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 37 No. 1 JANUARY 1969

The Changing Years

THE pattern of co-operation now emerging among the engineering societies in Great Britain is but one facet of the changes taking place in the engineering profession. The multifarious applications of electronics are particularly well suited to this changing pattern as evidenced by the increasing number of joint meetings between this Institution and other Societies. The benefits thereby given to all engineers, scientists and industry will be obvious and all over the world they have welcomed the advantages of the more broadly based planning of forthcoming Conferences and, even more, the avoidance of clashes of dates and duplication of subjects.

Immediate examples of 1969 planning are the three I.E.R.E. conferences to be jointly sponsored by sister institutions and other societies. The Conference on "Lasers and Opto-electronics" to be held at the University of Southampton in March will present the first meeting of its kind in Great Britain for nearly five years. Work in this field will be reviewed with special reference to the electronic and radio engineering implications. British reports will be complemented by papers from Czecho-Slovakia, France, Germany, Italy, Japan, Poland and Switzerland, the U.S.S.R. and the U.S.A. The conference on "Digital Methods of Measurement" at the University of Kent at Canterbury in July will aim to bring together techniques employed in many different branches of electronics and thus encourage a broader approach to 'digital engineering'.

Whilst the above conferences are concerned with techniques generally having applications in electronic engineering, the third Conference will cover an industrial application of electronics which is now almost a distinct discipline in itself. The "Industrial Ultrasonics" Conference is a subject which the Institution has promoted for many years and the three-day meeting at Loughborough University of Technology in September will provide a valuable forum for designers and users.

In turn I.E.R.E. is joining conferences initiated by kindred societies. In 1969 these joint meetings will fall into four main groups—computers and control engineering, communications, semiconductor techniques, and education. "Computer Aided Design" is in Southampton in April, "Computer Science and Technology" at Manchester is in July and "Industrial Applications of Dynamic Modelling" at Durham in September. In April "Switching Techniques for Telecommunication Networks" will be held in London and in September the First European Microwave Conference will also be held in London: this will, it is hoped, be the forerunner of a peripatetic conference organized by national bodies at different centres every two years, in a similar way to the established Microwave and Optical Generation and Amplification Conference. The two semiconductor conferences will be respectively on "Microelectronics"—at Eastbourne in June and "Solid State Devices"—at Exeter in September. It is probable that these two complementary conferences will become regular annual or biennial events. Finally, the education conferences are, respectively, "Measurement Education" at Warwick in July and "Education and Training Technology", a very broadly supported venture which will include an exhibition in London in September.

Space has permitted only the briefest of references to these many joint meetings—fuller details have been or will be published elsewhere in the *Journal*. There is, however, clear evidence of a desire for co-operation between the Institutions in order better to serve the advancement of knowledge.

As a postscript, but perhaps even more significant, it may be noted that the Seventh Commonwealth Engineering Conference will be held in New Delhi in November 1969; its theme will be "Priorities in Public Works in Developing Countries". The fact that this meeting will follow closely on the Second General Assembly of the World Federation of Engineering Organizations gives added point to regarding 1969 as a year for technological co-operation.

G. D. C.

INSTITUTION NOTICES

New Year Honours

The Council of the Institution has congratulated the following members whose appointments to the Most Excellent Order of the British Empire were announced in Her Majesty's New Year Honours List.

To be Commander in the Military Division (C.B.E.)

Colonel Royston Knowles, C.Eng. (Fellow), lately R.E.M.E.

(Col. Knowles is Commanding Officer, 38 Central Workshop R.E.M.E., Chilwell; he serves on the Institution's Technical Committee and Instrumentation and Control Group Committee.)

To be Officer in the Military Division (O.B.E.)

Lieutenant-Colonel John Loudon Purdon, C.Eng. (Fellow), R. Sigs.

(Lt.-Col. Purdon was appointed to H.Q. Army Department Radio Service in 1966.)

To be a Member in the Civil Division (M.B.E.)

George Fothergill Budden, C.Eng. (Member).

(Mr. Budden is Assistant Engineer-in-Charge, Operations, North Region, B.B.C.)

Symposium on Patient Monitoring

The Scottish I.E.R.E.–I.E.E. Joint Medical Electronics Committee are holding a symposium on Patient Monitoring on Friday, 28th March 1969, in the Department of Medical Physics, Royal Infirmary, Edinburgh. Papers will discuss systems, techniques and ward experience and will be supported by demonstrations and a small exhibition.

Further information may be obtained from the Honorary Secretary of the Joint Committee, Mr. I. B. White, B.Sc., C.Eng., M.I.E.E., c/o West Instrument, 21 Alva Street, Edinburgh 2. (Tel. 031-225 7102.)

Software in Process Control

The Institute of Measurement and Control is arranging a Symposium on 'Experiences with Software in Process Control', to be held on 1st July 1969. It will cover the difficulties encountered and overcome in programming process computer control schemes, and the requirements for future software development in this area. The I.E.R.E., the I.E.E. and the British Computer Society are co-sponsors of the symposium, which will be held in London.

Further information is available from the Institute of Measurement and Control, 20 Peel Street, London, W.8.

Sir Arnold Lindley

Sir Arnold Lindley, a Vice-President of the I.E.R.E., has been elected President of the Institution of Mechanical Engineers for the year 1968–69.

In his Presidential Address, 'On Becoming Wise After the Event', delivered on 23rd October, Sir Arnold made the point that engineering was always changing and that every change was based on some previous experience—on becoming wise after the event. He believed that engineering was an evolutionary process which followed very closely the Darwinian pattern of natural selection and progressed only according to the quality of those responsible for steering its destiny. He concluded: 'It is the responsibility of all of us to keep this quality at the highest level through careful selection, through training and by providing opportunity for adequate experience.'

On 2nd December Sir Arnold, who is Chairman of the Engineering Industry Training Board, was awarded the Honorary Degree of Doctor of Science by the City University.

Secretary to the Council of Engineering Institutions

Michael William Leonard, B.Sc.(Eng.), C.Eng., F.I.C.E., M.I.Mech.E., has been appointed Secretary to the Council of Engineering Institutions with effect from 1st January 1969.

Mr. Leonard, who is 51, is a Chartered Civil and Mechanical Engineer and graduated in engineering at University College London. He joined John Mowlem & Co. Ltd., in 1941 and became one of the early members of staff of Soil Mechanics Ltd. He was appointed to the Board of that Company in 1958 and became Managing Director.

Mr. Leonard has taken an active part in the work of the Institution of Civil Engineers, serving on a number of its committees and on British Standards Institution Codes of Practice Committees. He was appointed a member of the Advisory Committee on Tip Safety set up by the Ministry of Power after the Aberfan disaster.

Correction

The following correction should be made in the Clerk Maxwell Address on 'Spatial Communications', published in the November 1968 issue.

Page 261, Sect. 2.1.2,

Equation (4) should read

$$P_K = 1{\cdot}38 \times 10^{-23}\, T_K B \text{ joules/second.}$$

U.D.C. 621.317.733.029.55

The Use of a Radio Frequency Transformer to Increase the Sensitivity of a Radio Frequency Admittance Bridge

By

I. L. FREESTON, B.Sc., M.Sc.†

Summary: This paper describes a technique by which the sensitivity of a radio-frequency admittance bridge was increased by using an r.f. transformer, in order to measure small changes of admittance. The bridge measured the admittance of the primary circuit with the unknown admittance connected across the secondary, the nominal value of the secondary to primary turns ratio being $\sqrt{12}$. Using this method, measurements at frequencies between 5 and 30 MHz were made of the admittance of a plane capacitor of about 1 pF capacitance, to an accuracy of about 2%, and of changes in this admittance to less than 4%. The system was calibrated by a least-squares method, which also enabled the standard errors in the calibration to be evaluated. The results of measurements made on the plane capacitor in vacuum are presented, and the advantages and limitations of the method are discussed.

List of Principal Symbols

n	secondary to primary turns ratio of the transformer
Y	admittance at end of line
Y_o	reference value of Y
Y_m	admittance measured by bridge
Y_{mo}	bridge reading corresponding to Y_o
α, β, γ	circuit parameters
δ_{yi}	standard error in imaginary part of Y
δ_{yr}	standard error in real part of Y

1. Introduction

This paper describes a method by which a radio-frequency admittance bridge was used to measure admittances comparable to the resolution of the bridge. Although the technique is of general use, its application to the specific problem for which the method was devised is described as an example of the method. The method was developed to overcome the difficulties involved in making laboratory measurements over a frequency range from 5 to 30 MHz of the admittance of a plane-grid capacitor, with variable separation, immersed in a plasma, and having a capacitance of 1 pF or less.

A Wayne Kerr admittance bridge B801B was available for these measurements and this instrument was able to measure admittance in the frequency range 1 to 100 MHz; however, the discrimination on the vernier capacitance scale was only 0·2 pF, and 0·02 mΩ^{-1} on the conductance scale. An increase in the accuracy of the measurements could be achieved by inserting an isolating transformer between the bridge and the unknown admittance. If the turns ratio of the secondary winding of the transformer to the primary winding is n, a change ΔY in the admittance to be measured results in a change in the admittance measured by the bridge given approximately by

$$\Delta Y_m = n^2 \Delta Y \qquad \ldots\ldots(1)$$

Therefore, an unknown admittance comparable in magnitude to the bridge discrimination could be made to give a bridge reading considerably greater than the discrimination, by suitably choosing the value of n.

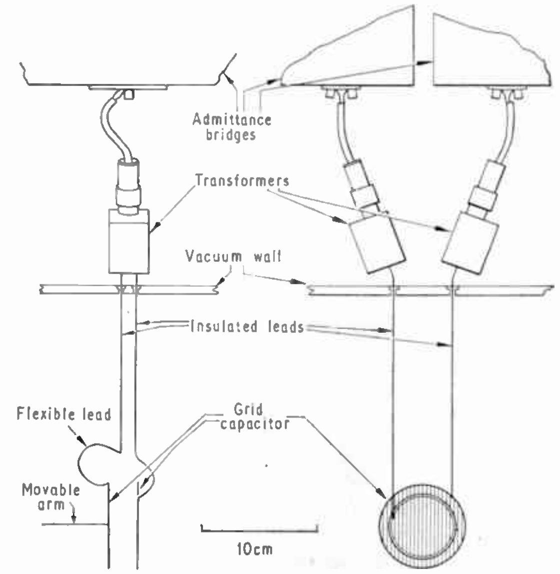


Fig. 1. Schematic arrangement of admittance measuring apparatus.

† Formerly at the Science Research Council, Radio and Space Research Station, Slough; now at the Department of Electronic and Electrical Engineering, The University of Sheffield.

In general, the bigger n the greater the increase in accuracy, but in practice a limit was set by the range of measurement of the bridge. A suitable transformer for the present purpose was the balun AT60N made by Hatfield Instruments, which had a frequency range of 3 to 300 MHz, and a value of n^2 of about 12.

The apparatus is shown in Fig. 1. The grid capacitor was protected by a guard ring, which was connected to a second Wayne Kerr bridge in an identical way to the capacitor to be measured. Both bridges were supplied by the same generator, so that the r.f. voltage on the capacitor and guard ring could be maintained equal in amplitude and phase. The vacuum wall and insulated leads were requirements of the plasma experiment.

2. Calibration

The object of the calibration was to find an accurate relationship between the admittance at the end of the line, Y, and the admittance measured by the bridge, Y_m. (The circuit between the point to which Y is connected and the point at which Y_m is measured will be referred to subsequently as the line.) If it is assumed that the transformer is a linear device (i.e. that the current through any element is directly proportional to the voltage across it), then the relationship between Y and Y_m is of the form

$$Y_m = \frac{a+bY}{c+dY} \qquad \ldots\ldots(2)$$

where, in general, the parameters a, b, c and d are complex and functions of frequency. To calibrate the system it is necessary to determine these parameters. It is not possible, or necessary, to determine all four unknowns, but only the ratio of any three to the fourth. If the numerator and denominator on the right-hand side of eqn. (2) are divided throughout by d, then the resulting equation can be written

$$\alpha+\beta Y-\gamma Y_m - YY_m = 0 \qquad \ldots\ldots(3)$$

where α, β and γ are complex. If Y_m is measured for three different known values of Y, then there are three complex equations for the three complex unknowns, and α, β and γ can be determined. However, if more than three sets of measurements of Y and Y_m are made, it is possible to use a least-squares method to determine the values of α, β and γ, and also to estimate the standard error in these parameters.[1, 2] It is shown in Appendix 1 that the values of the real and imaginary parts of α, β and γ given by this method are those for which

$$\Delta = \sum_{p=1}^{s} (\alpha+\beta Y_p-\gamma Y_{mp} - Y_p Y_{mp}) \times$$
$$\times (\alpha+\beta Y_p-\gamma Y_{mp} - Y_p Y_{mp})^* \qquad \ldots\ldots(4)$$

is a minimum, where s is the number of sets of measurements of Y and Y_m.

The decision to divide the numerator and denominator of eqn. (2) by d was arbitrary; the least-squares method would produce an equally valid, but different result if a, b or c had been chosen as divisor. The discrepancy between the results produced by these different choices is an indication of the accuracy with which eqn. (2) describes the data. The results are identical if the data are described exactly by eqn. (2). As will be seen later, division by d gave a result which was sufficiently accurate for the present purpose.

The measurements of Y and Y_m were made as follows. The grids were removed and the line lengthened to the point corresponding to the centre of the grids. A number of calibrating admittances were made from standard components; a typical value was a nominal 10 pF capacitor in parallel with a nominal 1000 Ω resistor. The value of this admittance was measured on the vernier scales of the bridge to determine Y, and then immediately attached to the end of the line to determine Y_m. The value of Y_m was found by taking the difference between the bridge reading with the line disconnected from the bridge terminals, and the reading with the line connected to the terminals and the admittance Y connected to the far end of the line. This procedure eliminated any effects due to changes in the value of components with time. Care was taken in connecting the components to the bridge and the line to ensure that the measurement procedure gave repeatable results.

In general the standard error in the circuit constants, α, β and γ, was reduced as the number of observations was increased, but a limit was set by the inherent accuracy of the bridge. Although as many as 20 observations were made at one frequency, it was found in practice that eight were usually sufficient to achieve the limiting accuracy. To explain this it should be emphasized that the values, Y, of the components used in the calibration were those values measured by the bridge. These values differ from the true value of the components by the amount by which the bridge scale is in error. This produces an error in the calculated circuit constants, which cannot be reduced by taking further measurements, because it is due to the inherent limitation of the bridge accuracy. For the present purpose, the standard errors in α, β and γ due to this limitation were acceptably small, but should greater accuracy be required it would be necessary to calibrate the bridge scales against absolute standards.

The values of Y and of the corresponding measured Y_m used in the calibration at 20 MHz are shown in Fig. 2. The values of α, β and γ calculated from these data, together with their standard errors, are

$$\alpha = (40{\cdot}34 \pm 0{\cdot}78) - j(30{\cdot}21 \pm 0{\cdot}78) \quad (\text{m}\Omega^{-1})^2$$
$$\beta = (-3{\cdot}03 \pm 1{\cdot}55) - j(61{\cdot}04 \pm 1{\cdot}55) \quad \text{m}\Omega^{-1}$$
$$\gamma = (-0{\cdot}06 \pm 0{\cdot}12) - j(6{\cdot}89 \pm 0{\cdot}12) \quad \text{m}\Omega^{-1}$$

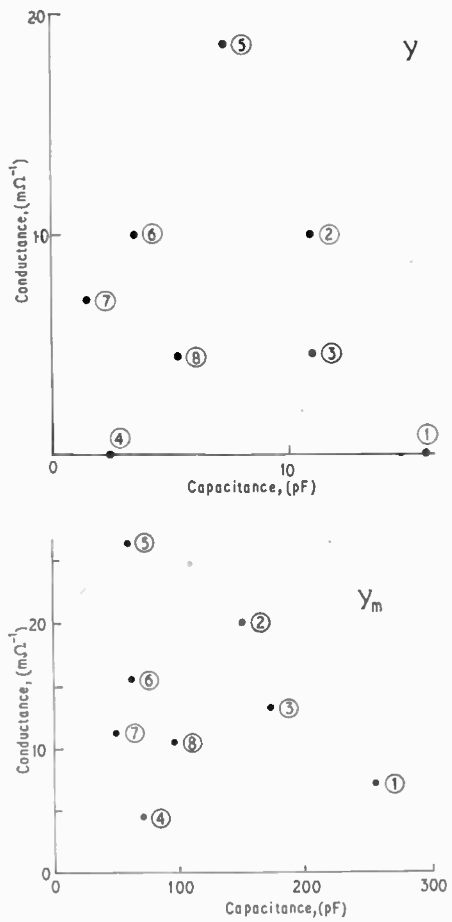


Fig. 2. Typical values of Y and Y_m used to calibrate circuit (at 20 MHz).

It can be readily shown from the method used to evaluate the standard errors[1] that the errors in the real and imaginary parts of each of α, β and γ are equal.

In addition to the uncertainty in the circuit constants due to the bridge limitations, there is the possibility of errors arising because of stray capacitance between the line and nearby electrodes. In accordance with the Wayne Kerr Instruction Manual,[3] one side of the leads from the bridge to the transformer was earthed. Therefore, any stray capacitance to earth from the other parts of the line connecting the bridge and the admittance to be measured are effectively coupled to this lead. The effect of this is to include the strays in the equations relating Y and Y_m, and as the circuit was used in the same conditions as it was calibrated, no error from strays should result.

The other source of error in the calibration is the random error involved in reading the bridge scales. This contribution to the error is reduced by increasing the number of readings but, as explained above, the standard error attains a value independent of the number of readings for other reasons and therefore, when this is the case, the contribution of the random error in the readings to the standard errors in the circuit coefficients will be small.

The assumption of a linear circuit was tested in the following way. Equation (2) can be rewritten as

$$Y = \frac{\gamma Y_m - \alpha}{\beta - Y_m} \qquad \text{......(5)}$$

and using the measured data for Y_m and the calculated values of α, β and γ, it is possible to calculate an

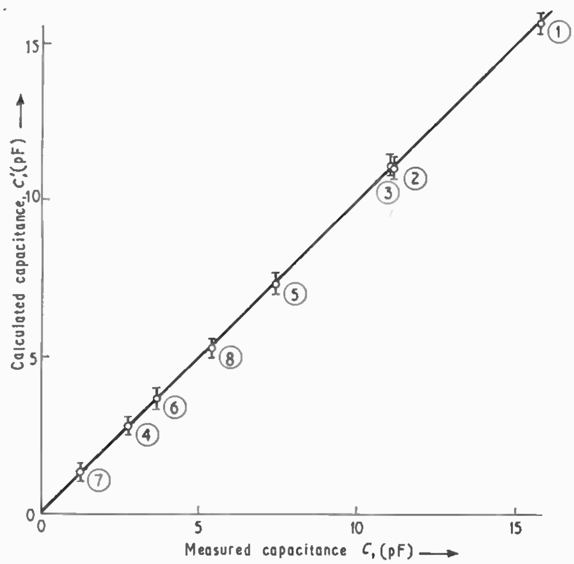


Fig. 3. Comparison of calculated capacitance C' with measured capacitance C.

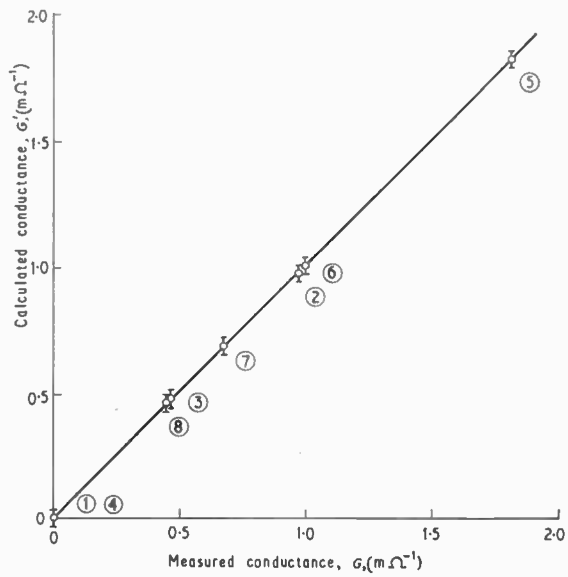


Fig. 4. Comparison of calculated conductance G' with measured conductance G.

admittance Y', which is an estimate of the original value used in the calibration. Comparison of Y and Y' is a test of whether or not the assumed relationship in eqn. (2) is accurate. Figures 3 and 4 show the real and imaginary parts of Y' plotted against the real and imaginary parts of Y respectively for the data shown in Fig. 2. It will be seen that the points lie very close to a straight line of slope unity and that the deviation of the points from the line is less than the standard error in the calculated quantities, G' and C'. These results confirm the adequacy of the assumption of a linear circuit and of the least-squares procedure outlined above.

The above procedure was repeated at all frequencies at which experimental measurements were to be made.

3. Admittance Measurements

The parallel grid capacitor was placed at the end of the line, and admittance measurements were made in free space. The admittance Y can be determined from the measured admittance Y_m and the calibration constants using eqn. (5).

Here Y_m was again the difference between the bridge readings with the line disconnected and connected, with Y connected at the other end of the line. Equation (5) was also used to calculate the standard errors in the real and imaginary parts of Y, δ_{Yr} and δ_{Yi} respectively, due to the standard errors in α, β and γ.[2] These errors were found to be unacceptably large, being comparable in magnitude to the admittance being measured, which was in general much smaller than the admittances used in the calibration. The reason for this is that the errors in Y are approximately proportional to the value of Y_m used in eqn. (5) to determine Y. It can be seen from eqn. (2) that this value of Y_m has contributions from terms that are independent of Y, and for small values of Y these terms are greater than the Y-dependent contributions, and therefore make a larger contribution to the standard error in Y. The most important of these Y-independent terms is that due to the admittance of the primary winding of the transformer.

Therefore, to overcome this an alternative method of determining Y was devised in which a change in Y was calculated from a change in the bridge reading. If it is assumed that at a particular grid spacing in vacuum the capacitor has a known value of admittance, Y_o, and that the corresponding bridge reading is Y_{mo}, then from eqn. (5) we get

$$Y_o = \frac{\gamma Y_{mo} - \alpha}{\beta - Y_{mo}} \qquad \ldots\ldots(6)$$

The admittance at the end of the line is then changed in some way to an unknown value Y, either by altering the spacing or by the introduction of a dielectric. It is shown in Appendix 2 that from eqns. (5) and (6), Y is given by:

$$Y = Y_o + \frac{(\gamma + Y_o)^2 (Y_m - Y_{mo})}{(\beta\gamma - \alpha) - (\gamma + Y_o)(Y_m - Y_{mo})} \qquad \ldots\ldots(7)$$

$(Y_m - Y_{mo})$ is the change in the bridge reading caused by the change in Y, and it could be measured very accurately on the vernier scales of the bridge without disconnecting the line from the bridge terminals. The standard errors in the determination of Y, δ_{Yr} and δ_{Yi} could also be found from eqn. (7).

To use eqn. (7) it was necessary to know the reference admittance Y_o and this was determined in the following way. The grid spacing, which could be accurately controlled, was set at 1·5 cm and the bridge reading at this position was used as Y_{mo}. The free space capacitance of the capacitor was calculated by using the formula for a plane capacitor, assuming a spacing of 1·5 cm, and this value was used as the reference admittance Y_o. The admittance of the capacitor was then determined by using eqn. (7) for various grid separations. A plot of the imaginary part of the admittance against the reciprocal of the grid separation is shown in Fig. 5, together with the standard errors, for measurements made at 20 MHz. A straight line fits the data accurately, showing that there was no appreciable effect from stray capacitance, or the grid-like nature of the capacitor, over the range of measurement. The area of the grids calculated from the slope agreed with the measured value of 20 cm^2, showing that the initial chosen value of Y_o was accurate. If this chosen value of Y_o had been in error, then the variation of capacitance with the reciprocal of grid separation would not have been a straight line through the origin, and it would have been necessary to choose a new value for Y_o. The conductance term calculated from eqn. (7) was zero to within the standard error.

The standard error in these measurements, due to the standard error in α, β and γ, was of the order shown in Fig. 5. This increase in accuracy over the previous determination of Y using eqn. (5), arises because the standard error in Y is proportional to Y_m using eqn. (5), but only proportional to $(Y_m - Y_{mo})$ using eqn. (7). The contributions to the standard errors in Y, of uncertainties in Y_o and $(Y_m - Y_{mo})$ have so far been ignored, but they can be included. However, it was found that these contributions were much smaller than those due to uncertainties in the circuit constants and so they have been ignored. Therefore, the standard error in these measurements is due to the errors in the circuit calibration, and as discussed in Section 2, these errors arise mainly because the bridge scales were not calibrated absolutely.

The accuracy of this method of determining Y depends on how accurately the difference from the known reference admittance Y_o can be measured, and on the accuracy with which Y_o can be determined. From Fig. 5, the largest error obtained in the capacitances measured at various reciprocals of the grid separation is about 0·025 pF, and therefore the percentage error in the chosen value of the reference admittance cannot be greater than $100 \times 0{\cdot}025/1{\cdot}2 = 2\%$. Further, taking the point at a reciprocal of grid

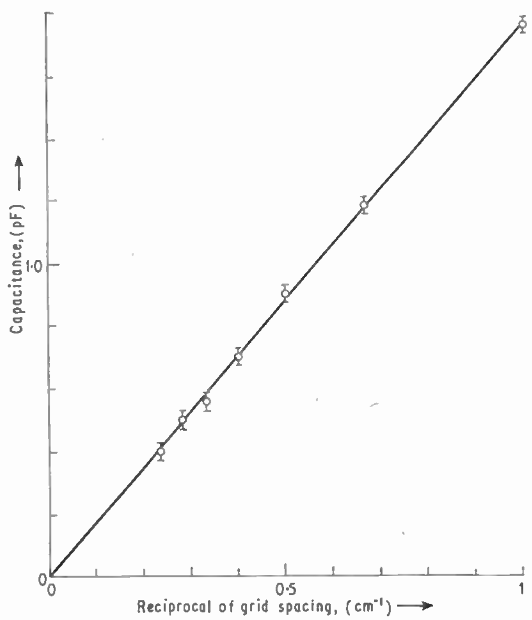


Fig. 5. Variation of free space capacitance of grids with spacing.

spacing of 0·28 cm^{-1} as typical, the percentage error in the difference from the reference admittance is $100 \times 0{\cdot}25/0{\cdot}7 = 3{\cdot}5\%$. In the plasma experiment to which this technique was applied only the ratio of the measured admittance to the reference was of interest and this could therefore be determined to 3·5%. The percentage error in the conductance term was of the same order when referred to ωC_o, where C_o is the capacitance of the reference admittance.

4. Conclusions

This method has been shown to provide a means of using an admittance bridge to measure an admittance comparable in magnitude with the bridge discrimination to a high degree of accuracy. This advantage is gained at the expense of two complications, the necessity for tedious calculations, and the requirement of an accurately determined value of Y_o. In certain circumstances the calculations may be simplified if the circuit parameters are such that some terms in eqn. (7) are negligible. Alternatively, a computer can be readily programmed to evaluate Y and its standard errors. The problems of an accurately-determined standard of admittance can be overcome if the admittance to be measured can be connected and disconnected easily. Then $Y_o = 0$ with the admittance disconnected and this can be used as the reference standard.

Using this technique, measurements were made between 5 and 30 MHz. The lower frequency limit was set by the inductance of the transformer primary behaving as a negative capacitance and producing a balance point outside the range of the bridge. If necessary a capacitor could have been connected across the primary winding to extend the range below 5 MHz. The upper limit of 30 MHz was set by the plasma experiment, but there appeared no reason why the technique could not be extended above this frequency. The admittance range which could be measured is roughly given by R/n^2, where R is the range of the bridge scales and n is the transformer turns ratio. With the particular transformer used, which had a nominal value of $n^2 = 12$, an admittance range between ± 40 pF and 0 to 8 mΩ^{-1} on the coarse bridge scales, and between $\pm 2{\cdot}5$ pF and 0 to 1 mΩ^{-1} on the vernier scales could be measured.

5. References

1. Whittaker, E. and Robinson, G., 'The Calculus of Observations', 4th edition, Chapter IX (Blackie, London, 1944).
2. Topping, J., 'Errors of Observation and Their Treatment', 3rd edition, p. 86 *et seq.* (Chapman and Hall, London, 1962).
3. Instruction Manual, V.H.F. Admittance Bridge B801B (Wayne Kerr Laboratories, New Malden, Surrey).

6. Appendix 1

Least-squares Method for Complex Equations

Suppose that s independent sets of measurements are made of Y and Y_m, and that the admittances are labelled Y_p and Y_{mp}, where $p = 1, \ldots, s$. Then from eqn. (3), there are s equations of the form

$$\alpha + \beta Y_p - \gamma Y_{mp} - Y_p Y_{mp} = 0, \qquad p = 1, \ldots, s \quad \ldots\ldots(8)$$

from which α, β and γ are to be determined. This set of equations is termed the equations of condition. These are not necessarily consistent, since the values of α, β and γ found from any three equations will not in general satisfy all the other equations. Consider first the case when all the terms in eqn. (8) are real. The method of least-squares is a way of using all the equations of condition to find the values of α, β and γ such that the sum of the squares of the left-hand sides of equations (8) is a minimum. That is

$$\Delta = \sum_{p=1}^{s} D_p^2 = \sum_{p=1}^{s} (\alpha + \beta Y_p - \gamma Y_{mp} - Y_p Y_{mp})^2 \quad \ldots\ldots(9)$$

is a minimum. The equations for α, β and γ are

$$\frac{\partial \Delta}{\partial \alpha} = \frac{\partial \Delta}{\partial \beta} = \frac{\partial \Delta}{\partial \gamma} = 0 \quad \ldots\ldots(10)$$

Now consider the case when all the terms in eqn. (8) are complex. The equations of condition are

$$(\alpha_r + j\alpha_i) + (\beta_r + j\beta_i)(Y_{pr} + jY_{pi}) - \\ -(\gamma_r + j\gamma_i)(Y_{mpr} + jY_{mpi}) - \\ -(Y_{pr} + jY_{pi})(Y_{mpr} + jY_{mpi}) = 0, \qquad p = 1, \ldots, s \qquad \ldots\ldots(11)$$

where the subscripts 'r' and 'i' indicate the real and imaginary parts respectively. The six unknowns, α_r, α_i, β_r, β_i, γ_r and γ_i have to be determined. Equation (11) can be separated into real and imaginary parts and be written as

$$\alpha_r + \beta_r Y_{pr} - \beta_i Y_{pi} - \gamma_r Y_{mpr} + \gamma_i Y_{mpi} - Y_{pr} Y_{mpr} + \\ + Y_{pi} Y_{mpi} + j(\alpha_i + \beta_i Y_{pr} + \beta_r Y_{pi} - \gamma_r Y_{mpi} - \\ - \gamma_i Y_{mpr} - Y_{pr} Y_{mpi} - Y_{pi} Y_{mpr}) = 0, \qquad p = 1, \ldots, s$$

or

$$D_{pr} + jD_{pi} = 0, \qquad p = 1, \ldots, s \qquad \ldots\ldots(12)$$

Since both the real and imaginary parts of eqn. (12) must be zero separately, there are $2s$ equations of condition

$$D_{pr} = 0, \qquad D_{pi} = 0 \quad p = 1, \ldots, s \ \ldots\ldots(13)$$

Then by the least-squares method, the values of the unknowns are those for which

$$\Delta = \sum_{p=1}^{s} (D_{pr}^2 + D_{pi}^2)$$

is a minimum. But

$$\Delta = \sum_{p=1}^{s} (D_{pr} + jD_{pi})(D_{pr} - jD_{pi}) \\ = \sum_{p=1}^{s} (\alpha + \beta Y_p - \gamma Y_{mp} - Y_p Y_{mp}) \times \\ \times (\alpha + \beta Y_p - \gamma Y_{mp} - Y_p Y_{mp})^* \ \ldots\ldots(14)$$

The equations for the six unknowns are found by partially differentiating Δ with respect to each of the six unknowns.

7. Appendix 2

Incremental Method for Y

The relationship between Y and Y_m is

$$Y = \frac{\gamma Y_m - \alpha}{\beta - Y_m} \qquad \ldots\ldots(15)$$

If under some repeatable reference conditions $Y = Y_o$, and the corresponding value of Y_m is Y_{mo}, then

$$Y_o = \frac{\gamma Y_{mo} - \alpha}{\beta - Y_{mo}} \qquad \ldots\ldots(16)$$

These two equations can be used to express Y in terms of Y_o and the change in the measured admittance $(Y_m - Y_{mo})$. From eqn. (16)

$$Y_{mo} = \frac{\alpha + \beta Y_o}{\gamma + Y_o} \qquad \ldots\ldots(17)$$

and therefore, the change in measured admittance can be written as

$$(Y_m - Y_{mo}) = Y_m - \frac{\alpha + \beta Y_o}{\alpha + Y_o}$$

This can be rewritten as an equation for Y_m,

$$Y_m = (Y_m - Y_{mo}) + \frac{\alpha + \beta Y_o}{\gamma + Y_o} \qquad \ldots\ldots(18)$$

Substitution of this expression for Y_m in eqn. (15) and rearrangement gives

$$Y = Y_o + \frac{(\gamma + Y_o)^2 (Y_m - Y_{mo})}{(\beta\gamma - \alpha) - (Y_m - Y_{mo})(\gamma + Y_o)}$$

Manuscript first received by the Institution on 22nd February 1968 and in final form on 27th September 1968. (Paper No. 1231/CC31.)

U.D.C. 621.382.23

A Theoretical Appraisal of the Quenched Multiple Domain Mode in GaAs Microwave Diodes

By

R. I. HARRISON, Ph.D.,†

S. P. DENKER, Ph.D.,†

AND

H. BERGER, Ph.D.†

Summary: Quenched multiple domain (q.m.d.) mode oscillations in bulk GaAs have been analysed to obtain the relationships between efficiency, fundamental-frequency power output, fundamental-frequency negative resistance, and d.c. bias voltage. It is found that this highly efficient mode does not have the limitations of transit-time devices and consequently has a frequency-independent power × impedance product. The results demonstrate that q.m.d. oscillators designed with multiple-tuned circuits can operate at twice the efficiency of those operated in single-tuned circuits.

1. Introduction

The observed properties of transferred-electron gallium arsenide oscillators, e.g. the oscillation frequency and diode waveform, are often determined by the external circuit or cavity rather than by the diode itself. Of particular importance is the highly efficient quenched multiple domain mode[1–3] (q.m.d.) which may occur if

$$fL > 10^{7} \qquad \text{......(1a)}$$

$$nL > 10^{12}\ \text{cm}^{-2} \qquad \text{......(1b)}$$

where f is the oscillation frequency in Hz, L device length in cm and n the doping density in cm^{-3}.

If the voltage across a GaAs diode, with nL product exceeding $10^{12}\ \text{cm}^{-2}$, is raised above a threshold value V_T, a dipole domain can be formed at the cathode contact or any other major doping fluctuation. This domain travels steadily from its nucleation site to the anode contact where it is absorbed, giving rise to periodic transit-time oscillations.

Devices having nL products exceeding $10^{12}\ \text{cm}^{-2}$ can also oscillate at frequencies several times their transit-time frequency, and can involve the formation of multiple domains, if they are imbedded in a cavity whose frequency satisfies relation (1a).[4] To achieve stable operation in the q.m.d. mode the semiconductor should be reasonably homogeneous with no major doping fluctuations.[3, 4] Small, random, doping fluctuations present in most GaAs diodes can, in some cases give rise to field inhomogeneities which rapidly grow into multiple dipole domains to produce q.m.d. oscillations.[3] However, multiple domains can be produced more reliably by an array of evenly-spaced nucleation sites, e.g. a grating or metal stripes, along the diode length separated by $L \lesssim 10^{7}/f$ or $10^{12}/n$ cm. These domains would be more or less identical. In steady state, multiple domains can grow and quench synchronously throughout the entire device length, leading to the terminal behaviour described in this paper.

Since at most one domain can form if nL equals $10^{12}\ \text{cm}^{-2}$, the total number of domains is given by

$$N \simeq nL \times 10^{-12} \qquad \text{......(2)}$$

Although in practice a device need not be entirely filled with domains, we will assume an ideal situation in which eqn. (2) is satisfied.

Multiple dipole domains form more rapidly than a single domain since the time-constant for forming and quenching one domain is N times that for quenching N domains.[3] A single domain must be discharged through a positive resistance whose value depends on the sample length, whereas N domains discharge through a resistance whose value is approximately N^{-1} times as large.[3]

Quenched multiple domain (q.m.d.) mode oscillations in bulk GaAs have been analysed to obtain the relationships between efficiency, fundamental-frequency power output, fundamental-frequency negative resistance, the d.c. bias voltage. It is found that this highly efficient mode does not have the limitations of transit-time devices and consequently has a frequency-independent power × impedance product. The results demonstrate that q.m.d. oscillators designed with multiple-tuned circuits can operate at twice the efficiency of those operated in single-tuned circuits.

Two analyses of the quenched multiple domain mode are presented here. In one case a sinusoidal terminal voltage is assumed. The results indicate that, with the approximations made, peak efficiency is

† Research Centre of General Telephone & Electronics Laboratories Inc., 208–20 Willets Point Boulevard, Bayside, New York 11360.

limited to approximately 7%, in agreement with the special case analysed by Warner.[6] This type of operation is relatively simple because it only requires a single-tuned load circuit and generates no harmonic power. The second case analysed involves a square-wave terminal voltage. This operation yields much higher efficiencies at the fundamental frequency (~18%) and produces some output power at the third, fifth, seventh, etc., harmonics. The necessary load circuit is more complicated in that it must include multiple-tuned circuits to filter the unwanted harmonic powers.

2. Quenched Multiple Domain Operation with Sinusoidal Terminal Voltage

The following assumptions are made in this analysis:

(i) The device consists of a uniform cross-section of bulk GaAs between two non-injecting contacts.

(ii) The carrier density within the bulk semiconductor is uniform except for purposeful or accidental inhomogeneities which give rise to multiple domains.

(iii) Equal multiple Gunn domains are simultaneously nucleated and grow to their steady-state condition in a time short compared with the microwave period.

(iv) The multiple domains are simultaneously quenched in a time short compared with a microwave period.

(v) The time-dependent part of the terminal voltage across the bulk semiconductor is purely sinusoidal.

(vi) The domain threshold field and the domain quench field are equal.

(vii) The current density during the interval when steady-state domains exist is constant.

(viii) The bulk semiconductor is fully packed with domains.

Figure 1 and eqns. (3–15) in this analysis define all the pertinent quantities:

$$V(t) = V_{\text{d.c.}} - V_1 \cos \omega t \qquad \ldots\ldots(3)$$

$$I(t) = \frac{V(t)}{R_o}; \quad -\tau < t < \tau \qquad \ldots\ldots(4)$$

$$I(t) = I_d; \quad +\tau < t < \frac{T}{2} \text{ and } -\frac{T}{2} < t < -\tau \qquad \ldots\ldots(5)$$

$$V_T = V_{\text{d.c.}} - V_1 \cos \omega\tau \qquad \ldots\ldots(6)$$

$$V_{\min} = V_{\text{d.c.}} - V_1 \qquad \ldots\ldots(7)$$

$$V_T = I_{\max} R_o \qquad \ldots\ldots(8)$$

$$V_{\min} = I_{\min} R_o \qquad \ldots\ldots(9)$$

where R_o is the device low-field resistance

$$V' = \frac{V_{\text{d.c.}}}{V_T} \quad (>1) \qquad \ldots\ldots(10)$$

$$I'_{\min} = \frac{I_{\min}}{I_{\max}} \quad (<1) \qquad \ldots\ldots(11)$$

$$I'_d = \frac{I_d}{I_{\max}} \quad (<1) \qquad \ldots\ldots(12)$$

$$\theta = \omega\tau \qquad \ldots\ldots(13)$$

$$\cos\theta = \left(\frac{V'-1}{V'-I'_{\min}}\right) \qquad \ldots\ldots(14)$$

$$P_T = V_T I_{\max} \qquad \ldots\ldots(15)$$

Since the number of multiple domains given by eqn. (2) is proportional to the device length, R_o and the voltages V_T and $V_{\min}$ can be viewed equivalently as being proportional either to N or L (cf. eqns. (8) and (9)).

By Fourier analysis of the current, the amplitude of the current flowing at the fundamental frequency I_1 is

$$\frac{I_1}{I_{\max}} = \frac{1}{2\pi}\left[\frac{(V'+1-2I'_d)\sin 2\theta - (V'-1)2\theta}{\cos\theta}\right] \qquad \ldots\ldots(16)$$

Similarly, the direct current $I_{\text{d.c.}}$ is given by

$$\frac{I_{\text{d.c.}}}{I_{\max}} = \frac{1}{\pi}\left[\pi I'_d + (V'-I'_d)\theta - (V'-1)\tan\theta\right] \qquad \ldots\ldots(17)$$

the fundamental-frequency power output (P_1) delivered

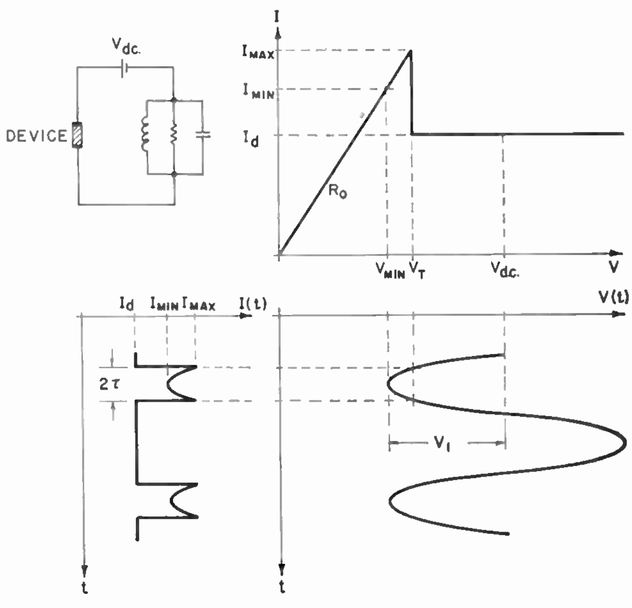


Fig. 1. Circuit and terminal current-voltage curves for quenched multiple domain operation with sinusoidal terminal voltage.

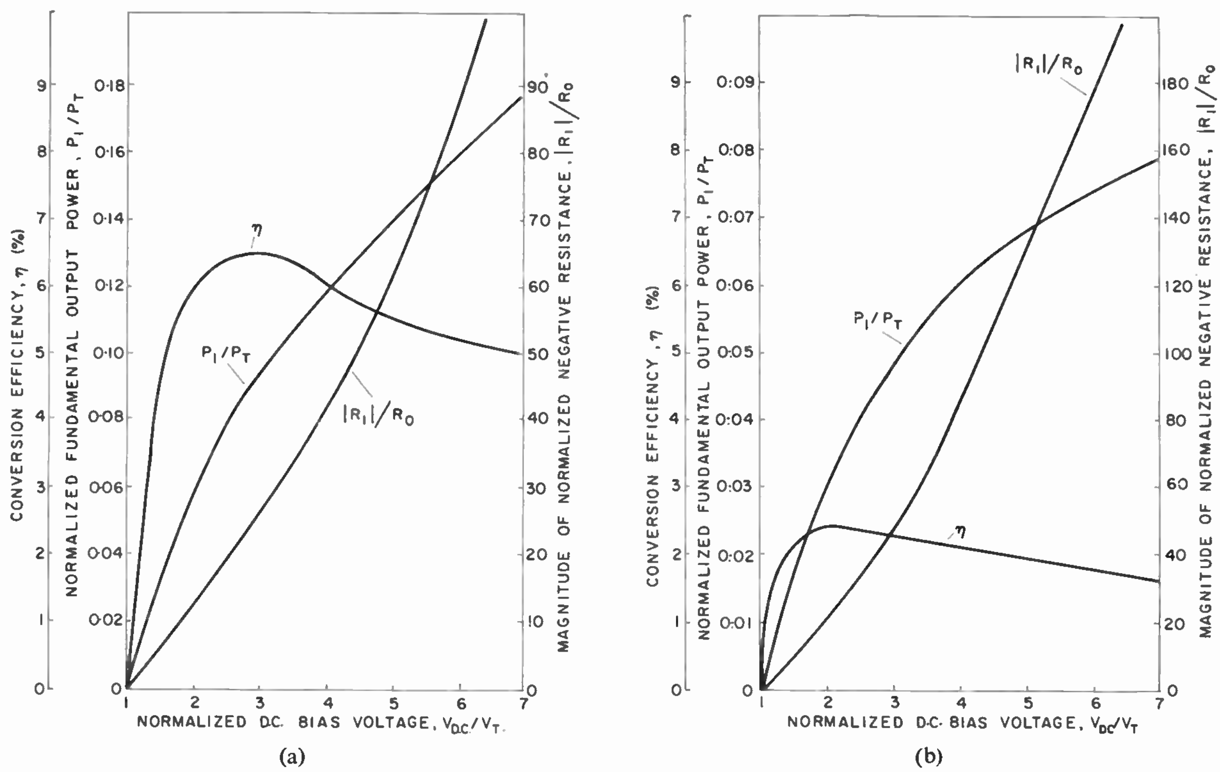


Fig. 2. Performance characteristics for q.m.d. oscillation with sinusoidal terminal voltage (a) $I_d' = 0{\cdot}5$, (b) $I_d = 0{\cdot}7$.

to a matched load is

$$\frac{P_1}{P_T} = \frac{1}{4\pi}\left(\frac{V'-1}{\cos^2\theta}\right)[(V'+1-2I_d')\sin 2\theta-(V'-1)2\theta] \quad \ldots\ldots(18)$$

and the d.c. power input ($P_{d.c.}$) to the device would be given by

$$\frac{P_{d.c.}}{P_T} = \frac{V'}{\pi}[\pi I_d'+(V'-I_d')\theta-(V'-1)\tan\theta] \quad \ldots\ldots(19)$$

From eqns. (18, 19) the conversion efficiency η is:

$$\eta = \frac{1}{4}\left(\frac{V'-1}{V'}\right)\left(\frac{1}{\cos^2\theta}\right)\times \left[\frac{(V'+1-2I_d')\sin 2\theta-(V'-1)2\theta}{\pi I_d'+(V'-I_d')\theta-(V'-1)\tan\theta}\right] \quad \ldots\ldots(20)$$

the magnitude of the negative resistance ($|R_1|$) of the device is:

$$\frac{|R_1|}{R_0} = 2\pi\frac{(V'-1)}{[(V'+1-2I_d')\sin 2\theta-(V'-1)2\theta]} \quad \ldots\ldots(21)$$

and the device d.c. resistance $R_{d.c.}$ is:

$$\frac{R_{d.c.}}{R_0} = \frac{\pi V'}{[\pi I_d'+(V'-I_d')\theta-(V'-1)\tan\theta]} \quad \ldots\ldots(22)$$

The characteristics can be obtained for this type of operation computed with the preceding equations, and are presented in Fig. 2. The maximum efficiencies obtainable are of the order of 7% when $I_d = 0{\cdot}5$ and 3% when $I_d' = 0{\cdot}7$. Practical devices operate with $0{\cdot}5 \le I_d' < 0{\cdot}7$. Somewhat higher efficiencies would be calculated if one took account of the finite domain growth and quench times. On the other hand, a domain quench voltage which differs from the threshold voltage would give lower efficiency. With a high nL product (cf. eqn. (2)) the domain quench field should be nearer $1/2V_T$ (see Ref. 2).

3. Quenched Multiple Domain Operation with a Square-wave Terminal Voltage

Operation with a square-wave terminal voltage could yield much higher efficiencies at the fundamental frequency. To achieve this higher efficiency, however, a more complicated microwave circuit is required to separate out unwanted generated harmonic power. This circuit is shown in Fig. 3. As before, performance is not transit-time limited.

In this analysis most assumptions are identical to those of the previous analysis. The exceptions are that the time-dependent part of the terminal voltage across the bulk semiconductor is a square-wave and the minimum total voltage equals the threshold voltage V_T, as shown in Fig. 3.

For convenience, we define the following normalized quantities:

$$V' = \frac{V_{d.c.}}{V_T} \quad (>1) \qquad \text{......(23)}$$

$$I'_d = \frac{I_d}{I_M} \qquad \text{......(24)}$$

k = harmonic number
(with $k = 1$ for the fundamental)

$$P_T = V_T I_M \qquad \text{......(25)}$$

Fourier analysis of the current and voltage waveforms yields

$$\frac{I_k}{I_M} = \frac{2}{k\pi}(1-I'_d) \quad k = 1, 3, 5 \ldots \qquad \text{......(26)}$$

$$\frac{I_{d.c.}}{I_M} = \tfrac{1}{2}(1+I'_d) \qquad \text{......(27)}$$

$$\frac{V_k}{V_T} = \frac{4}{k\pi}(V'-1) \quad k = 1, 3, 5 \ldots \qquad \text{......(28)}$$

The harmonic frequency output power P_k, delivered to matched loads, is

$$\frac{P_k}{P_T} = \left(\frac{2}{k\pi}\right)^2 (1-I'_d)(V'-1) \quad k = 1, 3, 5 \ldots \qquad \text{......(29)}$$

The efficiency at the fundamental frequency η_1 is

$$\eta_1 = \frac{8}{\pi^2}\left(\frac{1-I'_d}{1-I'_d}\right)\left(\frac{V'-1}{V'}\right) \qquad \text{......(30)}$$

The magnitude of the harmonic frequency negative resistance $|R_k|$ is

$$\frac{|R_k|}{R_o} = 2\left(\frac{V'-1}{1-I'_d}\right) \qquad \text{......(31)}$$

The d.c. resistance $R_{d.c.}$ is

$$\frac{R_{d.c.}}{R_o} = 2\left(\frac{V'}{1+I'_d}\right) \qquad \text{......(32)}$$

$R_{d.c.}$ is greater than 0, i.e. the device is d.c. stable.

It is of interest to find the total generated power which is the sum of the individual harmonic powers:

$$\frac{\sum_k P_k}{P_T} = \tfrac{1}{2}(1-I'_d)(V'-1) \quad k = 1, 3, 5 \qquad \text{......(33)}$$

The total conversion efficiency η_{tot} would then be

$$\eta_{tot} = \frac{\sum_k P_k}{P_{d.c.}} = \left(\frac{1-I'_d}{1+I'_d}\right)\left(\frac{V'-1}{V'}\right) \quad k = 1, 3, 5 \qquad \text{......(34)}$$

The ratio of the total generated harmonic power

$$\sum_k [P_k - P_1]$$

to the generated fundamental frequency power, P_1, can be found to be

$$\frac{\sum_k [P_k - P_1]}{P_1} = 0{\cdot}23 \text{ (or } -6{\cdot}38 \text{ dB)} \quad k = 1, 3, 5 \qquad \text{......(35)}$$

The ratio of a specific generated harmonic power P_k (k = 3, 5, 7...) to the generated fundamental-frequency power P_1

$$\frac{P_k}{P_1} = \frac{1}{k^2} \quad k = 3, 5, 7 \ldots \qquad \text{......(36)}$$

is as shown in Table 1.

Table 1

k	$\frac{P_k}{P_1}$ (dB)
3	−9·542
5	−13·98
7	−16·90

The circuit necessary to filter all the unwanted harmonics is shown in Fig. 3. This circuit consists of harmonic filters. In practice, the performance predicted by the above analysis should hold with very few tuned circuits. Figure 4 shows the computed values of η_1, $|R_k|/R_o$, and P_1/P_T as a function of V' with $I'_d = 0{\cdot}5$ and $I'_d = 0{\cdot}7$. Comparison with the results of the previous analysis shows that much higher efficiencies are attainable.

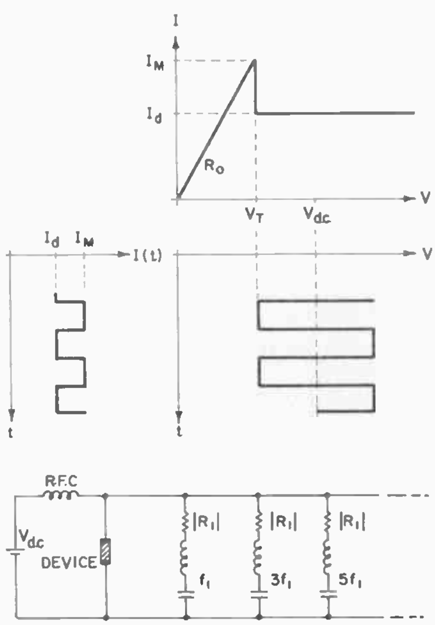


Fig. 3. Circuit and terminal current-voltage curves for q.m.d. operation with square-wave terminal voltage.

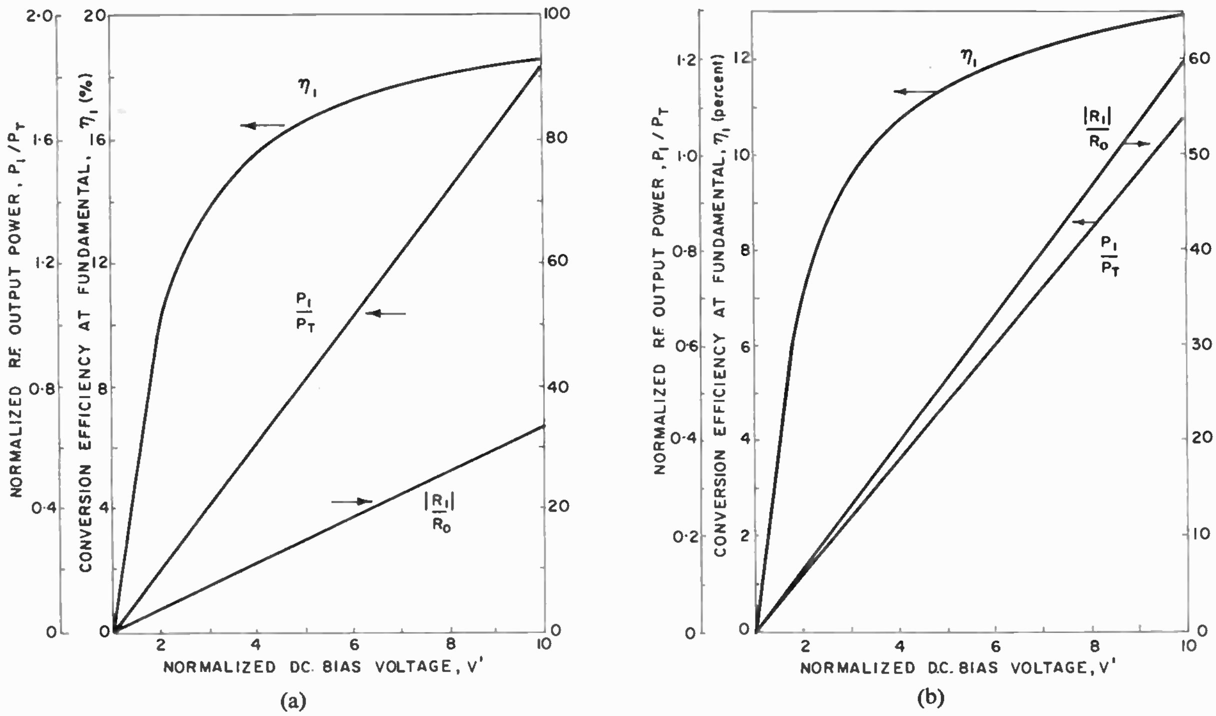


Fig. 4. Performance characteristics for q.m.d. oscillation with square-wave terminal voltage: (a) $I'_a = 0{\cdot}5$, (b) $I'_a = 0{\cdot}7$.

4. References

1. Carroll, J. E., 'Mechanisms in Gunn effect microwave oscillators', *The Radio and Electronic Engineer*, **32**, pp. 17–30, July 1967.
2. Barber, M. R., 'High-power quenched Gunn oscillators', *Proc. Inst. Elect. Electronics Engrs*, **56**, pp. 752–3, April 1968.
3. Thim, H. W., 'Computer study of bulk GaAs devices with random one-dimensional doping fluctuations', *J. Appl. Phys.*, **39**, pp. 3897–904, July 1968.
4. Thim, H. W. and Barber, M. R., 'Observation of multiple high-field domains in n-GaAs', *Proc. Inst. Elect. Electronics Engrs*, **56**, pp. 110–1, January 1968.
5. Harrison, R. I., Denker, S. P. and Hadley, M. L., 'Characteristic ranges for LSA oscillation', *Trans. I.E.E.E. on Electron Devices*, **ED-15**, October 1968.
6. Warner, F. L., 'Extension of the Gunn-effect theory given by Robson and Mahrous', *Electronics Letters*, **2**, pp. 260–1, July 1966.

Manuscript first received by the Institution on 19th August 1968 and in final form on 31st October 1968. (*Paper No. 1232/CC32.*)

Letters to the Editor

Experiences in Tropospheric Scatter Propagation Experiments

SIR,

Widespread use is made of radio communication by the tropospheric scatter mode of propagation although a detailed understanding of the scatter process is far from complete. Indeed, communication engineers depend heavily on propagation loss calculations using semi-empirical formulae relating operating frequency, antenna dimensions, link geometry and (some times) meteorological environment of the link.[1]

The prediction accuracy for annual median received signal level and seasonal variations in level is often inadequate. Where static terminals are to be installed the predictions are often refined by preliminary propagation surveys in an attempt to ensure economic achievement of the desired link performance.[2] This refinement (which is costly and time-consuming) is impracticable for military communicators who need to move their terminal equipments frequently.

Data from which prediction formulae can be developed are extremely limited in quantity, scope of climatic variation and quality. For example, Larsen[1] found data on 15 links only which were suitable for comparing various prediction methods. On the other hand, Gunther[3] has listed 94 world-wide systems (367 links) which had been installed (or were about to be) in 1966 and few of these have been used to provide published propagation information.

Setting up tropo links solely for the purpose of collecting long-term propagation data is extravagant in cost, time and personnel required. We suggest that much-needed information can be obtained economically by using existing operational tropo links. The measurements reported by Fitzsimons[4] were produced in this way. The receiver terminal was provided with a simple analogue signal level recorder which could be serviced easily and the record annotated with timing information as well as equipment malfunctions which might otherwise have invalidated the results. These tasks (together with occasional signal level calibrations) were performed by relatively unskilled operators without impairment to the operational role of the link and comprised a trivial addition to their normal operating duties. Indeed we believe that the operational efficiency of the link was improved because the recordings sometimes gave warning of incipient failures so that remedial action could be taken in time. Occasional visits were paid to this overseas link to check the validity of the recordings but the majority of the highly-skilled effort was concentrated on analysis of the recorded data at the home station. We are now setting up a similar experiment on another overseas link and have had no difficulty in securing the ready co-operation of the managers of this link as well as those responsible for the meteorological services in the area.

We hope our experiences will encourage other workers in the field to seek the co-operation of existing tropo-link operators in mounting and *publishing* the results of similar experiments. It would be particularly valuable if planners of such experiments observed certain criteria for the reported propagation data:

(a) Signal to be measured on a single receiver channel before combining and detection.

(b) Any record of signal level should be continuous for not less than one month.

(c) Numerical sampling of the signal record, for the purposes of calculating median values should be at least every six hours.

(d) Details of operating frequency, antennae dimensions and link geometry (including changes during the test period) should be reported.

(e) Information on the climate should, wherever possible, include monthly or yearly median values of surface radio refractivity and the radio refractivity gradient at the height of the common volume, obtained from the nearest radio sonde.

Boithias[5] has published similar criteria to be observed, when reporting on aerial gain loss, so that experimental results can be compared effectively.

The foregoing recommendations have been devised to cover the case where there is access to an operational link. One would, of course, seek more data of greater detail where a link exists for experimental purposes or where considerable effort is available for the analysis of the experimental data. The following information might then be sought:[6]

(f) Hourly median or 10–15 minute average values of signal.

(g) Complete coded sonde data.

(h) Details of inversion heights.

(k) Details of constant-pressure surfaces.

(l) Changes in the common volume.

(m) Identification of large scale weather changes (frontal passages, monsoon effects, etc.).

A. M. J. MITCHELL, B.SC., A.INST.P.
T. K. FITZSIMONS, BS.C.

Ministry of Technology,
Signals Research and Development Establishment,
Christchurch, Hampshire. *4th December 1968.*

SIR,

The initiative of Mitchell and Fitzsimons is to be heartily welcomed. This problem was highlighted in the Joint I.E.E.-I.E.R.E. -I.E.E.E. Conference on Tropospheric Propagation in September, where Larsen[1] indicated the prediction uncertainties remaining with current calculation methods and also the paucity of entirely adequate pub-

(*continued on page 21*)

U.D.C. 621.319.43 : 666.29

The Design, Construction and Performance of Thick Film Cermet Trimming Potentiometers

By

B. S. METHVEN, B.Eng.†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E.–I.S.H.M. Conference on 'Thick Film Technology' held at Imperial College, London, on 8th to 9th April 1968.

Summary: The cermet trimming potentiometer is a miniature light-weight control having an extremely rugged construction. The nature of the resistive element provides an inherent freedom from catastrophic failure and offers a wide range of possible ohmic values. The design and its influence on some of the more fundamental properties of this control is discussed. The effects of both electrical and environmental stresses liable to be imposed in practice are given in some detail. This form of control is shown to have a number of advantages over the more conventional wire-wound device, some of which can play an important role in present-day high-speed, high-reliability electronic circuits.

1. Introduction

Prior to the introduction of thick film cermet (ceramic metal glaze) trimming potentiometers, the electronics industry relied mainly on wire-wound, conductive plastic and carbon composition devices. The wire-wound trimmer, though having generally excellent performance characteristics, is known to suffer from a number of limitations that include a small ohmic range, limited resolution and high end-resistance. Conductive plastic and carbon composition types on the other hand do not suffer these limitations but their overall performance leaves much to be desired for precision and semi-precision applications. With the advent of cermet technology it was recognized that a new range of trimming potentiometers could be designed which combined the majority of the desirable features of other types without introducing any new deficiencies.

2. Design Considerations

A trimming potentiometer resistance element is one of the more complicated resistive devices to design and manufacture. The specification of this element contains not only those parameters normally associated with fixed resistor design but also a number of important requirements that are peculiar to trimmer design. For a cermet element these requirements place severe restrictions on the choice of inks and manufacturing methods. Film dimensions directly influence trimmer performance with respect to such parameters as electrical travel, end-resistance, linearity, etc., as well as ohmic value. As a result, strict dimensional control is necessary in order to obtain consistent performance over the whole ohmic range. The condition of the film surface is also important in that it controls such parameters as rotational noise, contact resistance and rotational life, which necessitates a surface that is both hard and uniformly smooth; hence conductor particle size and dispersion in the glass matrix are important.

Mechanical design problems are in general similar to those encountered in most trimmer designs with the exception that all piece-parts are required to withstand very high temperatures (e.g. up to 200°C) in order to make use of the full high operating temperature capabilities of cermet. The wiper is also required to have a multi-finger contact arrangement to reduce contact resistance and noise to a minimum.

3. Manufacture and Construction

The main requirements of the substrate are a high thermal conductivity, a thermal expansion commensurate with film and case materials and consistent manufacture to within a few thousandths of an inch of the nominal dimensions. Steatite was selected as being the most suitable substrate. During manufacture, precious metal pins are inserted for contact to the film and external terminations. The ink system used is formed from a mixture of glass and precious metals including palladium, gold, silver and other metals in the platinum group. In contrast to most other systems, the metals are first mixed with the glass powder in the organic state. This liquid phase ensures the proper dispersion of metals in the glass. After heat treatment at approximately 500°C to convert the metals to the inorganic state, the mixture is fired at temperatures up to 1000°C for several hours to obtain the required resistivity. After firing, the mixture is milled to a very fine consistency and then suspended

† Morganite Resistors Ltd., Jarrow, Co. Durham.

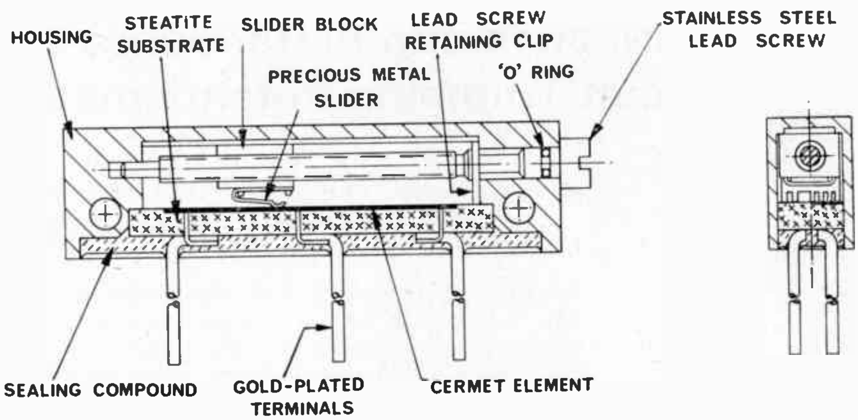


Fig. 1. Constructional details of cermet linear motion trimming potentiometer.

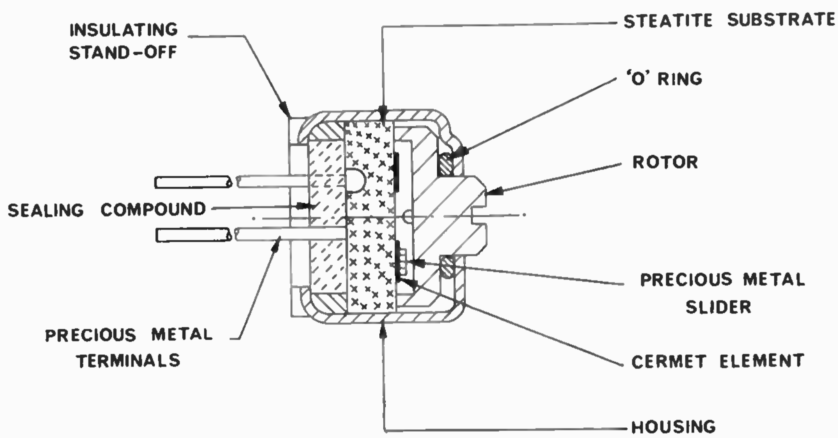


Fig. 2. Constructional details of cermet rotary trimming potentiometer.

in a liquid carrier. A standard silk-screening technique is used for printing and is followed by a short firing cycle in a programmed tunnel furnace at temperatures up to 1100°C. Atmospheric control is not as critical as in the simpler palladium silver/glass systems since oxidation of the palladium is part of the reaction to achieve the required resistivity. End terminations of platinum/gold or silver are printed and fired in a similar manner at a previous stage to complete the resistive element. At this or at a separate stage, a metal track is deposited that is effectively parallel to the resistance track. This is used to make contact through the wiper to this resistive track.

Two important points arise in this method of manufacture.

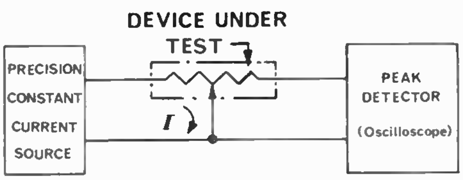
(i) The same screen pattern is used for all resistance values on identical models. This ensures that the dimensions and position of the track are always the same, hence maintaining continuity of those performance parameters which are dependent on dimensions.

(ii) The resistance value is controlled predominantly by variation in the choice of metals and not by major changes in metal content. As a result, the thermal conductivity of the film remains substantially unaltered over the whole ohmic range which in turn ensures continuity of wattage rating and other parameters. Some variation in temperature coefficient (t.c.) does however occur in that the sign of the t.c. changes from positive at the very low values to negative for very high values.

The element can now be installed into its appropriate housing containing the adjusting device and wiper assembly. Figures 1 and 2 illustrate the construction of a linear motion and rotary trimmer respectively. In the manufacture of any multi-component device it is necessary to ensure that no one material used can degrade performance. All the component parts used have been chosen to have high operating temperature capabilities and to be minimally affected by moisture. The housing material of the linear motion trimmer, for example, is diallylphthalate which is dimensionally stable up to 200°C, the slider block is p.t.f.e. and all metal parts either stainless steel or precious metal alloys. Sealing is achieved by a silicone rubber O-ring for the lead screw and a silicone and/or epoxy sealing compound for the element.

4. Performance Capabilities

The use of a cermet resistance element for the manufacture of trimming potentiometers has proven to be successful in that its performance is adequate for most precision applications. Apart from the obvious capabilities of improved resolution compared to wire-wound devices the more outstanding capabilities of cermet are immunity to catastrophic failure, high temperature and wattage ratings, chemical and electrical resistance to water vapour and long life during which it experiences little degradation in functions such as total resistance rotational noise, electrical travel and linearity.



Input $Z = 10 \times R/T$
Bandwidth 100 Hz to 50 kHz

Fig. 3. Noise test set.

Table 1 illustrates the electrical ratings of a typical linear motion trimmer and is followed by a discussion of some of the more important performance characteristics of such a trimmer.

Table 1

Typical electrical ratings of a cermet linear motion trimming potentiometer

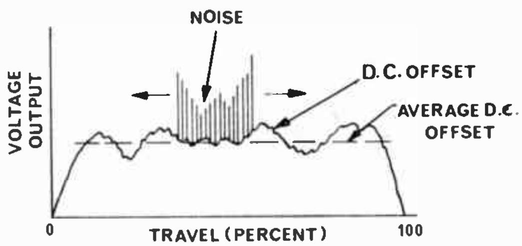
Resistance range	10 Ω to 2 MΩ
Tolerance	±10% and ±20%
Ambient temperature range	−65°C to +175°C
Power dissipation	1 W at 85°C derating to zero at 175°C
Maximum working voltage	200 V d.c. or r.m.s.
Voltage proof	900 V r.m.s.
Insulation resistance	1000 MΩ
Maximum slider current	$\sqrt{1/R_t}$ or 100 mA, whichever is lesser
End-resistance	Less than 2 Ω for all values
Voltage coefficient	Typically 0·02% per volt.

4.1. *Temperature Coefficient*

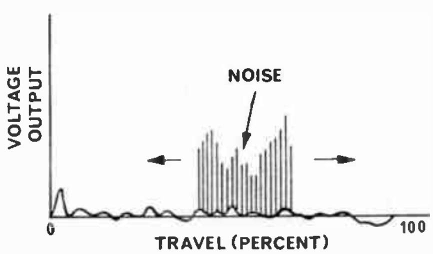
The temperature coefficient of cermet trimmers varies from a positive value for low ohmic values to a negative one for high values. This is caused by a change in the conduction mechanism, low-value resistance elements are predominantly metallic and therefore tend to have a positive t.c., whereas high values closely resemble a semiconductor as far as the conduction mechanism is concerned and therefore have a negative t.c. The t.c.s obtained by natural yield from the process are less than $+500$ parts/10^6/degC for values below 500 Ω and -250 parts/10^6/degC for values above 50 kΩ. Over the majority of the range the t.c. is typically within ± 100 parts/10^6.

4.2. *Rotational Noise and Resolution*

The method normally employed to measure rotational noise for trimming potentiometers involves the measurement of the variation of the contact resistance along the length of the track. The test set-up used for cermet trimmers, shown in Fig. 3, differs from that used for wire-wound trimmers in that the detector is a.c.-coupled to the device. If direct coupling were used, the output waveform observed on the oscilloscope would be as shown in Fig. 4(a). This shows the presence of a small d.c. offset and its low-frequency variations, which are repeatable, together with the noise output. It must be emphasized that the d.c. variations occur only as a result of testing the trimmer potentiometer as a rheostat. A true noise measurement should exclude any repeatable output and for this purpose the detector is a.c.-coupled to the trimmer giving the output waveform shown in Fig. 4(b). The



(a) Measurement under d.c. conditions.



(b) True noise measurement.

Fig. 4.

noise is given in terms of equivalent noise resistance (e.n.r.) which is the peak variation in contact resistance expressed as a percentage of the total resistance of the trimmer.

The typical e.n.r. of cermet potentiometers is less than 1% for all values in the range of 10 Ω to 2 MΩ. The d.c. offset and its variation are of no consequence for the majority of potentiometer or rheostat applications since they are maintained at a very low level. The average d.c. offset is less than 0·5% of R_t and its variations less than 0·2%. This has been achieved by the high particle dispersion and small granularity of particles used in the ink manufacture. Particle granularity is also the limiting factor controlling resolution and theoretically negates the claim of infinite resolution. However, with practical measuring equipment and accuracies the resolution of cermet devices has been found impossible to measure. The effective resolution obtained in practice is found to depend entirely on operator capabilities.

4.3. *Power Rating and Load Performance*

The power rating of a resistive element is normally limited by its permissible operating temperature. To obtain long life and high stability a component is often derated resulting in a poor power handling capacity with relation to its physical size. The quoted power dissipations and operating temperatures for

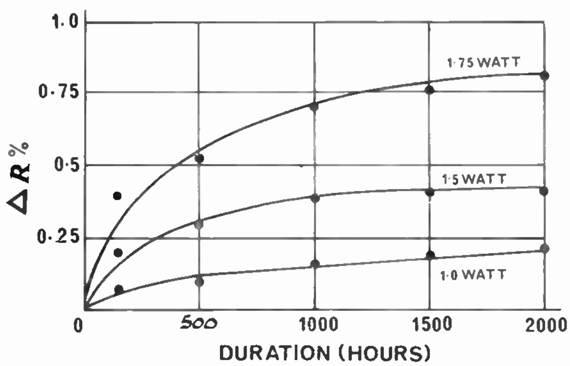


Fig. 5. Load performance at 85°C of a cermet trimming potentiometer normally rated at 1 watt.

cermet trimmers are the highest normally called for by circuit engineers. Their performance at these ratings is such that they may still be classified as high-stability devices. This is due to the capability of the cermet element to operate at temperatures far in excess of the maximum required (175°C) and in consequence the cerment element has been derated by the manufacturers. This derating is necessary to limit the maximum body temperature of the control to a level that is accepted by circuit designers and does not overstress housing materials, etc. Figure 5 shows the normal load and overload performance test results obtained from an evaluation of a 1¼ in linear motion trimmer rated at 1 W at 85°C. The graph illustrates the conservative power rating of this device and shows that even operating at one and a half times maximum power, the rate of change of resistance has nearly decreased to zero after 1500 hours indicating that very little further change should take place throughout the devices' lifetime.

4.4. *Rotational Life*

The expected rotational life of a trimming potentiometer as judged from the various specification requirements is in the range of 200 to 500 cycles. The hard smooth surface that is obtained with a cermet ensures that track wear is kept to a minimum. Table 2 shows the typical performance that is obtained for up to 1000 cycles with a linear motion trimmer.

Table 2

Rotational life of a cermet trimming potentiometer under full load conditions

Measured parameter	Number of operational cycles		
	200	600	1000
Total resistance change	0·02%	0·02%	0·5%

4.5. *Environmental Stability*

The environmental operating conditions to which a device is subjected can often seriously reduce its life. Both the cermet film and steatite substrate are capable

Table 3

Environmental performance of a cermet trimmer potentiometer

Test	Parameter	Performance
Moisture resistance	Total resistance change	0·10%
	Final rotational noise	0·5% of R_t
Vibration	Total resistance change	0·05%
	Setting stability change	0·05%
Thermal shock	Total resistance change	0·15%
	Setting stability change	0·01%

of operating at temperatures in excess of 200°C and are substantially impervious to moisture. As mentioned previously all other component parts have been selected so as not to degrade the performance of the element. A sealed housing is provided for both military and commercial models to exclude dirt from entering and to provide overall protection from moisture. Tests conducted on controls to establish their performance with respect to moisture resistance, vibration, etc., have repeatedly shown that very little degradation takes place in device performance indicating that a long, trouble-free life can be expected. Table 3 shows the typical performance achieved by a linear motion trimmer under the influence of various stresses. All tests were carried out according to the requirements of MIL 22097C Characteristic C on a full range of ohmic values.

4.6. *Reliability*

The evaluation of a new device with respect to its expected life and reliability creates problems for both manufacturers and circuit engineers. Extended life tests are costly, time-consuming and normally give data for a set of semi-ideal operating conditions. In an attempt to simplify lifetime evaluation, the concept of built-in reliability was used in the design and manufacture of the present range of trimming potentiometers. The performance data given in the preceding sections show that these devices exceed the requirements necessary for most applications and the requirements of the appropriate MIL specification indicating a high potential for a long trouble-free life.

5. References

1. Bainbridge, P. L., 'Cermet trimming potentiometers', *Electronic Components*, 9, pp. 671–5, June 1967.
2. Schneider, S. and Silverman, D., 'Testing non-wirewound potentiometers for resolution and noise', *Electro-Technology*, **76**, pp. 38–41, October 1965.

Manuscript received by the Institution on 7th February 1968. (Paper No. 1233/CC33.)

(*continued from page 16*)

lished results, Anderson[7] indicated the cost of increasing system gain and Boithias[8] presented a new calculation method which everyone concerned must be eager to see tested on an adequate variety of paths.

Mitchell and Fitzsimons give a modest list of the basic data required and it seems worthwhile to emphasize the value of this limited information. Might there not be paths for which various valid reasons make the publication of a full paper inappropriate but for which these essentials could be published in a brief note?

G. C. Rider, B.Sc.

Tropospheric Section,
Propagation Group,
Great Baddow Laboratories,
The Marconi Company Ltd.,
Great Baddow, Chelmsford, Essex. *31st December 1968.*

[*Editorial Note.* The Institution's Paper Committee will gladly co-operate in considering for speedy publication in *The Radio and Electronic Engineer* information on tropospheric propagation either in the form of papers or as brief notes as suggested by Mr. Rider.]

References

1. Larsen, R., 'A comparison of some troposcatter prediction methods', Conference on Tropospheric Wave Propagation, London, October 1968. (I.E.E. Conference Publication No. 48, pp. 110–7).
2. Anderson, E. W., 'Tropospheric scatter surveys', *Point to Point Telecommunications*, **12**, No. 7, pp. 4–14, January 1968.
3. Gunther, F. A., 'Tropospheric scatter communications, past, present and future'. *I.E.E.E. Spectrum*, **2**, pp. 79–100, September 1966.
4. Fitzsimons, T. K., 'Observations of tropospheric scatter path loss at C-band frequencies and meteorological conditions in Cyprus', *Proc. Instn Elect. Engrs*, **115**, No. 7, pp. 898–902, July 1968.
5. Boithias, L., 'An experimental study of antenna gain loss', *Trans. I.E.E.E. on Antennas and Propagation*, **AP-15**, No. 4, pp. 578–9, July 1967.
6. Samson, C. A., E.S.S.A. Research Laboratories, Boulder, Colorado. Private Communication. 7th November 1968.
7. Anderson, E. W., 'The economics of propagation surveys for tropospheric scatter systems', Conference on Tropospheric Wave Propagation, London, October 1968. (I.E.E. Conference Publication No. 48, pp. 118–124.
8. Boithias, L., 'Calculations of attenuations in over-the-horizon propagation based on radiometeorological parameters', *ibid.*, pp. 238–246.

U.D.C. 621.374.5

Zero Quiescent Current Monostable Multivibrator

By

C. F. HO, M.Sc.(Eng.),
C.Eng., M.I.E.R.E.†

Summary: A simple monostable multivibrator circuit using a complementary pair of transistors is described. In comparison with the conventional configuration, this circuit has such features that it would need no standing power and a smaller number of components. The common-emitter current amplification factor, h_{fe}, of the transistors may be made as the overriding factor for the pulse width control and thus, enables a small combination of C-R to be used for long time-delay. However, the sensitivity of the pulse width on the transistor current amplification factor, h_{fe}, may be reduced, if desired, in the improved circuit.

1. Introduction

A survey of the literature on the design of the transistor monostable multivibrator reveals the fact that the circuit is generally operated under the following conditions:

The biasing network is so arranged that one transistor is held ON and the other is OFF all the time.

When an input pulse is applied, the ON transistor is turned OFF.

The OFF transistor becomes active and charges up the timing capacitor.

Timing is determined by the discharging rate of the combination of a C-R network.

When the charge on the capacitor is reduced to the 'cut-in' voltage level of the initially ON transistor, then the circuit switches back.

It will be noted that whether the circuit is in operation or not, one transistor is always ON. Since most switching circuits are generally operated in saturation mode, the quiescent current is quite appreciable. In the design of circuits for aerospace systems, minimum power dissipation and least number of components are prime objectives. The monostable multivibrator circuit proposed in this paper draws power only when it works, and in its simplest form this circuit operates with only two resistors and one capacitor.

2. Circuit Description

Figure 1 shows the basic circuit of the monostable multivibrator. Normally, both transistors TR1 and TR2 are inactive because there is no biasing network to cause them to function. If a positive trigger pulse is applied to the base of TR1 through the coupling network C_1, R_1 and the diode D, both transistors will start to conduct. These transistors will saturate rapidly since their bases and collectors are cross-coupled and regeneration takes place.

† Department of Electrical Engineering, University of Hong Kong.

As TR2 saturates, the capacitor C begins to charge. The base voltage of TR2 decreases exponentially and TR2 comes out of saturation. Again, the action is regenerative. Both transistors are switched back to OFF state.

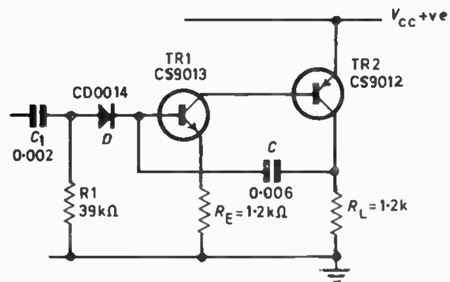


Fig. 1. Basic monostable multivibrator circuit.

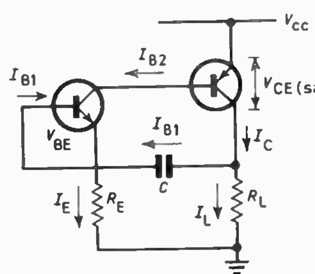


Fig. 2. Circuit for analysis of conduction period.

3. Conduction Period

The conduction period of the monostable circuit can be evaluated by referring to the direction of current flow in the circuit shown in Fig. 2. The voltage across the capacitor C is given by the following equation:

$$V_C = [V_{cc} - V_{CE(sat)}][1 - e^{-t/CR}] \qquad \ldots\ldots(1)$$

From the circuit, the final voltage on the capacitor C can be seen to be:

$$V_C = V_{cc} - V_{CE(sat)} - V_{BE} - I_E R_E \qquad \ldots\ldots(2)$$

But

$$I_E = I_{B1} + I_{B2}$$
$$= I_{B1}(1 + h_{fe1}) \qquad \ldots\ldots(3)$$

and

$$I_{B2} = \frac{I_C}{h_{fe2}}$$
$$= \frac{I_{B1} + I_L}{h_{fe2}}$$
$$= \frac{I_{B1}}{h_{fe2}} + \frac{V_{cc} - V_{CE(sat)}}{h_{fe2} R_L}$$
$$= h_{fe1} I_{B1} \qquad \ldots\ldots(4)$$

Rearrangement of eqn. (4) gives

$$I_{B1} = \frac{V_{cc} - V_{CE(sat)}}{(h_{fe1} h_{fe2} - 1) R_L} \qquad \ldots\ldots(5)$$

Substituting eqn. (5) in eqn. (3) we get

$$I_E = \frac{[V_{cc} - V_{CE(sat)}][1 + h_{fe1}]}{(h_{fe1} h_{fe2} - 1) R_L} \qquad \ldots\ldots(6)$$

Now, from the equivalent circuit of the transistors shown in Fig. 3, the effective timing-resistance, R_{eff}, is given by

$$R_{eff} = r_{bb} + r_{be} + R_E(1 + h_{fe1}) \qquad \ldots\ldots(7)$$

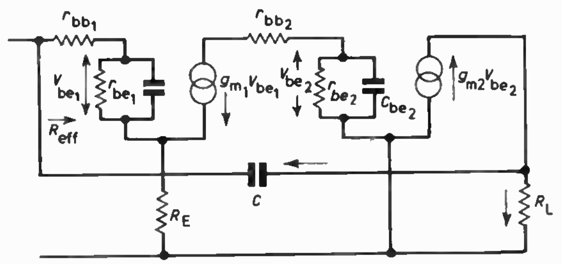


Fig. 3. Equivalent circuit of Fig. 2.

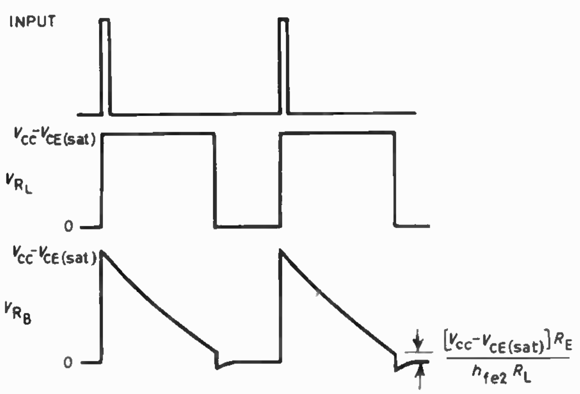


Fig. 4. Output waveforms of the monostable circuit shown in Fig. 2.

Combining eqns. (6), (7) and (1) and assuming that h_{fe} is much greater than unity, the conduction period is expressed by:

$$t = C[r_{bb} + r_{be} + (1 + h_{fe1}) R_E] \times$$
$$\times \ln \left[\frac{V_{cc} - V_{CE(sat)}}{V_{BE_1} + \frac{V_{cc} - V_{CE(sat)}}{h_{fe2}} \left(\frac{R_E}{R_L} \right)} \right] \qquad \ldots\ldots(8)$$

4. Experimental Results

An experiment on the basic circuit was performed using the following parameters:

$V_{cc} = 12$ V $\qquad R_L = 1{\cdot}2$ kΩ

$R_E = 1{\cdot}2$ kΩ $\qquad C = 0{\cdot}006$ μF

TR1 = planar silicon transistor, type CS9013

$h_{fe} = 53$ at $I_C = 50$ mA, $V_{CE} = 1$ V

$V_{CE(sat)} = 0{\cdot}2$ V at $I_C = 120$ mA, $I_B = 2$ mA

$V_{BE} = 0{\cdot}7$ V

$h_{ie} = 600\ \Omega$ (approximately)

TR2 = planar silicon transistor, type CS9012

$h_{fe} = 120$ at $I_C = 50$ mA, $V_{CE} = 1$ V

$V_{CE(sat)} = 0{\cdot}2$ V at $I_C = 120$ mA, $I_B = 2$ mA

The output waveforms are shown in Fig. 4. The conduction period as calculated by using eqn. (8) is 1·92 ms while the measured value is 1·9 ms.

5. Improved Circuit

It is seen that from eqn. (8) the pulse width of this circuit may be varied by the common-emitter current amplification factor (β) of the transistors. This property enables a long delay time to be obtained without using a large capacitance. However, it is preferable to arrange the circuit such that the pulse duration may be made to depend on the β-factor of one transistor only. This is done by connecting a resistor from the base of TR1 to ground as shown in Fig. 5. In this case, the effective timing resistance is the parallel combination of R_B and the input resistance, $r_{bb} + r_{be} + (1 + h_{fe1}) R_E$, of TR1. The input resistance of TR1 can be made large so that the effective resistance, R, approximates to R_B.

Analysis of the circuit of Fig. 5 shows that the conduction period of the circuit is given by:

$$t = C R_B \ln \left[\frac{V_{cc} - V_{CE(sat)}}{V_{BE_1} + \frac{V_{BE_1}}{R_B h_{fe2} - R_E} + \frac{V_{cc} - V_{CE(sat)}}{R_B h_{fe2} - R_E} \frac{R_B R_E}{R_L}} \right]$$
$$\simeq C R_B \ln \left[\frac{V_{cc} - V_{CE(sat)}}{V_{BE_1} + \frac{V_{cc} - V_{CE(sat)}}{h_{fe2}} \frac{R_E}{R_L}} \right] \qquad \ldots\ldots(9)$$

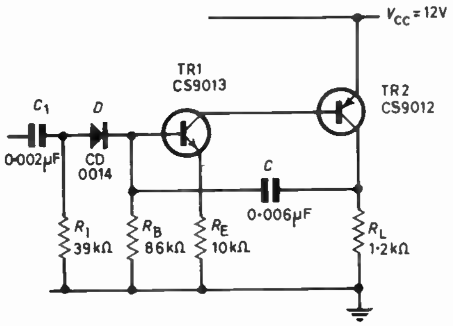


Fig. 5. Improved monostable circuit.

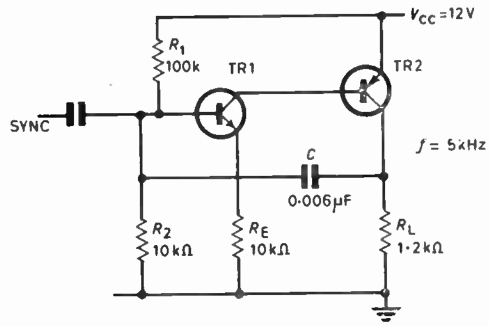


Fig. 6. Low-dissipation astable multivibrator.

With the components values indicated in Fig. 5 and the same transistors, the calculated pulse width is 1·02 ms, whereas the measured value is 1 ms. In both experiments the results obtained agree very well with theory.

6. Trigger Level

It may be suspected that this circuit would be more sensitive to spurious trigger conditions than the conventional arrangement would be. However, this is not so because the circuit will not conduct until the input to the base of TR1 is greater than or equal to

$$V_{BE_1} + \frac{V_{cc} - V_{CE(sat)}}{h_{fe2}} \frac{R_E}{R_L}$$

The trigger voltage for the circuit of Fig. 5 is approximately 2 V. It may be varied by adjusting the ratio R_E/R_L for a given supply voltage, V_{cc}, or by using transistors with different characteristics. This has been verified experimentally.

7. Astable Configuration

The monostable circuit can be converted into an astable one by providing a d.c. bias to the transistor TR1 as shown in Fig. 6. The principle of operation is analogous to that of the monostable multivibrator. The frequency of oscillation is given by:

$$f = \frac{1}{2t} \qquad \ldots\ldots(10)$$

where t is obtained from eqn. (9) with

$$R_B = \frac{R_1 R_2}{R_1 + R_2}$$

This circuit is inherently self-starting and also has the feature of the monostable circuit of having minimum power dissipation. Both transistors are either ON or OFF simultaneously. During the OFF state the only current drain is that flowing through the potential divider R_1 and R_2 which can be made reasonably large to minimize the power dissipation.

8. Conclusion

A monostable multivibrator circuit using a complementary pair of transistors has been described. The outstanding features of this circuit are its simplicity and the fact that it draws power only while it is in operation. The circuit may be easily converted into an astable multivibrator with a minimum of additional components. The rise-time of the output pulse is independent of the time-constant of the timing network. The circuit is useful over a wide range of frequencies if transistors with low values of C_{ob} are used.

9. References

1. Millman, J. and Taub, H., 'Pulse, Digital and Switching Waveforms', Chapter 11 (McGraw Hill, New York, 1965).
2. Joyce, M. V. and Clarke, K. K. 'Transistor Circuit Analysis', Chapter 15 (Addison-Wesley, Reading, Massachusetts, 1961).
3. Wolfendale, E., 'The Junction Transistor and Its Applications', Chapter 9 (Heywood, London, 1958).
4. Hakim, S. S., 'Junction Transistor Circuit Analysis', Chapter 12 (Iliffe, London, 1962).
5. French, R. C., 'A high duty cycle high stability monostable', *Electronic Engng*, **39**, pp. 372–3, June 1967.
6. Kitai, R., 'Emitter coupled monostable multivibrator design', *Electronic Engng*, **38**, pp. 642–7, October 1966.
7. Rakovich, B. D., 'A high-efficiency monostable circuit using complementary transistors', *Electronic Engng*, **38**, pp. 384–7, June 1966.
8. Rakovich, B. D. and Tesic, S. L., 'A 5-MHz switching multivibrator using a complementary pair of transistors', *The Radio and Electronic Engineer*, **35**, pp. 297–305, May 1968.

Manuscript received by the Institution on 12th August 1968. (Contribution No. 110/CC34.)

U.D.C. 621.37/9.002

Active and Passive Components for Attachment to Thick Film Circuits

By

A. F. DYSON, Dipl.Eng.†
C.Eng., M.I.E.R.E.

AND

D. GROSVENOR, B.Sc.†

Reprinted from the Proceedings of the Joint I.E.R.E.–I.E.E.–I.S.H.M. Conference on 'Thick Film Technology' held at Imperial College, London, on 8th to 9th April 1968.

Summary: This paper discusses the development of suitable passive components (resistors and capacitors) for use with thick film circuits. The techniques of attachment of these components to the circuits are described. Experience with the attachment of certain types of semiconductors to thick film circuits is also described.

1. Introduction

The thick film circuit has now developed from a relatively simple resistor network to a complex form in which film resistors and capacitors are combined with discrete passive elements together with sophisticated semiconductor devices. This latter development has been particularly fast and it is useful to study the most recent designs so that the limitations of each element of a hybrid thick film module become more widely understood. This knowledge will enable the designer to specify in detail the most suitable techniques for his requirement.

Progress is such that it is rapidly becoming apparent that using a combination of the thick film circuit and the semiconductor flatpack, an operating circuit, having some superior parameters to an equivalent silicon integrated circuit in the same physical package, will be possible.

2. The Passive Module

The first circuits designed for and manufactured with precious metal thick films were used in large numbers by the computer industry (Fig. 1). They were used in conjunction with a hybrid thick film module which contained discrete transistor devices (colloquially 'flip-chips'). The great majority of these early networks were purely resistive, the resistors of which were designed to the parameters as shown in Table 1. There were two major limitations to this resistor element:

(a) Limited ohmic range.

(b) Difficulty with production yields was found when the number of close tolerance resistors ($\leqq 2\%$) on a single substrate was greater than approximately 12.

Limitation (a) has largely disappeared as the number of thick film resistive paints has increased. This has not only increased the ohmic range to 1 MΩ but has also shown a worthwhile improvement to the temperature coefficient of resistance achieved. Limitation (b) is still a problem. It can be seen that if one reject resistor occurs during manufacture, the entire substrate is rejected as useless, causing the loss of eleven or more perfectly good resistors. To overcome this particular problem and to provide greater design flexibility, a resistor 'chip' was designed, the design and application of which is discussed in detail later in Section 3.

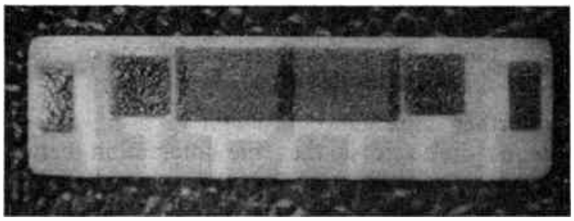


Fig. 1. Standard resistor module.

The initial passive networks to contain capacitive elements used the network substrate as the dielectric. This, of course, has the obvious advantage of a good form-factor and simplicity of manufacture. The important limitations in this approach were:

Table 1

Parameters of resistors and capacitors

	Resistors	Capacitors
Range	50 Ω–50 kΩ	10 pF–10 000 pF
Minimum tolerance	±2%	±1%
Temperature coefficient	±500 parts/10⁶/degC	All temperature-compensated and high-*K* bodies

† Erie Electronics Ltd., South Denes, Great Yarmouth, Norfolk.

(i) Limited capacitance range.

(ii) Interaction between capacitors.

(iii) Stray capacitance between closely screened conductors owing to nature of substrate.

Limitation (i) occurred because the capacitance achieved was dependent on the following factors:

dielectric (substrate) thickness,

effective electrode area, and

effective dielectric constant of substrate.

The effect of these factors was increased by production and economic considerations. For example, for ease of production, it is convenient to confine the thickness of the substrate between the limits 0·020 in (0·5 mm) and 0·030 in (0·75 mm). It is also desirable, for good economics to be achieved, that the electrode areas for each capacitor should be as small as possible. An example of these influences is shown (Fig. 2) in which a customer requested a large value capacitor to be placed on a standard substrate. Although the required value was achieved, it can be seen that the substrate could accommodate hardly any other circuitry and five lead-outs were effectively wasted. It is of course possible by varying the dielectric composition to obtain a quite reasonable spread of dielectric constants but the customer more often than not will specify that the capacitors should have a specific temperature characteristic which, of course, imposes a further limitation.

A further capacitance problem in using the substrate as dielectric is that the interaction between two or more capacitors on the same substrate depends to a large extent on their proximity to one another. Figure 3(a) shows an initial design in which the electrode areas had been calculated theoretically. A sample to this design was manufactured and measured and it was found that the inner pair of capacitors achieved far too high a value due to interactions between themselves and the outer pair. The design was then modified (Fig. 3(b)) resulting in the required values being achieved.

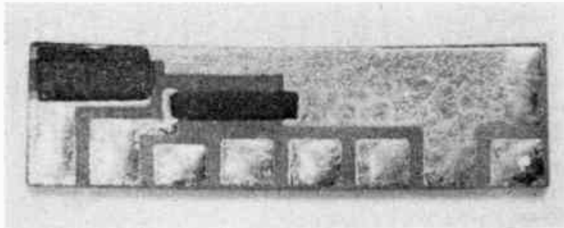


Fig. 2. Large-value capacitor on a standard-size substrate.

To overcome these limitations and to provide greater flexibility to the designer, discrete capacitor elements were developed for use within the thick film circuit package and these are discussed as part of Section 3.

3. Discrete Passive Components for Thick Film Circuits

3.1. *The Chip Resistor*

The design of the chip resistor is shown in Fig. 4. It is manufactured in an identical manner to a standard thick film resistor but has two end clips mounted at either end to facilitate attachment to the substrate.

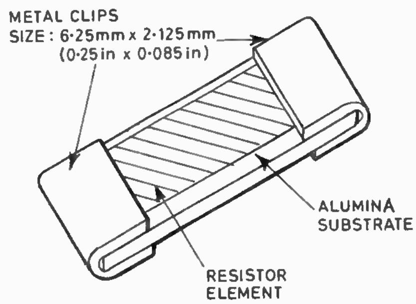
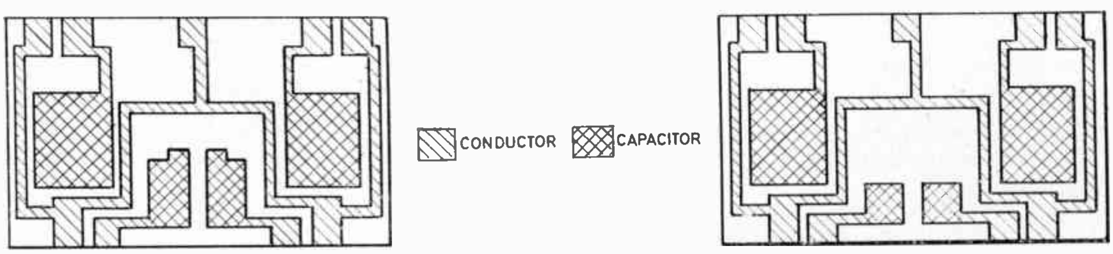


Fig. 4. Chip resistor.

The substrate and the chip resistor are solder-tinned separately and then the resistor is mounted on the substrate using a solder reflow technique. If the chip resistor is to be used to replace a faulty resistor on an existing substrate, the faulty resistor is first removed



(a) Electrode areas as calculated theoretically.

(b) Diagram showing final capacitor areas.

Fig. 3. Substrate with two pairs of capacitors.

by an air-abrasive technique leaving the 'lands' intact. The circuit is then tinned and the chip resistor mounted in position.

3.2. *The Chip Capacitor*

The simplest form of chip capacitor is obtained by dicing available dielectrics currently used in the manufacture of ceramic disk capacitors. This gave an immediate increase in available capacitance in thick film work to approximately 1000 pF in a chip size approximately 0·150 in × 0·050 in (3·8 mm × 1·3 mm) as well as the advantage of a choice of dielectric bodies. The chief disadvantage is that a separate connection has to be made to the top surface of the capacitor after mounting the capacitor on the substrate. The mounting is carried out in an identical manner to the chip resistor. Further capacitance was still needed and this has now been achieved with the use of a ceramic thin film chip capacitor having a thickness of 0·005 in (0·13 mm). This enables capacitor values of up to 10 000 pF to be obtained in a size 0·150 in × 0·150 in (3·8 mm × 3·8 mm). This method still unfortunately requires the use of a separate connection to be made to the top surface

Further increase of capacitance was still required and this coincided with the development of the monolithic ceramic capacitor (monobloc). This capacitor is manufactured from thin ceramic sheets upon which precious metal electrodes are printed. The sheets are assembled into a layered block and fired at a high temperature. In the firing cycle the dielectric and electrodes are fused into a homogeneous block, the resulting capacitor having excellent reliability and environmental performance, a capacitance density of 400 $\mu F/in^3$ and, important from the thick film circuit point of view, connections brought out to each end of the block. By using an external electrode connecting paint such as liquid silver, the chip capacitor may be directly bonded to a thick film substrate (Fig. 5). In the simplest method of attachment connection may be made between the capacitor

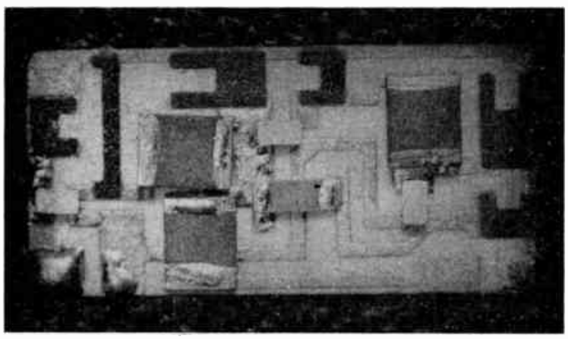


Fig. 5. Module with 'monobloc' capacitors and LID transistors.

and the termination pattern on the substrate with the use of eutectic solder and a miniature soldering iron. Solder reflow techniques, similar to that described for the chip resistor, may also be used.

A further development and one as yet untried in production concerns the use of a high capacitance barrier layer ceramic type of device (Transcap). If a chip capacitor is diced out of a Transcap dielectric in a similar manner to a standard capacitor chip, it is possible to place a further cut in the bottom face of the capacitor as shown in Fig. 6. This does, of course, give a simplified theoretical capacitor of two capacitors in series. This method has the advantage

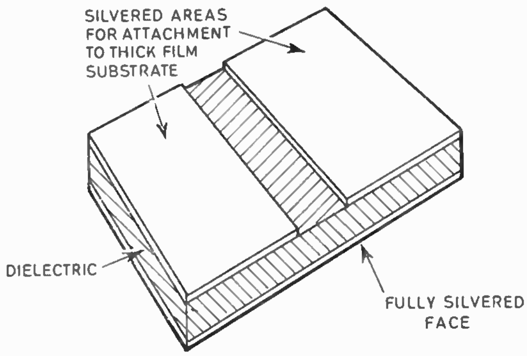


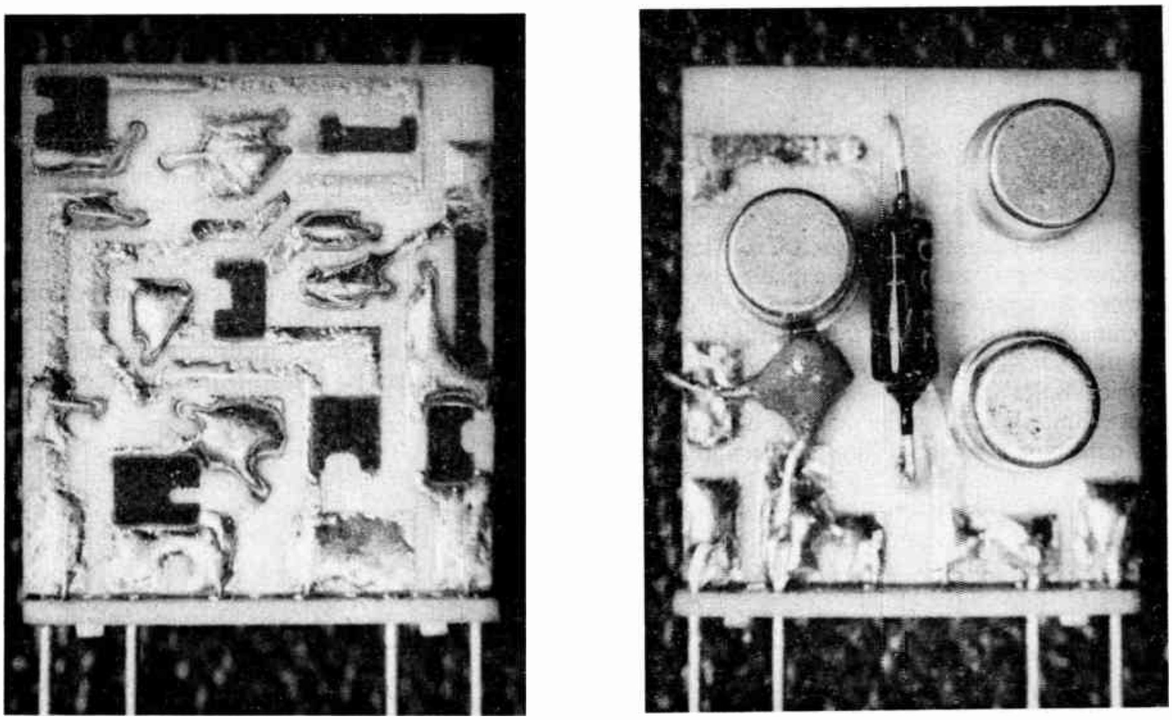
Fig. 6. Barrier layer chip.

that the capacitor chip may be mounted in a similar manner to a monolithic chip, i.e. two direct substrate bonds. It has the disadvantage that the capacitance value will be reduced to half that of the uncut chip. This technique may also be used with standard ceramic capacitor chips. It does not have the same utility of the barrier layer device due to its low initial capacitance.

4. Semiconductor Devices

The simplest method so far used for the attachment of semiconductor devices to thick film networks has been to use standard discrete transistors hand soldered to the substrate. This has the attraction of using a very low priced semiconductor device but the disadvantage is the time and skill required to connect the transistor lead wires to the substrate.

A more elegant but unfortunately more expensive solution, especially where higher powered semiconductors are involved, is to use a pre-drilled substrate (Figs. 7a and b). This provides a good location for the device and an excellent anchorage for the lead wires but still requires a hand soldering



(a) Front face. (b) Reverse face.

Fig. 7. Active module using a pre-drilled substrate.

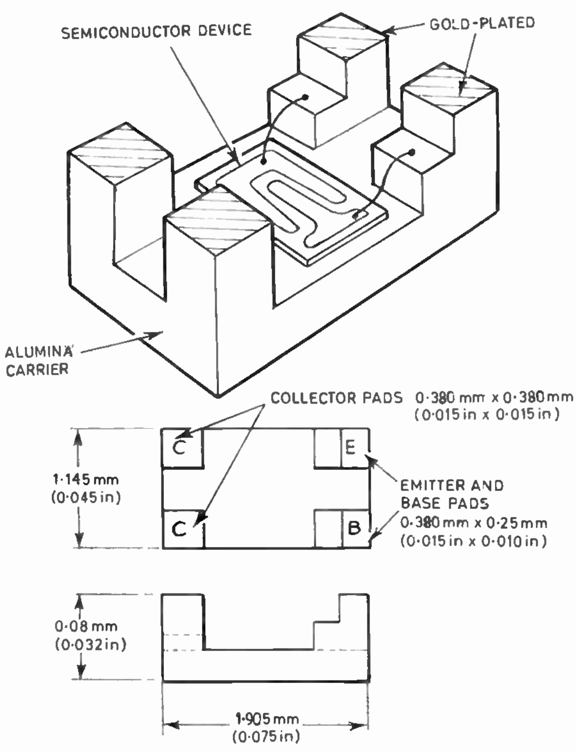


Fig. 8. LID transistor.

process to make the lead connection. Devices specifically designed for thin and thick film work such as leadless inverted devices (LID), are extremely effective. The outline dimensions are shown in Fig. 8 and a photograph of devices used in a typical application is shown in Fig. 5.

The best method of attachment used to date uses solder reflow techniques. The substrate is placed on a temperature-controlled hotplate and the LID is placed in position on the connecting lands. Local heat is applied to each connection in turn at which the solder on the substrate land flows and wets the LID. The local heat is then removed and the bond solidifies. Utilizing this technique it has been found possible to attach and remove LID transistors to and from thick film substrates without deterioration of either the thick film termination pattern or the transistor parameters.

The most important advance in semiconductor packaging which is applicable to thick film circuitry is the availability of the flatpack. It is possible to obtain in this package combinations of diodes and transistors suitable for the particular applications. The package itself is most suitable for solder reflow techniques and has overall dimensions most suitable for thick film work.

The thick film industry has borrowed from the semiconductor industry the concept of the dual-in-line (d.i.l.) package and it is possible, with a suitable combination of thick film resistors, chip capacitors and suitable semiconductor devices now available, to achieve a complete operating network such as video amplifiers in a d.i.l. package. This has advantages over its semiconductor counterpart—the silicon integrated circuit—of being complete, i.e. feedback resistors and large value capacitors are included and that stage isolation is easily achieved. These advantages are, of course, of much greater importance in linear amplifiers than in digital circuits where it is possible that the silicon integrated circuit may have an economic advantage. The strength of the hybrid thick film circuit may well lie in the linear amplifier field.

5. Conclusion

The design and manufacture of thick film conductors and resistors is now an established art. An extension of this is the availability of a thick film resistor as a discrete element. The design and choice of capacitive elements to be used on thick film circuits requires considerable care if the best economics and performance is desired but apart from the substrate dielectric-based capacitor there is available in chip form a wide choice of values and temperature characteristics in several different mechanical configurations.

There is also now considerable choice in the method of attaching an active semiconductor to a thick film substrate which is going to enable the hybrid thick film circuit to compete on at least equal terms with the silicon integrated circuit.

Manuscript received by the Institution on 9th February 1968. (Paper No. 1234/CC35.)

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

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

December 1968	24-hour mean centred on 0300 U.T.			December 1968	24-hour mean centred on 0300 U.T.		
	GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz		GBR 16 kHz	MSF 60 kHz	Droitwich 200 kHz
1	− 300·1	− 0·1	0	17	—	− 0·1	− 0·1
2	− 300·1	− 0·1	0	18	− 300·1	− 0·1	− 0·1
3	− 300·1	− 0·1	0	19	− 300·1	0	0
4	− 300·1	0	0	20	− 300·1	0	0
5	− 300·1	0	0	21	− 300·1	− 0·1	0
6	− 300·1	− 0·1	0	22	− 300·2	− 0·1	0
7	− 300·1	0	0	23	− 300·1	− 0·2	0
8	− 300·1	0	0	24	− 300·1	− 0·1	0
9	− 300·1	− 0·2	0	25	− 299·9	0	0
10	− 300·0	0	0	26	− 300·1	− 0·1	+ 1
11	− 300·0	0	0	27	− 300·1	0	+ 0·1
12	− 300.1	0	0	28	− 300·2	− 0·1	0
13	− 300·1	− 0.2	0	29	− 300·1	− 0·2	0
14	− 300·1	− 0·1	0	30	− 300·1	− 0·2	0
15	− 300·1	0	− 0.1	31	− 300·2	− 0·2	0
16	− 300·1	− 0·1	− 0·1				

The fractional frequency offset of co-ordinated transmissions for 1969 will be -300×10^{-10}.

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.

All measurements are made in terms of H.P. Caesium Standard No. 334, which agrees with the N.P.L. Caesium Standard to 1 part in 10^{11}.

U.D.C. 681·52

Control: Past, Present and Future

By

Professor

H. H. ROSENBROCK,

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

An Address given at the Inaugural Meeting of the Institution's Instrumentation and Control Group in London on 19th March 1958.

Summary: Automatic control developed historically as a necessary consequence of replacing human or animal power by wind and water power and later by steam power. Its beginnings were empirical, but an important contribution to the theory was made by Clerk Maxwell, and later the development of feedback amplifiers led to intensive theoretical work in the U.S.A. By 1945 a well-developed and highly successful theory was available, and wide applications have been made.

In more recent times, theoretical effort has been concentrated on problems of guidance, using the calculus of variations. The incentive here was to solve aerospace problems, but there have been some applications to industrial problems. This theory seems to have a narrower range of applications than the earlier work.

The development of the digital computer has placed in the hands of engineers a tool of unrivalled flexibility and power for the control of processes. By contrast, our theory is at present unable to exploit this power with full effect. The most promising field for control theory and application in the next five to ten years seems to be the application of computer control to all types of industry, and the development of the corresponding theory.

What I would like to do in this Address is to make an assessment of the situation which we have reached in the application of automatic control to industrial processes, and to suggest the directions in which we are likely to proceed. To do this with any success it is first necessary to say something of the way in which the present situation has been reached, so I hope you will excuse some preliminary remarks on the history of control.

Though the rudiments of control can be traced back almost as far as we wish, the true beginnings of the subject as it exists today can be placed in the eighteenth century. Windmills had been used in Europe for several centuries at that time, and the wind is a notoriously uncertain source of power. Among the adjustments that the miller had to make was the adjustment of the gap between the grindstones as the speed of the mill changed. About 1730 the flyball governor was invented and was used to change the gap between the stones as the speed increased. Technically this is not a difficult problem in control, because the system (as we should now say) is open-loop.

Towards the end of the eighteenth century, the flyball governor was borrowed from the windmill and applied to the steam engine. As the load on the engine increased, the speed tended to fall. The flyball device was used to open the steam valve as the speed decreased, so that the speed remained more constant than it would otherwise have done. The speed naturally did not remain entirely constant: if the valve was to open, the flyballs had to drop, and this they could only do if the speed fell somewhat. Nevertheless, the fall in speed could be made much less than it would otherwise have been.

Technically, this is a much more advanced form of control than the previous one: the governor is now being used in a closed-loop system. Consequently the possibilities of misbehaviour of the system are much greater. The device was made to work by practising engineers and was gradually improved. Attempts to improve it to the point where it would keep the speed quite constant regardless of the load were constantly frustrated however by instability: instead of keeping the speed constant, the governor would cause the speed to vary cyclically.

The reasons for this behaviour were not understood, and progress in improving the device was consequently impeded. The explanation was finally provided by Clerk Maxwell in 1868, just one hundred years ago. In his paper to the Royal Society 'On Governors' he put forward a linearized model for the system and showed how attempts to increase the gain of the control system could lead to instability. He

† Control Systems Centre, The University of Manchester Institute of Science and Technology, Sackville Street, Manchester, 1

was able to analyse the second-order system and the third-order system in detail, but was not able to handle the fourth-order system because he did not have available results on the location of the roots of a polynomial. This generalized problem of stability for higher-order systems was proposed later, by a committee of which Maxwell was a member, as the subject for a prize essay at Cambridge. The prize was won by Routh with his now well-known analysis. Independently the same problem was suggested to Hurwitz by the steam-turbine engineer Stodola and was solved in a way essentially equivalent to that of Routh.

There is here an interesting lesson on the relation between mathematics and engineering. Whereas in the physical sciences it is common for new results to be predicted by a mathematical analysis and subsequently confirmed by experiment, the analogous procedure is excessively rare in engineering. Typically, new engineering devices do not arise from a detailed analysis followed by experimental verification. Rather the device is conceived and made to work by an inventor who is close to the application and who works in an intuitive way, guided by existing engineering knowledge. Later the device is gradually improved, but there may come a stage where the existing theory fails to provide sufficient guidance to the engineer's intuition, and progress will then become slow. It is at this stage that the theorist can make his contribution: once the observed complications have been brought into a simple intellectual framework, further progress can again be rapid. It is still likely to be made, however, by engineers who are closely in touch with the application, rather than by the theorist who explained the previous difficulties.

This is illustrated again by the next chapter in the story. Leaving aside some work on ships' steering-gear, the next important development was the invention of the electronic feedback amplifier by Black, in response to a demand for amplifiers of very high quality for the U.S. telephone system. The invention arose from a simple intuitive idea, but its development and refinement again ran into problems of stability. Nyquist's investigation of these difficulties (very close in some ways to that of Routh) is still the foundation of the modern theory.

One of the most fruitful generalizations in all engineering was the recognition that the problem of the feedback amplifier and that of the steam-engine governor are technically the same. At the time when this was recognized, the present-day subject of control can be said to have been born. The subsequent history is well-known: during the Second World War, Nyquist's ideas and those of the many other engineers who contributed to their development were applied to the design of control systems for more and more complicated devices. By the end of the war, there existed a well-developed body of knowledge concerning control systems and their design. The theory underlying this branch of engineering is concerned chiefly with mechanical or electrical systems (or a combination of both), chiefly with single-loop feedback systems, and chiefly with problems of stability and speed of response. It is significant that the theory arose from a clear and urgent demand for solutions to practical problems.

This development, of what we now call the 'classical' theory, or 'servo' theory, is by a long way the oldest. More recently, however, it has been combined with a number of other strands of quite separate origin, so that the present theory of control contains a number of somewhat unlike parts. The strand which has absorbed most effort in the theory of control in the last ten years is what is called optimal control. This again arose from a practical demand—that of computing the best way of controlling the thrust of a rocket to perform a desired task. For example, the rocket might be required to reach a certain point in the least time, or with the least consumption of fuel. The mathematical techniques appropriate to such problems are those of the calculus of variations, and optimal control theory is largely a restatement of this calculus in terms appropriate for engineering purposes: the most famous of those who have contributed to this process is the Russian mathematician, Pontryagin.

A third strand which can be identified is concerned with the control of industrial processes of the type found in the chemical industry—processes of heating, cooling, mass transfer, chemical reaction, etc. Such processes are found in very diverse industries—chemicals, oil, cement, paper, glass, steel-making, food preparation, sugar, etc., etc. An empirical art of controlling such operations grew up from, say, 1920 and by 1940 it was well developed. In contrast with the two other developments which have been mentioned, this field never gave rise to a large body of well-developed theory. Possibly this was because the art was generally successful and did not seem to need such support. More fundamentally, processes of the type mentioned are at once more difficult to describe mathematically, and more stable, than those considered in the 'classical' theory.

As a fourth and last strand we may identify the influence of the electronic digital computer. This was not in its origins intended as a device for control purposes, but it became clear at an early stage that a sufficiently fast and reliable computer would be able to do anything which could be done by the control devices used previously, and would be able to do other desirable things at the same time. Once again the military uses are probably the most advanced, but in 1958 a digital computer was used for the first time in the control of an industrial process. Since then there has been a continuing and increasing use for this purpose, though the acceleration in the number of applications was for a long time less than might have been expected. For this the relative unreliability and high cost of the early computers were largely responsible, together with the fact that the engineering effort involved in a successful application was large and expensive. All of these difficulties are decreasing as computers (and their peripheral equipment) become cheaper and more reliable, and as ways of simplifying the task of programming are developed.

Although one could identify many other topics, these four strands are probably the most important in the present fabric of control as a theoretical and practical subject. The past achievements are sufficiently impressive, but what is likely to be the future? If it is necessary to pick one field as the most likely to show intensive and rapid development in the next five years, my own choice would be the control of industrial processes. By this I do not mean only those 'process' industries which I have mentioned before, but also those concerned with the shaping and assembly of materials: steel-rolling, manufacture and assembly of motor cars, and so on. In all these processes the digital computer gives us a tool of incomparable power, and its influence on industry in general is unlikely to be less than that brought about in an earlier industrial revolution by the use of steam power.

In this development, many of the problems will be special to each industry, and they will be overcome as such problems have always been overcome by the ingenuity of the engineers who meet them. Some of the problems, however, will be more general ones in the area of control. These also will call for the ingenuity of engineers trained in the disciplines of control, and we can expect the development of new devices and of new ways of achieving control. Will these developments call for new theoretical developments? I believe myself that they will. Already there are signs that the full potential of the digital computer cannot at present be exploited for control and this is only partly because we do not fully understand the dynamical behaviour of many of the traditional processes. Partly it is also because we do not have a coherent theory at present which has been developed in response to the particular demands of industrial process control. For some questions the existing theories provide a satisfying answer—problems, for example, of stability or of optimization, though no one will pretend that we know as much as we would like to about these two questions. For other questions, however, present theories give little guidance: examples are, how to choose most effectively what it is necessary to measure and to manipulate in order to control a process effectively; or what limits are set to the goodness of control by interaction between various control actions. In such areas, any guidance which we are able to obtain by the development of theory should be of considerable help to engineers in explaining the properties of large control systems and guiding the inventiveness which will lead to better systems.

It will be clear, I think, that in my view the impact of control on industrial technology has only just begun. The next ten years will be a time of great activity in this subject, and it is therefore much to be welcomed that the Institution has chosen this time to set up a Group to mark its interest in this subject. Control is a subject which needs engineers with every kind of background, not least those whose chief concern is with electronics. The greater the activity in each of the areas which contribute to the application of control, the greater the progress will be.

At the same time, may I make a plea for active collaboration with engineers of all those other disciplines which are concerned in Control? The vigour of the whole is encouraged by the vigour of each component part, but the whole is still greater than the parts. Control as a subject must suffer if it becomes fragmented to the point where communication of ideas becomes difficult, and this could happen either through the growth of many small specialized groups or through continual subdivision of the channels of communication. A continuing effort is needed by all of us if these dangers are to be avoided. Fortunately, this fact is now widely recognized, and I believe that any efforts which the Institution makes in this direction will meet with warm encouragement and corresponding efforts from the other bodies which have an interest in control.

Manuscript received by the Institution on 29th July 1968 (Address No. 41).

U.D.C. 621.376.56:621.395.74

The Impact of Pulse Code Modulation on the Telecommunication Network

By

Professor
K. W. CATTERMOLE
B.Sc., C.Eng., M.I.E.E.†

Reprinted from the Proceedings of the Institution's Convention on 'Electronics in the 1970s', held at Cambridge on 2nd to 5th July 1968.

Summary: The basic processes of transmission and switching are cheaper for multiplex digital signals than for their voice-frequency equivalents, and introduce little or no impairment of the message. Conversion between analogue and digital signals introduces much of the cost and almost all of the impairment in a p.c.m. communication system. The long-term plan for a telephone administration should therefore be to introduce an integrated digital network in which only a minimum number of conversions and reconversions is required: ideally only one per connection. Transmission between exchanges should use multiplex p.c.m. and switching of concentrated traffic should be digital (either analogue or electromechanical). A planned transition from present practices towards a digital network is possible, because the partial use of p.c.m. is technically and economically beneficial provided that the applications be carefully chosen. Junction transmission, trunk transmission, tandem switching, group selection in terminal exchanges and integration over limited areas are considered. A digital network whose main traffic is telephony can also carry other messages such as telegraph and data.

1. Introduction

A new technique of communication poses problems of two kinds: how best to make it work, and how to use it to best advantage. In connection with pulse code modulation (p.c.m.) the first problem has received much more attention. The purpose of this paper is to cast some light on the second problem. For what parts of a communication network is p.c.m. most suitable; what type of p.c.m. equipment is needed to provide such service; how much expense will be saved by using it?

A distinctive property of p.c.m. is this: the more extensively it is used, the more benefits it offers. It is tempting to describe a Utopian vision of an entirely digital communication network having all the advantages and none of the penalties. But the existing network can evolve towards such a state only gradually, and it will not do so unless the intermediate stages are viable: that is, unless they offer adequate service at tolerable cost. Three stages must be considered:

The use of existing p.c.m. equipment.

The use, in isolation, of new equipments which could emerge during the next few years.

The extension towards an integrated network of mutually compatible digital systems.

† Department of Electrical Engineering Science, University of Essex; formerly with Standard Telecommunication Laboratories Ltd., Harlow, Essex.

These stages are not strictly chronological eras to which one could assign dates; they will, in general, overlap even in one locality, and the pace of introduction could vary considerably even between different localities within the same major territory. Moreover, progress depends critically on decisive action by the administration concerned.

It is likely that the digital network will ultimately carry many types of message: telephony, telex, data at several rates, video telephone, facsimile, television. It will offer new facilities such as conference calls, location of mobile subscribers, automatic message accounting. In the earlier stages, however, the main demand will be for telephone service of essentially present-day characteristics; and it is preferable that this type of service be proved viable on its own.

A framework for evolution of a network is set by the existing routing plan, which defines a hierarchy of switching centres and transmission circuits. This may well change in response to the new technology, but on present evidence the changes are likely to be of proportion rather than fundamental topology. A generalized hierarchy is shown in Fig. 1, which is reproduced from the C.C.I.T.T. handbook on National Telephone Networks. Most national networks fit this pattern, with the following variations:

(i) Terminology in the national language may be quite arbitrary. (In this paper, the terms given in the C.C.I.T.T. publication will be used as far as possible.)

(ii) Some of the higher-order centres may not be needed: most developed countries go as far as tertiary, only a few have quaternary centres for national traffic.

(iii) The proportion of final circuits (which follow the hierarchical arrangement, primary–secondary–tertiary, etc.) and of high-usage or direct circuits (which bypass part of the hierarchy) depends on the traffic distribution and switching technology of the territory. Certain types of direct route not shown in the Figure may also exist, for example, from a primary to a tertiary centre.

(iv) Some territories or localities make little or no use of tandem exchanges.

(v) A high-order centre may, or may not, incorporate the functions of lower-order centres.

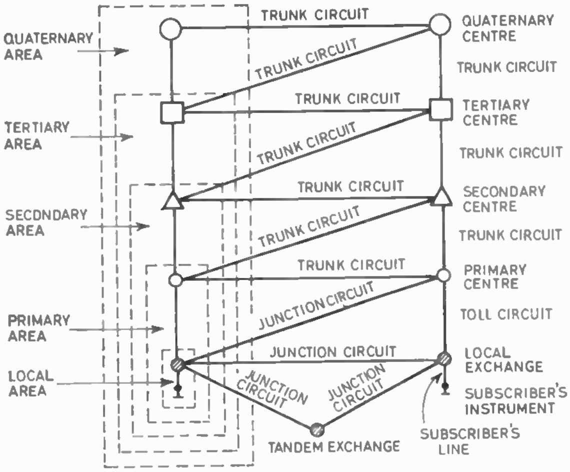


Fig. 1. Generalized telephone network.

At present, p.c.m. transmission systems are used for junction and toll circuits in many countries, as an alternative to voice-frequency cable pairs. Transmission systems for trunk routes (as an alternative to frequency-division-multiplex carrier systems) have been developed in several laboratories but are not in service at the present time. P.c.m. switching is under study by several authorities, especially for tandem centres, and small-scale trials are in progress.

It will be argued in this paper that p.c.m. is technically and economically suitable for virtually all junction, toll and trunk routes requiring more than a few circuits: and for all switching stages carrying concentrated traffic. The full benefits, both in economy and performance, will be achieved only when an extensive digital network comes into being, whose parts are designed for mutual compatibility. However, the benefits at intermediate stages are quite substantial, and the gradual introduction of p.c.m. is a reasonable and feasible policy.

2. Principle of P.C.M.

There are two types of message which may be sent by telecommunication: the *analogue* and the *digital*, of which the classic examples are, respectively, telephone and telegraph. The analogue message is an essentially continuous signal, such as the electrical output of a microphone which corresponds to variations in sound pressure. A digital message is drawn from a finite repertoire, such as the letters of the alphabet, and the resulting electrical signal may consist of discrete elements such as impulses of current.

P.c.m. may be briefly defined as *the conveyance of analogue messages by digital means.*

A major item in any p.c.m. system is the interface equipment which converts analogue signals to digital form, and vice versa. This introduces virtually all the impairment (noise, crosstalk, distortion) suffered by the signal: and it is also a major part of the cost. Once the signals are in digital form, they may be conveyed through transmission lines or switching matrices with negligible impairment, and at a cost which is very low compared with voice-frequency transmission or switching. Analogue transmission, on the other hand, introduces attenuation, noise, crosstalk and distortion which accumulate along the length of the route.

Pulse code modulation was invented in 1938 by A. H. Reeves, then at the Paris Laboratories of I.T. & T. and now at STL. The tools for its economical realization did not exist at that time. With the advent, first of the transistor, then of semiconductor integrated circuits, it has become a practical and highly competitive means of communication.

3. Telephone Transmission on Junction and Toll Cables

The first major commercial use of p.c.m. was for the transmission of telephone channels over junction and toll cables. Equipment for this purpose, introduced by the American Telephone and Telegraph Corporation in 1962, has been spreading rapidly in the United States ever since, and it is only a matter of time before p.c.m. becomes the dominant method. Similar systems have been developed elsewhere, notably in Great Britain, Italy and Japan, whose administrations have each ordered substantial quantities of equipment.

Now the significant property of p.c.m. signals is that, being regenerable, they need not suffer accumulation of impairments as they traverse a lengthy route. Thus it may appear paradoxical that their first extensive usage should be over short distances. This application does, however, follow quite logically from a technical and economic comparison of p.c.m. with other modes of transmission.

Before the introduction of p.c.m., it was usual to employ 2-wire voice frequency (v.f.) cable circuits for junctions, and f.d.m. (frequency-division-multiplex) carrier systems for trunks. Now the v.f. circuits are, except for short distances, inferior to typical carrier channels. The latter have additional modes of impairment, but these are closely controlled: their attenuation is low, and the overall quality is good. The reason why carrier systems are not used on junctions is economic: they are expensive, and much of the cost is in the terminals. On a long route, the cost of the terminals is offset by the low cost per channel-km of multiplex transmission. No junction circuit is long enough to supply this justification.

One major advantage of p.c.m. is that the terminals were, even at an early stage in manufacturing development, cheaper than f.d.m. carrier terminals. Thus p.c.m. could compete with v.f. cable transmission on shorter junction area. The pay-in distance was originally about 20 km, but within the next few years is likely to fall to 10 km as the cost of equipment falls.

A second factor is that this same application can benefit from the distinctive technical potentiality of p.c.m.: the freedom from impairment of digital signals. A junction cable is not a very long nor, for v.f. traffic, an unduly bad transmission path. But it is a poor medium for wideband signals, on which it superimposes a large and variable amount of noise and crosstalk. Digital signals, with the aid of regenerative repeaters, can withstand these impairments much better then f.d.m. or other analogue signals. Hence p.c.m. can be used to convey multiplex groups over cable pairs intended for single v.f. channels, and so increase the capacity of existing cables.

The conventional junction system has 24 speech channels (some authorities prefer 30), each sampled at 8 kHz. The samples are assembled in time-division multiplex and encoded into a 7-bit binary code by a common encoder. An eighth bit is appended to each channel for telephone signalling. The product of these numbers indicates a gross bit rate of 1536 kbit/s, to which may be added further digits for synchronizing the operations of the receiver with the transmitted signal. The quantum intervals are graded in an approximately logarithmic manner, equivalent to companding of about 24 dB, usually by incorporating a non-uniform scale of quanta in the encoder and decoder. Similar apparatus is provided for each direction of transmission, and the two directions are combined at voice-frequency by a hybrid for each channel.

The speech quality of such systems has been investigated by several authorities using such various languages as English, Italian and Japanese. It has been proved adequate over a wide range of talker volume (30 to 40 dB, according to the exact standard demanded) and with up to four or five encoding/decoding operations applied in tandem. Usage beyond these limits is by no means ruled out, since degradation is gradual rather than sudden. This means that such systems are suitable for junction or toll service, in any public network. Most telephone connections introduce not more than two such circuits, and none more than four.

Signals are conveyed over two cable-pairs, one for each direction, which may or may not be within the same cable sheath. Regenerative repeaters are inserted at intervals of about 1830 metres, and supplied with power over the signal pairs. The cable is equalized at frequencies up to about half the digit rate, with a controlled rate of cut-off for another octave. The loss at the half-digit rate is about 20 dB (with conductors of 0·9 mm diameter) to 32 dB (with conductors of 0·64 mm diameter). The dominant limit to repeater spacing, for any given amount of traffic in the cable, is set not so much by loss as by crosstalk between the several pairs carrying p.c.m. The 1830-metre distance is determined mainly by the existence of convenient access points, due to the practice of inserting loading coils in v.f. junctions at this spacing; it is a fortunate coincidence that it is also a feasible repeater spacing for 24-channel p.c.m.

Since junction p.c.m. systems are now being installed between conventional electromechanical exchanges, they must accept the mode of signalling (usually d.c.) already in use. The signalling interface is a substantial part of the equipment both in size and in cost. Moreover, the signalling format of present-day exchanges varies widely, and it is usual to provide several interface options. The signalling apparatus is mounted on a plug-in card for each channel, so that an arbitrary mixture of (say) incoming and outgoing circuits can be accommodated.

4. High-capacity Transmission

Twenty-four-channel transmission would not be economic over long distances (say more than 80 km). Moreover, in large cities and conurbations it is not unusual for some thousands of circuits to be needed over a short distance such as 10–20 km. Thus for general use of p.c.m. a higher-capacity system will be needed.

A large multiplex assembly can be built up by using the 24-channel junction equipment as a basic group, and combining such groups in time-division multiplex. This procedure is analogous, in the time domain, to the frequency-domain methods used to build up conventional groups, supergroups, and mastergroups. It is preferable to the use of independent or incompatible high-capacity systems for several reasons:

Economy in manufacture, ordering and stocking of a limited range of standard equipment.

Economy and freedom from impairment of patching through groups at network nodes, without the intervention of a voice frequency interface.

Ability to accommodate non-telephone traffic within the multiplex structure.

Compatibility of all transmission systems with digital switching, whether introduced at the same time or subsequently.

From the point of view of telephony, the successive multiplexing factors are to some extent arbitrary. It is therefore possible to accommodate other classes of message without detriment to telephone practice. In developments at STL, successive multiplexing factors of 4 have been used: the resulting bit rates and traffic capacities are displayed in Table 1. It is to be understood that the higher bit rates may convey any mixture of messages so long as the gross bit rate is suitable.

Table 1

Multiplex hierarchy for digital network

Bit rate Mbit/s (approx.)	Number of p.c.m.–t.d.m. telephone channels	Other types of message accommodated
1·5	24	4 p.c.m. music channels Various data multiplexes Encoded 12-channel f.d.m.
6·2	96	P.c.m. video telephone Encoded 60-channel f.d.m.
25	384	Encoded 300-channel f.d.m.
100	1536	P.c.m. television (625-line)
400	6144	

Various other multiplexing factors have been proposed: small changes in the numbers are unlikely to affect either the technology used, or the overall cost per channel.

Multiplexing and transmission equipment can be designed to convey any digital sequence of the appropriate rate, regardless of geographical origin, message content, frame structure or code. Digital multiplexing and demultiplexing equipment can be designed to accept signals with a certain tolerance on phase and frequency, so that exact synchronism of signals is not essential. Within a limited area, it is practicable to derive all signals from a common clock or frequency standard; but interconnection between autonomous areas is likely to occur and should be provided for. The technique known as 'digit stuffing' has been developed, both at Bell and at STL, for just this purpose. It depends on the insertion of redundant digits in message streams, in whatever quantity is needed to raise the aggregate bit rate to a standard value.

The types of cable suitable for the various bit rates proposed are shown in Table 2.

Table 2

Bit-rate capacities of cables

	1·5 Mbits	6·2 Mbits	25 Mbits	100 Mbits	400 Mbits
V.f. pair or quad	×	×			
Carrier pair		×			
Possible cheap coaxial		×	×		
4·4 mm coaxial			×	×	
9·5 mm coaxial				×	×

Transmission over coaxial cable has been studied by STL at 25 and 100 Mbit/s, and by Bell at 224 Mbit/s. Laboratory models of regenerative repeaters exist for these rates. Repeater spacing on 4·4 mm coaxial cable is about 4 km for 25 Mbit/s and 2 km for 100 Mbit/s. Various line codes have been proposed: in general some digital translation is required to convert an arbitrary bit sequence into some form of restricted signal more easily regenerated at the repeaters. The cost of such a translation is only a small fraction of the total terminal cost.

Radio transmission is less immediately attractive than line, since the bandwidth required for simple p.c.m. signals (about 8 times the bandwidth of the channels conveyed) is not always available. If this factor be ignored, short-distance radio links for say 96–384 channels could be very cheap. Techniques for the conveyance of digital signals within a restricted band-width (about equal to that required for s.s.b.) are being investigated but must still be regarded as speculative.

High-capacity p.c.m. transmission invites comparison with the f.d.m. systems which are now widely used. The advantages of p.c.m. are:

The equipment cost (forecast for early 1970s) is lower for p.c.m.

On paired cable, p.c.m. can carry more traffic.

P.c.m. provides a cheap and simple outband signalling channel.

P.c.m. is susceptible to integration with switching.

P.c.m. is suitable for all categories of message. In principle, f.d.m. techniques can provide circuits of various bandwidths: but the impairments tolerated by telephone, television, music, data, etc., are quite different. In practice carrier systems are designed to suit the statistical properties of telephony, and mixed traffic is a difficult problem.

P.c.m. is suitable for new transmission media such as waveguide and optical fibres which are likely to constitute the very-wide-band routes of the future.

The disadvantages of p.c.m. are:

F.d.m. systems are already available for 2700 channels. By the time p.c.m. has been developed for say 3000 to 6000, f.d.m. may well be up to 8100 or 10 800 channels. The number of channels which can be carried by a good-quality coaxial cable or radio link may always be a little larger using f.d.m.

F.d.m. is already widely used, and any new system such as p.c.m. raises problems of incompatibility.

Since the advantages are mostly long-term, it may be asked whether piecemeal introduction of p.c.m. on trunk circuits would be worth while. The crucial factors, as ever, are performance and cost. Speech quality in an integrated digital network would be better than in an analogue network: but extensive use of p.c.m. transmission links with analogue switching would introduce many quantizing impairments in tandem. There are two ways of overcoming this problem. It would be possible to use 8-bit encoding for trunks (as compared with the present 7-bit for junctions). But even if this were not done, the 7-bit quality would probably be good enough for 6 to 10 links in tandem. Opinions differ on the figure, but the limited subjective tests which have been made over tandem connections are, on the whole, favourable. It could be argued that there is no real need for more than say 7 links in tandem: but this too is a controversial matter. In the writer's view piecemeal installation could begin in the UK and in other countries already using p.c.m. as soon as the equipment could be manufactured (say 1969–70) and would be overtaken by the growth of digital switching before any serious problem arose.

Cost forecasts are favourable to p.c.m., if one ignores compatibility problems and compares only equipment, cable and installation. The cost of either p.c.m. or f.d.m. systems is made up of:

The line system, comprising cable and repeaters.

The terminals, comprising primary multiplexing (channelling), secondary or higher order multiplexing, and signalling interfaces.

A large part of the terminal is the primary multiplex, which is essentially similar to the junction terminals now in production. Higher-order multiplex equipment exists in the form of laboratory models. Thus there is enough information for cost estimating. Even allowing for cost reduction in f.d.m. equipment, it is likely that in say 1970–72 the p.c.m. terminal will be some 25% cheaper. This is on the assumption that channel-associating signalling is required: common-channel signalling could alter both the f.d.m. and the p.c.m. system significantly.

The line system, for a given number of channels, is likely to be about equal in cost whichever technique be used. P.c.m. uses more but cheaper repeaters, and the cable is assumed to be the same. Because the number of channels available is smaller with p.c.m., long links of high growth rate could be more expensive: but over the distances and capacities required in the U.K. (and most European countries) expense on the lines is outweighed by savings on the terminals.

This reckoning refers to single routes, but of course practical telephony requires many branching routes. If a system is installed to carry say 1536 ($= 64 \times 24$) channels from London to Leeds, it is unlikely that all channels terminate at these points. Maybe 16×24-channel groups are required at Leeds, another 16 groups at Harrogate, and a few each at many places such as Bradford, Halifax, and Huddersfield. The compatibility of a high-capacity p.c.m. system with the 24-channel toll and junction systems now becomes valuable. The higher-order multiplex equipment is placed at Leeds, together with primary channelling only for locally-terminating traffic. Through groups are sent over junction cables to their ultimate destinations. True, a similar thing could be done in f.d.m.: but there is no cheap low-capacity f.d.m. system comparable with the p.c.m. junction equipment. Here is an economical use of p.c.m. in an extensive network, not dependent on the introduction of digital switching.

5. Switching Principles

The switching field is remarkable for the number of different techniques which have evolved, and for the fact that most of them survive to this day. Many of them have introduced successive advantages either in the switches which connect the message circuits, or in the methods used to control the operation of those switches. A classification, in approximately chronological order of origin, is shown in Table 3. P.c.m. appears as a logical successor to the others, having most of the advantages listed for the preceding techniques plus several unique benefits of its own. It does have one significant disadvantage—the cost of interface equipment—whose effect on its usage is described later.

Table 3
Classification of switching techniques

	Type of switch	Type of control	Examples	Advantages over forerunners	Disadvantages compared with forerunners
1.	Wiping contact	Directly actuated	Up-and-around step-by-step, some rotary 7D		
2.	Wiping contact	Indirect (register)	Most rotary, including all 7A, 7B, 7E marker-uniselector systems U.K. director system	Closed numbering scheme More economic routing Alternative routing Efficient use of common equipment Higher speed (usually)	
3.	Pressure contact	Indirect	Crossbar-Pentaconta, Ericsson 41, Plessey 5005, WE No. 5, etc.	Ease of maintenance Reliability Higher speed Economy	
4.	Reed	Indirect (register) or central processor	He60, TXE series, Artemis, 10C	Sealed crosspoints Additional facilities (calling line identification, transferable numbers, etc.)	Cost of crosspoints
5.	Electronic t.d.m.–p.a.m.		G.P.O.—Highgate Wood experiment Soviet of Economic Mutual Aid experiment	Crosspoint economy Reduced blocking	Crosstalk Cost of subscriber line equipment
6.	Electronic t.d.m.–p.c.m.	Indirect (register) or central processor	G.P.O.—Empress experiment Socotel Lannion experiment Various laboratory developments	Crosstalk immunity Best compatibility with digital processors Integrated network possibilities	Cost of analogue/digital interface equipment

The distinctive properties of the p.c.m. switch stem from the facts that it is digital and that it passes *multiplexed signals*. In the field of transmission, multiplexing came into use before p.c.m., and its basic features are well known. In the switching field, the use of multiplex technique is more complex but lesser-known, and calls for some preliminary explanation.

A multiplex highway carries samples of many channels, arranged in time sequence. There are two modes of switching which can be employed, usually known as space switching and time switching.

Space switching is the interconnection of several multiplex highways by electronic switches acting instantaneously (or nominally so: in practice there is a small fixed delay). Thus a signal appearing in channel x of highway a can be connected into the same channel x of some other highway b. In general there will be several input highways, say $a_1 \, a_2 \ldots a_m$ and several output highways, say $b_1 \, b_2 \ldots b_n$, interconnected by an $m \times n$ matrix of switches. The interconnection patterns in each time slot are independent, because the electronic switches can be reset in an interdigital interval. Thus a multiplex space-switching matrix carrying, say, 96-channel signals is equivalent to 96 conventional space-switches each of availability n.

The economy due to multiplex usage is especially striking in space switches. For instance, one 96-channel highway can carry the traffic originated by perhaps 700–1000 subscribers. Fifteen such highways can carry the concentrated traffic of a large terminal exchange. A 15×15 space switch for 96-channel signals can be built with cheap integrated logic to occupy about 20 cm of rack space.

Time switching is the rearrangement of the channel sequence in the multiplex frame, so that the signal in channel x of the (possibly unique) input highway is transferred to channel y of the (possibly unique) output highway. This is effected by writing the appropriate digits into a digital store at time slot x and reading them out at the next occurrence of time slot y. The store must, in general, have at least one word for each channel in the multiplex. One store may be used at complementary times for the two directions of the same connection. Use of a large number of channels does not save stores but does lead to high availability.

It is possible to utilize space switching alone or time switching alone but either has severe limitations. More usual is a tandem connection of space and time switches, in sequences such as T–S, T–S–T or S–T–S. The time switch can be considered as a link which interconnects the effectively separate matrices embodied in a single t.d.m. space switch. This combination enables a large array, with high availability, to be built in a compact and economical manner.

It is likely that the multiplex highways will be limited to 96 channels (about 6·2 Mbit/s with serial code format), partly because this falls comfortably within the technological bounds but mainly because it is necessary for security to distribute traffic over many highways. In small exchanges a lower highway capacity such as 24 channels might well be used.

A telephone exchange is much more than a switching system for interconnecting subscribers, and in many cases less than half its cost is due to the speech-switching equipment. The establishment and supervision of a conversation, charging, call tracing and ancillary functions of many kinds are equally essential to but quite different from the conveyance of speech in its conventional analogue mode.

A speech-path switch is usually conceived as a contact whose closure enables an arbitrary signal current to pass. There is little resemblance between the speech message, the signalling information which accompanies it, and the controlling currents which actuate the selectors. Digital switching uses a more unified technology. The crosspoints of a matrix are logical gates, whose several inputs—digitized messages, coordinate selection, timing waveform—are pulses of similar amplitudes and durations. The apparatus for supervision, routing and control uses similar digital signals and similar digital technology to the speech-path switch. The entire equipment is susceptible to integrated embodiments, whether the now conventional semiconductor integrated circuits or the newly-developing practice of large-scale integration.

6. Switching Applications

At first sight, it is natural to consider p.c.m. and the various alternatives as contestants, and to ask which will win or lose. But this is not a realistic question. Detailed study suggests that, in the near future at least, no one technique will offer clear superiority in every field of application. In any case, equipment of many types, and manufacturing capacity for several of them, already exists and will persist for many years. Thus the true questions are:

In what field of applications is each type advantagious?

Which types can coexist and fulfil complementary roles?

The answer given here is that p.c.m. does have a distinctive role: that it can coexist with space-division systems: and that its most natural partnership is with the more modern space-division switches, which are amenable to indirect control. If this were not so, the introduction of p.c.m. would, despite its many advantages, be very much more difficult.

Speech-path switching equipment normally falls into one of two categories:

(a) Switches giving access to a group of trunks within the exchange, or a group of external circuits, each circuit being used intensively.

(b) Switches giving access to individual circuits such as subscribers lines, each carrying relatively light traffic.

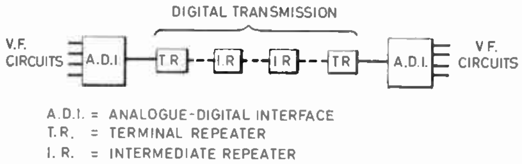
In terms of crossbar switching, these are (a) trunk link frames and (b) line link frames. In step-by-step, they are (a) group selectors and (b) final selectors plus subscribers' uniselectors or line finders. These categories will be called here (a) group selectors (b) line selectors.

A terminal exchange has switches of both categories. A tandem exchange has only group selectors. Pure tandem exchanges are not very common. Far more numerous are exchanges with a mixed terminal and tandem function. These have line selectors for their local subscribers, together with more group selectors than the locality would need, shared between traffic with a local origin or destination and traffic which is passing through. The quantity and description of such exchanges varies from one geographical location to another, but two general statements can be made: almost all primary centres are of this type, and so are many local exchanges in large cities or conurbations.

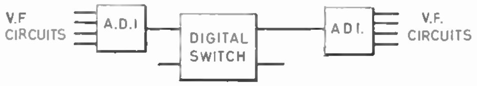
It has often been said, and rightly, that p.c.m. is suitable for tandem exchanges. A more general statement, which includes this as a special case, is: p.c.m. is suitable for group selectors. The justification for this statement will emerge in the following section.

7. Integrated Switching and Transmission

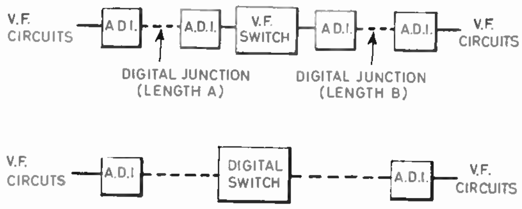
It is now well known that p.c.m. transmission can be economical in its own right, and it is contended here that the range of suitable applications is wider than present usage would suggest. Similarly p.c.m. switching on its own can be economical in certain applications, of which the large tandem exchange is the most generally agreed. In either case part of the cost resides in the basic mechanisms for digital transmission or switching, and part—a very large part—in the interfaces between the voice frequency world outside, and the new digital equipment (Fig. 2(a), (b)).



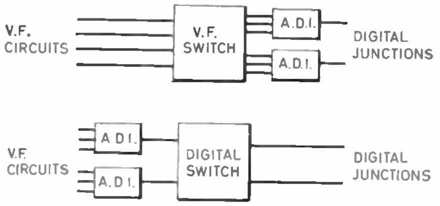
(a) p.c.m. transmission.



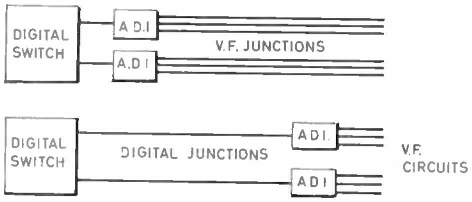
(b) p.c.m. switch.



(c) v.f. versus digital tandem switch.



(d) v.f. versus digital switch: digital junctions postulated.



(e) v.f. versus digital junctions: digital switch postulated.

Fig. 2. Integrated switching and transmission.

The basic concept of the integrated network is that both transmission and switching should be carried out in the digital mode, and the interfaces introduced only where essential. In situations where p.c.m. switching and transmission could be justified separately, the integrated system is obviously more economical becauses interfaces are saved. More important in the long run, integration extends the usefulness of p.c.m. to situations where neither element would pay in on its own. And, since most of the impairment in a p.c.m. system is introduced at the interfaces (not along the route), the integrated method leads not only to lower cost but to better performance.

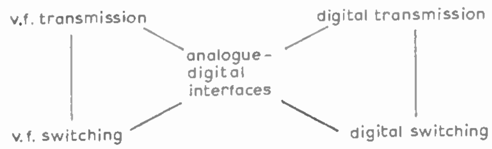
The comparison shown in Fig. 2(c) has often been adduced as an example of integration. Since much of the cost of a p.c.m. junction system is in its terminals (interfaces), separate systems are justified only if the lengths A and B exceed a certain pay-in distance, say 10 km. But if the digital switch (together with a minimal interface to line) costs no more than a v.f. switch, p.c.m. junction transmission pays in if the total length $A+B$ exceeds the pay-in distance, say if each is only 5 km.

Again, if p.c.m. junctions be postulated as a source of digital traffic, introduction of digital switching can be viewed as a shifting of interfaces (Fig. 2(d)) and the digital switch as such has an easy cost target. Note that the switch in this example is not necessarily a tandem exchange, but may be any group selector whether in a tandem or terminal exchange.

If p.c.m. switching be postulated (Fig. 2(e)), then introduction of digital transmission amounts to installing at a distance interfaces which would otherwise be needed locally, and so it pays in over distances or capacities which would otherwise be too low.

The common factor in all these approaches is that once some digital item has been accepted, it becomes easier for some other digital item to pay in. This is perfectly true, and will have its influence on the evolution of practical integrated networks. But any such viewpoint is unduly tarnished by the traditional dichotomy between switching and transmission. With the advent of digital methods their technology is becoming more alike, their economics and operating principles ever more closely intertwined. The optimization of an integrated network requires a new approach, from first principles.

The major elements in a network may be displayed as follows:



The basic digital processes are cheaper than their v.f. counterparts and introduce less (virtually no) impairment. But both v.f. and digital elements must coexist, and they can only be connected through interfaces which introduce much of the cost and almost all the impairment. Digital-to-digital or v.f.-to-v.f. interface equipments are relatively simple, though not negligible (especially signalling interfaces between v.f. switching and transmission). The crucial question is then: where should the interfaces be placed in order to minimize cost or impairment or some acceptable compromise between them?

Figure 3 shows an outline telephone network, with some alternative dispositions of interfaces. In successive members of the series, p.c.m. extends further through the network. Scheme (a) uses only isolated transmission links, such as the present junction systems. If two or more links occur in tandem, then there are two or more pairs of interfaces, with the cost and impairment that this implies. Scheme (b) introduces tandem switching: if this were used extensively the total number of interfaces would be reduced, and only one pair would ever occur in any connection. Scheme (c) uses also digital group selectors at the terminal exchanges. The number of interfaces is much the same as in (b), because the total traffic over junctions is being routed through them in each case, and the circuits concerned are all likely to be used with about the same efficiency (say 0·6 erlang per circuit).†

So long as the traffic is great enough to justify multiplexing, this is a sound solution both technically and economically. The number of interfaces is reduced to the minimum, and the digital mode—which is basically cheap and free from impairment—is used for all but subscriber-line transmission, and for all but subscriber-access (line selector) switching.

The next scheme, shown at (d), carries the digital mode further into the terminal exchange. There should be some saving on switching cost, but the number of interfaces is increased by 8 to 10 times because they are now inserted in subscriber circuits with light average loading. Whether this scheme is cheaper or dearer than (c) depends on the balance of these two factors. Perhaps it will be cheaper ultimately, but this is unlikely over the next several years. Circuit quality may be improved somewhat by cutting out electromechanical switches completely but, on the other hand, a reed or crossbar line link frame would more readily admit several categories of service with different bandwidths or digit rates (e.g. normal speech, wideband speech for conference circuits, facsimile, high-speed digital data, perhaps even video telephone).

† The erlang is the unit of traffic flow; the flow in erlangs for a period is the number of calls originated during the period multiplied by the average holding time of a call (expressed in terms of the period).

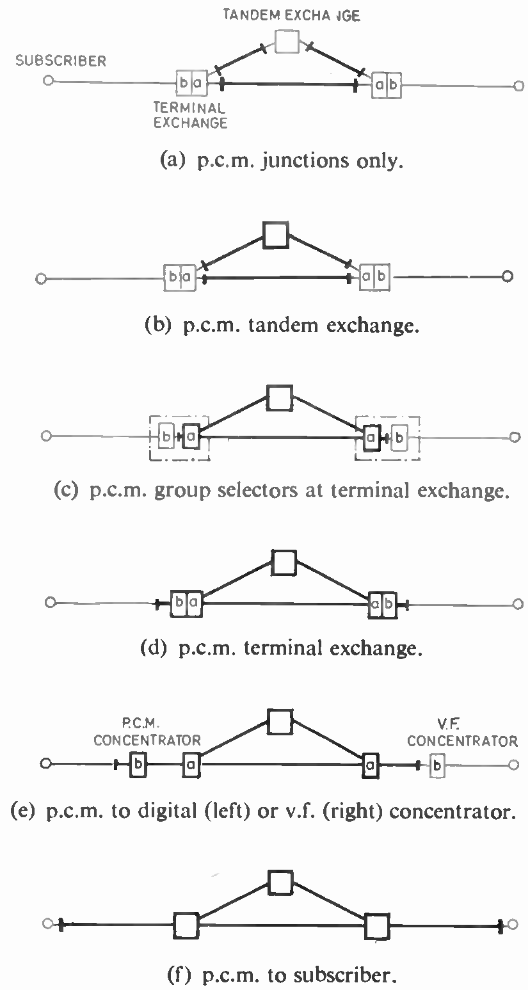


Fig. 3. Alternative interface locations in a telephone network

Extension of p.c.m. to concentrators is shown in Fig. 3(e). The switching stage in the concentrator, which may be either v.f. or digital, is equivalent in function to the line selector of a normal terminal exchange. The performance and economics of this arrangement are consequently not very different from those of (c) and (d) respectively, except insofar as the principle of concentration (not that of p.c.m.) saves subscriber lines.

Finally, scheme (f) shows digital transmission extended to the subscriber. Transmission at single-channel digit rate over subscribers' lines is feasible, likewise the construction of a compact encoder/decoder which would fit in a subset of normal size. But this arrangement, now and in the near future at least, is unduly expensive for normal telephone service. The number of circuits to be interfaced is the maximum, and each interface is necessarily single-channel rather than multiplex. The possible justification is that new categories of service could be provided, such as high-speed data, digital facsimile, or digitally scrambled speech. The demand for such services will probably necessitate digital extension to a limited class of subscribers within the next few years. Their subsequent spread to a majority of subscribers is a matter for conjecture.

The control functions of many modern exchanges are centralized in a computer-like digital processor, for three main reasons:

It is then easy to incorporate new facilities such as calling party identification, abbreviated dialling, automatic routining, etc.

Complex routing and translation procedures are facilitated.

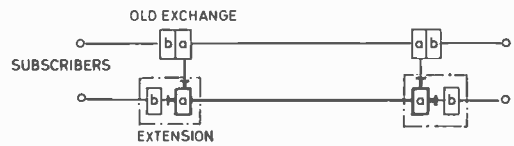
Central control is cheap and reliable if carried out on a sufficiently large scale.

These arguments apply to a network just as strongly as to an individual exchange, and it has been suggested that an extensive area could be controlled by a common processor. Information from and commands to the remote locations would be conveyed by digital data links, essentially similar to those previously proposed as a means of common-channel signalling. This conception is well suited to the digital integrated network, where the transmission links and interface equipment for control purposes would be highly compatible with the techniques used for digital signals. Certainly in a configuration such as Fig. 3(e) a concentrator could be dependent on the main exchange, and would not need a complex autonomous control system of its own. Similarly, in a larger multi-exchange area many functions could be centralized in the major exchange.

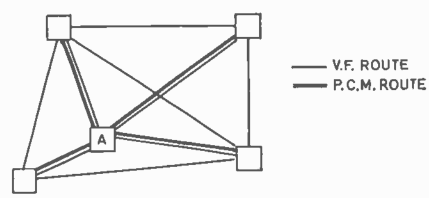
8. P.C.M. Extensions to V.F. Networks

The schemes shown in Fig. 3 imply that a whole network, or a substantial sector of it, is treated in much the same way. This may be possible in newly-developing areas with little or no investment in old plant, or occasionally in long-established areas which present problems of such magnitude that complete replacement is necessary. But most telephone equipment is used to extend the existing network of an area in an intermediate stage of development. The next problem, therefore, is to see how a partly-integrated p.c.m. network, not necessarily optimal, can coexist with a v.f. network of predetermined form.

It is assumed here that p.c.m. is to be used mainly or exclusively for multiplex line transmission, tandem exchanges and group selectors. Further, it is to be used as an extension, to carry growing traffic, with a minimum scrapping of existing plant. The general arrangement is shown in Fig. 4(a), for an elementary configuration. A hybrid p.c.m. v.f. network exists in parallel with the original v.f. network, and has interconnections which carry (at least) traffic originating in one half and terminating in the other. If traffic is increasing faster than the number of subscribers (a common situation) more interfacing may be necessary since the new p.c.m. routes will act as an overflow path for the old routes.



(a) simple network extension.



(b) more complex extension.

Fig. 4. P.c.m. extension to v.f. network.

In a more extensive network, the optimal routing arrangement for the p.c.m. part may differ from the original one. An example is shown in Fig. 4(b). Here the five exchanges are almost fully meshed with v.f. circuits, as is usual in a multi-exchange area with linked numbering. The p.c.m. extension is star-shaped, with tandem switching at A. The basis for this is that:

P.c.m. switching and transmission are more economical in large quantities than small.

Tandem switching introduces no impairment and little cost.

With suitable control, routing need not be tried to the numbering plan.

The major technical problems concern the control of the composite network. As compared with a fully integrated network designed as such, there are two extra difficulties.

Firstly, it is likely that some (at least) of the extensions will be too small to justify individual processors for control. There are two solutions. The small stations can use wired logic rather than program control: the basic telephone service is economically provided in this way, though certain special facilities may be excluded. Otherwise, the control of a multi-exchange area may be centralized for instance at A in Fig. 4(b). These are not strictly alternatives: the best solution may be a mixture of the two.

The second difficulty is less tractable. In a new network designed to use a hybrid technology, with p.c.m. group selectors and space-division line selectors, the same method of control should be applicable to both types of selector. Indeed, the selectors must be designed with this in mind. But in general it will not be possible to apply to a pre-existing exchange an over-riding control from a new processor. The old- and the new equipment must be treated as autonomous units, and this restricts the efficiency with which the combination can be used.

As an example, consider the method of path search for an originating call in a hybrid exchange. A free path must be found from a calling party, through the line selector, to a link between line and group selectors: thence through the group selector to a junction on the required route. In a centrally-controlled system it is possible to connect the calling party only to a link which can in turn be connected to a free junction. But, if the control of line and group selectors cannot be coordinated (either because one or both uses directly-controlled switches, or because the control techniques are incompatible) this cannot be done. The caller is connected to any free link, and suffers the risk of blocking in the group selector even though the latter may have several free paths via other links.

This type of limitation is familiar in old switching systems, but could be avoided in new ones. One possible way out for mixtures of old and new may be suggested. Digital switching permits the construction, at moderate cost, of large switching arrays of high availability. It would be possible therefore, to reduce blocking in a digital group selector to a very small value normally regarded as uneconomic (say 10^{-4}). Then almost all the blocking is due to the old line selectors. Even in a new hybrid exchange, the economic balance between space-division and digital switching would allot a larger share of the blocking probability to the former.

9. New Messages and Media

For a long time to come, the major business within the scope of this paper will be the application of p.c.m. to telephone networks recognizably similar to those of the present day. On a broader view, the advantages of digital technology are enhanced by the facts that:

It can convey messages of many other categories.

It is applicable to the wideband communication media of the future.

The importance of these messages and media will undoubtedly increase in later years, and it is reasonable to take them into account in present-day planning. In general, this can be done with little or no detriment to the short-term provision for telephony. For instance, it would seem reasonable to:

insist that multiplexing and transmission facilities be tolerant to a wide range of code sequences and frame structures;

choose a multiplex hierarchy to accommodate the major known sources of non-telephonic message, such as television;

utilize for the line selector function switches which can carry a variety of messages for which subscribers are likely to need the services of a switched network;

utilize control and signalling techniques capable of distinguishing several categories of service.

The messages to be considered are: digital data at all rates from 50 bauds up to several megabits per second: encoded television to broadcast standards: encoded picturephone: encoded facsimile: encoded music or other wideband audio channels. The term 'data' embraces many different categories of message, with different rates, lengths, and modes of operation. Some of these, probably including encoded visual displays, are potentially compatible with encoded speech. It is suggested that they should use rates such as 40–48 kbit/s which can be carried either by a 56 kbit/s digit stream equivalent to one p.c.m. speech channel, or a 48 kHz analogue path equivalent to a 12-channel group in frequency-division multiplex. The development of a p.c.m. integrated network for telephony will make available, not just point-to-point links, but a switched network capable of carrying such data channels. This would differ from the present telephone network in several ways, namely:

Rapid response time of digital switches (with suitable control, only one frame of delay, i.e. 125 μs.

Possibility of fast signalling between exchanges, e.g. with common channel using 48 kbit/s).

Rapid path search and translation using processor control.

Economical achievement of very low blocking in large digital switches.

The complete circuit between line selectors could be established in, say, 40 ms, given adequate processing capacity in the digital exchanges. The proviso is important, because a network used only for telephony might be deliberately allowed to take, say, 200–400 ms, which is fast enough for telephone users but can be provided a little more cheaply. However, the extra speed is technically feasible and could be provided were a sufficiently important category of traffic to demand it.

The advantage of a mixed-traffic network over a separate data network is one of economy. Transmission, switching, control processing are all cheaper in large quantities than in small. A nation-wide network for data alone would inevitably be rather 'thin', especially at first, and correspondingly expensive per circuit or traffic unit. The marginal cost of adding extra circuits or traffic to a large network would be much less.

Other categories of 'data', including much alphanumeric information, would be conveyed less efficiently by 48 kbit/s, circuit-switched channels: either because the data rate is too low, because the average message duration is very short, or because more rapid set-up is demanded than even the newer telephone practices could provide. For these a message-switching network has been advocated. Suitable equipment based on this principle would be an efficient means of assembling diverse messages into continuous digit streams of relatively high rate. These digit streams could share trunk and junction transmission facilities with the other categories of digital message discussed here, and it is suggested that their gross rates could usefully conform to the telephone multiplex hierarchy.

In the long-term future, the spread of a videotelephone service will demand wideband transmission media such as waveguide optical beams or optical fibres. These differ in detailed technology, but have one important property in common: they are much better suited to digital than analogue transmission.

10. Some Remarks on Economic Comparisons

To be widely used, a new technique must provide good enough service at low cost: but it is not easy to define what 'low enough cost' really means. Obviously, it must not cost more than a conventional method of providing the same service. But there are two complications. The service is not exactly the same: usually some improvement is introduced, whose value is difficult to quantify. And the temporal pattern of expenditure may vary: it may be concentrated into a large initial outlay, or spread over a long period.

It is usual to allow for the temporal factor by estimating the *present value* of future expenditure. That is to say, an amount A to be paid in n years' time is taken as equivalent to a smaller amount $A(1+i)^{-n}$ paid now, where i is a rate of interest or discount. This factor is helpful to some p.c.m. applications and a hindrance to others.

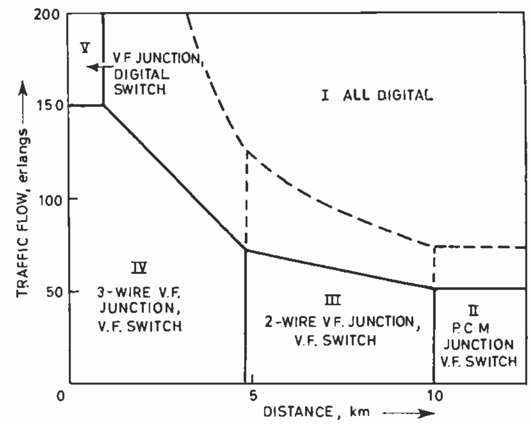


Fig. 5. Economic partition of voice frequency and digital equipment in a star network.

In comparing p.c.m. with v.f. junction transmission it *is* helpful. Factors adverse to the p.c.m. equipment —notably, a 20-year nominal life compared with 40 years for cable—are cancelled out by the spreading of investment over many years. It would be usual to install enough cable for, say, 6 years' growth or more, because small installations are relatively expensive. On the other hand, p.c.m. terminals and repeaters can be installed as required, the main initial outlay being the provision of repeater housings and the splicing of cables for perhaps 6 years' growth. Comparison of present values leads to much the same result as comparison of installed capital costs.

In comparing digital with conventional switching, the temporal factor may be adverse. This is because a small installation may need some disproportionately large initial expenditure on interface and control equipment. However, digital switching *per se* is cheaper by such a large factor that, in appropriate usage, it pays in even on the less favourable reckoning. This is especially true of integrated areas.

As a somewhat idealized example of integrated system economics, let us consider a star network, such as the one shown in bold lines in Fig. 4, with a tandem exchange at the central node A and junctions radiating to dependent exchanges. (Local subscribers at A can be considered as a dependent exchange connected to the tandem by junctions of zero length.) The economic distribution of analogue and digital techniques within such a network can be determined if certain assumptions be made about their costs. Average values must be taken, though of course in practice the unit cost can vary with number, location, and exact facilities required.

Digital switching, even with the addition of analogue/digital interfaces, is cheaper than space-division for group or tandem switches. However, the addition of a central processor to the digital system makes it relatively expensive for small installations though still economic in large ones. The dividing line in the near future is estimated to be at about 150 erlangs of traffic, assuming about 0·6 erlang per through circuit.

For junction circuits, a typical result is that digital transmission plus interfaces will be cheaper than 2-wire voice frequency junctions for distances over 10 km, while 3-wire junctions will be cheapest for very short distances (say less than 5 km).

It is also necessary to make some assumption about the distribution of traffic in the network. The most important parameter is the proportion of local traffic, which is taken as 20%, leaving 80% to pass over one or more junctions. This is approximately true for many urban areas.

The economic partition between voice-frequency and digital equipment may now be given as a function of two parameters: the traffic through the tandem switch (E) and the average length of junction circuit (L). These are respectively the vertical and horizontal co-ordinates of Fig. 5. This chart is divided by the solid lines into 5 regions, within each of which there is a most economic solution, namely:

I: integrated digital transmission and switching.

II: p.c.m. junction system, v.f. switching.

III: 2-wire v.f. junction, v.f. switching.

IV: 3-wire v.f. junction, v.f. switching.

V: 3-wire v.f. junction, digital switching.

The regions II and V reflect the fact that long p.c.m. junctions, or large p.c.m. exchanges, are economical even in isolation. The most interesting feature of the chart is the large extent of region I, which shows that the combination of digital transmission and switching can be economical even in networks where neither would be justified in isolation. Consider, for example, a tandem exchange of 75 erlangs (half the break-even traffic for an isolated exchange) with junctions of average length 5 km (half the break-even length for an isolated junction system).

It is possible that the cost of space-division switching will be reduced, and (in the unlikely event that digital equipment were not reduced so rapidly) the economic size of digital exchanges would rise. Suppose, as an extreme example, that it were to increase to 1500 erlangs. The boundary of region I would then move to the position shown by broken lines. (Region V is well above the upper edge of the chart.) Even in this case, the combination of 150 erlangs and 5 km is economic for an integrated digital network.

The explanation for the large extent of region I is simple: with compatible switching and transmission techniques, very little is spent on interface equipment.

This example is of course idealized and simplified. Many realistic cases need to be studied in detail.

11. Conclusions

The foregoing view of p.c.m. and digital communication may sound rather partisan. However, most of the statements are based on careful analysis and comparison, too long to record here.

In the writer's view it is feasible and economic to use digital methods for virtually all multiplex transmission and group selection. It is preferable to use space division for line selectors, which constitute some two-thirds of all selectors, and thereby achieve intensive use of analogue-digital interfaces. Digital signals can be extended to those subscribers, probably a small minority though an important one, who need them. Control and signalling can be cheaper, more uniform, and more rapid in action. Transmission impairment can be reduced virtually to that of one encoder-decoder pair. Blocking can be reduced virtually to that of the two line selectors. Many categories of message, both analogue and digital, can be accommodated.

But this is not a prediction. It is a scenario for a possible future. The impact of p.c.m. is already significant, but its extension depends on decision by the operating authorities. A decision in its favour would be amply justified.

12. Acknowledgment

The work on which this paper is based was carried out while the author was at STL, and he would like to acknowledge his indebtedness to his former colleagues.

Manuscript first received by the institution on 31st May 1968 and in final form on 16th August 1968. (Paper No. 1235/Com 8.)

(For discussion on this paper, see page 51.)

British Radio and Space Research

The triennial (1965–67) report of the Radio and Space Research Station† (the first to be published since the transfer of the Station in 1965 to the aegis of the Science Research Council) shows that the Station's activities will continue to be in the broad area of atmospheric physics and radio wave propagation.

Studies of the physics of the ionosphere have continued using data obtained from the Canadian satellites *Alouette I* and *II* received by telemetry at R.S.R.S. outstations situated at Winkfield (Berkshire), in Singapore and in the Falkland Islands. Many aspects of ionospheric behaviour have been studied with these data and in particular the knowledge acquired has indicated the need for combined studies of the behaviour of the ionosphere and the neutral atmosphere. Work done in this connection has indicated that atmospheric winds at heights between 100 and 700 km are responsible for many well known ionospheric anomalies.

Further work has been done on relating meteorological data to the characteristics of u.h.f. signals propagating in the troposphere. This work will be extended with the aid of the 25 m (82 ft) diameter steerable aerial at Chilbolton, Hants, which was opened in 1967.

† Report of the Director of the Radio and Space Research Station. (H.M.S.O.) Price 7s. 6d. net.

In anticipation of the commercial use of millimetre waves for purposes of communication investigations have begun of their transmission characteristics. The scintillation of signals arising from variations in atmospheric refraction has been studied and measurements made of attenuation caused by rain on 2·9 mm waves propagated over a path of 300 metres length. Research at 0·783 mm wavelength has been concerned with the development of the techniques of generation and detection and also the measurement of absorption by water vapour.

It is only in recent years that the importance of the millimetre wave component of the Sun's radiation has been realized. A practical result of the present studies may be to provide accurate predictions of radiation from solar flares which could be harmful to passengers in the very high flying aircraft of the future, or a hindrance to the engineer operating high frequency communication links.

The Station's experiment to measure high frequency radio noise distribution over the world, launched in May 1967 in the *Ariel III* satellite, has worked well. Noise intensities above major storm areas are in broad agreement with expectations and much new information has been obtained from areas not previously covered by ground stations.

Camera Tube for Ultra Violet Radiation

A new type of camera tube, called the Uvicon, is being used in the National Aeronautics and Space Administration's large unmanned satellite, the Orbiting Astronomical Observatory, *OAO-A2*, launched recently. Developed by Westinghouse Electric Corporation, four of the Uvicon tubes will map the stars and interstellar space in a space experiment—called Project Celescope—conducted for N.A.S.A. by the Smithsonian Astrophysical Observatory, Cambridge, Massachusetts. The mapping will be done by means of far ultra-violet radiation. These ultra-violet rays, emitted by celestial bodies, never reach the Earth's surface because of atmospheric absorption.

Thus, the Uvicon will give pictures of the heavens that have previously been screened from Earth-bound observations. Such pictures offer promise for gaining new insight into theories concerning the atmospheres of young, extremely hot stars, including their age and their chemical composition. Such information might lead to new discoveries about the origin and evolution of the universe.

Project Celescope is a pioneering space experiment dating back to the late 1950s when the Smithsonian Observatory proposed it to N.A.S.A. and began work with Westinghouse on development of a suitable camera tube. The project will survey some 100,000 ultra-violet stars at a rate of up to 700 per day. An ultra-violet map of the entire sky will require about a year to complete.

Basically, a Project Celescope telescope consists of three components: an optical system (Schwarzschild telescope) to form the ultra-violet star images and focus them onto the camera tube; the Uvicon tube, which responds to the u.v. images, amplifies them electronically stores them and converts them into television-type electrical signals; and a digital television system that processes these signals for transmission back to Earth, where the original u.v. pictures are reconstructed. There are four high-resolution telescopes, each equipped with a Uvicon camera tube, in the project.

Ultra-violet images are focused on the input surface of the Uvicon through a window of a material such as lithium fluoride. Glass is not used because it absorbs the u.v. wavelengths of interest. The input windows and photo surfaces of the four Uvicons are designed to respond to four bands of ultra-violet radiation in the wavelength range from 1050 to 3000 Å, or about 0·1–0·3 μm. The back surface of the window is coated with a photo-sensitive material that releases electrons where the u.v. image strikes it. These electrons are accelerated by high voltage toward the rear of the tube, where they strike a thin-film target consisting of a porous layer of potassium chloride.

The impact of each electron on the target causes it to emit as many as 200 to 300 secondary electrons. The flow of secondary electrons is conducted through the porous layer to a metal film. Here the electrons create an electric charge pattern that is the exact reproduction of the original ultra-violet image. This charge pattern, amplified about 200 times by the target, is read out as an electrical signal by scanning the target with a beam of electrons. These are amplified, coded and stored until released by command from the ground.

U.D.C. 621.394/5.74:681.32

A Communication Network for Real-time Computer Systems

By

D. W. DAVIES,

B.Sc., A.R.C.S.†

Reprinted from the Proceedings of the Institution's Convention on 'Electronics in the 1970s', held at Cambridge on 2nd to 5th July 1968.

Summary: Digital communication between computers and their users has until now been provided by exploiting the telephone network to carry digital data because of the wide coverage of the telephone network and the high capital cost of establishing any new, generally-available telecommunication service. In the long run it is expected that a specialized digital communication network will be required, designed around the requirements of computer services which interact in 'real-time' with their remote users. The system design of the proposed new network is outlined and some estimates of performance are given.

1. Introduction

The majority of computers in use today operate 'off-line' and are local in their operation, having no communication network attached. Operation of computers in this way detaches them in time and space from their *users*, avoiding the problems of 'real-time' operation or remote access.

A small number of computers exist which are both 'real-time' and connected through a communication network to their remote users. The pioneers in the civil application of this kind of computer system were the airlines for whom efficient seat reservations were such an important aspect of their economy that the cost and complexity of a 'real-time' system could be faced.

'Real-time' is an imprecise term and is used to express the fact that the computer system can react sufficiently rapidly that the interaction with the user has something of the nature of a conversation. This has a dramatic effect of the kind of service that a computer can offer enabling man's intelligence to be combined with the speed and accuracy of the computer with very useful results. Conversational use of a computer can often be justified economically where the urgency of the transaction is not in itself an important factor. In one of the existing real-time services the important factor was the validation of the input data and the service in question would not work (and make money for its sponsors) without conversational interaction.

These systems are increasing in number with the realization that new and profitable computer-based services can be developed using 'real-time' methods. It has been predicted that during the 1970s a change-over will take place so that 'real-time' systems with communication networks attached will become the majority of all computer applications. A major factor in promoting the remote, real-time use of computers will be cheap and convenient communication facilities.

Digital communication between computers and their users has until now been provided by services such as DATEL in which the telephone network is exploited to carry digital data. This is an obvious and sensible development because of the wide coverage of the telephone network and the high capital cost of establishing any new, generally-available telecommunication service. In the long run it is expected that a specialized digital communication network will be required. This paper contains a specific proposal for a digital communication network designed for computers offering real-time services to remote users.

It has been argued that the immense capital cost of the existing telephone network implies that any new telecommunication facilities must be provided by gradual development from it. This is by no means obvious, and the argument put forward in this paper is that a specialized network operating as a separate system is necessary to provide the facilities required by real-time computers. To further the argument it is necessary to give a full account of the system design of the proposed new network and make rough estimates of its economic viability. At the present time, the missing element in the argument is a firm estimate of the amount of traffic that will be generated for the new network, and work is in hand to produce this estimate.

2. Present-day Systems

Data transmission facilities based on the use of the telephone network are merely one component of the communication systems now used between real-time

† Division of Computer Science, National Physical Laboratory Teddington, Middlesex.

systems and their users. Typically these systems are based on 'private wires'. For a widespread system these private wires are arranged with 'way stations' at many points along them in order to serve a number of subscribers. The terminal equipment contains storage and the terminals are polled, that is tested in turn until one is found with a message to transmit, when this is sent rapidly in order to free the wire for the next user. A number of these 'multi-drop' lines is usually connected to the central computer where the communication controllers are sited.

In another kind of network, the computer provides service at a few distant towns with a number of terminals located in each of these towns. In this case, a small message-concentrating computer in each of the remote towns is used to assemble messages. Medium-speed digital links, provided by private telephone wires, are used between the message concentrators and the centre.

In these systems, the cost of the telecommunication equipment is considerable and hire of the private wires is only part of this cost. The systems are vulnerable to failure and this causes demands on the telecommunication authority which may become embarrassing if the number of these private networks increases. Looked at overall, the covering of the country with a mass of such private networks seems absurd and this fact gives us an incentive to design a digital communication network general enough to cover all these different requirements by one public network.

3. Types of Service that will be Provided through the Network

To avoid misunderstanding it must be emphasized that the proposal made here is for a *communication* network. The users of the network will be many and various, and, by providing generally useful services to network subscribers from attached computers they will enhance its value. The concept, however, is different from that of the 'computer grid' or 'computer utility' in which one authority is supposed to provide both communication facilities and computer power. A 'grid', if it later comes into operation might well use the network for access to its subscribers and might provide raw computer power to the 'middlemen' who sell special services though the network.

The kind of real-time computer-based service for which the network is designed can be classified roughly in the following way:

(i) Services to a group of users in a given line of business sharing a data-base. Examples would include services to estate agents, bookmakers, stockbrokers, travel agents and libraries.

(ii) Computer systems private to organizations such as large firms, banks, insurance companies, government departments, the police service and the G.P.O.

(iii) Computer services such as language translation, computer-aided design and scientific calculations.

(iv) Facilities to remote computers, for example, a disk file system or program compilation. Such facilities will improve the usefulness and economy of small computers in many applications. The provision of bulk store facilities through the network could be very profitable.

(v) Correspondence between computers owned by and operated for separate organizations in order to give a rapid response to a complex enquiry or make a complex transaction.

(vi) Remote control, telemetry or data gathering. Examples are the control of a pipe line, the signalling system for a telephone network and the collection of meteorological and hydrological data. These applications will require the fullest development of the local distribution network and may come later than the others.

Many services are in operation, have been proposed or are in the course of design in the categories (i), (ii) and (iii) above. These are the ones that have guided the design of the proposed network.

4. Principles of the Proposed System

It is well known that there are two kinds of switched communication network—the 'circuit-switched' and the 'store-and-forward' network. The circuit-switched network is the most familiar and is typified by the present telephone network. It provides the effect of a circuit from one subscriber to another. The network provides no storage for the data being transmitted except in the form of the inevitable delay in transmission.

In a store-and-forward network the information being transmitted is divided into 'messages' which are stored at each switching point before retransmission. The use of storage means that there is no fixed data-rate and information may be sent at whatever is the best speed for each stage of its journey. Correspondingly, it means that there is delay at each switching node. This delay depends on the number of queues which the message visits, the average length of these queues, the rate of transmission and length of message.

The apparent simplicity of a circuit-switched network is illusory. A well-designed circuit-switched network contains a store-and-forward system for handling the routing information. Allowing for verification of the switched path a double transit of the network is needed before transmission of the data begins. This is justified only if the amount of data sent while the path is held is large compared with the routing information.

A circuit-switched system gives no help at all in matching the characteristics of different subscribers. If more than one data rate is provided this merely divides the subscribers into mutually incompatible classes. The 'transparency' of the network to data rate and control procedure proves to be a limitation.

The disadvantage of circuit-switched systems may appear to the user in a way that is not obvious. An example is the high cost of being connected to a real-time service irrespective of the use being made of it. The cost is levied to prevent the tying-up of communication equipment at the computer centre, and it defeats the aim of ready accessibility which real-time services should be able to give. These bad features of circuit-switched communication systems must not be accepted if a better alternative exists, as we hope to prove it does.

The proposed data network has as its 'back-bone' a high-level or trunk network operating entirely in the store-and-forward mode in which the length of messages is kept within bounds and the rate of transmission is high. By these means, the delay through the high-level network is kept to a low value.

The assumption of an upper limit to the size of a message is the critical one. If a long message has to be sent (and the network design should be general enough for this to be possible) it must be broken into shorter messages for transmission. To avoid confusion we refer to the messages carried by the high-level network as 'packets'. The user's message may therefore be broken into packets for transmission.

In a real-time or conversational system the information sent between users has a natural message structure. Typically, real-time systems employ relatively short messages from the terminal to the computer centre. For example it is usual, when a teleprinter is employed, to treat each line of type as a separate message, with an opportunity for the computer centre to respond at the end of each line. In the return direction (computer to terminal), messages are typically longer but can be broken into packets without inconvenience. More detailed study of the use of the high-level network, and the control procedures employed shows that a large number of short messages will be sent through the system. The format of the packet has therefore been devised so that its message content can vary efficiently from a few bytes to about 100 bytes.

The high-level network is like a postal system in which messages are carried in standard packets. The sorting of packets at each node of the network is done by using only the data carried in standard format on the 'envelope'. The message itself can be in any code, and is not examined at the nodes.

Starting with a high-level network which will carry short messages rapidly, it is next necessary to consider how this will be employed by the subscribers. It is proposed that they should gain access to the high-level network only indirectly through computers which have been called 'interface computers'. These are responsible for forming packets in the correct format for launching in the high-level network and for dealing with standard packets coming from that high-level network in the way required by the subscriber. These interface computers have some similarity with the message-concentration computers already in use.

Interface computers may be responsible for handling a variety of terminal equipment such as keyboards, displays, line-printers and graph-plotters. For each of these terminals the interface computer acts as a 'transformer' between the local conventions required by the terminal and the packet format in the high-level network.

5. Detailed Design Considerations for the High-level Network

A design for a switching node was made in some detail, comparing three versions in order to get a proper apportionment of its tasks between the general purpose computer (which is its main controlling element) and the special hardware associated with the links. For these designs, five links were assumed each with a digit rate of 1·5 Mbit/s. Such rates can be assumed to become commonplace with the further development of pulse code modulation for telephony but the cost of providing them with present day f.d.m. techniques, though higher, is not a dominant part of the cost of the whole network.

The packet format is chosen for ease of handling the packets by computer and is unlike the well-known message formats customary in telegraphy. The method of routing chosen is adaptive so that failure of a single node or link does not deny service to any user and the network is overconnected so that alternative routes are always available.

Packets are assembled from the serial link, two bytes at a time, by the link hardware and transferred directly into the computer store. Space is reserved in the store for the maximum size of packet and only when the whole packet has arrived is an interrupt generated so that the computer can act on the packet. Within the computer, the only data handled consists of pointers to the beginning of each packet. From these pointers the address of any relevant part of the packet's heading can be determined. These pointers are queued in 'circular buffers' since it was found that list handling methods were less efficient.

For the design found to be optimum, a small computer of known design with a 0·6 μs, 16-bit, store was assumed and the link hardware and critical parts of the operating programs were designed. An approximate estimate of its performance was made by a simplified queueing model.

The performance of the switching node was found to be computer-limited at about 2600 packets per second. At 80% of this limiting rate, the mean delay to a packet through a node was 1·5 ms. Therefore the delay through the whole high-level network could easily be arranged to be imperceptible in relation to most of the user's requirements. It seems feasible, then, to employ the network for access from a computer to a remote file store.

6. Interface Functions

The store-and-forward network handling short messages provides a general purpose communication system which can be exploited by interface computers to give specialized services. Among the functions of the interface computer are the following:

(i) Assembly of data in the correct format for the high-level network and treating formatted packets from the network. This process typically involves the assembly of packets from characters and the distribution of characters to terminals.

(ii) Control of access through the network. This includes administering the routing information (source and destination) in packets. It also may involve examining incoming packets to see if the source numbers are acceptable. It is possible to control access to a centre from distant points by the use of 'keys' stored at the distant points which are not accessible to the subscribers.

(iii) Accounting for the charges for using the network.

(iv) Splitting long messages, numbering the parts and reassembling these at the far end.

(v) Control of data rate. In a network in which the rates of operation of terminals are not fixed, rate of transmission from the fast to a slow device can be controlled by acknowledgement from the far end.

The interface computer gives a special kind of service to multi-access computers attached to it. Pre-assembled packets are taken from a queue and have only source information to be added to them before launching into the high-level network. Such a facility costs less per packet than assembly from characters.

7. The Local Network

Slow terminals in the local network operate mainly in units of one character or byte and the need therefore arises for a multiplexed system handling bytes to reduce the wire cost. Such a system has been devised and is being put into operation experimentally at the N.P.L. It is arranged not only to be transparent to the 8-bit byte but also to handle status and control information to computer peripherals while remaining transparent to the data.

The way in which local networks will develop on a national scale must be closely related to the existing telephone local networks. In total cost, they dominate the economics of the whole system, and local network design is therefore a subject worthy of careful study. There are indications that bit rates of the order of 50 kbit/s are feasible over wires primarily designed for telephony.

Using the random access byte multiplexing technique adopted for the N.P.L. experiment, a form of local network can be postulated to suit the early conditions of a data network. (As the number of subscribers increases, the need for multiplexing in the local network decreases.) Such a scheme would allow all terminals to employ peak rates of at least 500 characters per second, which is probably the highest rate that would be needed in interaction with a display, where human engineering considerations are allowed to decide the rate, rather than supposed communication costs. Such a multiplexing system depends on high diversity, that is, a low mean rate. Subscribers needing steady high rates would be provided with 'private wires'.

8. Conclusions

Requirements for the communication of digital data in connection with real-time systems are developing rapidly. Some of these systems are, in effect, private networks operated for one organization such as a bank. Some of them, like the multi-access scientific computer services, offer a service to any paying subscriber wherever he is located, but their use at long distances tends to be limited by the cost of communications. There are other requirements of a special kind such as those for meteorological data and the signalling and control requirements of the telephone network itself.

If the common features of all these real-time communication requirements are examined, it is found that they can be met by a single high-level network employing short messages in a store-and-forward system. Such a system, with its fail-soft provisions and its common use of equipment by a large number of users, would be very economic and convenient. The requirements for privacy of certain groups of users can easily be met. Furthermore, the special requirements of groups of users for different bit rates, particular kinds of protocol or particular schemes of routing messages can be provided by a public network, wherever a requirement is sufficiently widespread to warrant the necessary software effort.

Local distribution networks for data based on byte multiplexing are certainly feasible and will provide flexibility for a variety of types of terminal.

The achievement of a public service for data communication, designed for the purpose and not simply adapted from a telephone network, is a problem for which the technical solutions can now largely be understood. The difficulties of establishing a network on an economic basis are still very real ones, because a national network giving widely available access will only pay its way when the number of subscribers and the traffic they generate, hence the revenue, exceeds a certain critical point.

9. Acknowledgment

The work described in this paper was carried out at the National Physical Laboratory, Teddington, Middlesex, England.

Manuscript received by the Institution on 13th March 1968. (Paper No. 1236/Comp. 110).

Points from a Discussion on Communication Systems†

The Chair was taken by Professor G. B. B. Chaplin

Mr. W. P. Nicol: In the classification of organizations needing a real-time computer-based service, Mr. Davies lists in his paper (Section 3 (ii)), the police. This is a correct identification of a need. In the final sentence of Section 3 he states, 'these are the ones that have guided the design of *the* proposed network'. It is these two statements that have inspired my contribution to this discussion.

The characteristics of the provision of civil telecommunications facilities for such public safety/security networks as those of the police, fire or civil defence services in democracies are that:

(a) They have been and are still to a degree planned and, in whole or part, financed by municipal or local government resources. This is not, I emphasize, confined to the U.K. but applies also in many countries of the western world, including the U.S.A.

(b) Superimposed upon these local networks, there are national or federal government networks which provide the necessary integration.

It must be accepted that civil populations in complex societies are now directly exposed to a variety of threats to their way of life and well-being. Thus, it is essential to a democracy that its civil institutions be provided with effective systems of communication, including data transfer, of a type which has been more usually understood in the past to be appropriate to military operations. Disasters can occur as a result of man's fallibility, his deliberate intent or by natural causes. In such circumstances, it is essential to establish control—even by a thin thread—of reliable, survivable communications at the earliest possible moment. If this is not done the scale of any disaster cannot be contained and will escalate.

Mr. Davies has referred to problems associated with the interface of different systems and networks, e.g. digital systems of data transfer with radio networks. He has stated that no two applications are the same. With these observations I am in full agreement. Reference has also been made, as a point for discussion, to the production of a public data transmission network based on advances in telephone techniques using p.c.m. Mr. Davies represents that the production of such a valuable facility need not wait on this development but can be organized now as a national grid, provided the customers indicated will support the project.

Public security organizations will always use, in large measure, normal public communication facilities whether ordinary telephone, telex or more modern data transmission systems; good use might indeed be made of such a national data grid. However, in the two papers which we have just heard, there is perhaps more than an inference that private networks should be discouraged, because covering the country with a mass of private networks would be absurd. I cannot accept that in the context of the police, which has been specifically included in the classification of potential users. It is *not* absurd.

† Following the presentation during the 1968 Convention at Cambridge of the papers 'A communication system for real-time computer systems' by D. W. Davies and 'The impact of pulse code modulation on the telecommunication network' by K. W. Cattermole.

Moreover, I am concerned about trends, already evident, to discourage development of private networks on the hypothesis of some national data grid becoming available at some time in the future. I am sure that these proposals for the 1970s are forward-looking, yet they are already causing concern and may well constitute a brake on the planning of private networks. If there were a hard project for a national data grid, if we knew when it would become available in relation to particular requirements and if we knew how much it would cost, then the customer might weigh and judge. We might here recall with advantage why aviation authorities and airlines, for example, embark upon the provision of their own automatic switching centres and data transmission networks. Their business is essentially to provide transportation. The answer is simple: they have undertaken these projects, financed and developed them, because there was and is nothing else equally efficient for their purpose on offer.

The intelligence content, the need for engineering reliability and the need for survivability, are among the factors to be considered when meeting public security network requirements. These obviously cannot be met in full by any public network, which is almost bound to have its nodes located in areas of high-density population. The determining factors will be existing plant capacity, line availability, finance, and commercial return on investment.

My contribution is therefore not so much a question as an observation on the papers from Professor Cattermole and Mr. Davies. As engineers and scientists, we may rightly consider how desirable it would be if everyone would just get together and use one big system. However, we all live in an imperfect world and I frankly cannot see certain major organizations (such as are listed in the classification referred to) willingly placing all their vital communications in the hands of any organization over which they—the users, have little direct control. Furthermore, there is always the danger that a few disoriented individuals could exert undue influence.

In a variety of ways, private networks will always be required by government organizations and by such services as the police. It is the inclusion of these categories as *public* network requirements that has brought forth this comment.

Mr. D. W. Davies (in reply): Mr. Nicol raises the question of data networks separate from a public data network of the kind described in my paper. It seems that there are three possible reasons why a separate network for a large organization might be needed, security, survivability and time scale.

In the transmission of data it is relatively easy to include moderately good security against interception of the messages by well-known means, based on pseudo-random numbers. This will be sufficient for many purposes such as, I suspect, many of the requirements of the police. Some transmissions will require greater security than these simple measures can provide but, without knowing in detail what is needed, I believe that a public data network could well be employed. There will be obvious advantages in keeping the equipment that needs to be secure out of the network itself. An additional facility that might be considered for security purposes is to give the subscriber some control over the routing of messages if he needs it for this purpose. Components of secure messages could then be sent over different and variable paths if the network is sufficiently over-connected.

Survivability is a more difficult problem and I would agree with Mr. Nicol that a separate 'thin' survivable network would be needed for military and some civil purposes. In including the police in my list of users I had in mind the functions of crime prevention and detection which, if more automation of the information-handling aspects comes about, may generate a moderate amount of data traffic.

The question of time scale is a real worry. Suppose that a firm plan for a data network had been announced, there would still be a number of years of development, planning and installation before it came into use and the problem for organizations with urgent requirements would become more critical. The decision would have to be taken on whether to go ahead with a private network or an interim scheme or to save cost at the expense of delay. It would be absurd, I believe, to allow private networks to proliferate indefinitely and the increasing rate at which these networks are being announced, all with good reasons for going ahead now, underlines the urgency of a decision about a national network. This is not to minimize the difficult market, system engineering and economic questions which arise in making this critical decision.

I must emphasize that my comments on these matters are personal ones and not expressions of official policy.

Professor K. W. Cattermole (in reply): I do not think I have ever suggested that private networks are *absurd*. I have said that they can be *inefficient* and *uneconomic* and that their *limited access* may be a drawback for their own users: so that many potential users of a private network would be better served by a public network, assuming the latter to be adequately provided. Certainly this is suggested by the history of telephone and telegraph services, for which (in virtually all countries, not just the U.K.) a public or common-carrier network has been found the most useful. These public services are extensively used by police, fire brigades, and airlines, among others. I believe that at least some categories of 'data' could usefully develop in the same way.

It must be admitted, however, that (i) the traffic distribution of certain users could be well suited to a private network (though I doubt whether this is true of the police), (ii) security and urgency will sometimes override other factors.

Mr. R. W. Cannon: Can Professor Cattermole please explain the economic comparisons between p.c.m. and traditional f.d.m. systems?

Professor Cattermole (in reply): A more detailed economic comparison between f.d.m. and p.c.m. would require up-to-date information on manufacturing costs of which I cannot claim any special knowledge. The statements in the paper were broadly true when I last attempted any detailed comparison (mid-1967) and I am not aware of any significant change in the balance of costs between the techniques.

U.D.C. 621.37 : 629.113

Thick Films in Automotive Equipment

By
R. W. NOLAN†

Reprinted from the Proceedings of the Joint I.E.R.E.-I.E.E.-I.S.H.M. Conference on 'Thick Film Technology' held at Imperial College, London, on 8th–9th April 1968. Also presented at the Joint I.E.E.-I.E.R.E. Symposium on 'Applications of Microelectronics' held in Birmingham on 27th March 1968.

Summary: The problems facing the manufacture of microelectronic equipment for motor vehicles are outlined. These arise from the physical and electrical environments, the power levels to be controlled which result in high currents, a need for very low saturation voltages, low thermal resistances and efficient heat sinks. These items are then considered with reference to a hybrid integrated circuit alternator voltage regulator, and the design employed to develop an economic unit is discussed. The methods include the use of a printed and fired thick film circuit, with dice-mounted directly to the circuit. The assembly is encapsulated into a combined heat sink and cover. Manufacturing costs are discussed in relationship to those of equivalent electromechanical and discrete component assemblies.

1. Introduction

The year 1967 saw the introduction of microelectronics into the automotive charging systems on both sides of the Atlantic. This is a highly significant step for it means that microelectronics could not only compete with discrete component electronic assemblies but also with the tradition methods of manufacturing electromechanical voltage regulators. The motor industry is highly cost-conscious and will not pay a premium for a new component unless it offers some outstanding technical advantage. Under these latter conditions, the component would be fitted to high performance or luxury cars. The microelectronic regulator described in this paper does not fall into this class. It is intended to control a new range of alternators designed specifically for the average family car, that is, the most cost-conscious area in the automobile industry.

The vibrating contact electromechanical voltage regulator has given long and satisfactory service controlling d.c. generators. Because of the different field current/speed characteristics of an alternator resulting in extremely arduous conditions at the regulator contacts, the decision was made to introduce an electronic voltage regulator. From 1959 to 1963, voltage regulators employed a transistor-assisted contact system.

By 1963, the economics of transistor manufacture were conquered and an all-electronic voltage regulator was introduced. This regulator employed silicon devices throughout the circuit, giving the vehicle manufacturer considerable freedom in the choice of mounting position within the engine compartment.

† Semiconductor Division, Joseph Lucas (Electrical) Ltd., Birmingham 19.

2. Advantages of Electronic Regulators

The advantages of electronic voltage regulators over electromechanical voltage regulators in controlling alternators are:

Improved life

More precise regulating voltage

Accurate control of voltage against variations in temperature

Freedom from adjustment in service

Freedom from the effects of mechanical vibration on regulator operation

A unit of reduced size and weight.

When development commenced on the integrated circuit voltage regulator, the aim was not just to integrate the existing circuit, but to incorporate further improvements to overcome certain problems that had shown up during the manufacture of the electronic regulator and to incorporate features which cost had excluded from the discrete electronic voltage regulator. These improvements included:

Reduced manufacturing costs

More controlled amplifier gain by the use of negative feedback

Limited protection against external faults

Elimination of contacts from the voltage sensing circuit by connecting the regulator sensing terminal permanently to the battery

Temperature compensation to be achieved without the use of thermistors

Reduction in size of the regulator so that it could be incorporated within the alternator.

In order to reduce the power dissipation to a minimum, it was decided to employ a switching mode regulator. The maximum field current to be controlled was 4 A which occurs immediately after switching on at very low temperatures. Under normal stabilized operating conditions, the maximum field current would be 2·5 A. The maximum saturation voltage that could be permitted under stabilized conditions was 0·8 V. Hence, the maximum power dissipation in the control stage during stabilized operation would be 2 W. Immediately after switching on at low temperatures, the power dissipation could be up to 6 W because the saturation voltage of the output transistor would increase.

Thus, the integrated circuit voltage regulator presented a considerable power dissipation problem together with the following environmental problems.

3. Environmental Conditions

3.1. *The Physical Environment*

The under-bonnet environment in a motor car is extremely severe for electronic assemblies. In the case of a voltage regulator which is intended for world wide usage, it was assessed that it must be capable of withstanding the following environmental conditions:

Ambient temperature range −40°C to +110°C with rapid fluctuations in temperature

Humidity to K1007 Method B

Vibration, up to 30 *g*

Contamination from a wide variety of materials including petrol, diesel fuel, detergents, antifreeze, dust, salt, etc.

Sand abrasion

Mycological growth.

3.2. *The Electrical Environment*

Equipment connected to a vehicle's electrical system must be capable of surviving the numerous transient voltages that occur on the system. Transients are generated by wind-screen wiper motor, the ignition system, etc., and the cost of suppressing all the individual units is prohibitive.

The short-term transient voltages present one set of problems, but another condition that can exist during high-rate charging of a discharged battery, is a prolonged high voltage. System voltages up to 50% above normal occur under these conditions. By careful choice of the charging system it is possible to provide protection to the regulator against these conditions.

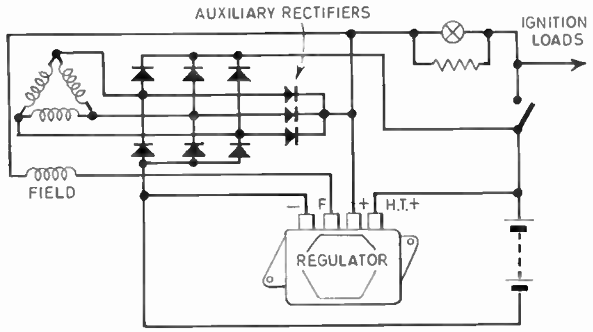


Fig. 1. The complete alternator and regulator circuit.

Figure 1 illustrates the system employed. The regulator is energized from three auxiliary rectifiers which together with the three rectifiers connected to earth for a three-phase full-wave bridge. The sensing circuit in the regulator is high resistance so that prolonged high voltages during high rate charging will not damage the regulator sensing circuit. Because of the isolation provided by the alternator main rectifier and the auxiliary rectifiers, system transient voltages cannot reach the regulator and cause damage.

4. Selection of the Type of Integrated Circuit

The requirements of the integrated circuit for the regulator were assessed as follows:

(i) It must be capable of dissipating power.

(ii) It must become economic to manufacture at reasonably low volumes, thereby minimizing losses during a long slow introduction period.

(iii) It must be capable of being manufactured in slightly different versions thereby permitting the technique to be used across the whole range of regulator applications.

(iv) Rectification must be possible in order to achieve very high yields.

(v) It must be easy to adjust the regulation voltage to very fine limits by varying a resistor.

The system selected was the printed and fired thick film circuit. The monolithic s.i.c. was rejected because it was not compatible with items (ii) to (v) above. Thin film techniques were rejected because of non-compatibility with items (i) and (ii).

Additional advantages from adopting the thick film circuit were that the tooling cost was low and the lead-in time was very short. A disadvantage was that for minimum cost it was necessary to limit the number of resistor screenings to one and use unadjusted resistors.

An overall problem facing the designer of integrated circuit is the very limited range of low-cost capacitors available.

4.1. *Selection of the Type of Semiconductor Devices*

One naturally assumes that planar type devices should be employed with hybrid integrated circuits. In the case of the regulator, investigations had shown that planar devices possessed features which increased the cost of the complete circuit.

A major problem with the switching-mode regulator is the generation of radio frequency interference which can cause severe interference to radio receivers fitted to the vehicle. The problem is greater in Britain than in America for here we have low-frequency broadcast stations which are relatively close to the frequency at which the regulator can switch. When planar transistors are employed, two additional, expensive capacitors are required to reduce the level of radio interference.

Double-diffused homogeneous base transistors were selected because in addition to having relatively slow switching speeds, these devices have excellent second breakdown characteristics which permit them to withstand certain system fault conditions where high power dissipation can occur with voltage present.

In order to minimize cost, it was decided to mount the dice direct to the thick film substrate. Because of the power dissipated within the devices, it was not possible to employ flip-chip techniques. The dice are mounted direct on to the substrate and bonds are made across to lands adjacent to the dice.

5. Circuit Development

At this point, it was possible to commence development of the regulator circuit. The circuit finally developed is shown in Fig. 2. It comprises a three-stage transistor amplifier providing a very high overall gain. The Zener reference diode is connected to a

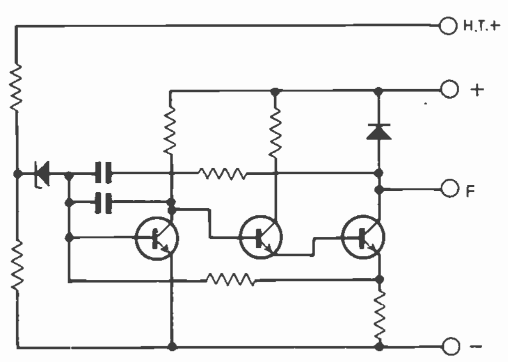


Fig. 2. Basic regulator circuit.

point on the input potentiometer chain. The output transistor is protected from high voltages when the regulator switches off by a field recirculation diode. The voltage developed across the emitter-resistor of the output transistor provides negative feedback which defines the overall gain of the amplifier. This is such a low-value resistor that it is provided by a short length of the normal circuit conductor. The circuit is made to oscillate by providing positive feedback via a C-R circuit from the collector of the output transistor to the base of the input transistor. A Miller capacitor across the input transistor reduces the switching speed of the regulator in order to minimize radio frequency interference. The regulator is adjusted by abrading the lower resistor in the input potentiometer chain.

5.1. *Joining Process Development*

Before the layout of the regulator substrate could commence, the joining techniques had to be decided.

A wide range of conductor inks are available which may be employed with different bonding processes. Gold inks are available for eutectic bonding, while platinum gold and palladium silver are suitable for soldered conditions. The two capacitors employed in the circuit had silver electrodes and had to be mounted by soldering. This dictated that the conductor pattern must be suitable for soldering. Palladium silver was selected on the basis of cost and the better conductivity it possesses over platinum gold.

As one set of bonds had to be soldered, it was eventually decided to adopt soldering for all the connections. This avoided an additional screening of conductor for eutectic bonding.

5.2. *Regulator Design and Assembly*

With the circuit development completed and the bonding processes established it was possible to lay out the substrate.

The power dissipation in resistors was limited to 10 W/in^2 (about 1·5 W/cm^2) where very high stability was required. In other resistors, the power dissipation was increased to 40 W/in^2 ($\sim$6·2 W/cm^2).

The substrate layout is shown in Fig. 3. All the external connections are brought out to one edge. As many resistors as possible were orientated in the same direction in order to achieve the best 'as screened' resistance tolerance. A standard length of 0·15 in (3·8 mm) is adopted for internal bonds. The resistor that is abraded to set the regulator is placed as close as possible to the reference corner. This resistor can be abraded from 800 Ω to over 2000 Ω.

The substrate assembly operation consists of attaching pretested semiconductor chips and the capacitors to the substrate and making the internal connections.

The whole substrate is then tested. Should a fault be indicated, the substrate is inserted in automatic fault analysis equipment which indicates where the

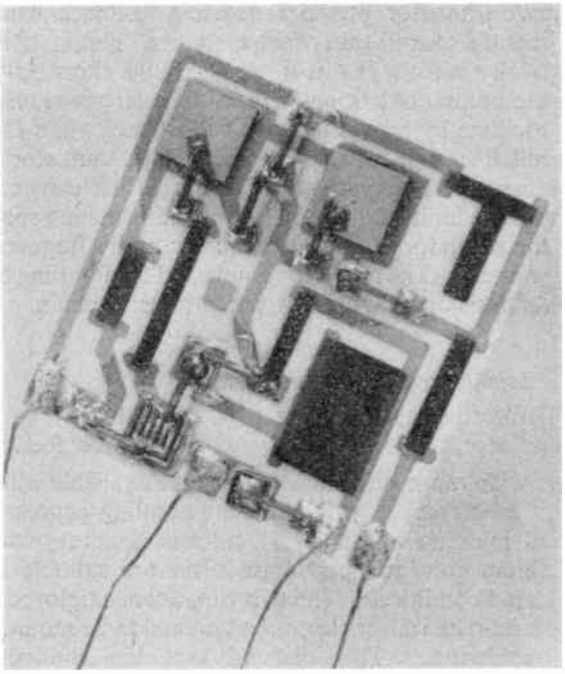


Fig. 3. The thick film regulator unit.

fault is located. The fault would then be rectified and the substrate retested. Substrates which do not have a fault or which have been rectified next have the make-off leads attached to them. This is followed by protecting the dice by a layer of resin. The setting resistor is then abraded so adjusting the regulator to the correct regulating voltage. This is followed by fixing the substrate down to the heat-sink, encapsulating in silicone rubber, welding the blade connectors to the make-off leads, then riveting the blade connector supports and bottom cover to the heat-sink. The regulator then receives its final testing operation.

6. Relative Manufacturing Costs

It is very difficult to compare the manufacturing cost of regulators made by different techniques for features are frequently incorporated within the regulator which are dictated by other parts of the system. An example of this would be components which are part of the warning-light system. Another confusing factor is the difference that exists between the specifications of regulators made by different techniques. A regulator made from discrete components has a higher specification than an electro-mechanical vibrating contact regulator, while the regulator described in this paper has a higher specification than the discrete component regulator. The following relative cost values are based on the assumption that the regulators are built to an average specification for the type of assembly employed and only refer to the voltage regulator function.

	Relative cost
Vibrating contact electro-mechanical regulator	100
Discrete component electronic regulator	180
Hybrid integrated circuit regulator	120

These costs are slightly misleading, for such a low-cost mechanical regulator would probably have an unsatisfactory service-life. If the specification of the mechanical regulator was such that service-life should be comparable with that of the electronic regulators, the manufacturing cost relationship becomes:

Vibrating contact electromechanical regulator	100
Discrete component electronic regulator	150
Hybrid integrated circuit regulator	100

One other comparison is of interest, namely the relative cost of manufacturing regulators by the various techniques to the hybrid integrated circuit regulator specification. As a mechanical regulator could not be made to meet this specification, the comparison is limited to the other two regulator types.

Discrete component electronic regulator	100
Hybrid integrated circuit regulator	40

All the above figures are estimates based on similar manufacturing quantities.

The last two groups of figures indicate that as the specification and hence the complexity of a regulator increases, so the hybrid integrated circuit approach becomes more attractive.

7. Other Applications

The next application of hybrid integrated circuits will probably be to the ignition system. In this application the voltage and power levels to be controlled are very much higher than those of the regulator.

In the future, automatic speed controllers, windscreen wiper control systems and control of a petrol injection system may be performed by electronic means. However, this will only happen if the electronic solution is the most economic.

8. Acknowledgment

The author wishes to thank the Directors of Joseph Lucas (Electrical) Limited for the permission to publish this paper.

Manuscript received by the Institution on 30th January 1968. (Paper No. 1237/CC36.)

U.D.C. 518:621.37

A Fast Binary to Logarithmic Analogue Converter

By

J. A. LADD, B.Sc.

(*Graduate*)†

Summary: The paper describes a circuit configuration which produces an output proportional to the logarithm of binary numbers presented at its input. The circuit is designed to process the first six significant digits only and can deal with binary numbers up to $2^{20}-1$.

The maximum output settling time, to within $\frac{1}{2}$% of the final output level, is about 2 μs. The resolution of the circuit is limited by the number of digits taken; in the unit described it varies from one part in 32 to one part in 64 depending on the value of the binary number. This uncertainty amounts to a maximum of 0·22% of full-scale deflection. Other inaccuracies are small compared with the resolution errors. The circuit employs digital switching and uses linear components only. No devices with special characteristics are required.

1. Introduction

The converter‡ to be described in this paper is designed to produce the logarithmic analogue of a binary number within approximately 2 μs from the time of presentation of the number to the converter. It is intended mainly for use in display, plotting or hybrid analogue computing systems where a precise logarithmic presentation is required. This form of presentation is most useful where it is necessary to compare the shapes of histograms such as are obtained in biological, statistical or nuclear studies. The shape of a logarithmic histogram does not change with the amount of data collected; the histogram merely rises up the scale, becoming smoother as the data become more refined. Histograms of different scale can therefore be superimposed and compared merely by shifting the histograms vertically until they coincide.

2. Logarithmic Devices

Various logarithmic elements have been used in systems designed for display and analogue computing purposes. These elements include valves with special voltage–current relationships between certain electrodes, p-n junction diodes[1] and the emitter-base junctions of transistors.[2]

The difficulty with most of these elements is that they are only accurately logarithmic over a limited range. For example, most logarithmic diodes do not maintain reasonable accuracy over more than four decades. In addition, temperature variations cause shifts in d.c. position and slope, particularly with the semiconductor elements, and temperature control has to be incorporated to produce a stable output.[3]

† EMI Electronics Ltd., Wells, Somerset.
‡ The device is the subject of U.K. Patent Application No. 31204 of 1968.

In the design presented in this paper, the aim has been to avoid the use of any elements with special characteristics, and to eliminate the need for temperature control of any part of the circuit.

3. Mode of Operation

The method of obtaining the logarithm of a binary number is similar in principle to that employed for a decimal number. The logarithm consists of two parts, the characteristic and the mantissa, and in the circuit described these two parts are separately evaluated.

The characteristic is obtained by determining the position of the first significant figure within the number. The characteristic is then equal to one less than the number of significant figures. This is illustrated by the examples in Table 1.

Table 1

Binary logarithmic characteristics

Binary number	Significant figures	Characteristic
1	1	0
10	2	1
100	3	2
1000	4	3
—	—	—

To generate the analogue of the characteristic therefore we merely need to know the number of significant figures presented. If a current I is generated for each significant figure except the first, a correct analogue for the characteristic may be obtained. This is shown in Fig. 1. In the application envisaged, all numbers will be integers and no negative characteristics are required.

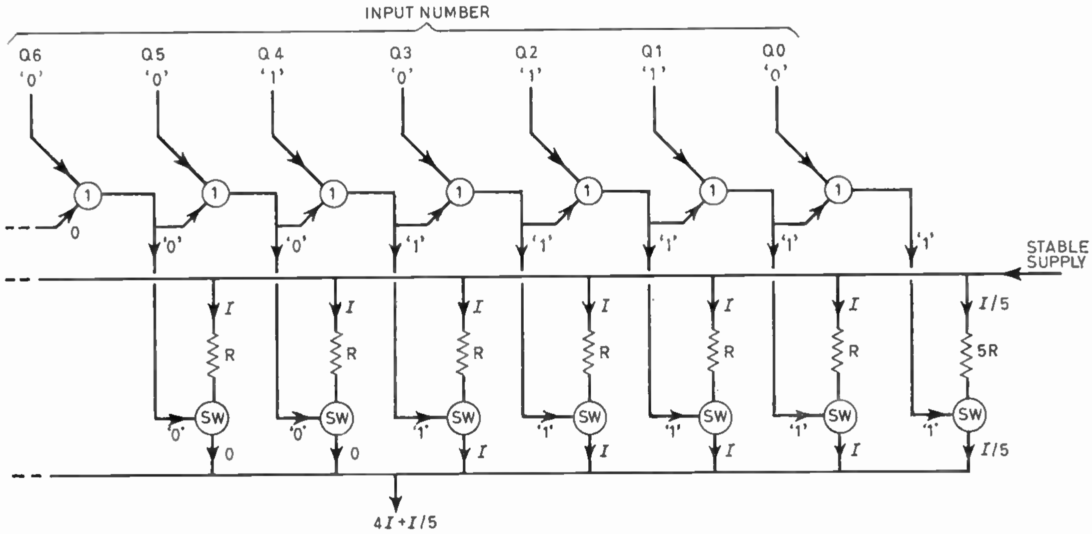


Fig. 1. Derivation of characteristic analogue.

It is seen that with OR gates and resistors arranged as shown, an output of $4I+I/5$ is obtained. The last term $I/5$, is arbitrary and is generated merely to distinguish the integer 1 from an all-zero input number. This is necessary because the logarithm of 1 is 0 and theoretically no current should be generated for this number.

An ambiguity thus arises because an all-zero number also generates no current and would normally be indistinguishable from 1. The additional current $I/5$ is therefore introduced and the effect of this is to start the display deflection at a distance of one-fifth of a characteristic increment from the true zero of the display. The ambiguity is therefore eliminated.

Ignoring the current $I/5$, which is generated for any non-zero number, the output current from the circuit of Fig. 1 is $4I$. The correct characteristic is 4, and from an examination of the circuit it is apparent that the output is always $(N-1)I+I/5$, where N is the number of significant figures presented at the input.

This particular circuit can therefore be used for characteristic generation, and can be adapted to deal with numbers having a radix other than binary if required.

The derivation of the characteristic of the logarithm is therefore logically quite simple. To obtain the complete logarithm it is necessary to add the mantissa. This is a more complicated procedure, and the following section describes the method adopted.

4. Mantissa Figure Selection

The first six significant figures are taken into account in the generation of the complete logarithm. In binary code the most significant figure is always a 1. It is not necessary therefore to consider this figure in the generation of the mantissa. The five figures following the first are used for this purpose. These figures must be separated out as a block from the complete number.

The logical arrangement for the selection of the five figures is shown in Fig. 2. The logical state of the circuit is indicated with a typical input number for the purpose of illustration.

The first logic layer is that used for characteristic generation, as shown in Fig. 1. Layer 2 (Fig. 2) consists of a row of inverters and layer 3 consists of 2-input AND gates. The logical arrangement is such that the output from a gate in layer 3 can be a '1' only if the gate lies in the column where the outputs from layer 1 change from a '0' sequence to a '1' sequence, i.e. if the gate lies in the same column as the most significant figure.

There is therefore only a single '1' output from layer 3. All other outputs are at '0'. In this case the '1' output emerges from layer 3 in the Q7 column.

The outputs from layer 3 are connected to an array of 2-input AND gates, arranged in columns. Now each of the digit inputs from Q0 upwards is connected diagonally to five successive AND gates in the array. For example, input Q2 is connected to gates in column 3 layer 4, column 4 layer 5, column 5 layer 6, column 6 layer 7 and column 7 layer 8.

The outputs of the AND gates in layers 4, 5, 6, 7 and 8 are connected as shown in Fig. 2 to the distributed OR gates on lines A, B, C, D and E respectively.

Examining the conditions in the array (layers 4 to 8), it can be seen that the only gates which can possibly produce an output of 1 are in the Q7 column. All

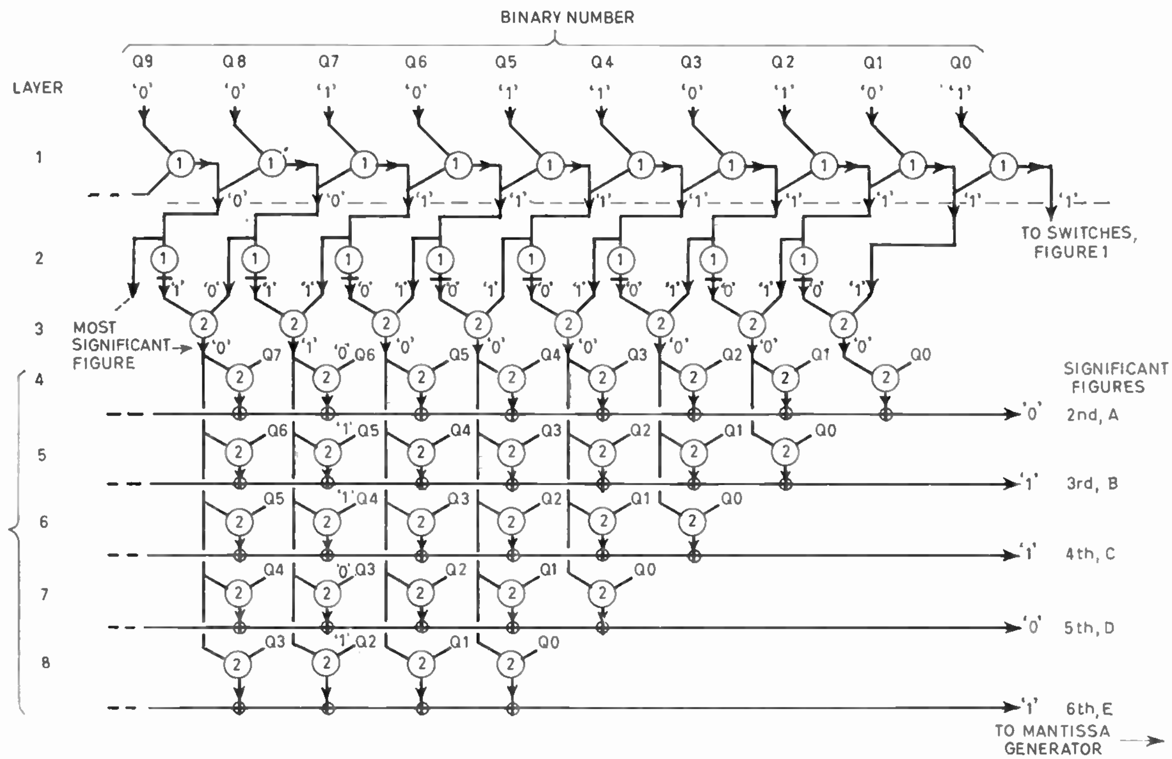


Fig. 2. Mantissa figure selector.

other gates are disabled by the 0 outputs from layer 3 and their outputs are therefore also at 0.

Considering column 7 only then, the outputs from the AND gates in this column will be controlled by Q6, Q5, Q4, Q3 and Q2 because of the diagonal arrangement of the input connections to the array. The states of the OR gate lines A, B, C, D and E will therefore also be similarly controlled.

The outputs from the lines A to E are therefore the same as the input digits Q6 to Q2. The five significant digits following the most significant digit have therefore been extracted by the array. In the example shown the outputs ABCDE will be 01101. These significant digits are passed to the mantissa generator, which is now described.

All other digits of lesser significance in the input number are ignored.

5. Mantissa Generator

The block of five digits fed to the mantissa generator can range from 00000– – – to 11111– – –. The number of discrete steps in the mantissa output is therefore $2^5 - 1$ or 31 and the main objective of the mantissa generator design must be to ensure that each of these steps lies as close as possible to the ideal logarithmic curve.

The method chosen involves the selection of two groups of fixed points over the range of 31 steps. Interpolation between these points is then arranged by means of a cascaded switching network.

The first set of fixed points is selected by choosing currents which can be switched in directly so that the five most obvious outputs are correctly generated. The numbers corresponding to these points, and the currents required, are set out in Table 2.

The currents in the table are marked on the mantissa curve in Fig. 3.

It is not possible simply to add current contributions together as in a linear converter, since the slope of the mantissa curve changes, the slope of the upper portion being only half that of the lower portion.

This decrease of slope with increase in numerical value implies that the current contribution of any particular digit must be reduced if other more significant digits are present. Furthermore, this reduction must be cumulative, so that the presence of any digit will modify the current contribution of all following digits.

One way of achieving this modification of current contributions would be to take a linear digital-to-analogue converter, and arrange that the presence of a '1' at position A, for example, would reduce the

Table 2

Mantissa current cardinal points using fixed currents

First figure	Mantissa figures	Decimal equivalent	Current increment required
- -001	10000- -	48	$\log_2 48 - \log_2 32 = 0{\cdot}5850 = I_A$
- -001	01000- -	40	$\log_2 40 - \log_2 32 = 0{\cdot}3219 = I_B$
- -001	00100- -	36	$\log_2 36 - \log_2 32 = 0{\cdot}1699 = I_C$
- -001	00010- -	34	$\log_2 34 - \log_2 32 = 0{\cdot}0875 = I_D$
- -001	00001- -	33	$\log_2 33 - \log_2 32 = 0{\cdot}0444 = I_E$
- -001	00000- -	32	$I = 0{\cdot}0$

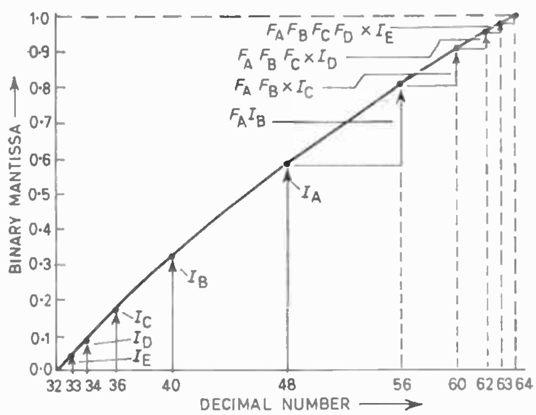


Fig. 3. Mantissa curve showing fixed points.

contribution made by a '1' at position B. A simple circuit which could be used for this purpose is shown in part in Fig. 4. The resistors, RA1, RA2 are chosen so that when switch SA is operated, a current I_A flows out of the 'mantissa current output' terminal.

At the same time, the voltage at the junction of RA1 and RA2 is reduced from V_R to a lower voltage, determined by the ratio of the two resistors. This reduced voltage is fed through a buffer to the next section of the converter. Section B will therefore make a reduced contribution to the total output current when a digit is present at position A.

In semiconductor practice it is most convenient to use a diode current switching pair in place of switch SA, which implies a continuous current flow through resistors RA1, RA2. The voltage at the junction of RA1, RA2 would therefore not vary with the digit presented to the switch SA.

This difficulty is overcome by adopting the slightly more complicated configuration shown in Fig. 5. Here the current-controlling and voltage-controlling functions are separated, so that two switching components are required for each digit except the last. The first of the two switch components, such as SA1, controls the current fed to the 'mantissa current output', and the second, SA2, controls the voltage fed forward to the next section, such as B.

The operation of the circuit shown in Fig. 5 is as follows:

Switches SA to SE are controlled by the outputs of the mantissa figure selector of Fig. 2, and resistors RA to RE are chosen to pass the currents I_A to I_E from the reference voltage source V_R to the output line whenever any one of the switches is operated.

The buffers BU merely reproduce at low impedance the voltages at their inputs.

Let the first mantissa digit be a 1. Switch SA is then operated. A current I_A will flow through contact SA1 to the mantissa output line. Similarly, any other

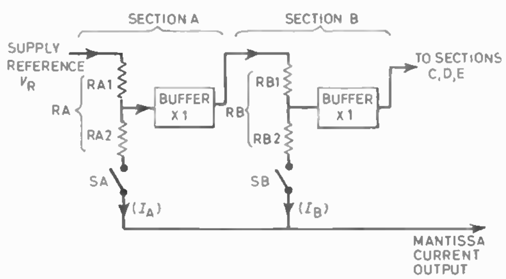


Fig. 4. Elementary network of a mantissa generator.

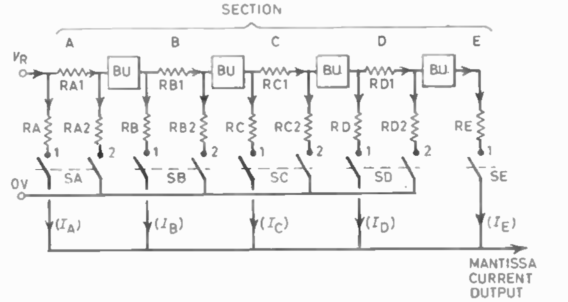


Fig. 5. Complete cascaded modifying network of a mantissa generator.

Table 3

Mantissa current cardinal points using modifying factors

First figure	Mantissa figure	Decimal equivalent	Current increment required
--001	11000---	56	$\log_2 56 - \log_2 48 = 0{\cdot}2224 = F_A \times I_B$
--001	11100---	60	$\log_2 60 - \log_2 56 = 0{\cdot}0995 = F_A F_B \times I_C$
--001	11110---	62	$\log_2 62 - \log_2 60 = 0{\cdot}0473 = F_A F_B F_C \times I_D$
--001	11111---	63	$\log_2 63 - \log_2 62 = 0{\cdot}0231 = F_A F_B F_C F_D \times I_E$

switch, when operated alone, will give a current output as listed in Table 2.

Let switches SA and SB now be operated together. Current I_A will flow, but the current I_B will be modified by the action of contact SA2. Current I_B is thus reduced by a factor F_A where

$$F_A = RA2/(RA1+RA2)$$

Similarly, switch SB can modify I_C by a factor F_B, where $F_B = RB2/(RB1+RB2)$.

All elements are cascaded and, as an example, let switches SA, SB and SC be operated together. The output will then be:

$$I_{OUT} = I_A+(F_A.I_B)+(F_A.F_B.I_C)$$

The modifying process extends right through the network to section E and the introduction of modifying factors allows the selection of another set of points on the mantissa curve. The points selected are listed in Table 3, and are marked on the curve in Fig. 3.

From Table 3 the factors F_A to F_D may be found. Thus:

$$F_A = 0{\cdot}2224/I_B = 0{\cdot}6908$$
$$F_B = 0{\cdot}0995/F_A \times I_C = 0{\cdot}8478$$
$$F_C = 0{\cdot}0473/F_A F_B \times I_D = 0{\cdot}9234$$
$$F_D = 0{\cdot}0231/F_A F_B F_C \times I_E = 0{\cdot}9618$$

In the circuit of Fig. 5 the values of factors F_A to F_D are set by suitable choice of resistors RA1, RA2 to RD1, RD2.

Having fixed the values of I_A to I_E and F_A to F_D, there are now nine cardinal points which lie exactly on the required mantissa curve. The accuracy of interpolation for other possible numbers within the mantissa interval is not so far known.

Using the values of I and F already determined, the outputs obtained for other combinations of the five mantissa digits may be calculated. This has been carried out on a computer and the complete results are plotted in Fig. 6.

The points on Fig. 6 are plotted simply in the form of departures from the ideal curve. The fixed points lie by definition on the ideal curve and exhibit zero error. However, the interpolated points all exhibit positive errors and the largest of these appears at

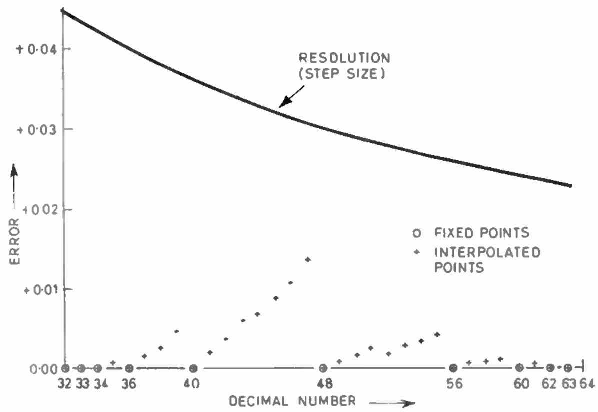


Fig. 6. Mantissa interpolation errors.

the mantissa combination 01111, corresponding to the decimal number 47. The error here is $+0{\cdot}0133$ with respect to the ideal curve and amounts to 0·437 of the unit step at that point.

From the curve it can be seen that the interpolation errors are generally reasonably small compared with the resolution. The resolution is worst at the beginning of the curve where a unit step is equivalent to a current increment of 0·0444, whereas the maximum interpolation error is 0·0133.

The mantissa generation network shown in Fig. 5 is therefore theoretically capable of producing outputs of more than sufficient accuracy for the purpose envisaged.

It is nevertheless apparent that the network could not profitably be extended to deal with further digits, since the maximum interpolation error would then become comparable with the size of the associated unit step.

6. Practical Circuits

6.1. *Construction*

A complete logarithmic converter has been constructed using integrated circuits for all logical operations. A binary number of up to 20 digits can be presented at the circuit inputs.

The 'characteristic generator' of Fig. 1 and the 'mantissa figure selector' of Fig. 2 are combined and the whole structure is made up of NAND gates. The use of this type of gate involves several interlayer logic inversions, hence the elements of Fig. 2 require a total of 10 layers of 2-input NAND gates in the unit constructed.

To obtain the characteristic current, the outputs of the characteristic generator are used to actuate diode switches, controlling currents of 250 μA per switch. These currents flow into an output amplifier which incorporates feedback to produce a low input impedance. Since there may be up to 20 input digits, the full-scale current output is 20 × 250 μA, that is 5 mA. This current produces an amplifier output swing of 0 to +5 V at an impedance of 100 Ω.

The mantissa generator is constructed using discrete components. The reference voltage V_R of Fig. 5 is −10 V and the buffers BU are simple emitter-followers. The contacts such as SA1 are diode pairs which route the currents flowing through resistors such as RA into the current output line. Contacts such as SA2 are 'on-off' common-emitter transistor switches.

The total mantissa excursion is adjusted to the correct scale with respect to the characteristic increment by setting the reference voltage V_R. The latter is adjusted to give a monotonic transition at the output as the input number changes from, say, 111111 (decimal 63) to 1000000 (decimal 64).

The mantissa current is added to the characteristic current at the input to the low impedance amplifier.

The Q0 input is arranged to make a small contribution of 50 μA, or $I/5$ in Fig. 1, to the output current. This displaces the whole scale upwards by 50 mV whenever any input number other than zero is presented. This eliminates the ambiguity described in Section 3.

6.2. *Calibration*

Seven inputs are provided which allow the user to calibrate the converter for use on decimal displays or plotters. The inputs are labelled from 10^0 to 10^6 and logic elements are provided within the unit first to generate a 1 at the appropriate point in the 'characteristic generator' and second, to present an artificial combination of figures to the 'mantissa generator'.

For example, taking the decimal 10^2 calibration input, a 1 level applied to this input produces a 1 level at the Q6 position in the 'characteristic generator', Fig. 1, giving a characteristic current output of 6 units plus the small non-zero offset previously mentioned.

At the same time, other gates are arranged to present the combination 10010 to the mantissa generator, Fig. 5, and the current generated is added to the characteristic current to produce the complete analogue of the logarithm to the base 2 of decimal 10^2.

The converter logic is thus set up internally in the same way as if the binary number 1100100 were presented at its input. This number represents, in decimal form, 64+32+4 or 100.

All the other decimal calibration points are synthesized in a similar way.

7. Performance

7.1. *Speed*

A plot of the output waveform showing the speed of response is given in Fig. 7. When a large binary number is presented to the input, the logical sequence of operations may have to traverse as many as 40 NAND gates, and this accounts for the relatively long logical settling time t_l. This may amount to 0·6 μs.

The rest of the response lag is caused by the finite rise time of the amplifier which is about 1 μs from the 10% to 90% points.

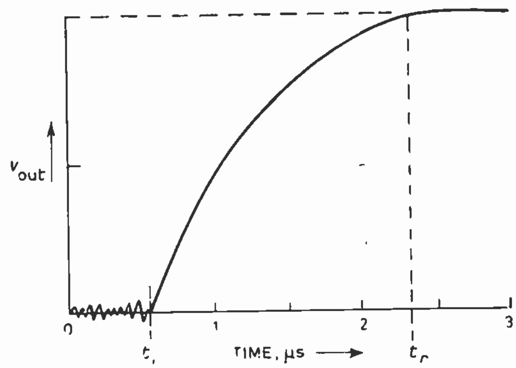


Fig. 7. Transient response of a logarithmic converter.

The total time taken to reach a final deflection, say to within one-half of a cathode-ray tube spot width, is about 2 μs.

7.2. *Accuracy*

As previously stated, the main limitation on the accuracy is the resolution, which is determined entirely by the number of significant figures processed. In this case, where binary numbers up to 20 digits long may be presented, the uncertainty caused by selection of the first six digits only is 0·222% of full scale deflection.

Static tests show departures from ideal output of 0·2% of full scale over most of the range. At very nearly full-scale deflection, the error is as high as 0·5% of full scale, and this is believed to be caused by slight non-linearity in the output amplifier.

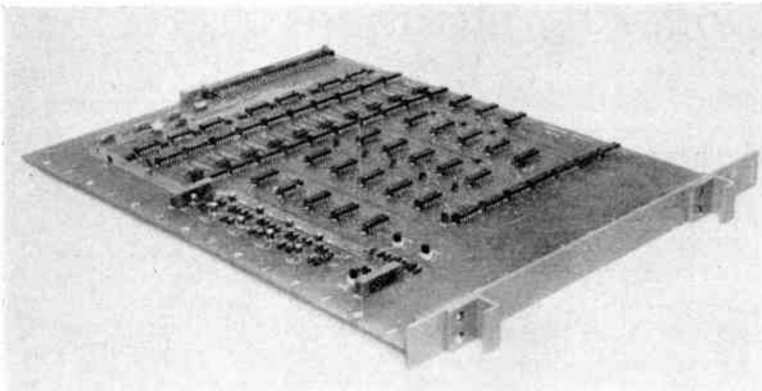


Fig. 8. Logarithmic converter module.

8. Physical Construction

The micrologic circuits, which are quadruple 2-input NAND gates in dual-in-line packages, are mounted on a double-sided epoxy fibre-glass printed board. In all, 57 packages are mounted, together with discrete components for the mantissa generator and output amplifier. The whole assembly is designed as a single module which can be plugged into a chassis for connection to a digital display, plotting or computing system.

A photograph of the converter module is shown in Fig. 8.

9. Conclusion

A fast logarithmic converter of sufficient precision for use in accurate display, plotting and computing systems has been designed.

The interpolation method used has been shown to be satisfactory for the number of digits processed. A practical version of the unit has been constructed and its speed and accuracy demonstrated.

10. Acknowledgments

Thanks are expressed to Mr. S. A. Hickman for his encouragement during the design of the converter, to Mr. M. J. Crockett for carrying out detailed testing, to Mr. L. R. Bulmer for layout work, and to Mr. B. Burton for programming advice.

Thanks are also extended to the Directors of Nuclear Enterprises Limited, EMI Electronics Limited and to the U.K. Atomic Energy Authority for permission to publish this paper.

11. References

1. Cowell, T. K. and Gordon, M., 'LOPAD, a logarithmic on-line processing system for analogue data', *Electronic Engng*, **37**, p. 146, March 1965.
2. Deighton, M. O., Sayle, E. A. and Kestin, C. S., 'Analogue pulse arithmetic using logarithmic p-n junctions', *Proc. Instn Elect. Engrs*, **113**, p. 225, February 1966.
3. Ladd, J. A. and Davidson, A. J. S., 'A Data Plotting System for use with Punched Paper Tape', Chalk River Report, No. CREL–1009, May 1962.

Manuscript received by the Institution on 13th September 1968. (Paper No. 1238/Comp. 111.)

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.

IMPROVED CIRCUIT TECHNIQUE FOR F.E.T.

The field-effect transistor (f.e.t.) has the characteristic feature that the input resistance is much higher than that of other semiconductor devices. At low frequencies the f.e.t. input impedance is generally high and the output impedance low. The current through the device is due to the transport of only majority carriers. The cut-off frequency of the f.e.t. is rather low because larger input capacitance is present, and also the characteristic of high input resistance disappears at high frequencies. The input capacitance is subject to increase to obtain larger transconductance, g_m. These faults of the f.e.t. can be improved when it is used in a travelling-wave circuit of either the constant-*k* or *m*-derived type. Using the travelling-wave method, the degree of freedom in design can be increased, and this method is very useful for fabricating an integrated circuit with the f.e.t., since the circuit used in this case is very simple. In a Japanese paper, the design methods used in the travelling-wave type of circuit are given and good agreement between theory and experimental results is shown. From the design theory, the upper-frequency limit of the f.e.t. was deduced using the structure and material constants of the f.e.t. as parameters, and the minimum number of f.e.t.s to obtain the necessary voltage gain was investigated as a function of bias voltage.

'Travelling-wave field-effect transistors', A. Abe, H. Tango, H. Furukawa and J. Nishizawa. *Electronics and Communications in Japan* (English language edition of *Denshi Tsushin Gakkai Zasshi*), **50**, No. 12, pp. 1–10, December 1967.

EARTH-STATION AERIALS FOR TELEVISION BY SATELLITES

Within the framework of television transmission by satellite, indirect distribution, that is to say, the exchange of programmes between Earth-stations connected to the television broadcasting station by terrestrial circuits, very probably constitutes the next stage of development to be achieved, following upon the satellite point-to-point circuits such as the *Intelstat* already in use. This will be followed in due course by direct distribution to the broadcasting transmitters, and after that transmissions from satellites intended for direct reception by small communities, or even individual viewers themselves.

A paper written jointly by three engineers working with the Group d'Etudes de Technologie Spatiale, S.p.r.l., Brussels, describes such indirect distribution systems by means of some typical examples of the geographical coverage of Europe, Africa, the Middle East and the eastern part of the Americas that can be achieved by means of geostatic satellites. The factors that have to be taken into account when designing the receiving aerials for the Earth-stations, when satellites of that type are used for the indirect distribution of television programmes (power balance, characteristics of the f.m. radio circuit, quality standards to be attained) are considered. For each of the systems described, the results are given in graphical form on the basis of power outputs of the satellite transmitter between 10 and 100 W using carrier frequencies in the 4, 8 and 12 GHz bands.

The paper then examines the performance of the system on the basis of the characteristics of the satellite-to-Earth link only, disregarding the possibilities of interference arising from the existence of terrestrial radio-relays using the same frequency bands or from the existence of other satellites in the same region of the equatorial belt.

'Distribution of television by satellite with special reference to the size of the Earth-station aerials', A. Benoit, H. Godfroid and P. Kuypers, *E.B.U. Review*, **110-A**, pp. 162–72, August 1968.

MULTI-PATH PROPAGATION IN P.C.M. SYSTEMS

Pulse code modulation communication techniques have been studied extensively in recent years in conjunction with line-of-sight microwave transmission. Such systems are quite different from f.m. systems and possess distinct advantages with respect to noise and crosstalk. However, in the use of free space as transmission medium a number of problems, such as multi-path propagation due to ducts, ground reflection and attenuation of the transmitted wave due to rain, are encountered. Such phenomena coupled with the filter frequency bandwidth characteristics produce intersymbol interference or interchannel interference, and impair the noise tolerance. Therefore, in order to design p.c.m. radio communication systems it is necessary to analyse the relationship between waveform distortion due to multi-path propagation and the band-pass characteristics and the error rate in the presence of fading.

Multi-path propagation in microwave p.c.m. systems has considerable effect on the mean error rate. A paper from the Japanese N.T.T. Company assumes a Gaussian-type filter and two-path model in examining the wave distortion due to multi-path propagation and the error rates for three different types of modulation and demodulation are found as functions of the filter bandwidth, bit rate, and the difference in mean propagation time between the primary and secondary waves. The results not only show clearly the effect of the propagation distortion on mean error rate but also show the possibility of determining the maximum bit rate for various ranges of parameters and given circuit specifications.

'Analysis of error rates due to multi-path propagation in microwave p.c.m. systems', F. Yoshida, *Electronics and Communications in Japan* (English language edition of *Denshi Tsushin Gakkai Zasshi*), **50**, No. 11, pp. 80–89, November 1967.