is to be measured from irrelevant influences is immensely facilitated by three and four terminal measuring techniques, which in fact, are almost necessary for uncertainties of less than 1%.

Recent developments employing the above techniques for measuring capacitance and conductance between 100 KHz and 1 MHz with an uncertainty of 0·1% are described.

The Use of Noise Measurements in Radar Receiver Analysis

E. W. Houghton, R. S. Peters and M. W. Sinclair. (*Ministry of Technology, Royal Radar Establishment, Great Malvern, Worcestershire.*)

The characteristics of band-limited Gaussian noise are employed to measure the signal transmission properties of non-coherent pulse radar systems used for evaluating backscatter signals from ground, sea and precipitation clutter targets.

Both backscatter and noise signals can be characterized by their amplitude distributions and their auto-correlation functions. The amplitude distribution and auto-correlation function of band-limited Gaussian noise can be calculated at various points along a receiver system and these characteristics can be compared with measured results. This comparison between measured and expected results can be used for tests of radar receiving performance.

A scheme using a calibrated microwave or i.f. noise source and amplitude and spectrum (in a stationary process like band-limited Gaussian noise, the spectrum function is the Fourier transform of the auto-correlation function) measuring techniques will be described. These methods will be used to measure the characteristics of radar receiving systems fitted with logarithmic amplifiers. Theoretical and measured results for such receiving systems will be compared, and effect of distortions explained.

Noise Source Calibration in the Decimetre Band

G. J. Halford, B.Sc. and E. G. Robus. (*Services Valve Test Laboratory, Haslemere, Surrey.*)

This paper discusses the problems involved in the design of noise comparators and standards for the frequency range 300–3000 MHz.

The problems are illustrated by reference to an existing noise comparator built for the frequency range 300–1000 MHz, and to an experimental system for the range 1–2 GHz.

Thermal noise standards are considered with reference to an existing 1000°C coaxial design, employed at the lower frequencies. A projected 400°C model, based on an S-band design, and more suitable for the higher frequencies, is also discussed.

Experimental results are given for the calibration of various types of broadband coaxial noise sources, demonstrating the need for accurate calibration facilities.

The Ion Pair Method of Specifying Noise in a Nuclear Pulse Amplifier in Relation to the Noise Figure and Noise Temperature

M. O. Deighton, B.Sc., C.Eng., M.I.E.E. (*Atomic Energy Research Establishment, Harwell, Berkshire.*)

For many years communication engineers have described the noise level of their amplifiers in terms of a noise figure, expressed in decibels, or latterly in terms of noise temperature; both are related to source resistance. Nuclear engineers, on the other hand, normally operate their low-noise pulse amplifiers with an almost purely capacitive source, like the particle detector. Here the noise figure or temperature is an apparently meaningless concept, hence the customary use of equivalent noise charge, usually quoted in ion pairs, as a measure of amplifier noise.

The two-generator (e_n, i_n) model for noise, which has gained wide acceptance, does not depend on the source impedance and thus provides a means for comparing the two methods of specifying noise. This paper derives the direct quantitative relationships between noise figure and noise ion pairs and represents a first step to breaking down the language barrier, which has persisted between nuclear and communication engineers on this subject. Numerical values underline the very low noise levels demanded for present-day nuclear particle spectrometry and suggest that measurement of noise charge is inherently more accurate than noise figure, because the source itself is noiseless. The method might therefore be usefully employed in fields other than nuclear physics.

A Coaxial Primary Standard for Noise Source Calibration

R. W. Murray, M.A. (*Ferranti Ltd., Edinburgh.*)

Constructional details are given of a coaxial hot load noise source designed to act as a primary standard for noise calibration at frequencies in the 1 to 3 GHz range. Several samples were made and the best of these gave a match at the output terminals better than 1·25 v.s.w.r. over the frequency range d.c. to 5 GHz at its operating temperature of 1225°K. Covering this frequency range it can thus be compared with low-frequency precision noise diodes and high-frequency gas discharge noise tubes. The device consists of a pyrolitically-deposited carbon film resistor matched to a 50-ohm coaxial line by means of a taper and heated in a vacuum envelope by means of a molybdenum wire-wound heater. The stability of the resistor film is better than 0·1% over prolonged periods of operation and temperature cycling. By careful arrangement of heat shields and the use of thin-walled tubing to form the coaxial line, heat losses were reduced to a minimum and the operating temperature could be achieved with the dissipation of only 90 watts of heater power. The overall dimensions of the device are 12 inches long by 3 inches diameter.

Standardization of Radio Interference Measuring Equipment and Techniques

G. A. Jackson, B.Sc., C.Eng., M.I.E.E. (*The Electrical Research Association, Leatherhead, Surrey.*)

Radio interference investigations involve the measurement of the effects of complex waveforms on circuits in which the r.f. impedances are often ill-defined and may vary considerably with frequency. It is necessary to specify certain parameters of the measuring receivers within close limits and also to define the conditions under which measurements are made, in order that consistent and repeatable results can be obtained. The paper surveys types of measuring equipment which are in common use such as peak and quasi-peak sets and briefly refers to receivers having r.m.s. and averages responses which may be used for specialized investigations.

The ultimate calibration of all radio interference measuring receivers is made in terms of sine-wave signals and the standard signal generators used for this purpose can in turn be checked to a high degree of accuracy by thermistor bridge methods. The performance of the interference measuring equipment is also specified in terms of the response to regularly repeated pulses and it is in this respect that the parameters of the receivers such as bandwidth, time-constants and overload factor are extremely important.

The paper considers the essential characteristics of the calibrating pulse generators and the methods which are used to ensure that the spectral content of the pulse output is satisfactory for the frequency range under consideration.

Finally, in the measurement of conducted or radiated interference it is essential that field or circuit conditions must be specified, especially where compliance with regulations is required. The methods by which the specified conditions are achieved are described and discussed.

INSTITUTION NOTICES

The Institution in Africa

Mr. Graham D. Clifford, C.M.G., the Director and Secretary of the Institution, left London for Nairobi on 15th September. His visit to East Africa will now be confined to Kenya where he will visit broadcasting and telecommunications organizations and educational establishments, and discuss with members the feasibility of establishing an Institution Section in Nairobi. Mr. Clifford will also address a joint meeting of members of the I.E.R.E. and other Institutions.

The Institution has not been able to arrange for any officer of the Institution to visit South Africa since 1949, when Mr. Leslie McMichael, then an Immediate Past President of the Institution, undertook an extentensive tour of the main centres of South Africa. It is recorded in the history of the Institution† that Mr. McMichael, in a written report to the Council dated 24th July 1949, stated that "Interest in the work and future of the Institution is extremely keen and on my return [to Great Britain] I would like to assist in establishing an Institution Section in the Union'. Later that year, the Council did, in fact, authorize the formation of an Institution Section in South Africa.

From Nairobi Mr. Clifford will proceed to Johannesburg and whilst in South Africa will visit the South African Broadcasting Corporation, as well as other organizations which Mr. McMichael mentioned in his report. The South African Section Committee has also arranged meetings in Pretoria, Johannesburg and Cape Town.

Mr. Clifford will return to London on 16th October.

Forthcoming Conference on Electronics Design

The valuable contributions and discussions at the recent Joint I.E.R.E.–I.Prod.E.–I.E.E. Conference in Nottingham on 'The Integration of Design and Production in the Electronics Industry' suggested to many of those attending the desirability of following up some of the subjects of especial interest, for instance design management, by holding further meetings on a particular range of topics.

This possibility had in fact been discussed informally between the I.E.R.E. and the Institution of Electrical Engineers during the planning period of the Nottingham Conference, and it has now been agreed to promote a Joint Conference on Electronics Design during 1968. Members who attended the recent Conference will have received a questionnaire from the I.E.E., inviting suggestions for topics to be included within the scope of the 1968 Conference. Return of the completed questionnaires to the I.E.E., who will be responsible for the secretariat of the Conference, will materially assist the Joint Organizing Committee in planning a programme which will be of the widest possible value.

Further information, including date and venue and outline programme, will be published in due course by the two Institutions.

† 'A Twentieth Century Professional Institution: The Story of the Brit. I.R.E.' Published by the Institution 1960, price 30s.

Second N.Z. National Electronics Convention

The Second National Electronics Convention, organized by the New Zealand Section of the I.E.R.E. and the New Zealand Electronics Institute, Inc., will be held from 20th to 23rd August, 1968, at the University of Auckland.

The Convention will include five sections dealing respectively with Components and Instruments, Applied Electronics, Research Electronics, Communications, and Data Handling; and three Symposia entitled 'Communications for Computers', 'Integrated Circuits and the Electronics Designer' and 'Automation for Primary Production in New Zealand'.

Offers of papers are invited and authors are asked to indicate the particular section for which they consider their paper would be suitable. Details of papers should be sent to Mr. C. W. Salmon, New Zealand Section, Institution of Electronic and Radio Engineers, P.O. Box 3381, Auckland, from whom further information may be obtained.

Corrections

The following correction should be made to the paper 'Amplitude Limiting Applied to a Sensitive Correlation Detector', which was published in the July 1967 issue of *The Radio and Electronic Engineer*:

Page 8: Equation (13) and the caption to Fig. 2(c) should both read:

$$I = I_i \sigma_i \operatorname{erf} \frac{V}{\sqrt{2}\sigma_i}$$

Some of the copies of the August 1967 issue of *The Radio and Electronic Engineer* contain errors which arose during printing. In the paper 'Voltage Stabilized Sinusoidal Inverters using Transistors' the following corrections may, therefore, be necessary:

Page 122, left-hand column, Appendix 1: In the third and sixth equations, the letter f is missing in two places, and the equations should read, respectively,

$$f_a = f_0(1+x)$$
$$f_b = f_0.$$

U.D.C. 621.382.2:621.373.51

The Step-recovery Diode in Harmonic Generation: A Linear Time-varying Circuit Model

By

J. G. GARDINER, B.Sc., Ph.D.

(*Graduate*)†

Summary: A circuit model of the step-recovery diode is proposed based on the known conduction properties of the device. This model permits analysis of harmonic generation circuits with greater generality than has hitherto been possible and is used to estimate the efficiency and optimum terminating conditions of a high-order frequency multiplier.

List of Principal Symbols

i_d	total current in diode
i_0	component of i_d at input frequency
i_m	component of i_d at harmonic frequency
m	order of multiplication
n	quality factor of diode $= \sqrt{r_b/r_f}$
r_f	resistance of diode in 'on' condition
r_b	resistance of diode in 'off' condition
$r_d(t)$	time-varying resistance of diode
s	ratio τ/t_0
t_0	period of input cycle
τ	duration of 'off' condition
V_0	source generator voltage
$Z_0 = R_0+jX_0$	impedance of source generator
$Z_m = R_m+jX_m$	impedance of load
ω	input angular frequency
ϕ	phase angle separating input current and fundamental component of $r_d(t)$

1. Introduction

Several analyses of harmonic generation circuits using the step-recovery diode have been made recently[1-3] but these have been to some extent restricted in their generality of application by their dependence upon some knowledge of the waveforms of currents flowing through the diode.

In the case of the series-diode arrangement of Fig. 1, the mechanism of harmonic generation is apparently to be found in the abrupt snap-off of reverse current which excites the output tuned circuit. From analysis of the harmonic content of the current 'step' predictions of efficiency can be made. However, in the case of the shunt configuration of Fig. 2 more difficulty arises since the tuned circuits employed restrict the harmonic content of the diode current too severely to allow a current step to exist.

The latter circuit configuration is of particular practical significance for two reasons; first, it permits use of the diode with one electrode grounded, a valuable feature in high power multipliers, and second, it is in this arrangement that step-recovery enhancement of conventional varactor harmonic generation is usually obtained.

A recent analysis has considered a similar circuit arrangement from the voltage and current waveform point of view,[4] but it is none the less apparent that if useful predictions of optimum terminations, bias conditions etc are to be made, a more general representation of the diode is needed which permits analysis without detailed knowledge of the current and voltage waveforms at the diode terminals. It is the purpose of this paper to indicate the possibilities of a very simple representation based on what is known as the conduction process in step-recovery diodes.

2. The Step-recovery Diode as a Switch

The essential feature of the step-recovery phenomenon is to be found in the turn-off transient of a suitably graded p-n junction. When a high frequency supply is connected to the device, conduction takes place while the diode is forward biased and minority carriers are being stored in the base region and continues after the bias polarity is reversed until all the stored carriers have been reclaimed. If the minority carrier lifetime in the base semiconductor is long relative to the cycle time of the applied signal very few carriers are lost by recombination.

The device behaves, in fact, as a switch, 'on' under forward bias or when containing stored carriers and 'off' at all other times. Let us suppose therefore, that this switching action takes place between two states of linear resistance, r_f when turned on and r_b when turned off, and re-examine the circuits of Figs. 1 and 2 considering the diode as a switch and that the time of switching is determined by the flow of charge through the diode. It is apparent that in the circuit of Fig. 1 the combination of d.c. bias and input and output signals results in forward bias for some portion of the input period followed by reverse conduction. When the diode turns off a step of current is produced.

† Formerly at the Department of Electronic and Electrical Engineering, University of Birmingham; now with Racal Research Ltd., Tewkesbury, Gloucestershire.

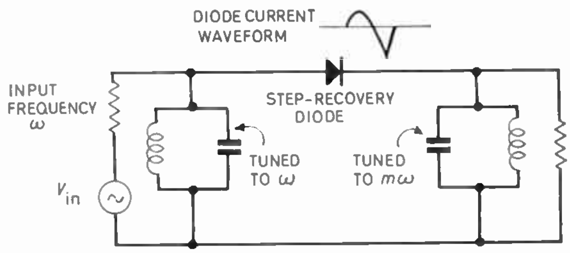


Fig. 1. Series diode harmonic generator.

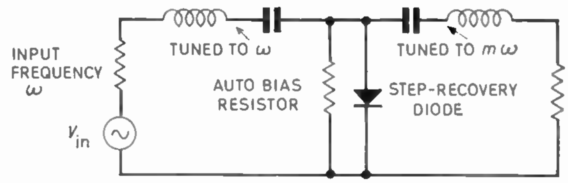


Fig. 2. Shunt diode harmonic generator.

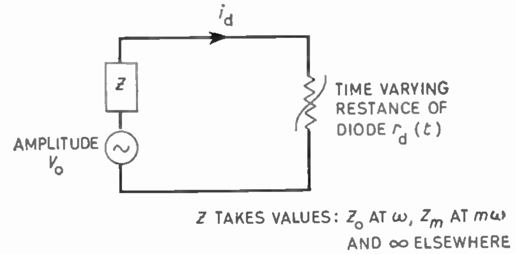


Fig. 3. Equivalent circuit of generator shown in Fig. 2.

In the circuit of Fig. 2 when conduction ceases no step-in current occurs since the current components are restricted by the tuned circuits, instead the change of diode state is accompanied by a voltage step across the diode.

In other words, the change of state of the diode may be regarded as the primary source of harmonic generation and the steps in voltage or current which appear in the circuitry are merely manifestations of this mechanism. In the following Sections it will be shown that the performance of the circuit of Fig. 2 can be analysed solely in terms of the diode switching function and the number of currents allowed to flow by the tuned circuits.

3. Circuit Analysis of the Shunt-diode Harmonic Generator

3.1 *Calculation of Efficiency*

The circuit of Fig. 2 may be represented by the equivalent circuit of Fig. 3. The diode current is restricted by the tuned circuits to be made up of only two frequency components at the input and wanted harmonic frequencies. This current may be written

$$i_d = i_0 e^{j\omega t} + i_0^* e^{-j\omega t} + i_m e^{jm\omega t} + i_m^* e^{-jm\omega t} \quad \text{...(1)}$$

where m is the order of multiplication.

The diode is considered to be switched off for a fraction s of the input signal period and therefore possesses a switching function

$$r_d = r_f + s(r_b - r_f) + \frac{2(r_b - r_f)}{\pi} \sum_{m=1}^{\infty} \frac{\sin m\pi s}{m} \cos m\omega t \quad (2)$$

The time of switching may be related to the input current waveform by a phase angle ϕ so that the time variation of the diode becomes

$$r_d(t) = \sum_{m=-\infty}^{\infty} r_m e^{jm(\omega t+\phi)} \quad \text{......(3)}$$

Considering i_0 to be the phase reference we may define this as an even function, i.e. $i_0 = i_0^*$. Therefore, multiplying eqns. (1) and (3) together and collecting terms at the two frequencies in question the circuit-equation for Fig. 3 becomes

$$V_0 = i_0(R_0 + jX_0 + r_0 + r_2 e^{j2\phi}) + i_m(r_{m-1} e^{-j\overline{m-1}\phi}) + i_m^*(r_{m+1} e^{j\overline{m+1}\phi}) \quad \text{......(4)}$$

$$0 = i_0(r_{m-1} e^{j\overline{m-1}\phi} + r_{m+1} e^{j\overline{m+1}\phi}) + i_m(R_m + jX_m + r_0) + i_m^*(r_{2m} e^{j2m\phi}) \quad \text{......(5)}$$

Since the phase of V_0 relative to i_0 is unknown it is necessary to define

$$V_0 = V_R + jV_I \quad \text{......(6)}$$

and

$$i_m = a_m + jb_m$$
$$i_m^* = a_m - jb_m \quad \text{......(7)}$$

The real and imaginary parts of eqns. (4) and (5) may be equated to give

$$V_R = i_0 A_1 + a_m B_1 + b_m C_1 \quad \text{......(8)}$$
$$V_I = i_0 A_2 + a_m B_2 + b_m C_2 \quad \text{......(9)}$$
$$0 = i_0 A_3 + a_m B_3 + b_m C_3 \quad \text{......(10)}$$
$$0 = i_0 A_4 + a_m B_4 + b_m C_4 \quad \text{......(11)}$$

where A, B and C coefficients are defined in the Appendix. Solving for i_0, a_m and b_m gives

$$i_0 = \begin{vmatrix} B_3 & C_3 \\ B_4 & C_4 \end{vmatrix} \frac{V_R + V_I}{\Delta_1} \quad \text{......(12)}$$

$$a_m = \begin{vmatrix} A_3 & C_3 \\ A_4 & C_4 \end{vmatrix} \frac{-(V_R + V_I)}{\Delta_1} \quad \text{......(13)}$$

$$b_m = \begin{vmatrix} A_3 & B_3 \\ A_4 & B_4 \end{vmatrix} \frac{V_R + V_I}{\Delta_1} \quad \text{......(14)}$$

where

$$\Delta_1 = \begin{vmatrix} A_1 + A_2 & B_1 + B_2 & C_1 + C_2 \\ A_3 & B_3 & C_3 \\ A_4 & B_4 & C_4 \end{vmatrix} \quad \text{...(15)}$$

or

$$i_0 = K_{11}(V_R + V_I) \quad \ldots\ldots(16)$$

$$a_m = K_{21}(V_R + V_I) \quad \ldots\ldots(17)$$

$$b_m = K_{31}(V_R + V_I) \quad \ldots\ldots(18)$$

Similarly from eqns. (8), (9), (10) and (11) it is possible to write

$$i_0 = K_{12}(V_R - V_I) \quad \ldots\ldots(19)$$

$$a_m = K_{22}(V_R - V_I) \quad \ldots\ldots(20)$$

$$b_m = K_{32}(V_R - V_I) \quad \ldots\ldots(21)$$

where K_{12}, K_{22} and K_{32} are defined by

$$K_{12} = \begin{vmatrix} B_3 & C_3 \\ B_4 & C_4 \end{vmatrix} \Big/ \Delta_2 \quad \ldots\ldots(22)$$

$$K_{22} = -\begin{vmatrix} A_3 & C_3 \\ A_4 & C_4 \end{vmatrix} \Big/ \Delta_2 \quad \ldots\ldots(23)$$

$$K_{32} = \begin{vmatrix} A_3 & B_3 \\ A_4 & B_4 \end{vmatrix} \Big/ \Delta_2 \quad \ldots\ldots(24)$$

and

$$\Delta_2 = \begin{vmatrix} A_1 - A_2 & B_1 - B_2 & C_1 - C_2 \\ A_3 & B_3 & C_3 \\ A_4 & B_4 & C_4 \end{vmatrix} \quad \ldots(25)$$

Now from eqns. (17) and (20)

$$V_I = \gamma V_R \quad \ldots\ldots(26)$$

where

$$\gamma = \frac{K_{12} - K_{11}}{K_{12} + K_{11}} \quad \ldots\ldots(27)$$

From eqn. (27), the efficiency of harmonic generation may now be written as:

$$\eta = \frac{\text{power in load at harmonic frequency}}{\text{maximum available power from source}}$$

i.e.

$$\eta = 4R_m R_0 \left| \frac{\sqrt{a_m^2 + b_m^2}}{\sqrt{V_R^2 + V_I^2}} \right|^2 \quad \ldots\ldots(28)$$

which expressed as a percentage is

$$\eta = 400 R_m R_0 \left[\frac{(1+\gamma)^2 (K_{21}^2 + K_{32}^2)}{(1+\gamma^2)} \right] \% \quad (29)$$

3.2 *Calculation of Input and Output Impedances*

The input impedance is readily obtained from eqns. (26), (27) and (16) since

$$Z_{in} = \frac{V_R + jV_I}{i_0} \quad \ldots\ldots(30)$$

$$Z_{in} = \frac{1}{K_{11}(1+\gamma)} + j\frac{\gamma}{K_{11}(1+\gamma)} \quad \ldots\ldots(31)$$

The output impedance is obtained in similar fashion by impressing a voltage (V_m) at the harmonic frequency, then

$$V_m = V_{mR} + jV_{mI} \quad \ldots\ldots(32)$$

$$i_0 = K_{13}(V_{mR} + V_{mI}) \quad \ldots\ldots(33)$$

$$a_m = K_{23}(V_{mR} + V_{mI}) \quad \ldots\ldots(34)$$

$$b_m = K_{33}(V_{mR} + V_{mI}) \quad \ldots\ldots(35)$$

where

$$K_{13} = \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix} \Big/ \Delta_3 \quad \ldots\ldots(36)$$

$$K_{23} = -\begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix} \Big/ \Delta_3 \quad \ldots\ldots(37)$$

$$K_{33} = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} \Big/ \Delta_3 \quad \ldots\ldots(38)$$

Similarly

$$i_0 = K_{14}(V_{mR} - V_{mI}) \quad \ldots\ldots(39)$$

$$a_m = K_{24}(V_{mR} - V_{mI}) \quad \ldots\ldots(40)$$

$$b_m = K_{34}(V_{mR} - V_{mI}) \quad \ldots\ldots(41)$$

where

$$K_{14} = \begin{vmatrix} B_1 & C_1 \\ B_2 & C_2 \end{vmatrix} \Big/ \Delta_4 \quad \ldots\ldots(42)$$

$$K_{24} = -\begin{vmatrix} A_1 & C_1 \\ A_2 & C_2 \end{vmatrix} \Big/ \Delta_4 \quad \ldots\ldots(43)$$

$$K_{34} = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} \Big/ \Delta_4 \quad \ldots\ldots(44)$$

and

$$\Delta_3 = \begin{vmatrix} A_3 + A_4 & B_3 + B_4 & C_3 + C_4 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix} \quad \ldots(45)$$

$$\Delta_4 = \begin{vmatrix} A_3 - A_4 & B_3 - B_4 & C_3 - C_4 \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix} \quad \ldots(46)$$

Then as before

$$\frac{V_{mI}}{V_{mR}} = \gamma' = \frac{\Delta_3 - \Delta_4}{\Delta_3 + \Delta_4} \quad \ldots\ldots(47)$$

and

$$a_m = K_{23} V_{mR}(1+\gamma') \quad \ldots\ldots(48)$$

$$b_m = K_{33} V_{mR}\ 1+\gamma') \quad \ldots\ldots(49)$$

which results in

$$Z_{out} = \frac{V_m}{i_m} = \frac{1}{(1+\gamma')(K_{23}^2 + K_{33}^2)} \times$$

$$\times \left[(K_{23} + K_{33}\gamma') + j(K_{23}\gamma' - K_{33}) \right] \quad \ldots\ldots(50)$$

4. Evaluation of Maximum Efficiency and Optimum Terminations

So far the analysis has been concerned with a device switching between two linear resistance states but it must be remembered that the multiplier itself is a non-linear circuit and there are three important ways

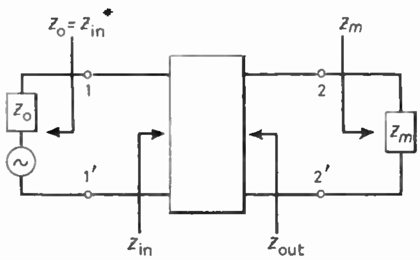


Fig. 4. Optimum matching conditions.

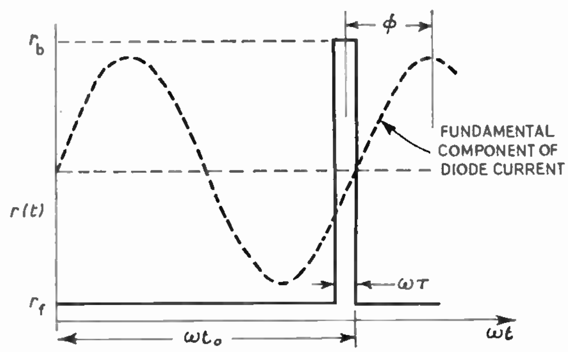


Fig. 5. Switching function of diode.

in which the necessity of maintaining this non-linearity influences evaluation of eqns. (29), (31) and (50).

(i) The point at which the diode switches is determined by the signals across the terminals so that s and ϕ in eqns. (2) and (3) are interdependent.

(ii) The generated current at the wanted harmonic circulates through the switch and this together with the phase relationships of the practical circuit prevents the output impedance from being defined solely in terms of the input termination and switch parameters. In other words the output impedance looking into the terminals 2 2′ in Fig. 4 is dependent upon the load which the circuit works into. This in turn means that a conjugate match at the output is not (in general) the condition which will result in maximum efficiency of generation.

(iii) The value of the phase angle ϕ is also dependent on the applied signals and d.c. bias.

The conditions under which the maximum efficiency is obtained for a given set of conditions of diode quality and bias are summarized in Figs. 4 and 5. It is considered necessary to specify a conjugate match at the input in order to obtain maximum transfer of power from the source (assumed to possess linear output impedance) into the multiplier circuit. The diode itself is assumed to be auto-biased so that some d.c. leakage takes place and this controls the amount of stored charge lost from the diode before being reclaimed during reverse conduction. Thus, the combined input and generated signal currents are assumed to cause the switch to be turned on for most of the input cycle, the transition to high resistance occurring, as shown in Fig. 5, almost at the completion of this cycle. In this way ϕ is defined for a given value of s.

In view of the complexity of the expressions involved analytical optimization of eqns. (29), (31) and (50) is clearly not a practical proposition. In practice the process of tuning up a multiplier on the bench is an iterative one, successive tuning of output termination and input match, and this process has been simulated on the computer. The program used adjusts first the reactive part of the output termination for maximum efficiency, then the resistive part, then restores a conjugate match at the input and repeats the cycle until no further improvement in performance can be obtained.

5. Results

Theoretical evaluation has been concentrated on high-order generation in the main, i.e. 4th, 8th, 10th, 14th and 18th harmonics and the results are summarized in the graphs of Figs. 6–11.

Provided the generator terminations are tuned to optimum values the efficiency of generation is governed by the conduction time and diode quality factor, n. The latter quantity is a measure of the resistance change offered by the diode at switching and is defined by

$$n = \sqrt{r_b/r_f} \qquad \ldots\ldots(51)$$

Thus for a diode of given n-factor an optimum conduction time exists which is different for each order of multiplication. These conditions are indicated in Figs. 6 and 7.

The graph of Fig. 8 shows the dependence of available efficiency at a given harmonic on the diode conduction time and indicates the sensitivity of the multiplier to this parameter. A change of 1% in conduction-time results in a loss of efficiency of 10% for the $n = 100$ diode. (This is in agreement with the experimentally observed sensitivity of step-recovery multipliers to changes in ambient temperature etc.).

In Fig. 9 the maximum efficiencies obtainable from the generator at various harmonic orders are plotted for diodes of various n-factor. It is of particular interest to compare the results of the theory for an $n = 50$ diode with the dotted curve. This curve represents the published experimental results obtained by Hall[5] for a step-recovery generator using HPA 0241-type diodes in a circuit similar to that shown in Fig. 1.

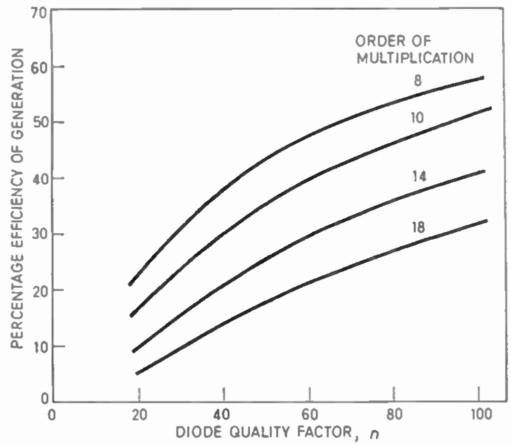


Fig. 6. Variation of maximum efficiency with n-factor of diode.

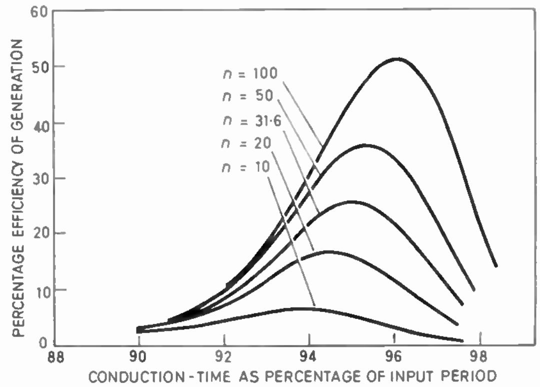


Fig. 8. Generation of 10th harmonic with diodes of various values of quality-factor.

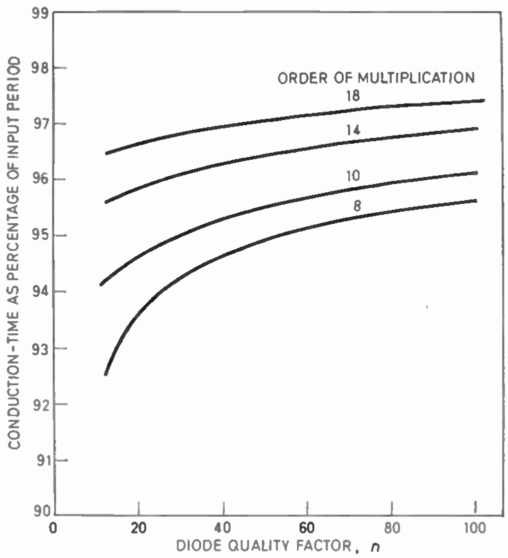


Fig. 7. Variation of optimum conduction time with diode quality factor.

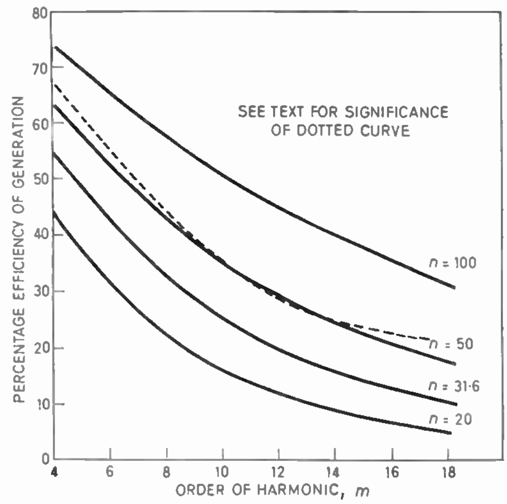


Fig. 9. Maximum efficiencies with increasing harmonic order for different values of quality factor.

The remaining curves of Figs. 10 and 11 indicate the input, output and optimum load impedances of a generator using diodes with 1-ohm forward resistance. The interesting features of these graphs are:

(i) The input impedance is almost entirely capacitive even for a diode of $n = 100$.

(ii) Significant differences exist between input and output impedances and between output and optimum terminating impedances. It was indicated earlier that a conjugate match should not be expected to yield the optimum condition and appreciable departures from this are seen to appear particularly in the cases of high-order generation using low n-factor diodes.

6. Conclusions

A number of aspects of performance of practical step-recovery diode multipliers, particularly those using the circuit configuration treated here, have been difficult to explain in the light of previous theories. These are:

(1) the shape of the curve of efficiency against harmonic number,

(2) the unusually low input impedance apparently presented by the multiplier to the source,

(3) the sensitivity of this type of multiplier to changes in ambient conditions.

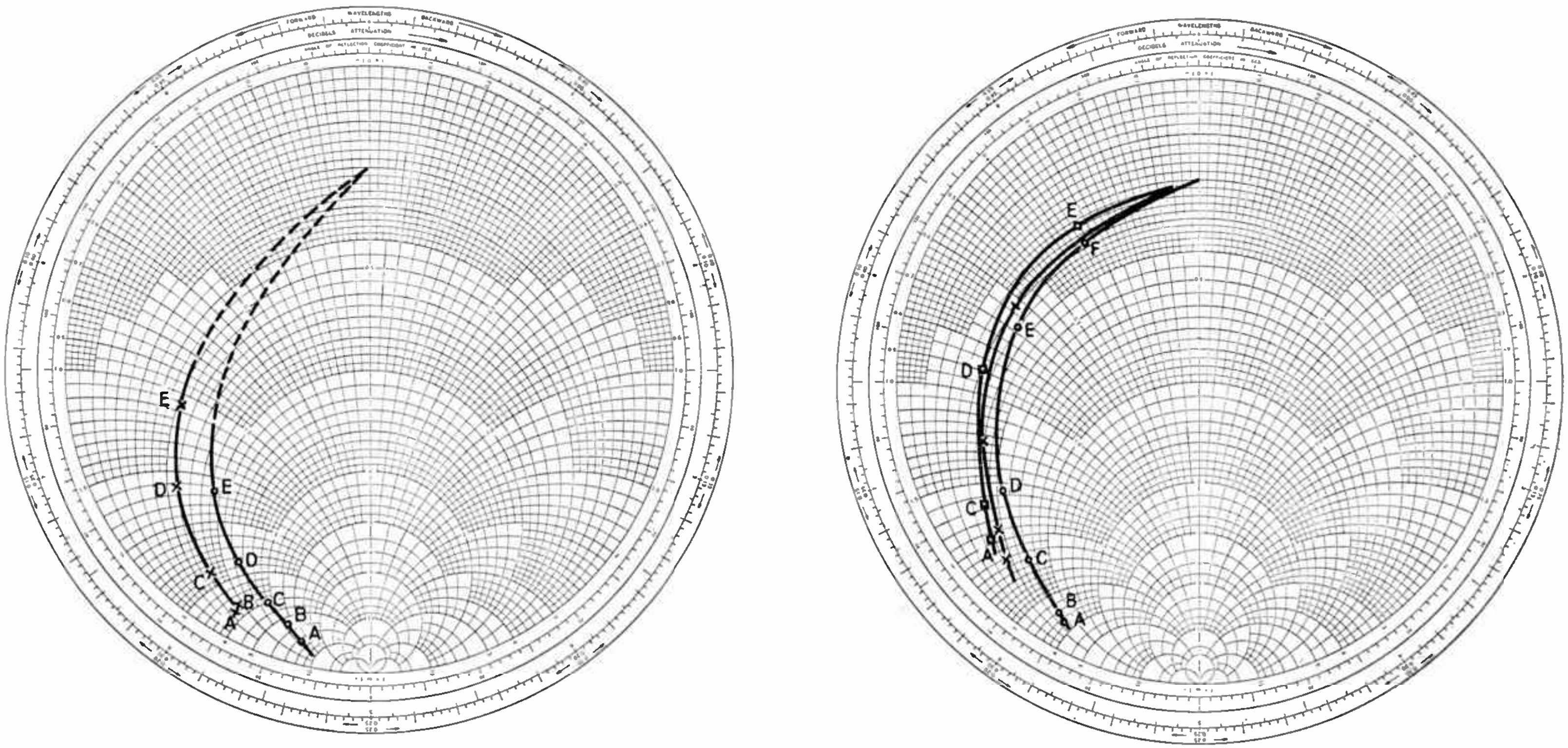
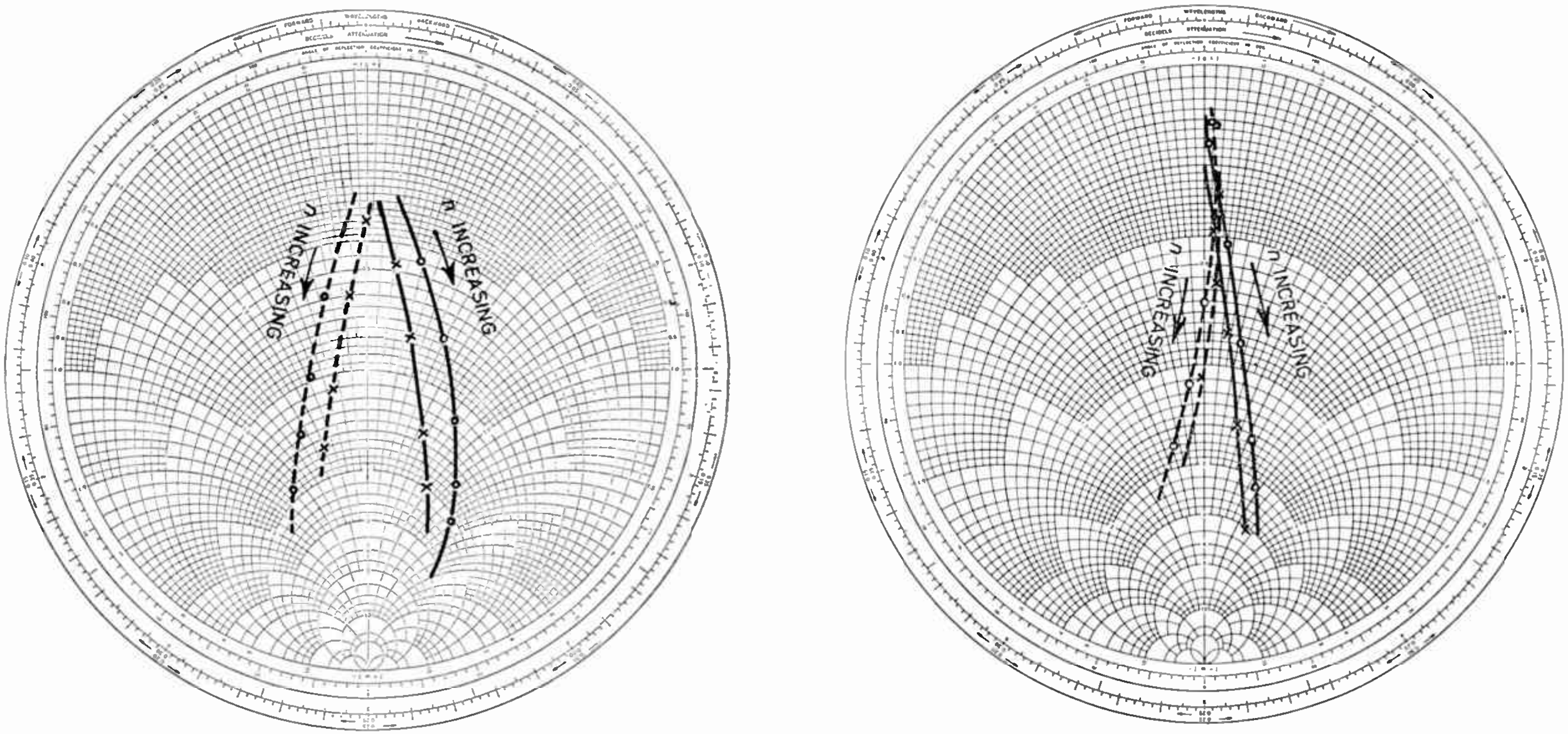


Fig. 10. Multiplier input impedance



—o— OUTPUT IMPEDANCE 4th HARMONIC --o-- OPTIMUM TERMINATION 4th
—x— OUTPUT IMPEDANCE 8th HARMONIC --x-- OPTIMUM TERMINATION 8th
NORMALIZED TO 50Ω

—o— OUTPUT IMPEDANCE 10th HARMONIC --o-- OPTIMUM TERMINATION 10th
—x— OUTPUT IMPEDANCE 14th HARMONIC --x-- OPTIMUM TERMINATION 14th
NORMALIZED TO 50Ω

Fig. 11. Multiplier output and optimum load impedances.

These features are, however, seen to be clearly indicated by the present analysis. It has been the purpose of this paper to indicate that the simple circuit model proposed is capable of performing these predictions while possessing the very great advantage of being independent of the voltage and current waveforms present in the circuit. Moreover, this approach lends itself readily to the incorporation of such effects as variations in transition time and calculation of the effects of idlers.

Investigations are currently in progress to specify the model in terms of step transition time and stored charge (the parameters normally available in data sheets and readily measured) rather than *n*-factor and it is expected that this will permit more complete analysis of practical step-recovery generators than has hitherto been possible.

7. Acknowledgments

The author is indebted to Mr. D. P. Howson and Mr. R. J. Stinchcombe for valuable discussions and comment. Mr. Stinchcombe also wrote the computer programs.

8. References

1. D. L. Hedderly, 'An analysis of a circuit for the generation of high-order harmonics using an ideal non-linear capacitor', *Trans. Inst. Elect. Electronics Engrs on Electron Devices*, **ED-9**, p. 484, 1962.
2. S. M. Krakauer, 'Harmonic generation, rectification and lifetime evaluation with the step-recovery diode', *Proc. Inst. Radio Engrs*, **50**, p. 1665, 1962.
3. D. J. Roulston, 'Frequency multiplication using the charge-storage effect: an analysis for high efficiency, high power operation', *Intl J. Electronics*, **18**, (Series 1) No. 1, p. 73, January 1965.
4. R. Thompson, 'Step-recovery diode frequency multiplier', *Electronics Letters*, **2**, No. 3, p. 117, March 1966.
5. R. D. Hall, 'Step-recovery diodes add snap to frequency multiplication', *Microwaves*, pp. 70–75, September 1965.

9. Appendix

The coefficients A, B and C in equations (8)–(11) are defined as follows:

$$A_1 = R_0 + r_0 + r_2 \cos 2\phi$$
$$A_2 = r_2 \sin 2\phi + X_0$$
$$A_3 = r_{m+1} \cos \overline{m+1}\phi + r_{m-1} \cos \overline{m-1}\phi$$
$$A_4 = r_{m-1} \sin \overline{m-1}\phi + r_{m+1} \sin \overline{m+1}\phi$$
$$B_1 = r_{m-1} \cos \overline{m-1}\phi + r_{m+1} \cos \overline{m+1}\phi$$
$$B_2 = r_{m+1} \sin \overline{m+1}\phi - r_{m-1} \sin \overline{m-1}\phi$$
$$B_3 = R_m + r_0 + r_{2m} \cos 2m\phi$$
$$B_4 = r_{2m} \sin 2m\phi + X_m$$
$$C_1 = r_{m-1} \sin \overline{m-1}\phi + r_{m+1} \sin \overline{m+1}\phi$$
$$C_2 = r_{m-1} \cos \overline{m-1}\phi - r_{m+1} \cos \overline{m+1}\phi$$
$$C_3 = r_{2m} \sin 2m\phi - X_m$$
$$C_4 = R_m + r_0 - r_{2m} \cos 2m\phi$$

Manuscript received by the Institution on 24th April 1967. (Paper No. 1138.)

U.D.C.37:621.37/9

The Future Education of Electronic Engineers

By

Professor G. D. SIMS, M.Sc., Ph.D., D.I.C., C.Eng., F.I.E.E., M.I.E.R.E.†

Professor W. A. GAMBLING, B.Sc., Ph.D., C.Eng., M.I.E.R.E., M.I.E.E.†

AND

B. H. VENNING, B.Sc.(Eng.), Ph.D., A.C.G.I., C.Eng., M.I.E.E.†

Summary: The broad aims of a university course are reviewed and the ways in which these can best be met at the present day are considered. The present structure of the electronics industry within which the students are most likely to find their employment in the immediate future is discussed, and an industry/university course on Product Technology is described. Further education and re-training are discussed. A plea for an improved relationship between industry and the universities is made.

1. Introduction

Edward Lear once wrote that 'Electricity is of two kinds—positive and negative. The difference is, I presume, that one comes a little more expensive but is more durable. The other is a cheaper thing but the moths get into it.' Anyone who has had reason to be concerned about electronics over the last few years must be strongly aware of the accidental truth contained in the part of this statement which relates to the quality of the latter kind. Indeed, just as the moths have got into our thermionic valves and the termites into our transistors with the advent of micro-electronics, so also have they threatened our university courses as well. We have, in consequence, to be continually re-examining the foundations that we lay and the material which we teach, in order to ensure that they satisfy both the needs of today and the prospective requirements of the future. Any entrant to the profession of electronics must be prepared to face a situation in which the basic technologies which he is to use are likely to change a number of times during his professional life, and thus in endeavouring to educate him for the future, we must try to ensure that the foundations laid will be sufficiently secure and broad intellectually to bear the changes which he will encounter. Also his education must be of such a nature as to encourage adaptability, which is far more easily learned early in life than later on. Any satisfactory university course must thus prepare to meet not only the short-term needs, but the long-term ones as well. There are, perhaps, six main aims which should guide the design of any university course. A satisfactory course must set out:

(1) to develop a critical approach to the solution of scientific problems of any nature,

(2) to develop the student's capacity to learn by his own resources,

(3) to provide a sound knowledge of the scientific principles underlying any chosen field of technology,

(4) to provide, in more detail, over a more restricted field, some knowledge of current practice—including design at a suitably advanced level,

(5) to introduce the student to the sociological implications of his work,

(6) to develop the latent individual character of the student.

These broad aims are of vital importance, and if it can be accepted that all are equally important, many of the criticisms which tend to be made of our current university courses can be shown to be without substance. How can we best meet these aims at the present day? In what follows we shall try to analyse the problem and suggest what we see as being a satisfactory solution. Before doing so, however, it is helpful to have in mind some picture of the present structure of the industry into which our students are most likely to find their employment in the immediate future.

2. The Electronics Industry

According to the report on the industry, published in 1966[1] by the Economic Development Committee for Electronics (set up by the National Economic Development Council), the breakdown of employment in the industry is as shown in Table 1.

Table 1

Employment in the electronics industry 1963

Consumer Goods	46 (000s)
Capital Goods	110
Active Components	38
Passive Components	78
Total	272

Source: Trade Association Submissions, Basic Statistics of Ministry of Labour and Ministry of Aviation.

† Department of Electronics, University of Southampton.

Combining the figures for consumer and capital goods, we may conclude that one 'active device' man *currently* supports approximately four 'systems' men. There are, further, those who argue that the advent of microelectronics will make this ratio *many times* its present value in the near future, possibly as much as one to fifty. Assuming therefore that the proportion of graduates and equivalent professional personnel required by each sector is constant, we should clearly be emphasizing 'systems' to a greater extent than we do at the moment. We shall return to this point later, however.

It would be a bold man who would predict what the pattern of industry will be in ten years' time. Already the boundary between devices and systems is blurring with the coming of microelectronics technology. Also, with the decrease in price as well as size of modules, there will undoubtedly be a greater demand for 'systems', both as we know them today and on a more ambitious scale as well. Many of the 'systems' of today will become the 'devices' of tomorrow, but the ratio of 'systems men' to 'device men' may well remain the same—though the basic education which will suit best for either area may well differ.

Much of present device development has stemmed from the application of new materials technology. The future will see more of such development. In no field is this likely to be more true than that of optical systems. Laser techniques and opto-electronics systems are bound to make their impact particularly in the camera tube, display systems, and storage areas. Thus we could agree that there is a strong case for ensuring that our students are aware of these possibilities and of the underlying physical basis. One is tempted to wonder if the field of acoustics, particularly in respect of transducer and acoustic devices generally, would have become the Cinderella that it was before its recent renaissance, had it not suffered from almost total exclusion from physics courses and neglect by electrical engineers. It would be unfortunate to miss our optical opportunities by making the same mistake again.

Whatever the present state of systems design it is clear that much of the future depends upon digital methods for transmission and processing of information. This is as true in the field of control as it is in communications. The electronic engineer who is not able to think in digital terms and in terms of large systems, comprising substantial computer units, will be at a disadvantage in whatever area he seeks employment.

These are just a few of the factors which are affecting the shape of our subject *outside* college and university. There are many others, but it is our responsibility to ensure that the education that we give inside is not only relevant to this picture but also meets the general aims set out in Section 1.

We shall now return to our principal theme for while university courses remain of only three years length, the pressure to meet all of the requirements discussed in Sect. 1 raises a critical question.

3. The First Year Course

Can we, as electronic engineers, still afford the time in a university course for the general engineering studies which for so long have been compulsory for all electrical engineers? Those who advocate the retention of fluids, elasticity, thermodynamics, mechanics etc., in the Part 1 Engineering course are generally those who claim that engineering is one, but surely we must acknowledge at the present time that it is as meaningless to speak about 'engineering' as a specific discipline as it is about 'science'. Does the marine engineer or the gas engineer really need the same basic training as the engineer who will design a microelectronic computer? To argue that he does, is rather like arguing that all scientists should take a first-year course made up of Physics, Chemistry, Physiology, Botany and Geology.

We would hasten to acknowledge, however, that all engineers do have much in common, in that they are all concerned from the outset with the end-product of their work, which must generally be designed to meet a pre-agreed specification, which must be reliable, and further, so designed that it will be safe and environmentally, and as far as possible aesthetically acceptable, once it has been created. On the other hand, no one could argue that atomic physics is of less importance for the electronic engineer than is engineering thermodynamics to an engine designer. Since it is clearly not possible to include *all* of the foundation subjects which *all* of the manifold branches of engineering require in any first-year course, we must conclude that the idea of a completely common first-year course for all engineers is untenable in the modern world. Provided that we ensure that the first-year course does offer a broad introduction to the general principles underlying the particular branch of engineering with which the man is to be mainly concerned, then we must consider that we have done our work satisfactorily.

We recognize implicitly in this statement, however, that there is no *unique* satisfactory pattern and that it is not in the *titles* of courses that the magic reposes, but more in the syllabuses and methods of presentation that the titles represent. Generally, there will be local constraints upon the form of the first-year course and the course as a whole must try to integrate with these. We feel, however, that to subject the student to a variety of courses for which we, ourselves, find it difficult to argue the relevance, is unwise as it is not only a waste of valuable time, but, more significantly, has a disastrous effect on a student's morale.

3.1 *A Preferred Pattern*

Our preference in first-year courses is one which lies rather between the 'pure science' and pure engineering approaches, being made up of:

hours	
4	Mathematics (including statistics and use of computers)
2	Physics *or* Engineering Studies
2	Materials Science
2	Electricity
2	Electronics and Circuit Theory

The course thus divides into four blocks which might be categorized as Mathematics, Physics or Engineering, and Electrical Studies. It is desirable that students should also have some training in the preparation of working sketches and electronic drafting but we have purposely avoided using the word drawing, because of the sinister shades of the past which it conjures up.

Laboratory work at this stage should be straightforward—largely concerned with set experiments designed mainly to introduce the student to the principles and modern methods of measurement. These experiments must complement the lecture course and be framed to aid the understanding of the course material.

Ability in presentation of technical information is of immense importance and lectures and seminars in the subject can well be introduced at this stage. It is important, however, that it should be recognized that in developing the ability to write 'for a specific purpose', exercises other than the writing of formal reports are essential. The amount of work associated with formal report writing should be minimized other than where it fits with this aim or where it is essentially concerned with training the student to record and criticize results systematically.

Of the subjects listed above, we would wish to comment only briefly on the contents of the courses. How far one can go in a Mathematics course depends very much upon the calibre of the students, and entry qualifications vary widely from one Institution to another. Bearing in mind, however, the future interests of the student, we have found that there are great advantages in introducing the students to computer methods at an early stage, in order that they may be able to appraise the relative merits of the analytical and numerical methods of approach. In addition, a treatment of Boolean Algebra at this stage enables the introduction of the ideas of logical design of systems to be introduced in second-year with obvious advantages.

Materials Science is of the greatest importance, for most major advances in engineering over the last few decades have resulted from developments in materials. Materials Science, at this level, however, must be considered in its broadest sense and *not* merely as a study of semiconductors. We should aim to lay a foundation in this subject which can lead later on to the student understanding the limitations which materials technology imposes upon his design at the present time. He should be able to recognize a materials problem when he encounters one—for very often it is such problems which set the limit on good engineering rather than problems more obviously endemic in, or germane to, the design subject.

The Physics or Engineering studies need to be of as general a nature as possible while the two topics comprising the Electrical studies group must be concerned with basic circuit and network theory, electromagnetism, semiconductor devices and transistor circuit theory.

3.2 *Tutorials and Supervisions*

It is vital in any engineering subject that the student should be able to apply the principles and methods that he learns in his courses. Few, however, are inclined to do this without some coercion. Moreover, if the only measure of the student's 'ability to apply' is the sessional or final examination, he will inevitably emerge from the experience with shaken confidence, if indeed he survives in the university at all.

We consider it essential, therefore, to supplement the lectures with suitably conceived and regularly issued examples sheets wherever appropriate. Tutorial classes are also necessary to give the student the opportunity to ask questions and to help him to reserve time for his applications work. We regard these classes as being complementary to, rather than a poor man's substitute for, regular individual or small group supervision. We operate within a Faculty where supervisions in groups of up to five are given at the rate of two per week to all students, and which in any four-week cycle cover all subjects of the student's course. The object of the supervision is not to provide time for solving exercises but more to provide time for general discussion of the subjects covered. Students are given set work and may read essays, give talks, or engage in other activities designed to supplement, or increase the understanding of, the material of the lecture courses. These sessions also afford an excellent opportunity to discuss with the student the industrial significance of a subject or the ways in which the industrial approach to a particular sector will differ from the academic, and why.

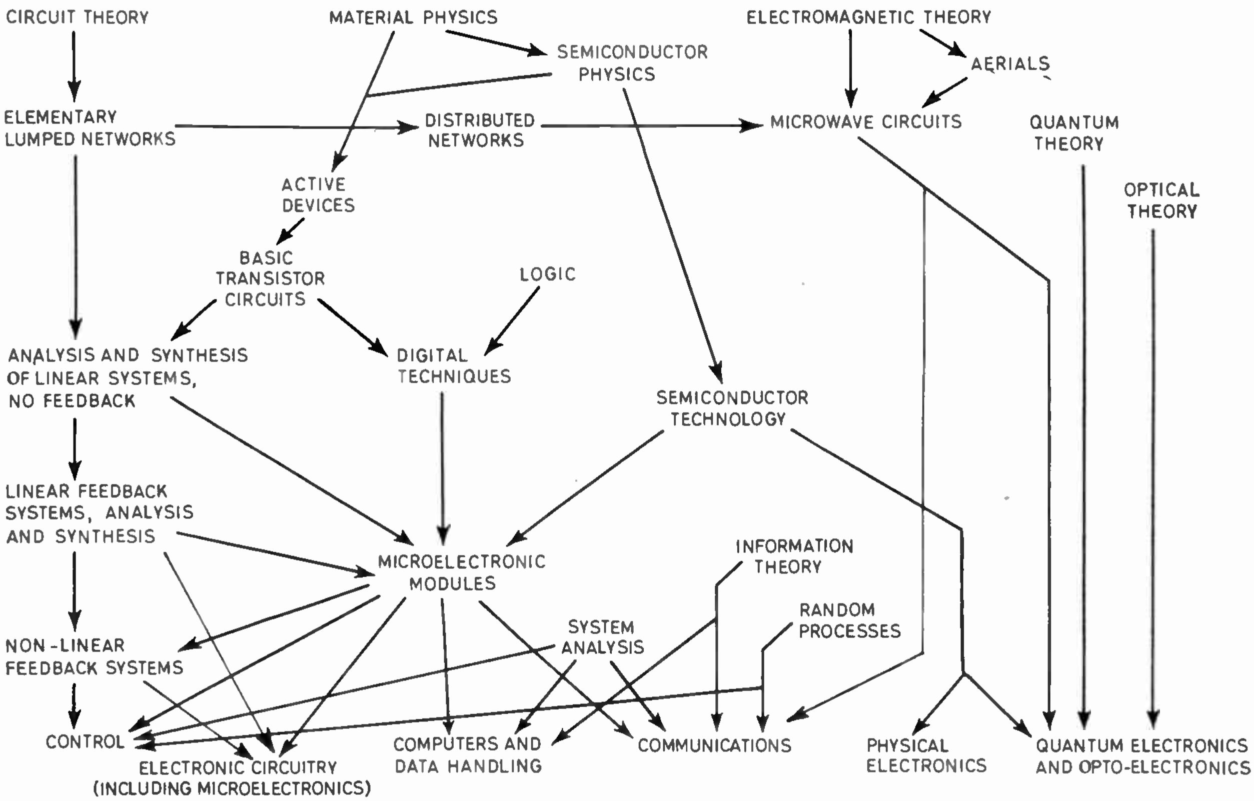


Fig. 1. Course structure.

We use the tutorial class/supervision approach throughout the first two years with encouraging results. We are in no doubt that if the student is really to learn how to *use* his knowledge, this, or some equivalent method *must* be used. Its absence often succeeds only in producing 'parrots'. These matters will, however, be more fully discussed in Sect. 5.

4. Second and Third Year Course Structure

Looking ahead to the probable pattern of employment in the electronic engineering industry in a few years time, when microelectronic modules will be available at prices which will be at least competitive with corresponding piece-part assemblies, the picture as we have argued in Sect. 2, must be one of a large number of *system engineers*, utilizing for most of their work ready-built modules incorporating a large number of active and passive components designed and manufactured by *device engineers*. The former are only concerned with the use of individual transistors and components when they can no longer meet their requirements from standard modules, and they have to know something of the techniques of producing integrated circuits because they have to communicate their wishes to the designers, and also to repel the attacks of salesmen. Conversely, the device engineers are intimately concerned with solid-state manufacturing techniques and circuit design, but need to have a less extensive knowledge of the systems where their devices eventually function. For economic reasons alone, equipment specifications that can be satisfied by micro-modules will have no choice but to employ them.

The training we provide must therefore be suitable for device engineers and also for those who will be concerned with such systems as control, communications (including ranging techniques and telemetry) and computers. In addition there are other areas of lesser, but by no means negligible, commercial significance and special research fields, other than those mentioned above. Not all such latter topics can, or need, be concerned in any one university so that there is bound to be some variation in the material covered. Furthermore, since our products do not normally become effective engineers for several years and may not fully mature for perhaps between five and ten years, our selection should be as far-sighted and as fundamental as we dare make it because of the very rapid rate of progress in the techniques and applications of electronics.

Looking at the later stages of the degree course first, to consider the possible range of final-year options from which a student may select perhaps three or four, the following are suggested as being likely to meet the needs of the electronic engineering industry and research organizations in the near future:

Device Technology,
Advanced Circuit Design,
Systems Analysis,
Computer Technology,
Control,
Communications,
Radar and Navigation Systems,
Microwaves,
Physical Electronics,
Quantum Electronics.

These are not mutually exclusive, and many topics would be common to more than one option.

The basic first-year features discussed in Sect. 3 may be generalized under the three headings of Circuit Theory, Electromagnetic Theory, and Materials Physics. A typical (or suitable) course structure leading from these fundamental topics to the options quoted above is given in Fig. 1. Some brief details of the individual courses are given in the Appendix, but these could obviously be discussed at much greater length. It will be seen that one chain leads from elementary circuit theory to linear and non-linear systems, linking at a fairly early stage with transistor circuits. Much system analysis can be covered very satisfactorily from a 'black-box' approach, treating the elements as ideal amplifiers of certain input, output and transfer characteristics. Digital systems can be treated similarly from ideal switches, and the increasing use of integrated circuits, where the internal design of the building bricks is no longer under control of the systems engineer, means that such a treatment allows more emphasis to be placed upon system aspects. The content of transistor circuit design in courses is therefore likely to diminish in future years, but it must be emphasized that from an educational point of view, it is wise to encourage students to enquire deeply into the contents of the 'black-boxes' at some stage, rather than accept such devices without questioning the circumstances under which their characteristics can be modified. Further, with the rapid advance of large-scale-integration techniques it is becoming progressively more difficult to distinguish the systems aspects from the underlying circuit design. The appropriate mathematical background is assumed to be provided separately by a continuous course throughout the three years, suitably phased with the other lectures. Much of the lecture material is also suitable to act as a common core for an electrical power engineering course.

5. Laboratory Work and Materials Physics

The topic of laboratory work in general has received its fair share of attention over the years, and it is not the intention here to become involved in any arguments over the relative importance of the subject, or about possible methods of implementation. We believe, however, that in planning suitable laboratory work it is essential to arouse the students' interests in the creative aspects of the subject. Individual projects, group projects etc. in the final year each help to satisfy this need as well as helping to encourage self-reliance and co-operation. We have found also that 'design and build' type projects in the earlier years provide a good outlet for enthusiasm as well as helping the student to realize the limitations of theory and experiment when viewed individually. It is as well that this lesson is learned at as early a stage as possible.

It is also of great importance to give the able student the opportunity to use his originality wherever possible and 'open-ended' experiments, though inevitably making more work for the staff, create more interest for the student than too-closely-specified investigations. Interest is also encouraged by ensuring that the practical work really does teach something worthwhile rather than becoming a mere re-iteration of points which the lecture courses have already covered adequately. In that much of the student's early professional life is likely to be concerned with measurements in one form or another the advisability of introducing the student to a sufficiently broad range of measurement techniques in the early part of the course is another point to be considered when designing the pattern of laboratory work.

The one point, further, that has hitherto been little discussed, concerns the provision of laboratory work appropriate to materials physics which is now a very significant proportion of the total undergraduate course. Much of the lecture courses would deal with technologies, such as the planar process of manufacturing transistors, which cannot usually be demonstrated within the economic restrictions prevailing in a university laboratory. But many of the basic physical processes encountered in the solid-state physics course *can* be catered for, as demonstrations or experiments, and moreover they introduce the student to the manufacturing techniques used to obtain the devices that he uses. One approach, used in the University of Glasgow, utilizes fairly simple equipment, but covers some of the fundamental aspects of their comprehensive Materials Physics course[2]:

These experiments include:

(*a*) The construction by students, within a three-hour experiment, of several alloy-junction diodes, and the measurement of junction capacitance, and hence its physical size.

(*b*) The use of a diffusion oven (again in a three-hour experiment) to deposit a silicon dioxide layer on a silicon substrate, with measurement of layer thickness, and also to diffuse phosphorus into silicon to obtain n-type material.

(*c*) The use of bevelling and staining techniques to measure diffusion depths, and the use of four-probe measurements to obtain electrical characteristics.

Many other possibilities exist, and even if the laboratory equipment is simple, it offers an introduction to the complex techniques which the student will encounter in his lectures, or in industry.

If one wishes to tackle more advanced devices, particularly to supplement a course on micro-electronics, then the cost of equipment inevitably rises. Even without photographic equipment, the minimum outlay to deal with monolithic silicon and thin-film circuits is of the order of £10 000 and the cost of the various consumable materials is also high. Provision of bonding and dicing facilities and the manufacture of masks adds to the above, and the processes may well involve provision of clean-room facilities.

Such an outlay can only be justified if it is to support a reasonably large and active interest in integrated circuit technology at undergraduate and/or post-graduate levels, both for teaching and research. An alternative course of action is to omit much of the production equipment, obtaining partly processed elements from industry and concentrating the university laboratory resources upon equipment to evaluate the devices, physically and electrically. Such measurements could involve:

(*a*) the input, output and transfer characteristics of each complete circuit, over a wide frequency range, to characterize the functional block,

(*b*) a more detailed electrical study of the individual elements, to measure parasitic circuit features,

(*c*) a physical study of the various regions, measuring the geometry of the diffusion or deposition patterns, film thicknesses and diffusion depths, relative conductivity etc.

Such an examination can obviously not be carried out on an entire completed circuit, so provision must be made of chips in suitable stages of partial preparation. As the equipment can be acquired in stages as opportunity and finance allows, the laboratory facilities can be provided more economically and the observations and measurements made can provide the necessary supplementation to the lectures.

An alternative, or supplementary scheme to the above, would entail the organization of joint university (or college)/industry projects, at undergraduate and postgraduate levels, with the latter not necessarily tied to a 'Product Technology' course (see Sect. 7). Co-operation from industry in such a field would be invaluable, providing both technical expertise and a useful atmosphere to introduce the student to the commercial world. There are many instances where close co-operation between industry and academic establishments will be of great benefit to both sides in overcoming some of the educational problems arising from the introduction of integrated circuits.

It is apparent therefore that the laboratory facilities necessary for the study of materials science can take many forms, of varying complexity. Something of use can, however, be accomplished with even minimal facilities, so the substantial content of solid-state physics lectures should no longer be left completely without laboratory support. The sight of a suitably treated p-n junction under an optical microscope or, even better, a probe electron microscope, brings a touch of reality to the subject, and 'microns' begin to mean something.

6. Teaching and Assessment Methods

University staff usually devote considerable attention to the content of their lecture courses, but seldom match this with any comparable effort to assess the lecture-effectiveness. With university courses, of necessity, much of the appreciation and understanding must come from a subsequent effort on the part of the student, for a course taken at such a pace as to allow a *full* understanding of *every* point by the slower students, would be intolerable to the brighter ones. An effort is, however, required to ensure that assimilation is made as easy as possible and this requires more than the handing out of examples sheets at infrequent intervals. Sheets of such examples tend to be used more as a preparation for the types of questions to be found in examinations than as a means of reinforcing knowledge of the course content as a whole. A serious attempt to obtain question-and-answer work sheets for use at frequent intervals during a course can be an invaluable aid to the student. These should not confine themselves to long numerical questions taking thirty minutes or so to answer, which seems to be the pattern usually encountered. A series of well-chosen short questions can often be far more effective in testing the understanding of a particular topic, while more descriptive matter, such as an explanation of physical processes, can be tested by a series of multiple-choice answers. A further point in favour of the use of work sheets of some form is that these necessitate individual effort on the part of the student, who otherwise tends during his early years to work only as one of a group, until examinations come along to expose his weaknesses.

It is probably undesirable to use too many exercise sheets for staff assessment of a student's progress during the year, and the major value is an indication to the student himself of his areas of weakness. They can serve as a valuable feedback link to the lecturer, provided he does not get too discouraged at the apparent failure of many of his efforts. The traditional pattern of students' work so often appears to be one of almost complete neglect of lecture material during much of the year—other than pure note-taking—and then a period of intense activity before the examinations. Any scheme to spread this more evenly would be advantageous, and the introduction of continuous-assessment tests is certainly one way. On the debit side is the marking load involved, which could be reduced if some forms of teaching machines are considered—a technique used with considerable success in the U.S.S.R. where the machines are often built by the students themselves.

Teaching machines have found wide acceptance in many forms of education over the last five years or so, both in scholastic fields at primary and secondary levels, and in industrial training schemes. They involve programmed learning courses, which are themselves time-consuming and expensive to produce. It is doubtful whether machines themselves, or even programmed textbooks, have much of a part to play in the type of course that we associate with degree work, but they may well be of use in the following fields:

(i) Remedial work to ensure a common starting standard for students intake, to overcome deficiencies in basic knowledge of, say, Applied Mathematics, or Atomic Physics.

(ii) Remedial work at intervals in the course: giving a student an opportunity to re-inforce his background of certain basic features such as 'The Laplace transform' or 'The p-n junction'.

(iii) Preparatory work before commencing particular laboratory experiments where successful performance of the experiment requires material not covered in lectures and a short programme on a machine would save a laboratory supervisor repeating the same information week after week.

If we ask ourselves why such apparently desirable innovations have not become commonplace, the answer must come to one of time, money and 'is it worth it?'. With small classes requiring such facilities only once a year, the answer tends to be 'no', but with the expanded numbers now going through most departments it would be wise to see whether some at least of these ideas ought to be incorporated. This is particularly true of the early years of the course, where common subjects mean very large classes, and an outlay in time and money on some automated teaching methods shows a useful return.

Examination methods themselves tend to remain in the same traditional pattern of demanding a small number of written answers, which are then marked with a degree of efficiency varying inversely with the number of scripts to be tackled in a given time. There is no intention to argue here that such papers taken in conjunction with other criteria such as the quality of individual project work and oral examinations are not a satisfactory means of deciding the class of a student's degree. But are they the most efficient method of deciding that a student has satisfied a certain qualification level in a subject at an intermediate stage of the course, particularly if several hundred students are concerned with the same paper? Common courses, large lecture theatres, audio-visual aids and closed-circuit television *may* enable one to lecture more efficiently to large classes, but the sessional examinations arising from them produce their own headaches. Here we *must* employ alternative techniques of presenting questions, ending with answers in an easily marked form, and thus reducing the marking load to one of reasonable proportions. We end these remarks, however, on a note of caution, for there is a real danger, which is very evident here, that the regular issuing of example sheets and setting of tests, can result in the type of question, even in final examinations, which is easy to mark rather than a test of understanding. The two are not necessarily incompatible but are often so if insufficient care is taken.

7. Postgraduate Courses

At the present time, the electronics industry is far from well supplied with graduates with adequate electronics background. It recruits a fair number of physicists, and some chemists and mathematicians, but complains that

(*a*) insufficient students from these disciplines consider careers in industry to be sufficiently satisfying academically,

(*b*) those who do enter industry are often not of the highest calibre.

Moreover, it is becoming increasingly clear that a three-year graduate course of any kind does not often give, however hard we try, an adequate preparation for professional life in industry. There is, therefore, an increasing tendency for students on graduation to take postgraduate courses of one kind or another and although there is undoubtedly a need for these courses, they are often neither well conceived, patronized nor undertaken for the right reason! Postgraduate courses of one year or more duration can serve a

Table 2

Postgraduate courses in British Universities in Electronics, 1966-7

Electronics and Electrical Engineering 6 + Control 2	Applied 2	Physical 2	Solid-State 3	Pulse and Digital 1
Control and Control Systems 5		+ Instrumentation 2	+ Systems 2	+ Machines 1
Systems 5		+ Information 1	+ Circuit Engineering 1	
Communications and Communications Systems 2		+ Physical Electronics 1	+ Electronics 1	
Microwaves† and Microwave Engineering 5				
Materials 2		+ Solid-State Electronics 1		
Quantum Electronics 3				
Other: Photoelectronics 1				
Other: Instrument Technology 1				

† Includes one course 'for graduate physicists'.

variety of purposes, for example they can:

(i) give further education beyond degree level in a particular specialist branch of a discipline,

(ii) enable those graduating in some other discipline to 'convert', for example, there are several electronics courses which are run with the main purpose of 'converting' graduate physicists into electronic engineers,

(iii) be of a non-technical nature, for example, in management, aiming to supplement the knowledge gained in undergraduate courses in cognate fields.

The M.Sc. conversion course, intended for those who have not read electronics as a first degree subject, fulfils a useful purpose in enabling the physics student or other suitable graduate to gain a sufficient background to make them into useful electronics engineers in one year and, in our experience, the demand for courses of this kind is substantial.

Of the more specialized 'deep' M.Sc. courses, there is a greater proliferation as shown in Table 2 and not surprisingly student interest is less marked. Very often these courses do serve a national need, but frequently they are merely mounted as outlets for the enthusiasms of the departments running them. There is a strong case for a more active national co-ordinating mechanism for university postgraduate courses in general, but especially in the more specialist fields. Industrial views are of great importance in this connection, but again in the experience of many of us, an industrial expression of faith in a course is often not backed by an influx of industrially sponsored students.

Far more care needs to be taken too to provide suitably informative titles for these courses as it is plain to anyone that many of the courses given in the table are inadequately described at present. Many of the courses merely titled 'electronics' offer options in the more specialized fields, while for example, though only one course is entitled 'pulse and digital' electronics it is clear that the subject is covered in many of the other courses in which it is not specifically mentioned. Some national co-ordination here could prove advantageous both to the intending student and to industry as well.

While we feel that the general conversion-type of course should be encouraged therefore, and that where there is an obvious national need, specialized postgraduate courses should be run, there is a very strong case for courses of the 'product technology' type proposed in the Bosworth report.[3] These 'product technology' courses are so conceived that they are specialized to the needs of a particular industrial product group, for example, the semiconductor device industry. Courses of this kind are intended to be run from centres which are established in areas where an academic body of suitable competence and a complementary industrial body exist in proximity to one another. Much of the teaching in the course is intended to be done by members of industry and much of the practical work is conducted in the industrial environment. Such a course would meet at least some of the objections which are made to our present system including giving real experience of technological processes, production lines, working to target dates and working within cost limits! Another feature of these courses would be group projects in which a number of students would collaborate to undertake some industrial task under industrial conditions, again involving cost and target dates. This latter exercise is likely to be particularly valuable as this is, perhaps, the first time that the student becomes aware that most of his life will be spent working in collaboration with others rather than working completely on his own, as he has always been encouraged to do at school and probably at university as well! Courses of this kind will serve a most valuable purpose, not the least part of which could be to attract the better quality graduates into industry, via the higher degree which he will obtain on satisfactory completion of the course. The fact that the course was recognized by an academic body would go a long way to persuade the intending student that it was worthwhile, and while pursuing it he would discover that there was as much challenge in the whole range of industrial activity as he previously imagined to exist perhaps only in the research field. Courses of this type are only just beginning to be talked about, but they do raise in a very well-defined way the advantages which could accrue from improved university/industry collaboration. We have already mentioned the advantages and difficulties of undergraduate students undertaking project work in industry. It has to be remembered also that although many university staff at the present time undertake consulting work in industry they are often still, to a large extent, shielded from everyday industrial problems. Participation in courses of this kind would bring them face to face with these problems and into closer contact with those whose work it was to deal with them. This, in turn, would help to add a substantial realism to their teaching and to the academic courses with which they were concerned.

As we have stressed earlier, the main purpose of the academic course must be to develop critical ability and scientific method in the student, but in our efforts to do this there is no reason whatsoever why the student should not be brought into contact with *realities*! If courses of the Product Technology type come to be generally accepted, they may well make the ideal 'matching' unit between the university graduate and industry, and the student once thus trained will not only be immediately useful in an industrial environment, but will further have received a broader view of the industrial approach than most receive at the moment.

The first industry/university centre to be formed for the purpose of running a 'Product Technology' course has been created as a joint venture by Associated Semiconductor Manufacturers Ltd., and the University of Southampton.

The first course in semiconductor device technology will commence in 1967 extending over a period of fifteen months and will lead to an M.Sc. degree. The general pattern of the course is set out in Table 3:

Table 3

General pattern of Product Technology course

Base			
U and I	Introductory Course	2 months	September–October
I	Production experience	3 months	November–January
U	Product Technology Course		
	Phase 1	~7 weeks	February–end of Spring term
	Reading and Private Study		Easter vacation
U	Product Technology Course		
	Phase 2	~10 weeks	Summer term
	Reading and Private Study		Summer vacation
I	Group Project	3 months	August–October
U and I	Product Technology Course		
	Phase 3	~6 weeks	November–mid-December
U	Final examination and assessment		

U = University or College; I = Industry.

The university part will comprise lectures and associated laboratory work along the lines set out in Table 4 in which the columns included cater for a number of specific ranges of industrial occupation—from device research to production.

8. Further Education and Re-training

It is often implicitly assumed that education finishes when a student leaves the university. However, whether or not he has undergone postgraduate university training of any of the forms outlined above, this is clearly not the case. In future, industry must accept a greater responsibility for ensuring that a man is adequately trained in those aspects of their requirements which are not covered at the universities. Moreover, they must see that technical 're-training' is available in whatever form is necessary to meet the needs of the time. For example, there are at present

Table 4
University part of the Product Technology course

Group	(a)	b(i)	b(ii)	b(iii)	b(iv)
Course number	Device applications	Device research	Pre-production	Production	Process development
1		Industrial Organization and Management			
2	← Industrial Services →				
3	← Commercial Operations →				
4	← Engineering Materials →		← Production Engineering →		
5	Systems	Device Design	Physics of Failure	Reliability and Quality Control	Mechanization
6	Device Design	Systems		Systems and Device Design	
	← Device Technology →				
			Statistics and Use of Computers		
			Special Invited Lectures and Seminars		

many men in industry who have failed even to come to terms with transistor technology and this has resulted in a perpetuation of the use of vacuum tubes in many applications for which they are far from being the best elements.

How should this re-training be done? Again the universities can play their part. The point has often been made that the university teacher should be expert in the *presentation* of material: he is, in fact, a *teacher* as well as a skilled research worker. It would thus be logical to ask the Universities to undertake re-training where necessary. However, this does present a problem, for frequently the universities are not, as we have argued, in the forefront of new technology and their teachers therefore need instruction themselves. However, in many cases industry would be prepared to give this teacher training—for example, at the present time in modern microcircuit design—and having allowed the teacher to work in an industrial group for a short spell he would then be able to go back and present his subject in a way which people at least could understand and digest. Generally speaking, this approach to the subject is far more satisfactory than for industry itself to try to put on courses with unskilled instructors. It does, however, mean that universities must be prepared to release to industry their teachers for short periods.

Re-training courses present a quite different problem from that presented by normal undergraduate training. Many of the people to be re-trained will have graduated some years ago and will thus be out of touch with the academic approach and may need some revision of relatively simple studies before proceeding to the re-training course. For this reason, careful thought needs to be given to the duration of such courses and also to their content, if they are to make the maximum impact. Implicit in this is the recognition that the people to be re-trained are often of mature years and will therefore need to be approached in a completely different way from that used with a normal undergraduate population. Whether or not these courses are residential again depends very largely on the nature of the subject to be studied, although generally speaking residential courses of more than a few weeks duration do not prove possible due to industrial pressures, apart from the difficulties of providing short-term university accommodation. It would seem therefore that the universities must accept responsibility in the future on a far wider scale than at present for running many more *short* courses.

9. Conclusion

The essence of most of the argument of these last sections has been that of an improved relationship between industry and the universities. Given this partnership in the first place, there is no reason why considerable progress should not be made towards a solution of these problems. However, it must first of all be countenanced that these problems exist, and secondly, that they are worthy of solution. The solution will not come without the expenditure of a considerable amount of effort on both sides and the responsibility here lies as much with the universities as with industry. They must try to break away from their established patterns of long-term courses at both undergraduate and postgraduate level and realize that a more flexible role is needed if the maximum use is to be made of the considerable investment of intelligence and money which is made in them at the present time! On the other hand, industry must recognize that merely to criticize the universities is not enough. Industry has the resources to help the universities perform a more widely useful function than at present and we would suggest that they would find the universities not unwilling to collaborate if the right kinds of approach were made. Frequently the criticisms currently encountered are based upon impressions of the universities gathered when the critics themselves were undergraduates several decades ago. Were they perhaps to go inside a modern university they would find that conditions and attitudes have changed out of all recognition!

10. Acknowledgment

The authors acknowledge with gratitude the many discussions with their colleagues of the Department of Electronics, University of Southampton, and elsewhere, over many years, which have contributed to the ideas set forth in this paper.

11. References

1. 'Electronics and the Future', Electronics EDC Report (H.M.S.O., London 1966).
2. J. H. Collins, P. Hlawiczka, R. Hutchins and J. Lamb, 'University teaching and laboratory work in electrical materials science', *Internat. J. Elect. Engng. Educ.*, **3**, p. 175, July 1965.
3. 'Education and Training Requirements for the Electrical and Mechanical Manufacturing Industries' (H.M.S.O., London 1966).

12. Appendix

Course Details

It is unnecessary here to specify details of all the component parts of the courses shown in Fig. 1, as many have featured for a number of years. The content of some parts may not, however, be so obvious, and suggestions of the course content of these are given below.

(i) *Materials Science—1st Year Subject*

Crystal structure, x-ray diffraction, Brillouin zones, reciprocal lattice, crystal defects.

Basic atomic physics, atomic structure, wave nature of electronics, quantization of energy and radiation, uncertainty principle, Schrodinger wave equation (concept only).

Molecular bonding. Nuclear structure and forces. Polymers. Mechanical properties of materials—elastic forces, plastic deformation, fracture.

Thermodynamics of solids, internal energy, entropy.

Simple treatment of electrical and magnetic properties of materials.

(ii) *Physics Studies—1st Year*

This will include:

Optics: Image formation with lenses and mirrors, interference and diffraction, optical instruments.

Atomic structure, atomic spectra, elementary ideas of quantum theory, periodic table.

Black-body radiation, laws of radiation, radiance, etc.

Waves and vibrations.

Elementary electricity and magnetism.

Treatment of experimental observations.

(iii) *Solid-State (or Semiconductor) Physics (or Electronics)—2nd Year*

Elastic properties and elastic waves in crystals, thermal vibrations and phonons, free electron gas models, specific heat, electrical and thermal conductivity.

Particle statistics, Boltzmann and Fermi-Dirac.

Further treatment of wave functions, zone theory and energy bands, effective mass for holes and electrons.

Semiconductors, extrinsic and intrinsic, thermal ionization, carrier lifetimes, mobility, diffusion, experimental techniques.

Junctions, rectifying and ohmic; detailed treatment of transistor operation, minority and majority carrier devices.

Thermoelectric effects.

(iv) *Physical Electronics—3rd Year*

Superconductivity—simple theory and applications.

Dielectric properties of materials, dielectric constant, polarizability, relaxation processes, ferroelectronics.

Magnetic properties of materials; para-, dia-, and ferro-magnetism, domains, hysteresis, spin waves, magnons.

Motion of charges in E and H fields, magnetron principle, cyclotron resonance, lenses and electron microscope.

Simple treatment of magnetic resonance.

Simple treatment of plasmas including electromagnetic interactions.

(v) *Quantum Electronics and Opto-electronics—3rd Year*

Interaction of radiation with matter; amplification of stimulated emission; gaseous and solid-state masers; gaseous, liquid and solid-state lasers; design and application of lasers and laser systems; non-linear optics; optical harmonic generation and parametric amplification, quantum counters.

Optical interactions in solids, colour centres, photoconductivity luminescence, absorption and emission spectra.

Display systems, optical detectors.

Optical coupling in circuits.

(vi) *Microelectronics—3rd Year*

The advantages and problems of miniaturization.
Thin film circuits: production technology: passive and active devices: circuit layout and interconnections.

Semiconductor integrated circuits: production technology: epitaxy and the planar process: passive and active devices: hybrid circuits.

System design.

Reliability and failure physics.

(vii) *Digital Techniques—3rd Year*

Logical design via Boolean Algebra: gate and bistable circuits: computer arithmetic, information processing.

Circuit details of gates and bistables, memory elements and peripheral equipment.

Computer organization.

(viii) *Semiconductor Technology—3rd Year*

Crystal production, zone melting and refining: Alloying, diffusion and epitaxial techniques as applied to various semiconductor devices and materials.

(ix) *System Analysis—3rd Year*

Graph theory: critical path methods: flow maximization: minimum cost flows in networks: synthesis of networks (to include manual and computer calculations).

Manuscript received by the Institution on 30th May 1967. (Paper No. 1139/Ed.13.)

Fourth International Vacuum Congress

Preparations are well advanced for the Fourth International Vacuum Congress to be held at the University of Manchester Institute of Science and Technology from 17th to 19th April 1968. This Congress, like its predecessors at Namur in 1958, Washington in 1961 and at Stuttgart in 1965, will take place under the general auspices of the International Union for Vacuum Science Technique and Applications (IUVSTA); the detailed organization is the responsibility of the Joint British Committee for Vacuum Science and Technology on which the Institution is represented.

The subject of vacuum science and technology is vast and diverse, and it can only be adequately covered by means of parallel conference sessions. The groupings are:

'Theory and Fundamental Processes' (e.g. Electron Emission and Ionization Phenomena and Surface Phenomena).

'Design and Operation of Vacuum Devices' (e.g. Vacuum Gauges and Process Control Devices).

'Vacuum Systems and Applications' (e.g. Space Simulation, Particle Accelerators and Nuclear Fusion Apparatus, Electrical Breakdown and Vacuum Insulation, Vacuum Deposition Systems and Related Properties of Thin Films, and Electron Tube Systems and Devices).

Papers are invited on all the subjects to be treated at the Congress. The Committee request that in dealing with those subjects with rather limited vacuum content, contributors will frame their papers so as to bring out the vacuum aspects.

Offers of contributions of about 20 minutes' presentation time (say 3,000 words) are welcomed and should be sent not later than 30th September 1967, to: Mr. J. Yarwood, 4th International Vacuum Congress Technical Secretary, 47 Belgrave Square, London, S.W.1.

U.D.C. 621.396.96

Efficiency of Attenuation of Moving Clutter

By

J. KROSZCZYŃSKI,

Dr.Eng.†

Summary: The Doppler-frequency shift resulting from clutter motion degrades the efficiency of moving target indication (m.t.i.) radar systems. Although the degradation is much less when some compensation methods are used, it still exists because of the finite tolerances of the Doppler-compensation frequency shift. This effect is analysed for single- and double-delay-line m.t.i. cancellers and fluctuating clutter.

1. Introduction

The m.t.i. systems used in radio location are operating in principle against clutter echoes, reflected from stationary objects. In some cases, however, the clutter spectrum may be Doppler-shifted. This can occur either when the interfering reflecting objects are moving: for example, the water droplets in a drifting cloud, or the radar itself is in motion as in the case of a shipboard or airborne radar. Since the relative velocity of motion between radar and target will usually be greater than the relative velocity between clutter and radar, the clutter may still be attenuated, but the efficiency of attenuation will suffer. It is the purpose of this paper to analyse the degradation of clutter attenuation, resulting from the Doppler shift of the clutter spectrum. This analysis can be useful in the case when no Doppler-frequency compensation is used, or in the case when the finite tolerances of the Doppler-compensation frequency shift are considered.

2. Attenuation of Stationary and Moving Clutter

The attenuation of the clutter signal by the m.t.i. delay-line canceller (clutter attenuation, CA) can be written, after Grisetti *et al.*[1], as the ratio of input power divided by the output power:

$$CA = \frac{\int_{-\infty}^{\infty} W(f)\,\mathrm{d}f}{\int_{-\infty}^{\infty} W(f).|\mathrm{F}(f)|^2\,\mathrm{d}f} \qquad \ldots\ldots(1)$$

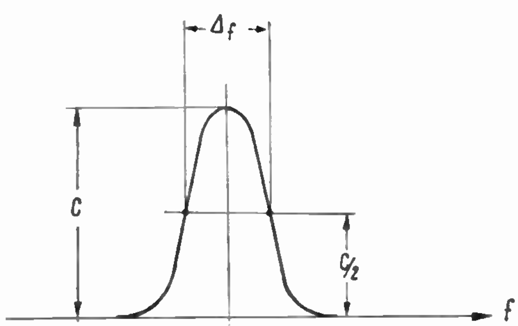


Fig. 1. Clutter power spectrum bandwidth.

† Przemysłowy Instytut Telekomunikacji, Warsaw, Poland.

where $W(f)$ = the clutter power spectrum; $\mathrm{F}(f)$ = the frequency-response function of the canceller.

For the single-delay-line canceller[1] with the delay time T:

$$|\mathrm{F}_s(f)| = 2\,|\sin(\pi f T)| \qquad \ldots\ldots(2)$$

Using the expression given by Barlow[2]:

$$W(f) = W_0 \exp[-a(f/f_0)^2] = W_0 \exp(-q^2 f^2) \ldots\ldots(3)$$

where f_0 is the radar carrier frequency and a is a parameter dependent upon clutter. The expression for CA_s may now be obtained in the form[3]:

$$CA_s = \frac{0{\cdot}5}{1-\exp(-3{\cdot}56\,k^2)} \qquad \ldots\ldots(4)$$

where k is the ratio of clutter power spectrum bandwidth Δf (see Fig. 1) to the pulse repetition frequency f_r:

$$k = \frac{\Delta f}{f_r} = \Delta f T \qquad \ldots\ldots(5)$$

When the clutter spectrum is shifted by an amount f_D due to Doppler effect, the minima of the delay-line canceller response do not coincide any more with the maxima of the interfering spectrum (see Fig. 2). This results in a degradation of the clutter attenuation efficiency. In this case, the value of CA will be, according to eqn. (1), given by the expression:

$$CA = \frac{\int_{-\infty}^{\infty} W(f)\,\mathrm{d}f}{\int_{-\infty}^{\infty} W(f-f_D).|\mathrm{F}(f)|^2\,\mathrm{d}f} = \frac{\int_{-\infty}^{\infty} W(f)\,\mathrm{d}f}{\int_{-\infty}^{\infty} W(f).|\mathrm{F}(f+f_D)|^2\,\mathrm{d}f} \qquad \ldots\ldots(6)$$

Substituting eqns. (2), (3) and (5) in the denominator of eqn. (1) yields:

$$\int_{-\infty}^{\infty} \exp(-q^2 f^2)\,4\sin^2[\pi(f+f_D)T]\,\mathrm{d}f = 2\int_{-\infty}^{\infty} \exp(-q^2 f^2)\{1-\cos[2\pi(f+f_D)T]\}\,\mathrm{d}f \qquad \ldots\ldots(7)$$

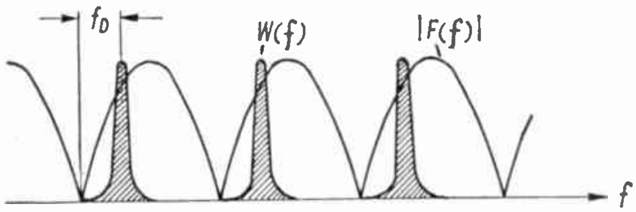


Fig. 2. Doppler-shifted clutter power spectrum and the frequency response of the single-delay-line canceller.

As shown by Gradshteyn and Ryzhik[5]

$$\int_{-\infty}^{\infty} \exp(-q^2f^2)\cos[b(x+\lambda)]\,dx = \frac{\sqrt{\pi}}{q}\,e^{-\frac{b^2}{4q^2}}.\cos(b\lambda) \qquad \ldots\ldots(8)$$

and

$$CA_s = \frac{0{\cdot}5}{1-\exp[-(T/4q)^2]\cos(2\pi f_D T)} \qquad \ldots\ldots(9)$$

Comparing eqns. (4) and (9), it may finally be shown that:

$$CA_s = \frac{0{\cdot}5}{1-\cos(2\pi f_D T)\,\exp(-3{\cdot}56\,k^2)} \qquad \ldots\ldots(10)$$

For the double-delay-line canceller

$$|F_d(f)| = 4\sin^2(\pi f T) \qquad \ldots\ldots(11)$$

The efficiency of stationary clutter attenuation for the double-delay-line canceller is[3,4]:

$$CA_d = \frac{0{\cdot}5}{3-4\exp(-3{\cdot}56\,k^2)+\exp(-14{\cdot}24\,k^2)} \qquad \ldots\ldots(12)$$

Using eqns. (6), (8) and (11), and the relation

$$\sin^4\alpha = \tfrac{1}{8}(3-4\cos 2\alpha+4\cos 4\alpha)$$

the expression for the case of Doppler-shifted clutter spectrum may be obtained. Thus

$$CA_d = \frac{0{\cdot}5}{3-4\cos(2\pi f_D T)\,e^{-(T/4q)^2} + \cos(4\pi f_D T)\,e^{-(T/2q)^2}} \qquad \ldots\ldots(13)$$

Comparing eqns. (12) and (13), one finally obtains:

$$CA_d = \frac{0{\cdot}5}{3-4\cos(2\pi f_D T).\exp(-3{\cdot}56\,k^2) + \cos(4\pi f_D T).\exp(-14{\cdot}24\,k^2)} \qquad \ldots\ldots(14)$$

3. Conclusions

The general dependence of CA_s upon k and $\omega_D T$ (where $\omega_D = 2\pi f_D$) may be best visualized with the help of a three-dimensional diagram (Fig. 3). The values of CA_s can be read more exactly from Figs. 4 and 5. Similar diagrams for the double-delay-line canceller are shown in Figs. 6 and 7. It can be seen that although the double-delay filter gives much higher absolute clutter attenuation values, the curves for both filter types are quite similar. It may also be noted that

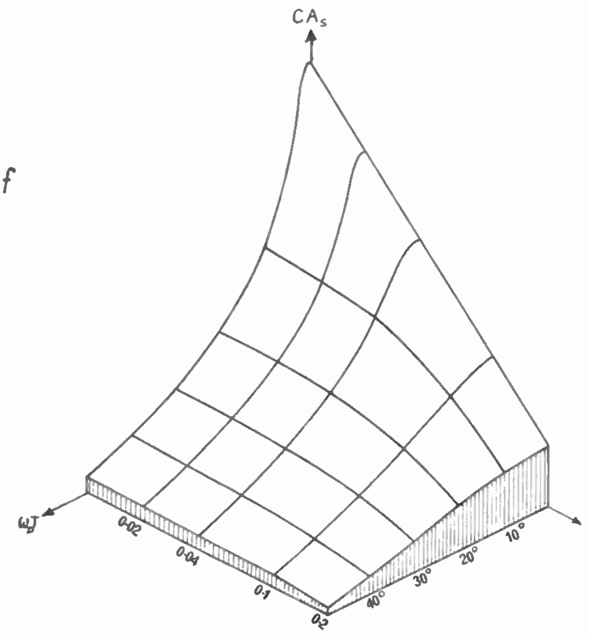


Fig. 3. Three-dimensional CA_s diagram.

$$\omega_D T = 2\pi\frac{v}{v_{b1}} \qquad \ldots\ldots(15)$$

where v = relative radial velocity of clutter drift; v_{b1} = first m.t.i. blind speed. For small relative clutter drift speeds, clutter attenuation depends largely on the relative clutter bandwidth k; for small k, the attenuation is limited mainly by the ratio v/v_{b1} (see Figs. 5 and 7).

4. References

1. R. S. Grisetti, M. M. Santa and G. M. Kirkpatrick, 'Effect of internal fluctuations and scanning on clutter attenuation in m.t.i. radar', *Trans. Inst. Radio Engrs on Aeronautical and Navigational Electronics*, ANE-2, pp. 37–41, March. 1955.
2. E. J. Barlow, 'Doppler radar', *Proc. I.R.E.*, **37**, pp. 340–55, April 1949.
3. J. Kroszczyński, 'Tłumienie ech stałych w radiolokacji' (The attenuation of clutter in radiolocation), Volume 1, (PWN, Warsaw, 1965).
4. J. Kroszczyński, 'Tłumienie ech stałych w pojedynczym i podwójnym układzie kompensacyjnym' (Efficiency of attenuation of clutter in simple and double cancellation apparatus), *Prace Przemyslowy Inst. Telekomun.* **8**, No. 24, pp. 41–6, 1958.
5. I. S. Gradshteyn and I. M. Ryzhik, 'Tables of Integrals, Series and Products' (Academic Press, New York and London, 1965).

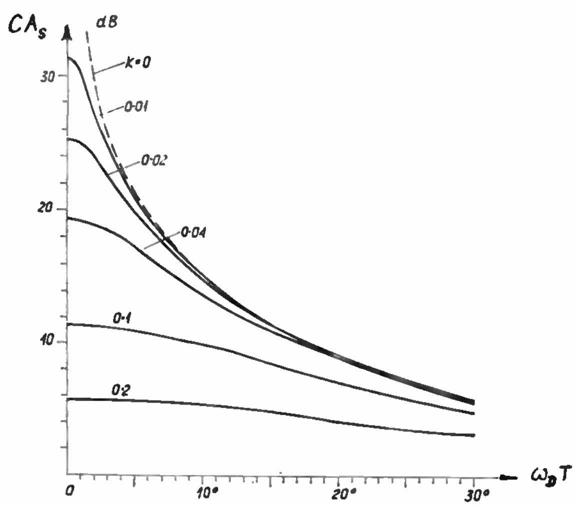


Fig. 4. CA_s as a function of $\omega_D T$.

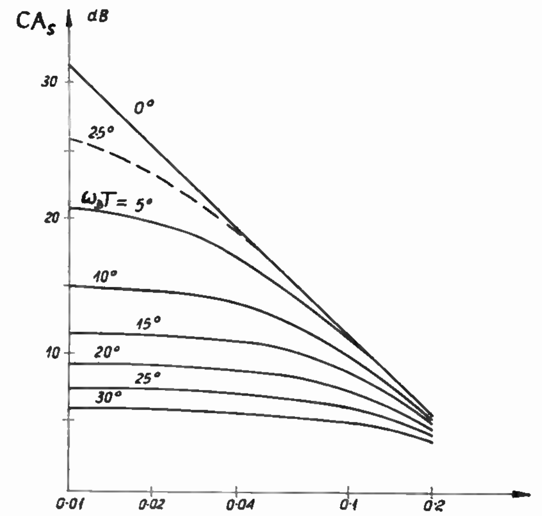


Fig. 5. CA_s as a function of k.

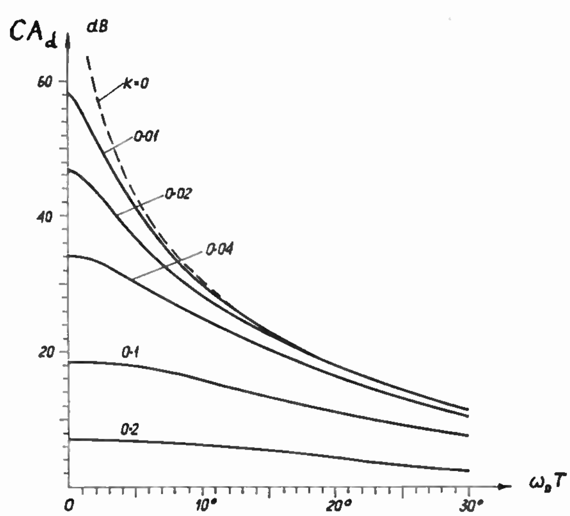


Fig. 6. CA_d as a function of $\omega_D T$.

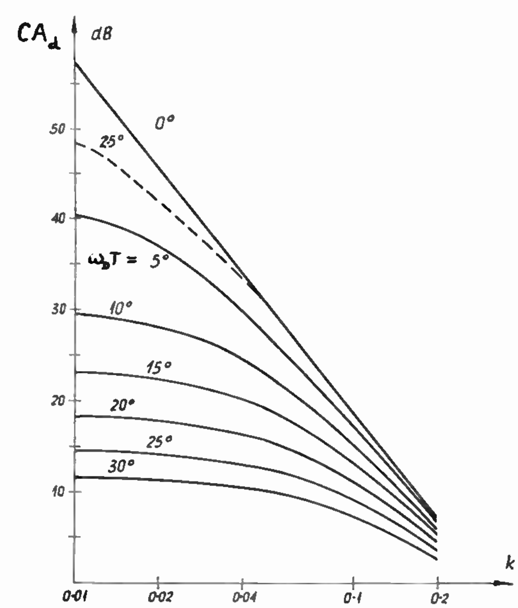


Fig. 7. CA_d as a function of k.

Manuscript received by the Institution on 28th April 1967. (Contribution No. 99/RNA72.)

B.B.C. Electronic Field Store Colour Television Converter

A demonstration was recently given in London by the B.B.C. Engineering Division of the first purely electronic standards converter, shortly to go into operational use. This is capable of converting colour television pictures instantaneously from the 525-line 60-field N.T.S.C. system used in America and other countries, to the 625-line 50-field PAL system used in Great Britain and much of Europe and in countries in other parts of the world; if required the converter can be arranged to provide pictures on the 625-line 50-field SECAM system used in France and some other countries. Electronic standards conversion is a technique in which the B.B.C. has achieved considerable success and ascendancy. In 1963 the first all-electronic standards converter was developed for monochrome television† and this provided the necessary basis for the line-store converter, which is a component unit of the new field store converter.

The task of converting a colour picture originating on the 525-line, 60-field N.T.S.C. American system to the 625-line, 50-field PAL or SECAM European systems involves three major problems: changing the number of lines; changing the colour television system; and changing the number of fields per second. Of these, the changing of the number of fields per second is very much the most difficult, particularly in the case of colour pictures.

The principle on which the new electronic converter operates may be appreciated by considering a period of one-tenth of a second. It is clear that within this period 6 fields will have elapsed in the American system whilst 5 fields will have elapsed in the European system. It will also be obvious that, if the two systems start scanning simultaneously the first American field will have been completed slightly before the first European field, and thus the information can be transferred directly as regards time.

The second American field will, however, begin slightly before the second European field, 16·6 ms after the start of the one-tenth second period as opposed to 20 ms. A delay of just over 3·3 ms is, therefore, introduced into the converter circuits between the incoming and outgoing field scan. Thus American field No. 2 is delayed until field No. 2 European is starting, and so on with increasing delays until No. 5 American becomes No. 5 European.

It may now appear that field No. 6 of the incoming 525-line picture has to be discarded. If this were really the case, movement in the picture, particularly of fast-moving objects, would appear to be jerky as every 6th field would be completely missing in the converted picture. To overcome this disadvantage, a further delay of 3·3 ms making a total of 16·6 ms or one complete field, is introduced between field No. 6 of one series and field No. 1 of the following series so that these two are superimposed as far as the delay section output is concerned. The result of this technique is to mitigate the jerky effect to an extent at which it is barely noticeable except when the movement being portrayed is extremely rapid.

As the line scanning rate has not been altered by the introduction of the delay, the height of the delayed 525-line picture will be reduced and the picture geometry will be wrong at this stage in the conversion. The effect of this is that there will be a space at the top and bottom of the screen and any circular object in the picture would appear oval.

To rectify the geometry, to change the number of lines, and also to deal with the colour content of the picture, further stages are necessary to complete the conversion. The output from the delay section of the converter is fed to an American type (N.T.S.C.) decoder which separates the luminance and chrominance components of the delayed incoming signal. An extra unit in the decoder breaks down the chrominance signal into the I and Q signals which together form the coded N.T.S.C. colour-signal. The luminance component is fed through a line-store converter which reconstitutes the delayed black and white picture to 625-line standards and, in doing so, corrects the picture geometry. The I and Q signals are also passed through a second line-store converter and emerge as I and Q signals on 625-line standards. The luminance and chrominance components, all three now at 625-line standards, are fed into a PAL colour coder the output of which is a European 625-line, 50-field/s PAL colour signal. The same three signals could, alternatively, be coded to provide a 625-line SECAM picture. If only monochrome signals are incoming, the output of the luminance line-store converter alone provides the 625-line picture as no colour information is present. In correcting the picture geometry the line-store converters compress the picture in a horizontal direction and the fully-converted picture therefore has a small space on the right- and left-hand edges as well as at the top and bottom.

At the demonstration, pictures on American N.T.S.C. standards, from videotape recordings as well as by direct transmission from Toronto via *Early Bird* satellite, could be compared side-by-side with the 'converted' 625-line PAL pictures. These showed excellent quality with little degradation, only a few vertical striations appearing to indicate the severe electronic processing that the signal had undergone.

The fact that the area of the converted picture is somewhat smaller than usual and that its aspect ratio was slightly changed (5 : 4 instead of 4 : 3) was not too disturbing—less so indeed than the alterations in area which are to be seen in the transmission of wide-screen films. Some operational problems are presented to the Broadcasting Authority by the nature of the output signal, depending on whether it is to be fed to a transmitter or recorded, and due to the small changes in sub-carrier which arise. These disadvantages are a small price to pay in exchange for the ability to provide viewers with colour pictures of excellent definition directly from American-type pictures.

† 'Television standards conversion: an electronic switching system designed by the B.B.C.', *The Radio and Electronic Engineer*, 26, No. 3, pp. 242–4, September 1963.

U.D.C. 621.396.677 : 621.396.96

Radar Signal Processing for Angular Resolution beyond the Rayleigh Limit

By

A. A. KSIENSKI, Ph.D.†

AND

R. B. McGHEE, Ph.D.‡

Summary: The problem of angular resolution is formulated in a statistical decision theoretical context. A tutorial discussion of decision theory and its application to resolution is presented, and an explicit signal processing radar system is derived, the resolving properties of which approach an optimum. Computer simulation tests show that the system is capable of reliably resolving targets as close as a quarter of a 3 dB beam-width in the presence of noise and system errors.

1. Introduction

The problem of angular resolution has been of interest for many years. Its significance in the fields of radar, radio astronomy, optics, sonar and seismology has prompted researchers in all of these fields to devise various techniques for the improvement of resolution. Often, techniques developed in one field were borrowed by others.

Let us define first what we mean by resolution. It is the ability to identify the presence of two (or more) closely spaced point objects. The objects may be active sources such as in radio astronomy, or may be reflectors such as radar targets. The 'point' term in the definition may be generalized to extended objects by subdivision of the extended object into smaller areas, each representing a point object. Now the resolution limit could be stated as the maximum number of (equal) subdivisions per unit solid angle which may be allowed without causing two or more neighbouring subdivisions to merge in the output or image.

The conventional, or Rayleigh, resolution limit may be stated as follows. Two point sources can be resolved by a lens if they are separated in angle by at least half a null-to-null beam-width of the lens.[1] The resolution limit may then be related to the lens size or, in the case of linear arrays, to the length. If a uniformly illuminated line source of length D is considered, its far-field pattern is given by

$$E(\theta) = C\frac{\sin\left(\dfrac{kD}{2}\sin\theta\right)}{\sin\theta} \qquad \ldots\ldots(1)$$

where C is a proportionality constant, $k = 2\pi/\lambda$, and θ is the observation angle. The half null-to-null beam-width is then given by

$$\sin\theta = \frac{\lambda}{D} \quad \text{or} \quad \theta \simeq \frac{\lambda}{D} \quad \text{for} \quad \frac{\lambda}{D} \ll 1 \qquad \ldots\ldots(2)$$

This relationship links rigidly the resolving power of a radar or other mapping or searching device to the size of the available aperture and the operating wavelength.

Various attempts have been made to overcome this limitation, the most famous technique to date being the supergain or superdirectivity approach. Bouwkamp and de Bruijn[2] have shown that any desired gain or beam-width can be obtained from an antenna of given aperture size by appropriate excitation. This approach, although of academic interest, is of little practical use when arrays larger than a couple of wavelengths in size are considered.[3,4]

Another approach, which is applicable to arrays of any size, involves what is known as multiplicative, correlation or nonlinear processing arrays which were originally conceived by radio astronomers and later utilized in acoustic, sonar and radar applications.[5-9] This technique provides an improvement over the Rayleigh limit by about 20 to 50%, depending on the particular array geometry, its application, and its signal and noise characteristics. A third approach is the use of non-reciprocal arrays when one has control over the illumination of the object as well as the receiving array as in radar.[10] With this technique, the improvement in resolution is about 15%.

Multiplicative antennae are useful only when (input) signal/noise ratio (S/N) is high or when large integration times are available, i.e. when output S/N is high. This limitation is caused by the degradation of the S/N by the detection process, i.e. the multiplication of the output of two or more sub-apertures. The supergain array is even more sensitive to noise or random error problems.

The inherent failure in achieving significantly improved resolution (by, say, an order of magnitude) leads to several basic questions. First, what is the actual resolution limit and what does it depend on? Second, how do we attain the limit, or approach it closely?

If we are to base our approach on a deterministic picture, then the answers to both questions would be

† Aerospace Group, Hughes Aircraft Company, Culver City, California; now at the Electro Science Laboratory, Department of Electrical Engineering, The Ohio State University, Columbus, Ohio.

‡ University of Southern California, Los Angeles, California.

given by the supergain technique. It is quite clear, however, that random errors and noise make it impossible to realize this approach. Another limitation of the above techniques is that resolution is dependent on a rather primitive data extraction method, namely an amplitude display of the received signal on an oscilloscope or recorder. The appearance of one peak or two peaks decides whether there are one or two targets, respectively. In the presence of noise these peaks may fluctuate, merge or be suppressed and the onus is on the observer to decide what the display indicates. Judging from past experience with optimum data extraction methods, such as in target detection, the limits of performance are decided by signal and noise characteristics and by our *a priori* knowledge of these.[11] Thus, if the signal and noise characteristics are given, the probability of target detection and the miss probability for a given false alarm rate can be determined. It can also be learned with what accuracy the target's angular location can be determined, as well as its other parameters.[12]

The problem of resolution may be formulated on the same fundamental basis as the single target detection problem, and it will be shown that the limit of resolution is indeed dependent on the signal and noise characteristics of the targets and the environment, and our prior knowledge of them; in addition, resolution is, of course, dependent on aperture size in wavelengths.

2. Decision Theoretic Formulation of the Resolution Problem

The basic problem may be stated as follows: For an aperture of a given physical size and for specified signal and noise characteristics (to the extent they are known), how closely may n targets be spaced (in angle) and still be correctly identified as n distinct targets (with a given probability)? The next question is: How can a system that may achieve that limit, or at least closely approximate it, be constructed? Answers to both questions may be obtained by means of statistical decision theory. Various simplifications and assumptions will be made until the final results are obtained. However, none of them is essential and each may, and hopefully will, be gradually removed in future developments.

Let us begin by examining the information available at the aperture regarding the unknown target distribution. All that is known after the measurements are made may be expressed as a conditional probability density $p(s|r)$ where s is the signal representing exactly the target distribution, such as the number of targets and their parameters (range, angle, Doppler frequency, etc.), and r is the received signal as measured at the antenna aperture. Because of the presence of noise and measurement errors, the relationship between the measured values and the exact signal is given only in terms of a probability distribution. The explicit form of s may be obtained by specifying the target distribution and the illuminating waveform characteristics such as centre frequency, f_0, and bandwidth, $2W$. Since the form of s is required at the measuring points across the aperture, these points have to be specified. If Shannon's sampling theorem is extended to spatial sampling, it can be shown[13] that $\lambda/2$ ($\lambda f_0 = c$) spaced samples will represent the required spatial information.† The temporal sampling is at $\Delta t = 1/2W$ where $2W$ is the signal (and noise) bandwidth. Thus the jth temporal sample at the ith antenna element due to the αth target will be given by

$$s_{ij\alpha} = \gamma_\alpha S\left(t_0 - \frac{2R_\alpha}{c} - \frac{i\lambda \sin\theta_\alpha}{2c} + \frac{j}{2W}\right) \quad \ldots\ldots(3)$$

where R_α is the range to the αth target, θ_α is its angular location, and γ_α is its relative reflection amplitude. Since the exact waveform has not been assumed, the signal is given by $S(t_0+t)$, with the argument t representing time referred to some reference time, t_0. The noise and errors are assumed to be Gaussian and additive so that the measured values obtained over an observation interval T are related to the noiseless (exact) values by the conditional probability

$$p(r|s) = p_n(r-s) = \frac{1}{(2\pi)^{TWK}|\Lambda|^{1/2}} \times$$
$$\times \exp\left\{-\frac{1}{2|\Lambda|}\sum_{i=1}^{K}\sum_{j=1}^{2TW}\sum_{k=1}^{K}\sum_{l=1}^{2TW}|\Lambda|_{ijkl}(r_{ij}-s_{ij})\times \times (r_{kl}-s_{kl})\right\} \quad \ldots\ldots(4)$$

where s_{ij} contains the summation over α, the targets present (or assumed). The total number of elements in the array is given by K; hence, the aperture size is assumed to be given by $(K-1/2)\lambda$. A one-dimensional array is assumed; consequently, a single variable θ is used for the angle parameter. Λ_{ijkl} is the co-variance between the (ij)th noise sample and (kl)th one, $|\Lambda|$ is the determinant of the covariance matrix, and $|\Lambda|_{ijkl}$ is the cofactor of the element. The assumption

† This representation is approximate and, in fact, one could improve further the resolution by closer spacing. However, the amount of improvement obtainable may not warrant the effort of combating the mutual coupling effects. The formulation presented in this paper nevertheless permits any desired spacing, and moreover one could incorporate the mutual coupling effects in the signal s and thus obtain a more general expression for the likelihood function described below. Such a formulation might be desirable for antennae of very small electrical size, perhaps up to two wavelengths. The percentage improvement gained in the case of larger arrays using spacing smaller than $\lambda/2$, is negligible. For a detailed discussion on the amount of spatial information gained by close element spacing and its relation to supergain, see Reference 13.

regarding the noise (and error) distribution represents the overall 'system errors' which corrupt the signal. Any knowledge regarding this signal corruption can and should be incorporated into the distribution of $p(r|s)$ as in eqn. (4). From this distribution, which is often called the likelihood function, the *a posteriori* distribution $p(s|r)$, which is discussed above, can be obtained by means of Bayes's equation:

$$p(s|r) = \frac{p(r|s)p(s)}{p(r)} \qquad \ldots\ldots(5)$$

where $p(s)$ is the *a priori* distribution of the signal and $p(r)$ may be obtained from $p(r|s)$ and $p(s)$ by

$$p(r) = \int p(r|s)p(s)\,\mathrm{d}s \qquad \ldots\ldots(6)$$

Any information regarding any of the signal parameters is incorporated in $p(s)$ to reduce the uncertainty regarding s and thus improve the resulting accuracy, or reliability, of the estimate, or decision, regarding the target situation. The *a priori* information may include the size of the expected targets, excluding for example very large targets, or may relate to target velocity or angular location, assigning different probabilities to various ranges of the variable. In the absence of any prior information, or in the event of unwillingness to commit oneself for fear of wrongly biasing the outcome, a uniform distribution may be assumed.

The question of extracting the available information or converting it to a statement regarding the target distribution should now be considered. We are thus faced with the question of a strategy or decision rule, i.e. how to go about making an optimum decision (regarding s) based on the available information. The answer to this question will be obtained once we have decided what relative costs or penalty we will assign to wrong decisions. Since our knowledge of the true situation, represented by s, is incomplete, it is obvious that we may reach the wrong conclusion; that is, our estimate of s, which we will call $\hat{s}$, may or may not be equal to s.

Deviations from the correct value may be weighed in accordance with an arbitrarily assigned cost function. For example, the relative cost may be specified to be proportional to the square of the deviation,

$$C(\hat{s}, s) = (s-\hat{s})^2 \qquad \ldots\ldots(7)$$

or may be raised to a higher power. It may be weighed uniformly outside a certain 'distance' from the true value. Thus,

$$\begin{aligned} C(\hat{s}, s) &= 0 \qquad |s-\hat{s}| \leq \varepsilon \\ &= 1 \qquad |s-\hat{s}| > \varepsilon \end{aligned} \qquad \ldots\ldots(8)$$

This is called a zero-one cost-function which means that as long as the estimate is within a certain range of true value, it will be considered as good as if the true value were found; beyond that range, however, all deviations are equally bad.

Although the above estimates and deviations refer to the signal itself, it may be considered a representation of the various parameters involved. Thus, s is a multi-dimensional vector and so is $\hat{s}$, and ε, the permitted error, may, for example, have different values assigned to its different components. Error in target phase may be almost irrelevant while range or angle may be critical. Once a cost function is chosen, the optimum procedure would be decided from the following considerations: It is desired to minimize the average, or expected, loss (or cost)

$$R(r,\hat{s}) = \int p(s|r)C(s,\hat{s})\,\mathrm{d}s \qquad \ldots\ldots(9)$$

for any measured r. The only variable at our disposal is $\hat{s}$, the estimate which has to be chosen so as to minimize this average cost or risk. When $\hat{s}$ is so chosen, the minimum resulting risk is called Bayes's Risk and the method through which it is obtained is called Bayes's strategy or Bayes's or optimum decision rule.[14] That this is the optimum decision rule is intuitively obvious since we are choosing an estimate which will minimize the expected cost of the decision, where the expectation is obtained by using the probability distribution of the signal conditioned on the measurements made. We have thus incorporated all known information, expressed our feelings as to how to penalize wrong decisions, and chosen the procedure that will minimize the expected penalty for any measurements made.

Now, regarding the actual procedure of obtaining $\hat{s}$, since we are looking for an $\hat{s}$ that will minimize R in eqn. (9), we will equate the derivative of R with respect to $\hat{s}$ to zero and solve for $\hat{s}$. If a squared error cost function is assumed, this solution may be easily obtained:

$$\begin{aligned} \frac{\partial R}{\partial \hat{s}} &= \frac{\partial}{\partial \hat{s}} \int p(s|r)(s-\hat{s})^2\,\mathrm{d}s \\ &= 2\left[\int sp(s|r)\,\mathrm{d}s - \hat{s}\int p(s|r)\,\mathrm{d}s\right] = 0 \end{aligned} \qquad \ldots(10)$$

Since

$$\int p(s|r)\,\mathrm{d}s = 1$$

$$\hat{s} = \int sp(s|r)\,\mathrm{d}s \qquad \ldots\ldots(11)$$

This derivation is quite simple for the cost-function chosen. It may be considerably more complicated if the cost-function involves s and $\hat{s}$ in a more complicated relationship, such as the information theoretical cost-function.[15] Although eqn. (11) represents a simple mathematical relationship, it is extremely difficult to mechanize.[16] A simpler cost-function was therefore chosen for the computation and implementation of an actual radar system based on this approach.

The zero-one loss-function will be used and the signal s will be explicitly stated as dependent on the number of targets, n, and their parameter vectors, ξ_n. The loss-function is thus given by

$$C[(n,\xi_n),(\hat{n},\hat{\xi}_{\hat{n}})] = 0 \qquad |n-\hat{n}| \le \delta; \quad |\xi_n - \hat{\xi}_{\hat{n}}| \le \varepsilon$$
$$= 1 \qquad |n-\hat{n}| > \delta; \quad |\xi_n - \hat{\xi}_{\hat{n}}| > \varepsilon \qquad \ldots\ldots(12)$$

and for small ε and δ, $R_{\min}$ is given by

$$\min_{\hat{n},\hat{\xi}_{\hat{n}}} R[r,(\hat{n},\hat{\xi}_{\hat{n}})] = \min_{\hat{n},\hat{\xi}_{\hat{n}}} \int p(n,\xi_n|r) \times C[(n,\xi_n),(\hat{n},\hat{\xi}_{\hat{n}})]\,\mathrm{d}n\,\mathrm{d}\xi_n$$
$$= 1 - \left[\max_{n,\xi_n} p(n,\xi_n|r)\right]\varepsilon\delta \qquad \ldots\ldots(13)$$

since the value of the integral will be minimized by choosing $\hat{n}$, $\hat{\xi}_{\hat{n}}$ equal to the values of n, ξ_n for which $p(n, \xi_n|r)$ is maximum. (This choice will remove from the integrand the highest portion of the function.)

The signal description in terms of the number of targets and their parameters requires some comments. The dependence of $p(n, \xi_n)$ on n is quite different from its dependence on ξ_n. One difference is in the discrete nature of n as compared with the continuous values assumed by ξ_n. This discreteness not only converts the integral over n in eqn. (13) into a sum but also specializes the minimization of risk into a set of comparisons called 'hypothesis tests' with the cost function turning into a cost matrix in which a specific penalty is assigned to each wrong decision. This process is also termed 'detection' as compared with estimation of the values of ξ_n (for a given n) which is termed 'extraction'.[17] In the present case we are faced with both problems, i.e. joint estimation and hypothesis testing.

The second and most important difference between the dependence of the probability function on n and ξ_n is that n changes the dimensionality of the distribution. If ξ is assumed to represent M parameters (angle, range, Doppler resolution, etc.), each requiring a dimension in the probability distribution, it is apparent that the total number of dimensions required to describe the statistics of the target distribution is given by Mn. Thus, as n changes by one integer, the dimensionality of ξ_n changes by M. This difference is the reason for separating the variables of s into n and ξ_n which thus naturally separates the problem into detection and extraction phases.

Returning to eqn. (13) and using eqn. (5), we obtain

$$R_{\min} = 1 - \left(\max_{n\,\xi_n}\left\{\frac{p(r|n,\xi_n)p(\xi_n|n)p(n)}{p(r)}\right\}\varepsilon\delta\right) \qquad \ldots\ldots(14)$$

or

$$R_{\min} = 1 - \frac{\varepsilon\delta}{p(r)} \max_{n}\left\{\frac{p(n)}{V_n} \max_{\xi_n}\,[p(r|n,\xi_n)]\right\} \qquad \ldots\ldots(15)$$

where the separation between the detection and extraction phases is evident. Note that $p(\xi_n|n)$ has been assumed to be uniformly distributed over a hyper-volume, V_n, indicating a lack of more specific target information. Equation (15) describes the processing that is to be carried out to determine the target situation (n, ξ_n).

The first step is to choose the maximum expected number of targets within the search range, which will be (in angle) the beam-width of the illuminating radar. This number will therefore be rather small.† Once the maximum n is picked, an estimation procedure is carried out n times to determine the respective values of ξ_n.

The estimation is carried out, as indicated in eqn. (15), by maximizing the likelihood function $p(r|n, \xi_n)$. This process is called maximum likelihood estimation.‡

After the estimates for the various n are carried out, a multiple hypothesis test may be performed to determine the largest of the (normalized) maxima. This step leads, however, to serious difficulties that have not been solved in an optimum way. A satisfactory procedure may be devised to obtain n, the number of targets; such a procedure will be discussed in Section 3.

3. System Implementation

The most complicated part of the system described by eqn. (15) is the estimation process which is carried out by maximizing the likelihood function given by eqn. (4) with

$$s_{ij} = \sum_{\alpha=1}^{n} \gamma_\alpha S\left(t_0 - \frac{2R_\alpha}{c} - \frac{i\lambda\sin\theta_\alpha}{2c} + \frac{j}{2W}\right) \ldots\ldots(16)$$

where n is the number of targets assumed, γ_α are target amplitudes, R_α are target ranges with the term representing target phase as well (in fact, for targets at equal range, phase will replace the range term which is then lumped with t_0), θ_α are the angles, and the assumption has been made that no relative Doppler frequency is present in the returns.

At this point several additional assumptions will be made partly to reduce computation complexity and partly to facilitate experimental implementation. The noise added to the signal at each element of the array will be assumed independent of all other noise sources, since the output of each element in the array will be individually amplified. If receiver noise predominates,

† Practical considerations such as equipment complexity will also limit this number.

‡ Bayesian estimation reduces to a maximum likelihood estimate when a zero-one loss function is used and the signal distribution is assumed to be uniform, i.e. no *a priori* information available.

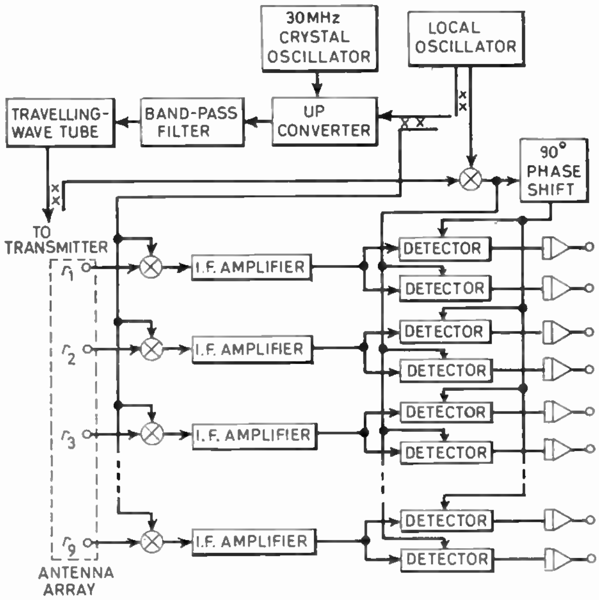


Fig. 1. System block diagram.

this assumption is quite realistic. The individual noise samples will similarly be independent because the sampling distance is $1/2W$. This independence reduces the covariance matrix to a diagonal matrix. The multi-dimensional distribution is thus converted into a product or into a summation in the exponent of the Gaussian distribution. This exponent, which we will term $L(r/n, \xi_n)$ is given by

$$L(r/n, \xi_n) = \frac{1}{2|\Lambda|}\sum_{i=1}^{K}\sum_{j=1}^{2TW}\sum_{k=1}^{K}\sum_{l=1}^{2TW}|\Lambda|_{ijkl}(r_{ij}-s_{ij})\times(r_{kl}-s_{kl}) \quad \ldots\ldots(17)$$

$$= \frac{1}{2\sigma^2}\sum_{i=1}^{K}\sum_{j=1}^{2TW}(r_{ij}-s_{ij})^2 \quad \ldots\ldots(18)$$

where the variances of the noise samples are assumed equal and given by σ^2.

It will be noted from eqns. (17) or (18) that minimizing L is equivalent to maximizing the likelihood function; hence, we will be concerned with the minimization of eqn. (18). The assumptions regarding the independence of noise sources are not essential and are unjustified for strong non-uniform external noise. (The computational problem of minimizing L under such conditions is more complicated but certainly feasible.) Several methods may be used in carrying out the minimization of eqn. (18). The approach taken here is that of converting the received r_{ij} into d.c. levels and performing a regression analysis in a digital computer. The reason for this approach is to minimize processing errors and r.f.–i.f. equipment complexity. Thus the received signal is coherently detected, digitized and fed to a digital computer which is programmed to search in the multidimensional parameter space until an absolute minimum is found for the quadratic function given by eqn. (18). The largest errors, except for noise, are encountered in the amplification and detection process; beyond that errors may be assumed relatively small. In the presently considered c.w. system without relative target Doppler, only two d.c. outputs are required per channel to represent the amplitude and phase of the signal at each element of the array. The system is shown in Fig. 1 and depicts part of an experimental system that has been implemented and tested. The required information is provided by two quadrature components per channel. The received signal is given by

$$r_i(t) = S_i(t)+n_i(t)$$
$$= \sum_{\alpha=1}^{n}\gamma_\alpha\cos(\omega_0 t-\phi_\alpha-i\pi\sin\theta_\alpha)+n_i(t) \quad \ldots\ldots(19)$$

where range difference as well as target phase are included in ϕ_α. The detector outputs are then

$$u_i(t) = \sum_{\alpha=1}^{n}\gamma_\alpha\cos(i\pi\sin\theta_\alpha+\phi_\alpha)+x_i(t) \quad \ldots\ldots(20)$$

$$v_i(t) = \sum_{\alpha=1}^{n}\gamma_\alpha\sin(i\pi\sin\theta_\alpha+\phi_\alpha)+y_i(t) \quad \ldots\ldots(21)$$

where $x_i(t)$ and $y_i(t)$ are the cosine and sine components of the pass-band noise represented by $n_i(t)$ and may be assumed to be zero mean, uncorrelated Gaussian processes.[18] Except for noise, $u_i(t)$ and $v_i(t)$ are two d.c. voltages which may now be digitized and fed to a computer for processing.

The required number of parameters to be determined is $3n$, and $2n$ of them appear in a non-linear form, in the phases and angles of the targets. This situation implies a non-linear regression problem in $2n$ dimensions plus a linear one in n dimensions. There are techniques that can handle multi-dimensional non-linear regression problems. However, it is preferable from the computational complexity and computer memory requirements to avoid or at least minimize the number of dimensions for which a non-linear regression analysis must be performed. In the case of the phase parameter, the non-linearity may be removed by an appropriate co-ordinate transformation:

$$u_i(t) = \sum_{j=1}^{n}[\alpha_j\cos(i\pi\sin\theta_j)-\beta_j\sin(i\pi\sin\theta_j)]+ +x_i(t) \quad \ldots\ldots(22)$$

$$v_i(t) = \sum_{j=1}^{n}[\alpha_j\sin(i\pi\sin\theta_j)+\beta_j\cos(i\pi\sin\theta_j)]+ +y_i(t) \quad \ldots\ldots(23)$$

where $\alpha_k = \gamma_k\cos\phi_k$, $\beta_k = \gamma_k\sin\phi_k$ are the new parameters appearing in linear form.

The remaining parameter, θ_j, cannot be linearized because of the i dependence of the argument. Rather than perform a continuous search over this parameter, it was decided to perform a discrete search, or scan over a grid of discrete values of θ_j separated by a tenth of a beam-width. For each such value assigned to each one of the targets, the values of α_j and β_j were computed that minimized eqn. (18). This procedure is very fast for a small number of targets; in the case of two targets the total computation including the discrete search requires a fraction of a second on an IBM 7094.

The method of computation will be now briefly described. Let us denote the measured or observable values by a vector [6]:

$$z = \begin{pmatrix} \int_0^T u_1(t)\,dt \\ \int_0^T v_1(t)\,dt \\ \cdot \\ \cdot \\ \cdot \\ \int_0^T v_N(t)\,dt \end{pmatrix} \qquad \ldots\ldots(24)$$

where integration over a period T is provided for preliminary smoothing. The assumed values or estimates will be denoted by

$$\hat{u}_i(t) = \sum [\hat{\alpha}_j \cos i\pi\hat{x}_j - \hat{\beta}_j \sin i\pi\hat{x}_j]$$

and

$$\hat{v}_i(t) = \sum [\hat{\alpha}_j \sin i\pi\hat{x}_j + \hat{\beta}_j \cos i\pi\hat{x}_j]$$

where $x_j = \sin\theta_j$. Thus the estimates vector is given by:

$$\hat{y} = T \begin{pmatrix} \hat{u}_1 \\ \hat{v}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{u}_N \\ \hat{v}_N \end{pmatrix} \qquad \ldots\ldots(25)$$

and the parameter vector (for a given set of angle variables $x_1, x_2, \ldots, x_n$) is given by

$$p = (\alpha_1, \beta_1, \alpha_2, \beta_2 \ldots \alpha_n, \beta_n) \qquad \ldots\ldots(26)$$

The error vector implicit in the Gaussian likelihood function (eqns. (17) and (18)) is (again for a given set of x values)

$$e(p) = z - y(p) \qquad \ldots\ldots(27)$$

and the exponent (eqn. (18)) is equal to

$$L(r/n, \xi_n) \simeq L(z/p) = \frac{1}{2\sigma^2}(e'e) \qquad \ldots\ldots(28)$$

The change of notation from r to z is due to the fact that r represents instantaneous (received) signals while z represents actual measured outputs which undergo filtering or integration. To stress the difference between theoretical considerations and actual computations which will be described from here on, the vector y is used instead of s, and p instead of ξ_n. It can be noted that the summation in eqn. (18) is replaced by an integral from zero to T. This is justified by the sampling theorem. We are thus looking for a parameter $\hat{p}$ which minimizes the least squares objective function:

$$Q(p) = e'e \qquad \ldots\ldots(29)$$

Let η be an arbitrary parameter change vector. Then

$$y(p+\eta) = y(p) + X\eta \qquad \ldots\ldots(30)$$

where

$$X = [X_{ij}] = \left(\frac{\partial y_i}{\partial p_j}\right) \qquad \ldots\ldots(31)$$

Note that a set of values for $\hat{x}_k$ is assumed in accordance with a discrete grid search. Hence,

$$p = (\alpha_1, \beta_1, \ldots, \alpha_n, \beta_n)$$

and does not involve x_k, the angular location. It is because of this relation that eqn. (31) completely characterizes the variation of y. Now,

$$e(p+\eta) = z - y(p) - X\eta \qquad \ldots\ldots(32)$$

and

$$Q(p+\eta) = (z - y - X\eta)'(z - y - X\eta) \qquad \ldots\ldots(33)$$

Equating the gradient of Q to zero and solving for η gives

$$\eta = [X'X]^{-1}X'e \qquad \ldots\ldots(34)$$

and

$$\hat{p} = p + \eta = p + [X'X]^{-1}X'e \qquad \ldots\ldots(35)$$

Since the initial value of p is arbitrary, assume $p = 0$. Hence, $y(p) = 0$ and $e = z$, so

$$\hat{p} = [X\ X]^{-1}X'z \qquad \ldots\ldots(36)$$

The estimated parameter values may be obtained by operating directly on the measured signals represented by z. It may also be noted that $[X_{ij}]$ depends only on $\hat{x}_k$ and not on α_k and β_k, which means that the matrix $[X'X]^{-1}X$ can be computed just once and stored for subsequent use in eqn. (36).

Following the estimation phase, the hypothesis testing phase should be carried out to determine which one of the estimates is the most acceptable or, in the case of uniform *a priori* distribution and zero–one loss function, which is the most likely state. This test, although quite simple to carry out, leads to substantial difficulties. The problem is, as pointed out by Root,[19] that the hypothesis that there are n targets includes the m target hypothesis where $m < n$, as a particular case where $(n-m)$ of the targets are of zero amplitude. In fact, in the presence of noise we can obtain a better fit to the signal by assuming more targets than are actually present, since this procedure is equivalent to permitting more terms in an approximation to an arbitrary curve. Root[19] suggests that the probability of resolution may be obtained by assuming that the target distribution is known exactly and then computing the probability that the likelihood function based on the correct hypothesis is higher than the one based on the wrong hypothesis. For the specific case of two targets against one, the target's amplitudes, angular locations, and all other parameters are assumed to be known exactly. Then the likelihood of the measured signal based on these assumptions is compared with the maximum likelihood based on a single target assumption. Now, if the probability,

$$P\left\{\frac{\max\limits_{\gamma,\varphi,x} p(z/1,\gamma,\varphi,x)}{p(z/2,\gamma_1,\varphi_1,x_1,\gamma_2,\varphi_2,x_2)} > 1\right\} > \eta \quad \text{......(37)}$$

where γ_1, φ_1, x_1, γ_2, φ_2, x_2 are the known two target parameters, and γ, φ, x are the single target parameters, z is the received signal, and η is some assigned threshold, then the targets are considered not resolvable.† This approach relies very heavily on the estimation accuracy. In the limit when the S/N ratio is very high it will yield the resolution limit for each target situation. To obtain the average resolution limit we would have to average the above over the various parameter values and weight each contribution with its probability of occurrence.

A similar approach was taken by O'Sullivan[20, 21] who computed the required S/N ratio for various target separations to insure 90% probability of resolution for two equal targets with identical parameters except for their angular locations. The required S/N ratio for the resolution of two targets separated by half a 3-dB beam-width was 13 dB. This result, however, is based on the assumption of exact knowledge of the two-target situation. O'Sullivan[20] justifies the assumption by specifying that the target angular parameters be estimated to within 1/10 of their separation. This estimation accuracy requires 28·5 dB S/N ratio which is then stated to be a sufficient S/N ratio for target resolution. There is little doubt that this approach yields a safe bound on resolution.

Root argues that there is no point in asking for fine resolution if the S/N ratio is not high enough so that the parameter estimates are indeed highly accurate. This argument, however, is not entirely justified. There is no question that the S/N ratio should be higher than that required for single target detection, but the next order of precision is not estimation but resolution.

These remarks bring up the question of what is meant by resolution. Roughly, what is expected is a statement that there are, say, two objects rather than one in some specified region of space. The parameters of those objects are of secondary importance and may be quite poorly estimated. In fact, resolution may be termed a very coarse estimation in which all that may be reliably said is that there are two objects but their parameters cannot be reliably determined. It is apparent therefore that the required S/N ratio for resolution should at most be that for relatively poor estimation. We are again led to the inherent relation between resolution and estimation which indicates that perhaps we should re-examine Root's dilemma (n targets state includes m targets state, where $n > m$) rather than avoid it by specifying very high values of S/N ratio. In fact, the dilemma is the result of an unrealistic implied definition of the number of targets present. The assumption is made that one will encounter a radar return corresponding to one, two or n distinct targets, and consequently, a hypothesis test may be carried out to determine which state corresponds to the true situation. But in reality the usual radar return is a composite wave which requires an infinite number of targets to be exactly represented. Usually, however, a relatively small number of targets will adequately describe the situation while the rest of the targets may be considered insignificant. Which of the targets are significant may be determined by their amplitudes relative to the noise variance. The situation is quite similar, as mentioned above, to the approximation problem in which an arbitrary curve is fitted by a polynomial. If the curve represents a noisy signal, we would normally ignore terms with absolute values which are small compared with the noise variance, since these terms are more representative of the noise than the signal.

Coming back to the dilemma of how to carry out a hypothesis test for the number of targets, it is clear that at least in the noiseless case the choice $m < n$ will be made if the auxiliary condition exists that target amplitude be bounded away from zero. In other words, n targets cannot be made to fit an m target situation if all n targets are required to be larger than some constant $\varepsilon > 0$. It thus becomes apparent that the presence or absence of targets is a quantitative matter: a definition is required of what constitutes a target and what is just noise. This definition must be related to the amplitude of the reflecting objects since, in general, radiation will be arriving from all angles;

† This equation is Root's equation (II) modified for angular resolution rather than range and Doppler resolution. Root's approach provides a resolution probability but does not provide a method for hypothesis testing, which is required for the design of a system.

radiation intensity is a continuous function of angle, and any discretization is arbitrary and must be related to some parameter of the distribution. One parameter of the distribution is amplitude.† We shall define a target as an object whose reflection amplitude is $\gamma > \gamma_{min}$, where γ_{min} is specified in accordance with the *a priori* knowledge regarding target size expected. A target situation of m targets implies that there are m targets, each larger than γ_{min}.

When what constitutes a target has been defined in terms of its reflecting amplitude, there remains the question of distinctness. Since each reflecting object has a finite angular extent (subtends a finite angle at the antenna's site), another dilemma may again arise because it can be validly stated that the finite object consists of an infinite number of point objects. Thus we must specify a minimum angular distance between reflecting objects, below which the returns will be considered as coming from one object. Let us specify this distance as $|x_1 - x_2| > \Delta x_{min}$ where $x_1 - x_2$ may in general be a vector distance in a parameter space of any desired number of dimensions.

From the above discussion, it follows that the number of targets present is determined by the parametric values of the object distribution, and consequently, the decision regarding the number of targets present must be based on the estimated values of these parameters. Specifically, the following hypotheses must be compared:

H_0: no target present, i.e.

$$\gamma_j < \gamma_{min}, \quad \text{all } j \qquad \ldots\ldots(38a)$$

or all targets are smaller than γ_{min}.

H_1: one target present, i.e. one and only one of the following inequalities is satisfied,

$$(\gamma_1 > \gamma_{min}), (\gamma_2 > \gamma_{min}), \ldots, (\gamma_N > \gamma_{min}) \ldots\ldots(38b)$$

or the following inequalities are all jointly satisfied:

$$(\gamma_1 > \gamma_{min}), (\gamma_2 > \gamma_{min}), \ldots, (\gamma_k > \gamma_{min}) \qquad k \leq N$$
$$|x_1 - x_2| < \Delta x_{min}, |x_1 - x_3| < \Delta x_{min},$$
$$|x_2 - x_3| < \Delta x_{min}, \ldots, |x_{k-1} - x_k| < \Delta x_{min},$$
$$(\gamma_{k+1} < \gamma_{min}), \ldots, (\gamma_N < \gamma_{min})$$

That is, one of the targets is larger than γ_{min} or two or more of the targets are larger than γ_{min}, but those that are larger are separated by less than the minimum distance Δx.

H_N: N targets present, i.e. all of the following inequalities are satisfied:

$$(\gamma_1 > \gamma_{min}), (\gamma_2 > \gamma_{min}, \ldots, (\gamma_N > \gamma_{min}),$$
$$|x_1 - x_2| > \Delta x_{min} \ldots$$
$$|x_1 - x_3| > \Delta x_{min}, |x_2 - x_3| > \Delta x_{min}, \ldots,$$
$$|x_{N-1} - x_N| > \Delta x_{min} \qquad \ldots\ldots(38c)$$

That is, all N targets are larger than γ_{min} and all target separations are larger than Δx_{min}.

We have described a set of non-overlapping and exhaustive hypotheses which uniquely define the alternative situations that may prevail. We will now describe the appropriate tests to be performed on the received data to determine which hypothesis is to be chosen.

In order to test the hypotheses H_i against one another, it is necessary to define a loss-function which depends only on n, the true number of targets present, and $\hat{n}$, the estimate of n. Once again a one–zero loss-function will be chosen. For example,

$$C(n, \hat{n}) = 0 \quad \text{if} \quad \hat{n} = n$$
$$C(n, \hat{n}) = 1 \quad \text{if} \quad \hat{n} \neq n \qquad \ldots\ldots(39)$$

With this (or any other) loss-function, eqn. (9) reduces to the sum

$$R(r, \hat{n}) = \sum_{n=0}^{N} P(n/r)C(n, \hat{n}) \qquad \ldots\ldots(40)$$

where P denotes a discrete distribution.

The probability distribution-function, $P(n/r)$, is given by,

$$P(n/r) = \int \frac{p(r/\xi_n)p(\xi_n/n)P(n)}{p(r)} \, d\xi_n \qquad \ldots\ldots(41)$$

where ξ_n again refers to the parameters (such as γ, φ, x, etc.) of the target complex.

With the loss-function given by eqn. (39), $R(r, \hat{n})$ is minimized by choosing for $\hat{n}$ the value of n which maximizes $P(n/r)$. Thus $\hat{n}$ is again a maximum likelihood estimator. Equation (41) must be evaluated for each value of n. While this is conceptually straightforward, such a computation is extremely lengthy due to the high dimensionality of the integral and, consequently, this method of choosing the best hypothesis does not appear to be practical at the present time.

In view of the difficulties associated with eqn. (41), it is necessary to choose a simplified criterion for selecting the best hypothesis. Since maximum likelihood estimates are known to be functionally related to sufficient statistics (when sufficient statistics exist), a decision can be made on the maximum likelihood estimates for γ, φ, x, etc., just as well as from the raw data, r.[22] Thus the problem of finding an optimum test of hypotheses can be reformulated as the problem of defining optimum *acceptance* regions for each hypothesis in the $\hat{\gamma}$, $\hat{\varphi}$, $\hat{x}$ space. For the remainder of

† Polarization may be another parameter. In pattern recognition, the presence of cultural objects might be characterized by polarization properties.

this paper, we will restrict ourselves to consideration of such regions for the two-target case.

Specifically, we would like to determine whether there are two targets, one target, or none. We will determine that by comparing the amplitudes of the estimated targets to a threshold, $\gamma_{thresh.}$, and (when γ_1 and γ_2 are larger than $\gamma_{thresh.}$) target separation, Δx, to $\Delta x_{thresh.}$. The threshold values chosen are not necessarily γ_{min} and Δx_{min}, respectively. Thus $\gamma_{thresh.}$ should be related to the receiver noise level and to measurement errors. For example, in the presence of noise with standard deviation, σ, several times greater than the value of γ_{min}, we would end up with numerous erroneous decisions of the false alarm type if $\gamma_{thresh.}$ were chosen equal to γ_{min}. On the other hand, setting the threshold too high will cause excessive miss probabilities.

The two limits lead to the question of an optimum threshold which may be determined by either minimizing one type of error (say failure to detect), subject to a specified other type of error (such as false alarm), as in the approach of Neyman and Pearson, or by minimizing total error. A detailed study of this type is now in progress but is not completed yet. Therefore, we will choose a threshold which is twice the standard deviation of the noise, a somewhat conservative threshold, but in the absence of an adequate false alarm study, it was felt that a lower threshold might yield a too optimistic picture of resolution. We will then show the probability of correct resolution for that threshold.† As to Δx_{min}, similar considerations apply although in the present study other considerations such as computational complexity led to a choice of $\Delta x_{thresh.}$ to be one-tenth of the conventional antenna beam-width. This value coincides with the grid size used in the angle search of the estimation process. Thus to be recognized as two targets, the estimated angular locations are required to be separated further than one grid quantum. This angular separation appeared sufficiently large not to be too vulnerable to noise, and on the other hand was smaller than the expected minimum resolvable distance, even for large values of S/N ratio.

It might be appropriate at this point to re-emphasize that the objective of this study was not only to obtain theoretical resolution limits but also to develop realistic methods for realizing them (at least approximately) and moreover to design and build an experimental model to demonstrate the results. Thus, simplifications were made at various points in order to obtain a realizable system. An important merit of this study, it is felt, is the explicit way in which these simplifications were made. This allows future studies gradually to remove the simplifications and approach more closely the optimum.

† As discussed in Section 4, Computer Results, this choice of threshold proved adequate to ensure low false-alarm rates.

4. Computer Results

To assess the performance of the proposed system, two alternatives are available. One is to compute analytically the probability of resolution, and the other is to carry out simulation experiments. The analytical approach is extremely difficult because of the large dimensionality and non-linearity of the problem. The simulation experiment, on the other hand, carries an added advantage. The computer algorithm not only provides the required results but can also be utilized for the actual system implementation. Thus, the simulation provides an actual test of part of the experimental system.

In the first part of the computer experiment, the output of an 11-element $0{\cdot}5\lambda$ spaced array was simulated on an IBM 7094 computer. Thus 22 d.c. levels, corresponding to the quadrature components of the coherently detected array output (see Fig. 1 where a 9-element system is shown) were generated to represent both signal and noise of varying characteristics. In this simulation, the relative cross-sections of two targets, their relative r.f. phases, and their angular spacings were all varied systematically. For every set of target parameters, ten statistically independent trials were made using a different simulated receiver noise† error for every trial. This experiment was repeated for several different signal/noise ratios. Altogether, 4200 simulated resolution experiments were carried out.

The simulated output was then subjected to the parameter estimation and hypothesis testing to determine the presence of the targets and their parameters. Figures 2 and 3 summarize the results of these resolution experiments. As indicated on these curves, each data point represents an average over five different relative phase shifts or fifty trials altogether. Since each trial results in either resolution or no resolution, the observed relative frequencies must follow a binomial distribution. The 95% confidence intervals for the true probability of resolution for a sample of fifty are $\pm 0{\cdot}14$ in width at $\hat{P} = 0{\cdot}5$ and slightly less than this value at larger or smaller values of $\hat{P}$. These confidence intervals are sufficient to account for the small anomalies appearing on some of the curves.

The ordinate of all three figures is the observed relative frequency of resolution, $\hat{P}$:

$$\hat{P} = \frac{\text{number of trials in which resolution occurred}}{\text{total number of trials}} \qquad \ldots\ldots(42)$$

† This receiver noise may also correctly represent additive random system errors which are independent from channel to channel.

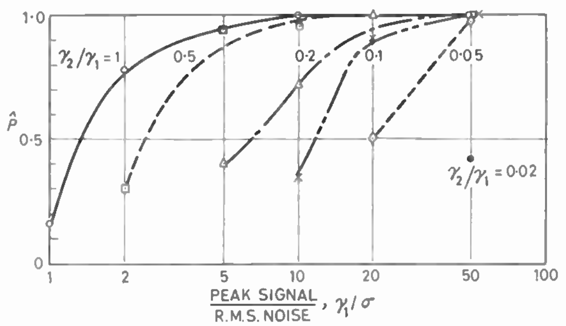


Fig. 2. Graphs showing $\hat{P}$ against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}10$ averaged over $\delta = 0, \pi/4, \pi/2, 3\pi/4, \pi$; 11 elements $\lambda/2$ apart.

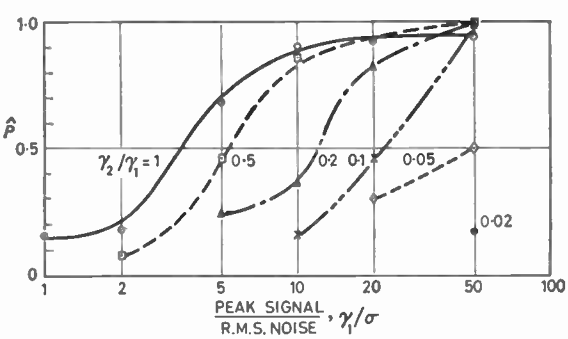


Fig. 3. Graphs of $\hat{P}$ against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}05$ averaged over $\delta = 0, \pi/4, \pi/2, 3\pi/4, \pi$; 11 elements $\lambda/2$ apart.

The γ and Δx thresholds discussed in the previous Section of this paper were used to define resolution in every case. The abscissa on the curves is the peak signal amplitude to r.m.s. noise ratio, γ_1/σ (expressed linearly rather than in dB).

The curves for $x_1 = 0$, $x_2 = 0{\cdot}10$ (Fig. 2) are perhaps the most remarkable of the results obtained. This target spacing corresponds to about one-half of the conventional half-power beam-width. Yet, even when the signal/noise amplitude ratio after integration is as low as 2 (3 dB power ratio), the probability of resolution is 0·78 for two targets of equal strength. At signal/noise ratios greater than 10 (17 dB power ratio), this probability rises to essentially unity.

Figure 3 describes the resolution probabilities for two targets separated by approximately one quarter of a 3 dB beam-width. The somewhat erratic spread of resolution points indicates that the error variance is increasing as resolution of closer targets is attempted. Also a larger number of trials may be needed to obtain the means of the resulting distributions. At any rate it is clear that resolution does occur with a high probability for a larger value S/N ratio. Thus, for a peak signal/r.m.s. noise ratio of 10 (or 17 dB mean power ratio), two targets of equal size will be resolved with 90% probability.

To evaluate the effects of a reduced number of elements in the antenna array, the simulated resolution experiments with the 11-element array were repeated for an antenna with nine elements. Once again 50 trials were carried out on the digital computer for each experimental condition. In addition to the reduced number of elements, the particular array used for the new experiments had an element spacing of $0{\cdot}8\lambda$ rather than the $0{\cdot}5\lambda$ value of the earlier array. This spacing corresponded to the actual inter-element spacing of the experimental system under construction. With this spacing, the antenna (employed in the conventional linear mode) 3 dB beam-width is approximately 0·14 radian. As before, the resolution experiments made use of a simulated strong target at $x_1 = 0$ and a weaker target located either at $x_2 = 0{\cdot}05$ or $x_2 = 0{\cdot}10$. The grid of values used for searching over x_1 and x_2 was based on a net with spacing of 0·02 covering the range from $x = 0{\cdot}00$ to $x = 0{\cdot}18$ inclusive. Thus, the points searched encompassed both pairs of values for x_1 and x_2 used in the simulation. In all other respects, the experiments performed with the 9-element array were identical with those performed for the 11-element array.

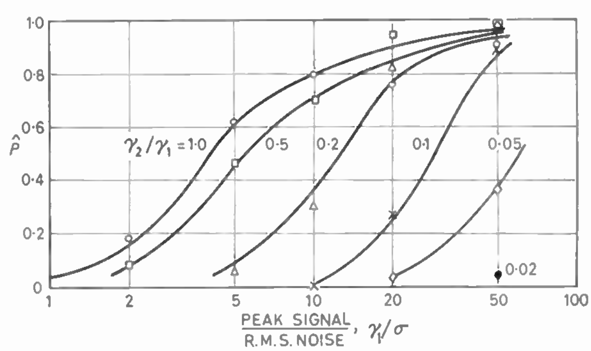


Fig. 4. Probability of two target discrimination against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}05$, 9 elements $0{\cdot}8\lambda$ apart.

Figure 4 shows the resolution probabilities observed for a target spacing of $\Delta x = 0{\cdot}05$, which corresponds roughly to one-third of a 3 dB beam-width for this array. Comparison of Figs. 3 and 4 shows that there is a small degradation in performance resulting from reduction of the array size. Since both sets of curves are best defined in the vicinity of $\hat{P} = 0{\cdot}5$, a reasonable quantitative comparison can be made by tabulating the value for γ_1/σ at which $\hat{P}$ takes on this value. Table 1 presents this comparison.

An important, although expected, property of decision theoretic arrays is illustrated by the results of

Table 1

Relative resolution capabilities of 9-element and 11-element arrays for small target separation.

γ_2/γ_1	γ_1/σ for $\hat{P} = 0{\cdot}5$	
	9-element array	11-element array
1	4·0	3·6
0·5	5·5	5·5
0·2	13	12
0·1	30	22
0·05	60	50

$x_1 = 0,\ x_2 = 0{\cdot}05$

Table 2

Relative resolution capabilities of 9-element and 11-element arrays for larger target separation.

γ_2/γ_1	γ_1/σ for $\hat{P} = 0{\cdot}5$	
	9-element array	11-element array
1	2·1	1·3
0·5	3·6	2·4
0·2	8·5	6·5
0·1	17·5	11·7
0·05	36	20

$x_1 = 0,\ x_2 = 0{\cdot}10$

this experiment. The 3 dB beam-width of the 11-element array with $0{\cdot}5\lambda$ spacing is approximately 0·18 radians. According to the Rayleigh criterion, the 9-element array with its smaller 3 dB beam-width ($\theta = 0{\cdot}14$) should resolve targets better than the 11-element array. Table 1 shows that this prediction is not borne out by experiment. This apparent contradiction comes about because the optimal quality of the decision-theoretic array permits it to make optimum use of every observation, and it is therefore able to separate targets more accurately with eleven observations than with nine. The number of antenna elements and the accuracy with which the radiation from each target is detected affect the resolution capability of the array as much as its overall size.

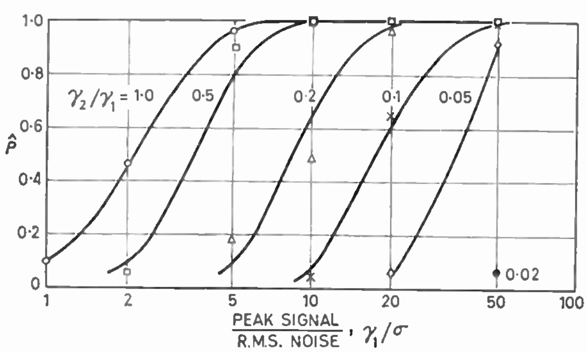


Fig. 5. Probability of two target discrimination against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}10$, 9 elements $0{\cdot}8\lambda$ apart.

Figure 5 presents observed detection probabilities for targets separated by $\Delta x = 0{\cdot}10$ or slightly more than two-thirds of the 3 dB beam-width of the array. Again, the performance of the 11-element array is better than that of the 9-element array. Table 2 provides a quantitative comparison analogous to Table 1. In every respect, the results obtained for $\Delta x = 0{\cdot}10$ are in agreement with those for $\Delta x = 0{\cdot}05$.

From the above results it is quite clear that one may expect to resolve targets separated by half of a 3 dB beam-width with S/N ratio of the same order of magnitude required for single target parameter estimation rather than that required for two-target parameter estimation where an increase of an order of magnitude is required for estimating the location of two targets spaced a half 3-dB beam-width apart (Figs. 2–16 of Ksienski[21] †; Sklar and Schweppe[23]). It is interesting to compare the resolution requirements in S/N ratio with those obtained by Root and previously obtained results (Figs. 2–20 of Ksienski[21]) based on the assumption that the exact target situation is known and a hypothesis test is carried out comparing the likelihood function based on the (correct) two-target assumption with the maximum likelihood of one target assumption. For 90% probability of resolving two targets half a 3 dB beam-width apart, the required S/N ratio from Figs. 2–20 of Ksienski,[21] is 13 dB; from Fig. 2, the ratio γ/σ is approximately 4. For a quarter beam-width resolution these figures[21] indicate 23 dB; Fig. 3 shows γ/σ equals approximately 15. To obtain a fair comparison we should take into account the fact that there are eleven elements. The effective S/N ratio for the array is thus improved by that ratio; hence the figures to be compared with the curve of Figs. 2–20 of Ksienski[21]; are 16 dB for the half-beam and 22 dB for the quarter beam-width separation.

In view of the high resolution probabilities achieved, it might be suspected that the false alarm probabilities are also quite high. This, fortunately, is not the case as can be shown by an examination of the various types of false alarms and miss probabilities that may occur in a multiple target situation. In the case of a system capable of detecting N targets, there are N^2 relevant conditional probabilities that should be

† The S/N ratio requirements for two-target resolution and the estimation of their parameters were computed by M. R. O'Sullivan who was one of the contributors to Reference 21.

considered. These probabilities can be presented in a matrix form.

$$P = \begin{bmatrix} P(0/0) & P(0/1)\ldots P(0/N) \\ P(1/0) & P(1/1) \quad P(1/N) \\ \cdot & \cdot \quad \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \quad \cdot \\ P(N/0) & P(N/1)\ldots P(N/N) \end{bmatrix} \qquad \ldots\ldots(43)$$

The diagonal elements in the matrix are 'detection' probabilities or probabilities of correct decision. The off-diagonal elements are false alarm and miss probabilities, i.e. probabilities of wrong decision. The objective, of course, is to maximize the diagonal elements and minimize the off-diagonal ones. The weights attached to the various errors will usually not be uniform since a decision, for example, of 'no targets' when N targets are present appears to be a more serious mistake than a decision of 'N targets' when $(N-1)$ targets are present. Thus minimizing risk may involve a stronger suppression of some errors than others.

The study of optimum thresholds that would minimize risk for various target–noise environments is a subject of future studies and will not be attempted presently. It is, however, important to assess the miss and false alarm probabilities for a selected threshold, in addition to the detection probabilities, to prevent an overly optimistic picture of the system's performance. Considering the specific two-target situation presently of interest, there are two conditional probabilities which may be termed miss probabilities (in the presence of two targets): $P(0/2)$ and $P(1/2)$. In view of the fact that $P(0/2)+P(1/2)+P(2/2) = 1$, it is apparent that the miss probabilities approach zero as $P(2/2)$, the detection probability, approaches unity (see Fig. 5).

There are also two probabilities which represent the false alarm error, $P(2/0)$ and $P(2/1)$. An estimate of $P(2/1)$ may be obtained from Fig. 5. The curve of $(\gamma_2/\gamma_1) = 0{\cdot}1$, for example, represents essentially a single target for $(\gamma_1/\sigma) < 1{\cdot}0$. It can be noted that $P(2/1)$ is then quite small and grows smaller as γ_1/σ is reduced. It is thus apparent that for $\gamma_1/\sigma \ll 1$, $P(2/1)$ will be negligibly small. $P(2/0)$ cannot be obtained from the curves of Fig. 5 but has been found to be vanishingly small for $\gamma_{\text{thresh.}} = 2\sigma$. From the above discussion and from further tests which have been recently performed, it appears that the threshold chosen (2σ) in the present experiment is adequate for insuring small false alarm probabilities.

4.1 *Estimation Errors*

The main objective of the present study was to find the limit of resolution and attempt to realize it by the design of an appropriate system. However, in the process of resolution an estimation had to be performed. Consequently, it was interesting to examine the results of this process as well and find the errors incurred in the estimation. The most significant parameter is, of course, the angular location of the targets, x_1 and x_2.

The r.m.s. error in the estimates of x_1 and x_2 for the trials in which resolution occurred at $\Delta x = 0{\cdot}05$ are presented in Figs. 6 and 7, respectively. Since a total of 50 simulated resolution experiments were carried out under each experimental condition, these errors represent an average of approximately 50 trials, the exact number depending on the number of cases in which resolution occurred. Missing data points on these curves correspond to cases in which the number of resolutions was too small to give a statistically meaningful estimate of the true r.m.s. estimation error.

Figure 6 shows that the error in estimating the stronger target position, x_1, is more or less independent of the weaker target amplitude providing that resolution has occurred. Figure 7 appears to indicate a strong connection between the estimation errors for the weaker target position, x_2, and the ratio of the weak to strong target amplitudes, γ_2/γ_1. This effect is only apparent, however, since it is primarily the result of plotting against γ_1/σ rather than γ_2/σ. Replotting can be effectively accomplished by merely sliding each curve horizontally to the left by an amount equal to γ_2/γ_1. When this change is made, the dependence on γ_1 largely disappears.

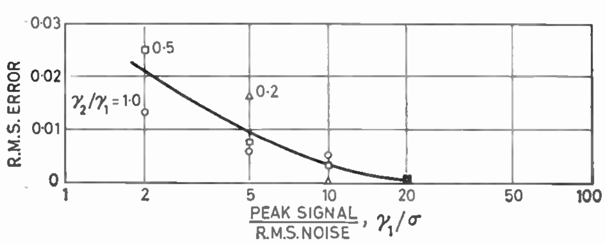


Fig. 6. R.M.S. error in estimation of x_1 against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}05$, 9 elements, $0{\cdot}8\lambda$ spacing.

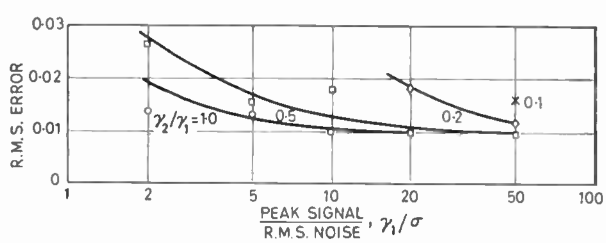


Fig. 7. R.M.S. error in estimation of x_2 against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}05$, 9 elements, $0{\cdot}8\lambda$ spacing.

As noted previously, the value for $x_2 = 0{\cdot}05$ lies midway between the adjacent values $x_2 = 0{\cdot}04$ and $x_2 = 0{\cdot}06$ used in the search for the angle variables. For this reason, the r.m.s. error in x_2 is asymptotic to 0·01 rather than to zero for large signal/noise ratios.

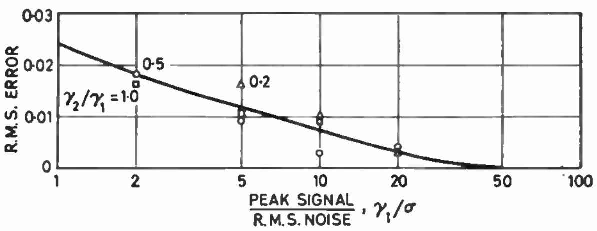


Fig. 8. R.M.S. error in estimation of x_1 against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}10$, 9 elements, $0{\cdot}8\lambda$ spacing.

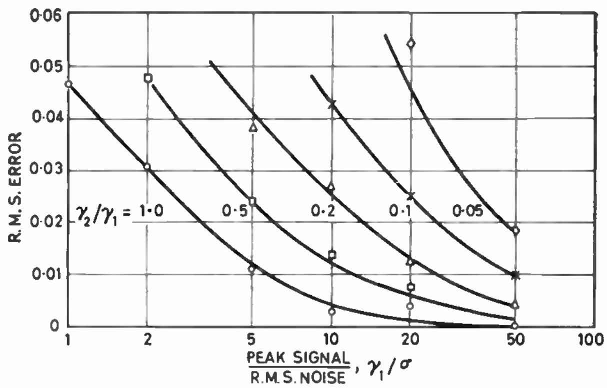


Fig. 9. R.M.S. error in estimation of x_2 against γ_1/σ for $x_1 = 0$, $x_2 = 0{\cdot}10$, 9 elements, $0{\cdot}8\lambda$ spacing.

Plots of average estimation error for the case when $x_1 = 0$ and $x_2 = 0{\cdot}10$ are presented in Figs. 8 and 9. Once again, the estimation error for each target position appears to be independent of the relative strength of the other target. Moreover, comparison of these curves with those of Figs. 6 and 7 reveals that, unlike the detection probability, the estimation error for resolved targets apparently does not depend significantly on target separation. This condition may not be altogether unexpected since, as mentioned, only the resolved cases are tested for estimation errors. These cases probably correspond to conditions in which the particular noise sample was weaker than the average and would thus permit resolution as well as small estimation error.

Taken all together, the results of the experiment indicate that the detectability of a target or the accuracy with which its parameters may be estimated, is relatively insensitive to the strength of a neighbouring target. This result is contrary to what would be expected from classical optics, but in agreement with decision theory and regression analysis.

5. Conclusions

The results obtained by means of the decision theoretic approach are encouraging in that a resolution improvement of at least three to one and perhaps four to one over the Rayleigh limit is feasible with reasonable S/N ratio requirements. Certain significant differences between the present approach and those previously attempted to improve resolutions should be stressed.

First, compared with the supergain or superdirectivity approach, the present approach does not require an exponential rise in accuracy and signal/noise ratio when the number of elements in the array increases. In fact, for a given percent of resolution improvement, the required S/N ratio is independent of array size. Also, the high-Q problem or large reactive energy and extremely narrow bandwidth is not present.

Compared with other approaches, such as multiplicative arrays, the present approach does not cause multiple target interactions and distortions since the present detection process is linear. It does not cause any sacrifice in the S/N ratio or the gain. In fact there is no compromise involved except perhaps in the processing complexity.

The most significant advantage of the decision theoretic approach is that it provides an explicit method for obtaining the optimum resolution utilizing to best advantage all given information and the measured data.

The particular system developed in this paper involves several simplifications as pointed out in the body of the paper. For example, no *a priori* information regarding target distribution was assumed and a very simple cost function was used; given more information and a more sophisticated cost function, a better resolution performance may be expected. In addition, the hypothesis testing procedure utilized in the experiment is non-optimum since it was not obtained from minimum risk considerations. An optimum procedure would therefore further improve the system's performance.

On the other hand, the number of parameters which were varied was reduced from the most general case by assuming no relative Doppler or range. Neither scintillation nor jamming were considered, and both factors may degrade resolution. A detailed investigation of a more general target and noise environment is needed. These more involved investigations are now justified in view of the significant resolution improvements obtained for two targets at identical ranges and no relative Doppler.

In closing, we would like to report an additional very encouraging fact. The system described in this paper has been built and initial tests have been performed. Two targets at a quarter of a 3 dB beam-width apart were indeed resolved quite well and with relatively small estimation errors. A thorough study of the system's performance has just begun and will be reported as soon as it is completed.

6. Acknowledgment

The work reported in this paper was sponsored by the Air Force Cambridge Research Laboratories, Office of Aerospace Research, Bedford, Massachusetts, under Contract AF 19(628)–4358, and in part by the U.S. Army Electronics Command, Fort Monmouth, N.J., under Contract DA 36–039 SC–90772.

7. References

1. M. Born and E. Wolf, 'Principles of Optics', p. 333 (Pergamon Press, London, 1959).
2. C. J. Bouwkamp and N. G. de Bruijn, 'The problem of optimum antenna current distribution', *Philips Res. Reports*, **1**, pp. 135–58, 1946.
3. L. J. Chu, 'Physical limitations of omnidirectional antennas', *J. Appl. Phys.*, **19**, pp. 1163–75, 1948.
4. R. F. Harrington, 'Effect of antenna size on gain bandwidth and efficiency', *J. Res. Nat. Bur. Stands*, **64D**, pp. 1–12, 1960.
5. A. E. Covington and N. W. Broten, 'An interferometer for radio astronomy with a single-lobed radiation pattern', *Trans. Inst. Radio Engrs on Antennas and Propagation*, **AP–5**, p. 247, 1957.
6. V. G. Welsby, 'Multiplicative receiving arrays: the angular resolution of targets in a sonar system with electronic scanning', *J. Brit. Instn Radio Engrs*, **22**, pp. 5–12, July 1961.
7. E. Shaw and D. E. N. Davies, 'Theoretical and experimental studies of the resolution performance of multiplicative and additive aerial arrays', *The Radio and Electronic Engineer*, **28**, pp. 279–91, November 1964.
8. R. Blommendaal, 'A note on multiplicative receiving systems and radars', *The Radio and Electronic Engineer*, **28**, pp. 317–25, December 1964.
9. A. A. Ksienski, 'Multiplicative processing antenna systems for radar applications', *The Radio and Electronic Engineer*, **29**, pp. 53–67, January 1965.
10. R. L. Mattingly, 'Non-reciprocal radar antennas', *Proc. I.R.E.*, **48**, 795, 1960.
11. D. A. Middleton, 'Introduction to Statistical Communication Theory', pp. 801–940 (McGraw-Hill, New York, 1960).
12. L. E. Brennan, 'Angular accuracy of a phased array radar', *Trans. I.R.E.*, **AP-9**, pp. 268–76, 1961.
13. A. A. Ksienski, 'Spatial frequency characteristics of finite aperture antennas', Electromagnetic Theory and Antennas, pp. 1249–67, Ed. E. C. Jordan (Pergamon Press, New York 1963).
14. A. Wald, 'Statistical Decision Functions, (John Wiley, New York, 1950).
15. D. A. Middleton, loc. cit., pp. 1008–24.
16. C. Giese and R. B. McGhee, 'Estimation of nonlinear system states and parameters by regression methods', Trans. 1965 Joint Automatic Control Conference.
17. D. A. Middleton, loc. cit. pp. 940–1008.
18. J. L. Lawson and G. E. Uhlenbeck, 'Threshold Signals', pp. 152–3, Vol. 24, M.I.T. Radiation Laboratory Series (McGraw-Hill, New York, 1950).
19. W. L. Root, 'Radar resolution of closely spaced targets', *Trans. I.R.E. on Military Electronics*, **MIL-6**, pp. 197–204, March 1962.
20. M. R. O'Sullivan, 'A statistical decision theory approach to optimum processing of the outputs of a receiving antenna array', International Conference on Microwave Circuit Theory and Information Theory, Tokyo. (For details see ref. 21.)
21. A. A. Ksienski, 'Very High Resolution Techniques', Eighth Quarterly Report, Contract No. DA 36(039)–SC90772, Hughes Aircraft Company, Culver City, California, 1964.
22. H. Cramer, 'Mathematical Methods of Statistics', p. 499, (Princeton University Press, 1951).
23. J. R. Sklar and F. C. Schweppe, 'The Angular Resolution of Multiple Targets', M.I.T. Lincoln Laboratory Group Report, No. 2, 1964.

Manuscript first received by the Institution on 6th March 1967 and in final form on 15th May 1967. (Paper No. 1140/RNA 73.)

U.D.C.621.398:627.92

The Remote Control of Lighthouses and Beacons

By

A. C. MACKELLAR,
C.Eng., F.I.E.E.†

Presented at a meeting of the Institution's Radar and Navigational Aids Group, held in London on 15th February 1967.

Summary: The general philosophy behind the remote control of lighthouses and the fundamental principles involved in operating such a system where the maintenance of the navigational service is of paramount importance are outlined. The characteristics of plant suitable for remote control and its integration into a system are described. The number of control functions and monitoring facilities required for each item of plant are discussed, together with an outline of a comprehensive scheme designed to accommodate a number of fully-operational stations operating on a group basis. The technical merits and reliability of various systems of communication are considered and a comparison made on an economic basis for rented circuits, the public telephone network and radio links of differing path length. Finally, details are given for a group of stations which are now being brought under remote control as an experiment to establish the feasibility of the scheme and to determine whether there is a significant economic advantage in operating lighthouses in this way.

1. Introduction

In this paper a general outline of the philosophy behind modern remote control systems for lighthouses is given and the minimum requirements in control and monitoring for a fully equipped station are discussed. Consideration is given to the economics of different communication systems.

The remote control of lighthouses and other aids to marine navigation has become increasingly common in recent years for a number of reasons, usually concerned with economics and the difficulty of recruiting staff who are prepared to spend long periods at isolated sites with few of the amenities normally associated with modern living.

There are basically two modes of operation employed by lighthouse authorities, depending on the location of the site and the complexity of the plant which is provided.

Scheme 1: Automatic operation with remote monitoring and with facilities for remote control.

Scheme 2: Full remote control.

Because of the importance which is attached to maintaining the navigational service at all times, Scheme 1 is the system which has been adopted in England as it is felt that this arrangement gives the highest probability of maintaining the service. Each installation is designed on a completely automatic basis with each function controlled by a sensing element actually situated at the lighthouse. Under these circumstances the remote control system merely assumes the role of a monitoring system to provide information at the control centre that the station is operating satisfactorily. Control facilities are, of course, provided so that action may be taken should the sensing element fail to initiate the appropriate function at the desired time.

By adopting this system, the satisfactory operation of the remote station is no longer dependent upon the communication path which inevitably carried a significant probability of failure.

Scheme 2, here designated 'Full Remote Control', has been adopted by some lighthouse authorities but generally for stations where the facilities provided are few and where the remote station is relatively close to the control point, usually within visual distance. The advantages of this system are usually of an economic nature, but these are offset by the difficulty of ensuring that an installation of this type is 'fail safe'.

2. Navigational Facilities

The navigational facilities which could be required at a major lighthouse and the impact of each on the remote control system are as follows:

(1) Navigational lights

(2) Fog signal

(3) M.F. radio beacon

(4) Radar beacon

(5) Stand-by power plant

† Engineer-in-Chief's Department, Corporation of Trinity House, London.

2.1 *Main Navigational Light*

As the word 'lighthouse' implies, the navigational lights are considered to be the most important navigational aids provided and, in consequence, considerable care is taken to ensure that the probability of failure is low and that the probability of not exhibiting a light during the hours of darkness, virtually non-existent. The navigational lights provided can comprise, in addition to the main navigational light, one, or in some cases two sector lights which are operated at the same time as the main navigational light. It has therefore been the practice to arrange for the control function which controls the main navigational light to exercise control of the sector lights with a resultant economy in the number of control functions required.

At most lighthouses the main navigational light will have an apparent intensity which lies between 300 000 candela and 5 000 000 candela, depending on the light source employed and the type of optical system used to concentrate the light output from the lamp into an acceptable beam.

Most installations employ filament-type lamps in a rotating optical system, but in recent years a number of systems have been installed using xenon arc discharge lamps with fixed drum lenses and also xenon capacitor discharge tubes with reflecting surfaces in place of lenses. Invariably it is arranged that duplicate equipment is provided and in the case of the filamentary lamps there is usually a third lamp of lower power rating designed to operate directly from the stand-by battery power source. With a rotating optical system it is not considered likely that a fault will occur in the optical system which will impair its performance so that the stand-by takes the form of a lamp exchanger which automatically brings into the focal point of the optic a new lamp in the event of failure of the operational lamp. In the case of failure of all power, i.e. mains and stand-by generator, the lamp exchanger is designed to bring to the focal point a low-powered battery lamp under the action of gravity, a sufficient power reserve being available in the stand-by battery to maintain the light until a maintenance engineer can be transported to site.

With xenon lamps it is not usually practicable to provide a lamp exchanger, so in consequence, because of the comparatively low cost of the optical system, a complete duplicate installation is provided with automatic changeover in the event of failure of the equipment in service. Once again a battery-operated standby, using a filament lamp and a different optical system, is provided in the event of failure of all means of power generation on site.

A time switch or photo-electric light switch is normally provided to bring the main navigational light into operation at the appropriate time.

The basic facilities required from the remote control system to give adequate control and monitoring of the navigational lights are therefore:

Controls	ALL LIGHTS 'ON'
	ALL LIGHTS 'OFF'
Indications	ALL LIGHTS 'ON'
	LIGHT ALARM
	LIGHT FAIL

2.2 *Fog Signal*

The operafion of a fog signal under a system of remote control is complicated by the need to have a reasonably accurate indication of the visibility at a remote station so that the signal is only brought into operation when it is actually required.

As an alternative, it would be possible to operate the fog signal continuously but this would be costly in power and undesirable because of the likelihood of complaints from people living in the vicinity of the lighthouse. A pre-requisite, therefore, of any remote control scheme has been the development of a satisfactory fog detector; a number of systems are now available operating on the back scatter principle. Quite briefly, such a system relies for its operation on the projection of a beam of light into the atmosphere, where it is scattered by the fog particles. The amplitude of the signal scattered back to the receiver is a measure of the density of the fog, and therefore the visibility. By setting a trigger level to be equivalent to the back scatter signal received from a homogeneous fog of the required visual range it is possible to initiate the fog signal when the visibility drops below the pre-determined range, usually two miles. This inevitably means that the greatest return signal is obtained from a volume of the atmosphere which is within a hundred feet or so of the lighthouse, depending on the physical layout of the detector and the geometry of its optical system.

The systems now available are only satisfactory in uniform fogs but as there is a finite possibility of a fog bank some distance from the lighthouse a system of ranging out to about two miles would be more suitable. For this reason consideration is currently being given to a laser system of fog detection. Any failure of the fog detector is arranged to bring the fog signal into operation automatically.

Returning now to the fog signal, this will generally be of an electric type using rotating machines such as triple frequency alternators, solid-state generators or static frequency multipliers using the mains as the basic frequency. Pneumatic signals are used but due to the complication of remote control of these signals wherever possible they are replaced by electric signals at stations to be remotely controlled.

A total of two control functions and five indications are required for the operation of the fog signal and the fog detector, and are as follows:

Controls	FOG SIGNAL 'START'
	FOG SIGNAL 'STOP'
Indications	FOG SIGNAL RUNNING
	FOG SIGNAL ALARM
	FOG SIGNAL FAIL
	FOG
	FOG DETECTOR FAIL

2.3 *M.F. Radio Beacon*

A considerable number of lighthouses are provided with omnidirectional radio beacons which are operated on a time-sharing basis in groups of three or six. By taking a bearing on three or more of these beacons a ship can determine its position with reasonable accuracy, using a relatively cheap radio receiver and a conventional direction-finding loop aerial.

As a usual practice in lighthouse installations, these beacons with their timing and coding facilities are provided in duplicate to ensure a high grade of service. Local monitoring circuits check the timing and ensure that the beacon does not remain either in the 'key-up' or 'key-down' position for longer than would be warranted for the code being transmitted.

Should either of these conditions occur, the beacon would be taken out of service automatically and the stand-by beacon brought into operation. In addition, at certain sites it may be necessary to change the mode of operation of the m.f. radio beacon from the normal programme which is designated 'beacon' service to a 'calibrate' service, which requires some or all of the following changes, reduction of power, change of frequency and change of coding. The normal radio beacon control panel provides these facilities, but provision must be made for the remote selection of either beacon or calibrate service. One further complication at a radio beacon station is the need to zero the timing clock to ensure that the timing of the transmissions are within the agreed international standard of ± 1 second/day. To complete this operation satisfactorily it is necessary to hold the control circuit open so that the clock can be zeroed just prior to a time signal and then released at the appropriate moment. The need to perform this function by remote control has not proved particularly convenient and it seems likely that a local circuit will eventually be provided to reset the clocks, using one of the broadcast time signals. Experiments have been carried out on this system in Scotland with, it is understood, satisfactory results.

The control and indications required for the m.f. radio beacon are therefore as follows:

Controls	SELECT BEACON SERVICE
	SELECT CALIBRATE SERVICE
	BEACON ON
	BEACON OFF
	TIME RESET
Indications	BEACON SERVICE ON
	CALIBRATE SERVICE ON
	BEACON ALARM
	BEACON FAIL

2.4 *Radar Beacon*

During the past year a number of X-band racons have been provided at various lighthouses and lightships around the coast of England and Wales, and it is possible that in the future it will be necessary to provide a service from stations operating under remote control. The racon equipment is a transponder beacon which is designed to provide both range and bearing information on particular radar targets, the presentation being in the form of an identifying paint on the p.p.i. which appears just beyond the radar target being marked. The present racons have been designed to be compatible with all marine radar equipments operating in the band 9300–9500 MHz by sweeping slowly through the band once every 90 seconds. This feature assists in reducing the amount of interference experienced from side lobe triggering of the racon and by employing a relatively long pulse as the identifying mark the radar set differentiator can be brought into operation to remove the racon paint from the p.p.i. in cases where heavy interference with the normal radar picture is being experienced. The standard installation comprises duplicate racon equipments complete with a local control and monitor panel; once again any faulty operation of the equipment in service will be detected by the local monitor circuits and the stand-by equipment brought into operation.

The control and indications required are therefore:

Control	RACON ON
	RACON OFF
Indications	RACON ALARM
	RACON FAIL

2.5 *Stand-by Power Plant*

Whenever possible stations are equipped to operate from the commercial mains supply, but provision is made for the generation of power on site in the event of a mains failure. Where commercial mains supply is not available it is normally the practice to provide at least three generators; this allows for a working machine, a stand-by machine and a machine which could be under overhaul.

The controls and indications will differ with the installation, but for a mains station they would be as follows:

Controls	MAIN SUPPLY 'ON'
	MAIN SUPPLY 'OFF'
	STAND-BY SET 'START'
	STAND-BY SET 'STOP'
Indications	MAINS HEALTHY
	MAINS FAIL
	STAND-BY SET RUNNING
	STAND-BY SET FAIL

The foregoing has detailed the items of plant which could be required at the lighthouse and has indicated the controls and indications necessary to ensure the satisfactory operation of that plant and the provision of adequate supervisory indications at the control centre. There is, however, one other control function which is necessary, particularly after a period when the communication path has failed, i.e. a check request which enables the latest state of plant to be observed by the control station:

Control CHECK REQUEST

It will be seen that a total of 16 control functions and 18 indications would be required to provide what is considered to be the minimum by way of control and supervision at a fully equipped station. Should, however, additional facilities be required, more complex systems of remote control would be provided to suit the particular installation. Figure 1 shows a control desk used by Trinity House for the control of a number of stations. The capacity of this equipment is the control of 16 stations, providing a maximum of 16 control functions and 32 indications.

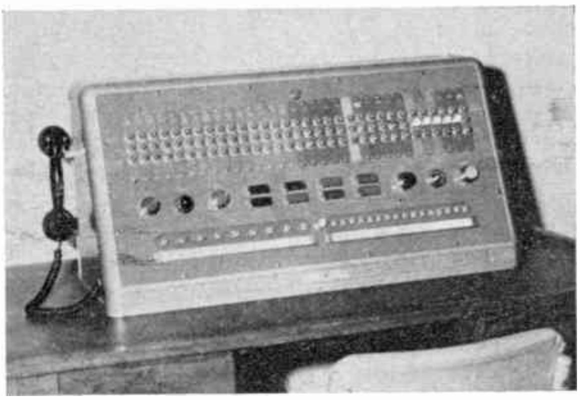


Fig. 1. Control desk used by Trinity House.

3. Economic Considerations

It can be stated that, apart from certain situations where an environmental hazard exists, the reason for providing remote control is either to improve the grade of service being given or to reduce the operating costs whilst maintaining the grade of service; in some fortunate cases it may be possible to improve the service and reduce operating costs. It is therefore important to consider the economics of operating a remote control system to ensure that the most economical mode of operation is being employed. Circumstances will differ from country to country, depending on such factors as labour costs, distances involved, form of communication path, grade of service required, availability of staff, etc.

Of these various factors, one of particular importance is the communication path, as this can be expensive and can have a considerable influence on the grade of service provided. With this point in mind, it is proposed to consider the various systems available and to compare their costs for different path lengths. The figures employed in these calculations only apply to operation in the United Kingdom and are based on the scale of charges applied by the British Post Office for rented circuits and for use of the public telephone network.

There are, of course, two basic communication systems, land line and radio link; in general for land-based situations a land-line connection will be provided but during recent times there has been an increasing tendency to use u.h.f. radio links for some applications.

Let us consider, in the first instance, the land-line connection; this can either be by private wire where the Post Office provides and maintains a pair between two specified points for the exclusive use of the subscriber, or by use of the public telephone network where the cost of connection is at the rate prevailing for ordinary telephone subscribers, the circuit being in all respects the same as a standard exchange line, in both cases the control and supervisory information being passed to line as voice frequency tones within the level tolerances laid down by the Post Office.

3.1 *Private Wire*

This system has the advantage of being provided for the exclusive use of the subscriber, and under these circumstances is available immediately at all times. Circuits of any length can be provided with a known and substantially constant transmission characteristic. Maintenance is the responsibility of the Post Office but because an exclusive circuit is set aside for the subscriber's use the probability of failure is rather higher than for a system using the public telephone network where in the event of a cable breakdown re-routing can quickly re-establish contact with the distant out-station. The use of private wires is probably the most expensive form of communication path available for remote control, but in cases where a continuous scan system is used either to monitor the communication path or to transmit information on rapidly changing conditions it may be the only scheme available as the use of radio links is still limited to special situations. Figure 2 shows how the annual cost of a private wire circuit varies with path length and can be seen to be roughly proportional to the length of the circuit.

3.2 *Public Telephone Network*

The use of the public telephone network lends itself to remote control schemes where information or control is required at infrequent intervals and where a time delay can be accepted between an event taking place at the out-station and the information being passed to the control point. Schemes of this type have been in use in the United Kingdom for some years by the Electricity Authorities for the control of ancillary generating plant, and by the Meteorological Office for the reporting of weather data from selected points. Quite briefly, contact with the out-station is made by calling the required number in the usual way, the out-station answers by means of a recording on tape or disc and, after announcing its identity, permits the control station to transmit a control tone which switches the circuit from the answering equipment to the remote control equipment; from this point on the operation is similar to that undertaken when using the private wire communication path. In the reverse direction when a change of state takes place at the out-station an initiate signal is given to the automatic calling equipment, which by using a different part of the recording either dials the control station or asks the operator for the required number. Once the call has been established the operator at the control station can send the necessary tone to connect the line at the out-station to the remote control equipment which then sends its stored information. It is usual to program the out-station to make repeated attempts to call the control station in cases where the first attempt to make contact fails.

By referring to Fig. 2, it will be seen that this system can, in certain circumstances, prove very economic in operation, particularly in regions where the subscriber trunk dialling (s.t.d.) system is being used. The actual cost will, of course, depend on the distance of the out-station from the control centre, the duration of each call and the number of calls made. For the purpose of preparing the costs shown in Fig. 2 it has been assumed that each out-station either makes or receives a total of 10 calls in each 24 hours, each call lasting 30 seconds, half the calls being at the cheap rate and a 25% increase has been added to cover for maintenance. The cost of the automatic answering and calling equipment has been amortized over a 10-year period. In addition to the economy of operation the use of the public telephone network offers a fairly high standard of reliability and is also flexible in that it enables control to be exercised from any point equipped with the necessary equipment, so that in the event of a break-

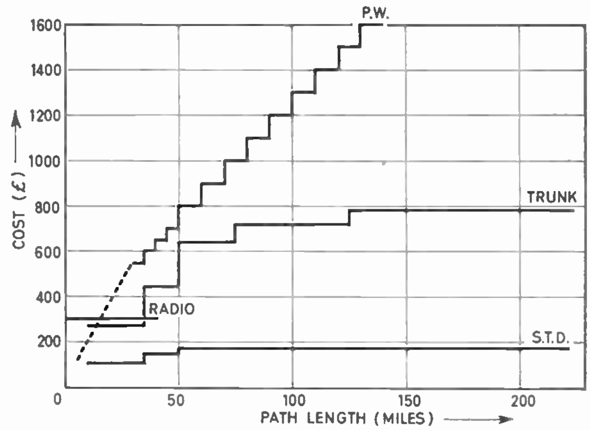


Fig. 2. Annual cost of communication path.

down at the normal control point a stand-by control centre could be brought into operation immediately. During an experiment to determine the reliability of this system a link was established between Trinity House in London and a research station on the South Coast and a total of 200 calls were made in each direction over a period of 12 months. On only two occasions was it not possible to make contact, once due to a Post Office cable breakdown and once due to a faulty motor in the automatic calling and answering equipment.

The out-station was programmed to make four attempts to call the base station by dialling '0'. The analyses of the success of these calls was as follows:

1st attempt	45% successful
2nd attempt	40% successful
3rd attempt	15% successful

The delay was due to the manual operator not answering immediately.

Calls made from the base station in London were dependent on the time of day, the results being as follows:

1st attempt	70% successful s.t.d. code
2nd attempt	20% successful s.t.d. code
3rd attempt	5% successful s.t.d. code
4th attempt	3% successful s.t.d. code
5th attempt	2% by routing via operator

3.3 *Radio Link*

The use of a radio link for remote control purposes is, of course, limited to the special situations agreed by the licensing authority, a typical instance being an off-shore island where alternative means of communication are not available. As the frequencies allocated for this type of service are usually in the v.h.f. or u.h.f. bands, the path length is limited to slightly greater than line-of-site distance. Where radio links have been used for the control of a lighthouse it has been a practice to install duplicate equipment to reduce the probability of losing contact with the out-station. This arrangement inevitably increases the capital cost of providing the communication path and as can be seen from Fig. 2 the cost of a radio link can be the most expensive of the systems considered for path lengths of less than about 15 miles. Again, the capital cost of providing the duplicate radio equipment and radio terminal equipment has been amortized over 10 years. No figure has been included for maintenance purposes but with transistorized equipment this should be a relatively small addition.

4. Typical Remote Control Scheme for Lighthouses

As an experiment the Corporation of Trinity House has decided to bring into a remote control scheme a group of stations on the North West coast of England, the intention being to control four out-stations from the Holyhead Depot. The stations concerned are South Stack by land-line, private wire, North Stack private wire, St. Bees private wire and Bardsey Island by radio link and land-line. Figure 3 shows the relative geographical situation of the stations involved.

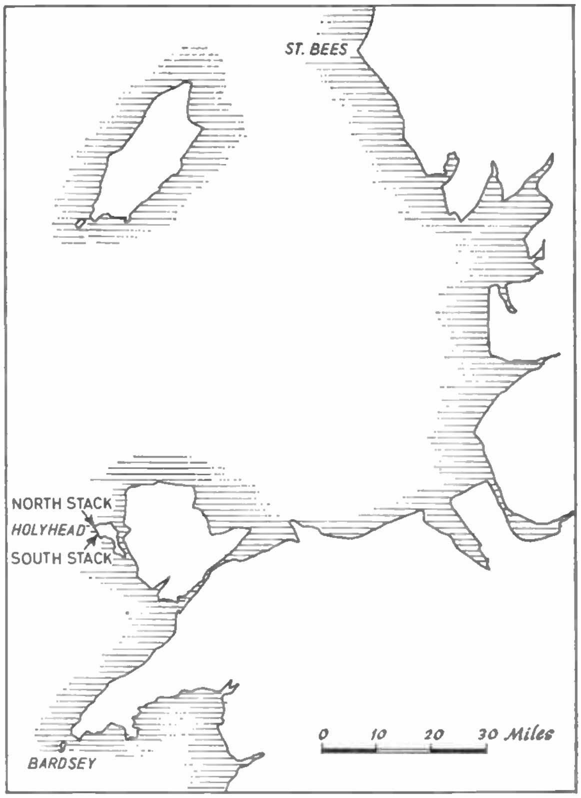


Fig. 3. Holyhead remote control scheme.

The base-station equipment which is sited at Holyhead is shown in Fig. 1. This equipment employs cold-cathode tubes for operation and display and has been designed to control up to 16 out-stations with an operational capability of 16 control functions and 32 indications. The display, apart from indicating the station which is in contact, comprises three horizontal lines of cold-cathode tubes designated 'selected store', 'register' and 'discrepancy'. The selected store indicates the previous state of plant which has been stored at the base-station, the register indicates the latest plant state, whilst the discrepancy indicates where the change of state has taken place.

A block diagram of the system as operated over land-line is shown in Fig. 4 and the signal transmitted

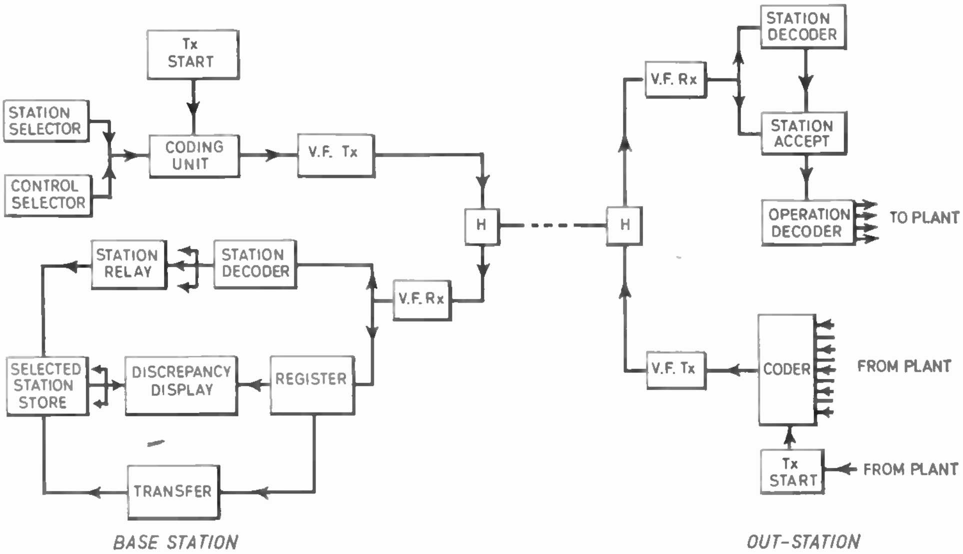
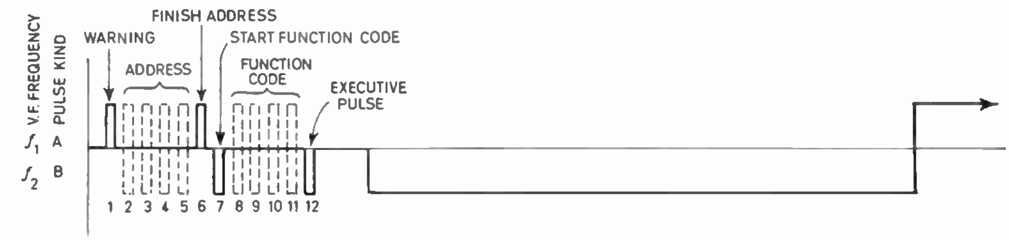


Fig. 4. Block schematic of remote control system.



(a) Control-station pulse train.

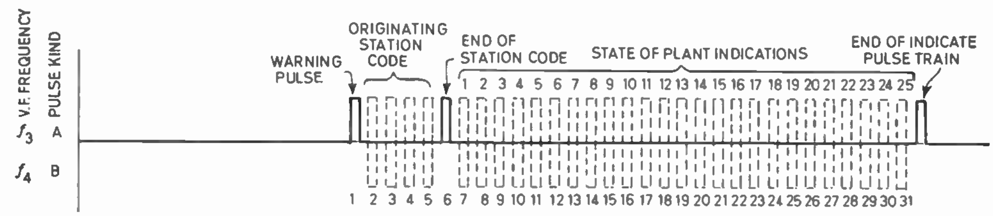


Fig. 5. (b) Out-station pulse train.

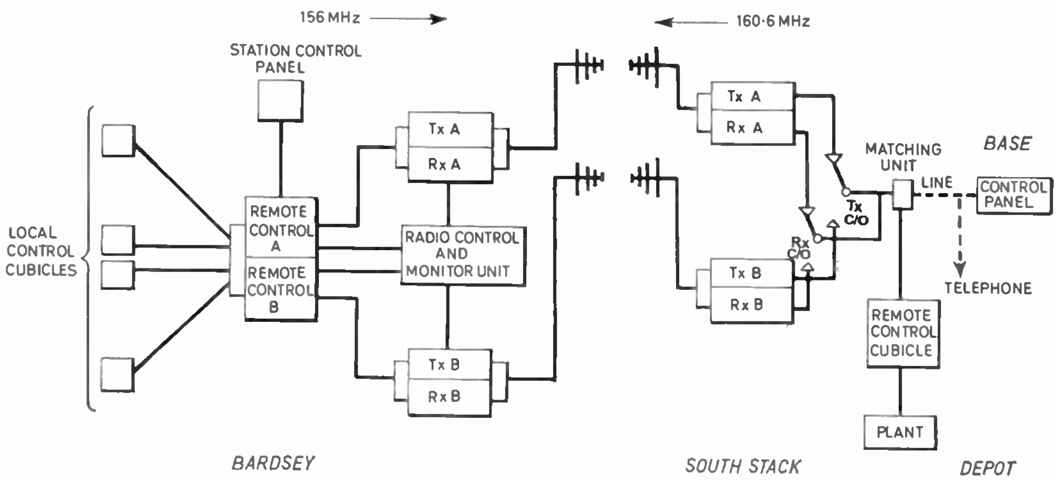


Fig. 6. Basic outline for radio remote control.

to line is indicated in Fig. 5(a), the A and B pulses being different voice-frequency tones.

The out-station equipment is relay operated, a typical signal transmitted to line being shown in Fig. 5(b).

The operation sequence is as follows:

(a) *Passing of Control Function from Base-station*

The operator would first select the out-station by operating the appropriate key in the 'station selector' and then the desired control function in the 'control selector'. By operating the transmitter 'start' push-button the coding unit would transmit the coded address and control to the v.f. transmitter which sends the message to line in the form of v.f. tones, as shown in Fig. 5(a). This signal would be received by all out-stations but only the station with the address appearing in the signal will respond and route the control signal via the 'station accept' to the 'operation decoder'. There the signal is decoded and passed to the plant in the appropriate form to carry out the required operation.

(b) *Passing of Change of State from Out-station*

A change of state taking place at the out-station will program the out-station coder and operate the transmitter 'start', thereby feeding the coded signal to the v.f. transmitter. The signal transmitted to line will comprise the address of the out-station originating the call together with the state of each item of plant at the out-station (Fig. 5(b)). On receipt of this signal at the base-station, the output of the v.f. receiver will be routed to the station decoder, which will establish the identity of the out-station originating the call and operate the corresponding station relay. The signal indicating the state of the plant at the out-station will then be routed to the register, the previous state of plant being called from store and indicated on the selected station store; where a difference exists between these two states an indication will be given in the discrepancy display. Operation of the transfer push-button will transfer the latest state of plant into the station store and clear the equipment down to await the next operation.

In the case of Bardsey Island, whilst the remote control equipment is basically the same as at the other stations in the group, the need to operate over radio link created problems when remote monitoring of the radio equipment itself was considered. For this reason, and because of the inaccessible location of the site, it was decided to provide duplicate radio equipment and also complete duplication of the remote control equipment; the final arrangement is shown in Fig. 6. By this arrangement it can be seen that faults on the A radio equipment can be reported over the B remote control channel and vice versa, the two remote control equipments being connected to operate in parallel, each having a different address, so that either equipment can be used for the remote control of the Bardsey station. A fault in either channel will be reported by the remaining channel so that the probability of a complete failure to pass and receive information from Bardsey is very low indeed.

5. Conclusions

The unattended operation of fully equipped lighthouses is technically possible provided the plant installed has been suitably designed. Where the greatest reliability is required, the lighthouse should be automatic in operation, the remote control equipment providing what is, in fact, a remote monitoring system with facilities for remote control.

For the most economic operation of the system very careful consideration must be given to the type of communication path provided as this choice could well influence the type of remote control system provided; for example, continuous scanning or an 'on demand' system. Where a delay of a few minutes can be accepted in performing a control function or receiving an indication of a change of state at the out-station the use of the public telephone network could be both economic and reliable, but some knowledge of the likely traffic would be required before a final decision is taken.

6. Acknowledgment

The author would like to thank the Elder Brethren of Trinity House for permission to present this paper.

Manuscript received by the Institution on 15th February 1967. (Paper No. 1141/RNA74.)

Discussion on
The Remote Control of Lighthouses and Beacons

Under the Chairmanship of Mr. A. Harrison

Mr. C. E. Goodison: At what rate does Mr. MacKellar expect lighthouses to be made automatic? Since the Meteorological Office use certain lighthouse reports, when the lighthouses lose their keepers it will probably be necessary to instal some automatic weather reporting system.

Mr. MacKellar (*in reply*): It is extremely difficult to predict the rate at which lighthouses will be made unattended. The present programme is by way of being a large scale experiment to establish the advantages and disadvantages both operationally and economically, the future conversion to unattended operation will depend on the outcome of this experiment.

Lt.-Com. R. B. Richardson: Could Mr. MacKellar tell us what sort of power supply is needed to operate his fog detectors? Also, is he able to grade the settings of his detector so as to provide indications of, say, four different set visibility distances? We, in the Port of London, have a definite need for such instruments at remote out-stations, relaying in to the operations centre, but have so far got no further than exploratory stages.

Mr. A. C. MacKellar (*in reply*): The power supply required for the Trinity House transistorized fog detector is at present 50 volts d.c. but this could quite easily be changed to 12 volts d.c. as this is the operating voltage for the circuitry. The 50 V supply was selected as this was a voltage available at most Trinity House sites, the total power demand is 8 watts. The present design of fog detector cannot be arranged to indicate when the visibility reaches one of a number of different ranges as this is not a design requirement in the Trinity House service. Modifications could, however, be incorporated to provide this facility.

Captain W. R. Colbeck: Mr. MacKellar referred specifically to a 2-mile range for the automatic fog detector. Is this the maximum or maximum-efficient range for this instrument? In certain areas fog signals are started, usually, on visibility falling to 3 miles.

Mr. MacKellar (*in reply*): The range of 2 miles specified in the paper is the visibility generally taken at Trinity House sites as requiring the sounding of the fog signal. The instrument used to detect this visibility can be adjusted to start the fog signal at any visibility up to 5 miles.

Mr. C. P. Hamilton: Are fog detectors, so far developed, able to rotate and scan 360° or are they limited to scan a fixed sector angle? If so, what is the reliable angle of scan on the horizontal and vertical scale and what is the range?

Mr. MacKellar (*in reply*): It is assumed that Mr. Hamilton is referring to fog detectors operating on the back-scatter principle, in which case, because of the preference given to returns from the near atmosphere, i.e. the first 150 ft or so, it is not considered necessary to scan an arc extending to 360°. The important factor is to sample as large a volume of atmosphere as possible to minimize the effect of any local variations and in addition to ensure that the sampled volume is sufficiently distant from any source of heat which might affect the visibility locally, i.e. buildings. It is for these reasons that the latest Trinity House fog detector samples an arc of 135° simultaneously, the sampling volume commencing at a range of 50 ft.

In contrast a number of commercial fog detectors operate in a fixed direction over an arc of 2° in both the vertical and horizontal planes, so it seems that reasonably satisfactory results can be achieved for very narrow sampling angles.

With regard to range, again all back-scatter fog detectors assume a homogeneous fog and can be set to indicate when the visibility drops to any specified range between the limits of 5 miles and ¼ mile. It must, of course, be understood that no significant back-scatter signal is being received from ranges much in excess of ½ mile.

Lt.-Com. F. M. Berncastle: I would like to know whether Trinity House has done any work in connection with the control of fog signals on buoys or floats. These would have to be battery operated but the degree of reliance could be relaxed as it was for any unmanned floating mark. My particular problem is that I want to operate an electric fog signal on one, and if successful two, unmanned lightships from a third stationed between them at a distance of about 4 miles. The system would need to be 'fail safe' and, of course, the aerials would need to be omnidirectional. Monitoring would not be essential as the vessels would be more or less within audible distance.

Mr. MacKellar (*in reply*): Trinity House has recently carried out experiments on the radio control of a fog signal mounted on a buoy using a temporary experimental frequency in the 29 MHz band. This experiment proved the possibility of operating a fog signal in this way but at the present time the only permanent frequency allocation which can be obtained for such a service is in the u.h.f. band. In addition at the present time the transmission of a continuous carrier is not approved by the licensing authority so that other methods of 'fail safe' must be involved. Current thinking tends towards the use of a delay at the output of the receiver on board the buoy. The expiry of the delay will bring the fog signal into operation, the delay being reset before expiry by transmissions from the control transmitter at intervals in excess of 5 minutes. By making this delay, say 15 minutes, the fog signal will automatically come into operation not longer than 15 minutes after the failure of the communication system. The system proposed by Lt.-Com. Berncastle is therefore possible but at the moment it is difficult to obtain transistorized u.h.f. equipment with which to implement the service.

Mr. P. F. Cook: Reliability should be relative to the overall integrity of fundamental function. What order of reliability is achieved? Secondly, can Mr. MacKellar give more information on 'fail safe' aspect of fog detection?

Mr. MacKellar (*in reply*): The figure of reliability of the light at a major light, i.e. a lighthouse or light-vessel, has been established by taking an analysis over a period of several years, a failure being taken as the light extinguishing for a period in excess of 5 minutes. Under these circumstances the m.t.b.f. has been found to be 55 000 hours.

The 'fail safe' facilities provided in the fog detector are designed to bring the fog signal into operation in the event of a fault developing in the fog detector. In practice this is achieved by selecting the fog condition to provide no output from the equipment, or in other words the fog signal is held off by the clear indications obtained from the fog detector. Compensation is also provided for the gradual reduction in light output from the lamp used to transmit a light pulse into the atmosphere and for condensation, moisture and dirt on the optical surfaces.

Mr. C. E. Goodison: Since many questions have been about the fog detector, I would like to emphasize that the measurement of visibility or fog was only taken in this particular system within a fairly short radius around the instrument, say 150 feet. To extrapolate the measurement of this small sample to longer distances within what is normally a quite inhomogeneous mass could be quite misleading. For instance, if the fog detector measures, say, a 1 mile visibility this means that the air would have to have the same qualities over 1 mile point to point as that air within sampling range of the detector.

The Meteorological Office have a visibility measuring system which uses a projector lamp and a photo-electric cell receiver spaced over a fairly long baseline. Of course, this would not be suitable for lighthouses.

Mr. MacKellar (*in reply*): I would like to comment on the statement regarding, the operation of fog detectors made by Mr. Goodison, and although from his point of view as a meteorologist the results obtained from back-scatter fog detectors could be misleading, in practice at lighthouses the back-scatter system has proved most effective. An analysis carried out over a six-month period at five lighthouses equipped with back-scatter fog detectors showed a very good correlation between the periods when the keeper actually sounded the fog signal, based on his own estimation of the visibility, and the periods the fog detector would have sounded the fog signal had it been so connected. I believe that a back-scatter system of fog detection can provide an adequate means of determining when a fog signal should be sounded.

Mr. G. R. Barnes: In considering stand-by power supplies, can Mr. MacKellar provide details of the reliability of conventional motor-generator sets in this role, and is it likely that static power supplies, such as fuel cells, may be used in the future?

Mr. MacKellar (*in reply*): In practice it has been found that the diesel alternator has been most reliable and where failures have occurred they have been concerned with the fuel or ancillary equipment such as starter batteries, etc. Probably the most serious problem has been contamination of the fuel by water, but with means available to obviate this problem it can be said that the reliability of stand-by diesel alternators has been satisfactory. There is, of course, a fairly high maintenance commitment with diesel engines and in consequence a constant watch is being maintained for other power supply systems which might offer advantages. The fuel cell is one such system, but from discussions with different research centres and manufacturers it is quite clear that commercially available systems rated in tens of kilowatts are still some years off.

Lt.-Com. Richardson: Concerning the m.f. beacons so widely fitted round the coast, could Mr. MacKellar say whether we could not use the 'silent' periods on these to transmit an indication of the actual visibility at the specified beacon? It would seem to me that the mariner using certain of these beacons on his passage might well find positive value in receipt of intelligence of this sort.

Captain Colbeck: The 5-second silent period at the end of the beacon transmission was adopted to ensure that there was no overlapping of the transmissions of the next beacon in the group. The time allocated to each beacon is one minute and there are commonly six stations in a group operating on the same frequency transmitting in sequence. In some cases, as in the English Channel for instance, two stations may be run by Trinity House and four by French authorities. Clocks can easily gain or lose one or two seconds in a day and if transmissions, whether of the beacon signal or other information, did overlap, the user might have difficulty in appreciating the change to the next station in the sequence.

Mr. MacKellar (*in reply*): The specification for beacon characteristic was agreed on an international basis during the I.T.U. Paris Convention of 1951 and it would require international agreement to change the present characteristic to include information on visibility. There would, however, be no technical difficulty in including such information in the transmission but there may be considerable difficulty in acquiring the information on visibility, particularly at sites which are unattended.

Mr. N. D. Clotworthy: In the automated lighthouse of the future there will be considerable investment in remote control equipment of very high reliability. Are we spoiling the mariner by trying to obtain too high reliability? Could a reduced standard be accepted, at lower cost?

Mr. MacKellar (*in reply*): It is the Trinity House policy to provide a service 'to the mariner' with the highest degree of reliability which may be achieved by modern techniques and practices. Any relaxation from this position would not only be unacceptable to Trinity House but also not in the best interest of the mariner. I think that to accept a policy of providing a lower grade of service, by reducing the reliability of plant installed, will create doubt in the mind of the mariner and a lowering of the morale of the personnel operating and maintaining the service. Such an outlook inevitably leads to less reliance being placed on the service provided and to a gradual run down to a point where a useful function is no longer being served.

Captain Colbeck: The highest degree of reliability is still required for navigational lighting. Electronic aids may fail without warning and in spite of regular maintenance, and there are very many small vessels not fitted with such devices who rely entirely on visual observation of shore lights to determine their position at night.

Mr. P. F. Cook: The figures quoted by Mr. MacKellar are about equivalent to the probability of a civil aircraft crashing on landing. What is the probability of a ship catastrophe when one light fails?

Mr. MacKellar (*in reply*): Unfortunately I am unable to provide information on the probability of a ship catastrophe in the event of failure of a single navigational light. In general the major hazard to shipping is the collision risk, which is not related to the performance of the navigational lights. I can, however, state that I know of no case where the failure of a navigational light was the direct cause of a ship foundering.

U.D.C.621.37/9.004.6

The Use of Standard Sequences in Environmental Testing

Presented at the Joint Letchworth College of Technology—I.E.R.E. Conference on 'Reliability and Environmental Testing of Electronic Components and Equipment' held at The College of Technology, Letchworth, on 3rd May 1967.

By

R. T. LOVELOCK,
C.Eng., F.I.E.E.†

Summary: Laboratory tests to assess component behaviour in adverse environments are artificial, and designed to be consistent on repetition, but accelerated in attack. They each test for a particular weakness of design or of fabrication. In order to obtain a result quickly it is usual to apply a number of independent test procedures in parallel to as many separate groups of similar components. There are, however, two types of failure where it is advantageous to employ two or more individual procedures in sequence to a single group of components.

(i) Where a sequence of adverse conditions, often encountered in nature, can lead to a failure which would not occur where any one of the conditions had been applied above.

(ii) Where damage due to one procedure is not apparent immediately, but can be detected readily by the subsequent application of a second procedure.

Practical examples of both types of failure are examined in detail. Both climatic environments and mechanical environments are considered, and also the combination of the two. The more specialized cases where the simultaneous application of two or more different procedures is essential, are considered. Such combined tests usually contain also an element of sequence, and these much more complicated cases are considered briefly.

1. The Object of Environmental Testing

When a long established product continues to be sold in a market where it has been in use for several years, the problem of reliability does not often arise. Any faults in the design or construction will have been established empirically, brought to the notice of the manufacturer via customer complaints, and rectified by a suitable change in design. Some faults are completely random, and require a large number of items to have been operative if a true picture of reliability is to be obtained such that effective counter-measures may be adopted. Others take a prolonged period to develop, and require that a number of items shall have been in regular use for several years before they become sufficiently evident for remedial action to be decided and taken.

This empirical method of detecting weakness in design is an undesirable one, because the major portion of the experiment is carried out at the expense of the customer, and where serious faults are brought to light, it inevitably engenders a considerable amount of bad feeling in the customer. When therefore a new material becomes available, or a new method of construction is adopted, or an entirely new product is devised, it is very desirable that a method should be available which will enable the major sources of trouble to be located and rectified before tooling is finalized and the product placed on the market. It is obviously impossible for the manufacturer to carry out an empirical test on his own premises, because not only would the expense of destroying a large number of items be prohibitive, but the waiting period of several years before long-term faults develop would allow a competitor to steal the market, and the design to become out-dated.

Even when a manufacturer has been in an analogous field for a considerable time, and his staff have accumulated sufficient experience to foresee any major troubles and design them out in the initial stages, the opening up of a new export market in a climate different from that of the home country can bring a fresh crop of trouble to light. This type of trouble is particularly unfortunate because the customer is so far away, often speaks another language, and is apt to view his problem as an offshoot of international relationships rather than a normal instance of design teething troubles.

† Belling & Lee Ltd., Enfield, Middlesex.

It is to obviate these undesirable events that a series of artificial tests has been devised which may be carried out in the manufacturer's own laboratory, and which can detect faults, which would take months to develop, within a few days. The whole procedure has become known as 'environmental testing', and has been slowly but successfully developed over many years. As long ago as 1935 manufacturers exporting radio receivers to the tropical areas were using special climatic cabinets in Great Britain to find faults which might develop later. With the advent of a war fought in tropical jungles, and dependent upon electronic equipment for communication and location of the enemy, these techniques were developed rapidly and effectively. Since the war, the need for free interchange of manufactured products between countries has indicated the need for international standardization of such testing procedures, and the International Electrotechnical Commission is currently engaged in negotiating and publishing such an agreed system in its Publication 68.

2. Subdivision of Testing

A complex electronic equipment is costly, and is manufactured in comparatively limited quantities. To take a large number of such equipments therefore and subject them to destructive tests in order to locate the random type of fault would impose a heavy cost loading on those equipments which were sold. These equipments are an assemblage of much smaller and cheaper components however, such as resistors, capacitors, transistors, etc., and the majority of the random faults in the equipment will be a failure of one or more of these components. It is usual therefore to test separately the individual components in quantity, and to test only a few of the complete equipments.

Such a practice saves overall cost in two ways. Firstly, since the same component will be used on many types of equipment, the one test will serve to examine possible faults on many different equipment designs, and the cost, distributed over the very many components sold, will be shared by all of the equipment types instead of being allocated wholly to only one. Secondly, the cost of the component is much less than that of an equipment, and the test equipment necessary to determine its behaviour is smaller and simpler than that required for a large equipment.

For these reasons it has become standard to test separately the components and the equipments. It follows that the object of the testing procedure will be different in the two cases. The component tests will not only probe all modes of failure while operative in any of a number of very different equipments, but will also investigate the special stresses involved during assembly, and never subsequently experienced by the component during operation of the equipment.

3. Testing Conditions

It is possible to obtain a considerable measure of acceleration in the laboratory merely by reproducing the extreme conditions of service. To take an over-simplified example: a structure may contain a stressed elastomer which will suffer brittle failure at $-50°C$. If, during operation in the field, it may experience such a temperature only once per year, and may require exposure for at least two hours before failure occurs, it is possible for it to operate for two or three years before a prolonged exposure to a very low temperature initiates a failure. The item may be placed in a refrigerator at $-55°C$ for several hours, however, and such failure, if it is going to occur at all, will be found within a single day.

A second simplified, but equally realistic case, concerns electronic components exposed to high humidity at high temperatures. Where protection against moisture ingress is almost good enough for the operating conditions, because the occurrence of high humidity is periodic and only the occasional prolonged exposure causes failure, it may require more than a full year in the tropics before such failure occurs. Continuous exposure to these conditions for only a few weeks in the laboratory may well suffice to detect the weakness however.

For these reasons it is usual to expose both components and equipments to the various extreme conditions which they will encounter in practice, and thus find the types of failure likely to be experienced while they are still under skilled observation in the laboratory. In the interests of standardization, I.E.C. Publication 68 specifies the various conditions which should be used for such testing. This standardization is highly desirable for two reasons. Firstly, it minimizes the quantity of expensive testing equipment needed by ensuring that similar conditions on various items will all be investigated in the same manner. Secondly, it ensures that the test results from one manufacturer's test house will be directly comparable with those for his competitors' products, thus allowing the equipment designer to choose the better component for his purpose, and the equipment user to select that most suited to his requirement.

The conditions standardized for these testing procedures embrace all the limiting environments encountered during service. Climatic testing includes high and low temperatures, high humidity at various temperatures, low air density, and leakage rate past a gasket across which there is a pressure differential. Mechanical conditions include shock, steady acceleration, sinusoidal vibration, random vibration (white noise), acoustic (noise) environments, mechanical stress, and strain. Operational tests include continuous mechanical operation, and prolonged application of electrical and thermal loading. Manufacturing conditions include soldering, bending of leads, torsion of leads, etc.

The climatic characteristics of all portions of the Earth's surface have been analysed, and a series of standard severity levels in discrete steps has been standardized which will cover the practical working requirements for electronic parts and equipment. Because the markets for items other than electronic equipment are the same, it follows that although it is the needs of the electronic manufacturers which have initiated this standardization, the same conditions are the optimum for other classes of equipment also, and many types of product are beginning to adopt the same standardized testing procedures for assessing performance to be expected in the field.

4. Varying Conditions

Where a single simple mechanism of failure is involved, it is obvious that the application of a single prolonged stress will suffice as a test. Where, for example, a chemical change which is rate-dependent on temperature is the sole mechanism of failure, continuous exposure to a constant elevated temperature is sufficient, and where transmission of moisture through a barrier is dependent only on temperature and difference of vapour pressure across it, continuous exposure to a constant temperature and relative humidity is adequate to assess failure due to this cause. In practical use however an equipment is rarely exposed to a constant environment, and a part which could successfully withstand any one of the several limit stresses applied alone, may well fail by a different mechanism when exposed to the mixed and continuously varying environmental conditions of nature.

It is necessary therefore when testing such parts to probe for all probable mechanisms of failure, including those which are due to sequential, varying, or combined types of stress. Once one has devised a complex testing sequence involving several levels and types of stress, it is a controversial point of terminology whether a particular procedure is called a 'cyclic test', a 'sequence of tests' or a 'combined test'. International discussion within I.E.C. is still proceeding on the question of which of the above terms shall be used for a particular procedure. In this paper all three will be considered, and in dubbing them all as 'standard sequences' there is no intention of prejudging the terminology question: the name is only a convenient label under which the important principles of environmental testing which are common to all three may be considered.

If one attempts to sub-divide this general testing principle into over-simplified water-tight compartments, the following three general types emerge:

(*a*) Cyclic tests, in which one condition such as temperature, or two conditions related in nature such as temperature and humidity, periodically oscillate between two extremes. This test may, in its simplest form, comprise exposure to only one cycle of conditions, but usually a number of cycles applied sequentially are involved.

(*b*) Sequential tests, in which a number of unrelated stresses, such as mechanical stress, local heating, intense cold, humidity, etc. may follow one another in a meaningful sequence. Some of these stresses may occur once only in the sequence, and others may occur several times, either following one another, or else interspersed with others.

(*c*) Combined tests in which two or more individual stresses are applied simultaneously to the item under test. While there is a period during which several conditions are operative together, it may be necessary to apply them initially in a sequence, as, for example, when low temperature and low air-density are to be combined, it is preferable to take the temperature down to the low value before the air density is reduced.

It is not very advantageous to attempt to categorize these test procedures into neat classes as above. Not only are some of them a combination of two, or even of all three categories, but there is no logical coherence between the members of any one category which assists in understanding the testing principles involved. What is of major importance to the designer is the mechanism of failure, and in the following sections of this paper consideration will be based upon the similar modes of failure rather than upon the arbitrary methods of combining simple test procedures.

5. Temperature and Humidity

In the tropical jungle where combined extremes of temperature and humidity obtain, the conditions are varying semi-cyclically, usually with a period of 24 hours over the major cycle. Superimposed on this in most districts is a yearly cycle, and an irregular variation common to weather in nearly all parts of the Earth's surface. When the mechanism of failure is a molecular diffusion of moisture into or through an organic barrier, continuous exposure to a constant value of temperature and humidity is the ideal test condition. Where however variation of temperature causes breathing past a gasket due to differential air pressure across it, and where the thermal inertia of a specimen can cause condensation of water upon its surface, a cyclically-varying combination of temperature and humidity is essential to probe all the possible mechanisms of failure. In environmental testing it is common practice to employ both constant and cyclic conditions.

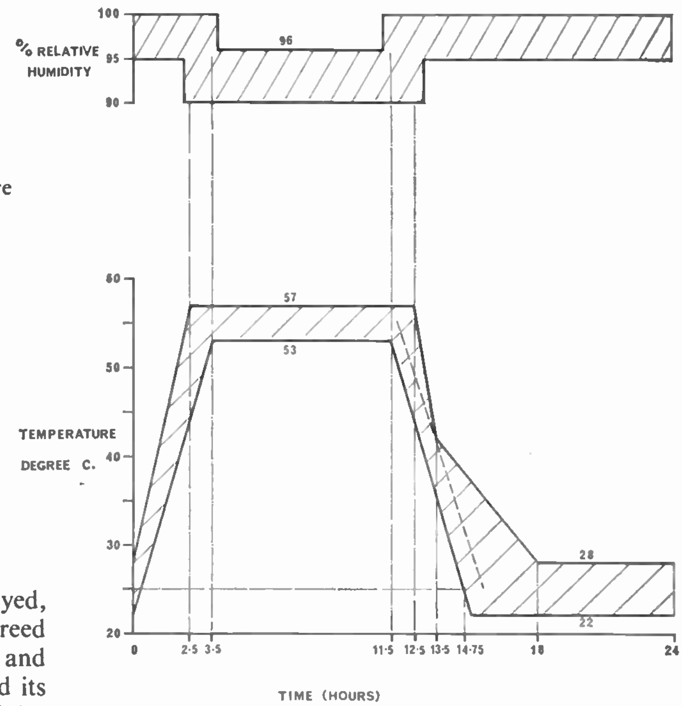


Fig. 1.
Typical test cycle of temperature and humidity.

In order to rationalize the test equipment employed, a small number of standard cycles has been agreed and used for all tests on electronic components and equipments. One of these is shown in Fig. 1, and its characteristics will serve to illustrate the principles involved in them all.

The aim behind this type of test is not to produce the random and irregular cycle experienced in nature, but a closely-controlled regularly repeating condition. This is extremely desirable because a main use of the test is to compare the performance of a new design with that of an old one tested a year or more ago. Unless therefore the effects of the test on the specimen are identical for every application, no close comparison of performance between two designs not exposed simultaneously to the cycle is possible. It is this requirement to compare degrees of failure rather than to detect a tendency to failure which necessitates such expensive test equipment and such closely controlled and monitored conditions.

It is common knowledge that the effects of a liquid film on the surface may produce very different results from exposure of the surface without a film to a high vapour pressure of humidity. The effects of the condensation film is often less severe than that of high vapour pressure without it. For this reason the limits of relative humidity are closely controlled during the cycle to ensure that vapour pressure shall always be high, but condensation will only occur on the surface of the item being tested during the period when exposure temperature is rising and thermal inertia causes the temperature of the specimen to lag below the dew point. The upper and lower temperatures are fairly closely controlled, not only as a control on vapour pressure, but also as a control of breathing rate resultant upon the variation of air pressure across seals and barriers in the test specimen.

6. Thermal Shock

When the constructional materials of an item have coefficients of thermal expansion which differ from one another, serious stresses can be generated by change of ambient temperature. Even where a wide difference of thermal coefficients does not exist, if a large mass of material has a low thermal conductivity, local heating of one portion only can set up stresses within the homogeneous mass. For this reason it is desirable to expose items to the same general type of condition which they will experience in practice.

Three basic testing procedures have been evolved which together cover most of the practical hazards of service. The first involves placing the specimen in an enclosure and varying the air-temperature cyclically at a controlled rate, and halting at clearly controlled maximum and minimum values for a specified time. This covers such conditions as operation within an aircraft which may climb rapidly from the surface of the Earth to a high altitude, and descend equally quickly later.

The second involves transferring a specimen within a few minutes from a chamber with air temperature at one extreme to a chamber with air temperature at the opposite extreme. Besides being an artificial method of tracing possible weaknesses of design, this is a realistic simulation of the practical condition where a component may be removed from an equipment which has been running hot into the outer air which may be below freezing point. Application of both the first and second of these procedures usually involves subjection to several sequential cycles.

The third method is entirely artificial, and is intended to locate undue stresses on the surface of such items as glass-to-metal seals. It involves transferring the specimen rapidly back and forth between immersion in a hot liquid and immersion in a cold one.

7. Mechanical Stress and Strain

The various mechanical stresses applied to an item frequently cause some measure of strain, and may, as a result, cause a permanent change leading to failure during life. Most defects of this nature can be detected either by visual examination following application, or by some simply applied measurement. There is one type of damage however which is not so readily detected, and which is most easily found by application of a standard sequence.

Many electronic components are protected from the ingress of moisture by a barrier surrounding them. Through this barrier must be brought out electrically-insulated leads, and where such metallic leads penetrate the barrier a seal against ingress of moisture must be provided. This seal may well be damaged due to mechanical stresses, and allow moisture to enter. The entry of moisture will be a slow process where vapour pressure is moderate and temperature cycling low, and being only small in quantity may take a long period before causing failure. In order to detect such failure it is usual to perform the mechanical tests on the same samples which will subsequently be subjected to a cyclic humidity test, and to rely on the humidity cycling to initiate failure if the mechanical tests have damaged the seals.

Where the structural members of an item are likely to suffer fatigue due to the cyclic stresses imposed by vibration, and the particular frequencies most likely to cause damage are unknown, there are two basic methods used to explore the portion of the spectrum which is of interest. One is to subject the specimen to a sinusoidal acceleration of controlled value, varying the frequency at a logarithmic rate from minimum to maximum value, and back to minimum again. This double sweep through the spectrum is a single cycle of application, and a large number of such cycles is applied. The second method is to subject the specimen for an agreed time to a vibration containing simultaneously components through the spectrum distributed in a random manner (white noise) and with an agreed value of spectral density.

8. Climatic Failures

When a moisture barrier surrounds an item, it is possible, if its contains flaws, cracks or interstices, for water to be trapped therein by surface tension, and if later the temperature falls below freezing point for the expansion to launch cracks through the structure. A standard climatic sequence has been agreed to test for such complex modes of failure.

After subjecting the specimens to the mechanical and thermal stresses likely in service to initiate such minor damage, they are first subjected to a period of exposure at maximum working temperature. They are then subjected to a single cycle of temperature and humidity variation to fill all small crevices and flaws with water. They are next transferred to a refrigerator where they are held at minimum working temperature for a period to turn the water into ice and open up any cracks in weak spots due to expansion of water when freezing. Finally the specimens are exposed to several cycles of temperature and humidity to allow any resulting failure of sealing to be detected by damage caused to performance through entry of moisture. A typical testing sequence of this type is shown in Table 1.

Table 1

Typical sequence of tests for electronic components

Lead deformation
Soldering
Bumping hot and cold
Vibration hot and cold
Shock
Acceleration
Temperature rapid change
High temperature
Cyclic humidity, 1 cycle
Low temperature
Cyclic humidity, 5 cycles
Final assessment of performance

9. Combined Stresses

It is possible for the simultaneous application of two or more stresses to cause failure where either, alone or in sequence, would leave the specimen undamaged. A typical example often met in practice is where an elastomer is used because its flexibility is essential to allow mechanical movement. Even if the application of a low temperature reduces the flexibility of the material to a point where movement would cause brittle fracture, unless mechanical stress is applied while the temperature is low, no failure will occur. To probe for this mode of failure it is usual to subject the specimens to repetitive cycles of mechanical stress while they are cold. To test for an analogous failure due to weakening of structural materials at high temperature, it is usual to perform the mechanical tests also while at maximum working temperature.

Another typical example of combined tests causing failure where sequential applications would not do so is found in the operation of airborne equipment which is not contained within the pressurized cabin. As an

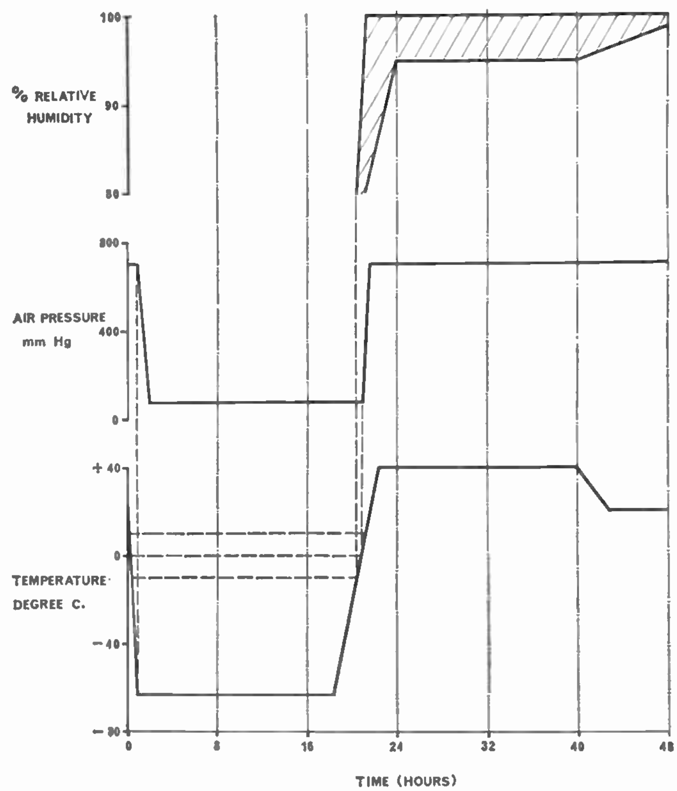


Fig. 2. Typical combined altitude cycle.

aircraft descends from high altitude and passes through heavy cloud an equipment is swept from low pressure to high pressure causing breathing past the gaskets, and by passing through cloud is simultaneously immersed in an atmosphere of high relative humidity. This condition is simulated in the laboratory by placing the equipment in an enclosure which can be evacuated, which can be frozen, which can be heated, and in which high values of humidity may be generated. The standardized test is a sequence of combined conditions, and a typical example is shown in Fig. 2.

10. The Economics of Testing

There is a further reason in some circumstances for adopting a particular sequence of tests. Some of the tests may damage the item to such an extent that it is no longer saleable, and hence will represent a considerable expense if the equipment is valuable. It is desirable therefore to reduce to a minimum the number of identical items tested, and where the specimen is a very complicated equipment it is usual to apply tests to one item only.

While it is to be expected that the tests will result only in degradation of performance, it is possible, if a design fault is present, that a particular test may cause catastrophic failure which renders the equipment unsuitable for any further testing. Since more than one fault may be present in a single equipment, and it is desirable to locate them all at the expense of the one item, it is wise to perform those tests least likely to result in catastrophic failure before those which are most likely to do so. In this way the maximum amount of information may be obtained with a minimum expenditure.

Where, in addition, the testing procedure is long and complicated, and in consequence very expensive, it is desirable that any fault which will necessitate such modification of design as to require complete re-testing, should be found as early in the procedure as possible. For this reason it is customary to perform such quick and inexpensive tests as may locate serious faults as early in the sequence as possible, thus obviating the necessity for completing the more expensive tests which would have to be repeated in any case.

The differences between equipments are so great that almost every case has its own special features in which it differs from others. For this reason it is not possible to standardize on a general testing sequence for all equipments, but each should be considered on its own merits. On the other hand, with small components the likeness is sufficiently great, and the price sufficiently low, for a few standard sequences to be adopted. Extensive experience in the United Kingdom has already resulted in a fair degree of standardization along these lines, and the whole matter is at present under active discussion within the International Electrotechnical Commission. Expensive as such complete testing procedures may be, the adoption of an international standard can materially reduce the overall cost of approval testing by obviating the need for repeating tests to a slightly different requirement for other markets to which it may be desired to export.

11. Acknowledgments

Thanks are due to Belling & Lee Ltd., and to the staff in their test house, for much of the experience which lies behind this paper.

12. References

1. British Standards on Environmental Testing Procedures—BS, 2011, BS.2G. 100.
2. I.E.C. Standard Recommendations on Environmental Testing Procedures—I.E.C. 68.
3. British Services Standards on Environmental Testing Procedures—DEF. 5011, DEF. 133 and DTD. 1085.
4. U.S.A. Military Standard on Environmental Testing Procedures—MIL.STD. 202.
5. British Services Standards on Application of Testing Procedures—DEF. 5001, DEF. 131.
6. British Common Standards on Application of Testing Procedures—BS. 9000 series.

Manuscript received by the Institution on 25th May 1967. (Paper No. 1142)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **August 1967**

August 1967	24-hour mean centred on 0300 U.T.			August 1967	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	− 299·8	+ 0·1	− 0·2	**17**	− 300·0	0	0
2	− 299·8	+ 0·2	− 0·2	**18**	− 300·0	0	0
3	− 299·8	+ 0·2	− 2·1	**19**	− 299·7	0	0
4	− 299·9	+ 0·2	− 0·1	**20**	− 300·1	+ 0·1	+ 0·1
5	− 300·0	0	− 0·1	**21**	− 300·0	+ 0·1	+ 0·1
6	− 300·1	0	0	**22**	− 300·0	0	+ 0·1
7	− 300·0	− 0·1	0	**23**	− 299·9	0	+ 0·1
8	− 300·1	0	0	**24**	− 299·9	+ 0·1	+ 0·1
9	− 300·1	− 0·2	0	**25**	− 300·0	0	0
10	− 300·1	− 0·1	0	**26**	− 300·0	0	+ 0·1
11	− 300·1	0	0	**27**	− 300·0	0	+ 0·1
12	− 299·9	0	0	**28**	− 300·0	+ 0·1	+ 0·1
13	− 300·0	0	0	**29**	− 299·8	0	+ 0·1
14	− 300·0	0	− 0·1	**30**	− 299·9	+ 0·1	0
15	− 300·0	0	0	**31**	− 300·0	0	+ 0·1
16	− 300·0	0	− 0·1				

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.

1. The measurements were made in terms of H.P. caesium standard No. 134 which agrees with the N.P.L. caesium standard to 1 part in 10^{11}.
2. A phase change of 6·8 microseconds occurred on the MSF 60kHz transmission at approximately 1215 G.M.T. on the 14th August and has been allowed for in obtaining the 24-hour mean frequency deviation for that day.
3. From 1300 G.M.T. on the 3rd August, for an experimental period, the carrier frequency of the Droitwich 200 kHz transmission has been controlled by a rubidium gas cell frequency standard.

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.

REFLEX-TYPE SOLID-STATE MASER

The bandwidths of reflex-type solid-state masers are increased by providing multi-cavity systems with active material in all the cavities. The approximation of uniform negative losses gives (by contour integration) an estimate of the bandwidth of such systems. The authors of two Russian papers have examined theoretically and experimentally the special case of a three-cavity maser, and it is shown that in the region of decimetre wavelengths (21 cm), this maser has a bandwidth larger than that of a travelling-wave maser.

In order to check experimentally the results of these investigations, the authors designed and investigated two models of a three-cavity solid-state ruby maser, which operates at a temperature of 4·2°K and has a bandwidth of 18 MHz with a gain of 20 dB. The small-size cavity system is formed by three closely-spaced quarter-wave strips. The size of the three-cavity system differs but little from the size of a single cavity. A permanent magnet is fixed directly in the cavity. With a cryostat capacity of 5 litres, the maser can operate for over 72 hours without replenishing the liquid helium.

'Reflex multi-cavity solid-state masers containing the active material in all the cavities', and

'A three-cavity solid-state maser at a wavelength of 21 cm', M. Y. Zhabotinsky and A. V. Frantsesson, *Radio Engineering and Electronic Physics* (English language edition of *Radiotekhnika i Elektronika*), **12**, No. 1, pp. 48–53, and pp. 54–57, 1967.

TRANSISTOR POWER SUPPLIES

In transistor d.c. power supply equipment electronic circuits are needed for safeguarding the supply units and the circuits connected to them against damage due to overloading or short-circuiting. Systems which cut off the power supply when overloading occurs can be made up from simple electronic circuits, but such systems have several disadvantages in practice. If a pure current-limiting system is used, there is a large additional power dissipation in the control transistor. Circuits have been developed by two Dutch engineers which hold the power loss in the control transistor to a constant value. Hybrid circuits of these systems have also been made. Two examples of these circuits used in production equipment are described.

'Current-limiting circuits for transistorized power supplies', R. Gasser and R. Hug, *Philips Technical Review*, **28**, No. 8, pp. 251–57, 1967.

'MOS' INTEGRATED CIRCUITS

After a short review of the advantages of MOS transistors in integrated circuits, a recent French paper presents two circuits realized at the integrated circuits laboratory of the Grenoble Nuclear Centre.

The first circuit contains eight MOS transistors which can be interconnected by thermo-compression of 25 μm gold wire in order to obtain such functions as a three inputs NOR gate, a memory cell, or a stage of a shift register.

The second circuit is a six-channel multiplexer with $R_{on} < 60\ \Omega$ for $V_{gs} - V_s = -10$ V, $I_{DSF} < 0{\cdot}1$ nA under 10 V reverse bias.

'MOS transistors integrated circuits', J. Borel, J. Lacour, and M. Verdonne, *L'Onde Electrique*, **47**, No. 484–5, pp. 944–49, July-August 1967.

EFFECT OF RADIATION ON 'MOS' DEVICES

A French paper shows the present state of knowledge concerning the effects of radiation on metal-oxide-semiconductor (MOS) devices. The interest in using such components in space circuitry is first discussed. The behaviour of electrical characteristics degradations under radiation is analysed and the variation of the turn-on threshold voltage is specially considered. A tentative physical explanation of the phenomena is proposed, based on ion displacements in the oxide during irradiation. Special experiments are indicated to verify this theory. In conclusion, the possibilities of using MOS transistors in an irradiated environment are reviewed.

'Radiation effect on MOS devices', P. Glotin, *L'Onde Electrique*, **47**, No. 484–5, pp. 950–53, July-August 1967.

MODEL AERIAL TESTING TECHNIQUES

Measurements performed on model aerials are made for a number of related reasons. In some fields the measurements provide a means of showing how closely a design conforms to the intended specifications, without the need for possibly expensive and time-consuming full scale tests. Other uses are to establish validity of theory in actual conditions.

The first part of an Australian paper consists of a discussion of the principles of model aerial testing techniques and the factors to be considered in the operation of a model aerial range. In particular, the development of the model aerial range used by the P.M.G. Research Laboratories is discussed.

This is followed by the description of two projects which illustrate the application of these techniques. One discusses the tests performed to increase the fading range of an existing medium frequency mast radiator by top-loading it rather than increasing its height, and the other reports development of a vertically polarized loft mounting television aerial specifically for use in Canberra.

'Model aerial testing techniques', O. F. Lobert and W. S. Davies, *Proceedings of the Institution of Radio and Electronics Engineers Australia*, **28**, No. 6, pp. 181–200, June 1967.