The Journal of the

BRITISH INSTITUTION OF RADIO ENGINEERS

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 24 NOVEMBER 1962 NUMBER 5

COLLABORATION BETWEEN ENGINEERS

MANY cogent arguments have been advanced in favour of qualified engineers working more closely together. Not the least important is that of the status of the engineer—a title which is so often wrongly used today. The setting-up of the Engineering Institutions Joint Council (E.I.J.C.) is therefore a step to be warmly welcomed as providing a means whereby these and other problems may be discussed. The Joint Council comprises the thirteen major British professional engineering bodies whose combined membership exceeds 200,000; a note about its organization appears on page 342. The need for engineers to be able more easily to exchange information is obvious and this will certainly be one of the Joint Council's early tasks.

Comment has been made in the Technical Press as to whether the definition of an engineer will be attempted by the E.I.J.C. and this could well lead to considering the vexed question of the registration of engineers—a matter which has frequently been referred to in this *Journal*. Codes of professional conduct, acceptance of common examinations in fundamental subjects, and standards for entry to the engineering institutions will no doubt also be discussed by the new Council.

During the past decade there have been set up several co-ordinating bodies within the overall field of engineering whose purpose has been to establish channels for collaboration and communication between engineers who have a common interest. Notable features of modern technological progress have been developments which may be regarded as being "horizontally" organized. For instance, development and applications of automation, computers, non-destructive testing and nuclear energy call for contributions from professional engineers of many different specializations. On a wide basis and bringing in the users of technical equipment some progress has been made in the last decade by the establishment of such co-ordinating bodies as the British Conference on Automation and Computation, the British National Committee for Non-Destructive Testing, and the British Nuclear Energy Conference (now Society).

The principal function of these co-ordinating bodies to date has been to sponsor meetings and thus encourage the exchange of views not only between engineers but also with users. The need for wider dissemination of such information has been met, usually, by the publication of proceedings of conferences and symposia under the aegis of one or other of the participating institutions and the risk of proliferating new journals thereby avoided.

The national co-ordinating bodies are well suited to ensuring adequate British representation at international conferences. This international role is a most important function of the co-ordinating bodies, since such international conferences as those held in 1960 on Non-Destructive Testing and on Automatic Control provided unrivalled opportunities for engineers of all nationalities to meet and discuss new discoveries and developments in their specialized fields.

The radio and electronic engineer has a particularly important contribution to make to the effectiveness of these co-ordinating bodies, since the techniques of radio and electronics have extensive "horizontal" applications to many diverse fields of engineering. The Institution's links with these bodies are therefore essentially two-way in the advantages brought about, all of which promote the development of the engineering profession in this country.

G. D. C.

B

INSTITUTION NOTICES

Birthday Honours List

The following reference was inadvertently omitted from the June *Journal*.

The Council of the Institution has congratulated the following member whose name appeared in Her Majesty's Birthday Honours List:

Major Frederick Vernon Baines, Royal Australian Corps of Signals, on his appointment as an Ordinary Member of the Military Division of the Most Excellent Order of the British Empire.

Major Baines is at present a technical staff officer with the Australian Army Design Establishment. He spent 13 months in Antarctica as leader of one of the four expeditions sent out by Australia during the International Geophysical Year, and his appointment as M.B.E. recognizes this service.

He was elected an Associate Member of the Institution in 1954.

Professional Engineers and the Common Market

A Conference on "Professional Engineers and the Common Market" is being arranged by the Engineers' Guild in London on Thursday, 6th December. Its purpose is to provide information and encourage thought about the issues raised by the prospect of membership of the Common Market, as they affect the engineering profession of the United Kingdom. It is intended for professional engineers, for those who use their services and for those concerned with their education and training.

Held at the Connaught Rooms, Great Queen Street, W.C.2, the chair will be taken by Sir Hugh Beaver, K.B.E., M.I.C.E., M.I.Chem.E., President of the Guild who will open the Conference. Subsequent speakers and the subjects of their addresses will be:

"Colleagues or Competitors"—Sir Walter Puckey, M.I.Prod.E., Chairman, Management Selection Ltd.

"Education for Europe"—Sir Willis Jackson, F.R.S., Professor of Electrical Engineering, Imperial College of Science and Technology.

"Challenge and Opportunity"—Dr. F. Q. den Hollander, Chairman of the Board, Delft University, Holland.

There will be ample time for discussion.

The Conference fee is four guineas for members of the Guild and five guineas for non-members, to include morning coffee, lunch and tea, and a printed copy of the proceedings which will be sent afterwards to all those attending.

Applications for accommodation, which is limited to 200 places, should be sent, with Conference fee, to The Secretary, Engineers' Guild Limited, 201 High Holborn, London, W.C.1.

The Engineering Institutions Joint Council

The leading British engineering Institutions have announced the formation of the Engineering Institutions Joint Council. The Joint Council will be under the Chairmanship of Sir Kenneth Hague, LL.D., M.I.Mech.E. The combined membership of the thirteen Institutions represented is well over 200,000.

The purpose of the Joint Council is to enable representatives of the bodies concerned to meet together to consider and take action upon matters of common interest relating to the advancement of engineering and the dissemination of knowledge in that field. The Joint Council also intends to establish a sound and useful channel of communication and understanding between the professional engineers, the Government, the public, and national and international bodies concerned with engineering.

The bodies represented on the Council are:

The Royal Aeronautical Society;
The Institution of Chemical Engineers;
The Institution of Civil Engineers;
The Institution of Electrical Engineers;
The Institution of Gas Engineers;
The Institute of Marine Engineers;
The Institution of Mechanical Engineers;
The Institution of Mining Engineers;
The Institution of Mining and Metallurgy;
The Institution of Municipal Engineers;
The Institution of Production Engineers;
The British Institution of Radio Engineers;
The Institution of Structural Engineers.

The Joint Council has appointed as its Honorary Secretary Mr. K. H. Platt, M.B.E., B.Sc., M.I.Mech.E., and it will have its office in The Institution of Mechanical Engineers, 1 Birdcage Walk, Westminster, S.W.1.

Comments on the establishment of the E.I.J.C. and other co-ordinating bodies in specialized fields of engineering are made in the Editorial on page 341.

Corrections

The following amendments should be made to the paper "Automatic Scanner Loggers for Small Installations", published in the October issue of the *Journal*.

Page 270, Fig. 2. The two transistor switches TS1a and TS1b should both be shown as "push to make" (*not* "push to break").

The contacts R and A of the transistor switches TS2b and TS3b should be linked.

Addenda to the paper "The Physical Meaning of Formulae for Excess Noise in Composition Resistors" are given on page 403.

U.D.C. 621.317.72 : 621.385.113

Electron Transit Time and other Effects in a Valve Voltmeter Operating at Extremely Low Current

By

I. A. HARRIS,

(Associate Member)†

Summary: The theory of the effect of electron transit time on the comparison of v.h.f. voltage with a direct voltage, using a thermionic diode operating in the exponential region of its characteristic, is presented. The resulting transit time correction factor is shown to be dependent on the value of the peak voltage measured as well as on frequency, except for the lowest quarter of the usable frequency range where the voltage level has little effect on the correction factor.

Agreement between theory and experiment is sufficiently close to enable the particular diode used (EA 52) to measure peak values of sinusoidal voltages between 0·1 and 2 V to about 3% at frequencies up to 1200 Mc/s, under laboratory conditions using a new method of comparing the r.f. voltage with a direct voltage that is independent of such variable things as total emission and contact potential. At frequencies up to 300 or 400 Mc/s, an accuracy in voltage measurement as good as ± 0·3% can be obtained using the diode.

It is concluded, however, that apart from frequencies up to about 400 Mc/s where the frequency correction is small, it is simpler to determine the frequency correction experimentally if suitable apparatus is available. This is because the accurate calculation of transit time effect which is necessary for the larger correction factors at frequencies around 1000 Mc/s requires difficult subsidiary measurements to determine all the constants required for the particular diode used.

1. Introduction

Thermionic diodes operating largely in the exponential region of the characteristic, which can be described broadly as the region of negative anode potential, have long been used in valve voltmeters for the measurement of alternating voltages over a wide frequency band with a moderate degree of accuracy. These valve voltmeters are usually calibrated with a frequency that is low enough to enable other, more familiar means to be used to measure the voltage. It is then assumed that the diode will operate in the same way at higher frequencies. At very high frequencies, the precise value depending on the type of diode, it is well known that two things affect the results: (a) the onset of series resonance in the mounted diode and (b) electron transit-time in the diode. The first effect is the better known of the two and when the series resonant frequency has been measured the effect can readily be taken into account. The second effect has not received adequate treatment. A theoretical treatment based on the positive voltage characteristic of an idealized diode has been made,[1] but the conditions assumed were very different from those associated with the characteristic at very small currents.

† Ministry of Aviation, Electrical Inspection Directorate, Harefield, Middlesex.

Because the negative-voltage region of the characteristic can be represented by a well-defined mathematical function, a thermionic diode operating wholly in this region may be used as a precise laboratory instrument to compare r.f. voltages with direct voltages that can in turn be referred to standards. To enable this advantage over the usual valve voltmeter to be applied to radio-frequency voltage measurement to the fullest extent, the correction required by the effect of electron transit-time must be determined for the actual state of affairs in a diode with negative applied anode potentials. This determination forms the major part of the present paper, preceded by an outline of the process of rectification of oscillations of moderate frequency by a diode operated in the exponential region of its characteristic.

2. Rectification at Moderate Frequencies

2.1. *The Exponential Region of the Diode Characteristic*

In most of the region of negative anode voltage, the diode current is governed by the selective action of the retarding electric field on the distributed velocities of the thermionic emission from the cathode. This distribution is conveniently expressed in terms of the initial kinetic energy of the electrons, directed normally to the planar cathode surface. Thus the kinetic energy of a particle with the charge e and mass m of an electron is expressed

$$\tfrac{1}{2}mu^2 = eU$$

where U is the retarding potential difference (volts) which would just bring the particle with initial velocity u (km/s) to rest. Thus U expresses the magnitude of the kinetic energy in electron-volts.

The known Maxwellian distribution of the initial emission velocities shows that the part dJ_S of the total emission-current density J_S, which has initial kinetic energy with the component directed normally to the cathode surface between U and $U+dU$ electron-volts, is

$$dJ_S = J_S \exp(-U/V_T).dU/V_T \qquad \ldots\ldots(1)$$

where eV_T is the component of average energy of emission normal to the cathode surface.

Ideally, $V_T = T/11600$ where T is the cathode temperature in degrees Kelvin, but in practice V_T may appear to be slightly larger for reasons discussed later. Electrons with normal kinetic energy less than eU will not go beyond the place which has a negative potential $V = U$ relative to the cathode, but will be halted and will return to the cathode. Then, if V_a is the *negative* anode voltage, only those electrons with initial energies greater than eV_a will strike the anode and form the steady-state anode current density J_a.

Then

$$J_a = \frac{J_S}{V_T}\int_{V_a}^{\infty} \exp(-U/V_T).dU$$

and

$$J_a = J_S \exp(-V_a/V_T) \quad (\mathrm{A/m^2}) \qquad \ldots\ldots(2)$$

This is the density of the convection current carried to the anode by the moving space-charge of electrons. It is unaffected by the form of the potential curve between the electrodes, provided there is no minimum of depth lower than the negative anode potential V_a.

In an actual diode, the anode current may be affected additionally by several things. First, the voltage V_a is the sum of the applied voltage V_b and the contact potential V_c which, with an oxide coated cathode, is usually in the range 0·5 to 1·0 volt negative relative to the cathode. Secondly, a small voltage drop, proportional to the anode current, occurs across any interface resistance that forms between the cathode core and the emissive coating. Thirdly, a voltage drop occurs in the cathode material that usually contaminates the anode surface, but the resistance of this increases with decrease in current.[4] The effect is to reduce the slope of the static characteristic in the negative voltage region and to result in a value of V_T higher than the ideal if it is obtained from this characteristic curve. Both cathode interface and anode resistances are shunted by relatively large capacitances so that at radio frequencies these resistances are effectively short-circuited. This important fact must be borne in mind when considering the diode as a voltmeter. Fourthly, there is the elastic reflection of electrons at the cathode and anode surfaces. The major effects of this are on the form of the electric field distribution and on the maximum current of the exponential region of the characteristic, i.e. on the maximum value of J_a for which eqn (2) is valid.

In spite of these things, provided the cathode interface resistance is relatively small, the anode current may be expressed in the appropriate region by

$$I_a = I_z \exp(-V_b/V_T) \qquad \ldots\ldots(3)$$

where $I_a = J_aS$ is the anode current when the electrode area is S, I_z is the anode current (real or virtual) at zero applied voltage. With close-spaced electrodes in planar diodes, the fact that the contact potential is between 0·5 and 1·0 volt negative often results in the upper limit of the exponential region of the characteristic (i.e. where eqn (3) applies) being at a slightly positive value of applied anode voltage.

2.2. *Rectification of Oscillations of Moderate Frequency*

The rectification by the diode of oscillations of period long compared with the time of transit of electrons depends on the quasi-static characteristic (eqn (3)). The electric field distribution in the diode is not critical in this case. The values of the "constants" I_z and V_T in eqn (3) may be determined for a

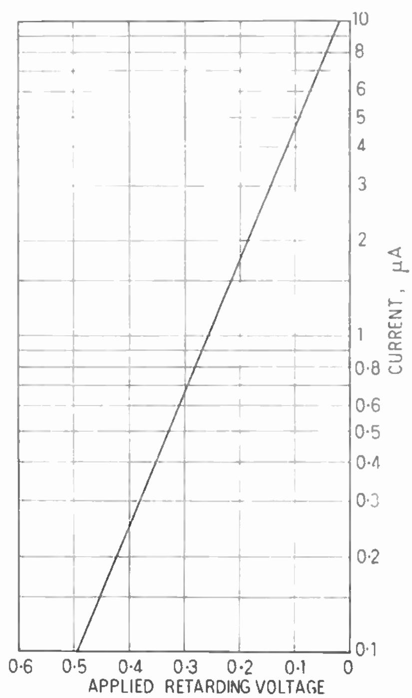


Fig. 1. Diode characteristic with negative applied anode voltage showing "exponential" characteristic.

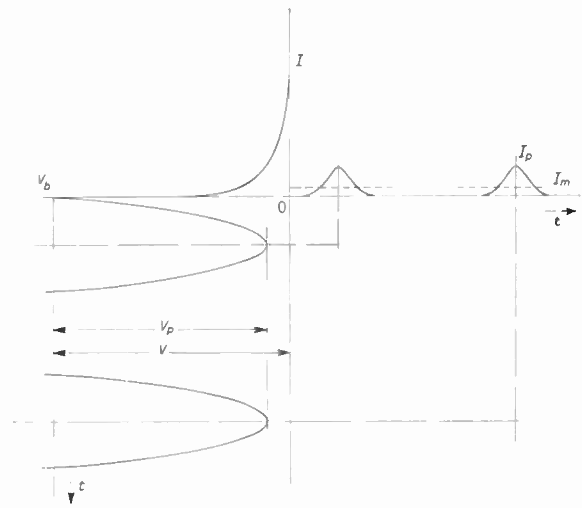


Fig. 2. Rectification of large signals.

particular diode from the measured $I_a - V_b$ characteristic curve, plotted on a logarithmic scale for I_a, such as that shown in Fig. 1. Caution must be exercised in the determination of V_T because of the possibility of resistances which are present at d.c. but which are shunted by capacitances at r.f. as explained in section 2.1. The current I_z varies considerably with small changes in cathode temperature, so that the cathode heating supply must be maintained at a constant level.

In rectification, the resultant applied voltage comprises the steady retarding voltage V_1 and an alternating voltage $V_p \cos \omega t$. For moderate frequencies, eqn (3) can be applied to the instantaneous voltage and current, giving

$$I_a(t) = I_z \exp[(-V_1 + V_p \cos \omega t)/V_T] \quad \ldots\ldots(4)$$

with the conditions $V_p < V_1$ and I_a does not exceed the maximum for the exponential region of the diode characteristic. The mean of eqn (4) over a complete cycle is the mean rectified current I_m. Thus

$$I_m = I_z \exp(-V_1/V_T) \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(\frac{V_p \cos \omega t}{V_T}\right) \cdot \mathrm{d}(\omega t)$$

and the result is

$$I_m = I_z \exp(-V_1/V_T) . \mathrm{I}_0(V_p/V_T) \quad \ldots\ldots(5)$$

where $\mathrm{I}_0(x)$ is the modified Bessel function of first kind and zero order. This equation which has been given before,[3] forms the basis of precise methods of determining V_p in terms of direct voltage or direct current measurements. Two extreme cases arise. When $V_p \gg V_T$, the modified Bessel function may be expressed in an approximate form, giving

$$I_m = I_z(2\pi V_p/V_T)^{-\frac{1}{2}} \exp[-(V_1 - V_p)/V_T] \ldots\ldots(6)$$

The rectification in this case is shown graphically in Fig. 2, in which it is seen that only the peak of the alternating voltage affects the mean current, enabling the diode to compare substantially the *peak* voltage with a direct voltage. When $V_p < V_T$,

$$I_m = I_z \exp(-V_1/V_T).[1 + V_p^2/(4V_T^2)] \ldots\ldots(7)$$

and the rectification is shown graphically in Fig. 3. If

$$I_{m0} = I_z \exp(-V_1/V_T)$$

is the current with the direct bias V_1 alone applied, then from eqn (7),

$$\frac{I_m - I_{m0}}{I_{m0}} = \left(\frac{V_p}{2V_T}\right)^2 \quad \ldots\ldots(8)$$

and

$$V_p = 2V_T[(I_m - I_{m0})/I_{m0}]^{\frac{1}{2}} \quad \ldots\ldots(9)$$

It is seen from Fig. 3 that the diode current in this case is affected by the whole of the oscillation and eqn (8) shows that the increment in mean current resulting from application of the oscillation is proportional to the square of the amplitude. Equation (9) suggests a method of determining V_T: by measuring a known V_p at a moderate radio frequency by two methods. Because eqns (8) and (9) depend only on the ratio of I_m to I_{m0}, the precise values of total

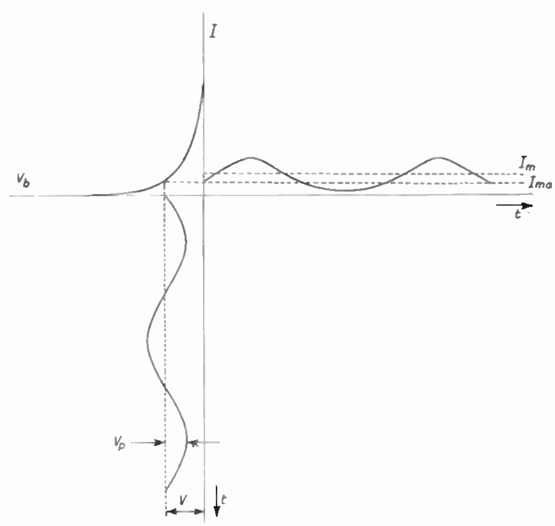


Fig. 3. Rectification of small signals.

emission and contact potential do not enter into the result. It is merely necessary that they are constant during the measurement. Similar methods can be devised to eliminate the total emission and contact potential in applying eqn (5) or (6) to the determination of the peak voltage; these will be described in section 4. The thermionic diode used in any one of these ways, i.e. as a voltage balance r.f.–d.c., is a precise instrument with many advantages over the semiconductor diode. Thus it can deal with a far wider range of voltages, it is not susceptible to variation in ambient temperature and it is more predictable in performance.

3. The Effect of Electron Transit Time on Voltage Measurement

3.1. *Statement of Problem*

When the electron transit time is not small compared with the period of the alternating component of applied voltage, eqns (4) to (9) require modification. Because only the mean rectified current is required, the convection current at the anode (i.e. the rate at which electrons are collected by the anode) is all that has to be calculated. In the steady state, the anode current in the exponential region of the characteristic is determined by the critical initial velocity (or kinetic energy) relative to the state (c.f. eqn (2) and the preceding text). With an alternating component of anode voltage, this critical initial velocity is a function of time, and one can define an instantaneous critical initial velocity such that all electrons leaving the cathode with velocities greater than this at the given instant will contribute to the mean current, while all electrons leaving the cathode at the same instant with velocities lower than critical will be turned back to the cathode. The problem here is to determine the instantaneous critical initial velocity under all the given conditions. Electrons reflected by the anode will all return to the cathode because, in the exponential region of the characteristic the field is always in a direction which accelerates electrons towards the cathode.

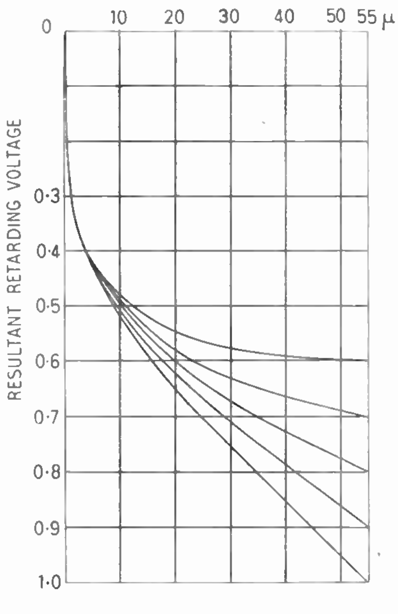


Fig. 4. Curves of potential versus position in the diode space.

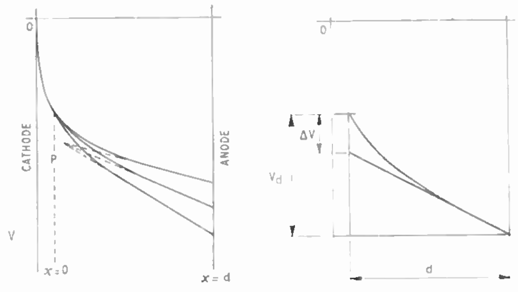


Fig. 5. (*a*) Approximate form of potential curves.
(*b*) Notation for the approximate formulation of the potential curves $V = (V_d - \Delta V)(x/d) + \Delta V.(x/d)^2$.

3.2. *The Electric Field Distribution in the Diode*

In the determination of electron transit-time and its effects, the electric field distribution has an appreciable effect on the results and must be known. A family of potential *versus* position curves for an actual planar diode operated in the exponential region of the characteristic has been calculated and is shown in Fig. 4. In calculating these curves, account has been taken of the fact that some 90% of the electrons returning to the cathode are reflected by it, with the result that there is a far greater number of low velocity electrons "bouncing up and down" from the cathode surface than would be indicated by the usual calculations which ignore reflections. This results in an extremely dense space-charge near the cathode which depresses the potential curve far more sharply near the cathode than the customary calculations would indicate. Thus, in Fig. 4 the space-charge between the cathode surface and point P is so dense that the anode voltage variation makes no appreciable change in the potential in this part of the space. Between P and the anode, the potential curves may be represented approximately by a linear function of distance from the anode plus a smaller quadratic function of distance from the anode, as illustrated in Figs. 5(*a*) and 5(*b*). It is seen that the potential relative to the anode may be represented as a function of distance x by

$$V = (V_d - \Delta V)(x/d) + \Delta V(x/d)^2$$

and the resulting time of transit of an electron from the plane through P, which acts as a virtual cathode for present purposes, to the anode is found to be expressed (in seconds) as

$$\tau_0 = \frac{3{\cdot}4 \times 10^{-6} d}{(V_d - \Delta V)^{\frac{1}{2}}}\left[1 - \frac{1}{6}\left(\frac{\Delta V}{V_d - \Delta V}\right) + \frac{3}{40}\left(\frac{\Delta V}{V_d - \Delta V}\right)^2\right] \quad ......(10)$$

where d is in metres and V_d and ΔV are in volts.

To a fair approximation, the electric field between the plane through P and the anode may be regarded as the field of a planar diode in which the space-charge density is too small to influence the potential-curve shape which is effectively a straight line, as in Fig. 5(*b*). In eqn (10), it should be noted that V_d is not the resultant negative anode voltage, but it is less than this by about 0·4 volt for the diode associated with the curves in Fig. 4.

3.3. *Calculation of the Instantaneous Critical Initial Velocity*

The model used to represent the diode is that shown in Fig. 5(*b*) where the "cathode" is taken to be the plane through P. The electric field intensity E (V/m) is assumed to be constant across the space for a given anode voltage V and equal to V/d where d is the electrode spacing of the model. With an alternating component of voltage $V_p \cos \omega t$ superimposed on V, the resultant anode voltage is

$$V' = -V + V_p \cos \omega t$$

when V is the retarding bias voltage. The suffix p signifies the peak value of the alternating component, the plain letter the steady component and the dash signifies the total variable value. If the usual complex notation is adopted, then

$$V' = -V + V_p\,e^{j\omega t}$$

The equation of motion of a particle with the charge e and mass m of an electron is

$$m\ddot{x}' = eE'$$

where x is the distance moved. Then the acceleration is

$$\ddot{x}' = -\left(\frac{e}{md}\right)(V - V_p\,e^{j\omega t}) \qquad \ldots\ldots(11)$$

On integration with respect to time t between the limits t_a (the instant at which the electron leaves the P-plane at $x = 0$ and t (the instant at which the electron reaches the plane at x) there results the velocity:

$$v' = u' - \left(\frac{e}{md}\right)\left[V(t-t_a) - \frac{V_p}{j\omega}(e^{j\omega t} - e^{j\omega t_a})\right] \qquad \ldots\ldots(12)$$

where u' is the initial velocity which has both steady and alternating components, the latter not necessarily being sinusoidal. The alternating component is a function of t_a and not t, however, so u' does not have to be expressed in detail for the present series of integrations.

Integration of eqn (12) with respect to t leads to

$$x = u'(t-t_a) - \left(\frac{e}{md}\right)\left\{\tfrac{1}{2}V(t-t_a)^2 - \frac{V_p}{(j\omega)^2}\left[e^{j\omega t} - e^{j\omega t_a} - j\omega(t-t_a)\,e^{j\omega t_a}\right]\right\} \qquad \ldots\ldots(13)$$

in which $x = 0$ at $t = t_a$.

So far the analysis has followed customary lines. A variable that plays an important part in this work, as in most similar work involving electro-mechanical theory, is the time of transit τ from the plane $x = 0$ to the plane at an arbitrary position x, which in general is a function of time. Thus $\tau = t - t_a$ is termed the instantaneous transit time to the position x. Associated variables are the transit angle $\theta = \omega\tau$ and the imaginary transit angle $\alpha = j\omega\tau$. With this notation, much time and space can be saved by using the following two complex transit-angle functions which belong to a set of six commonly used by British and European writers on the small-signal a.c. theory of electron devices[2]:

$$\Phi_1(\alpha) = \frac{1}{\alpha}(1 - e^{-\alpha}) \qquad \ldots\ldots(14)$$

$$\Phi_3(\alpha) = \frac{2}{\alpha^2}(1 - e^{-\alpha} - \alpha\,e^{-\alpha}) \qquad \ldots\ldots(15)$$

Then the foregoing equations can be written:

$$v' = u' - \left(\frac{eV}{md}\right)\tau\left[1 - \frac{V_p}{V}e^{j\omega t}\,\Phi_1(\alpha)\right] \qquad \ldots(16)$$

$$x = u'\tau - \frac{1}{2}\left(\frac{eV}{md}\right)\tau^2\left[1 - \frac{V_p}{V}e^{j\omega t}\,\Phi_3(\alpha)\right] \ldots(17)$$

The problem is to determine the initial velocity of an electron that just reaches the anode at the distance d from the plane $x = 0$. For such an electron, $v' = 0$ at $x = d$, so that eqn (16) gives:

$$u' = \left(\frac{eV}{md}\right)\tau_d\left[1 - \frac{V_p}{V}e^{j\omega t}.\Phi_1(\alpha_d)\right] \qquad \ldots\ldots(18)$$

where the subscript d denotes the value for $x = d$. Also, eqn (17) gives:

$$d = u'\tau_d - \frac{1}{2}\left(\frac{eV}{md}\right)\tau_d^2\left[1 - \frac{V_p}{V}e^{j\omega t}.\Phi_3(\alpha_d)\right] \qquad \ldots\ldots(19)$$

and with (18) substituted in (19)

$$d = \frac{1}{2}\left(\frac{eV}{md}\right)\tau_d^2\left[1 - \frac{V_p}{V}e^{j\omega t}.\Phi_2(\alpha_d)\right] \ldots\ldots(20)$$

in which

$$\Phi_2 = 2\Phi_1 - \Phi_3$$

or

$$\Phi_2(\alpha) = \frac{2}{\alpha^2}(\alpha + 1 - e^{-\alpha}) \qquad \ldots\ldots(21)$$

It follows from eqn (20) that:

$$\tau_d = \frac{2d}{(2eV/m)^{\frac{1}{2}}}\left[1 - \frac{V_p}{V}e^{j\omega t}.\Phi_2(\alpha_d)\right]^{-\frac{1}{2}} \ldots\ldots(22)$$

which is the instantaneous transit time of an electron that just reaches the anode. (It should be noted that the complex notation in the square brackets must be translated back into real notation before any operation is performed on it, because the notation $\cos \omega t = \mathrm{Re}\, e^{j\omega t}$ can only be used in linear operations.)

The instantaneous critical velocity is given by eqn (18), but a more useful quantity is the (electron) voltage equivalent to the initial kinetic energy, connected by the relation

$$U = (m/2e)u^2$$

Thus from eqn (18) with eqn (22) substituted for τ_d:

$$U'_{\text{crit}} = V\frac{[1-(V_p/V)\, e^{j\omega t}.\Phi_1(\alpha_d)]^2}{1-(V_p/V)\, e^{j\omega t}.\Phi_2(\alpha_d)} \quad \ldots\ldots(23)$$

When $\tau \to 0$, $U'_{\text{crit}} \to V[1-(V_p/V)\, e^{j\omega t}]$, the real part of which is $V - V_p \cos \omega t$, equal to the negative of the applied voltage.

Equation (23) expresses U'_{crit} as a function of t, the time of arrival at the anode, whereas it is required as a function of t_a, the time of leaving the cathode. The transformation is readily effected, starting from the relation $\tau = t - t_a$. Then

$$e^{j\omega t} = e^{j\omega\tau}\, e^{j\omega t_a} = e^{\alpha}\, e^{j\omega t_a}$$

which is substituted for $e^{j\omega t}$ in eqns (22) and (23). It can be deduced from the definitions of Φ_1, Φ_2 and Φ_3 that:

$$e^{\alpha}\, \Phi_1(\alpha) = \Phi_1{}^*(\alpha)$$
$$e^{\alpha}\, \Phi_2(\alpha) = \Phi_3{}^*(\alpha)$$

where the asterisk denotes the conjugate complex function in which $-j$ replaces j. Then

$$\tau_d = \frac{2d}{(2eV/m)^{\frac{1}{2}}}[1-(V_p/V)\, e^{j\omega t_a}\, \Phi_3{}^*(\alpha_d)]^{-\frac{1}{2}} \ldots(24)$$

$$U'_{\text{crit}} = V\frac{[1-(V_p/V)\, e^{j\omega t_a}\, \Phi_1{}^*(\alpha_d)]^2}{1-(V_p/V)\, e^{j\omega t_a}\, \Phi_3{}^*(\alpha_d)} \quad \ldots\ldots(25)$$

These two expressions may be transformed into real notation by first expressing the transit-angle functions as moduli and arguments and then using the relation $\cos\theta = \mathrm{Re}\, e^{j\theta}$. Thus, according to ref. 2,

$$\left.\begin{aligned} |\Phi_1{}^*| = |\Phi_1| &= \left[1-\frac{\theta^2}{12}+\frac{\theta^4}{360}-\ldots\right]^{\frac{1}{2}} \\ \arg(\Phi_1{}^*) &= \theta/2 \end{aligned}\right\} \ldots(26)$$

$$\left.\begin{aligned} |\Phi_3{}^*| = |\Phi_3| &= \left[1-\frac{\theta^2}{18}+\frac{\theta^4}{720}-\ldots\right]^{\frac{1}{2}} \\ \arg(\Phi_3{}^*) &= (2\theta/3)[1-\theta^2/540\ldots] \end{aligned}\right\} \ldots(27)$$

where $\theta = \alpha_d/j$. Then eqn (25) becomes

$$U'_{\text{crit}} = V\frac{[1-(V_p/V).|\Phi_1|.\cos(\omega t_a+\theta/2)]^2}{1-(V_p/V).|\Phi_3|.\cos(\omega t_a+2\theta/3)} \quad (28)$$

By differentiating with respect to ωt_a and equating the result to zero, the value which makes U'_{crit} a minimum may be found. For small values of θ and any value of V_p/V, or for small values of V_p/V and any value of θ, it is found that $\omega t_a = -\theta/3$ for U'_{crit} = minimum. It is concluded that this result is approximately true for all values of interest. Most of the convection current at the anode occurs at and near the minimum of the critical initial energy, so that this minimum is of most significance, and we replace ωt_a by $-\theta/3$ in eqn (28) and an equation derived from eqn (24). Thus

$$\tau_d = \frac{2d}{(2eV/m)^{\frac{1}{2}}}[1-(V_p/V).|\Phi_3|.\cos(\theta/3)]^{-\frac{1}{2}}$$

On squaring this and multiplying by ω^2 there results

$$\theta^2 = \frac{\theta_0^2(1-V_p/V)}{1-(V_p/V)|\Phi_3|\cos(\theta/3)} \quad \ldots\ldots(29$$

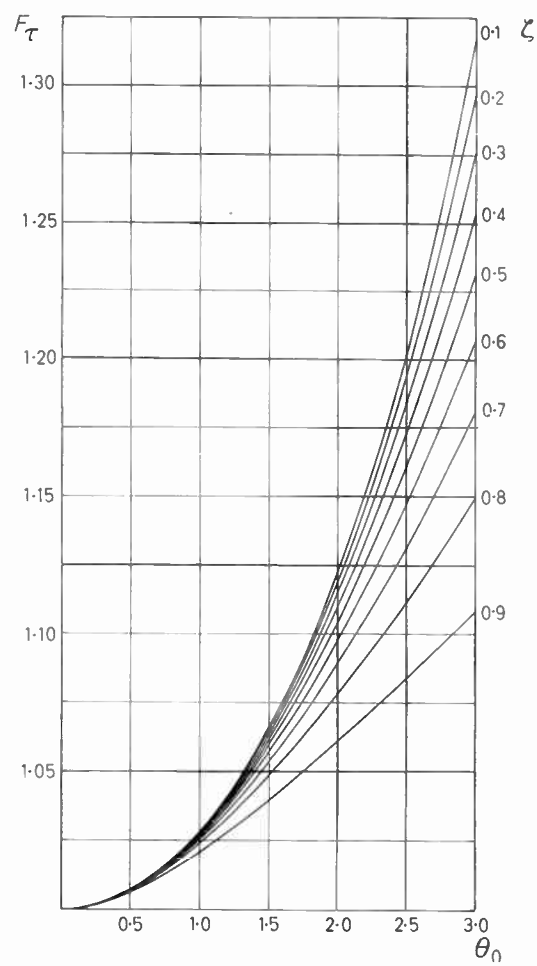


Fig. 6. Curves giving the transit-time correction factor F_τ as a function of the reference transit angle θ_0 (radians), with ζ (= peak applied voltage/bias voltage) as a parameter.

where ideally,

$$\theta_0 = \frac{2\omega d}{(2e/m)^{\frac{1}{2}}(V - V_p)^{\frac{1}{2}}} \qquad \ldots\ldots(30)$$

The quantity θ_0 is termed the reference transit-angle. Then finally,

$$U'_{\min} = V \frac{[1-(V_p/V).|\Phi_1|.\cos(\theta/6)]^2}{1-(V_p/V).|\Phi_3|.\cos(\theta/3)} \qquad \ldots(31)$$

The values of θ in eqn (31) have to be determined from the solution of eqn (29) as a function of θ_0 and V_p/V.

3.4. *The Transit Time Correction for a Voltage Measurement*

If V_τ is the apparent peak alternating voltage that would be measured according to section 2.2 with no correction made for transit-time effect, then:

$$U'_{\min} = V - V_\tau = V(1 - V_\tau/V)$$

Then

$$V_\tau/V_p = (V/V_p)[1-(U'_{\min}/V)]$$

and with eqn (31)

$$\frac{V_\tau}{V_p} = \frac{1}{\zeta}\left\{1 - \frac{[1-\zeta|\Phi_1|\cos(\theta/6)]^2}{1-\zeta|\Phi_3|\cos(\theta/3)}\right\} \qquad \ldots\ldots(32)$$

in which ζ is written for V_p/V.

The ratio V_p/V_τ is the transit-time correction factor F_τ to be applied to a voltage measurement. When $\theta \to 0$, $|\Phi_1|$, $|\Phi_3|$ and the cosines all approach unity, so that $V_p/V_\tau \to 1$. Values of F_τ have been determined for a range of values of ζ from 0·1 to 0·9 and for θ_0 from 0 to 3. The results are plotted as a family of curves in Fig. 6.

4. Experimental Verification

4.1. *Experimental Arrangement*

The basic circuit is shown in Fig. 7, in which the actual electrodes of the diode form the plates of a capacitor of capacitance C_v and where L_v represents the effective inductance of the connecting structure such as may be found in a diode with a disc-seal cathode connection. The d.c. blocking capacitor, of value C_b, that also serves as the reservoir capacitor, is in series with the diode in so far as r.f. current is concerned. In making a voltage measurement with the diode, it is the voltage between the actual electrode surfaces that is measured. It is therefore necessary to determine the voltage transformation, as a function of frequency, between the accessible terminals A′, B′ and the electrode surfaces A, B. An approximate relation, derived from the lumped circuit-element representation in Fig. 7, is:

$$V_{A'B'}/V_{AB} = V_p/V_{pe} = 1 + C_v/C_b - (f/f_0)^2 \quad \ldots(33)$$

where f_0 is the series resonant frequency of the diode

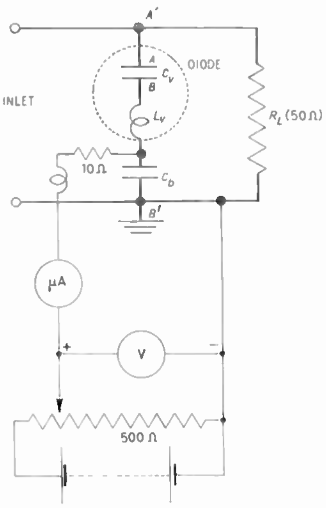


Fig. 7. Circuit diagram with equivalent circuit inside the diode. The active voltage in the diode appears across AB.

in its mount and f is the frequency at which the voltage ratio is required. A more accurate value of the voltage ratio is obtained when the distributed nature of the diode structure is taken into account. Thus for the diode type EA 52 used in the experiments, the series resonant frequency was 3120 Mc/s, C_v was 0·5 pF and the value of C_b used was 33 pF. The resulting voltage ratios, taking account of the distributed nature of the circuit constants, are given in Table 1.

Table 1

f (Mc/s)	300	500	700	900	1000	1100	1200
V_{pe}/V_p	0·995	1·012	1·040	1·079	1·101	1·130	1·162

At the d.c. end of the frequency scale there are effects of the cathode interface resistance and the resistance of the material deposited on the anode surface, both of which are shunted by capacitances, which short-circuit them at moderately high radio frequencies. The effect of the cathode interface resistance, which is nearly constant with change in anode current, is minimized by working with very low current (of the order of 1 μA or less); but the effect of the anode resistance, which increases with decrease in current, may not be minimized in the same way.

In order to make comparisons with a thermal wattmeter, the diode was shunted across a coaxial line which was terminated in a resistance of 50 ohms. The diode capacitance of 0·5 pF was neutralized in such a way that the voltage standing-wave ratio relative to 50 ohms, at the r.f. inlet, was not worse than 1·01 and that the points A′ and B′ were across a true 50 ohms resistance, at all frequencies up to at least 1200 Mc/s. With this apparatus, an accurate comparison between the power dissipated in a load of 50 ohms and the voltage between the diode electrodes can be made.

4.2. *Measurement of Radio-Frequency Voltage*

The peak r.f. voltage between the electrode surfaces (A′B′ in Fig. 7) is measured with the diode in accordance with a method mentioned in section 2.2. To eliminate I_z from eqn (5), the negative d.c. bias voltage V_0 required to obtain a small diode current I_m (such as 1 μA or less) with no r.f. applied, is noted. The r.f. is then applied and the negative bias voltage V_1 required to obtain the same mean diode current I_m is observed. From the difference $V_1 - V_0$ the peak r.f. voltage V_p can be obtained from the following equation, derived from eqn (5):

$$\frac{V_1 - V_0}{V_T} = \ln\left[I_0(V_p/V_T)\right] \qquad \ldots\ldots(34)$$

In this, V_p can be expressed numerically as a function of $V_1 - V_0$ for given values of V_T by using tabulated values of the logarithm and the modified Bessel function. It is important that a strictly sinusoidal oscillation is applied to the diode and for this purpose a harmonic filter must be used between the radio frequency source and the diode mount or the thermal wattmeter. As a typical example of the measurements made, the results in Table 2 are given. These were obtained with a nominal 1 mW dissipated in the load and a mean diode current of 1 μA.

Table 2

(Mc/s)	V_1	$V_1 - V_0$	V_p (eqn 34)	V_p (from power level)
900	0·420	0·160	0·310	0·309 ± 0·002
1000	0·405	0·145	0·288	0·289 ,,
1100	0·410	0·150	0·295	0·301 ,,
1200	0·410	0·150	0·295	0·302 ,,

The values of V_p obtained from the power level measurements have to be multiplied by the factors given in Table 1 to obtain the voltages across the electrodes, V_{pe}. The resultant value of the negative anode voltage relative to the cathode at the peak of the oscillation is then given by $V_1 + V_c - V_{pe}$ where V_c is the negative contact potential (0·75 volt in the diode used). This is the resultant retarding voltage shown as the ordinate in Fig. 4. From this resultant voltage, the electron transit time at the peak of the oscillation can be estimated by the method described in section 3.2. In eqn (10), the value of d is 5×10^{-5} m and V_d is given by $V_1 + V_c - V_{pe} - 0{\cdot}425$ according to the curves in Fig. 4. The curve in the family appropriate to the particular problem is that with the value $V_1 + V_c - V_{pe}$ at the anode (i.e. on the right of Fig. 4).

The values of V_{pe}, τ_0 and θ_0 ($= \omega\tau_0$) are given in Table 3.

Table 3

f (Mc/s)	V_{pe} (obs.)	$V_1 + V_c - V_{pe}$	τ_0	ω	θ_0 (radian)
900	0·334	0·836	$2{\cdot}96 \times 10^{-10}$	$5{\cdot}66 \times 10^{9}$	1·67
1000	0·318	0·837	2·96	6·28	1·86
1100	0·340	0·820	3·06	6·91	2·12
1200	0·351	0·809	3·15	7·54	2·37

From the values of θ_0 and $\zeta = V_{pe}/(V_1 + V_c)$ the transit-time correction factor F_τ can be obtained directly from the curves in Fig. 6 and the results can be compared with the experimental results when corrected by the voltage transformation factor (Table 1).

The various values are given in Table 4.

Table 4

f (Mc/s)	ζ	F_τ	V_{pe} (calc.)	V_{pe} (obs.)	Error
900	0·285	1·077	0·334	0·334	0
1000	0·276	1·100	0·317	0·318	−0·3%
1100	0·293	1·129	0·333	0·340	−2%
1200	0·302	1·160	0·342	0·351	−2·6%

The discrepancies are not greater than the maximum uncertainty between the two sets of experimental results of 3%.

More accurate measurements carried out at the frequency 300 Mc/s have shown agreement in voltage measurement between the diode and a thermal wattmeter to about ±0·3%, but the transit-time correction at this frequency does not amount to more than 1% so that it is not so effective a check on the theory as the results quoted for frequencies up to 1200 Mc/s.

5. Conclusions

The theory of electron transit-time effect on the measurement of v.h.f. and lower u.h.f. peak voltage by a suitable thermionic diode operated wholly in the exponential region of the characteristic has been worked out. A diode type EA52, selected for uniformity of electrode spacing and an exponential region of the characteristic that extends to an upper current of 20 μA was used to balance the peak r.f. voltage against a direct voltage by a method that eliminated the effects of total emission and contact potential. The voltage was also determined from the power dissipated in a precise 50 ohms r.f. resistor and comparison between the measurements, after allowing for the voltage transformation between the accessible diode terminals and the electrode surfaces, showed satisfactory agreement. It is concluded that suitable thermionic diodes can be used for the precise measurement of sinusoidal r.f. voltages (up to about 2 V peak at 1200 Mc/s, and 7 V peak at 300 Mc/s) under laboratory conditions, the uncertainty rising from about $\pm 0{\cdot}3\%$ at 300 Mc/s to $\pm 2\%$ at 1200 Mc/s.

The calculation of transit-time, however, is a tedious process that requires a number of subsidiary measurements which can only be made with a suitable diode, and it is probably better to measure the frequency correction factor of a particular diode by comparison of results with those obtained by other means, such as a thermal wattmeter with a v.s.w.r. better than 1·02. Up to about 400 Mc/s, a relatively rough calculated value of transit-time correction factor and a measurement of resonant frequency of the diode should not prove too difficult on a routine basis.

6. Acknowledgments

Acknowledgment is made to Mr. R. E. Spinney for measurements of the r.f. loss of the diodes at frequencies up to 1000 Mc/s and of the interface resistance and cathode coating resistance of the diodes. Also, to Mr. L. J. Graver for re-calculating the curves of Fig. 6 to a greater degree of accuracy than was originally attained by the author.

7. References

1. F. B. Llewellyn, "Electron Inertia Effects", pp. 87–91 (Cambridge University Press, 1941).
2. C. J. Bakker and G. de Vries, "On vacuum-tube electronics", *Physica*, **2**, pp. 683–97, July 1935,
   or W. E. Benham and I. A. Harris, "The Ultra High Frequency Performance of Receiving Valves", Appendix I. (Macdonald, London, 1957.)
3. E. G. James and J. E. Houldin, "Diode frequency changers", *Wireless Engineer*, **20**, pp. 15–27, January 1943 (Appendix II).
4. J. Seymour, "An Investigation of a Current-Decay Effect of Millisecond Duration in Thermionic Valves", M.Sc. Thesis, London University, July 1960.

Manuscript received by the Institution on 15th February 1962. (Paper No. 764.)

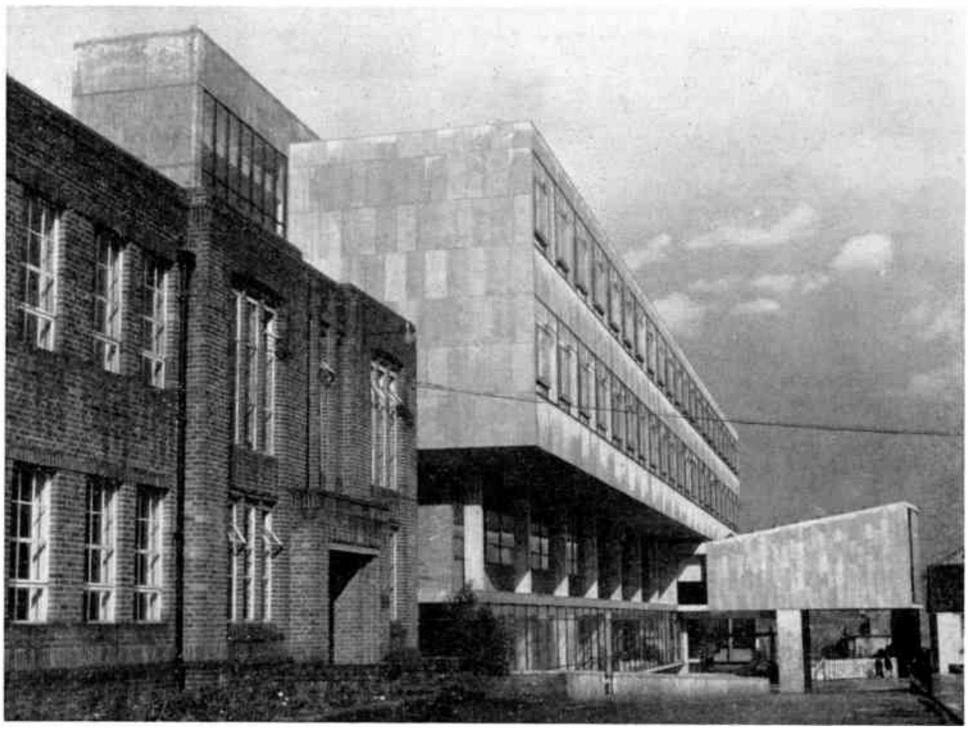
THE 1963 CONVENTION AT SOUTHAMPTON

April 16—20th

The 1963 Convention on Electronics and Productivity will be the second Institution function of its kind to take place at the University of Southampton. In 1951, two sessions of the Festival of Britain Convention, dealing with Radio Communications and Broadcasting, and Radio Aids to Navigation, were held at the then University College of Southampton. During the intervening twelve years, both the University College and the Institution have received the grant of a Royal Charter.

Yet another link between the University and the Institution is the fact that the University of Southampton is one of the very few universities in Great Britain which recognizes electronics as a subject in its own right distinct from either electrical engineering or physics. The first and present occupant of the Chair of Electronics is Professor Eric Zepler, who was President of the Institution from 1959 to 1961 and is now a Vice-President. Professor Zepler was elected Dean of the Faculty of Engineering earlier this year.

Founded in 1862 as the Hartley Institution, the University of Southampton recently celebrated its centenary. Its plans for expansion from a pre-war student population of 300 to the figure for 1961–62 of 1,816 and by 1980 to 4,000 have called for an extensive and imaginative scheme for erecting many new buildings. Already a number of these have been completed and members who attended the 1951 Convention will see a great change in the University site; more changes are scheduled before the completion of the development plan in the late 1970s. One of the most striking new buildings is the Lanchester Building, which houses part of the Faculty of Engineering and was designed by Sir Basil Spence, O.B.E., R.A., R.D.I., P.P.R.I.B.A., consultant architect to the University. The majority of the sessions of the Convention will be held in the excellently appointed large Lecture Theatre of this Building.



A view of the Lanchester Engineering Building at the University of Southampton.

Delegates to the Convention will be able to stay in the University's Halls of Residence—Connaught Hall and South Stoneham House—which are situated about a mile from the University site. Further details of the residential facilities in the Halls of Residence will be given in a later issue of the *Journal.*

U.D.C. 621.397.621.54

Dual Standard I.F. Amplifiers

By

A. C. BARTON †

Presented at a Television Group Symposium on "Switchable Standard Television Receivers" in London on 7th February 1962.

Summary: The problems of switching the amplifier between systems of different frequency, bandwidth and type of modulation are examined and it is shown that the characteristics of the amplifier must be altered, particularly when switching from a.m. to f.m. sound. Precautions are taken to avoid interference due to feedback. Circuit arrangements are given and detector designs for reception of f.m. sound are described.

1. Introduction

With the present 405-line system of television transmission in this country the picture modulation is positive, that is, maximum power output is with peak white. The 625-line C.C.I.R. system employs negative modulation, in which maximum power is transmitted during synchronizing pulses. The 625-line transmission requires greater bandwidth to take advantage of the higher line frequency and arrangements must be made in the i.f. amplifier for this, and for the change in modulation sense. A further fundamental difference is that the sound transmission with the 625-line system is frequency modulated and a frequency discriminator is required in the receiver instead of an a.m. detector.

The intermediate frequencies for the 405-line standard are established so that the interference possibilities are the same for all types of receiver, and similarly chosen i.f.'s have been suggested for the 625-line system.

Table 1

Dual System Requirements

	625	405
Modulation—Vision	negative	positive
Modulation—Sound	F.M.	A.M.
Channel Width	8 Mc/s	5 Mc/s
Carrier Separation / Inter-carrier Frequency	6 Mc/s	3·5 Mc/s
Sound/Vision Power Ratio	5/1	4/1
Vision I.F.	39·5 Mc/s	34·65 Mc/s
Sound I.F.	33·5 Mc/s 6·0 Mc/s	38·15 Mc/s
Adjacent Sound Frequency	41·5 Mc/s	33·15 Mc/s
Adjacent Vision Frequency	31·5 Mc/s	39·65 Mc/s

Table 1 shows some of the transmission characteristics and the intermediate frequencies. A 405-line channel covers 5 Mc/s with 3·5 Mc/s between carriers, while a 625-line channel covers 8 Mc/s with 6·0 Mc/s between carriers. For 405-line transmissions the sound carrier is at a lower frequency than the vision carrier, while for 625-line standards the reverse is the practice. If the local oscillator is set higher than the r.f. carriers, the frequencies for the i.f. amplifier are as shown in the table.

2. The I.F. Amplifier with Intercarrier Sound

The modulation characteristic of the sound carrier on 625-line transmissions is of similar form to that of the Band II f.m. broadcast transmissions. Amplification of the sound channel takes place in the i.f. amplifier at a frequency of 33·5 Mc/s. For the reception of f.m. sound transmissions a very stable local oscillator is required to prevent distortion in the frequency discriminator. To overcome this in the 625-line television system advantage is taken of the vision carrier to give a second i.f. of 6·0 Mc/s, that is, the intercarrier system.

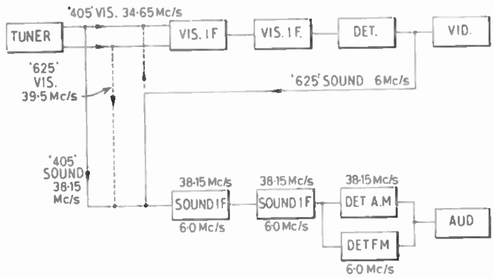


Fig. 1. Dual system block diagram.

Figure 1 shows a possible arrangement in the i.f. amplifiers. The aim should be to have as many common components as possible so that the dual system arrangement is near to its single system counterpart in price and size. Ignoring the dotted feeds, a tuner is shown feeding a vision i.f. amplifier common to both standards. Switching to 405 lines must select the appropriate tuned circuits in the amplifier for shaping and suppress the a.m. sound component at 38·15 Mc/s at the input to the vision

† Pye Ltd., Cambridge.

detector. The sound i.f. amplifier has two separate stages, and a standard a.m. detector. Switching to 625 lines must change the shaping circuits in the vision i.f. amplifier and reverse the polarity of the vision detector to give the same video sense into the common video amplifier. The sound carrier at 33·5 Mc/s is not suppressed completely in the amplifier but the difference frequency between it and the vision carrier, locked to the transmitter at 6·0 Mc/s and therefore independent of local oscillator drift, is taken from the vision detector and fed into the sound i.f. amplifier. Amplification takes place at 6·0 Mc/s and a frequency discriminator feeds a common sound output stage. Since the video amplifier can give a useful sound gain at 6·0 Mc/s the feed to the i.f. may be taken from its anode circuit. Extreme care is then required since non-linearity in this stage will produce unwanted vision buzz on the sound due to amplitude modulation while video cut-off will remove the sound component completely.

The dotted feeds show some of the interference possibilities in the receiver.

(1) Feedback from the vision detector may take place at the vision carrier frequency via the 625-line sound feed.

(2) The vision carrier on 625 lines at 39·5 Mc/s may feed into the sound i.f. amplifier and give cross-talk into the f.m. detector and buzz from the spurious a.m.

(3) There is an increased problem due to feedback from the detectors themselves, on the vision side since it has to be switched, and on the sound side because the detector region is more complex.

A 405/625 changeover means that the layout must be chosen so that the connections are short to the switches but that all these switches must be linked mechanically to the system switch. These two conditions are difficult to reconcile. It must be remembered that a switch at these frequencies may give an isolation of only 10 dB or so due to the capacitance across the contacts.

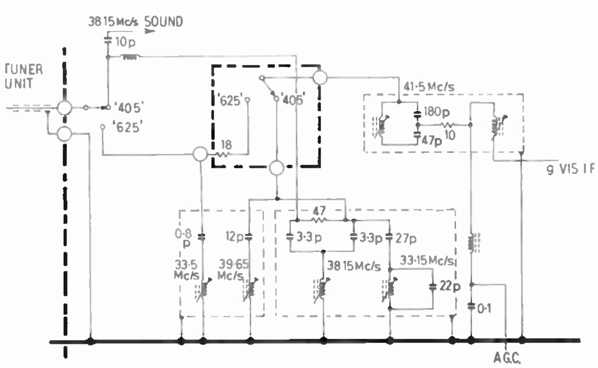


Fig. 3. 405/625 vision input.

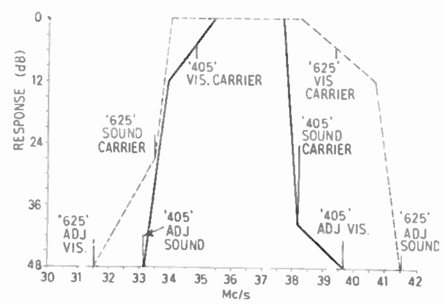


Fig. 2. Dual system i.f. characteristic.

In the block diagram shown which is suitable for v.h.f., there is a 60 dB amplifier in the vision i.f. amplifier and one of similar gain in the sound i.f. amplifier for either system. For intercarrier working on 625 lines there is normally some 30 dB rejection of the sound component in the vision i.f. amplifier, thus reducing the level to a small percentage of the vision carrier at the vision detector. This satisfies the requirement of keeping the sound amplitude well inside the unmodulated 10–12% region of the vision carrier, and avoids vision buzz on sound.

For fringe u.h.f. working, since the gain of u.h.f. tuners is at least ten times less than v.h.f. tuners, gains as indicated are too low and either—

(1) another stage is added for u.h.f. only

(2) another stage is added for u.h.f. and v.h.f. and the v.h.f. tuner gain is reduced, or

(3) the v.h.f. tuner is used as an i.f. amplifier on u.h.f.

Figure 2 shows the vision i.f. amplifier characteristic required, the aim being to achieve with monochrome transmissions over 40 dB rejection for adjacent channels, 34 dB for 405-line sound, and 20 dB for 625-line sound. The most significant region is above 37 Mc/s, where the 405-line response has to fall sharply to give sound rejection at 38·15 Mc/s in a region of maximum gain on 625 lines. The mean frequency in effect moves lower for 405 lines. If the smallest number of switches is desirable, then all the shaping should be in one part of the circuit, one switched interstage filter transforming the 625-line response into the 405-line version. Filters at extremes of the band may be left in on both systems.

Some typical dual system circuit arrangements can be shown in detail.

Figure 3 shows one form of a filter which is switched in the interstage connections. On 405-lines there is a bridged-T type rejector at the 38·15 Mc/s sound frequency, and two adjacent channel rejectors. The a.m. sound is taken off capacitively at the filter input. On 625 lines a 33·5 Mc/s filter is switched in, and the a.m. sound take-off for 405 lines is switched out since vision carrier on 625 lines would break through on to the f.m. sound as buzz. Some additional damping is switched in on 625 lines since with the increased bandwidth there would be a trough in the bandpass characteristic. An adjacent channel 41·5 Mc/s trap is shown in both systems. If a centre frequency change is required on changeover, series inductances may be switched in on the appropriate filter, in the same way as the 18-ohms damping component. A circuit of the form shown would determine largely the overall characteristic on either system, as indicated in Fig. 2.

sound take-off is shorted when the detector is changed to the 405-line position eliminating direct feedback from the vision detector.

In these particular circuits, the r.f. returns in the detector circuit are arranged so that no change in response takes place at i.f., only a reversal of polarity. If, as in Fig. 3, the shaping circuits are at the input, then the detector and interstage circuits must have a bandwidth wide enough for 625 lines and therefore without rejection at 38·15 Mc/s, which is the a.m. sound carrier.

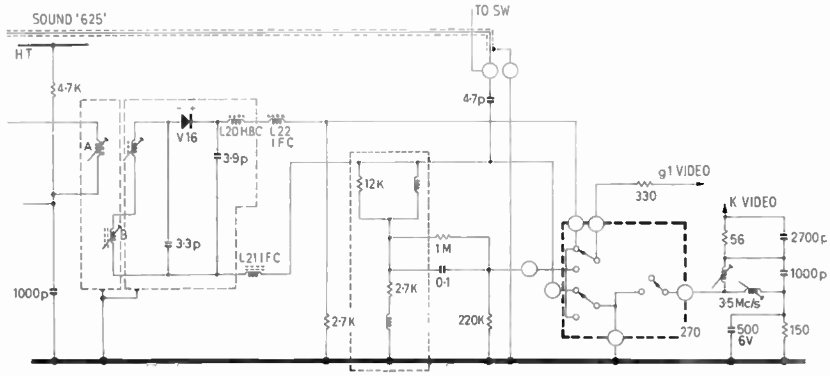


Fig. 4. Vision detector stage.

3. Vision Detector

Figure 4 shows a form of vision detector which may be used as changeover from positive to negative modulation. There is a bandpass filter with variable coupling to the diode detector. There is no r.f. switching, since reversal of the detector polarity is by means of a switch after the i.f. chokes. Either the anode or the cathode of the detector V16 is earthed via a choke, and a second switch on the video grid selects either 405-lines video with its positive d.c. component or 625-lines video with its negative d.c. component. Grid peaking components are switched in on 625 lines to improve the video response and the d.c. component is partially removed. Intercarrier traps at 3·5 Mc/s and 6·0 Mc/s are indicated in the video stage cathode although if the 625-line sound feed was from the video anode, the 6·0 Mc/s trap would be in the picture-tube feed. The 625-line

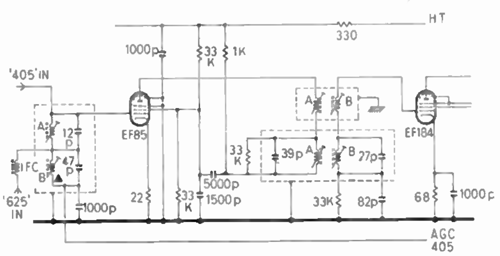


Fig. 5. Dual system sound i.f. stage.

4. Sound Detector

Figure 5 shows a sound input i.f. arrangement where the a.m. high frequency circuits and the f.m. lower frequency circuits are connected in series. Since the frequency difference is large, from 38·15 to 6·0 Mc/s, independence of tuning is obtained. The gain is good at both frequencies, that is, close to its 405-line counterpart. The bandwidth on 405 lines is 500 kc/s approximately to the detector, and on 625 lines must be wide enough to cover the deviation swing of ± 50 kc/s. The 625-line feed from the vision detector can be troublesome since it is also a feedback path for the a.m. sound carrier, vision carrier, or intercarrier harmonics. An r.f. choke reduces the feedback at the high-frequency end. Grid components are shown in the second stage, 33 kΩ and 82 pF, and with 25 mV or so at the first grid, limiting action at the second grid improves the a.m. rejection on 625 lines and backs up whatever discriminator is used.

Figure 6 shows one form of frequency discriminator and the usual 405-line circuit which could follow the arrangement shown in Fig. 5. The a.m. detector is shown in series with the ratio detector for 625 lines. In the latter case the de-emphasis components are switched in on the audio feed (47 kΩ and 1000 pF), and the 405-line detector is shorted to chassis together with the a.g.c. feed via the 1 MΩ resistor. Crosstalk in the detector stage is reduced by having separate

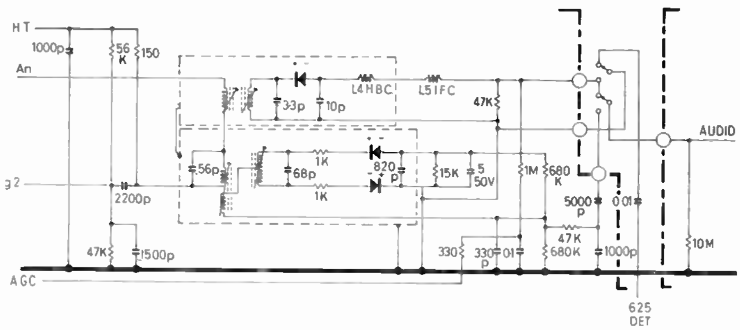


Fig. 6. Sound detector stage.

screening cans for the circuits, giving a figure of 60 dB suppression of the 405-line audio signal when the system switch is changed to 625 lines.

The arrangement shown is one of several possibilities for 625 lines. The Foster-Seeley circuit gives greater conversion efficiency and is good if preceded by a limiter for a.m. rejection. The conversion efficiency of the locked oscillator quadrature grid circuit, with the special pentode 6DT6 or EH90 arrangement, is even better, and very high audio outputs can be obtained with the i.f. circuits described. One disadvantage of the locked oscillator arrangement is that it has a definite input threshold below which it does not work and sufficient gain must be used after the video grid which has a sound component of about 25 mV to keep the working point above the threshold.

5. Cost Considerations

Some mention should be made regarding cost. For a dual standard receiver of comparable performance to its 405-line counterpart there can still be the same number of vision i.f. stages, although conversion of ordinary valve types to frame grid types is necessary to give an improvement of 6 dB approximately in gain bandwidth per stage. The same is largely true in the sound i.f. amplifier, where the introduction of frame grid valves makes up for slightly increased losses. The frequency discriminator for 625-line sound is an added requirement and six extra coils and screening cans are required, with associated components. A switch is required with something like eight functions for vision and sound response shaping and selection.

These practical solutions to problems in the design of the switchable i.f. amplifier stages suggest that the dual system arrangement need cost little more to produce in order to achieve the performance of a comparable single system i.f. amplifier.

6. Acknowledgment

Acknowledgment is made to the Directors of Pye Ltd., for permission to submit this contribution.

Manuscript received by the Institution on 5th March 1962. (Paper No. 765/T16.)

U.D.C. 621.397.22: 621.397.335

Video, Synchronizing and A.G.C. Circuits for Dual-Standard Television Receivers

By

J. H. HASLETT†

AND

P. L. MOTHERSOLE

(*Member*)†

Presented at a Television Group Symposium on "Dual Standard Television Receivers" in London on 7th February 1962.

Summary: The effects of different vision modulation polarities on the video amplifier and its associated a.g.c. system are discussed. It is shown that with mean-level a.g.c. some simple precautions are necessary to prevent blocking with negative modulation. Various circuit configurations are described and their limitations discussed. Flywheel oscillator circuit requirements are considered and a dual-standard line oscillator circuit is described.

1. Introduction

Television broadcast transmissions on a new standard may be introduced in the United Kingdom and as a result dual-standard receivers may be required. Although no definite information is yet available, it has been assumed for the purpose of this work that the standards on which the recent Stockholm agreements were based would be adopted, (i.e. 625-line, negative vision modulation, f.m. sound, 6 Mc/s vision-sound spacing in 8 Mc/s channels[1]).‡

It follows that a dual-standard receiver must incorporate facilities for appropriately converting: (1) tuner and i.f. amplifier response, (2) picture signal polarity and a.g.c. system, (3) line oscillator frequency and synchronizing circuit characteristic, (4) sound demodulator. This demands a multiple switching arrangement which will increase the cost of a receiver and which may lead to undesirable coupling and feedback between circuits. It has been the aim in this work to minimize the switching requirements.

It is also desirable that the user should not need to re-set any controls after system switching. In many sections of the receiver this conflicts with the aim of minimizing the number of switches and a compromise must be found.

2. Video Circuits

2.1. *Polarity of Modulation and D.C. Component*

The 405-line system currently in use employs positive modulation. This produces a positive d.c. component in the detector output, assuming the normal practice of applying a positive video signal to the video amplifier. Cathode bias is necessary for the video amplifier, generally with a current bleed from h.t., to accommodate this mode of operation (Fig. 1).

With a negative modulation transmission it is also common practice to apply a positive-going video signal to the video amplifier, but in this case the d.c. component is negative, that is, the entire positive-going video signal has a negative potential with

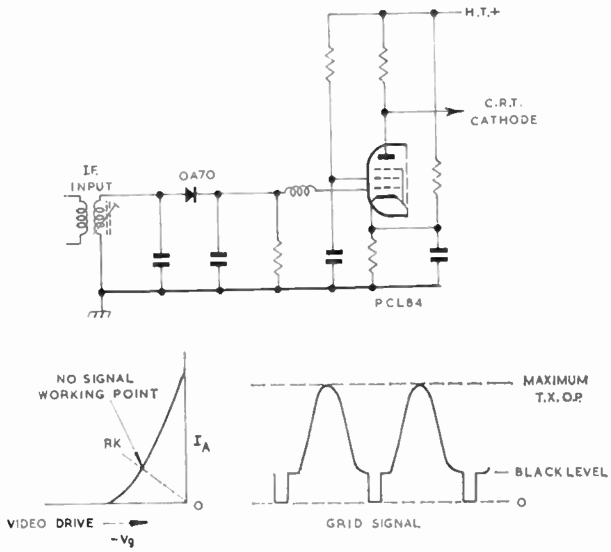


Fig. 1. Basic video detector and amplifier for positive modulation.

respect to earth. In this case no cathode bias is required on the video amplifier (Fig. 2).

If the d.c. component is to be retained in the video signal throughout the receiver it is thus necessary to switch the video amplifier cathode bias, and to proportion the bias so that, for a given signal amplitude, the d.c. level at the amplifier anode is the same on both systems. At the same time the polarity of the video detector must be reversed.

2.2. *Video Amplifier, D.C.-Coupled*

In the basic circuit of a dual-standard video detector and amplifier, the polarity of the detector output may

† Mullard Research Laboratories, Redhill, Surrey.

‡ In the recent Government White Paper (Cmnd. 1770) the change of standards has been confirmed.

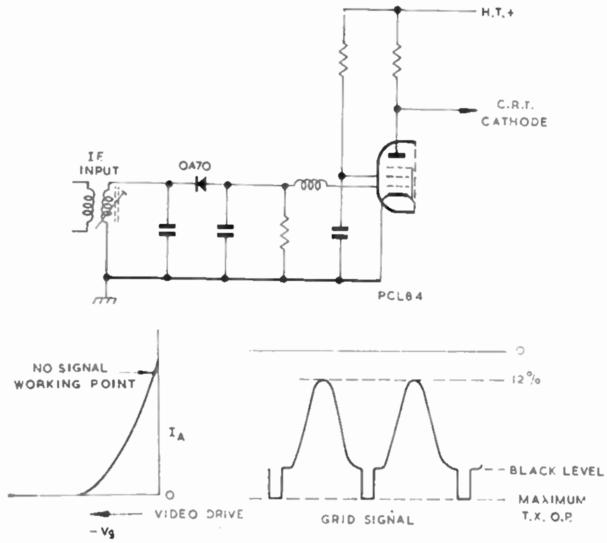


Fig. 2. Basic video detector and amplifier for negative modulation.

be changed by means of a two-pole switch, while the amplifier bias is switched out for operation with negative modulation. The video signal at the amplifier anode retains the d.c. component, and can be used to operate a black-level a.g.c. system.

With normal operating conditions the video amplifier in this circuit arrangement handles a similar signal in each case. However, maladjustment of the contrast control produces different types of overload distortion. With positive modulation peak white crushing occurs and the synchronizing pulses are unaffected. With negative modulation the synchronizing pulses are driven beyond cut-off causing the picture synchronization to fail. This condition is aggravated by mean-level a.g.c. and will be discussed later (Section 5.2).

Other differences in performance may arise from cathode degeneration on 405-line working, due to the cathode bias resistor, resulting in lower gain and improved amplifier linearity. The cathode resistor could be used for frequency response compensation but this may not be important as the amplifier must be compensated for the wider-band 625-line signal with the cathode earthed. Furthermore, the positive bias potential on the cathode detracts from the available anode swing.

2.3. *Video Amplifier, A.C.-Coupled*

The use of mean-level a.g.c. effectively removes the d.c. component from the signal and thus the possibility of a.c. coupling the video amplifier may be considered. With an a.c. coupling from the video detector as shown in Fig. 3, the signal handled by the amplifier is the same for both systems of modulation and a common bias circuit to give class A operation is possible. However, the 3·5 Mc/s sound trap must be switched out on 625-line operation and this trap is commonly incorporated in the cathode circuit. The same switch may be employed for modifying the bias circuit of the video amplifier and thus no significant saving in switching operations results from the use of a.c. coupling.

A further disadvantage of this circuit is that sudden changes in the picture signal average value cause the a.c.-coupled signal at the video amplifier grid to swing outside the normal operating limits. If objectionable overloading is to be avoided the normal working range must be restricted to about 70% of the usable amplifier characteristic. Alternatively, the d.c. component may be restored by a diode in the grid circuit.

2.4. *Contrast Control Considerations*

The requirements for a contrast control are to enable the user to adjust the amplitude of the video

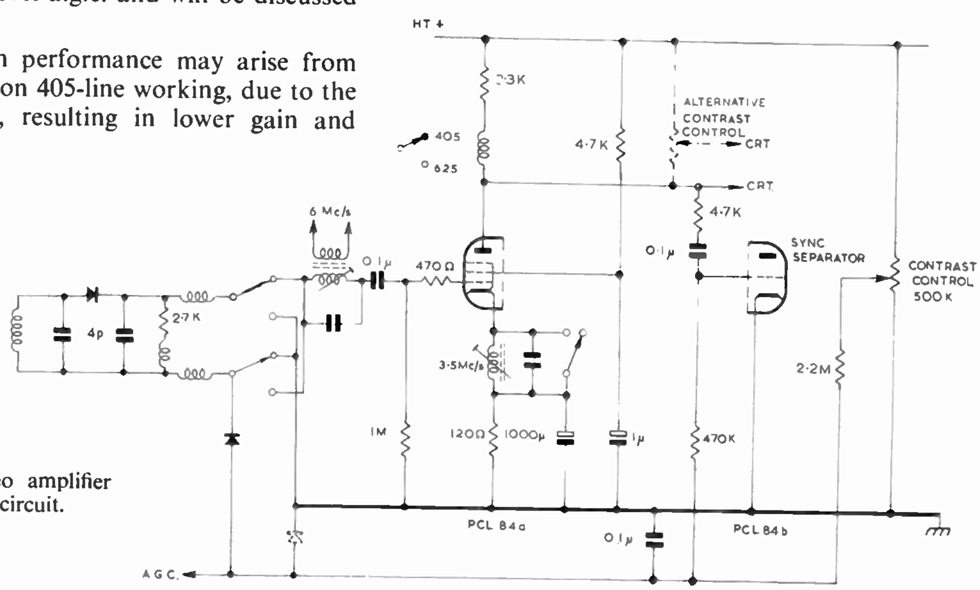


Fig. 3. A.c.-coupled video amplifier and mean-level a.g.c. circuit.

signal applied to the cathode-ray tube to any value up to the maximum available from the video amplifier.

A common method of contrast control involves the a.g.c. system, establishing a demand for a certain signal amplitude, which the a.g.c. system satisfies by adjustment of the receiver gain, i.e. the a.g.c. delay potential is used as a contrast control. With gated or black-level a.g.c. the respective requirements for positive and negative modulation systems cause a reversal in the sense of the manual control. As this is unlikely to be acceptable to the user it becomes necessary to introduce switching to reverse the connections to the contrast control.

This problem does not arise with simple mean-level a.g.c. since the a.c. coupling to the synchronizing pulse separator prevents the a.g.c. system responding to the difference in the sense of modulation.

A further complication arises with a dual-standard receiver when the d.c. component of the transmitted signal is retained by d.c. coupling through the video amplifier, together with the use of the a.g.c. delay potential as a contrast control. This is illustrated in Fig. 4. With positive modulation increasing contrast causes the black areas of the picture to become visible, i.e. brighter. With negative modulation, however, increasing contrast tends to cause the picture to black out. This difficulty is avoided by a.c. coupling the input to the video amplifier and using a d.c. restoring diode.

Alternatively, with mean-level a.g.c. the above effect may be prevented by reducing the d.c. component applied to the picture tube. With a gated or black-level a.g.c. system manual control of contrast may be effected by variation of the working condition of the video amplifier, e.g. by variation of the screen grid potential. The a.g.c. then operates effectively as a black-level clamp and this phenomenon is completely avoided. This feature is discussed fully in section 5.4.

Another alternative technique, applicable to any a.g.c. system, is the high-level contrast control. This consists of a potentiometer in the video amplifier anode circuit. The a.g.c. simply ensures the video amplifier receives a signal which fully modulates it. An advantage of this control technique is that the synchronizing pulse separator has applied to it the full video signal at all contrast control settings. The main disadvantage however, is that the control adds to the capacitive load of the video amplifier and may also limit receiver styling due to restrictions on positioning of the control. One possible solution to the latter problem is the use of a cadmium sulphide cell in a potential divider network, illuminated by a small lamp. The resistance of the cell is dependent on the illumination from the lamp which is adjusted by the manual contrast control.

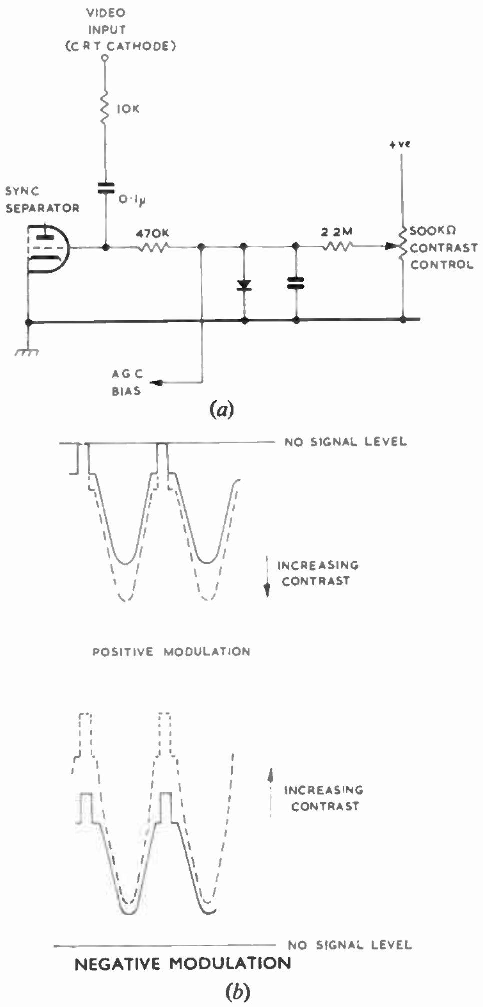


Fig. 4. (*a*) Simple mean-level a.g.c. circuit; (*b*) Video signal.

2.5. *Grid or Cathode Modulation of the Picture Tube*

An alternative to reversing the polarity of the signal applied to the video amplifier is to leave this unchanged and to switch the amplifier output to modulate the cathode-ray tube either on its cathode or on its grid (Fig. 5). In this way the d.c. polarity of the signal at the video amplifier grid is unchanged, but the sense of the video signal reverses from positive-going to negative-going with a change of system. Thus the video amplifier bias can remain unaltered on both standards, but any non-linearity of the amplifier will have different effects on the signals of different systems.

A black-level a.g.c. system can be operated from the pentode anode (Fig. 5) without any phase inverting problems since the sense of the signal change is always the same, i.e. an increase in signal causes the black level to go positive with either system. Some d.c. compensation is necessary however, due to the difference of the black levels.

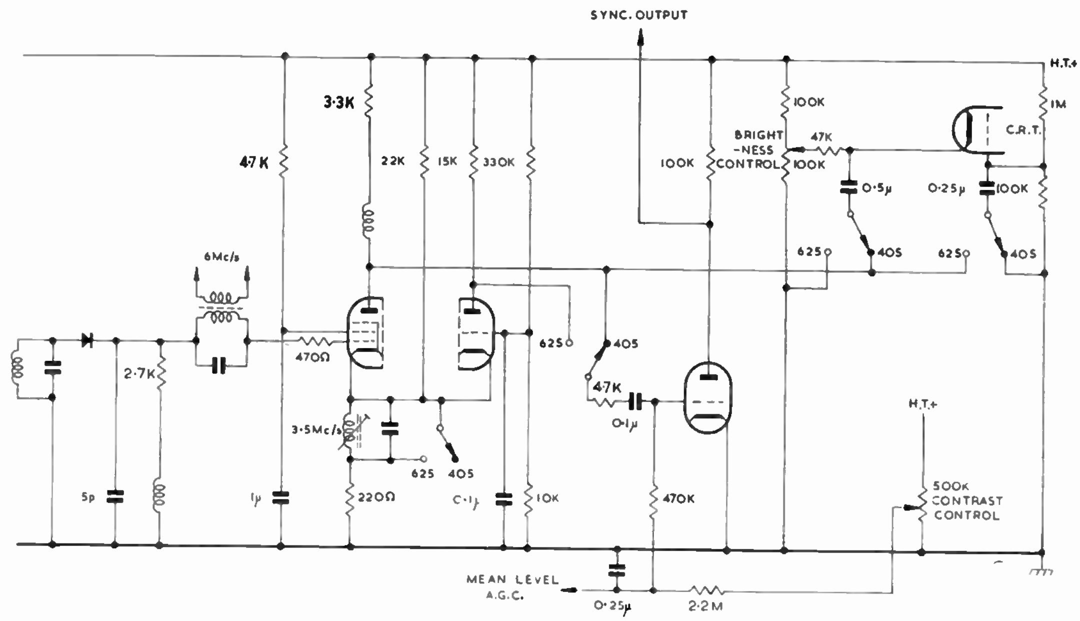


Fig. 5. Switched drive to cathode-ray tube with mean-level a.g.c.

In order to provide the same sense of signal to the synchronizing pulse separator a phase inverting stage is necessary. A mean-level a.g.c. system may then be operated from the synchronizing pulse separator in the conventional way.

Modulation sensitivity of the cathode-ray tube is greater when the signal is applied to the cathode than with the signal applied to the grid, since, in the former case, the potential between first anode and cathode is varied as well as the potential between grid and cathode. The difference amounts to some 30% and will result in a change of picture contrast with change of system even though the video signal amplitude is maintained constant by the a.g.c. action.

With d.c. coupling to the cathode-ray tube it has been found very difficult to provide a satisfactory control of brightness common to both systems, even with individual pre-set controls. As the provision of separate brightness controls for each system is unlikely to be acceptable to the user it becomes necessary to adopt a.c. coupling to the cathode-ray tube, as shown in Fig. 5. A single brightness control then operates in the cathode circuit.

When used in an experimental receiver with mean-level a.g.c., blocking was found to be a major problem. Large amplitude signals drive the pentode hard into conduction, cutting off the triode phase inverter. The loss of input to the synchronizing pulse separator prevents any a.g.c. potential developing and thus the receiver remains in an overloaded condition. A black-level or amplified mean-level a.g.c. system operated from the detector circuit can avoid this difficulty.

3. Synchronizing Pulse Separator Circuits

Synchronizing pulse separation can be achieved for both systems with common circuits, given suitable input signals to the separator stage. This is achieved basically by the reversal of the video detector output, but in the case of switched drive to the picture tube a phase inverting stage is necessary (Fig. 5). Time constants for line pulse differentiation and for frame pulse integration need not be changed, so switching in the synchronizing-pulse separator circuits is avoided.

It should be remembered that impulsive interference can cause severe disturbance of synchronizing pulse separator action with negative modulation systems. If negative modulation operation is confined to the u.h.f. transmission bands this problem is unlikely to arise,† but for dual-standard receivers operating on v.h.f. negative-modulation systems some form of noise protection for the synchronizing pulse separator is almost essential.[2] The heptode section of the ECH84 valve is designed specifically as a synchronizing pulse separator. In a dual-standard receiver, without noise gating, it can be used with advantage in a conventional circuit with g_1 earthed. Should the receiver be used in circumstances requiring noise protection the necessary components can be simply added.

Disturbance of line synchronization due to random noise will occur with both v.h.f. and u.h.f. transmission systems, and will be more noticeable on 625-line operation mainly due to the higher scanning speed and energy. This cannot be eliminated in a direct synchro-

† Based on previous limited Band V field tests.

nizing system but can be completely overcome by the use of a simple flywheel controlled line oscillator.

4. Flywheel Oscillator Circuits

4.1. *Oscillator Considerations*

The most commonly used oscillator circuits in flywheel systems are the anode or cathode-coupled multivibrators and the sine-wave oscillator. When a high impedance (i.e. diode) phase detector is used it is essential that grid current should not flow from the oscillator. A cathode-coupled multivibrator circuit or a sine-wave oscillator must therefore be used.

The free-running frequency of the sine-wave oscillator is dependent on the tuned circuit of which the reactance valve forms an important part. The free-running frequency of a multivibrator is governed by the grid circuit time-constants and the peak currents of both valves together with their cut-off potentials. To stabilize the long term oscillator frequency therefore, d.c. feedback or the operation of the valves in grid current is desirable.

These requirements cannot be simply satisfied with a high impedance diode phase detector and a stabilizing tuned circuit must therefore be incorporated into the multivibrator circuit. The most common arrangement, with a cathode-coupled multivibrator, is to connect a parallel tuned circuit in series with the anode load of the first triode.

The number of components required for both sine-wave stabilized multivibrator or sine-wave oscillator circuits is approximately the same. However, the frequency stability of the latter circuit tends to be higher since only one valve makes an active contribution. Furthermore, when a dual standard receiver (405/625 line) is considered, the simplicity of switching provided by the sine-wave oscillator is an important additional consideration.

4.2. *The Sine-wave Oscillator*

The circuit of a typical sine-wave oscillator is shown in Fig. 6. The pentode section operates as a screen coupled oscillator, the drive waveform for the line output valve being taken from the anode. The triode is used as a cathode driven capacitive reactance valve. The sine-wave potential developed across R3, connected in series with the tuned circuit, is in phase with the circulating current. This potential is applied to the triode cathode and is in quadrature with the anode potential. Thus the triode acts as a reactance valve, the reactance depending on the anode current. The internal resistance of the triode tends to damp the circuit. In order to compensate for this damping a potential must be applied to the triode to reduce its anode current as its anode potential swings positive. This may be achieved by applying a potential derived from the screen grid of the pentode to the triode grid, or alternatively from the cathode circuit of the pentode to the cathode circuit of the triode. In the circuit shown in Fig. 6 the latter technique is used, the value of the common cathode resistor R2 being chosen for optimum compensation.

4.3. *Dual Standard Circuit*

The main requirement, apart from achieving satisfactory stability of a line oscillator which is to function in a dual-standard receiver, is that the change from one standard to another should be accomplished with a minimum of switching operations. The frequency of the sine-wave oscillator may be changed over wide limits by simply switching an additional element to or from the tuned circuit with a switch that may be located at its "cold" end (Fig. 6).

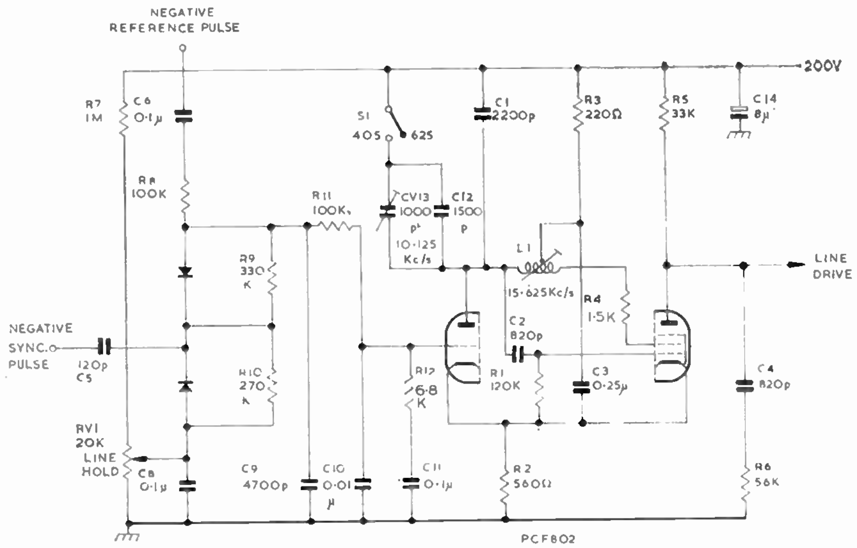


Fig. 6. Dual standard line oscillator.

A disadvantage of the simple single-ended phase detector circuit shown in Fig. 6 is that it produces a spurious output potential in the absence of the synchronizing pulses. This causes a deviation of the line oscillator phase whenever a block of synchronizing pulses is removed, e.g. during a severe burst of noise pulses. This phenomenon is unlikely to occur on u.h.f. and the simple circuit shown should be satisfactory for a dual-standard receiver. The values of R9 and R10 have been chosen such that the potential developed across the phase detector is zero with both the synchronizing and reference pulses applied.

5. Automatic Gain Control Circuits

5.1. *A.G.C. Requirements for Positive and for Negative Modulation*

An a.g.c. system is required to produce a negative control potential for the r.f. and i.f. stages dependent upon the magnitude of the received signal. This may be derived from the mean level of the demodulated signal or from the d.c. level of some specified part of the demodulated waveform. In either case some change of output signal amplitude with input signal level is unavoidable, in order to produce the required change of a.g.c. potential. The extent of the output change is dependent on the loop gain of the a.g.c. system; to achieve the most constant output the a.g.c. loop gain should be high.

An a.g.c. system dependent on the d.c. level of the demodulated signal normally needs to be switched between positive and negative modulation systems, since the d.c. potentials of the demodulated signals are opposite in the two cases. Such switching is not needed with mean-level a.g.c. systems or with a black-level system and a non-switched video detector (Section 2.5).

5.2. *Mean-level A.G.C. for Two Standards*

A mean level a.g.c. circuit is shown in Fig. 3, where switching of the video detector output is employed. In this case the synchronizing pulse separator operates in the conventional manner from signals coupled from the video amplifier anode, and no system switching is required.

Mean level a.g.c. systems are affected by overload of the video detector or amplifier. The effect is most severe with negative modulation, when synchronizing pulses may be cut off at the grid of the video amplifier in an overload condition. The synchronizing pulse separator then produces an unvarying d.c. potential at its grid, with the result that a.g.c. action is lost and the overload condition aggravated; the phenomenon is called blocking. One method of overcoming this difficulty is to provide a supplementary feed to the a.g.c. line consisting of the negative component of the video detector signal. This makes only a small contribution to the a.g.c. under normal working conditions, but provides sufficient control potential in the presence of large signals to prevent blocking.

With positive modulation this supplementary potential would be positive and would oppose the a.g.c. To avoid this a diode must be incorporated. The normal a.g.c. delay diode, shown dotted in Fig. 3, may be used for this purpose by connecting it as shown.

5.3. *Mixed A.G.C. Techniques*

In normal 625-line receivers it is usual to provide d.c. coupled a.g.c. which maintains the tips of synchronizing pulses at a constant level and thus maintains the gain independent of picture content; virtually constant black-level system. This method requires one valve function. A satisfactory black-level a.g.c. system is not possible with positive modulation using only one valve function. A possible solution is therefore to use a black-level circuit for negative modulation, switching it to an amplified mean level circuit for positive modulation.

A circuit is shown in Fig. 7, where it will be seen that the grid of the a.g.c. valve is d.c. coupled on 625-line operation to the video amplifier. The cathode is connected to a pre-set potential such that the valve just conducts on the positive-going tips of the synchronizing pulses. The a.g.c. circuit is operated by a floating d.c. potential obtained by rectifying the line flyback pulse with a voltage dependent resistor (v.d.r.). The a.g.c. valve operates as a controlled d.c. restorer to the potential developed across the v.d.r. As the valve is driven harder into conduction the potential developed across the v.d.r. increases. This d.c. potential is used as the a.g.c. control bias. An important feature of this circuit arrangement is that the a.g.c. potential is not affected by the line time-base synchronization.

On 405-line operation the video amplifier anode signal is integrated to produce a mean d.c. potential which is applied to the cathode of the a.g.c. valve. The grid is now taken to a pre-set positive potential, with the result that the anode current varies with the mean level of the signal and determines the negative a.g.c. potential produced.

Manual control of contrast is best effected by variation of the working point of the video amplifier. By this means the anode potential is varied causing the a.g.c. valve to adjust the a.g.c. potential. In the circuit in Fig. 7 the grid-cathode potential of the video valve is varied by means of a potentiometer in the cathode circuit.

The connections to the potentiometer must be reversed to obtain the same direction of rotation for increasing the picture contrast with the different polarities of vision modulation. This switching operation can also be combined with the sound

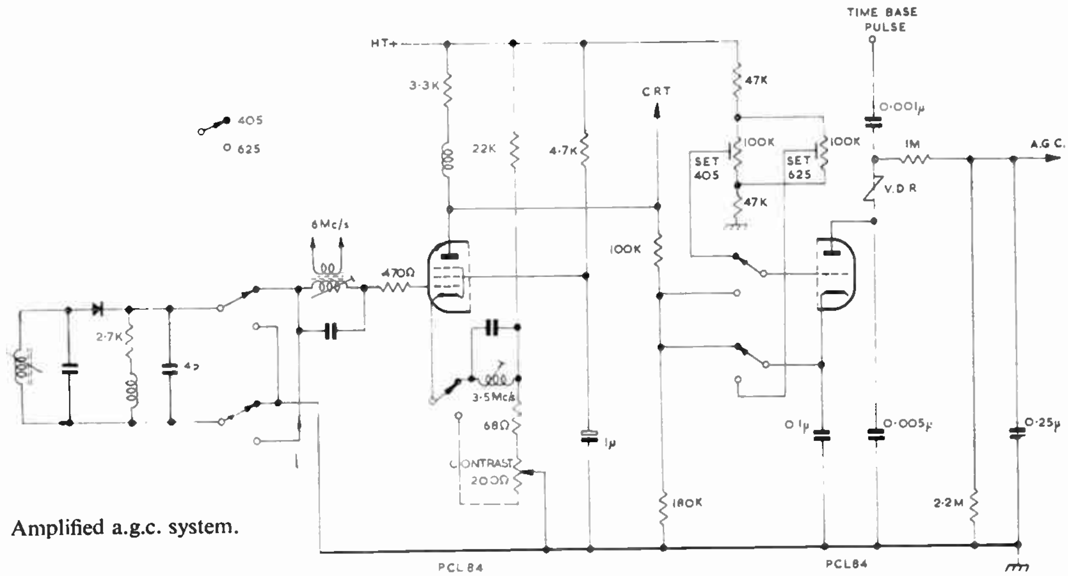


Fig. 7. Amplified a.g.c. system.

rejection circuit switching, as shown. When tested in an experimental receiver the circuit provided satisfactory a.g.c. operation on both standards.

5.4. *Dual-Standard Black-level A.G.C. System*

A double-triode a.g.c. circuit[3] which has been used in several commercial receiver designs can form the basis of a dual-standard black-level a.g.c. system.

A complete circuit is shown in Fig. 8. The video signal from the anode of the video amplifier has positive-going synchronizing pulses and is d.c. coupled to the grid of the black-level detector V2A via the 100 kΩ resistor. Also applied to this grid via the a.c. coupling and parallel RC network are the negative-going separated synchronizing pulses from the anode of the synchronizing-pulse separator valve. The resulting waveform at the grid of V2A has the synchronizing pulses cancelled and an overshoot added to the back porch of the video signal. V2A charges the cathode capacitance to a d.c. potential corresponding to the d.c. level of the back-porch overshoot. This d.c. potential is applied to the a.g.c. amplifier V2B.

With positive modulation increasing signal tends to drive the black-level detector output negatively and the a.g.c. amplifier is driven on its cathode. With negative modulation, due to the reversal of the video detector polarity, the output signal is positive-going and the a.g.c. amplifier must therefore be driven on its grid. Contacts D and E on the system switch S1

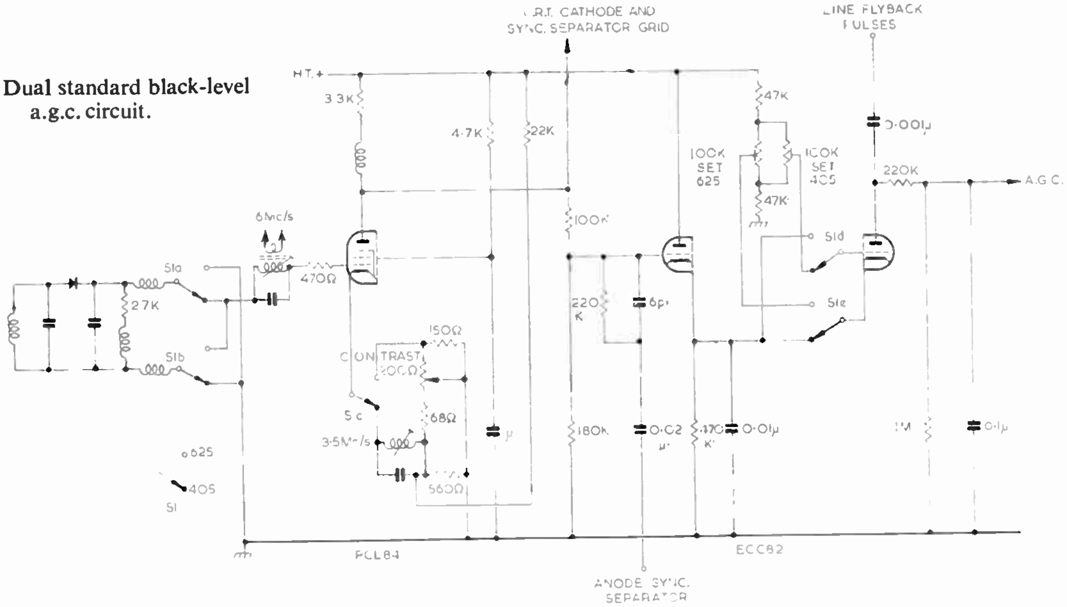


Fig. 8. Dual standard black-level a.g.c. circuit.

effect this change-over of drive and at the same time connect the grid or the cathode to an appropriate pre-set delay potential.

The a.g.c. amplifier V2B is driven into conduction by a pulse from the line output transformer. The valve acts as a d.c. restorer to this pulse, the anode current being controlled by the d.c. output of the black-level detector. The mean potential of the anode waveform, which is negative due to the d.c. restoration action of V2B, is used as the a.g.c. potential.

Manual contrast control as in the previous circuit is best effected by variation of the working point of the video amplifier. By this means the anode potential is varied causing the a.g.c. system to adjust the a.g.c. potential. This action is illustrated in Fig. 9. Variation of the video amplifier bias by the manual contrast control causes the signal level at the anode to change. The a.g.c. system acts as a black-level clamp, the shift in the black level being only that required to operate the a.g.c. circuit to modify the receiver gain. In practice, due to the high loop gain in the system, the black-level shift is very small. The separate pre-set controls enable the a.g.c. working point to be set to the same level for both systems, giving consistent operation of the common contrast control. The fixed series and shunt resistors in the contrast control circuit maintain a similar range of control for both systems.

The a.g.c. performance on both systems was adequate in an experimental receiver, the full range of automatic control being obtained for a very small change in the video signal amplitude. Some care is necessary in proportioning the feedback network from the anode of the synchronizing pulse separator valve and in the design of the separator circuit to avoid low-frequency instability at very low settings of the contrast control.

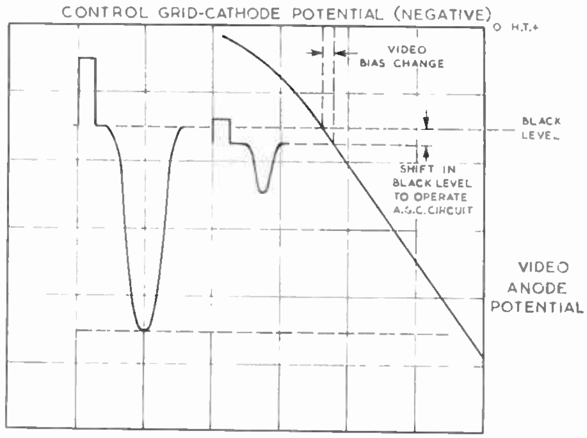


Fig. 9. Contrast control with black-level a.g.c.

6. Conclusions

Problems of dual-standard operation have been considered for the video, synchronizing and a.g.c. sections of the receiver. It has been found possible to combine satisfactorily established techniques currently used with both systems of vision modulation.

The only changes necessary in the video amplifier are the variation of bias and switching of sound traps between systems, provided the overall bandwidth is adequate for the wide-band modulation. These two functions can normally be combined in a single switch operation.

Simple unprotected synchronizing pulse separator circuits can be used for both systems. If negative vision modulation on v.h.f. is considered, however, some form of noise protection for the separator is essential.

Flywheel synchronizing may also be desirable for 625-line operation since "ragging", due to h.f. noise, is more noticeable at the faster scanning speed. It has been shown that the sine-wave oscillator is very satisfactory for use in a dual-standard receiver. A circuit for a dual-standard line oscillator, including a simple phase detector, has been described.

A simple mean-level a.g.c. circuit can be used on both systems provided precautions are taken to prevent blocking with negative modulation. It is also desirable to reduce the d.c. component of the video signal applied to the picture tube if d.c. restoration is not employed. A dual standard a.g.c. circuit has been described, based on a previous design for positive modulation, which maintains a constant black-level and prevents blocking.

Multiple switching is unavoidable in a dual-standard receiver, but the number of switch functions might be reduced at the expense of user convenience. Furthermore, in a complete receiver design some switches can be arranged to perform more than one function.

7. Acknowledgments

The authors would like to acknowledge the contributions made to the work described in this paper by their colleagues, in particular Mr. C. E. Longhurst. They also wish to thank the Director of the Mullard Research Laboratories and the Directors of Mullard Ltd., for permission to publish this paper.

8. References

1. Final Acts of the European VHF/UHF Broadcasting Conference, Stockholm 1961. International Telecommunications Union, Geneva.
2. M. Marks, "Noise immune sync. separator", *Electronics*, **25**, No. 4, pp. 124–27, April 1952.
3. P. L. Mothersole, "Automatic gain control circuits in television receivers for negative modulation systems", *J. Brit. I.R.E.*, **18**, pp. 307–16, May 1958.

Manuscript received by the Institution on 14th April 1962. (Paper No. 766/T17).

U.D.C. 621.375.9

Optical Masers

By

J. H. SANDERS, M.A., D.Phil.†

An introduction to the Symposium on "Masers and Lasers" to be held in London on 2nd January 1963.

Summary: This review of the present situation in the optical maser field discusses the principle of the maser and its extension to the infra-red and visible region of the spectrum. Various types of optical maser which have been successfully operated are described. The unique features of high spectral purity and narrow beamwidth are pointed out, and some present and future applications are discussed.

1. Introduction

The optical maser is an oscillator with an output in the visible or infra-red region of the spectrum. The radiation from an optical maser is produced by the process of stimulated emission, thus distinguishing it from conventional light sources in which the basic process is spontaneous emission. The latter is an incoherent process, that is, the atoms or other radiating systems emit their radiation at random times and in random directions. Stimulated emission under the conditions existing in an optical maser is, however, a coherent process; atoms emit in phase with each other and in a well-defined direction. This accounts for the striking features of the optical maser: the very narrow spectral range of the emission and the near-parallelism of the output beam. These, coupled with the very large power output obtainable from some models, lead to a number of unique applications.

Two different types of optical maser have been operated so far, those using a crystalline solid, and those using a gas discharge. The solid type is basically a pulsed device, though some of them have been operated continuously, whilst those using a gas discharge are normally run continuously. The basic principles of both are similar.

2. The Maser Principle

The light from a lamp of the discharge type arises from transitions between two energy states of an atom, molecule or ion. Figure 1 shows diagrammatically two allowed energy states 1 and 2 of such a system, the energy states being separated by an energy difference ΔE. The frequency ν of the quantum of radiation which has an energy ΔE is given by the well-known relation $\Delta E = h\nu$, where h is Planck's constant. Thus a system in the higher energy state 2, when left undisturbed, makes a transition to state 1 and emits a photon of radiation of frequency $\Delta E/h$. The average lifetime of the system in state 2 varies considerably with the nature of the energy levels under consideration and on the value of ΔE, but for levels which are relevant to this discussion typical values may be taken to be 10^{-8} second for an atom in a gas discharge and 10^{-3} second for an ion in a solid. This process is known as *spontaneous emission*, and excited systems emitting photons in this way do so completely at random.

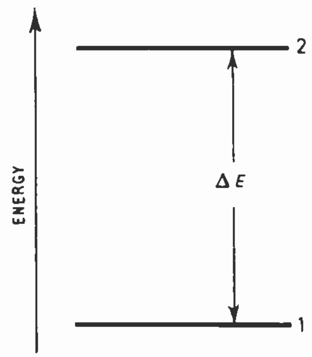


Fig. 1.

In practice the radiation emitted is not of a single frequency ν, but has a distribution over a band of frequencies $\Delta\nu$ centred about this frequency. In a gas discharge the width of the observed radiation is usually mainly due to the Doppler effect, as a consequence of the random thermal motion of the atoms in the discharge, and is in the most favourable case at least 1000 Mc/s. This is, however, a comparatively narrow spectral line, since the frequency of the line under consideration, in or near the visible spectrum, is between 10^{14} and 5×10^{14} c/s. In a crystal the energy levels are generally broadened by the internal fields of the lattice into bands covering a considerable region of the spectrum. In certain cases, however, the levels may lead to spectral lines as narrow as 100 000 Mc/s, and these are the favourable cases for optical masers, as will be shown later.

Spontaneous emission, so far considered, occurs when the system is completely isolated. Next, the situation will be considered in which the system is in an electromagnetic field of frequency ν, that is, it is

† Fellow of Oriel College, Oxford; University Demonstrator in Physics, Clarendon Laboratory.

irradiated with light of the same frequency as it would spontaneously emit in going from state 2 to state 1. If the atom is in state 1 it may absorb a photon of the radiation and be excited to state 2. This is the well-known *absorption* process responsible, for example, for the dark Fraunhofer lines which cross the continuous emission spectrum of the sun, the absorbing atoms in this case being situated in the outer layers of the sun. Absorption occurs to an appreciable extent only when the incident radiation has the frequency $\Delta E/h$, or is within the range $\Delta\nu$ discussed above.

A third process occurs when a system is in the higher energy state 2 and is irradiated with light of the correct frequency. The system may then fall to the lower energy state 1 and emit radiation of frequency $\Delta E/h$. This process of *stimulated emission* is the converse of absorption; the system loses energy and the radiation field gains energy. The probability of a system in state 2 giving up energy to a given radiation field is exactly equal to the probability of a system in state 1 absorbing the same amount of energy when placed in the same field. Thus if a beam of radiation were passed through a region containing N_1 systems per unit volume in state 1, and N_2 in state 2, the beam would be attenuated if $N_1 > N_2$, but would be amplified if $N_2 > N_1$. In the former case absorption predominates over amplification, and in the latter case the opposite is true. It is important to remember that spontaneous emission from state 2 occurs in addition to stimulated emission, but unlike stimulated emission, which adds its energy to the beam which causes it, spontaneous emission takes place equally in all directions from a randomly oriented collection of atomic, molecular, or ionic systems.

3. The Cavity

The microwave maser, as is well known, uses the principle described above to produce amplification and oscillation at frequencies of the order of 10–100 kMc/s, and gave the name to devices of this sort (*M*icrowave *A*mplification by the *S*timulated *E*mission of *R*adiation). In 1959 Schawlow and Townes[1] suggested the extension of the method to the infra-red and visible regions; the main interest here lies in oscillators rather than amplifiers and further discussion will be limited to such devices.

In the microwave maser, the amplifying medium, having $N_2 > N_1$, is placed in a cavity tuned to the frequency ν which corresponds to the energy difference ΔE between the levels 1 and 2. So long as the rate of stimulated emission is great enough to equal the power loss in, or from, the cavity the device oscillates. The provision of a suitable cavity for a wavelength of the order of 10^{-4} cm presents problems rather different from those encountered in microwave work. In order to accommodate sufficient amplifying medium

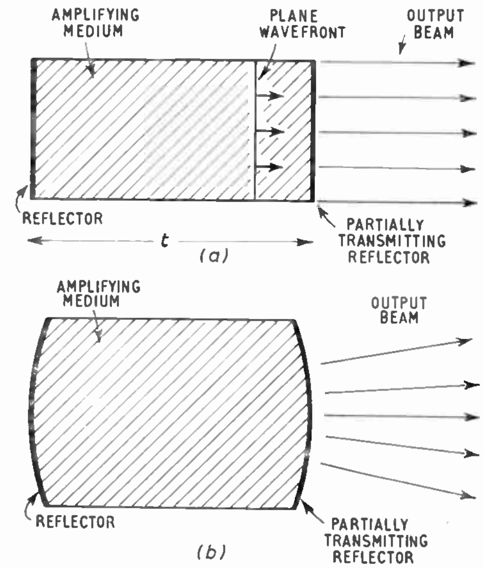


Fig. 2. Schematic diagram of optical masers using (*a*) plane reflectors and (*b*) confocal reflectors.

it is necessary to make the length of the cavity of the order of 1–100 cm, so that it is very many wavelengths long. A hollow cavity with a volume of 1 cm^3 operating at a wavelength of 10^{-4} cm has resonant modes separated by about 10 c/s, so that, within a band of, say, 1000 Mc/s over which the medium amplifies, the number of modes in which oscillation might occur is very large. Under these conditions the power in each mode would be a very small fraction of the total, and the radiation in the cavity would be essentially isotropic, and two advantages of the maser would be lost.

The form of the cavity used for an optical maser consists of a pair of reflecting surfaces, usually plane, placed parallel and facing each other as shown in Fig. 2 (*a*). This type of cavity has no reflecting sides, as in the microwave case, and consequently the only modes which have a high Q are those in which a wave is reflected to and fro, travelling normally, or nearly normally, to the reflecting surfaces. Fox and Li[2] have investigated the resonant modes of such a cavity. The modes of highest Q are those which approximate to a plane wave travelling normally to the plates, having almost constant amplitude except near the outer portion, where there is a loss of energy due to diffraction. The resonant wavelengths of such a mode are given very closely by

$$2\mu t = n\lambda$$

where λ is the wavelength measured in vacuum, μ is the refractive index of the medium between the plates, t is the separation of the plates and n is an integer. The frequency separation of these axial modes is thus $c/2\mu t$ where c is the velocity of light; for an evacuated cavity 1 cm long the frequency separation is thus 15 000 Mc/s, so that one, or only a few, modes fall within the line width of the amplifying medium. There are also modes with a more complicated distribution of amplitude and phase across the cavity, but these are of lower Q and consequently are less likely to be excited.

In an axial mode the Q of a cavity of this type having two surfaces with an intensity reflection coefficient R is equal to $2\pi t/[(1-R)\lambda]$, which may have a value between 10^8 and 10^9 for a cavity such as is used in the gas discharge maser described below. In this expression only reflection loss is considered. Diffraction loss is generally small, but in a solid maser there may be appreciable loss due to scattering of light in the solid.

The output from such a cavity is obtained by making one end, or both ends, partially transmitting. This is achieved, in the case of metallic reflecting coatings, by making the film sufficiently thin to allow partial transmission. When multi-layer dielectric coatings are used they are inherently partially transmitting, being composed of transparent materials. In practice a higher reflectivity is obtainable by the use of multi-layer dielectrics, but metallic films are often adequate and are certainly easier to produce. In the case of the plane-ended cavity operating in the simplest mode, the output is in the form of a plane wave, i.e. a parallel beam. It is this feature which gives the optical maser one of its striking properties, the small beam divergence. Ideally the output wave is coherent over the whole of the aperture of the maser, so that the angular divergence of the beam is limited only by diffraction and is of the order of λ/W radians, where λ is the wavelength of the radiation and W is the beam width. This divergence is typically very much smaller than can be achieved with any other source of comparable power, and the beam can be focused to a very small cross-sectional area where the power density is very high.

Another type of cavity consists of two spherical reflectors each of radius of curvature r placed a distance r apart, as shown in Fig. 2 (*b*). The resonant modes in this case have been investigated by Boyd and Gordon.[3] This confocal type of resonator has lower diffraction loss than the plane type, but unless the aperture is severely restricted there are a large number of high Q modes with complicated intensity distributions and the advantage of a single plane wave output is lost. It is, however, far less critical in alignment than the plane-parallel cavity.

4. The Condition for Oscillation

If we consider N_2V excited systems in state 2 (Fig. 1), distributed uniformly in a cavity of volume V, they may radiate spontaneously, or be stimulated to

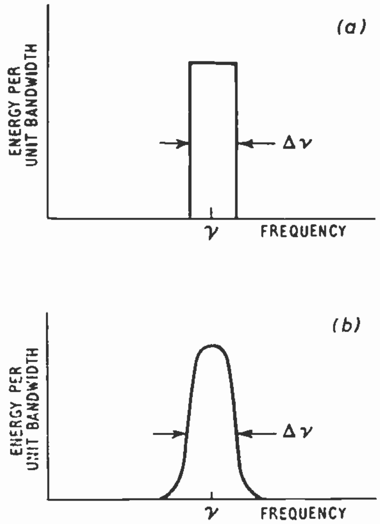


Fig. 3. (*a*) Hypothetical spectral line. (*b*) Typical spectral line as observed in practice.

radiate by any photons of the correct frequency which exist in the cavity. For simplicity we shall consider that the spectrum of the spontaneous radiation is rectangular in shape (Fig. 3(*a*)) and of width $\Delta\nu$. Theory[4] shows that the rate of spontaneous emission is equal to the rate of stimulated emission that would occur if there were just one photon in each resonant mode of the cavity; it also shows that the stimulated emission takes place into the same cavity mode as contained the stimulating photon. If τ is the lifetime of the excited state when spontaneous emission is the only process occurring, the rate at which the systems radiate energy spontaneously is $N_2Vh\nu/\tau$. Consequently, if $n\Delta\nu$ is the number of cavity modes in the frequency interval $\Delta\nu$, the rate of stimulated emission when there is just one photon in one of the cavity modes is $N_2Vh\nu/(n\Delta\nu\tau)$. If oscillation takes place in this mode the rate of stimulated emission must be at least equal to the rate at which energy is lost from the cavity. The latter, in terms of the Q of the mode, is $h\nu 2\pi\nu/Q$, so that the condition for oscillation is

$$\frac{N_2Vh\nu}{n\Delta\nu\tau} \geqslant \frac{h\nu 2\pi\nu}{Q}$$

or

$$N_2 \geqslant \frac{2\pi\nu n\Delta\nu\tau}{VQ}$$

For a cavity with a large number of modes we have[5]

$$n = \frac{8\pi\nu^2}{c^3} V$$

where c is the velocity of light. Thus the condition for oscillation becomes

$$N_2 \geqslant \frac{16\pi^2\nu^3\tau\Delta\nu}{c^3Q}$$

So far, a cavity containing only systems in the upper state 2 has been considered. If some systems in the lower state 1 are present, radiation is absorbed by them, and the rate of absorption due to each of these systems is equal to the rate of stimulated emission from each of the systems in state 2. If N_1 systems are present in unit volume of the cavity the condition for oscillation becomes

$$N_2 - N_1 \geqslant \frac{16\pi^2 \nu^3 \tau \Delta\nu}{c^3 Q}$$

For a case where the spectral line is not rectangular, but is of Gaussian shape (Fig. 3 (*b*)) such as is produced by Doppler broadening, this result must be multiplied by a factor of $(\pi/\ln 2)^{\frac{1}{2}}$ which is close to unity.

The expression shows how a spectral line must be chosen, in terms of its characteristic width $\Delta\nu$ and lifetime τ, in order to keep $(N_2 - N_1)$ to a minimum, and it also shows that Q must be high. Other things being equal, $(N_2 - N_1)$ at the threshold of oscillation rises rapidly with frequency, and this illustrates the difficulty of obtaining oscillation in the blue and ultraviolet compared with the red and infra-red.

Under conditions of thermal equilibrium an assembly of systems at an absolute temperature T have a population ratio given by the Boltzmann relation

$$N_2/N_1 = e^{-\Delta E/kT}$$

where k is Boltzmann's constant, so that it is clear that in this case N_2 is always less than N_1, and the condition for oscillation can never be satisfied. The non-equilibrium situation in which $(N_2 - N_1)$ is a positive quantity is often termed an *inverted population*, and, by reference to the Boltzmann relation, a condition of *negative temperature*. The achievement of an inverted population is one of the problems of maser design.

5. The Pulsed Ruby Maser

The first type of optical maser in which oscillation was produced used the energy levels associated with the Cr^{+++} ion in ruby. The type of ruby used consisted of a single crystal of aluminium oxide (Al_2O_3) containing about 0·05% molar of chromium oxide (Cr_2O_3). The energy levels of the chromium ions is shown in Fig. 4. When ruby is illuminated with white light it absorbs in the green part of the spectrum (hence the red appearance of ruby by transmitted light), and some of the Cr^{+++} ions are excited to the $^4F_2$ band. The excited ions decay by a non-radiative process (giving heat to the crystalline lattice) to a pair of sharp levels labelled ^2E. Decay from these levels gives, at room temperature, a pair of sharp emission lines at 6943 and 6929 Å, the fluorescence lines of ruby, and the ions return to their ground state. The ground state and the lower of the ^2E levels, responsible for the 6943 Å line, are the "maser" levels.

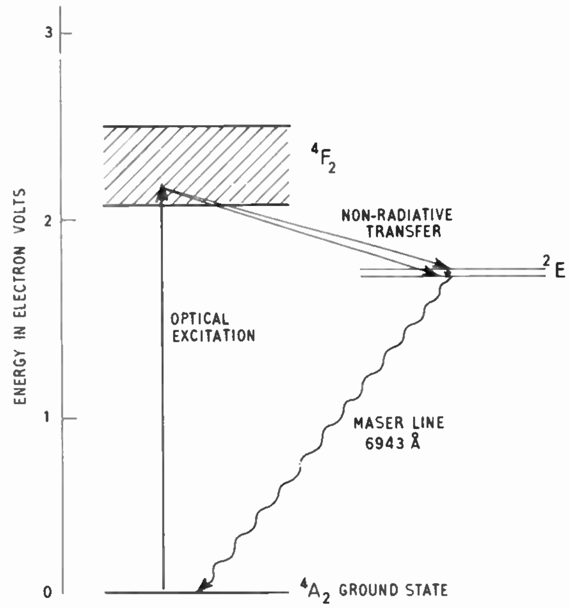


Fig. 4. Energy levels of the Cr^{+++} ion in 0·05% Cr ruby, and optical maser processes.

An optical maser of this type consists of a circular cylinder of ruby about 5 cm long and 5 mm diameter with its ends polished plane and parallel and silvered. In the pulsed version the ruby is illuminated by a flash of light from an electronic flash tube, the light entering the rod through the side. If sufficient light power falls on the ruby more than half the Cr^{+++} ions present are raised from the ground state into the ^2E levels via the $^4F_2$ band, resulting in an inversion of population between the ground state and the upper maser level. The ruby rod with its silvered ends acts as a resonant cavity, and as long as the inversion of population is maintained maser oscillation takes place and a coherent beam of visible red light at 6943 Å emerges from one or both ends, depending on whether one or both is partially transmitting.

The necessary light power to achieve the required condition for oscillation for a given specimen of ruby depends on the efficiency of light energy transfer from the lamp into the ruby rod. Some models of maser use a helical flash tube placed round the outside of the ruby (Fig. 5 (*a*)) but a reduction in the threshold energy needed for the lamp can be achieved by using a straight flash tube at one focus of an elliptical cylinder reflector, and the ruby rod at the other focus (Fig. 5 (*b*)). Typical flash tube energies range from 100 to 1000 joules, the duration of the flash being about 1 ms. With the higher flash energies the output light from the ruby maser may be as much as 30 joules in 1 ms, a power of 30 kW, emerging in a beam having a cross-sectional area of about $\frac{1}{4}$ cm^2. The light flux is thus about a million times that of sunlight.

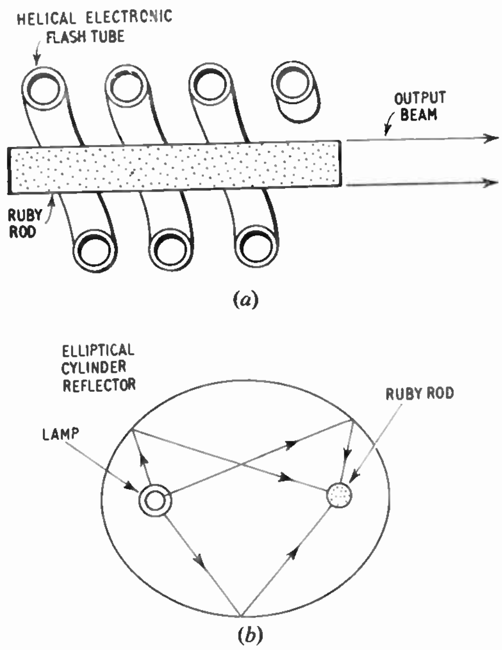


Fig. 5. (*a*) Ruby maser excited by helical flash tube. (*b*) Method of exciting ruby maser by use of linear flash tube in hollow elliptical cylinder.

The pulsed ruby maser may be operated at room temperature, and the allowable flash repetition rate is limited by the heating of the crystal when it is illuminated. Effective cooling permits a few flashes per minute. Reducing the temperature of the ruby narrows the line width of the spontaneously radiated light and so reduces the threshold flash intensity needed to produce oscillation. At room temperature the light output from the ruby is not a smooth pulse lasting approximately the same time as the illuminating pulse, but consists of a number of more or less random sharp pulses each about 1 μs long, known as "spikes" (Fig. 6). The mechanism responsible for these spikes seems to be a rapid building-up of excited atoms beyond the equilibrium density, followed by a large stimulated light intensity which reduces the density of excited atoms below the threshold, a process which has been likened to the "squegging" of a vacuum tube oscillator. The "spikes" disappear on cooling the ruby to liquid nitrogen temperature, and the light output is then a single smooth pulse of a duration which depends on the time during which the incident light maintains the inversion of population.

By using a technique involving a ruby rod which has one reflecting coating, and the other end clear but facing an electro-optical shutter and a mirror, very short pulses of light may be obtained. The ruby is illuminated in the usual way with a flash tube, but no oscillation occurs if the shutter is kept closed. When a large inversion of population has been built up in the ruby the shutter is opened, and the ruby is de-excited in one oscillatory burst. In this way flashes giving 0·2 joules of light in 3×10^{-8} seconds have been obtained at Hughes Research Laboratories. The light flux is about 10 MW/cm^2 during the flash.

The beam from a ruby maser is not coherent over the whole of the aperture of the crystal because of local variations of refractive index of the crystal, due mainly to variations in the distribution of the chromium in the ruby. At the normal concentration of chromium used (about 0·05%) a variation of 1% of this value gives a 25 parts per million variation in the refractive index, sufficient to change the optical length of the crystal by several wavelengths. Coherence is in practice maintained only over limited areas of the output wavefront, of the order of 1 or 2 mm in diameter. For this reason the beam divergence is limited to between 10^{-2} and 10^{-3} radian; nevertheless, the effect of focusing such a beam is to produce extremely high energy flux at the focus. In the two examples which have already been quoted, the maser giving 30 joules in 1 ms would give a flux of about 3×10^{10} watts/cm^2

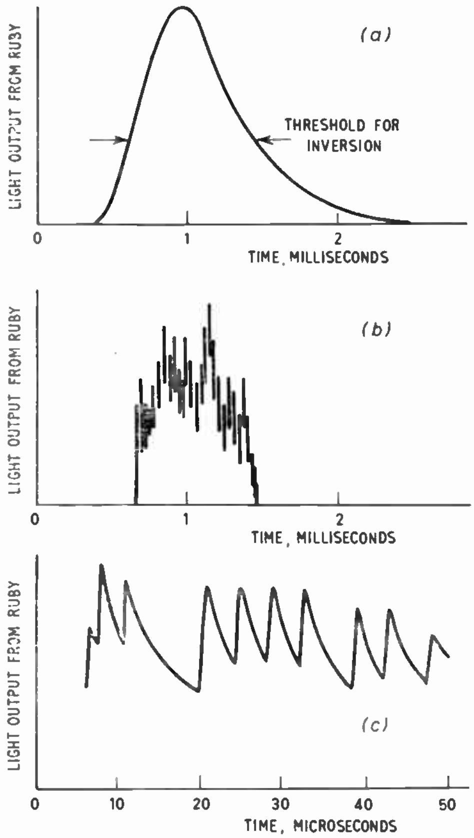


Fig. 6. (*a*) Typical light output curve from electronic flash tube. (*b*) Typical "spiky" output from ruby maser operated at room temperature.

(*c*) As (*b*), but with expanded time scale.

at the focus, while the output of 0·3 joule in 30 ns would give a flux of about 10^{13} watts/cm^2. The latter corresponds to a radiation pressure of over 1000 atmospheres.

The spectrum of the output from a ruby maser consists of a single very narrow component when the inversion of population is just above the threshold of oscillation. At higher levels of inversion a number of axial modes are excited, separated in frequency by, typically, about 2000 Mc/s. Each of these modes has been shown to be less than 12 Mc/s wide, compared with the width of the spontaneous emission from fluorescing ruby of about 200 000 Mc/s. This is an example of the line narrowing familiar in all oscillators, which in the absence of any disturbing effects such as noise would produce an output at a single fixed frequency.

The frequency of a ruby maser can be varied by a small amount by changing the temperature of the crystal, but in the range from room temperature to −180°C the change is only 600 000 Mc/s, or about 10 Å, a fractional variation of approximately 0·1%.

6. Other Crystalline Solid Masers

Besides ruby, a number of other solids have shown maser oscillation using excitation by irradiation with light. All these materials consist of a host lattice, such as aluminium oxide, calcium fluoride, calcium tungstate, calcium molybdate, or glass, containing a small percentage of a foreign ion. The ions which have been successfully used in this way, and the wavelengths produced, are summarized in Table 1. The production of crystals containing the required concentration of ions in the correct valence state, and having adequate optical homogeneity, is a considerable art, and there is scope for much development work in this field. Imperfections, such as bubbles, striations, and mechanical strains, scatter light and so increase the oscillation threshold, and distort the output wavefront. Typically, the ends of a maser rod are made parallel to a few seconds of arc and are flat to 1/10 or 1/20 wavelength. To a certain extent variations in the refractive index of the medium can be corrected by optically working the ends in order to achieve constancy of optical path through the crystal.

Table 1

Ion	Wavelength microns
Cr^{+++}	0·694
U^{+++}	2·61
Sm^{++}	0·703
Tm^{+++}	1·91
Ho^{+++}	2·05
Nd^{+++}	1·06
Pr^{+++}	1·05

Continuous operation has been achieved in some of these materials by illuminating the maser crystal with a very intense steady light source.[6] The power output in these cases is at most a few milliwatts. Materials which have been successfully used in this way are ruby, Nd^{+++} in calcium tungstate and U^{+++} in calcium fluoride, the first two having low enough thresholds at room temperature to achieve continuous operation when illuminated by a lamp dissipating 1–2 kW.

7. The Gas Discharge Maser

The pulsed type of solid maser is so far pre-eminent as a high power source of optical radiation. In its low power c.w. form it has the merit of simplicity and ruggedness. Both, however, have the disadvantage (which may well be overcome by technical advances) that the working medium, the crystal, is not optically perfect so that the line width and beam spread suffer as a result. This drawback is not present in the gas discharge maser, which demonstrates some potentialities of optical masers in a remarkable way.

Inversion of population has been achieved in a gas discharge, initially by Javan and his co-workers,[7] in a mixture of helium and neon. The energy levels of these two gases (Fig. 7) show a near-coincidence

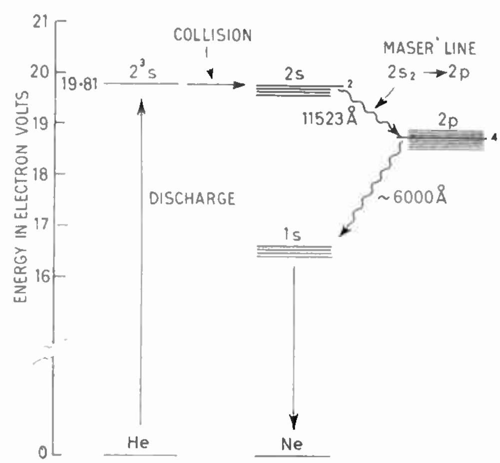


Fig. 7. Energy level diagram of helium (on left) and neon (on right) showing processes that occur in the He-Ne maser.

between the 2^3S metastable state of helium and the 2s group of levels in neon. In a discharge a large population density of metastable atoms builds up, and on colliding with unexcited neon atoms they transfer their energy to the neon atoms, selectively exciting the 2s levels. Among the transitions from these levels to the lower 2p states is one giving radiation at 11523 Å, approximately 3×10^{14} c/s; the upper state of the two involved has a longer lifetime than the lower, so an inverted population results. Under favourable conditions this can result in amplification of 11523 Å radiation of about 20% per metre of path

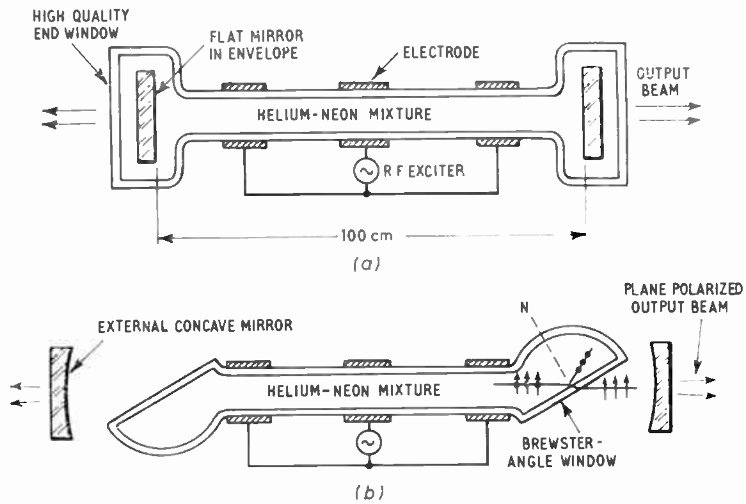


Fig. 8. Two types of He-Ne maser:
(*a*) internal reflector type using plane mirrors;
(*b*) external reflector type using confocal concave mirrors.

through the discharge, ample for maser operation. Several versions of optical masers have been constructed using a He-Ne discharge, using either reflecting surfaces inside the low pressure gas (Fig. 8 (*a*)), or external reflectors (Fig. 8 (*b*)). In the latter case the windows at the ends of the discharge tube have their surfaces oriented at the Brewster angle to the maser light so that, for one plane of polarization, loss due to reflection at the windows is eliminated. In the original version of this type of maser the reflecting surfaces were flat to better than 1/100 wavelength and had multilayer dielectric coatings giving 98·9% reflectivity in the region of 11500 Å, but subsequent experience shows that oscillation can be achieved with somewhat less rigid requirements. The use of confocal, rather than plane, reflectors reduces the requirements on reflectivity and alignment of the surfaces.

The spacing of the reflectors is commonly one metre, and the optimum filling seems to be 90% He, 10% Ne with a total pressure of 1–2 mm Hg. The gas has to be free of impurities in order to achieve inversion of population, and an r.f. electrodeless discharge is used, at a power input of between 10 and 100 watts. The power output is of the order of 10 mW in a beam about 1 cm diameter having a divergence of about ½ minute of arc. Oscillation occurs in one or more axial modes of the resonant cavity depending on the degree of excitation, and therefore the amplification, of the discharge. In the case of a 1 metre separation of the reflectors the mode separation is 150 Mc/s, and observation of beats between adjacent modes shows that the line width of the individual modes is, in the absence of vibration, less than 1 c/s. Vibration causes the separation of the reflectors, and hence the resonant frequency of the cavity, to change and it is difficult to reduce the frequency jitter due to this cause to below 1 kc/s, which results from a change in plate separation of less than 10^{-9} cm. Nevertheless, it is clear that the optical maser is potentially a source of radiation of extremely small spectral width.

Other gas mixtures in which successful maser action has been reported are neon-oxygen and argon-oxygen, in both of which the population of two levels in the excited oxygen atom has been inverted to give maser oscillation at a wavelength of 8446 Å, in the near infra-red. Visible light, at 6328 Å, has been obtained from a He-Ne maser by the use of confocal reflections with a high reflectivity at this wavelength.[8]

8. Optical Maser Applications

The special features of the optical maser may be summarized as

(*a*) high pulsed power;
(*b*) small beam divergence, and the consequent possibility of focusing the output to a very small spot;
(c) small line width.

The first two properties have recently been demonstrated in a striking way at M.I.T.: a ruby maser was used in conjunction with a telescope mirror to illuminate part of the moon's surface. At a distance of 250 000 miles the maser spot was only 2 miles in diameter, and a detectable amount of light was received back from the moon. Using this technique the moon–earth distance may be measured with higher precision than by radar or conventional optical methods.

The high power of the pulsed maser gives an electric field strength, both in the focused and unfocused beam, sufficiently high to give detectable harmonic generation due to the non-linearity of the polarization of dielectrics through which the beam passes.[9] Thus, by focusing the red 6943 Å radiation from a ruby maser into quartz or KDP, radiation at 3471·5 Å, in the ultra-violet, is produced. This shows, in addition, the possibility of producing, for example, a difference frequency in the submillimetre range by beating together two masers of different frequencies.

The high energy flux in the focused beam, which has already been mentioned in connection with the ruby maser, has been used to produce a high temperature locally near the surface of solids. Small holes may readily be made in metal sheets or even in diamonds, and the production of a temperature of 6000°C has been claimed. The maser has also been used as a substitute for the discharge lamp in the retinal coagulator used for the surgical treatment of detached retinas. Other applications of the intense pulsed light from a solid maser lie in the field of high-speed photography, and its use in a form of infra-red radar has been suggested.

In pure physics the study of the non-linear properties of dielectrics has already been mentioned. The high intensity and narrow spectral width of the maser output make it of value in Raman spectroscopy, where the wavelength of the light is of secondary importance. The development of masers as sources of intense monochromatic radiation would be particularly valuable for spectroscopy in the far infra-red and submillimetre range where no such sources exist at present.

One of the major applications of the optical maser in the future will, no doubt, be in the field of communications. The narrow beam property makes communication over astronomical distances possible, and the small line width allows exploitation of the very large available bandwidth. The highest frequency coherent r.f. generator has an upper limit of the order of 10^{11} c/s, but the optical maser beam operates at over 10^{14} c/s, so that in principle an optical maser beam has a much larger available bandwidth. At the moment this bandwidth cannot be used because no method is known for modulating a light beam at a frequency greater than about 10^{10} c/s. Once higher modulation frequencies have been achieved a superheterodyne technique could be used for channel separation, using fixed frequency optical masers at the receiver as local oscillators. For terrestrial use the beam would have to be piped in order to eliminate atmospheric absorption and scattering. Using the present gas discharge maser, calculations based on reasonable assumptions show that it is feasible to communicate over distances of the order of a million miles with a bandwidth suitable for speech channels. The flash from the ruby maser used for the moon reflection experiment would be visible with the naked eye at distances up to 10^9 miles. These examples illustrate the potentialities of the maser for communication purposes.

9. Summary of Performance

Table 2 gives a summary of the properties of present-day optical masers. In most cases the figures have been rounded off to the nearest order of magnitude.

Table 2

	Pulsed Crystal Maser	C.W. Crystal Maser	Gas Discharge Maser
Wavelengths, microns	See Table 1	0·694, 1·06, 2·61	0·633, 0·845, 1·15
Tuning range ..	0·1%	0·1%	Possibly 0·01%
Power output ..	Up to 10 MW	1–10 MW	10 MW
Pulse length ..	10^{-7}–10^{-3} s	C.W.	C.W.
Beam divergence	1–10 minutes	1–10 minutes	½ minute
Frequency width	10 Mc/s	10 Mc/s	1 kc/s (1 c/s†)

† In absence of vibration, etc.

10. References

1. A. L. Schawlow and C. H. Townes, "Infrared and optical masers", *Phys. Rev.*, **112**, p. 1940, 1958.
2. A. G. Fox and T. Li, "Resonant modes in a maser interferometer", *Bell System Tech. J.*, **40**, p. 453, 1961.
3. G. D. Boyd and J. P. Gordon, "Confocal multimode resonator for millimetre through optical wavelength masers", *Bell System Tech. J.*, **40**, p. 489, 1961.
4. W. Heitler, "The Quantum Theory of Radiation", 3rd ed., p. 177 (Clarendon Press, Oxford, 1954).
5. See, for example, J. K. Roberts and A. R. Miller, "Heat & Thermodynamics", 4th ed., p. 504 (Blackie, Glasgow, 1952).
6. D. F. Nelson and W. S. Boyle, "A continuously-operating ruby optical maser", *Applied Optics*, **1**, p. 181, 1962.
7. A. Javan, W. R. Bennett and D. R. Herriott, "Population inversion and continuous optical maser oscillation", *Phys. Rev. Letters*, **6**, p. 106, 1961.

   D. R. Herriott, "Optical properties of a continuous helium-neon optical maser", *J. Opt. Soc. Amer.*, **52**, p. 31, 1962.
8. J. D. Rigden and A. D. White, "Continuous gas maser operation in the visible", *Proc. Inst. Radio Engrs*, **50**, p. 1697, 1962.
9. P. A. Franken, A. E. Hill, C. W. Peters and G. Weinreich, "Generation of optical harmonics", *Phys. Rev. Letters*, **7**, p. 118, 1961.

   M. Bass, P. A. Franken, A. E. Hill, C. W. Peters and G. Weinreich, "Optical mixing", *Phys. Rev. Letters*, **8**, p. 18, 1962.

   J. A. Giordmaine, "Mixing of light beams in crystals", *Phys. Rev. Letters*, **8**, p. 19, 1962.

Manuscript received by the Institution on 27th July 1962 (Paper No. 767).

The Symposium on Masers and Lasers will be held in London on 2nd January 1963. In addition to the above introductory paper a further seven or eight papers will be presented on various specialized aspects of both these techniques. A list of some of the papers was given in the October Journal *(page 262) and the full programme will be published in the December issue. Registration will be necessary for both members and non-members.*

U.D.C. 621.315.213.24: 621.317.73

Automatic Measurements of the Electrical Characteristics of Telephone Cables

By

G. H. KING, B.Sc.†

AND

R. B. SUMNER ‡

Presented at the Symposium on "Recent Developments in Industrial Electronics" in London on 2nd–4th April 1962.

Summary: The paper first outlines briefly the manual methods used to measure the electrical characteristics of multiple pair telephone cables. Prior to the application of the final covering or sheath, a series of electrical tests are performed on the several conductors to ensure that no serious manufacturing errors are present, and automatic methods by which the measurements required at this stage can be carried out are described. When a trunk cable has been completed, a wide range of tests are carried out to determine the final electrical characteristics and to check the performance of the cable. Methods adopted on distribution cables where the test procedures are not quite so exacting, but more of a "go/no-go" procedure, are outlined.

1. Introduction

It may appear at first sight that telephone cables are simply a compact assembly of many separately insulated wires and that to obtain satisfactory transmission of speech signals it is necessary only to ensure that each wire is continuous and that the insulation between all wires is electrically sound. These are indeed vital features, the checking of which involves large scale measurements, but other important requirements in the construction of circuits in telephone cables are the equalization of the electrical characteristics of the line along its length, and the reduction of mutual interference or cross-talk between adjacent circuits. The former necessitates the measurement of the mutual electrostatic capacitance of each circuit or pair whilst the latter requires the measurement of the capacitance and resistance differences or unbalances between the wires of each circuit.

Thus, a telephone cable must be carefully designed to provide circuits to meet these requirements and it must then be manufactured with a high degree of precision.[1, 2] Electrical testing is necessary to control the quality of the product during manufacture and to measure the characteristics of the completed cables. Because the product is complex and in some respects unique, this testing is rigorous and extensive. A single cable length may contain several hundred or even several thousand wires and a full range of tests on each pair is often required. The problem is further complicated by the fact that many of the properties of the pairs or quads can only be determined after assembly and then it is possible for a cable to be rejected because of a single deviating result.

From the foregoing, it will be seen that the testing of telephone cables presents a problem of some magnitude. The use of manual test methods results in a heavy expenditure of time and effort and impinges adversely on the cost of the final product.

In recent years, much thought has been given to the development of mechanized or automatic test methods. This paper describes some of the test machines which have been produced for this purpose and gives some indication of the success achieved.

2. Test Requirements[3]

The telephone cables considered in this paper may be classified, for the purpose of describing tests and test methods, into two main types: trunk cables and distribution cables. The former, which are of quad construction, provide the comparatively long links between telephone exchanges. The latter are mostly of twin construction, and connect the subscribers to an exchange.

2.1. *Trunk Cables*

The electrical tests required on each completed length of trunk cable are shown in Table 1. The

Table 1

Electrical Tests on Trunk Cables

Characteristic	Measurements
Conductor continuity	Each wire
Conductor resistance	10 pairs
Resistance unbalance	Each pair
Insulation resistance	All wires in groups
Mutual capacitance	Each pair
Capacitance unbalance	
Within quad:	
side/side	Each quad
side/phantom (2 values)	Each quad
side/earth (2 values)	Each quad
Between quads:	Adjacent quads in same layer
pair/pair (4 values)	

† British Telecommunications Research Ltd., Taplow Court, Taplow, near Maidenhead, Berkshire.

‡ British Insulated Callender's Cables Ltd., Prescot, Lancashire.

recording of numerical values for all individual pair resistance unbalances and mutual capacitances and for all individual quad capacitance unbalances is required.

The meaning of the terms used to describe the various capacitance unbalances may be made clearer by a consideration of the resulting network of capacitances between the four wires of a quad which may be shown diagrammatically as in Fig. 1.

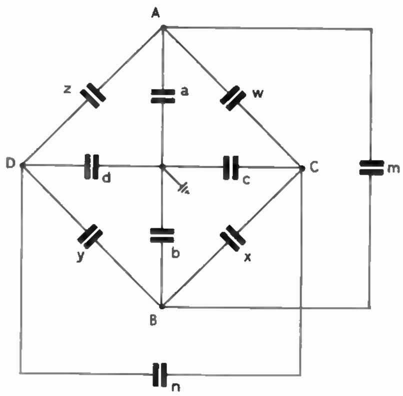


Fig. 1. Quad capacitance diagram.

With wires A and B forming pair 1 and wires C and D forming pair 2, the within-quad unbalances may be expressed as follows:[4, 5, 6]

side/side $= (w-x)-(z-y)$
side (pr. 1)/phantom $= w-x+z-y+\frac{1}{2}(a-b)$
side (pr. 2)/phantom $= w-z+x-y+\frac{1}{2}(c-d)$
side (pr. 1)/earth $= a-b$
side (pr. 2)/earth $= c-d$

The unbalances between pairs in different quads may be treated in the same way as side/side, there being four possible pair/pair combinations.

2.2. *Distribution Cables*

Although the testing of distribution cables is in some ways less onerous, the introduction on an increasing scale of polythene insulated cables into this class and the adoption of new processing techniques have brought new problems of multiple measurement.

The electrical tests required on each completed length of distribution cable are shown in Table 2.

2.3. *Measurements During Manufacture*

Once the outer covering or sheath has been applied to the cable, little can be done to rectify any fault or deviating characteristic subsequently discovered unless the sheath is removed. This is of course a delaying and costly operation and it is, therefore, considered necessary to carry out quite extensive electrical tests prior to the application of the sheath and at what is generally known as the stranded stage of manufacture.

Table 2
Electrical Tests on Distribution Cables

Characteristic	Measurements
Conductor continuity	Each wire
Conductor resistance	10 pairs
Insulation resistance	All wires in groups
Mutual capacitance	All pairs in groups
Capacitance unbalance pair/pair	According to type and construction
Voltage withstand: up to 10 kV d.c. (polythene)	All wires

The tests performed at this stage are shown in Table 3.

Table 3
Electrical Tests at Stranded Stage

Trunk cables	Distribution cables
Conductor continuity	Conductor continuity
Conductor gauge	Short circuit test 50 V d.c.
Resistance unbalance	Voltage withstand (polythene)
Short circuit test 250 V d.c.	
Capacitance unbalance:	
side/side	
pair/pair (2 values)	

3. Manual Test Methods

Manual test sets have been used for many years[7] and from time to time attempts have been made to speed up the testing operations by modifying bridge designs and layouts. Even so the process has remained time-consuming and expensive and always susceptible to the inaccuracies which can result from the human element and from operator fatigue.

Manual sets for dealing with one particular type of test may be portable bridges which are moved to the cable to be tested. With such an arrangement of testing on the factory floor, it is often necessary to employ visual indication of bridge balance. A single operator may be used to connect the cable pairs to the test set, to carry out the tests and to record the results. An alternative arrangement is for one operator to connect whilst another operator tests and records.

It is however more usual for the manual set and the test operator to be in a test room remote from the factory floor. Another operator connects the pairs of wires to be tested in turn into a test clip which is connected by screened and balanced cables to the test set. Communication between the two operators is maintained by telephone or by lights and buzzers.

Using existing manual methods it requires a number of separate testing operations to perform the principal tests. These will now be considered in more detail.

3.1. *Open and Short Circuits*

Tests for conductor continuity and for contacts between conductors, often referred to as "ringing-through" need not be carried out at a stage when certain other measurements such as resistance unbalance, insulation resistance or voltage withstand are to be made. An operator rings through by using a 50 V d.c. detector unit. Each wire is checked for open circuit or continuity and the process is then repeated for the short circuit test.

In an improved method, suitable for paper insulated cables, checks for continuity and contacts are made in one testing operation. The far end of the cable is impregnated in a semi-conducting solution to give a loop resistance within a range beyond the resistance of a contact path. Relays of different sensitivities are then used to operate at the currents in the respective resistance paths.

3.2. *Resistance Unbalance*

This is measured on a bridge with ratio arms made variable by connecting a potentiometer between the equal fixed resistors in each arm. The difference in resistance of the two wires of a pair is generally expressed as a percentage of the loop resistance and the potentiometer dial can be calibrated to read this percentage unbalance directly. Even with portable equipment it is found to be better to use a two operator team.

3.3. *Insulation Resistance*

The insulation resistance may be measured by conventional equipment comprising a standard megohm and galvanometer or by one of the electronic insulation resistance test sets.

The measurements are required to be made with not less than 500 V d.c. after steady electrification for one minute. To test each wire in this way would, of course, be impossible and thus it is customary to connect all the wires of the cable into pre-arranged groupings so that all possible wire-to-wire combinations are covered by a very limited number of measurements. Despite this, the test is still costly because the grouping calls for great dexterity and skill and, on a large cable, the operation can take several hours to complete.

3.4. *Voltage Withstand*

A test requirement for polythene insulated cables is that the dielectric between each wire and all the other wires in the cable shall withstand a flash test at a voltage between 5 and 10 kV d.c. according to the thickness of the insulation. A low energy ionization test set may be used for this purpose, operated remotely from the cable or as a portable instrument. High voltage switching devices with manually operated controls have been developed to provide limited facilities for sequential tests.

As when measuring insulation resistance, the wires of the cable may be grouped in order to reduce the number of tests. If, however, the fault incidence is high, this grouping method can be more time-consuming, because of the need to isolate the faulty wires from the groups. Thus, in certain circumstances it can be quicker, even by manual methods, to test each wire in turn.

3.5. *Mutual Capacitance and Capacitance Unbalance*

The alternating current bridge used to measure these characteristics is often located in a sound-proofed room remote from the cable under test and audible detection of bridge balance by headphones is used.

A special rotary switch enables the test operator to take the eleven readings of mutual capacitance and capacitance unbalance on each quad of a trunk cable. This operator also records these readings on a test sheet, but sometimes this job is done by a third operator. The output of a two-operator team is about 1000–1200 pairs per day and with a three-operator team this may be increased to about 1500 pairs per day.

To measure the mutual capacitance of a distribution cable, the wires of all pairs are connected together to form groups of parallel capacitances. From a few measurements the average pair mutual capacitance is easily derived, but the time spent in preparing the cable for test is considerable.

4. Automatic Measurements on Trunk Cables

4.1. *Electro-mechanical Machines*

An automatic machine, designed primarily for the testing of trunk cables, has been in use for some years. The machine is capable of selecting the measurements required, carrying them out and recording the results on a sheet of paper. Two portable versions of this machine exist, one for performing the stranded stage tests and one for testing the completed cable.

The stranded stage machine can perform in one operation all the trunk cable tests given in Table 3.

Two operators are needed because of the short circuiting requirements at the far end of the cable. The first operator connects the wires of the cable to the set and scrutinizes the printed results. The second operator connects the wires from the other end of the cable to a test jig, which is in turn connected by wires to the automatic machine. Under a command signal

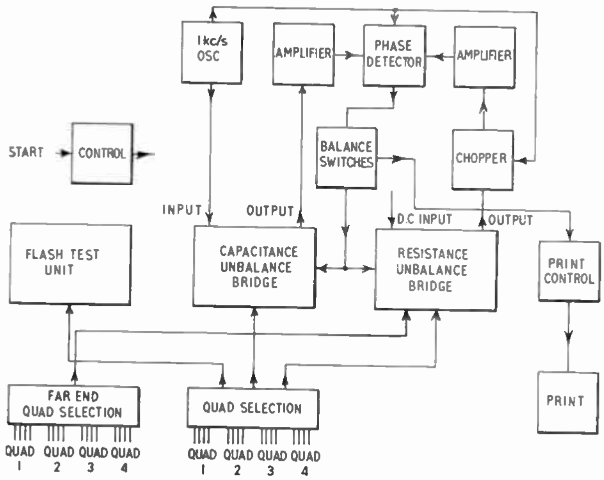


Fig. 2. Stranded stage cable test set. Block schematic.

from the main operator, the machine carries out the tests automatically and records the results. Synchronism between the two operators is ensured by control lamps.

A block diagram of this machine is shown in Fig. 2. The operation is predominantly electro-mechanical and Post Office type relays are used to connect the appropriate test circuits. Thermionic valves are used in the detector and oscillator circuits and the machine is cycled round the test programme by means of uniselectors. The output printer on this machine is of simple design and is an early version of what is known as the "Compur-printer". The printer has on its carriage a bar through which the print signals are directed. By this arrangement the layout of the test results is maintained in a constant pattern. A missed signal will cause the printer to be out of step and the machine will stop. However, the currents operating the electro-magnets have to be passed through this code bar and electrical wear is considerable. Sound preventive maintenance of all relays is necessary to ensure satisfactory operation.

Attempts have been made to dispense with the second operator by placing the far end of the cable into a vessel containing mercury. Signals from the machine cause the mercury level to rise and fall and these provide the short circuit and open circuit conditions at the appropriate time. With this method, however, there is some danger of mercury poisoning affecting the test personnel and alternative safer methods are being sought.

Five operators using manual test methods would be needed to maintain the same test output obtained from one stranded stage automatic machine of this type and furthermore the latter provides additional tests that would not normally be performed manually.

Only one operator is required for the machine testing the completed cable because at this stage measurements such as resistance unbalance which require the short circuiting of the far end are not repeated. The machine measures the mutual capacitance and capacitance unbalance characteristics listed in Table 1 and also has the facility for making a one second insulation resistance test on each quad. This short-period testing of insulation resistance will be dealt with later in more detail.

4.2. *Transistorized Machine*

A portable machine of more advanced design in which the control functions have been fully transistorized has now been produced for the testing of trunk cables at the stranded and final stages. The schematic of the stranded stage machine is similar to that shown in Fig. 2, with the difference that the electro-magnetic control functions are replaced by a transistorized logic using bi-stable Eccles-Jordon or flip-flop circuits as storage devices. The storage and gating

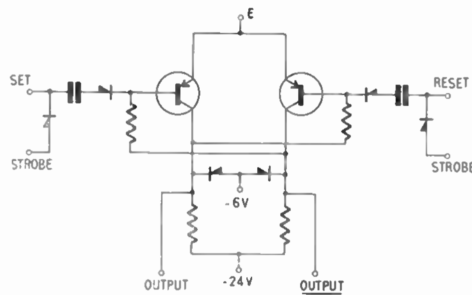


Fig. 3. Toggle-circuit. Simplified circuit diagram.

circuits and the logic methods used are of a standardized design, developed and produced at British Telecommunications Research Ltd., for a wide variety of control and telephone switching applications. The logic is of a step-by-step or strobed version and signal levels are fully clamped throughout. These two factors reduce drastically the possibility of indeterminate signal levels. The circuit diagram of the Eccles-Jordon storage circuit is shown in Fig. 3. Negative logic is used which means that a "0" state is represented by zero or earth volts and "1" state by -6 V. The output printer on this machine is a standard electric typewriter operated by electro-magnets.

All operating experience has so far been gained on the machine for testing the completed cables. A block schematic of this machine is shown in Fig. 4. Again the logical functions are performed by semi-conductor devices in the Eccles-Jordon storage circuits and diode gates. The machine is essentially of the fixed programme type, although individual tests can be omitted from the test sequence by the

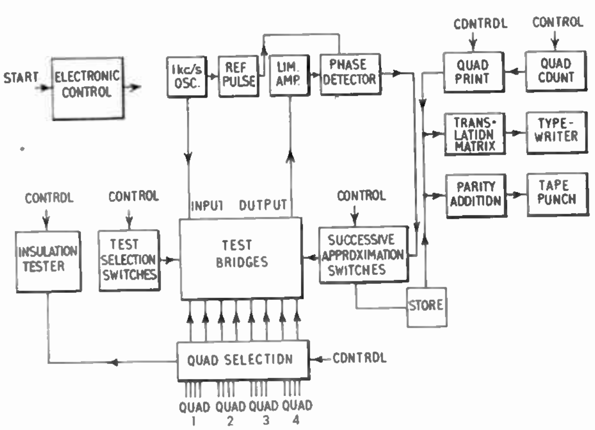


Fig. 4. Final stage cable test set. Block schematic.

operation of the appropriate keys on the control panel. The test results are again printed on a standard electric typewriter with a parallel output on punched paper tapes which enables any subsequent processing of the test results, if so required, to be carried out by a computer.

The final stage machine requires only one operator, who selects the wires to be tested in turn, and connects them to the test clips on the machine. On pressing the appropriate "start keys" the machine carries out the full sequence of mutual capacitance and capacitance unbalance tests and also checks each wire of the quad for insulation resistance. The measurement of insulation resistance after one second depends upon the establishment of some correlation between long and short time leakage current values. If then the insulation resistance of a few wires in the cable are measured after the specified electrification period of one minute, the results of the wires then known to be good may be compared at the one second electrification time with all the other wires in the cable. This can be done on a go/no-go basis and any wires worse than the standard wires may be individually checked.

A single operator on the final stage machine can test, in a standard eight-hour day about 2000 pairs for mutual capacitance, capacitance unbalance and insulation resistance. To achieve this same output by manually testing the a.c. parameters and by wiring and testing for group insulation resistance, five operators would be required.

5. Automatic Measurements on Distribution Cables

The testing of distribution cables is not so exacting, although some tests on individual wires or pairs are required. Test results are not required for subsequent processing and it is necessary only to determine that the measured characteristics conform to the specified limiting values. This enables go/no-go testing methods to be adopted and as there is no need to record test results the output devices and the circuits necessary for operating them can be dispensed with. A series of simple automatic test instruments is thus possible and some that have been developed will now be described.

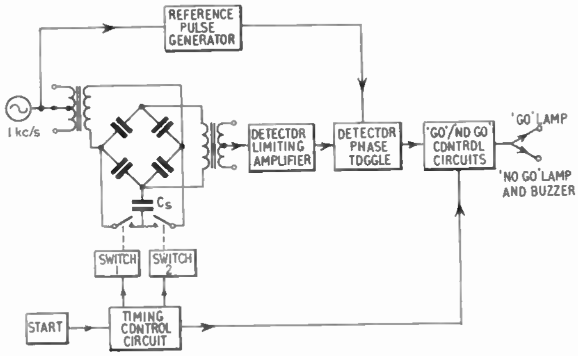


Fig. 5. "Go/no-go" capacitance unbalance test set. Block schematic.

5.1. *Capacitance Unbalance Tests*

A go/no-go type of automatic test set has been produced for the checking of pair/pair capacitance unbalance in twin distribution cables. An outline of the operation of this machine is shown in Fig. 5. The equipment is fully transistorized using gates and storage circuits as previously mentioned. As can be seen, the omission of printing-out devices has considerably reduced the bulk of equipment required, thus enabling the instrument to be mounted on a small self-contained trolley which can be moved easily to the cable under test.

The control panel for this machine is shown in Fig. 6. The operator sets into the control panel the maximum value of capacitance unbalance permitted by the specification or a lower value of unbalance according to the level at which it is desired to monitor

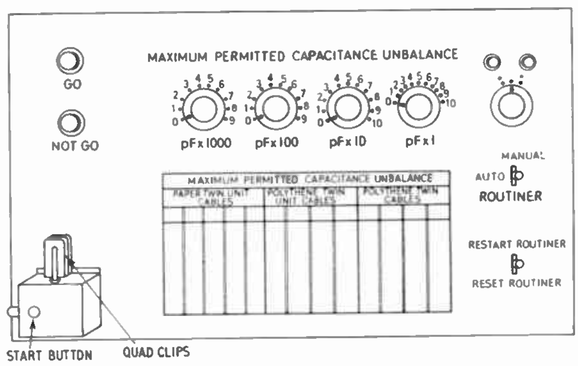


Fig. 6. "Go/no-go" capacitance unbalance test set. Front panel layout.

this characteristic. This capacitance is shown as C_s in Fig. 5. The pre-set value varies according to the length of the cable and is read from the chart provided. The pairs to be tested are clipped to the test set in turn and the start key depressed. If the value of pair/pair is less than that set into the machine, a green light is lit indicating that the test is satisfactory. If the unbalance exceeds the pre-set value a red light is illuminated and an audible warning is given. Furthermore, it is impossible to proceed with the testing sequence until the machine is reset.

The output of a two-operator team manually testing pair/pair unbalance in twin distribution cables can be achieved on this automatic machine by one operator in half the time.

5.2. *Mutual Capacitance and Insulation Resistance Tests*

A trolley similar to the one just described contains mutual capacitance and insulation resistance test sets. For both these tests the wires of the cable are connected together in groups as previously described. In the case of insulation resistance a go/no-go measurement is sufficient and each group is checked with the controls set to the specification limit or higher.

To change the wiring of the cable end to permit a single group mutual measurement, the operator simply snips a few of the connecting wires. The controls of the instrument are set according to the number of pairs in the cable and the length of the cable and a direct reading from digital indicators gives the capacitance of the cable in microfarads/mile.

5.3. *Conductor Resistance Measurement*

A highly mobile test set for automatic conductor resistance measurements has also been developed. The number of series connected wires under test, the cable length and the required temperature correction, are all set into the controls of the instrument. The operator connects the terminal wires to the set, presses the start key and then takes a direct reading from digital indicators of the resistance in ohms/mile.

6. Measurements with Automatic Individual Circuit Selection

All the automatic test methods so far considered involve a repetitive sequence of connect and test performed by the operator. In some instances, testing and connecting operations are performed simultaneously and when the speed of each operation, considered separately, is about the same, then the automatic machine is being fully utilized. When, however, it is necessary to test individually each wire or pair to all other wires or pairs, then this degree of automation is not sufficient. Mention has already been made of the need, when the fault incidence might be high, to test individually each wire of a polythene insulated cable for voltage withstand. This manner of testing for this type of cable may also be considered necessary when the method of stranding produces a random bunch of pairs and does not in any way guarantee a pair sequence. Individual wire tests for insulation resistance and voltage withstand and all pair combinations within a unit for pair/pair unbalance would result in a test program of considerable magnitude.

Equipment has been developed whereby groups or all of the wires in a cable can be connected to automatic test equipment via connecting frames, plugs and sockets and a device, often called a routiner, which selects in turn individual wire or pair combinations.

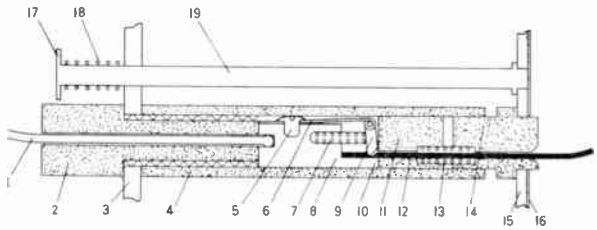


Fig. 7. Pierce and grip device for insulated wire.

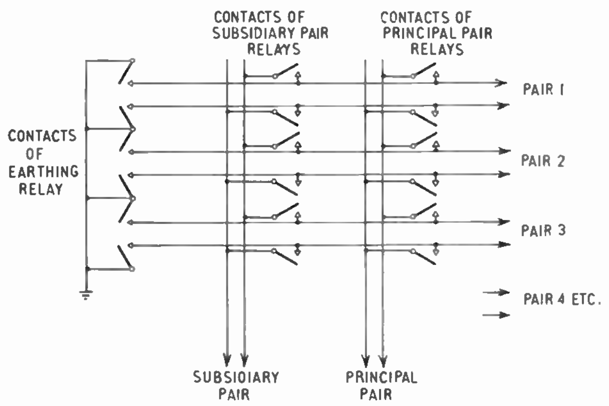
1. Lead wire
2. Securing screw (polythene)
3. Mounting plate
4. Case (polythene)
5. Hinge screw
6. Spring hinge
7. Return spring
8. Body
9. Knife
10. Anvil
11. Stop
12. Insulated wire, pierced and gripped
13. Compensating spring
14. Entry head
15. Cover
16. Cover insulation
17. Release button
18. Retaining spring
19. Release rod.

6.1. *Connection Frame or Block*

Before any tests can be carried out on a cable it is necessary to strip the insulation from the ends of the conductors. This is more difficult when the insulation is polythene and although the operators gain great skill and speed in stripping, it still remains time-consuming and expensive.

A simple device has been produced which enables connection to be made to the wire without first stripping the polythene insulation. The principle is shown in Fig. 7. The insulated wire is pushed through a hole into the device and slides past a knife edge. On pulling the wire back slightly the knife edge bites through the insulation and makes contact with the wire and in so doing it also locks the wire in position within the device. On the completion of all tests the wire is released by the operation of an eject lever.

Most important in the development of the connector has been the need to reduce its size so that it could be grouped in some convenient form and thus



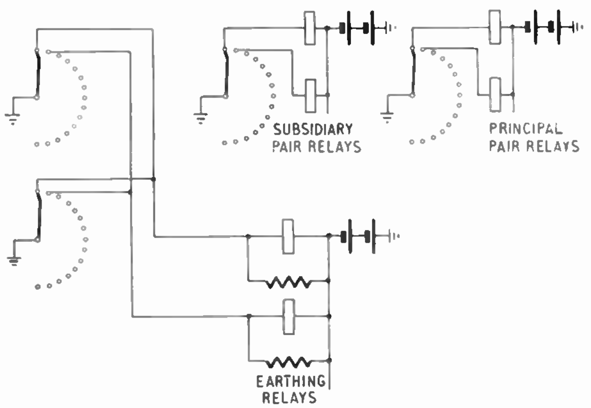


Fig. 8. Low voltage routiner. Outline of operation.

permit the connection without difficulty of many pairs at any one time. With the need to provide a 20 kV voltage-withstand between connectors and a separate guarding arrangement for each connector, this was indeed a problem. The connector now developed is assembled into blocks of 50 which are compact and light and the eject levers are made to release many wires simultaneously. The connection blocks are wired to 25 pair sockets, and to change from one test set to another it is merely necessary to insert the appropriate plugs into the 25 pair sockets.

6.2. *Routiners*

A routiner has been developed for automatically selecting each wire in turn for high voltage and insulation resistance tests. A routiner as shown in Fig. 8 has also been produced for the automatic selection of pair/pair combinations for capacitance unbalance tests. Fifty pairs are connected to the routiner by plug and socket. Each wire is connected to a pair of relay contacts, the appropriate contacts being operated as required to select the test pairs and to connect the remainder to earth. The relays used are of the single make contact sealed reed insert type, the program for opening and closing the reeds being arranged by means of uniselectors. A choice of four different programs is available:

(1) One pair in turn against all other pairs.
(2) One pair against adjacent pairs.
(3) Individual pair mutual.
(4) All pairs in parallel (bunched mutual test).

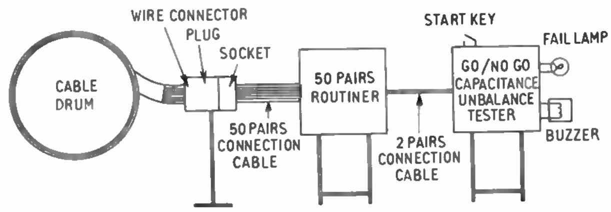


Fig. 9. Fully automatic testing. Typical layout.

7. Complete Automatic Testing

The way in which the various devices already described may be coupled together to produce fully mechanized testing is shown in Fig. 9. Cable connectors make contact with the wires of the cable and these are then routed via a plug and socket to a routiner. The routiner is connected to an automatic measuring set, in this case the capacitance unbalance test set already mentioned. The routiner program is set as required. The test instrument is set to automatic operation, and the start key depressed. Testing will then proceed automatically with all the required combinations being tested in sequence. If any one combination is unsatisfactory the machine will stop and give an audible warning.

8. Conclusions

We have seen that the testing requirements for telephone cables are most extensive and exacting. The performance of these tests by manual methods is expensive and time-consuming. By the use of suitable automatic machines it is possible to reduce the operator time for certain tests by a factor of four or more. Several other advantages also result from the use of such machines. Operator errors are reduced and in some respects even eliminated. Cable drums stand on the test floor for a considerably reduced time and this leads to an improved general operating efficiency of the test unit. The fitting, in certain cases, of mechanized output devices, such as punched paper tape, allows for the subsequent processing of results, required for the satisfactory installation of the cable in the field, to be done by a computer.

More automatic and semi-automatic testing methods are likely to be introduced and these will undoubtedly further reduce the amount of exacting work required from the test operators. Further development in the techniques for rapidly connecting the wires of a cable into frames also seems likely because with the adoption of a complete automatic testing system and a test plan which gives full utilization of automatic machines, connecting remains the only time-consuming manual operation.

Fig. 10. A final stage cable test set in use.

The outputs from automatic test machines may be produced increasingly in coded form and fed to computers in order to evaluate product quality trends and establish a rapid feed back of information to all stages of manufacture, thus reducing fault incidence and improving the quality of the product.

9. Acknowledgments

The permission of the Chief Engineer, British Insulated Callender's Cables Ltd., and the Director of Research, British Telecommunications Research Ltd., to publish this paper is gratefully acknowledged.

10. References

1. F. H. Buckland and R. H. Franklin, "Some notes on the design and manufacture of telephone cables". Institution of Post Office Electrical Engineers, Paper No. 144, 1932.
2. R. Bélus *et al.*, "Les cables de télécommunication", ***Câbles et Transmission***, **15**, No. 4, October 1961.
3. W. T. Palmer and E. H. Jolley, "Telephone cable testing". Institution of Post Office Electrical Engineers, Paper No. 138, November 1931.
4. A. Rosen, "A new network theorem", *J. Instn Elect. Engrs*, **62**, pp. 916–8, 1924.
5. R. Bélus and P. M. Prache, "Théorie élémentaire des circuits à 2 fils". *Annales de P.T.T.*, **20**, pp. 637–76, August 1931.
6. H. C. S. Hayes, R. A. Seymour and P. R. Bray, "Crosstalk". Institution of Post Office Electrical Engineers, Paper No. 109, November 1938.
7. A. Rosen, "Capacitance test set. Apparatus for use on trunk telephone cables", *The Electrician*, **122**, pp. 3–6, 6th January 1939.

Manuscript received by the Institution on 24th January 1962 (Paper No. 768).

POINTS FROM THE DISCUSSION

Mr. F. Gee-Wah: On insulation resistance measurement, the method used one second electrification instead of the standard one minute electrification. Can the authors give us more details of this method and has this method received official approval from the British Post Office or any other telecommunication concern for measuring insulation resistance of paper insulated and plastic insulated telephone cables.

The authors (*in reply*): The method of insulation resistance measurement outlined is at present being evaluated. In carrying out this test the ratio between the values after one minute and those after one second is determined using a few wires of a cable, and this ratio is used to set the machine for tests on the remaining wires. Any marginal results can always be checked over the full test period. The authors understand that a method similar to the one described has been in use by other manufacturers for some years.

Mr. A. B. Aldred (Associate Member): As the capacitance network of multiway cables under test is a complex network, what measurement system is used to eliminate the unwanted impedances?

What is the order of the lowest capacitance which can be measured?

The authors (*in reply*): All the capacitance bridges mentioned can be set to measure a capacitance or capacitance unbalance of 1 pF, although the standard range is 10 pF per step. To prevent erroneous measurement all conductors except those under test are connected to earth. Also where *in situ* type measurements are required a bridge using tightly coupled transformer ratio arms is used. Unwanted capacitances can then be shunted across the ratio arms and eliminated.

U.D.C. 533.27: 534.614: 621.317.772

Gas Analysis by Sonic Phase Measurement

By

A. M. REID

(*Associate Member*) †

Presented at the Symposium on "Recent Developments in Industrial Electronics" in London on 2nd–4th April 1962.

Summary: The relation between the phase difference of the two received signals in a twin tube analyser and the mean molecular weight of the gas mixture in one tube is developed. The advantages of the twin tube system for temperature compensation are shown.

The types of transducer used as both transmitter and receiver are discussed, and the advantages, particularly as regards the transducers, of using a continuous wave rather than a pulse or frequency modulated system to determine the sonic time delay are shown. Previous calculations take no account of the standing waves produced in the tubes by the reflection of energy from the receiving ends. The means of overcoming this effect are discussed. The practical requirements of the tube construction, and the minimizing of pressure, flow and temperature effects are considered.

The phase measurement of the received signals is done electronically. The salient features of the amplification, limiting, phase comparator and output circuits are discussed. The phase relationships, and the calibration checking facilities in the complete instrument are described.

List of Symbols

c	velocity of sound (cm/s)
c_t	velocity of sound in gas contained in tube (cm/s)
c_R	velocity of sound in reference gas (cm/s)
c_S	velocity of sound in sample gas (cm/s)
c_∞	velocity of sound in free space (cm/s)
γ_S	$\dfrac{\text{specific heat at constant pressure}}{\text{specific heat at constant volume}}$ (see reference 1, Appendix 1)
R	gas constant
T	absolute temperature (degrees Kelvin)
M_R	molecular weight of the reference gas
M_S	molecular weight of the gas sample
D	tube diameter (cm)
f	frequency of sound wave (c/s)
l	tube length (cm)
t	time delay
t_D	time delay difference
$d\theta$	phase difference (degrees)
df	frequency difference (c/s)
f_D	maximum frequency deviation
f_m	modulation frequency
r	proportion of gas in mixture (by volume)
R_c	reflection coefficient
ρ	density of medium (g/cm^3)

1. Introduction

When an industry requires the continuous analysis of a gas mixture for plant or process control, a method involving the measurement of one of the following properties could be used:

(1) electro-magnetic energy absorption.

(2) heat absorption or generation.

(3) chemical absorption.

(4) magnetic permeability.

(5) dielectric permittivity.

(6) mass, or density.

Some of these methods are very selective in their detection of gases, enabling a high sensitivity to be obtained in the presence of a large concentration of other gases to which the system is insensitive, e.g. the infra-red gas analyser (group 1). Other methods are less discriminating, having low sensitivity to a wider range of gases. The measurement of mass and density (group 6) includes the sonic gas analyser. Since mass is a property of all gases, and the components of a mixture are unlikely to be of the same molecular weight, a wide field of application is open to the sonic gas analyser. This almost universal sensitivity makes the sonic gas analyser only applicable to binary mixtures, in general. The potential simplicity, intrinsic safety, fast response, and the possibility of

† Sir Howard Grubb, Parsons & Co. Ltd., Walkergate, Newcastle-upon-Tyne, 6.

using the sonic gas analyser with corrosive gases, makes the development of this instrument attractive for general use.

A number of sonic gas analysers[1, 2, 3] have been developed for specific applications. It was essential, in this case, to accommodate a range of gas concentrations from a particular undiluted gas to the minimum detectable concentration of it, as limited by stability considerations.

For general applications it was considered preferable that temperature control should not be necessary, unless the absolute limit of sensitivity, or the nature of the gas or vapour, required it.

2. Sonic Time Delay

The velocity of sound (c_t) in a gas contained in a tube can be said to be:

$$c_t = \sqrt{\frac{\gamma_S RT}{M_S}} \qquad \text{......(1)}$$

when, (a) the sound wave is an axial one, there being no traverse waves if the diameter of the tube fulfils the condition:

$$D < \frac{c_t}{2f} \qquad \text{......(2)}$$

and (b) the wall effects are negligible. For smooth tubes[4] a simplified expression gives

$$c_t = c_\infty \left(1 - \frac{K}{D\sqrt{f}}\right) \qquad \text{......(3)}$$

where K is of magnitude 0·1–1·0 (approximately).

These conditions are easily fulfilled in practice.

The time of travel in a tube of length l is

$$t = l\sqrt{\frac{M_S}{\gamma_S RT}} \qquad \text{......(4)}$$

Since M_S/γ_S is characteristic of the gas mixture, a measurement of this delay is a measure of the concentration of one gas in the mixture.

M_S is the mean molecular weight of the sample gas mixture:

$$M_S = r_1 M_1 + r_2 M_2 \qquad \text{......(5)}$$

γ_S is the ratio of the specific heats of the sample gas mixture[1]:

$$\gamma_S = 1 + \cfrac{1}{\cfrac{r_1}{\gamma_1 - 1} + \cfrac{r_2}{\gamma_2 - 1}} \qquad \text{......(6)}$$

3. Twin Tube Analyser

3.1. *Temperature Compensation*

Although the change in the time of travel which is caused by the entry of the sample gas to the single tube analyser may be small, the temperature effect is large since it affects the absolute velocity.

If the sample gas is in a tube length l_S then the time delay

$$t = \frac{l_S}{c_S} = \frac{1}{\sqrt{T}} \cdot \frac{l_S}{\sqrt{R}} \sqrt{\frac{M_S}{\gamma_S}} \qquad \text{......(7)}$$

If a second, reference tube (fitted with transducers) of length l_R is placed so as to be at the same temperature as the sample tube, and this tube is filled with a reference gas which is not changed in composition during the measurements, the time delay difference t_D can be measured at the receivers of both tubes. Then:

$$t_D = t_R - t_S$$

or

$$t_D = \frac{1}{\sqrt{RT}} \left(l_R \sqrt{\frac{M_R}{\gamma_R}} - l_S \sqrt{\frac{M_S}{\gamma_S}} \right) \qquad \text{......(8)}$$

$$= 0$$

when

$$\frac{l_R}{l_S} = \sqrt{\frac{M_S \gamma_R}{\gamma_S M_R}} \qquad \text{......(9)}$$

If this condition is arranged to occur at zero deflection of the time interval meter, then there should be no zero drift due to gas temperature changes if the two tubes are always at the same temperature. The temperature effect then increases with the time delay. It is an interesting corollary that, with a time interval meter of given sensitivity, reducing this sensitivity does not alter the temperature sensitivity, since the time difference increases to give the same deflection of the meter, and the temperature sensitivity increases proportionately.

Although complete immunity to temperature changes affecting both sample and reference gases is only obtained at one meter reading, it can be seen from eqn. (8) that the temperature effect at finite values of t_D is much reduced. The actual improvement depends on the tube lengths employed and the major component of the sample gas.

It is not necessary that the reference gas be one of the components of the gas mixture. The only requirement is that the reference tube length be adjusted according to eqn. (9). It is often convenient to use air as the reference gas.

Since the thermal capacity of a gas is low compared to that of the analysis tubes, it is usual to bring the gas to the temperature of the tubes. A coil of copper pipe may be used to stabilize the gas temperature before admitting it to the sample tube.

3.2. *Tertiary Mixtures*

The twin tube instrument may be used to analyse tertiary mixtures, if one component can be completely removed without affecting the others. The tertiary mixture passes through one tube, and the binary mixture through the second tube, of equal length. The effects of the two components existing in both tubes cancel, leaving the instrument sensitive to the third. If the tertiary mixture in one tube is now replaced by the predominant one of the two components in the binary mixture, the instrument indicates the concentration of the second component without further modification.

4. Sonic Transducers

4.1. *Acoustic Matching*

The resistance offered to a moving diaphragm in a free gas is slight. The aim of any acoustic matching device is to ensure the movement of the maximum volume of gas at high velocity.

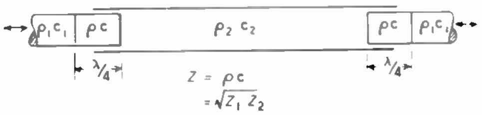


Fig. 1. Theoretical transducer arrangement with optimum acoustic matching.

Applying electrical transmission line theory to the acoustic case (Fig. 1), the application of a quarter-wave length of material of acoustic impedance which is the geometric mean between the gas and the transducers would give perfect matching, for one gas at least. Unfortunately such a material does not exist. Experiments with epoxy resin foam of suitable density to simulate such a material show little promise due to the attenuation caused by the lack of homogeneity of the material.

The dimensions of the transducers and analysis tubes are generally so similar as to make the use of a horn matching device of small advantage. When the transducer is larger than the analysis tube two methods may be used to increase the gas loading:

(a) Solid half-wave velocity transformer.[5]

By enabling an increased amplitude of the transducer to be transferred to the analysis tube (Fig. 2) the transformer presents a greater gas loading than would have been possible by direct connection to the transducer. This transformer is a high Q device, which may not always be desirable.

(b) Gas-filled throat.[6]

This is a convenient non-resonant method of coupling a large diameter transducer to a small analysis tube (Fig. 2).

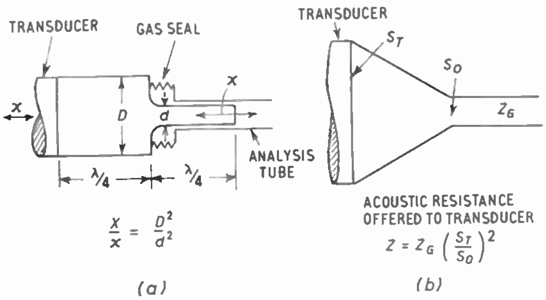
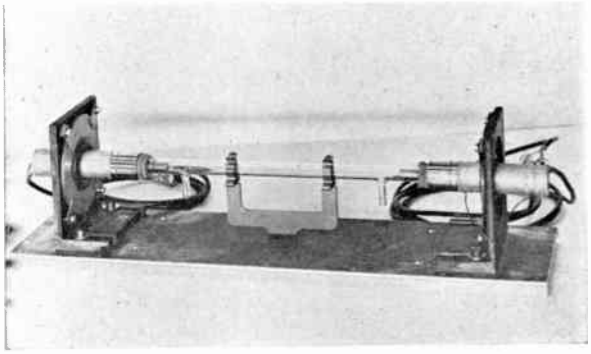


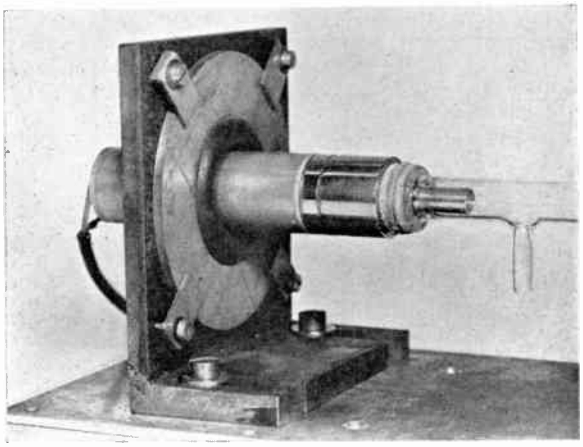
Fig. 2. (*a*) Double quarter-wave velocity transformer. (*b*) Gas-filled throat.

4.2. *Practical Transducers* (Fig. 3)

Miniature moving iron earpiece transducers (used with the integral throat, but without the ear plug) have been found to act as efficient transducers over the range 2500–6000 c/s when coupled directly to tubes of 2·5 cm diameter. They are not satisfactory when high pressures and temperatures are necessary in the analysis tubes.



(*a*) Analysis tube and transducers operating at 32 kc/s in the length mode of barium titanate cylinders.



(*b*) Close-up of transducer showing brass double quarter wave velocity transformer.

Fig. 3.

With the introduction of lead zirconate titanate with the Curie point at 300° C and available with various Q factors, the high pressure/high temperature transducers have become more feasible. Until the introduction of the bimorph transducer, the use of

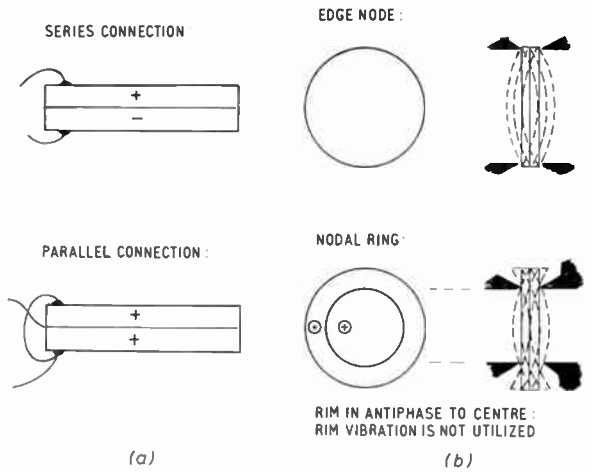


Fig. 4. The bimorph transducer.
(*a*) Alternative connections (poling of piezoelectric material is shown).
(*b*) Two simple modes of operation.

piezoelectric materials, such as lead zirconate titanate, at audio frequencies was impractical because of the size, cost, and the difficulty of energizing plates or cylinders of resonant dimensions.

Figure 4 shows the action of the bimorph transducer and the methods of mounting. The method using the nodal circle (*b*) is preferred, since it has the minimum effect on the resonant frequency, and the connections to the metal electrodes may be made outside the gas enclosure. Sound radiation from the edge of the disc is not passed to the gas, as it would be in anti-phase to that from the centre.

This arrangement gives a vibrator of minimum mass, yet robust enough to withstand high pressure operation. The high electrical capacity enables the transducer to be driven from a low voltage supply. By operating in a resonant mode increased acoustic efficiency results.[7] A compromise must be reached between phase stability and efficiency. The two transmitters are driven from a common oscillator and power amplifier.

4.3. *Time Measurement Systems*

Three methods are considered:

(1) Pulse

The travel time difference is measured, either by digital counting of a standard frequency, or by the measurement of the potential on a capacitor charged at constant current, in the period between the two received pulses.

(2) Frequency modulation

If the sound source is continuously varied in frequency, sound arriving at different times will be of different frequency, and a measurement of this frequency by digital counting, or the integration of electric charge switched at the frequency to be measured, will indicate the time of travel.

(3) Phase

Continuous sound of fixed frequency emitted by two transducers acting in phase will differ in phase

Table 1

Gas mixture (by volume)	c_t cm/s $\times 10^3$	t μs/cm	t_D(μs) 2 tubes $l = 1$ cm	df(c/s) 2 tubes $l = 1$ cm	dθ (deg) 2 tubes $l = 1$ cm
			(1)	(2)	(3)
Air	34·37	29·09	—	—	—
1% Hydrogen in air	34·53	28·96	0·13	0·27	0·059
1% Helium in air	34·54	28·96	0·13	0·29	0·059
1% Methane in air	34·42	29·05	0·04	0·10	0·014
1% Water vapour in air	34·42	29·05	0·04	0·10	0·014
1% Argon in air	34·32	29·17	−0·08	−0·07	0·029
1% Carbon dioxide in air	34·26	29·19	−0·10	−0·18	−0·036
1% Sulphur dioxide in air	34·14	29·30	−0·21	−0·38	−0·076
1% Chlorine in air	34·11	29·32	−0·23	−0·43	−0·083
1% Ethyl ether in air	34·87	29·53	−0·44	−0·72	−0·158
1% Chloroform in air	33·74	29·63	−0·54	−1·09	−0·194
1% Bromine in air	33·60	29·76	−0·67	−1·34	−0·241
1% Air in hydrogen	120·2	8·32	−0·47	−0·88	−0·169
Hydrogen	128·6	7·85	—	—	—

(1) is the time difference between reference and sample tubes.

(2) is the frequency difference between reference and sample tubes for frequency modulated sound.

$$\mathrm{d}f = \frac{2 f_D f_m l}{c_S - c_R} \quad \text{(sinusoidal modulation)}$$

$f_D = 10$ kc/s

$f_m = 10$ c/s

(3) is the phase difference between reference and sample tubes for continuous sound at 1 kc/s.

$\mathrm{d}\theta = 360\, f\, t_D$ deg

at the two receivers to an extent dependent on the difference in travel time.

From eqn. (8):

$$d\theta = \frac{360f}{\sqrt{RT}}\left[l_R\sqrt{\frac{M_R}{\gamma_R}} - l_S\sqrt{\frac{M_S}{\gamma_S}}\right] \text{ degrees} \quad \ldots(10)$$

Provided the phase excursions are limited to less than 180 deg they may be measured by a simple phase measuring circuit which will read accurately at a maximum full scale sensitivity of 20 deg at audio frequencies. For phase excursions greater than 180 deg an integrating phasemeter is necessary. If mechanical integration is used, the limited speed of response puts severe limitations on the rate of sample flow.

Table 1 compares the quantities to be measured in the pulse, frequency modulation, and phase systems for some common gases. Although the gas concentration/time delay relationship is not strictly a linear one, for purposes of comparison it may be assumed to be so.

It will be seen that both the pulse and frequency modulation require high fundamental frequencies, and wide bandwidths to obtain the same sensitivity that can be achieved by a phase measurement at a single audio frequency. The efficient transducers available, the low gas attenuation at an audio frequency, and the simplification of the ancillary equipment have made the phase measurement technique preferable for gas concentration measurement.

5. Standing Waves

The reflection coefficient at the junction of two media of characteristic acoustic impedances

$$\rho_1 c_1 \text{ and } \rho_2 c_2 \quad (\text{where } \rho_2 c_2 < \rho_1 c_1)$$

$$R_c = \left(\frac{\rho_1 c_1 - \rho_2 c_2}{\rho_1 c_1 + \rho_2 c_2}\right)^2$$

Since the acoustic impedances of solid and gas differ by a factor of 10^5 (approximately), almost complete reflection occurs at each transducer. The use of an absorbing material or labyrinth at the receiving end is not practical because the receiving transducer must intercept all the wavefront under the adverse matching conditions which it works. The result is that a standing wave can build up, as demonstrated in the classical experiment to determine the velocity of sound in gases by Kundt.[8]

Experiments have shown the nature of the standing wave effect. From Fig. 5 the effect can be seen to be:

(a) Curving of the phase/gas concentration characteristic with cusps at every half cycle of phase change (180 deg).

(b) The slope of the line joining the cusps represents the sensitivity without standing wave effects.

(c) The number of cusps corresponds to the loss or gain of half wavelengths in the standing wave pattern.

In a multi-purpose gas analyser the occurrence of the cusps in the calibration curve is highly undesirable, and the elimination of standing waves is necessary. Temperature compensation with two tubes cannot be effective. The temperature sensitivity will vary in the same manner as the gas sensitivity, and cannot therefore be matched in two tubes.

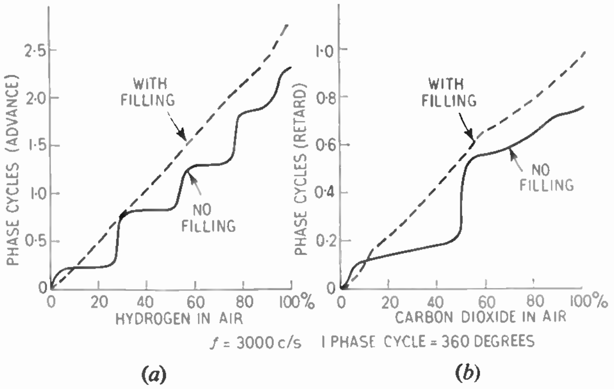


Fig. 5. Phase/gas concentration graphs showing the standing wave effect:

(a) For a gas lighter than air (hydrogen).
(b) For a gas heavier than air (carbon dioxide).

The method of removing the standing waves (Brit. Pat. No. 788,801) is to pack the analysis and reference tubes with suitable materials which ensure devious paths for both travelling and reflected waves, and so prevent their augmentation. By suitable choice of the filling a maximum/minimum transducer output voltage throughout the air to hydrogen change of 1·75/1 may be readily attained. The packing increases the effective path length in the gas causing a divergence of the packed calibration curve from the cusps of the unpacked calibration (Fig. 5).

The overall attenuation, under such conditions, at 10 kc/s using ceramic bimorph transducers, is less than 60 dB for a tube length of 32 cm.

The choice of packing material is governed by the gases and temperatures to be used, since adsorption on some materials can cause an objectionable time lag in response.

6. Tube Construction

To avoid unnecessary reflections the tube must present an unstepped bore to the transducers.

Since various lengths of sample and reference tube are required, plain drawn tube cut squarely to length is used in conjunction with two head units. These house the transducer elements, the electrical connections to them, and the gas connections to the tube. The head assembly used with moving iron transducers can be seen in Fig. 10.

When the bimorph transducer is used, knife edges must be provided coincident with the nodal circle on the transducer disc. One edge provides a gas seal. By suitable choice of materials such a head may be used to measure corrosive gas or vapour concentrations at high temperatures. When high pressures are used, gas or air may be fed to the back of the bimorph disc to balance the pressure on it.

7. Phase Measurement

7.1. *Basic Method*

The method adopted for phase measurement in the range between 0 to 160 and 0–20 deg is to use a bistable circuit driven alternately by unidirectional pulses derived from the squared and differentiated sine wave from each sonic tube receiver (see Fig. 6).

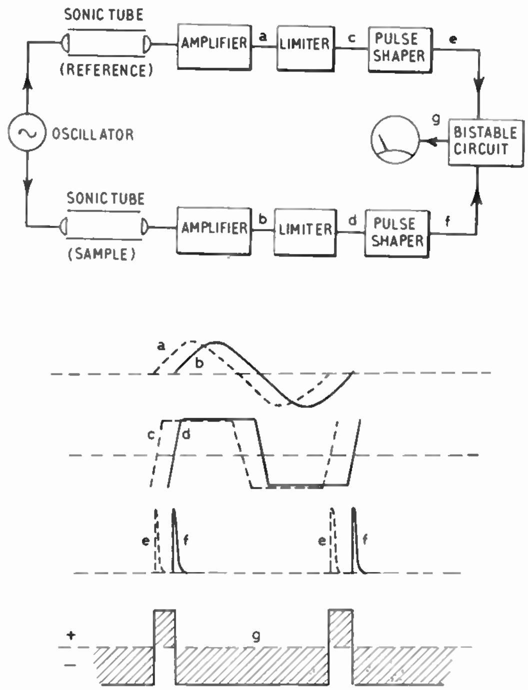


Fig. 6. Basic phase meter—block diagram and waveforms.

7.2. *Bistable Circuit*

A meter, or d.c. amplifier and meter, connected across the bistable circuit measures the mean d.c. component of the resultant square wave. Zero d.c. corresponds to 180 deg phase difference between the two sine wave inputs. The bistable circuit is then dwelling for equal times in each state, and the average d.c. to the meter is zero. A phase lag or lead of one signal then causes an unbalance of the dwell periods, and d.c. flows through the meter in proportion to the phase lag or lead. A reversing switch enables a side zero meter to be used for high or low molecular weight gases at will, although a centre zero meter can be used if this is desirable.

In use, the phase meter reading is compared with a calibration curve obtained by introducing metered gas mixtures. An electrical phase shift is used to enable the sensitivity of the phase meter to be reset to calibration conditions.

The unidirectional input pulses are formed by differentiating (CR) the limiter output, and removing the unwanted pulse with diodes D1 and D4 (see Fig. 7). Triggering occurs near the start of the current pulse so that the sensitivity to variations in the rise time of the pulse is small. To ensure that the output current is independent of the input pulse amplitude and width, transistors VT1 and VT4 are inoperative once they have received their respective pulses. Transistors VT1 and VT4 derive their collector potentials from the collectors of VT2 and VT3 respectively whose potentials are at a minimum after triggering by their respective pulses.

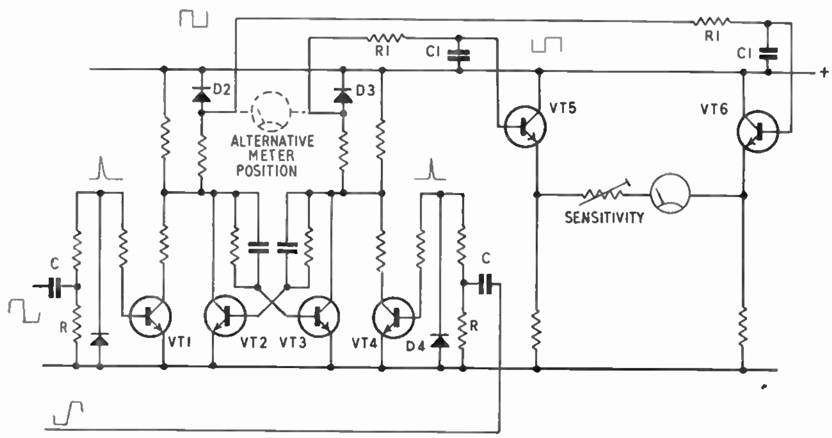


Fig. 7. Bistable phase measuring circuit and output stage.

As a precaution against variations of the peak excursions of the collector potentials of transistors VT2 and VT3, Zener diodes D2 and D3 limit the excursions of the output waveform to stable limits.

Although the square wave output may be fed directly to a moving coil meter, it is desirable, when the meter may be remote from the instrument, to feed only the d.c. component to it. A simple filter (C1,R1) followed by a d.c. amplifier (VT5, VT6), driven symmetrically, provides a maximum output of 1 mA to a load of 1000 ohms.

7.3. *Amplification and Limiting*

Figure 6 shows that the point of reference on each sine wave input should correspond to a negligible increment about the zero axis. This is accomplished by linear amplification followed by symmetrical limiting, before the differentiating circuit.

Variations of 40 dB in the input signal level must be accommodated without distortion in the amplifier. This is necessary to allow for various tube lengths and gas attenuation in the many varied applications of the analyser.

Symmetrical voltage limiting may be effected by using the forward current characteristic of a silicon diode as in Fig. 8 (*a*). This gives a lower peak to peak limited amplitude (2·6 volts) than could be obtained with currently available Zener diodes. The output is not a flat-topped waveform, but the limiting action obtained enables subsequent amplifiers to accept the waveform without shift of the zero axis.

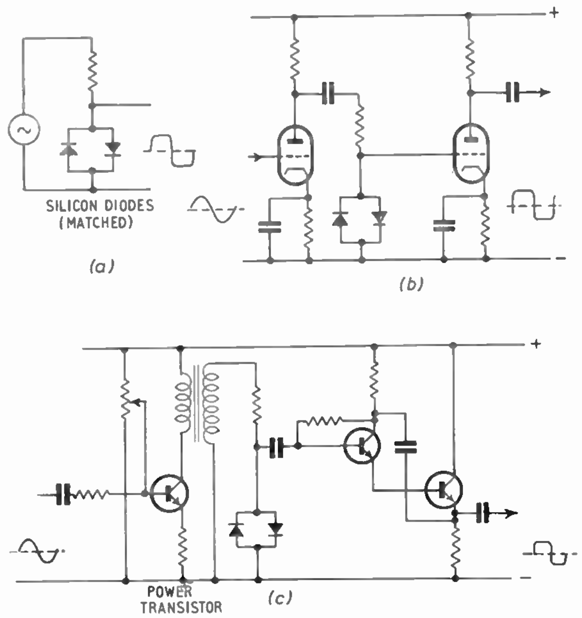


Fig. 8. Symmetrical voltage limiter.

(*a*) Basic circuit. (*b*) Applied to a valve amplifier.
(*c*) Applied to a transistor amplifier.

Since this limiter is essentially a voltage device, it is readily applicable to valve circuits (Fig. 8 (*b*)). For transistor amplifiers a low capacitance constant current diode[9] would be equally readily applicable, but these are not yet available. To utilize the voltage limiting diode, a small power amplifier driving a step-up transformer provides the high voltage (150 V r.m.s.) necessary for the limiter (Fig. 8 (*c*)). The limited output is then fed to an emitter follower stage which offers a high impedance to the limiter (1–2 megohms). Two limiters are used in cascade.

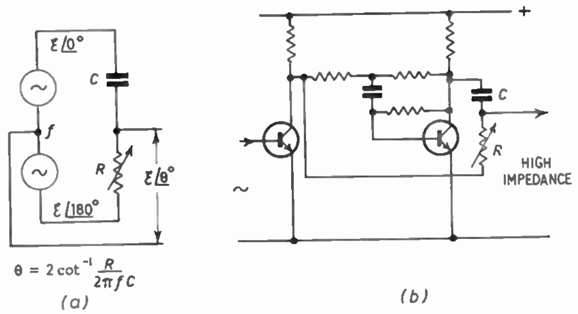


Fig. 9.

(*a*) Basic phase shifter circuit. (*b*) Phase shifter using transistors.

7.4. *Phasing*

The phasing facilities in the sonic gas analyser must provide the following facilities:

(a) Zero setting (for any tube or gas combination).

(b) Calibration.

(c) Range expansion.

The sine wave input to the limiters should be exactly in anti-phase at zero meter deflection. To set the zero for any input phase condition the phase shifter must be able to produce 180 deg phase lag or lead. The circuit as shown in Fig. 9 does not reach 180 deg, since R is never infinite, so a similar circuit is used in each channel. One is normally fixed at 90 deg phase shift, but this can be altered by changing a fixed resistor if necessary for setting up. Two other resistors may be switched to give fixed calibration shifts in both the high and low molecular weight senses.

Range expansion may be introduced by a further variable resistor in the second phase shifter. This may be used to reduce the deflection due to increasing gas concentration, and, if suitably calibrated, can greatly extend the range of the instrument.

The variable resistors used in the phase controls should be of a cotangent grading for smooth operation, although a logarithmic grading can be used. A reduction drive to the resistor is necessary.

7.5. *Transmitter Drive*

Driving power for the transmitting elements is supplied to them in parallel from a single transformer coupled power amplifier. This is fed, via a buffer amplifier, from a two-stage amplitude-stabilized RC oscillator. When very high phase sensitivity is called for, a crystal oscillator may be employed, although matching of the transducer and crystal resonant frequencies then becomes essential.

8. Conclusions

The use of directly air loaded transducers with packed tubes has produced a versatile analysis unit that can be readily adjusted for any application. The phase measurement technique can be used in conjunction with this analysis unit. The phase measurement circuit is simple and symmetrical, and can be adapted to transistor operation. The direct current output of the phase meter may be used to work meters, recorders and process controllers without modification.

The sonic gas analyser for normal temperatures and low pressures (Fig. 10) has been successfully employed to measure gas concentration in air from pure gas to the following minimum levels (at full scale deflection):

Hydrogen	4·5% (by volume)
Carbon dioxide	7·0% ,, ,,
Chloroform	1·1% ,, ,,
Methane	15·0% ,, ,,

Future development is aimed at reducing these minimum levels, operating at high temperatures and pressures, and with corrosive gases and vapours.

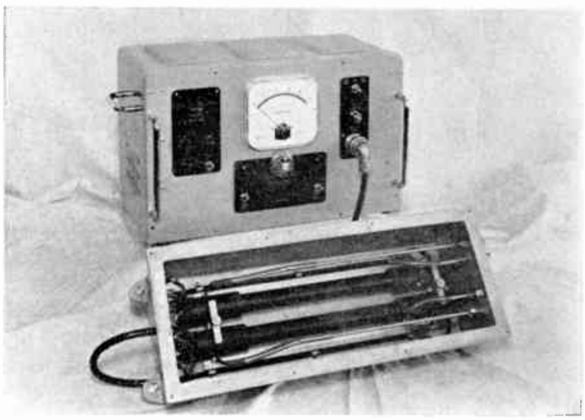


Fig. 10. The complete gas analyser for use at low temperatures and pressures.

9. Acknowledgments

The author gratefully acknowledges the permission of the Directors of Sir Howard Grubb, Parsons & Co. Ltd., to publish this paper, also the assistance of his colleagues, and in particular that of the late D. Mounfield. Thanks are due too to the Brush Crystal Co. Ltd. and Mr. R. F. J. Orwell for the design of bimorph transducers.

10. References

1. E. W. Pulsford, "The analysis of gas mixtures by a sonic method", *J. Brit.I.R.E.*, **15**, pp. 117–28, February 1955.
2. L. Molyneux, "Acoustic gas analyser", *J. Sci. Instrum.*, **36**, pp. 118–20, March 1959.
3. M. Kniazuk and F. R. Prediger, "Ultrasonic gas analyser", *Instruments & Automation*, **28**, pp. 1916–7, November 1955.
4. A. L. Foley, "Velocity of Sound", International Critical Tables, p. 466.
5. E. A. Neppiras, "Design of Ultrasonic Machine Tools", Institution of Mechanical Engineers, Session IV, Paper 6, March 1958.
6. F. E. Terman, "Radio Engineering", p. 879, 3rd Edition (McGraw-Hill, New York, 1947).
7. H. W. St. Clair, "St. Clair oscillator", *Rev. Sci. Instrum.*, **12**, p. 250, 1941.
8. C. J. Smith, "Intermediate Physics", p. 559, 2nd Edition (Arnold, London, 1944).
9. T. K. Hemingway, "Applications of the constant current diode", *Electronics*, **34**, No. 42, pp. 60–3, 20th October 1961.

Manuscript received by the Institution on 23rd February 1962 (Paper No. 769).

POINTS FROM THE DISCUSSION

Mr. M. C. Kearton (*Associate Member*): The system described eliminates temperature problems, but in the apparatus shown, Doppler effect due to gas flow may cause some concern. Has the author had problems of this type?

The Author (*in reply*): Firstly, it would be advisable to note that the two tube analyser reduces the temperature problems, rather than eliminates them. The improvement factor depends on the tube length and gas sensitivity, but is generally of the order of 50–200. Precautions must still be taken to ensure that the reference and sample gases are at the same temperature.

The Doppler effect due to the gas flow has not caused any problems with present instruments, but at higher gas sensitivities, and high flow rates in smaller diameter tubes, the effect can become apparent. The simple expedient adopted by Kniazuk and Prediger[3] was to feed the gas sample to the centre of the tube's length, and to allow it to travel in opposite directions to the open ends. Any Doppler effect will then cancel when the velocity is measured from end to end of the tube.

U.D.C. 621.385.62/3

Noise in Reflex Klystrons and Backward-Wave Oscillators

By

B. G. BOSCH, Dipl. Ing., Ph.D.,†

AND

W. A. GAMBLING, B.Sc., Ph.D.
(Associate Member)‡

Summary: The paper describes a.m. and f.m. direct-detection and superheterodyne noise spectrum measurements on various types of backward-wave oscillator and reflex klystron. A procedure has been developed by which the individual a.m., f.m. and background noise components can be separated. The results show that the predominant component at low sideband frequencies is f.m. noise and at high sideband frequencies is background noise. The a.m. noise is not a significant component. The presence of positive ions in the beam of a backward-wave oscillator causes noise peaks at frequencies near 1 Mc/s and an increase of modulation noise over a wide frequency range. Positive ions can be relatively easily removed by transverse ion drainage. Large noise peaks may also be generated at frequencies near those of spurious oscillations. The results of measurements of carrier frequency deviation, and of correlation between a.m. and f.m. noise, and between modulation and beam noise are also given.

1. Introduction

In the past few years great strides have been made in the development of new types of low-noise microwave amplifiers and oscillators. Furthermore substantial improvements have also taken place in the more conventional travelling-wave tube and klystron. As part of an investigation into the noise properties of travelling-wave-type devices noise measurements have been carried out at X-band on three types of backward-wave oscillator together with comparative measurements on reflex klystrons.

Various methods of microwave noise measurement have been described and analysed previously[1] and it is sufficient here to summarize the main points briefly. The noise output of microwave oscillators (and non-linear amplifiers) consists of two main components which may be termed modulation noise and background noise. Background noise is generated at microwave frequencies whereas modulation noise is generated at low frequencies and appears as sidebands of the carrier as a result of a mixing process. Modulation noise can be separated into amplitude-modulation (a.m.) and frequency-modulation (f.m.) components, and these may be correlated.

There are three distinct types of noise measurement which the authors refer to as the a.m. direct-detection, f.m. direct-detection, and superheterodyne methods.[1, 12, 13] The a.m. direct-detection method indicates[1] the noise power

$$N_{d(am)} = K\left[\overline{(2A)^2}+\overline{2B^2}\right] = 4N_{am}+2N_b \qquad \ldots\ldots(1)$$

where A and B are the instantaneous values of the a.m. and background fluctuations, respectively, at a particular sideband frequency

and N_{am}, N_b are the corresponding a.m. and background noise powers at the same upper or lower sideband frequency.

Correspondingly the f.m. direct-detection system measures a noise power

$$N_{d(fm)} = K\left[\overline{(2F)^2}+\overline{2B^2}\right] = 4N_{fm}+2N_b \qquad \ldots\ldots(2)$$

where F is the instantaneous value of the f.m. fluctuation

and N_{fm} is the f.m. noise power.

Finally, the noise spectrum can be investigated with the superheterodyne system. If it is assumed that the amplitude and frequency modulations take place at the same time instant so that the a.m. and f.m. components are in phase quadrature, then one obtains at the upper, as well as at the lower, sideband frequency the power

$$N_s = K(\overline{A^2}+\overline{F^2}+\overline{B^2}) = N_{am}+N_{fm}+N_b \qquad \ldots\ldots(3)$$

Combining equations (1) and (2), one obtains

$$N_{d(am)}+N_{d(fm)} = 4N_s \qquad \ldots\ldots(4)$$

Equation (4) shows that eqns. (1), (2) and (3) are not independent and that the three noise components cannot be separated by making the three types of measurement.

The paper describes a.m. direct-detection, f.m. direct-detection and superheterodyne measurements, together with measurements of carrier frequency deviation and correlation between a.m. and f.m. noise and between modulation and beam noise. The

† Formerly Department of Electronics, University of Southampton; now with Telefunken GmbH, Ulm/Donau, Germany.

‡ Department of Electronics, University of Southampton.

effect of transverse drainage of positive ions from the electron beam of a backward-wave oscillator has been investigated, and a large increase in noise near the frequency of the first higher-mode spurious oscillation has been observed. Superheterodyne measurements under pre-oscillation conditions have enabled a separation of the a.m., f.m. and background noise components to be achieved, it is thought for the first time.

There is not a great deal of published work on the measurement of low-power microwave oscillator noise. Kuper and Waltz[2] quote the a.m. plus background noise of a reflex klystron at three, much used, intermediate frequencies. Their measuring method suffers from the disadvantage that crystal mixer noise and conversion loss have to be known. Dalman and Ortiz[3] describe a method of measuring separately the upper and lower sidebands of klystron local-oscillator noise by suppressing one sideband at a time with a suitable filter. However, they do not seem to be aware of the fact that inaccuracies occur at lower spectral frequencies due to f.m. noise. In a further paper Dalman and Rhoads[4] quote values of reflex-klystron noise for sideband frequencies from about 5 Mc/s upwards. A superheterodyne measuring system was employed and carrier suppression was achieved in a rather crude filter of wide bandwidth, which even at frequency separations of 50 Mc/s from the carrier still had an attenuation of about 15 dB. The method, therefore, can only be used for rough measurements. The authors assume that this system measures local-oscillator noise contributions whereas, in fact, their results include the f.m. noise component which predominates at lower frequencies and which does not contribute to local-oscillator noise. Müller[5] also investigated reflex klystron noise and gives measured frequency deviation spectra as well as average a.m. (including background) noise in the band 300–5000 c/s. Gottschalk[6] measured the carrier frequency deviation of a c.w. magnetron and also correlation between a.m. and f.m. noise.

Krulee[7] investigated O-type backward-wave oscillators and states that the a.m. spectrum seems to be more than 100 dB below the carrier power and is practically unmeasurable. That his measuring equipment should be limited to such a level indicates that the sensitivity was inadequate. Noise measurements on an X-band backward-wave oscillator described by Cicchetti and Munushian[8] give average values of a.m. plus background noise over sideband frequencies up to 12 Mc/s, and of carrier frequency deviation due to noise in the same range. A further study of backward-wave oscillator noise has been made by Smith[9, 10] who measured a.m. plus background noise at spectral frequencies up to 10 Mc/s, observing discrete oscillations caused by positive ions.

Some of the results described here were presented in a summarized form in a short paper[11] read at the Third International Congress on Microwave Tubes in Munich in 1960.

2. Oscillators Investigated

Noise measurements have been carried out at X-band on three types of backward-wave oscillator and two reflex klystrons, details of which are given below. The three backward-wave oscillators represent examples of each of the three possible ways of focusing the electron beam.

2.1. *O-Type Backward-Wave Oscillators*

2.1.1. CO 43 (Services equivalent: CVX 2392)

This tube is a product of the Compagnie Générale de Télégraphie S.F., and covers the frequency range 7 to 11·5 kMc/s approximately. It contains a flat, interdigital, slow-wave structure and has two parallel strip-line electron beams originating in a tetrode gun. Focusing is achieved by cylindrical permanent magnets. The accelerating voltage for the electron beam depends on the frequency and is between 300 V and 1750 V. The output power varies between 35 and 200 mW.

2.1.2. BA 9–20

The BA 9–20 is a Mullard backward-wave oscillator the internal construction of which is rather similar to the CO 43 in that a flat interdigital structure and two strip beams are employed. The beams are, however, focused by an electromagnet. The required beam acceleration voltage is in the range 250 to 1400 V, the output power varies between 30 and 160 mW, and the frequency range is again from 7 to about 11·5 kMc/s.

The BA 9–20 oscillator investigated was a particular specimen modified at the Mullard Research Laboratories by the fitting of ion-drainage electrodes. For this purpose the sidewalls enclosing the slow-wave structure were removed and replaced by flat rectangular grids spaced from the sides of the structure by ceramic pillars. By making the grids negative with respect to the slow-wave structure, a transverse drainage of positive ions from the electron beams was achieved.

2.1.3. "Ophitron"

The "Ophitron" is a G.E.C. backward-wave oscillator employing electrostatic focusing. An experimental model was used in the measurements described here. A "ladder-line" periodic structure, together with two flat focusing plates, are used to set up the periodic electrostatic field which forms the electron beam into an undulating strip. The crests of the undulating beam are situated in regions of maximum field strength of the wave on the structure, so that a strong coupling

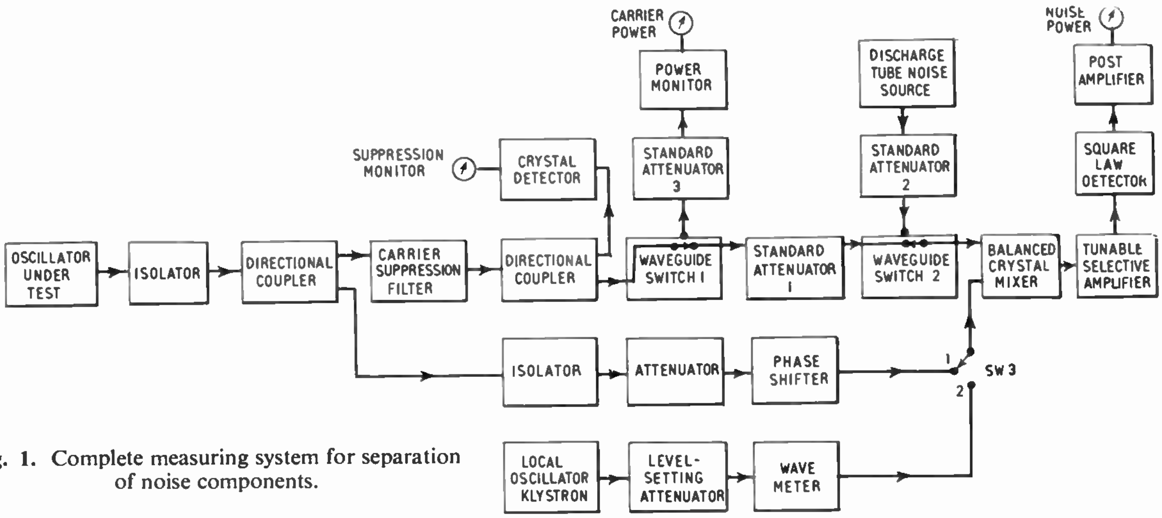


Fig. 1. Complete measuring system for separation of noise components.

exists between the wave and the beam. An advantage of this type of focusing is that positive ion drainage takes place from the beam to the negative focusing plate.

An accelerating voltage between 400 and 1500 V, an output power between 10 and 30 mW, and a frequency range from 7 to 12 kMc/s are typical for this tube.

2.2. *Reflex Klystrons*

2.2.1. CV 323

The CV 323 reflex klystron is a high-voltage, low-noise tube designed to withstand severe conditions of continuous tuning. It requires a resonator voltage of 1500 to 1600 V, a reflector voltage between −200 and −550 V, and delivers a microwave power of 80 to 180 mW.

2.2.2. V 58

The V 58 reflex klystron is manufactured by Varian Associates Inc. A relatively high power output of 300 to 600 mW is combined with a low resonator voltage of 500 V.

3. Measuring Techniques

The measuring techniques are similar to those which have been described previously.[1] However for the noise spectrum measurements the composite circuit shown in Fig. 1 was used. Part of the output from the oscillator under test is passed through an isolator, to a carrier suppression filter[13] (bandwidth < 1 Mc/s) and the noise sidebands remaining enter one arm of a balanced hybrid mixer. To obtain direct detection part of the test oscillator output is fed, via a phase shifter, to the other port of the balanced mixer in such a way that the accompanying noise is suppressed (by about 50 dB), thus providing the necessary "clean" carrier for the crystal drive. The phase of this re-introduced carrier is adjusted by means of the phase shifter so that it differs by $\pm n\pi$ from that of the residual, suppressed carrier component associated with the noise sidebands. For a.m. spectrum measurements $n = 0, 1, 2 \ldots$ and for f.m. spectrum measurements[1] $n = \frac{1}{2}, \frac{3}{2}, \frac{5}{2} \ldots$.

Superheterodyne detection is obtained by changing switch S3 to position 2 so that the mixer crystal drive is obtained from a local-oscillator klystron.

The oscillators were investigated under normal working conditions with the isolator providing a matched load. Particular care was taken to ensure a minimum of deleterious effects due to extraneous influence such as radio-frequency interference, microphony, and imperfect power supplies. For this reason all measurements were carried out inside an electrically screened room with the test oscillator resting on shock-absorbing material and with a flexible waveguide output coupling. Imperfect power supplies present a serious problem as supply voltage ripple may greatly contribute to oscillator noise. Attention must also be given to the long-term stability of the power supplies, especially when the measuring system contains a carrier suppression filter or a microwave discriminator. In the equipment used, the delay-line voltage for the backward-wave oscillators was stabilized to have a ripple of less than 1 part in 2×10^6, and a long-term stability of 1 part in 2×10^3 over a period of five hours.

4. Measured Noise Spectra

4.1. *Magnetically-focused Backward-Wave Oscillator*

Figures 2 and 3 are typical curves for the noise spectra of the CO 43 and BA 9–20 oscillators showing a.m. plus background, f.m. plus background, and the

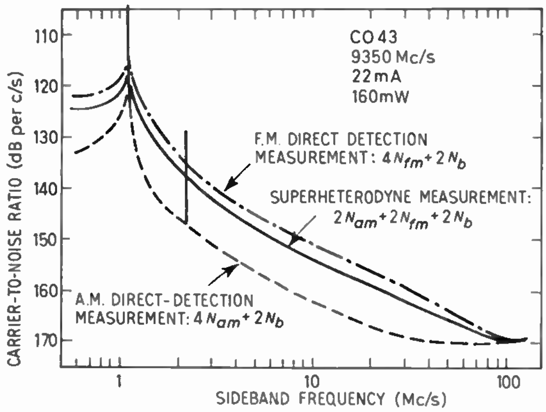


Fig. 2. Measured noise spectra of CO 43 backward-wave oscillator.

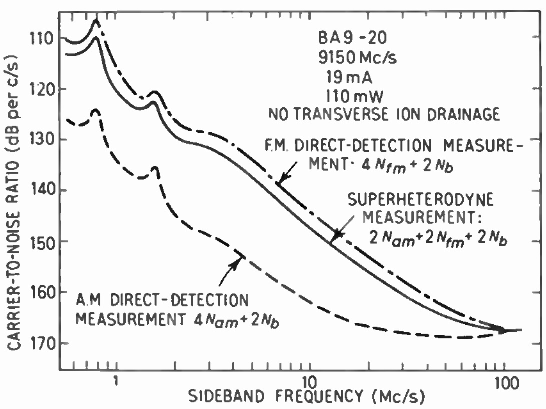


Fig. 3. Measured noise spectra of BA 9–20 backward-wave oscillator.

total noise. The total-noise curve indicates the sum of the upper and lower sideband values which were measured with the superheterodyne system.

The spectra of the two oscillators are of similar form. In the sideband range from a few hundred kilocycles per second to a few megacycles per second, noise peaks and sometimes discrete oscillations are superimposed on uniformly decreasing spectra. At higher frequencies the slope becomes less marked, and finally, for increasing frequencies near 100 Mc/s the noise level begins to increase again.

The shape of the noise peaks is very sensitive to changes in electron-beam current and magnetic focusing field, and by altering or distorting the magnetic field, a peak can be made to break into a discrete oscillation or, on the other hand, an oscillation can be replaced by a peak of random noise. The oscillator with permanent-magnet focusing (CO 43) proved to be particularly sensitive to distortions of the magnetic field. A relatively small piece of ferromagnetic material, such as a screwdriver, moved at distances of even a few feet from the oscillator, could influence the magnitude of the noise peaks and also, to a lesser degree, the frequency at which they occurred.

The noise peaks and oscillations observed at the higher frequencies seem almost certainly to be harmonics of the fundamental which occurs at a frequency of about 1 Mc/s. Search measurements made at frequencies down to 100 kc/s did not reveal any lower component than that near 1 Mc/s. A fundamental frequency of about 1 Mc/s suggests that an ion effect is responsible for the phenomenon, since the ion plasma frequency to be expected for the conditions in these tubes is of this order. Ions are formed by the collision of electrons with residual gas molecules. Positive ions tend to be trapped inside the electron beam causing it to become, at least partially, neutralized. The characteristic frequency at which ion plasma oscillations occur has been derived by Tonks and Langmuir[14] and is

$$f_i = \frac{r}{2\pi}\left(\frac{q\rho_i}{m_i\varepsilon_0}\right)^{\frac{1}{2}} \qquad \ldots\ldots(5)$$

where r is a geometrical factor equal to 1 and $1/\sqrt{2}$, respectively, for longitudinal and transverse oscillations,

q is the ion charge,

m_i is the ion mass,

ρ_i is the ion charge density,

and ε_0 is the dielectric constant of free space.

Assuming a fully neutralized beam, eqn. (5) can be written

$$f_i = \frac{r}{2\pi}\left(\frac{e\rho_e}{m_e N\varepsilon_0}\right)^{\frac{1}{2}} \qquad \ldots\ldots(6)$$

where e is the electron charge,

m_e is the electron mass,

ρ_e is the electron charge density,

and $N = m_i/m_e$.

We have further

$$\rho_e = \frac{J}{v_e} = \frac{J}{\left(2\frac{e}{m_e}V_0\right)^{\frac{1}{2}}} \qquad \ldots\ldots(7)$$

where J is the electron beam current density,

v_e is the electron velocity,

and V_0 is the beam acceleration potential.

Using m.k.s. units, the frequency for ion plasma oscillations then becomes

$$f_i = \frac{2{\cdot}92\times10^7\, rJ^{\frac{1}{2}}}{N^{\frac{1}{2}}\times V_0^{\frac{1}{4}}} \qquad \ldots\ldots(8)$$

For the BA 9–20 oscillator numerical values are: $J = 3\times10^3$ A/m^2, $V_0 = 800$ V, and $N = 5{\cdot}8\times10^4$ assuming singly-ionized diatomic oxygen. The corresponding values for the CO 43 are almost identical. For longitudinal ion plasma oscillations this gives the frequency

$$f_{i,l} = 1{\cdot}25 \text{ Mc/s}$$

and for transverse ion plasma oscillations

$$f_{i,t} = 0{\cdot}89 \text{ Mc/s}.$$

These figures are in good agreement with the sideband frequencies at which the fundamental oscillation and the fundamental noise peak have been observed.

The assumption of a fully neutralized beam is, however, questionable, and the electron-ion configuration in the tube does not represent what is normally understood by a plasma, since the electrons form a reasonably well-defined beam surrounding an ion core caused by the positive ions accumulating in the space-charge depression. On the other hand, it was observed that the frequency of the sideband oscillation, or noise peak, increased with increasing beam current, and decreased with increasing accelerating voltage, as would be expected from eqn. (8).

Oscillations in ion-neutralized electron streams have been dealt with in a number of papers.[9, 15–25] Pierce[15] has shown that wave amplification near the frequency for *longitudinal* ion plasma oscillations can exist in travelling-wave tubes. A feedback mechanism, provided by slow-speed secondary electrons moving in the backward direction, may then give rise to self-sustained oscillations at that frequency. However, the observed influence of the focusing field on the oscillations in the CO 43, and on the noise peaks in the BA 9–20 does not agree with a model based solely on longitudinal fluctuations. Rather it suggests that *transverse* effects play an important part. The existence of a transverse mode of oscillation in long electron beams, caused by the presence of positive ions, has been verified by Hernqvist,[17] Mihran[19] and Smith.[9]

In the present measurements it was observed that when the amplifier of the measuring system was tuned to a discrete oscillation and the output of the second detector was displayed on an oscilloscope, oscillograms similar to that shown in Fig. 4 (*a*) were obtained. When, on the other hand, a noise peak was examined in the same way, the oscillograms took various forms, of which those shown in Figs. 4 (*b*)–4 (*f*) are typical. The oscillograms shown were obtained for differently shaped noise peaks at different stages in the transition process between broad peak and discrete oscillation. However, because of the instability of the peaks, it was almost impossible to co-ordinate a particular oscillogram with a particular shape of a noise peak. As can be seen from the oscillograms, pulses and groups of pulses having various pulse repetition frequencies were observed. For example, the repetition frequency in Fig. 4 (*d*) is 3 kc/s, in Fig. 4 (*e*) 6 kc/s, and that in Fig. 4 (*f*) about 30 kc/s. Since a pulsed oscillation, such as those represented by the oscillograms, has a certain spectral width, the noise peaks might be due to ion plasma oscillations which are maintained only over certain time intervals. The effect then is similar to "squegging" in low-frequency oscillators; there, however, it is the carrier which is pulsed. Another, possibly more likely, explanation would be the existence of a form of relaxation effect in which the peaks observed consist of a continuous noise spectrum about the ion plasma frequency which oscillates according to the relaxation frequency.

The measured noise spectra given in Figs. 2 and 3 confirm the validity of eqn. (4) for the case of the two backward-wave oscillators. According to eqn. (4) the sum of the noise measured with the a.m. direct-detection, and the f.m. direct-detection, systems should be four times the noise value measured with the superheterodyne system at either sideband. Equation (4), however, is based on the assumption that the amplitude, and frequency, modulations take place at the same time instant. If a time lag exists, eqn. (3) does not hold, and hence eqn. (4) does not. Thus the measured spectrum curves show that no marked time lag exists between amplitude and frequency noise modulations caused by the same event.

4.2. *Electrostatically-focused Backward-Wave Oscillator*

The "Ophitron", which has an electrostatic focusing system, was available only for a short time, and for this reason only a total-noise measurement could be made.

Figure 5 shows that the spectrum does not contain any noise peaks or oscillations. The absence of these effects is probably due to the transverse drainage of positive ions which is an inherent feature of the tube. The total-noise spectrum is otherwise of the same order of magnitude as those obtained for the two magnetically-focused oscillators.

4.3. *Reflex Klystron*

In Fig. 6 noise spectra for the CV 323 reflex klystron are shown. The a.m. plus background, f.m. plus background, and total-noise spectra fall continuously with increasing frequency. No noise peaks or oscillations could be detected. Cursory measurements made on other klystrons (723 A/B, 2K25, CV 129, V–58) also failed to reveal any peaks or oscillations. Over the whole frequency range the spectra are of smaller magnitude than the corresponding ones for the backward-wave oscillators investigated. However, between 10 and 20 Mc/s the a.m. plus background noise spectrum approaches to within 3 dB of the corresponding backward-wave oscillator spectra.

5. Separation of Noise Components

It has been shown that it is not possible to make an exact evaluation of the a.m., f.m. and background

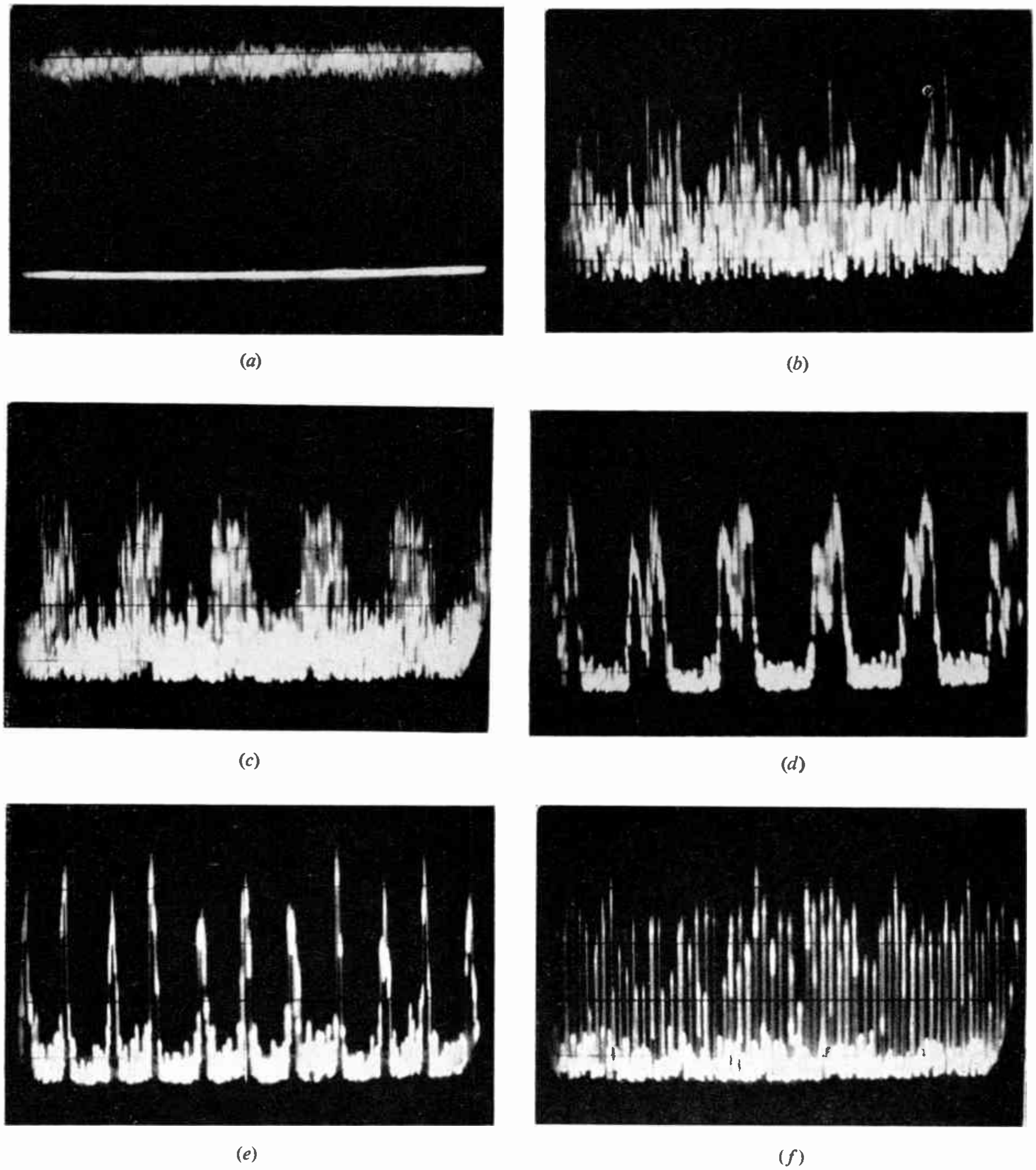


Fig. 4. Waveforms of (*a*) discrete oscillation, (*b*) to (*f*) noise peaks, caused by positive ions.

noise sideband components from a.m. and f.m. direct-detection, and superheterodyne spectrum measurements. Approximate values of the separate components can nevertheless be derived by the following method, where the background noise is first estimated with the help of noise measurements made when the tube is in a non-oscillating state.

5.1. *Separation Procedure*

If the electron-beam current of a microwave oscillator is reduced below a certain critical value, the starting current, the tube does not oscillate. It produces, however, a noise output with a maximum near the frequency where the oscillation is likely to occur at higher currents. The pre-oscillation noise spectrum

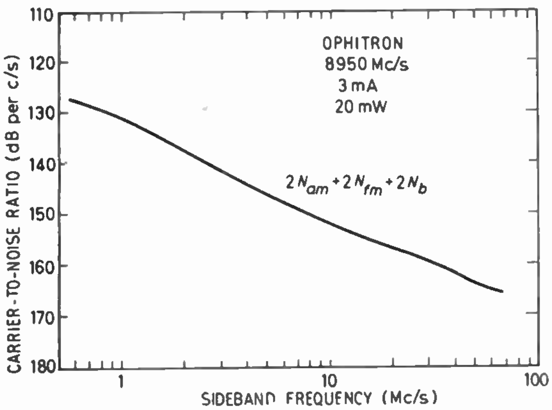


Fig. 5. Ophitron total-noise spectrum.

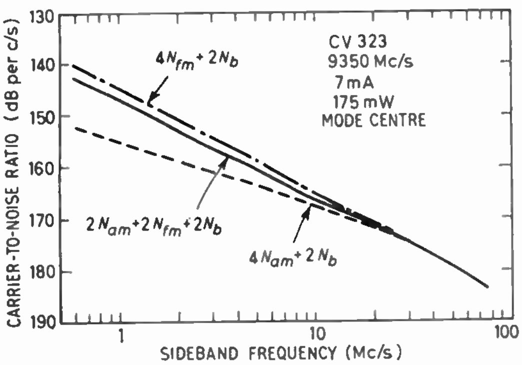


Fig. 6. Measured noise spectra of CV 323 reflex klystron.

can be measured with a superheterodyne system. Typical curves of such spectra are given in Figs. 7 and 8. It seems reasonable to assume that this noise, which appears when the tube is in the non-oscillating state, consists principally of background noise, certainly at the higher spectral frequencies. This point is discussed more fully in Section 5.2.1. The curves measured in the non-oscillating state can be used to estimate the slope of the background noise-frequency curve when the tube is oscillating since, at the higher sideband frequencies, the slope does not appear to change as a function of beam current. For the klystron the slope of the pre-oscillation noise-frequency curves varies with frequency but is independent of beam current for frequencies down to about 1 Mc/s from the centre frequency. With the backward-wave oscillators the curves are straight and parallel (in a logarithmic plot) at larger beam currents, and lie below those for the starting current, for frequencies down to 2 or 3 Mc/s.

To obtain the background noise component of the *oscillating* tube, measurements are taken at normal beam currents with the a.m. and the f.m. direct-detection systems up to and above that spectral frequency where both measurements yield the same reading. According to eqns. (1) and (2) the modulation components have become insignificant at this frequency, and only background noise is present. (Equal readings can also be obtained if the a.m. and f.m. noise components are equal in magnitude, but this possibility is discounted as being very unlikely.) With the help of this known value of background noise, and assuming that the slope of the background noise-frequency curve is essentially the same as when the tube is just not oscillating, the approximate background-noise spectrum at the lower frequencies can be determined. Knowing the background noise, it is then possible to calculate the a.m. and f.m. noise components from eqns. (1) and (2).

Apart from the measurements under pre-oscillation conditions, superheterodyne measurements are not essential for the evaluation of the separate noise components. They may serve, however, as a check on the results obtained from the two direct-detection methods by making use of eqn. (4).

One of the assumptions made here is that the upper and lower sidebands are of equal magnitude. If the background noise curves are asymmetrical about the carrier frequency, then the a.m. and f.m. components can still be determined but the background noise value obtained will be the mean of the values on either side of the carrier. Asymmetrical background noise is to be

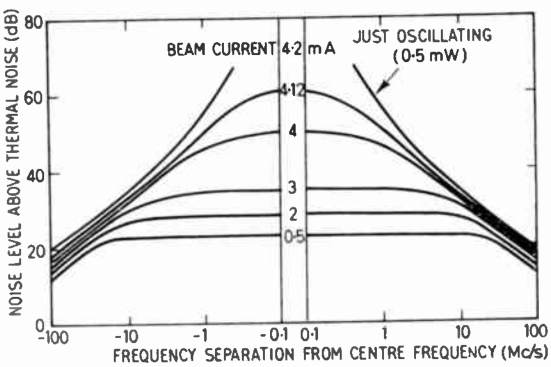


Fig. 7. Noise of CO 43 backward-wave oscillator in non-oscillating state.

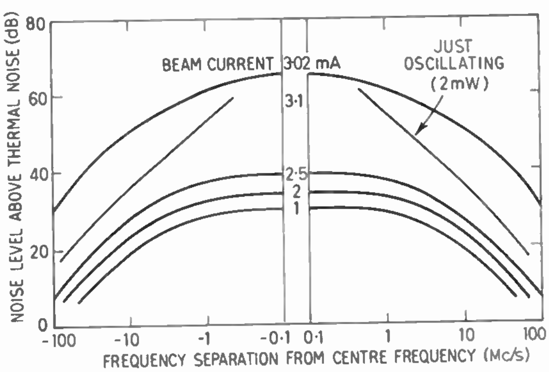


Fig. 8. Noise of CV 323 reflex klystron in non-oscillating state.

expected in backward-wave oscillators as will be seen in Section 7.

5.2. *Noise in Pre-Oscillation State*

The pre-oscillation noise measurements were carried out with the object of obtaining information about the frequency dependence of the background noise, in order to facilitate the separation of the noise components occurring in the oscillating state. At the same time a knowledge of the pre-oscillation noise level is of wider interest in that it may lead to a better understanding of the noise generation mechanisms.

5.2.1. Backward-wave oscillators

In Fig. 7 the noise curves obtained for the CO 43 backward-wave oscillator are given. They show that the noise level rises rapidly with increasing beam current, until at a current between 4·12 and 4·2 mA the tube breaks into oscillation. The curve for the current of 4·2 mA was obtained with the tube just oscillating. Even at a beam current of about one tenth of the starting current, the oscillator has a noise output approximately 20 dB above the thermal noise level. This noise is mainly due to the selective amplification of beam noise at microwave frequencies, by a process similar to that in backward-wave amplifiers. With increasing current the amplification increases and, finally, the oscillation starts.

At lower currents the noise spectrum is flat over the relatively wide frequency range of 10 to 20 Mc/s. The curves are straight and parallel at the higher frequencies, and the frequency range over which they remain parallel increases with increasing beam current.

The amplified beam noise is of the type which has hitherto, in these investigations, been referred to as background noise. No sinusoidal microwave component exists through which, and by a modulation process, low-frequency noise components can appear in the microwave spectrum. At higher beam currents, when the noise about the centre frequency rises to a relatively high level, it is possible that a modulation process occurs in which the low-frequency and the high-frequency components involved are both noise spectra. That a form of modulation possibly exists in the non-oscillating state, is suggested by the results of measurements made on the BA 9–20 oscillator both with and without ion drainage (Fig. 17). As will be shown in Section 6, transverse ion drainage in the BA 9–20 reduces the modulation noise components present in the oscillating state appreciably. Figure 17 indicates that even before oscillation a slight noise reduction takes place at the higher beam currents when ion drainage is applied. However, this reduction, which indicates the presence of a certain amount of "modulation noise", occurs only near the centre frequency, and thus it can be assumed that down to a few megacycles per second the curves given in Figs. 7 and 17 represent background noise only.

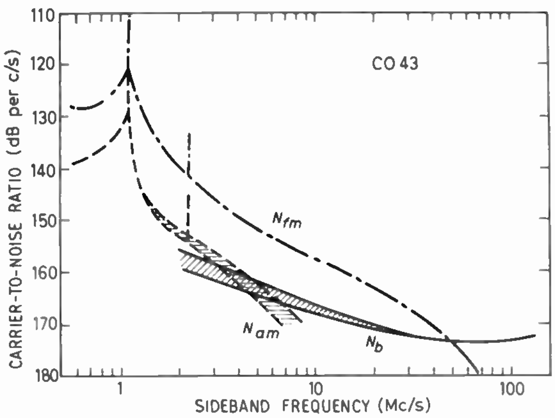


Fig. 9. Separated noise components of CO 43 backward-wave oscillator.

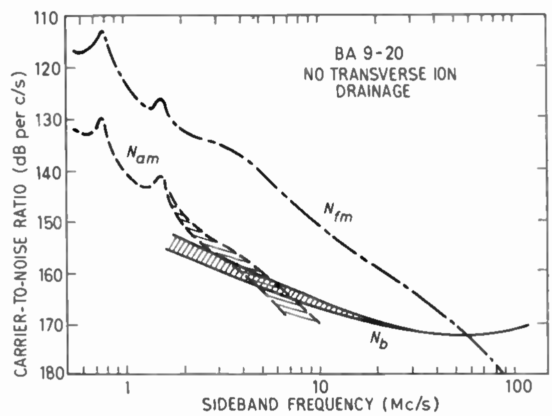


Fig. 10. Separated noise components of BA 9–20 backward-wave oscillator.

5.2.2. Reflex klystron

Figure 8 shows the pre-oscillation noise curves obtained for the CV 323 reflex klystron.

The behaviour of the klystron is different from that of the backward-wave oscillators in so far as the noise level rises with increasing beam current by a considerable amount, in fact to a maximum of more than 65 dB above the thermal noise level, before falling sharply at a beam current between 3·02 and 3·1 mA when the oscillation starts. As can be seen from Fig. 8, at a current of 3·1 mA the noise curve measured when the tube was just oscillating is markedly lower than the one measured just below oscillation. The slope of the curve is, except very near the centre frequency, reasonably independent of beam current, although it varies with frequency.

5.3. *Spectra of Separated Noise Components*

5.3.1. Backward-wave oscillators

Spectra of the separated noise components of the CO 43 and BA 9–20 backward-wave oscillators are shown in Figs. 9 and 10. The separation procedure is made more difficult in the case of the backward-wave oscillators by the fact that the a.m. plus background noise curves rise again towards higher spectral frequencies. Figures 9 and 10 show that as a result the determination of the a.m. noise and background noise becomes less accurate over the region indicated by the shaded areas. At first sight it may be surprising to find a minimum in a.m. plus background noise at about 75 Mc/s; the explanation for this is given in Section 7.

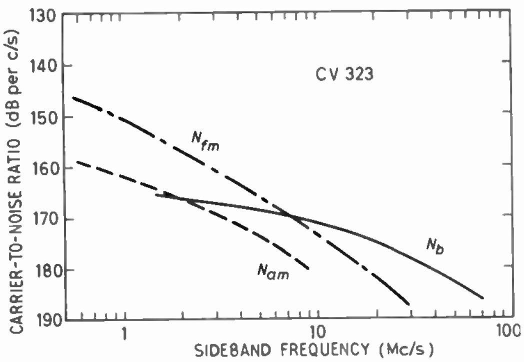


Fig. 11. Separated noise components of CV 323 reflex klystron.

The curves showing the separated components reveal that the f.m. component is the predominant one up to a frequency of about 50 Mc/s. The a.m. noise component becomes negligible above 10 Mc/s. This is of particular interest when these backward-wave oscillators are used as local oscillators in radar receivers, since it means that at the four intermediate frequencies commonly used—15, 30, 60 and 90 Mc/s—only background noise contributes to the overall receiver noise. Any attempt to reduce further the detrimental noise at these frequencies thus necessitates the reduction of microwave, and not modulation, noise.

5.3.2. Reflex klystron

Figure 11 shows the separated noise components. The f.m. component again predominates, although not so markedly as in the case of the backward-wave oscillators. The a.m. noise falls below the background noise at frequencies greater than about 2 Mc/s.

As one expects low-frequency beam noise to be mainly responsible for the modulation noise—this is confirmed by the correlation measurements quoted in Section 9—it is of interest to compare the "frequency pushing" characteristics of the oscillators with the modulation noise components obtained. The pushing behaviour of the CV 323 klystron and the CO 43 backward-wave oscillator is shown in Fig. 12. As can be seen, the change in output power for a variation in beam current is approximately the same for klystron and backward-wave oscillator, whereas the frequency change as a function of beam current is much more pronounced in the case of the backward-wave oscillator. Thus the fact that the f.m. noise is considerably stronger in the backward-wave oscillator than in the klystron, suggests that pushing effects due to microscopic factors, e.g. beam noise, and those due to macroscopic factors, are similar.

6. Noise Reduction by Transverse Ion Drainage

In several publications it has been proposed that beam focusing might be improved by the trapping of positive ions in the beam to neutralize the negative electron space charge.[26–29] Unfortunately the presence of positive ions usually results in undesirable noise effects as described above, and for this reason it is advisable to prevent any positive ions from forming or to remove any which are formed. The formation of ions can obviously be greatly reduced by sufficiently outgassing the tube parts and maintaining a good vacuum. On the other hand, the ions produced can be removed by a suitable drainage system.

The "Ophitron" backward-wave oscillator inherently possesses an ion drainage system formed by its focusing plates. In order to investigate the effects of ion drainage on the BA 9–20 oscillator the transverse ion drainage arrangement described in Section 2 was fitted. This type of drainage arrangement results in a marked noise reduction.[10] Ion drainage *along* the tube axis, suggested for example by Peter,[30] was not successful with this tube.[31]

Here we are concerned with the effect of this form of ion drainage on the individual noise components.

In Figs. 13 and 14 are shown the measured and separated BA 9–20 noise spectra when the drainage grids are 50 V negative with respect to the slow-wave structure. It can be seen that no peaks occur and that the spectra are generally lower than when there is no

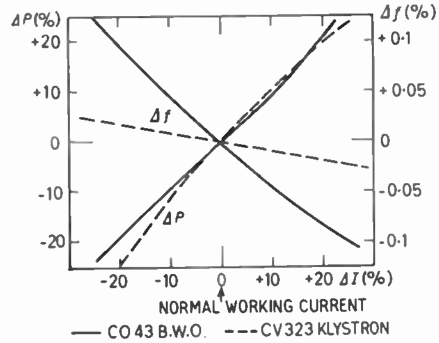


Fig. 12. Pushing characteristics.

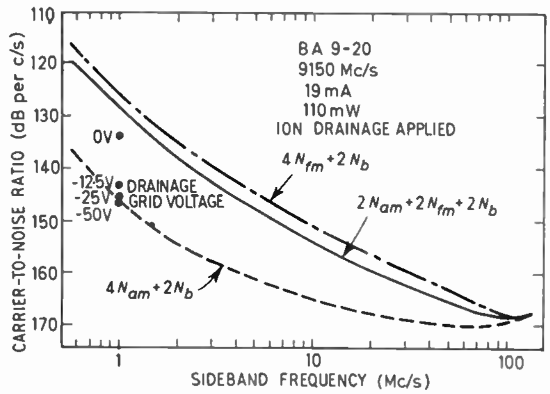


Fig. 13. Measured noise spectra of BA 9–20 backward-wave oscillator with transverse ion drainage applied.

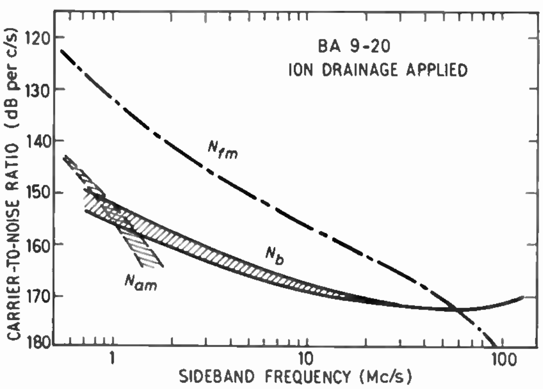


Fig. 14. Separated BA 9–20 backward-wave oscillator noise spectra with ion drainage applied.

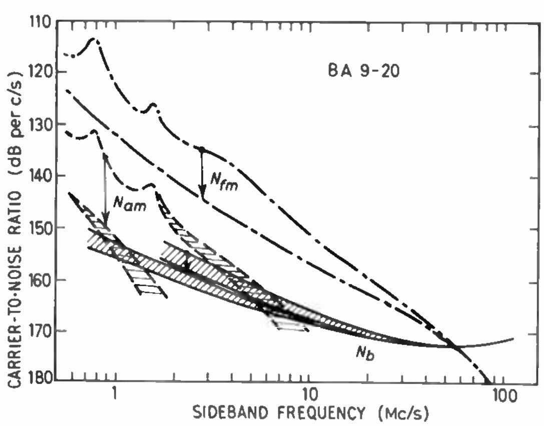


Fig. 15. Noise reduction by transverse ion drainage in BA 9–20 backward-wave oscillator.

ion drainage (Figs. 3 and 10). The noise reduction as a function of drainage grid voltage is shown in Fig. 13 for a.m. plus background noise at a frequency of 1 Mc/s. A (relative) grid potential of only −12·5 V decreases the noise by about 10 dB. Grid potentials lower than 50 V cause no further reduction beyond the maximum amount of 13 dB.

The degree of noise reduction becomes more obvious when the separated spectra with and without ion drainage are compared as in Fig. 15. The noise is reduced over a broad band of frequencies, the a.m. noise being decreased by 10 to 15 dB, and the f.m. noise at lower frequencies by approximately 10 dB. The reduction in f.m. noise becomes less marked at frequencies beyond about 10 Mc/s. This seems to be in agreement with the fact that ion noise occurs near the ion plasma frequency and possibly at a number of harmonics. It appears from the curves that the background noise is also slightly reduced, but this might be due to inaccuracies in determining the separated noise components. Generally, the drainage experiment indicates that noise due to ion effects is more pronounced than assumed so far. Not only oscillations and noise peaks, but also a broad band of relatively large magnitude noise can be attributed to the presence of ions. The modulation left with maximum ion drainage applied, is probably due solely to low-frequency cathode noise, and possibly also partition noise.

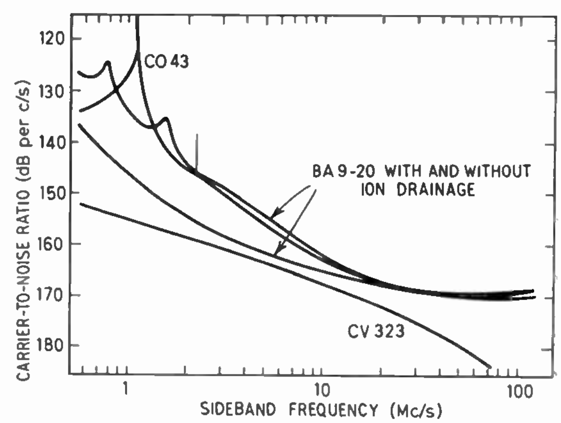


Fig. 16. A.M. plus background noise spectra for various oscillators.

In Fig. 16 all the a.m. plus background noise curves for the different oscillators are shown together. Although the noise reduction in the BA 9–20 by ion drainage is quite marked, the CV 323 reflex klystron has still the better performance over the whole frequency range. Ion drainage does not improve the backward-wave oscillator noise over the radar intermediate-frequency range as the spectrum at these frequencies consists of background noise only.

Figure 17 shows the results of measurements taken on the BA 9–20 tube under pre-oscillation conditions with and without ion drainage. The curves indicate that even in the non-oscillating state ion drainage can reduce the noise by about 3 dB near the centre

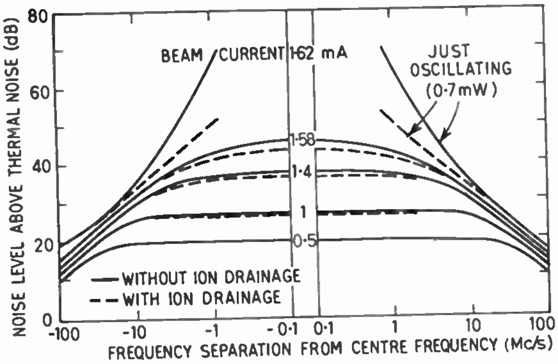


Fig. 17. Noise of BA 9–20 backward-wave oscillator in non-oscillating state with, and without, ion drainage.

frequency for the case when the tube is just not oscillating (1·58 mA beam current). This suggests that a form of modulation by ion effects occurs even in the non-oscillating state, since it seems unlikely that the microwave (background) noise can be affected by applying the drainage voltage.

7. Difference in Upper and Lower Sideband Magnitudes

7.1. *Backward-Wave Oscillators*

When measurements were made on the backward-wave oscillators with the superheterodyne system, it was found that for frequencies from about 30 to 40 Mc/s upwards the noise magnitude in the lower sideband exceeded that in the upper sideband. An extension of the measurements to a frequency of 200 Mc/s revealed the existence of a relatively broad noise peak, approximately 150 Mc/s below the carrier frequency. This frequency value of 150 Mc/s suggests that the observed peak may be due to the same mechanism that produces the higher-order, or spurious, oscillations, sometimes encountered in backward-wave oscillators, the strongest of which occurs at a frequency about 1 to 2% below the main oscillation frequency.

The theory of the backward-wave oscillator predicts the possibility of spurious oscillations,[32, 33] and other workers have made a special study of this phenomenon.[34, 35]

In a backward-wave tube oscillations can arise if the difference between the beam velocity and the phase velocity of the backward wave is such that the overall phase shift around the complete beam-circuit structure is given by

$$\left(\frac{\omega}{v_p} - \frac{\omega}{v_e}\right)L = (2m+1)\pi \qquad \ldots\ldots(9)$$

where ω is the radian oscillation frequency,

v_p is the phase velocity of the wave,

v_e is the electron velocity,

L is the geometrical length of the interaction region,

and m is an integer.

For the main oscillation $m = 0$, whereas for $m = 1$ an additional (first-order) spurious oscillation can occur, its frequency being approximately 1 to 2% lower than that of the main oscillation. Further higher-order modes are possible, and since intermodulation can occur as a result of the non-linear properties of the beam, apparent oscillations might also be observed on the high-frequency side of the main oscillation. The higher-order oscillations require a larger starting current than the fundamental. Since the starting current is affected by the magnitude and phase of reflection on the slow-wave structure, the occurrence of spurious oscillations depends both on the beam current and also to some extent on the matching conditions.

Figure 18 shows the result of total-noise measurements performed on the BA 9–20 oscillator (without ion drainage) up to a frequency of 200 Mc/s on either side of the carrier. Noise peaks occur in the frequency range 130 to 160 Mc/s below the carrier frequency, the peak height rising with increased electron-beam current. A slight increase in noise level can also be observed at frequencies above that of the carrier. It seems reasonable to assume that the noise peaks, and the broad-band increase in noise, consist predominantly of background noise, since they are similar to the noise observed at currents just below the starting current of the main oscillation.

A spurious oscillation could be induced in the BA 9–20 tube at a frequency 155 Mc/s below the main-oscillation frequency when, with a beam current thirty times the starting current, the output load was deliberately mismatched. This seems to confirm the suggested generation mechanism of the noise peaks.

An effect corresponding to the noise increase observed, occurs in backward-wave amplifiers. Currie

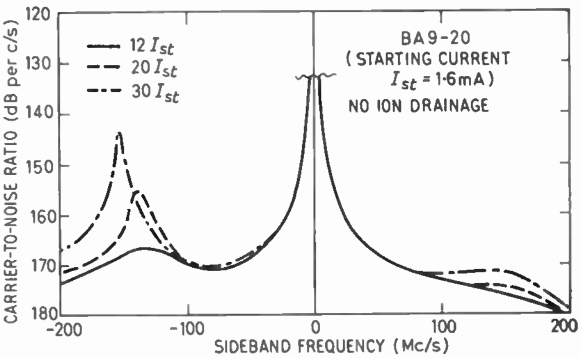


Fig. 18. Upper and lower sidebands of total noise of BA 9–20 backward-wave oscillator.

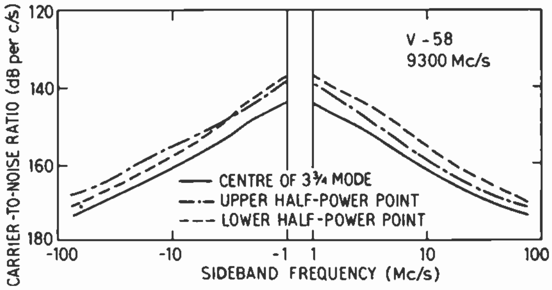


Fig. 19. Total-noise spectra of V-58 klystron as function of electronic tuning.

and Forster[36] report that the attenuation of a cascade backward-wave amplifier outside its passband is reduced considerably near the spurious-oscillation frequency. At this frequency electronic gain reduces the normally high isolation between input and output circuits of the tube.

7.2. *Reflex Klystron*

Figure 19 shows upper and lower sideband spectra of total noise for the V–58 klystron. The measurements were carried out for the 3¾ mode of operation, with the tube tuned electronically to (*a*) the centre of the mode, (*b*) to the high-frequency half-power point, and (*c*) to the low-frequency half-power point.

The results confirm the well-established fact that the noise is a minimum at a mode centre. At the mode centre the two sidebands are approximately equal in magnitude. At the upper half-power point the lower sideband is stronger for frequencies greater than a few megacycles per second, and at the lower half-power point the upper sideband contributes more noise. These observations agree reasonably well with the results obtained by Kuper and Waltz[2] for 723A/B reflex klystrons at the single sideband frequency of 30 Mc/s.

8. Carrier Frequency Deviation

Measurements of r.m.s. carrier frequency deviation were carried out on the CO 43 and BA 9–20 backward-wave oscillators and the CV 323 klystron, using a transmission-cavity discriminator.[1] The bandwidth of the amplifier employed was approximately 10 c/s to 4 Mc/s.

Typical values of r.m.s. carrier frequency deviation obtained were 6·5 kc/s for the CO 43 oscillator and 5 kc/s for the BA 9–20 (without ion drainage). The curve from which the value for the CO 43 has been deduced is shown in Fig. 7 of reference 1. Frequency deviations reported by other workers for backward-wave oscillators are 1 to 10 kc/s[7] and 8 kc/s.[8]

For the CV 323 klystron the values of deviation obtained are in the region of 7 kc/s. This is higher than the deviation of 2 to 3 kc/s measured by Gottschalk[6] for various klystrons. The discrepancy can probably be explained by the fact that Gottschalk's amplifier had a lower cut-off frequency of a few kilocycles per second, whereas the amplifier used here, with a cut-off frequency of about 10 c/s, responded to fluctuation components caused by the a.c. heater supply of the CV 323.

9. Correlation Measurements

The cavity-discriminator system has been used to determine the degree of cross-correlation existing between amplitude-modulation, and frequency-modulation, noise components, whereas the scatter-plot method was used to measure the correlation between the modulation noise and the low-frequency noise on the electron beam.[1] The degree of correlation was not determined as a function of frequency, but as an average value over the bandwidth (approximately 4 Mc/s) of the wide-band amplifier.

A correlation factor between a.m. and f.m. components of 0·35 and 0·4 was measured for the CO 34 and BA 9–20 backward-wave oscillators, respectively, whereas the corresponding figure for the CV 323 reflex klystron was about 0·6. Similar measurements made by Cicchetti and Munushian[8] on a backward-wave oscillator yielded a factor of 0·35, and Müller[4] obtained a figure of 0·8 to 0·9 for various reflex klystrons. The higher value given by Müller may be due to his amplifier having a bandwidth of only 0·3 to 5 kc/s, thus covering a range where the modulation components are by far predominant and background noise, which tends to reduce the measured correlation factor, is negligible.

The degree of correlation existing between the electron beam noise, and either f.m. or a.m. modulation noise, was found to be approximately 0·9 for both backward-wave oscillators investigated, and 0·9 to 0·95 for the CV 323 klystron. A strong correlation between beam and modulation noise is to be expected since modulation noise can only be caused by processes involving the electron beam with its non-linear properties.

10. Backward-Wave Oscillator Noise as Function of Focusing Field Strength

The focusing flux density of the BA 9–20 oscillator can be varied over a wide range by changing the focusing-coil current. A measurement of the noise dependence on the flux density was carried out at a particular sideband frequency, namely at 15 Mc/s. The a.m. plus background noise was determined by a direct-detection method, without applying transverse ion drainage. Figure 20 shows that the noise rises rapidly when the flux density is reduced below 0·05 to 0·06 Wb/m^2, whereas not much can be gained by increasing the flux density above 0·1 to 0·11 Wb/m^2. The noise spectrum measurements on the BA 9–20 oscillator described in the earlier sections were all

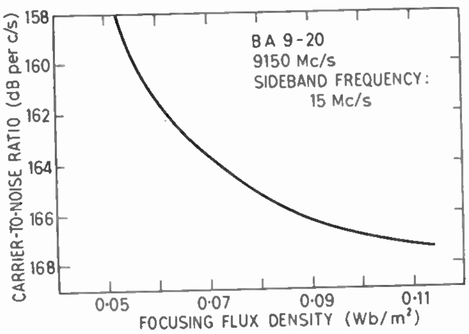


Fig. 20. A.M. plus background noise of BA 9–20 as function of focusing field at sideband frequency of 15 Mc/s.

performed at a flux density of about 0·08 Wb/m^2. The maximum noise reduction obtainable by an increase in focusing flux would at 15 Mc/s, for example, be approximately 2 dB. It is possible that the reduction in noise with increasing flux density is due to a reduction in partition noise. However, it is unlikely that there is any appreciable amount of partition noise at the relatively high flux density of 0·08 Wb/m^2, particularly as great care was taken in the alignment of the tube in the magnetic field. Unfortunately in the BA 9–20 the collector is internally connected to the slow-wave structure so that any possible interception current could not be measured.

In this context observations made on the influence of the focusing field on low-noise travelling-wave amplifiers are of interest. Currie and Forster[37] report that a satisfactory focusing of low-noise backward-wave amplifiers is possible with flux densities of only a few hundredths of 1 Wb/m^2, but a flux density of approximately 0·1 Wb/m^2 is necessary to obtain the lowest noise figure. The already-negligible interception current does not decrease further when the field is increased from the lower value and for this reason they conclude that the improvement in noise figure is not due to a reduction in partition noise. Other authors[38] have found that in a very-low-noise forward-wave amplifier an increase in flux density above 0·1 Wb/m^2 reduces the noise figure still further. However, the high flux density is only necessary in the electron-gun region. The cathode of a very-low-noise travelling-wave amplifier has normally a marked edge emission, so that the latter observations cannot be compared directly with those obtained for the backward-wave oscillator which has a fairly uniformly emitting cathode.

11. Reflex Klystron Noise for Different Modes of Operation

The quantitative theory of noise generation in reflex klystrons developed by Knipp[39] predicts that the noise power remains approximately constant from one operational mode to the next. Since the signal (carrier) output power normally falls with increasing mode number, the carrier-to-noise ratio should decrease from a low mode to a higher one. Kuper and Waltz[2] report that they could not observe this effect, but that the carrier-to-noise ratio remains relatively constant. Measurements made in the course of the investigations described here show the predicted behaviour, although the observed difference in carrier-to-noise ratio is small. Figure 21 shows the results of the measurements made on a V–58 reflex klystron for three different modes, the tube being tuned to the centre of each mode. It can be seen that the curves differ by between 1 and 3 dB. The measurements were carried out with the superheterodyne system, thus giving the total-noise spectrum.

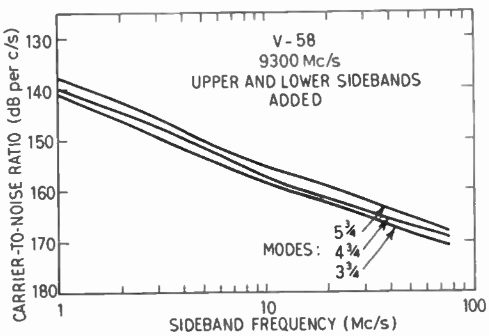


Fig. 21. V-58 klystron total noise as function of mode of operation.

12. Conclusions

The work described here shows that it is possible to make a.m. and f.m. direct-detection, and superheterodyne, noise spectrum measurements at sideband frequencies from about 0·5 Mc/s to 100 Mc/s and above. From measurements of the noise sidebands at beam currents less than the starting current an approximate evaluation of the separate a.m., f.m. and background noise components has been achieved. It is found that in the CV 323 klystron the f.m. exceed the a.m. noise sidebands by between 7 and 10 dB at frequencies at which the a.m. noise is detectable. At sideband frequencies below about 8 Mc/s the f.m. noise predominates but above this frequency background (microwave) noise becomes the only significant component. In the backward-wave oscillators the f.m. noise exceeds the a.m. noise by as much as 20 dB. Each of the components is greater than in the klystron and the background noise only exceeds the f.m. noise at frequencies above about 50 to 60 Mc/s.

When one of these devices is used as the local oscillator in a microwave receiver the f.m. noise is not important since the f.m. noise sidebands do not produce an output from the mixer, providing of course there are no frequency-sensitive elements present of bandwidth less than 100 Mc/s or so. The only significant contribution to local-oscillator noise at frequencies above about 10 Mc/s is thus background noise. It follows that for intermediate frequencies greater than 10 Mc/s ion modulation noise does not contribute to local-oscillator noise. The latter can be reduced only by reducing the microwave noise and ion drainage is of no advantage.

Not only do positive ions cause discrete oscillations and noise peaks, they also produce an increase in modulation noise over a wide frequency range. It appears that the background noise might also be increased slightly, particularly under pre-oscillation conditions. As shown by Smith[10] the fitting of drainage grids provides a very simple and effective way of removing positive ions from the electron beam of magnetically-focused oscillators.

Since eqn. (4) is found to apply, there is no marked time lag between amplitude and frequency noise modulations caused by the same modulating event.

It is of interest to note that at frequencies between 2·5 and 12 Mc/s the a.m. plus background noise (i.e. the local-oscillator noise) of the BA 9–20 with ion drainage is within 3 dB of that of the CV 323 klystron. At higher frequencies the backward-wave oscillator background noise reaches a minimum and then increases again, reaching a maximum at the frequency of the first spurious oscillation. Thus at the higher intermediate frequencies the local oscillator noise of backward-wave oscillators is likely to be greater than that of reflex klystrons unless the former are designed so that the frequencies of the carrier and the first spurious oscillation are widely separated. This effect might be particularly important at lower carrier frequencies. For example with S-band tubes the frequency of the first spurious oscillation is usually separated by between 30 and 60 Mc/s from the carrier and thus lies in the middle of the normal intermediate frequency range.

As expected there is a strong correlation between beam and modulation noise.

13. Acknowledgments

The work was carried out in the Department of Electronics of the University of Southampton and the authors would like to thank the Head of the Department, Professor E. E. Zepler, for his interest and encouragement. Grateful acknowledgment is also made to Mr. D. H. O. Allen and Mr. N. W. W. Smith for providing a BA 9–20 backward-wave oscillator and for arranging the fitting of ion drainage grids. The authors are also indebted to Mr. J. E. Ryley for carrying out the necessary modifications and mounting these grids in the oscillator.

14. References

1. B. G. Bosch and W. A. Gambling, "Techniques of microwave noise measurements", *J. Brit.I.R.E.*, **21**, p. 503, 1961.
2. J. B. H. Kuper and M. C. Waltz. Work summarized in D. R. Hamilton, J. K. Knipp and J. B. H. Kuper, "Klystrons and Microwave Triodes", Chap. 17 (McGraw-Hill, New York, 1948).
3. G. C. Dalman and E. Ortiz, "Measurement of microwave local-oscillator noise", *Proc. Natl. Electronics Conf.*, **9**, p. 833, 1953.
4. G. C. Dalman and A. S. Rhoads, "Microwave oscillator noise spectrum measurements", *Trans. Inst. Radio Engrs*, (*Electron Devices*), **ED-1**, p. 51, 1954.
5. R. Müller, "Noise measurements of microwave local oscillators", *Trans. I.R.E.*, **ED-1**, p. 42, 1954.
6. W. M. Gottschalk, "Direct-detection measurements of noise in CW magnetrons", *Trans. I.R.E.*, **ED-1**, p. 91, 1954.
7. R. L. Krulee, "Carcinotron noise measurements", *Trans. I.R.E.*, **ED-1**, p. 131, 1954.
8. J. B. Cicchetti and J. Munushian, "Noise characteristics of a backward-wave oscillator", *I.R.E. National Conv. Rec.*, Part 3, p. 84, 1958.
9. N. W. W. Smith, "Noise in backward-wave oscillators", *Proc. Instn Elect. Engrs*, **105B**, p. 800, 1958.
10. N. W. W. Smith, "Ion noise in backward-wave oscillators", Paper read at the International Congress on Microwave Tubes, Munich, June 1960. See "Mikrowellenröhren", p. 540, (Friedr. Vieweg & Sohn, Braunschweig, Germany, 1961).
11. B. G. Bosch and W. A. Gambling, "Noise measurements on low-power X-band oscillators", Paper read at the International Congress on Microwave Tubes, Munich, June 1960. Ibid. p. 436.
12. B. G. Bosch, "An Investigation of Noise in Microwave Oscillators", Ph.D. Thesis, University of Southampton, 1960.
13. B. G. Bosch and W. A. Gambling, "A microwave panoramic noise spectrum analyser", Paper presented at I.E.E. Conference on Microwave Measurement Techniques, London, September 1961. (*Proc. Instn Elect. Engrs*, **109B**, p. 658, 1962.)
14. L. Tonks and I. Langmuir, "Oscillations in ionized gases", *Phys. Rev.*, **33**, p. 195, 1929.
15. J. R. Pierce, "Possible fluctuations in electron streams due to ions", *J. Appl. Phys.*, **19**, p. 231, 1948.
16. W. E. Waters, "Observations on ion oscillations in a cylindrical-beam tetrode under hard vacuum conditions", *Trans. I.R.E.*, **ED-1**, p. 216, 1954.
17. K. G. Hernqvist, "Plasma ion oscillations in electron beams", *J. Appl. Phys.*, **26**, p. 544, 1955.
18. K. G. Hernqvist, "Plasma oscillations in electron beams", *J. Appl. Phys.*, **26**, p. 1029, 1955.
19. T. G. Mihran, "Positive ion oscillations in long electron beams", *Trans. I.R.E.*, **ED-3**, p. 117, 1956.
20. C. C. Cutler, "Spurious modulation of electron beams", *Proc. Inst. Radio Engrs*, **44**, p. 61, 1956.
21. T. Moreno, "Spurious modulation of electron beams", *Proc. Inst. Radio Engrs*, **44**, p. 693, 1956.

22. R. L. Jepsen, "Ion oscillations in electron beam tubes; ion motion and energy transfer", *Proc. Inst. Radio Engrs*, **45**, p. 1059, 1957.
23. M. Ettenberg and R. Targ, "Observations of plasma and cyclotron oscillations", *Proc. Symp. Electronic Waveguides*, April 1958, p. 379. Polytechnic Inst. Brooklyn.
24. R. Liebscher and R. Müller, "Frequency noise in travelling-wave tubes", *Proc. Instn Elect. Engrs*, **105B**, p. 796, 1958.
25. J. T. Senise, "Note on positive-ion effects in pulsed electron beams", *J. Appl. Phys.*, **29**, p. 839, 1958.
26. E. G. Linder and K. G. Hernqvist, "Space-charge effects in electron beams and their reduction by positive ion trapping", *J. Appl. Phys.*, **21**, p. 1088, 1950.
27. K. G. Hernqvist, "Space-charge and ion-trapping effects in tetrodes", *Proc. Inst. Radio Engrs*, **39**, p. 1541, 1951.
28. E. L. Ginzton and B. H. Wadia, "Positive-ion trapping in electron beams", *Proc. Inst. Radio Engrs*, **42**, p. 1548, 1954.
29. M. E. Hines, G. W. Hoffman and J. A. Saloom, "Positive-ion drainage in magnetically focused electron beams", *J. Appl. Phys.*, **26**, p. 1157, 1955.
30. R. W. Peter, Chap. 5 "Low-noise travelling-wave tubes" in L. D. Smullin and H. A. Haus (ed.), "Noise in Electron Devices", (J. Wiley & Sons, New York, 1959).
31. N. W. W. Smith, Mullard Research Laboratories, private communication.
32. H. R. Johnson, "Backward-wave oscillators", *Proc. Inst. Radio Engrs*, **43**, p. 684, 1955.
33. P. Palluel and A. K. Goldberger, "The O-type carcinotron tube", *Proc. Inst. Radio Engrs*, 44, p. 333, 1956.
34. R. E. Booth and R. G. Veltrop, "Detection and measurement of spurious signals generated by microwave oscillator tubes", Electronic Defense Lab. Techn. Memo. EDL-M72, January 9, 1958.
35. H. G. Kosmahl, "Influence of magnetic focusing fields and transverse electron motion on starting conditions for spurious oscillations in O-type backward-wave oscillators", *Trans. I.R.E.*, **ED-5**, p. 252, 1958.
36. M. R. Currie and D. C. Forster, "Conditions for minimum noise generation in backward-wave amplifiers", *Trans. I.R.E.*, **ED-5**, p. 88, 1957.
37. M. R. Currie and D. C. Forster, "Low-noise tunable pre-amplifiers for microwave receivers", *Proc. Inst. Radio Engrs*, **46**, p. 570, 1958.
38. M. Caulton and G. E. St. John, "S-band travelling-wave tube with noise figure below 4 dB", *Proc. Inst. Radio Engrs*, **46**, p. 911, 1958.
39. J. K. Knipp, "Theory of noise from the reflex oscillator", R. L. Report No. 873, 10th January, 1946, summarized in D. R. Hamilton, J. K. Knipp and J. B. H. Kuper, "Klystrons and Microwave Triodes", Chap. 17, (McGraw-Hill, New York, 1948).

Manuscript received by the Institution on 20th December 1961. (Paper No. 770.)

ADDENDA

The following list of references was omitted from the paper "The physical formulae for excess noise in composition resistors" by T. S. Korn, J. D'Hooghe and J. M. Kleinplac which was published in the September 1962 issue of the *Journal.*

1. J. Bernamont, "Experimental study of fluctuation in a metallic conductor of small volume", *C.R. Acad. Sci. (Paris)*, **198**, p. 2144, 1934 (In French).
2. J. Bernamont, "Fluctuations of potential at the edges of a metallic conductor of small volume traversed by a current", *Annales de Physique*, **7**, p. 71, 1937 (In French).
3. D. A. Bell and K. Y. Chong, "Current-noise in composition resistors", *Electronic Technology*, **31**, p. 142, 1954.
4. D. A. Bell, "Phenomenological approach to current-noise", *Brit. J. Appl. Phys.*, **6**, p. 284, 1955.
5. C. J. Christensen and G. L. Pearson, "Spontaneous resistance fluctuations in carbon microphones and other granular resistors", *Bell Syst. Tech. J.*, **15**, p. 197, 1936.
6. E. Meyer and H. Thiede, "Resistors of thin carbon films", *Elekt. Nachr Tech.*, **8**, p. 237, 1935 (In German).
7. M. Surdin, "Contribution to the study of fluctuations in potential at the edges of a thin conductor carrying a current", *Rev. Gen. Electricité*, **47**, p. 97, 1940 (In French).
8. A. Hettich, "Geometrical dimensions and resistor noise", *Frequenz*, **4**, p. 14, 1950 (In German).
9. P. L. Kirby, "Units for current-noise", *Electronic Engng*, **32**, p. 412, 1960.
10. G. T. Conrad, Jnr., "A proposed current-noise index for composition resistors", *Trans. Inst. Radio Engrs (Component Parts)*, **CP-3**, p. 11, 1956.
11. G. T. Conrad, Jnr., "Noise measurements of composition resistors", *Trans. Inst. Radio Engrs (Component Parts)*, **CP-4**, p. 61, 1955.
12. P. L. Kirby, "Current-noise in Deposited Carbon Film Resistors", Welwyn Electrical Laboratories, 1960.

APPLICANTS FOR ELECTION AND TRANSFER

The Membership Committee at its meeting on 25th October 1962 recommended to the Council the election and transfer of 24 candidates to Corporate Membership of the Institution and the election and transfer of 66 candidates to Graduateship and Associateship. In accordance with Bye-Law 21, the Council has directed that the names of the following candidates shall be published under the grade of membership to which election or transfer is proposed by the Council. Any communications from Corporate Members concerning these proposed elections must be addressed by letter to the Secretary within twenty-eight days after the publication of these details.

CORPORATE MEMBERS

Transfer from Associate Member to Member

HEAD, Leonard Russell. *Chelmsford, Essex.*
JOSEPH, David Peter. *Auckland, New Zealand.*

Direct Election to Associate Member

BALL, Peter William, B.Sc.(Eng.). *East Barnet, Hertfordshire.*
BURCHALL, Malcolm Arthur. *Harrow Weald, Middlesex.*
JEFFREYS, Dennis Clifford, B.Sc., M.Sc. *Hyde, Cheshire.*
McGLASHON, Squadron Leader, Donald Charles., R.N.Z.A.F. *Wigram, New Zealand.*
MARSH, Leonard Maurice. *North Watford, Hertfordshire.*
NEWMAN, Karel, B.Sc. *Guildford, Surrey.*
ROGERS, Arthur. *Chelmsford, Essex.*
STOKES, John Abel Brynley. *Chippenham, Wiltshire.*
TRIMBY, Roy. *Watford, Hertfordshire.*
WALKER, Kevin. *Bolton, Lancashire.*
WHITLOCK, Reginald H. *Rugeley, Staffordshire.*

Transfer from Associate to Associate Member

ALLCOCK, Michael. *Stevenage, Hertfordshire.*
STEELE, Gerald Frederick. *Shepperton, Middlesex.*

Transfer from Graduate to Associate Member

KINALLY, Dennis Raymond. *Guildford, Surrey.*
LEWIN, John Ernest. *London, N.W.7.*
LI YUAN-LU, B.Sc.(Eng.). *Birmingham.*
PETERSEN, Robert Laurence, B.Sc. *East Molesey, Surrey.*
RHODES, Alan Temple. *Fareham, Hampshire.*
WILLIAMS, Roger Morley. *Taplow, Bucks.*
WISE, Joseph. *Hitchin, Hertfordshire.*

Transfer from Student to Associate Member

HILL, Eric Sylvester. *Wells, Somerset.*
ROBERTS, Douglas Ian. *Cranford, Middlesex.*

NON-CORPORATE MEMBERS

Direct Election to Associate

BRUMBY, Derek. *Roxwell, Essex.*
SAUNTER, Harry Desmond. *Singapore.*
THOMPSON, Flight Lieutenant John Howard, R.A.F. *Henlow, Bedfordshire.*

Transfer from Student to Associate

BURTON, Alan Christopher Charles. *Stevenage, Hertfordshire.*
BURTON, Thomas Jeffrey. *Cheshunt, Hertfordshire.*

Direct Election to Graduate

ANSARI, Sayeedul Hasan, B.Sc. *Karachi, Pakistan.*
BELL, Robert. *Belfast, Northern Ireland.*
BRENARD, Julian Alan. *London, N.6.*
CHAMPION, Eric John. *Harrow, Middlesex.*
CHARLES, Elias Yehooda. *Stanwell, Middlesex.*
COLLINS, John Joseph. *Dartford, Kent.*
CRANFIELD, Ronald Frederick. *Bristol.*
CRITCHLOW, William Raymond. *Liverpool.*
DUERDEN, John. *Chandlers Ford, Hampshire.*
ELLIS, Peter James. *Balcombe, Sussex.*
ELSDEN, Charles March, B.Sc. *Basingstoke, Hampshire.*
EVANS, Gordon Jack. *Purbrook, Hampshire.*
FAITHFUL, Christopher Paul. *Bishops Stortford, Hertfordshire.*
FOULKES, Harry. *Manchester.*
GAULDER, Clifford Francis Kerry. *North Wilmington, U.S.A.*
GRIFFITHS, John Michael. *Dorchester, Dorset.*
GUIVER, Edward John. *Rushden, Northamptonshire.*
HAINES, Michael John. *Feltham, Middlesex.*
HARTNELL, David John. *Taunton, Somerset.*
HAWKINS, Alan Harold. *Rhyl, Flintshire.*
HOBSON, Lieutenant, David Hugh. *Horsham, Sussex.*
HUNT, Denis John. *Evesham, Worcestershire.*
KIMBER, Gerald Adrian. *Woking, Surrey.*
LEUNG, Daniel Iu. *London, S.W.12.*
McCULLOCH, Patrick. *Basingstoke, Hampshire.*
MERCIECA, Lewis. *London, W.12.*
MURTAGH, Maurice. *Arborfield, Berkshire.*
OLDROYD, Brian. *Leeds, Yorkshire.*
OXLEY, Kenneth James. *Mitcham, Surrey.*
PEARCE, Richard John. *Basingstoke, Hampshire.*
PIERSON, Patrick Blake Kirshaw. *Cheshunt, Hertfordshire.*
PRIOR, Alan Charles. *Ilford, Essex.*
RASHID, Abdul, B.Sc. *Wembley, Middlesex.*
ROBSON, Geoffrey Michael. *Farnborough, Hampshire.*
SARWAR SHAIKH, M.A.(Ph.). *Karachi, Pakistan.*
SCARGILL, Philip. *Oxted, Surrey.*
SHIPTON, Raymond Hubert. *Portsmouth, Hampshire.*
SMITH, Brian Frederick. *Rochester, Kent.*
TAYLOR, David Eric. *Aldershot, Hampshire.*
WHALLEY, Lawrence Albert. *Lytham St. Annes, Lancashire.*
WHITE, Robert George. *Farnborough, Hampshire.*
WILLIAMS, Roger David. *Portsmouth, Hampshire.*
WILSON, Richard Emile. *London, N.W.3.*
WOODE, Alan David. *Brighton, Sussex.*
WOOLLARD, Alfred James. *Chelmsford, Essex.*
WOODWARD, Stanley Frederick Henry. *Northolt, Middlesex.*

Transfer from Student to Graduate

ARAVAMUDHAN, Krishna, B.Sc. *Madras, South India.*
BAGULEY, James. *Hayes, Middlesex.*
CHADDA, Santosh Kumar, M.Sc., B.Sc. *New Delhi.*
CHRISTIE, Stanley. *Melbourne, Australia.*
DAY, Geoffrey George. *Croydon, Surrey.*
DUARTE-CATULO, Fernando Jose Euginio. *Lisbon, Portugal.*
ELLIOTT, Francis Hamilton. *Belfast, N. Ireland.*
FISHER, Vincent Colston Austin. *Dartford, Kent.*
HALE, Derek Stanley. *Jersey.*
HUNTER, Thomas. *South Shields, Co. Durham.*
MILLS, Raymond Hugh. *Cambridge.*
RANSLEY, John David. *Christchurch, Hampshire.*
SCOTT, Henry George. *London, W.10.*
SUDUL, Keith Murray. *Feltham, Middlesex.*
WORRELL, Brian Frank. *London, N.W.9.*

STUDENTSHIP REGISTRATION

The following students were registered on the 25th October.

ADBY, Paul Raymond, B.Sc. *High Wycombe.*
BALL, David. *Old Bexley, Kent.*
BATEMAN, Colin Herbert. *Norwich, Norfolk.*
BURGESS, Reginald Horace. *London, S.E.12.*
CLIFTON, James Thomas. *Leamington Spa.*
CLOKE, Victor C. *Newbury, Berkshire.*
COWLEY, Ivor Gordon. *Ashford, Middlesex.*
D'SOUZA, Clifford. *Surbiton, Surrey.*
DHARAM-DASS, Dosi, B.Sc. *London, N.W.3.*
DOSHI, Chandrakant. *London, N.W.2.*
EKWEOZOH, Patrick Obijiofor. *Lagos.*
ELLIS, Gordon H. *Chippenham, Wiltshire.*
EMOKPAE, Agho Ogieva. *London, S.E.20.*
EVANS, John Brinley. *Swansea.*
FERNANDO, Vincent Lalin. *London, W.11.*
GASTON, Samuel. *Ahoghill, Northern Ireland.*
HUAT, Ooi Beng, Dip.El. *Kuala Lumpur.*
JOHNS, David Winston. *Plymouth.*
JONES, Peter John. *Epsom, Surrey.*
KARPATI, Kalman. *Mitcham, Surrey.*
KWOK, Woon Hup. *Singapore.*
LASAKI, Muriseli Folorunso. *Lagos.*
LIEBENBERG, Abraham S. *P.O. Dett., Southern Rhodesia.*
LOWE, Anthony Brian. *Windsor, Berkshire.*
MOMEN, Abdul Md., B.Sc.(Eng.), *London, N.19.*
MORLEY, Gerald, B.Sc. *Wigan, Lancashire.*
MOTTRAM, Ivan Roy. *Southampton.*
OAKES, Stanley. *Rossendale, Lancashire.*
OCHIAGHA, Cornelius Chukwudi. *Southampton.*
OFOCHE, Emmanuel B. C. *London, S.E.27.*
OGBANGA, Ernest T. S. *London, S.W.16.*
OJEI, Jerry E. *London, S.E.27.*
PINTO, Raul Sebastian F. *London, W.3.*
POOLE, Robert E., B.Sc. *Basingstoke.*
PROCKTER, Adrian C. *London, S.E.23.*
PROUDMAN, Roy. *London, N.W.2.*
SARMA, D. Unnikrishna, B.Sc. *Hyderabad, India.*
SHEKAR, Krishnarao Chandra. *Bangalore.*
SLEET, Trevor W. *Wokingham, Berkshire.*
SLOANE, Captain Richard Michael J. D., A.O.C. *Athlone, Eire.*
TAN, Peter Keng H. *Singapore.*
WILLIAMS, Leslie John. *Cardiff.*
WONG, Andrew Kui H. *London, S.E.4.*

U.D.C. 621.394/5: 681.39

Some Recent Developments in Data Transmission

By

K. L. SMITH, B.Sc.(Eng.)†

Presented at the Symposium on "Data Transmission" on 3rd January 1962.

Summary: Much development has taken place recently in the use of the telephone system for data transmission. Special precautions have to be taken to minimize effects of certain characteristics of the telephone system and two types of modulation in current favour are reviewed in some detail.

The division of responsibility between the data terminal manufacturer and the communication authority requires the definition of a standard interface between the data terminal and modem. A typical range of data transmission systems designed to meet this interface specification is introduced.

The extension to broadband systems and the use of error correction and error detection systems are also discussed.

1. Introduction

The introduction of solid-state circuit techniques has enabled the systems designer to evolve larger and more powerful data processing systems. Not only can these systems handle greater volumes of data than their predecessors, but their increased logical power enables a much broader approach to be made on organizational procedures.

Furthermore, in recent years there has been a general acceptance and adoption by industry of these systems and their application to a wide range of problems, although the approach in general has been based on batch processing. Thus small systems are used in plants for such operations as production scheduling, production control, and payroll, and used in warehouses for stock control and invoicing. Larger systems can be used in a head office for optimizing the whole company operations by determining suitable strategies to achieve maximum profitability. A need exists for speedy communication between these systems, in order that all procedures are based on the true situation as expressed by up-to-the-minute, accurate data. Clearly the postal services are inadequate for such purposes, since the delays involved would necessitate unreasonable duplication of data at several points and so prevent true integration of the data processing system. A similar need exists for gathering data from the primary sources into the local data processing centres, and such systems are discussed in a companion paper,[1] a typical configuration being shown in Fig. 1.

To provide these communication facilities, use is made of the existing telegraph and telephone networks, although the present paper is concerned only with data transmission facilities based on the public telephone network because of its greater speed capabilities. Furthermore discussion will be restricted to medium-speed systems which attempt to obtain maximum possible data speed from the existing telephone channel. Low-speed systems for data gathering where the emphasis is more on reduction in cost of the terminal are discussed in detail in the companion paper.[1]

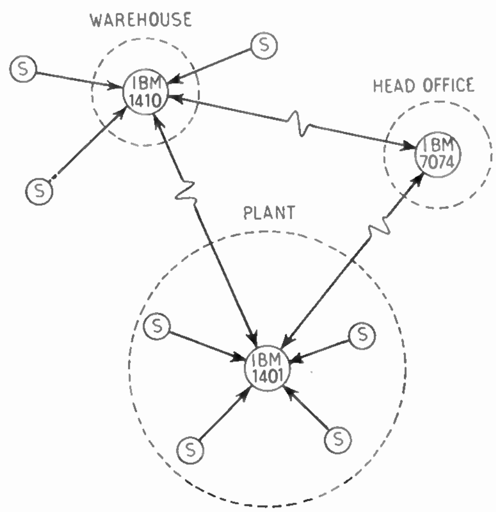


Fig. 1. Simplified data transmission network.

2. Characteristics of Telephone Network for Data Transmission

Although at first the telephone system would seem to offer exciting prospects for data transmission in terms of traffic capacity, reasonable cost and systems

† IBM United Kingdom Ltd, 101 Wigmore Street, London, W.1.

flexibility, this initial exhilaration is somewhat dampened by the sober realization that the network has been designed primarily for the transmission of speech, and that in order to provide an acceptable performance at minimum cost, advantage has often been taken of the insensitivity of the ear to certain phenomena to effect worthwhile economies in line and switching plant. Unfortunately, the types of data transmission system developed so far are particularly sensitive to these phenomena, and although future developments may reduce the sensitivity to some effects, others will inevitably remain an unsurmountable barrier to the full realization of the inherent capabilities of the telephone network. For example, the ear is insensitive to interruptions in the transmission path up to several milliseconds in duration caused by the use of cheap contact materials in the exchange switches, yet every one of such interruptions will inevitably cause the loss of several bits of data.

The principal causes of errors in data transmission which can be attributed to the telephone system are:

Circuit interruptions
Impulse noise
Bandwidth limitations
Frequency shifts
Amplitude distortion
Envelope delay distortion
White noise.

The major source of trouble is impulse noise caused by inductive disturbances from the exchange switching apparatus, or by crosstalk from dialling signals in adjacent circuits. The effect on speech reception is annoying; the effect on data transmission is disastrous, causing bursts of errors each of several bits duration. Furthermore, the occurrence of such bursts of errors is quite random and in order to characterize the telephone circuit for data transmission in the presence of noise, recourse has to be made to statistical methods based on many tests. The general problem is discussed more fully in a companion paper.[2]

Bandwidth limitations are caused mainly by the use of loaded cable in local line plant and restrict the speed of data transmission. The possible inclusion of such sections in a switched telephone call compels the designer of data transmission systems to provide variable speed facilities in his terminals, lest by catering for the worst circuit, he prevents the full exploitation of the majority of circuits which do provide quite a good bandwidth. For U.K. circuits the G.P.O., in effect, restricts data transmission signals to the range 900 to 2100 c/s plus a small band from 300 to 400 c/s, which allow data speeds of up to 1200 bauds. But the telephone authorities are continually upgrading their circuits by replacement of the older, heavily-loaded plant and so a continuous improvement in data speed performance can be expected.

Frequency shifts in telephone signals are introduced by lack of synchronism between multi-channel carrier-telephone terminals and together with amplitude and envelope-delay distortion hamper the reconstruction of the data carrier in certain types of phase-modulation system, although the total effect of all three phenomena seldom causes serious trouble, unless accompanied by excessive line loss.

It is unusual to meet appreciable levels of white noise on telephone circuits and then it is caused more by poor contacts than by the signal level approaching that of Johnson noise due to excessive line loss. It is, however, usual to make comparative and absolute tests of data transmission system performance in the presence of white noise, since such results can usually be compared with analytical assessments made possible by the facility of characterizing such noise by simple mathematical expressions.

3. Modulation Methods

In recent years a variety of modulation systems have been developed to provide a data transmission service over the public telephone network. Of the three classical systems of modulation only phase and frequency modulation have received much attention, due primarily to the protection they offer, with varying degrees of efficiency, against sudden level changes and impulse noise. Each method has its devotees and it is of some interest to consider the arguments submitted in support of each.

The supporters of frequency modulation argue that there is little difference between the performance of the two systems in the presence of impulse noise and level changes and that f.m. is superior to p.m. with its ability to work without need for synchronism between the data clock and the carriers. In fact, it is this ability to work asynchronously that has led to the use of f.m. in the service provided in the U.S.A. and its adoption in the United Kingdom by the G.P.O. for their proposed data transmission service.

The system used for the American service employs a fairly straightforward frequency shift circuit, in which the two frequencies are optimized for the particular line circuit and data speed employed. A zero cross-over principle is employed in the receiver detector.

On the other hand, the protagonists of phase modulation, counter that their systems are much simpler, involve the adjustment of only one carrier oscillator to match line conditions, require less bandwidth for the same data speed and have a distinct superiority over f.m. for performance in the presence of noise.

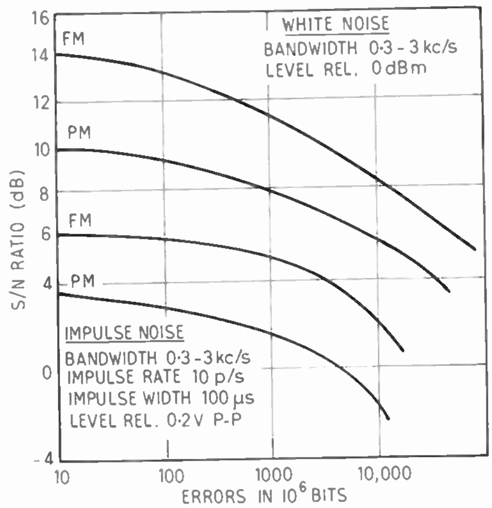


Fig. 2. Performance of modems in presence of impulse and white noise.

This advantage has a theoretical value of 3 to 5·5 dB (depending on the particular detector employed) in the presence of white noise, and has a measurable advantage in the presence of impulse noise. Figure 2 shows the results of tests made under identical conditions of Gaussian white noise and impulse noise for experimental f.m. and p.m. systems. A measured superiority for p.m. of at least 3 dB for a given error rate is apparent at most levels.

In terms of noise level this is not a very impressive improvement, but when expressed as a reduction in error rate at a given signal/noise ratio, the improvement is seen to be at least one and sometimes nearly two orders of magnitude. This has considerable advantage for the designer of data processing systems since his error rate with p.m. is reduced to less than 10% of the error rate with f.m. Furthermore, recent developments have produced p.m. systems capable of working asynchronously, i.e. without any synchronism between the data clock and carrier. In fact the curves of Fig. 2 relate to an asynchronous p.m. system, so that a direct comparison in performance with f.m. is possible.

It should be understood that this measurable advantage of the p.m. system over the f.m. was measured in the presence of impulse and white noise only. On some telephone circuits the effect of excessive envelope delay distortion can be to aggravate the carrier recovery in the p.m. system and reduce its advantage over f.m. and in some rare circumstances transfer the advantage to the f.m. system.

The evolution of this asynchronous p.m. system gives some idea of the rapid developments in data transmission over the last few years and the promise the future may hold in the pursuit of the ideal system.

In his original p.m. system[3] shown in Fig. 3, Hopner used a pilot scheme. Both clock and carrier were derived from a common pilot frequency, which was also sent along the line together with the modulated carrier. At the receiver the carrier required for synchronous detection was derived by modulating the delayed input (line) signal by the data output. Timing for the data was obtained by mixing the pilot with the derived carrier. The pilot was required primarily to derive a data clock at the receiver in order to squelch the output from the integrator at the correct frequency. This data clock could not be derived from the carrier because of the frequency shifts which are to be expected over carrier telephone circuits. It can be seen that the system readily corrects for these

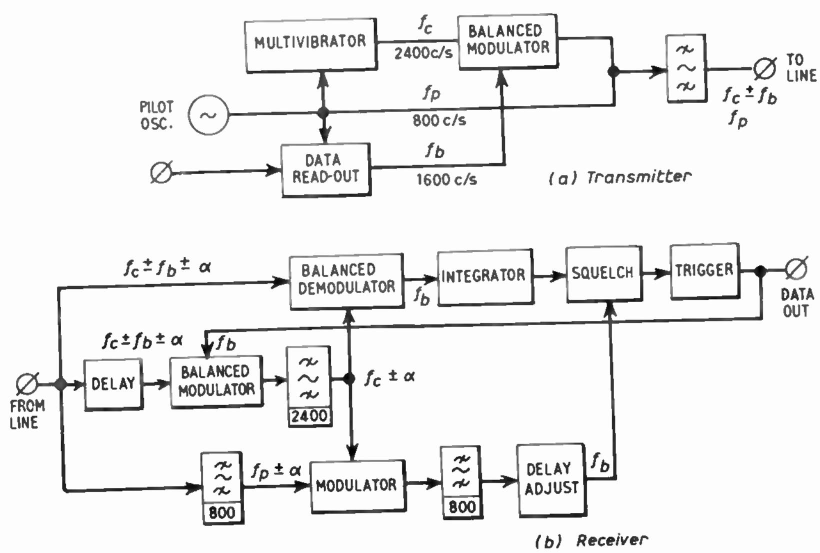


Fig. 3. Pilot synchronous phase reversal modulation system.

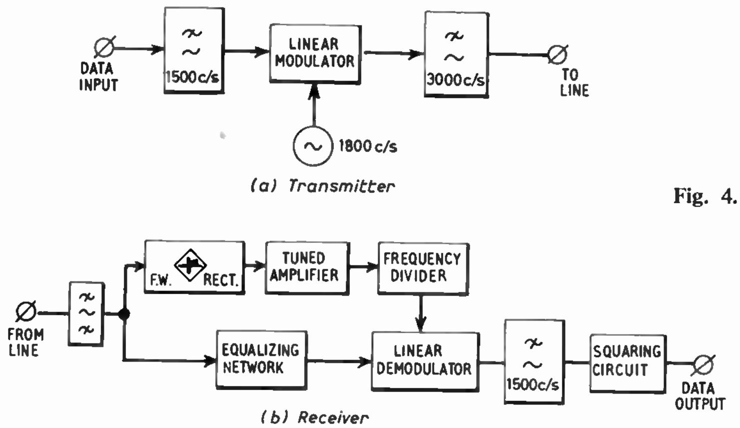


Fig. 4. Synchronous phase reversal modulation system.

frequency shifts. The major disadvantages of the pilot system were (*a*) the restrictions in range of data speed, since data, pilot and carrier frequencies had to be chosen so that pilot frequency always lay outside the signal spectrum and (*b*) the need to adjust the phasing of the data clock at the receiver to allow for envelope delay distortion along the line. This adjustment was tedious enough for private circuits, but would have necessitated an automatic facility for public network operation and caused an undesirable delay in operation. On the other hand the pilot system avoids any need to restrict code or information flow and could maintain synchronism even after the loss of eight consecutive bits. Synchronism between data clock and carrier oscillator was considered necessary to avoid modulation jitter which occurs as the bit rate approaches the carrier frequency.

For the same reason, Hopner's second system[4] still retained bit synchronism with the carrier, but eliminated the need for the pilot frequency by using a post-detection low-pass filter instead of the squelching circuit. The simplicity of this circuit is apparent from Fig. 4. The penalty paid for this simplification is the ambiguity of the phase of the received signal since the reference formerly provided by the pilot is no longer available. Consequently, a non-return-to-zero (or n.r.z.) system has to be used, in which a binary "1" is represented by a change in phase of the carrier, rather than by a particular phase condition. This causes little difficulty in system operation, but does necessitate suitable conversion circuitry in the data terminal. It will be seen that the carrier recovery circuit is identical to that used in the earlier system, and by using a narrow-band filter an essentially noise-free carrier is produced. This p.m. system has been used for most of the telephone circuit evaluation work undertaken by the author's organization in collaboration with many telephone authorities in Europe, which has shown conclusively that reliable and effective data transmission can be achieved over the public networks.

There still remained the problem of modulation jitter which necessitated the synchronism of the data clock and carrier frequency, with its difficulties of applications to a general data service and restrictions in the types of operation in the data terminals. Furthermore, the need for synchronism prevented the full realization of the data transmission capabilities of the circuits available, since at least one full cycle of carrier was needed for each bit of data. However, more recent work[5] has shown that this interference is caused largely by intermodulation products from the modulator producing "artificial carriers" at the receiver input which interfere with the carrier derived from the two major sidebands. A better understanding of this phenomenon is obtained by appreciating that phase modulation is broadly equivalent to amplitude modulation with the carrier suppressed. Thus the basic line signal is the two major sidebands, upper and lower, which between them adequately define both carrier and data signal. At the receiver, shown in Fig. 5, the carrier is derived by multiplying the two sidebands in a full-wave rectifier, which produces, *inter alia*, an output at twice the carrier frequency, from which the carrier is obtained by a simple frequency-halving circuit. As the data rate approaches the carrier frequency, intermodulation products are

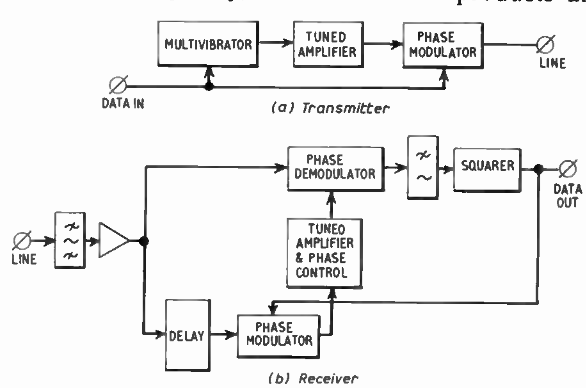


Fig. 5. Asynchronous phase reversal modulation system.

produced at near carrier frequency, which are sent along the line along with the major sidebands and interact with the carrier derived at the receiver to produce the characteristic modulation jitter. The principal cure to this problem adopted by Hopner in his latest circuit is to use linear modulation techniques and to restrict the harmonic content of the data signal to avoid these interference signals. Further protection against signal mutilation by the telephone circuit is provided by the equalizer network, which is a special filter arrangement converting that section of the receiver to vestigial sideband operation and using the least affected sideband from which to extract the data by synchronous detection with the locally derived carrier. Obviously, both sidebands are essential to extract this carrier, and jitter is further reduced by narrow-band filtering. Furthermore, frequency shifts along the line are automatically compensated by this system.

Thus this new circuit is a phase modulation system with all the facilities provided by frequency modulation, yet with a superior performance in the presence of noise and employing simpler circuitry, enabling an adjustment of carrier frequency to match line conditions to be achieved very easily. Such a system would allow data speeds of 600 and 1200 bauds over most circuits using a carrier of 1500 c/s. With a carrier of 1800 c/s a data speed of 2400 bauds requires a bandwidth from 600 to 3000 c/s which seems reasonable for most private wires.

These systems have all employed 180 deg phase reversal modulation, but it is clear that higher data rates could be achieved by using multi-phase modulation at the expense of circuitry complication and increased susceptibility to certain types of disturbance. Thus a four-phase system is available in the U.S.A. and has been used for an air line seat reservation system (SABRE) and the SAGE Defense Network. Data speeds of 2000 bauds are achieved over private telephone circuits, but the use of four-phase modulation involves the transmission of two adjacent bits of data simultaneously.

4. The Standard Interface Between Modulator and Data Terminal

Having seen that there are several satisfactory methods of transmitting data over the telephone network we next consider one of the major dilemmas faced by the manufacturers of data terminals and the telephone administrations. This is the placing and definition of the boundary between their respective responsibilities in the establishment of a data transmission link.

The monopolistic position of the telephone authority means that the telephone circuits will always be under the one control. On the other hand there will be a large number of suppliers and owners of data terminals, each with his own systems, designs and concepts, yet having a need to use the common telephone network. Between the data terminal and the telephone circuit a modulator/demodulator, or modem, is required, and the major problem is to decide who should supply this device.

One argument could be that, in view of the many forms of modulation, the best position for the interface, as the line of demarcation of responsibility is called, should be the exchange line change-over key; then complete freedom is given to all concerned.

Yet the signals produced in and used by the modem could have a serious effect on the telephone system, in both signalling interference and overloading of multichannel apparatus, and so there is considerable justification in making the provision of the modem the responsibility of the telephone authority and setting the interface accordingly. A further advantage of this arrangement is that communication between data terminals of different manufacture then becomes possible. Again, it is far easier to standardize the form of signal passing across this interface, since it is binary and serial in nature, so only the two voltage levels corresponding to the two binary states and the impedance level have to be defined. Other parameters such as pulse width are set by the line conditions through the data rate, etc.

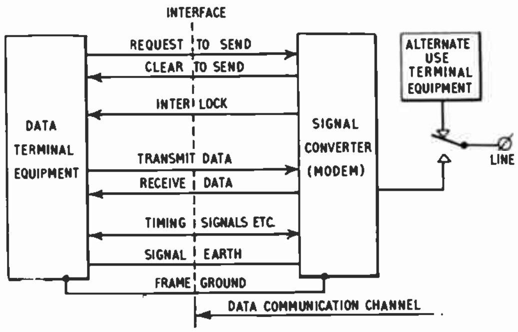


Fig. 6. Standard interface between data terminal and data communication channel.

An interface typical of those adopted in the United Kingdom and the U.S.A. is shown in Fig. 6 and it is seen that apart from the data signals and the various earth connections, certain supervisory and control signals are specified. For half-duplex operation the direction of transmission is switched to the transmit condition by the data terminal sending a "request to send" signal to the modem, which responds with a "clear to send" signal after a delay to allow for the turn-around of echo-suppressors, etc.

When data signals are received by the modem it holds the data terminal in the "receive data" condition via the "interlock" connection.

The adoption of such an interface has led to a straightforward introduction of data transmission services on the telephone networks. The modems are supplied by the telephone administration, who have complete control over the form of the signal which is sent over the telephone connection, and, what is even more important, can introduce up-to-date modems as developments in technology occur.

From the data terminal manufacturer's viewpoint he no longer has to concern himself quite so deeply with line transmission problems and can arrange for his data terminals to be suitably designed for connection to the modem most suitable for the circuit available.

5. The Synchronous Transmitter-Receiver Concept

The designer of the data terminal is faced with the need to handle data from the variety of data sources normally met in data processing systems. These range from punched card and paper tape units to magnetic tape and direct access to computer memories. Most of these sources make their data available one character at a time, and it is necessary to take this character and

(*a*) check the code validity on input and output and convert its code to a self-checking form to allow for the error characteristics of the transmission path;

(*b*) serialize the encoded character for transmission to line;

(*c*) provide synchronizing means between terminals to produce a stable and accurate clock for the data and so achieve an extremely low error rate;

(*d*) generate and receive certain supervisory or "red-tape" signals to direct the flow of data between terminals.

It is apparent that such a device will be needed for all types of terminal sending data over telephone circuits

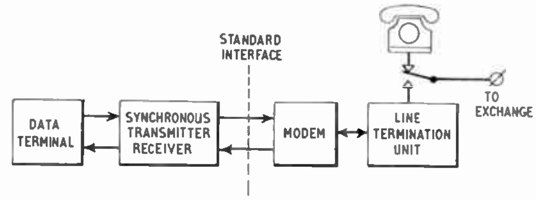


Fig. 7. Basic data transmission arrangement.

and so advantages accrue by adopting a common unit for all terminals. Figure 7 shows how a synchronous transmitter-receiver (s.t.r.) may be used with a range of data sources. In effect a new interface appears within the data terminal, corresponding to a single character data point and employing the normal binary-coded decimal code used in most data processing systems. Such an arrangement provides considerable system flexibility and enables the transmission of data between different types of data source and recording. Thus data read from a magnetic tape at one end of the circuit may be fed directly into a computer memory at the other end. Alternatively, data read from a paper tape can be transmitted directly to a magnetic tape, since each use a common design of s.t.r.

The s.t.r. itself is shown in schematic form in Fig. 8. It is designed to operate at transmission rates approaching the full capacity of the communications channel and incorporates checking facilities to maintain an extremely low undetected error rate. The maximum data speed is 4000 bauds.

The binary-coded decimal character from the data interface is first checked for validity via its odd parity check bit, and encoded to a 4-out-of-8 weighted code. This provides a more powerful self-checking code than the original single binary check bit and is itself further supported by a longitudinal check character generated in the s.t.r. and sent to line at the end of the data block. Seventy combinations are available from this

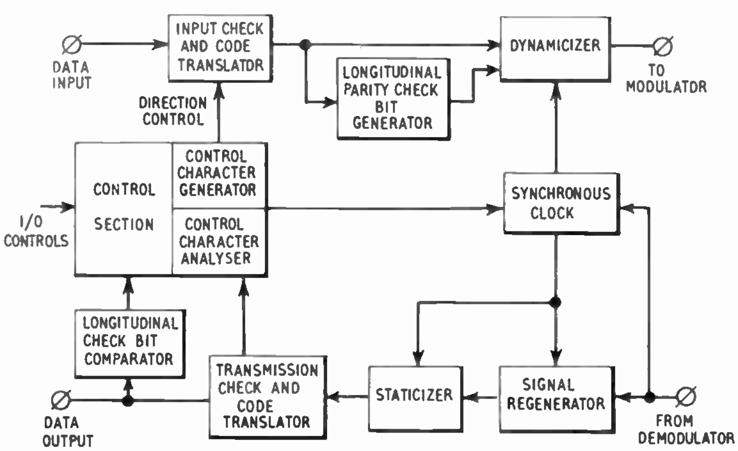


Fig. 8. Synchronous transmitter-receiver.

code, of which 64 are allocated to characters and 6 are used both singly and in pairs to provide 12 supervisory signals. From the encoder, the parallel character is dynamicized to the modem under the control of the clock and is again checked for 4-out-of-8 validity. In the receive sections, the serial data signal from the modem is staticized and checked for validity and converted back to its normal binary-coded decimal form. Simultaneously the longitudinal check character is developed from the received data and compared with the original, transmitted check character. Should an error in either a character or the block be detected, error signals are sent from the receiving terminal to the transmitting terminal which cause the block in error to be transmitted again automatically. Up to three attempts are made to retransmit the block in error, and if all are unsuccessful, then either an alarm is raised for operator attention, or that block is noted

Table 1

S.T.R. Operating Sequence—Half Duplex

Sending Station	Direction of Transmission	Receiving Station
Idle (1·3 seconds)	→	
End of Idle	→	
	←	Idle (1·3 seconds)
	←	End of Idle
(Start Input/Output Equipment)		
Idle	→	
Enquiry	→	
	←	Reply: ACK–2
Start of Record, S.O.R.–1	→	
Data Characters	→	
" "	→	
: :	: :	
End of Transmittal Record (E.O.T.R.)	→	
Longitudinal Check Character (L.C.C.)	→	
	←	Reply: ACK–1
S.O.R.–2	→	
Data Characters	→	
" "	→	
: :	: :	
E.O.T.R.	→	
L.C.C.	→	
	←	Reply: ACK–2

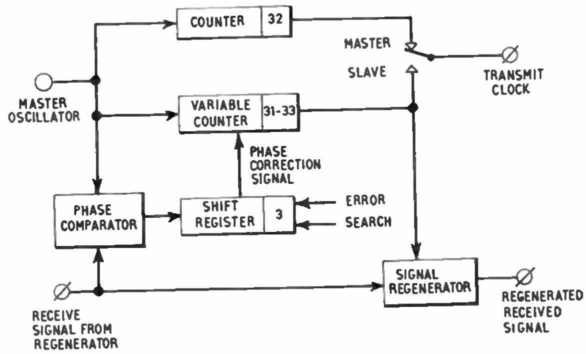


Fig. 9. Synchronous clock.

for attention later and the transmission of further data blocks is resumed.

The sequence of supervisory and control signals which pass between the data terminals and either precede or follow the data are shown in Table 1.

One of the most important circuits in the s.t.r. is the synchronous clock, since a stable and accurate clock is essential to achieve a low error rate. The schematic of the circuit is shown in Fig. 9. The receiving station clock is always phase-locked to the transmitting station clock, both being derived by counting down from a crystal oscillator, the counter radix providing a ready means of adjusting the data speed.

At the receiver the phase of the clock is adjusted with respect to the received clock by varying the counter radix from 30 to either 31 or 33 to provide either a phase advance or retardation respectively. The actual phase error is held in a three-position shift register and is not effective until three successive phase measurements agree on the direction. In this way the clock stability is unaffected by "jitter" in the received signals and by noise bursts of 10 seconds or more. A "search" mode allows rapid synchronization of the receive terminal when the first line signals are received, or following a circuit interruption.

6. Broadband Systems

So far the use of the public telephone network for data transmission has been discussed and it is clear that a considerable disparity exists between the data handling capacities of such circuits (250 characters per second) and the normal data rates employed in data processing systems. Typical magnetic tape speeds are from 15 000 to 62 500 characters per second.

Consequently in recent years attention has been directed to the use of broadband circuits based on the commonly occurring channel groupings of multichannel carrier telephone systems to match the data rates of computers and special data terminals have been developed to exploit these facilities. Figure 10

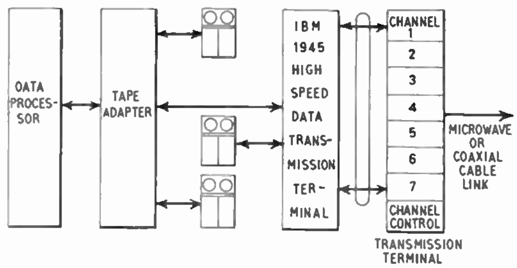


Fig. 10. High-speed data transmission terminal arrangement.

shows how a high-speed data transmission terminal is connected between a data processing system and a broadband circuit.

The terminal is connected into the data processing system by replacing a magnetic tape unit on the usual tape adapter, the displaced tape unit itself being connected back on to the terminal. With such a simple arrangement data can pass, under computer control, between magnetic tape units, directly between computer stores, or between store and magnetic tape.

Data flow is parallel by bit, serial by character using the standard data processing code and the transmission system provides a separate 96 kc/s path (equivalent to 24 telephone channels) for each bit of a character, so that a total of seven such channels are required, equivalent to 168 telephone circuits. An eighth 4-kc/s channel is used for supervisory and control purposes. Data speeds of up to 62 500 characters per second can be achieved with such a system and although it is expensive in communication bandwidth, nevertheless it involves only the addition of a simple unit to the data processing system itself.

One of the key design problems in such systems is to equalize the envelope delay distortion between all seven channels to within 6 microseconds, since any difference manifests itself as "skew" in the tape signal and could cause a failure in parity within a character. A typical delay through a single "channel" is 34 microseconds and occurs mainly in the filters.

7. Error Detection and Correction by Coding

The transmission medium used for data transmission is rather unreliable, and consequently precautions must be taken to ensure that the data sent through such systems arrives at the input to the processor at the receiving end with as few errors as present technology allows.

The precautions that are usually taken fall into two main categories:

(*a*) error detection by coding;

(*b*) error correction by coding;

and both are illustrated in Fig. 11.

In error detection systems redundant bits are added to the original message, which for convenience is usually grouped into discrete blocks of characters (or messages), and are chosen to represent the summation of the data bits in predetermined ways. At the receiving end these summations are repeated on the original data and compared with the received check bits. Should any discrepancies be detected, the receiving station informs the sending station and the block in error is sent once again, the process being repeated until it is correctly received.

In error correction systems the redundancy bits are related to the original data in a rather more sophisticated (and expensive) way so that not only can errors be detected at the receiving station, but by suitable examination of the data and redundant bits, the actual error can be pin-pointed and corrected without the need for retransmission.

It might appear from this brief comparison of the two methods that error correction systems are so much more powerful that the use of error detection seems hardly worthwhile. But a cursory examination of Fig. 11 will indicate the two major reasons which limit the application of error correction techniques. It will be seen that, in effect, the error correction system is a special network connected in tandem with the noisy transmission path, and so can be regarded as having the inverse error characteristics of that path. To be reliable, therefore, it must be designed with a full and complete knowledge of these error characteristics. This is a formidable task for a given network and an almost impossible task for a general purpose system. In practice the best that can be hoped for is to correct the majority of errors and to include a separate, simple error detection system for the uncorrected errors to initiate retransmission of the data. The second limitation is the cost of the error correction systems employed, although recent work on cyclic coding by Abramson[6] and others has greatly improved this position and has produced a class of error correcting codes capable of simple implementation.

The elegant simplicity of these cyclic block encoding and correcting methods can be seen from Fig. 12.

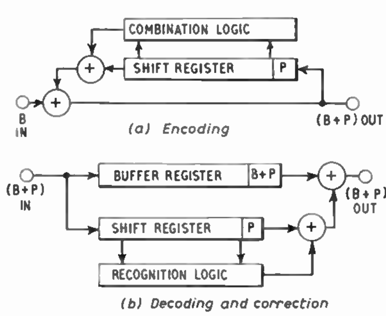


Fig. 11. Error correction and detection systems.

In the encoding process, data enter a P-stage shift register which in conjunction with the combination logic network assembles the P check bits and adds them to the original data bits. For decoding, the received message is held in a (B+P) stage buffer register, while a similar shift register-network combination, as in the encoder, checks its validity. Should any departures from the expected form occur, the combination produces a "1" in its output, which is added to the incident bit (assumed in error) of the received message and corrects it. The various cyclic codes which have been devised differ in the form of the combination logic and the number of check bits employed.

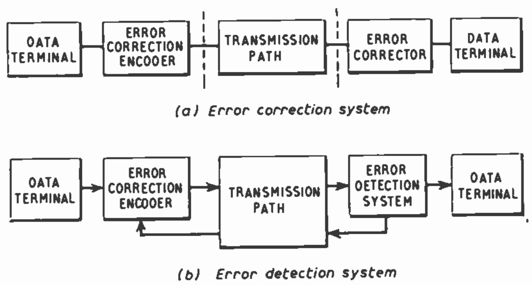


Fig. 12. Basic error correction methods.

Error detection with retransmission provides an excellent system if the transmission path introduces few errors and a good return-channel is available, but its performance deteriorates rapidly with increasing error rate. On the other hand error correction is very effective if the error characteristics of the transmission path are known. Systems will probably be developed in the future in which low cost error-correction systems correct most errors to convert a very bad physical channel to an effective channel having a relatively low error rate, which then allows error detection with retransmission to be used to its best advantage.

Alternatively, because of their critical importance, it may prove desirable to use error correction codes for the supervisory signals which signal a detected error, etc. Some idea of the sort of improvement which can be obtained by the use of error correcting codes over simple error detection can be obtained from the following data:

Over a typical telephone network, a block transmission system using error detection was producing errors which necessitated the retransmission of 5% of the blocks. By the introduction of an error correction code which increased the redundancy by 2½%, the retransmission rate could have been reduced to ½%—a net saving in transmission time of 2% at the added cost of the error correction equipment.

8. Future Developments

Having considered some of the recent developments in data transmission, it is as well to give some attention to future possible trends.

Perhaps the most rewarding future development will be the improvement of the telephone network itself. The present circuits have been designed almost exclusively for telephone traffic (although v.f. telegraph systems are used on certain selected carrier circuits) and could not be expected to provide a data transmission facility which did not evoke some desire for improvement. In fact, a good deal of time, effort and finance are devoted today to improving these circuits just for the transmission of speech. As the requirements for low-error, high-speed data transmission are better understood, so it is clear that, in the U.K. at least, these circuits will be greatly improved. The incentive today measured in terms of the ratio of data to regular telephone traffic is perhaps small, but a senior official of one American telephone administration has predicted that in 10 years time, 50% of their revenue from telephone traffic will come from the transmission of business data. Few people would accept his optimistic time estimate for the U.K., although they would agree with the concept. In fact, one could go further and predict that in 30–50 years time the basic network will be designed primarily for data, and speech be handled via pulse code modulation schemes. Not only should this lead to more efficient use of the available frequency spectrum but should simplify switching and routing of calls. We might even see a public data processing service integrated with such a communications network.

Specific improvements to the present network would be:

(*a*) the elimination of frequency shift, to allow vestigial sideband systems to take full advantage of the available spectrum, doubling of present speed being one possibility;

(*b*) an increase of this bandwidth to the full C.C.I.T.T. recommendation of 300–3400 c/s, to include local and toll circuits;

(*c*) a reduction in noise and circuit interruptions by improved switching systems.

Most data transmission engineers would agree that there is still a great deal of work to be done before the ideal modulation system is achieved and again much depends on the characteristics of the telephone circuits themselves. A C.C.I.T.T. channel having a usable spectrum of 3100 c/s has a theoretical data capacity of 6200 bauds, although the presence of in-channel signalling systems would probably restrict this to 5500 bauds. To date the best achieved by the author's Company over the public network is 2400 bauds, although elimination of frequency shifts should enable this speed to be nearly doubled. It is suggested that two-phase modulation is the best form of modulation available today, although some important work is being done on polyphase systems. These will lead to greater speed, but at the expense of circuit complexity and increased susceptibility to certain types of interference.

The incorporation of data transmission facilities has had a profound effect on the organization of data processing systems and, in effect, has added a whole new dimension to their configuration. With reliable data transmission facilities available, the designer of data processing systems can install the best type of processors at those points which have need of them and incorporate them all into a communications network designed to provide the desired processing function with the required availability and at minimum possible overall cost.

It is clearly not possible to embark on the installation of such complex systems without a careful and thorough study of the *total* systems organization to ensure a balance between the communications and data processing systems. This involves:

(*a*) the careful collection and reduction of traffic flow statistics to analyse the basic communications systems requirements;

(*b*) the use of simulation and like techniques to evaluate the systems performance at all stages of design to ensure adequate availability of the devised functions such as processing, permitted delays, etc.

For each study several professional specializations are involved and the need for the full and total integration of the data processing and data transmission equipment and the associated programming systems and associated procedures to produce a balanced system has led to the creation of a new profession, that of systems engineering. Systems engineers begin their professional careers as specialists in one particular field such as electronic design, communications engineering, programming or mathematical statistics, but they have to acquire quickly a full understanding of the other specializations so that a systems engineering group can operate as an effective and integrated team and produce the best possible system with function, availability and cost correctly balanced.

9. References

1. J. A. Pearce, "Data collection systems, their application and design", *J. Brit.I.R.E.* (To be published.)
2. K. L. Smith, J. Bowen and L. A. Joyce, "Telephone circuit evaluation for data transmission", *J. Brit.I.R.E.*, **24**, pp. 297–307, October 1962.
3. E. Hopner, "An experimental modulation-demodulation scheme for high-speed data transmission", *IBM J. Res. Devel.*, **3**, No. 1, pp. 74–84, January 1959.
4. E. Hopner, "Phase reversal data transmission system for switched and private telephone line applications", *IBM J. Res. Devel.*, **5**, No. 2, pp. 93–105, April 1961.
5. E. Hopner, "International Standards for Data Transmission", IBM Contribution No. 1 to Study Group A of the Data Transmission Sub-Committee of the C.C.I.T.T., June 21st, 1961.
6. N. M. Abramson, "A Class of Systematic Codes for Non-Independent Errors", Technical Report No. 51, Stanford Electronics Laboratories.
7. W. W. Peterson, "Binary codes for error control", *Proc. Natl Electronics Conf.*, **16**, pp. 15–21, October 1960.
8. W. W. Peterson and D. T. Brown, "Cyclic codes for error detection", *Proc. Inst. Radio Engrs*, **49**, pp. 228–35, January 1961.
9. P. Fire, "A Class of Multiple-Error Correcting Binary Codes for Non-Independent Errors", RSL-E2, Sylvania Reconnaissance Laboratory, California, March 1959.
10. C. M. Melas, "A new group of codes for correction of dependent errors in data transmission", *IBM J. Res. Devel.*, **4**, No. 1, pp. 58–65, January 1960.
11. W. R. Cowell and H. O. Burton, "Computer simulation of the use of group codes with retransmission on a Gilbert burst channel", *Commun. Electronics*, No. 58, pp. 577–85, January 1962.
12. J. J. Stone, "Multiple burst error correction", *Information and Control*, **4**, pp. 324–31, 1961.

Manuscript first received by the Institution on 15th February 1962, and in final form on 18th July 1962 (Paper No. 771/C16).

Points from the Discussion on this paper and others read at the Symposium on "Data Transmission" were published in the August issue of the *Journal*.

U.D.C. 534.88

The Future of Non-Military Sonar

An informal discussion in the final session of the Symposium on Sonar Systems, held at The University of Birmingham 9th–12th July, 1962

Chairman: Professor D. G. TUCKER, D.Sc. (Member) †

Professor D. G. Tucker: Up until now we have not specified, except in the session on fisheries, just what kind of application we have been dealing with. We are now, however, quite definitely discussing non-military sonar and I hope that this discussion will bring out some ideas as to the way sonar research and development should proceed for non-military applications, which are now rapidly increasing in importance. With a view to encouraging the discussion, I propose to say a few words by way of introduction and have listed a few headings (see Table).

The sort of applications that we might be concerned with are fisheries research, fisheries operations (that is, the actual operations of catching fish), hydrographic surveying, oceanographic research, navigation (of submarines and hovercraft and, I suppose, of ships too) and civil engineering. In case the last one is not self-evident, I should mention that we have reports of work done by civil engineers in studying the scour around bridge piers using sonar devices and it seems that there we have an interesting although not too easy field of application. Probably fisheries research, fisheries operations and hydrographic surveying are the most important from the point of view of sonar research.

Now for these various applications one wants to determine the direction in which system development and research has to go. What I shall say will amount really to a catalogue of headings and is based very largely on a discussion held a few weeks ago with people concerned with fisheries and oceanographical work.

By and large in non-military applications we are concerned with relatively short ranges. It is clear from some of the papers we have had during the symposium that in other fields of application very long ranges of 50 miles or more are of interest. I think the immediate concern of non-military applications is probably with ranges from a few feet up to a mile or so.

For fisheries research and possibly for other applications, one of the important lines of research and development is the question of "Higher resolution at longer and shorter ranges with electronic sector scanning". It does seem that, as Mr. Craig mentioned in his paper,[1] in fisheries research there is a need for a higher resolution device which will show the movements of individual fish and what we shall show you to-morrow at Belvide Reservoir is a first attempt at this; but it has a range of only 100 yards with an angular resolution of ½ deg covering with one pulse a sector of 30 deg. It will obviously be very useful if this sort of resolution—that is a 6-inch resolution roughly, could be obtained from ranges up to say a mile. You will appreciate, however, that there is quite a problem involved in this. It would mean maintaining a beam of width about 1/100th degree over that range. This is not quite the same kind of propagation problem as that which Mr. Weston talked about yesterday.[2] I think it has been fairly well agreed that these beams should ideally be pencil beams. (Up till now it is mostly fan beams that have been used in scanning systems.) There will be a considerable problem in obtaining a sufficiently high scanning speed for use with pencil beams on a raster basis.

Headings for Discussion

APPLICATIONS	RESEARCH
Fisheries research	1. Higher resolution at longer and shorter ranges, with electronic sector-scanning. Pencil beams for scanning systems. Acoustic camera.
Fisheries operations	2. Display and classification of target information to be improved. Understanding of echo-formation, portrayal of shape. Exploitation of frequency-response of targets.
Hydrographic surveying	3. System engineering—need for cheapness of equipment.
Oceanographic research	
Navigation (submarines and hovercraft)	
Civil engineering	

As regards shorter ranges, it has been suggested that there is a dead range for sonar and nobody seems to get below 50 ft or thereabouts; and it may be that more study has to be given to the possibility of scanning in the near field. This is a problem which has already been tackled in the field of ultrasonics in solids and I think it may have to come in sonar. The acoustic camera is another short-range device. There has been some work published of late on

† Electrical Engineering Department, University of Birmingham.

devices using acoustic image converters to form a picture in the same way as in a camera, and it may be that a further development of this idea would be worthwhile and that the fisheries and biological research people would find it valuable.

In hydrographic surveying there is definitely a use for a scanning echo sounder. At the moment very great precision is obtained in echo sounders using comparatively wide beams when the bottom is flat, but on steeply sloping bottoms these wide-beam systems fail rather badly. In fact we are quite convinced that an echo sounder using a within-pulse scanned highly-directional beam would give a very great advantage. In navigation, I believe the navigation of hovercraft is now a problem that is coming up and possibly a sonar system might serve the same sort of purpose there as the radio-altimeter in an aircraft, as used in the blind landing system, for example.

The second heading—"Display and classification of target information to be improved"—is rather vague perhaps. One of the things that might be included under that heading is the integration that we have been hearing about from Dr. Cooper[3]—pulse-to-pulse integration. There is probably the need for a three-dimensional display: work has been done on this over the years, but I fancy nobody has made one that has been thought really satisfactory. "Understanding of echo formation", "Portrayal of shape", are important topics for many applications which have had papers devoted to them in this symposium. They may be very important in the civil engineering application; they certainly should be in fisheries research.

"Exploitation of frequency response of targets" is a matter on which I feel some progress needs to be made. We at Birmingham are working on the development of a very-wide-band sonar covering a 10 to 1 frequency range in the hope of displaying the frequency response of targets. The critical requirement of this sort of wide-band system is to have a constant directivity. You can make a wide-band system using omni-directional transducers and you can make one using transducers whose directivity varies over the frequency range; but it does seem important that one should have a system with constant directivity. This would be particularly valuable for studying biological organisms in the sea, such as the deep scattering layer, where I understand that small fishes have resonances in the usual sonar frequency range. Constant directivity would be very valuable indeed since the limitation of most of the experimental work done up till now is the fact that it has had to be done with transducers whose directivity has varied with frequency. This makes the determination of frequency response of the targets very difficult unless one knows exactly where they are in the beam.

Work has been done, theoretical work I think mainly, on the frequency response of larger fish. Dr. Haslett has contributed to this and he might have something to say about it. The paper by Dr. Cushing and his colleagues[4] gave some limited experimental results. We are doing work ourselves but we are not quite sure what will come of it.

The third main heading—"Systems engineering, need for cheapness of equipment"—is a repetition of the point I made yesterday in connection with Mr. Craig's paper. It does seem that if these various non-military users (except perhaps fisheries research) are to be interested in more refined sonar, they will be more convinced by getting the sonar cheaply than by being told how much profit it will make them. You have to make better sonars for the same price as the present ones, and I think this does pose a very interesting problem.

Professor M. Federici †: There is an interesting instrument used in aerial navigation called the doppler navigator which gives an aircraft's course relative to the ground. One would think the same principle could be used from a ship using acoustic waves which would enable its course to be obtained relative to the bottom of the sea, i.e. relative to some immovable object. This would be independent of star observations, etc., and might be a new application for a special type of sonar.

Mr. R. Wyld ‡ said that the development of an acoustic doppler navigator would be valuable in providing speed measurement for hovercraft which, being amphibious, could not employ conventional techniques. A doppler navigator could also accurately measure hover height, an important factor in maintaining fore-and-aft attitude, and hence ensure that the stern would not drag in the water and increase operating costs.

Dr. L. Kay §: During the symposium much has been said about signal processing and the presentation of signals to the visual senses of an operator of equipment. Nothing, however, has been said about the use of the auditory senses which in the past played a most important role in sonar systems. With the advent of the chemical recorder the auditory presentation of signals was relegated to a secondary role in many systems, and as the duration of the transmission pulse became shorter and the use of c.r.t. displays became more common in echo-ranging equipments the use of the ears has been largely eliminated. In most systems, the information presented visually is superior to that which can be presented audibly, but it does not follow that this is always the case.

A feature of the auditory channel which is not normally exploited is its ability to analyse the frequency spectrum of a signal—not as an absolute measure but rather as a comparator in recognizing sound patterns. Work both in the United States of America and England has shown that frequency modulation echo-ranging systems can provide an auditory output which, for visual interpretation, requires a complex spectrum analyser and some form of display. What has not been established, however, are the relative merits of the visual and auditory channels in assimilating information. The advantage of the ear over the eye appears to be the ability to distinguish character in a signal. The classification of targets by their frequency response is not easily achieved by pulse systems and some advantage may be gained by using wide band frequency modulation systems with an auditory presentation, whereby the sound patterns from various objects may be recognized. At the same time an economy in equipment may be achieved.

† University of Milan.

‡ Vickers-Armstrongs (Engineers) Ltd., South Marston, Swindon, Wiltshire.

§ Electrical Engineering Department, University of Birmingham.

This brings me to the point of cheap equipment for the Fisheries Industry mentioned by the Chairman. Since fishermen will generally be free to use their ears, it would seem reasonable to employ frequency-modulation systems with an auditory presentation rather than expensive chemical recorders. Such systems would have a performance at least comparable with pulse systems—if not superior—since classification of objects may be more reliable by these means. There is evidence that frequency-modulation systems are good; bats emit sound waves which vary in frequency during the emission, and their interpretation of the returning echoes is remarkable as judged from their behaviour.

Mr. M. Schulkin †: Oceanographers and geophysicists have been slow in using sound to study the properties of the sea and the properties of the land underneath the sea. I would like to talk about a few of the techniques which could be used.

The first, which arises out of the paper which I gave,[5] would help oceanographers measure the salinity and chemical content of the water by using a salinometer to measure the conductivity of the water and an absorption meter to measure the magnesium sulphate content of the water. Then they would have the ratio of the four most important ions in sea water and with this they could probably learn a lot more about the identification of water masses and identification of the origin of water flowing into the sea, say along the coast. They would be able to trace water masses having slight differences in these four ions.

The sound speed meters that are now in existence could be used to measure and study fluctuations of the water elements in the sea. Turbulence, for example, could be studied if you had a way of studying fluctuations and this is a natural way of getting to know a lot about the spectra of turbulences by relating them to the spectra of the fluctuations.

Some work has been done and is continuing to be done on ocean wave spectra. Ordinarily we use wave staffs and the eye, or perhaps photographic measurements from the top of the sea, to learn something about wave heights, and this is very good; but sometimes they are very difficult to analyse. Dr. Marsh tells me that he feels he is able to determine the behaviour of the underside of the sea with regard to wave spectra from the back-scattering of sound. He says he is able to tell whether the power spectra of the water waves goes as ω^{-6} or ω^{-5}.

I feel that perhaps the reason we do not exploit listening sonar more is that we do not know enough about the characteristics of the sounds the fish make. Perhaps we are listening to the wrong frequency bands for them. For example, it is well known that porpoises give out sounds in various frequency bands up to 50 kc/s and higher. It would appear to me that if porpoises have a way of making these sounds, they have a way of detecting them too. Since porpoises, which are very large, use 50 kc/s sound, it is possible that smaller fish are actually using higher frequency sounds; these are subject to the greater attenuation of the sea and could escape easy detection. I think that passive sonar could be looked into further from the point of view of studying the characteristics of fish for fishery applications. In addition passive sonar can also be used in the geophysical application to locate underwater volcanoes. These could also be used as navigator beacons. I do not know how much people have thought about this but it is a natural extension of the listening technique.

Seismic methods have been used to find out about the sub-bottom and bottom but, in addition, there is a natural extension to underwater mining and oil prospecting which comes from the use of these methods.

Finally, extending the definition of sonar to wave periods of the order of 100 days, 100 weeks, and 100 years, Monk and Snodgrass at Scripps have been able to identify the origin of storms in the southern hemisphere and would be able to follow their motion and to detect arrival of waves, long ocean waves, coming from that direction.

Professor Tucker: Mr. Schulkin has mentioned a great many things. I wonder whether we could add to this list the determination of oceanic currents at given depths by measuring the doppler effect on the reverberation.

Mr. Schulkin: I believe the technique could be used for fast currents probably, but for the slower ones it is more doubtful.

Mr. D. Davies ‡: Much sonar work seems to be directed towards a study of the mathematical methods that are involved. It is impressive to see how these mathematical methods tie up with those used in other branches of physics. We ought to be aware of what contribution sonar studies could make to other fields and vice versa. The fields I am thinking of particularly are wave propagation in the ionosphere and radar; radio telescopes, of course, have been discussed here. Dr. Horton's mention of cubic arrays[6] obviously has analogies in X-ray crystallography and much of the mathematics that he used seemed to be closely connected with work done in that field. We might also mention optics, seismology and information theory. Acoustics, seismic prospecting and physical oceanography also seem to use very similar mathematical methods, and it is clearly to the gain of all the branches mentioned that they should maintain a keen interest in what is happening in their adjacent fields. Budden[7] has recently drawn attention to the fact that radio propagation investigators appear unaware of the work done in sound propagation and vice versa. It does seem that scientists need to get together all they can, and sonar is just one of those fields which can collaborate with many other branches of physics in which the problems are similar and in which the solutions will probably also be similar.

Dr. P. L. Stocklin §: We find a very strong lack of the realization on the part of those who are potential users and sponsors of the work that underwater acoustics is in fact a science all its own. We do not need to, and possibly we should not, borrow from such fields as electromagnetic theory and oceanography. Despite an appearance of

† Avco Corporation, Washington, D.C., U.S.A.

‡ Department of Geodesy & Geophysics, University of Cambridge.

§ Office of Naval Research, Washington, D.C., U.S.A.

multiple schizophrenia, it is difficult to deny that this symposium has been concerned with but a single science. Properly speaking, one would think it should be recognized as such, and have at least some formal recognition academically and otherwise recognized as a separate scientific field. A specific example of this is propagation. I have talked to people interested in radar about the extreme differences in the sorts of problems we face because of differences in the propagation of our different media. In many underwater propagation problems we experience changes of gradients, internal waves and boundary effects far greater than in corresponding radar problems. All this creates an extremely complex acoustic field quite atypical of radar conditions and for which usual radar analysis techniques are simply not useful. We should not neglect the purely scientific aspect of our work in looking for easy transfer of techniques from other fields.

Professor Tucker: I am a little disturbed that Dr. Stocklin thinks sonar should isolate itself at all. Perhaps I am overstressing the point he has made, but I think there would be a lot of benefit if sonar and radar and other branches of science and technology came together a bit more rather than separated themselves; but I do agree that it is not always appreciated that sonar has such a wide range of its own problems.

Mr. A. E. Crawford *: The science of Sonar has evolved from simple acoustics and this is primarily based on the conveyance of intelligence. It would appear that a field still requiring considerable development is the use of sonar principles for underwater communication. There are obvious military uses such as communication from one submarine to another, or submarine to surface ship.

There are also many non-military applications. With the increasing interest in skin diving a great many of the advantages are lost if telephone cables are required to be connected to the diver. Other uses include communications across harbours or rivers. A possible approach would be to use frequency modulation of the sonar signal and this could be done without undue complications of existing systems.

Lt. Cdr. R. H. Kerkhoven †: Perhaps acoustic analogue models should also be mentioned for solving radar problems. Several papers on this subject have been published lately, for instance, an acoustic model which was made to evaluate landing radar. There the problem cropped up of comparing acoustic scatterers with radar scatterers. Another application was where the interaction between elements of a radio array caused trouble in evaluating the individual elements. Although we have heard during the symposium of interaction between hydrophones in a sonar array, it seems that this is on a lesser scale.

Mr. P. G. Redgment: Earlier sessions of this symposium have dealt exclusively with systems operating in water, but the use of airborne sound has now been introduced in the suggestion for the measurement of hover height. If we admit such systems as coming within our subject, it seems to me we have opened another considerable field; that of devices for the aid of the blind.

Dr. L. Kay: It has been thought for many years that the principles of sonar systems should be adopted for the blind, but the difficulty in the past has been that of converting the information from the medium into a form suitable for feeding to the ears. This is why I raised the issue of frequency-modulation systems because this appears to be the only way of doing it. A guidance device based on these principles has been developed[8] and there is a fair amount of evidence now that it may be successful, in that blind people will accept this form of information.

Mr. J. O. Ackroyd ‡ suggested that some information about ultrasonics in medical diagnosis or on its destructive effect on tissue would be of interest.

Dr. A. Freedman § mentioned the use of echo equipment in which the beat of the heart was studied. A measure was obtained of the displacement of the heart walls to give information about certain medical conditions. It seemed to him that this was a field which would offer valuable scope for suitable equipment.

Dr. L. Kay: The medical aspects of ultrasonics have increased enormously. To give you some idea of the potentiality of this tool, it is now possible to obtain a cross-sectional picture of the eye with sufficient resolution to diagnose a cancerous growth between the retina and the retaining wall of the eyeball. A cross-sectional picture of the neck has also been obtained from which it was possible to diagnose an enlarged thyroid gland causing displacement of the sterno-mastoid muscle. Such pictures cannot be obtained with the use of x-rays. Ultrasonics is also used for surgical purposes. Lesions have been made in human brain to alleviate neurological disorders, and experiments are being carried out on animals to determine the full potentiality of this tool.

Dr. H. Maass ‖: One should not forget passive listening in the medical field; notably the intra-cardiac microphone. There has been made much progress in correlating certain murmurs in the heart with heart conditions such as mitral stenosis and septum defects. The microphone which has a diameter of only 2 mm is passed by means of a catheter into the heart to approach the spot where the murmur is created.

Dr. R. W. G. Haslett ¶: In work at the Western Infirmary, Glasgow,[9] the ultrasonic pulse method is used in a diagnostic instrument in gynaecology. Very interesting results have been obtained, particularly in the diagnosis of cysts, which are, of course, liquid-filled cavities. Also for example, the difference between twins and a single child may be seen readily. The instrument uses a special scanning technique and the display indicates a cross-

* Elliot Brothers (London) Ltd., Sonics Division, Reading, Berkshire.

† Royal Netherlands Navy, Oegstgeest, Holland.

‡ Admiralty Underwater Weapons Establishment, Portland, Dorset.

§ Admiralty Underwater Weapons Establishment, Portland, Dorset.

‖ Atlas-Werke A. G., Bremen, Germany.

¶ Kelvin Hughes Division, S. Smith & Sons (England) Ltd., Barkingside, Essex.

section through the body. A great advantage of this acoustic method is that the damage which may occur when using x-rays is avoided.

Mr. M. J. Daintith †: Dr. Stocklin and Mr. Davies have both implied that sonar is a major division of research. However, "sonar" as we normally use the term, refers essentially to an application, i.e. the use of equipment. Thus, in Professor Tucker's table, the items headed "Research" would be better termed "Development of instruments for research" (i.e. for research in the basic fields listed on the left). Dr. Stocklin suggested making sonar a discipline in its own right; I would support him if he had urged the formation of a Chair of Acoustics, say, but I do not feel that a Chair of Sonar would have any meaning; it would be like having a Chair of Voltmeters!

Mr. N. A. Anstey ‡: In seismic work we have applied pulse-compression techniques to echo-ranging. By pulse-compression techniques is meant the transmission of a very long signal and the subsequent cross-correlation of the received echo signal against the transmitted signal.

In seismic applications, this allows the replacement of a charge of dynamite (with a pulse duration of perhaps 100 microseconds) by a vibrator sweeping through a signal of perhaps 7 or 10 seconds in length. Then by subsequent cross-correlation the effective duration of the pulse is reduced to something of the order of 20 or 30 milliseconds. Our experience with this technique has been extremely good; it represents a practical as well as an elegant method, and it has a number of advantages. We have found also that the cross-correlation process, which looks rather fearsome on a theoretical evaluation, is in fact very easy in practice. It can be done quickly and continuously, in real time, with inexpensive and reliable equipment.

The pulse-compression technique has been reported in the literature in application to radar; it forms the basis of the "chirp" system used by the Bell Laboratories, and of some other radar systems using different types of signals. In some cases the pulse-compression of a swept-frequency signal is effected in a dispersive line; in other cases a random signal is used, which must then be compressed by cross-correlation. Certainly it is obvious that the radar people are very active in this field, and are realizing improvements in effective peak power of more than 100 times.

In the seismic applications we obtain improvements of much more than this—500 times or more. It is perhaps surprising, therefore, that in the literature on sonar, there are (to my knowledge) only two references[10, 11] to the application of random signals to sonar detection. This is despite the fact that these references make such application appear extremely attractive. They show, for instance, that a sample of sea-noise may be used as the transmitted signal, and that its amplitude can be adjusted so as to give "source search", that is, a search with the detected vessel having no knowledge that it has been detected. I had hoped that this had some application to fisheries, in that perhaps the fish did not like being detected; however, I now understand that they do not even know, so this is less attractive than I hoped!

Perhaps with this background I can provoke some of the experts on sonar to tell us what experience they have had in this field, and generally to describe their results.

Lt. Comdr. T. C. Line §: Could I make a plea to clear up the general matters of nomenclature and units? We have heard the same parameter called by six different names and the same variable measured in I think, five different units. In particular, absorption has been expressed in such units as decibels per kiloyard, decibels per nautical mile, nepers per metre, and so on. I do feel that some sort of standard in this matter would very much help the interchange of information between various countries and would simplify the subject generally.

Lastly, as a sea-going person, I would stress that all this sonar gear eventually must be made to work at sea by very unskilled people.

Dr. R. W. G. Haslett: Referring to Dr. Kay's remarks, although the human ear has not been mentioned in the formal sessions at this symposium, it is used in current equipments of various kinds, for example, fishing asdics, whale-finding equipments and deep-sea hydrographic echo-sounders.

Mr. Davies spoke of mathematical theory common to both electro-magnetic waves and acoustic waves. An example is the collection of mathematical tables due to Lowan *et al.*[12] for scattering by cylinders and spheres, but these only apply to ideal cases. In practical cases, for example, scattering by the different parts of fish,[13] it is difficult to apply the theory.

With regard to Mr. Anstey's comments, the pulse-compression method was suggested by Sproule[14] many years ago and I think one reason why it was not greatly developed for underwater devices was because the dispersive medium then available was rather large and expensive.[15] Another reason, is that, at any rate in hydrographic echo-sounding and in fish detection, it is possible to get adequate transmitted electrical power without causing cavitation[16] (e.g. 8 kW) so that pulse compression is not really an economic method to use.

Turning to the question of the echoes received from fish of various sizes at different frequencies, as mentioned in the discussion on the symposium paper by Cushing *et al.*,[4] some work on the frequency responses of fish has recently been published.[17] Looking to the future, if more were known of the frequency responses of different types of fish, it might be possible to determine fish size or, indeed, type of fish, from the received acoustic signals.

References

1. R. E. Craig, "The fisheries application of sonar", Sonar Systems Symposium paper, 1962.
2. D. E. Weston, "Propagation of sound in shallow water", Sonar Systems Symposium paper, 1962.
3. D. C. Cooper, "Visual display of integrated video waveforms", Sonar Systems Symposium paper, 1962.

† Admiralty Underwater Weapons Establishment, Portland, Dorset.

‡ Seismograph Services (England) Ltd., Keston, Kent.

§ Royal Naval College, Greenwich, London, S.E.10.

4. D. H. Cushing, F. R. H. Jones, R. B. Mitson, E. H. Ellis and G. Pearce, "Measurements of the target strength of fish", Sonar Systems Symposium paper, 1962.
5. M. Schulkin and H. W. March, "Absorption of sound in sea-water", Sonar Systems Symposium paper, 1962.
6. J. W. Horton, "Directional characteristics of volume arrays", Sonar Systems Symposium paper, 1962.
7. K. G. Budden, "The Wave-guide Mode Theory of Wave Propagation" (Logos Press, London, 1961).
8. L. Kay, "Auditory perception and its relation to ultrasonic blind guidance aids", *J. Brit.I.R.E.*, **24**, No. 4, pp. 309–17, October 1962.
9. Ian Donald, J. Macvicar and T. G. Brown, "Investigations on abdominal masses by pulsed ultra-sound", *Lancet*, **1958, i**, pp. 1188–95, 7th June.
10. J. L. Stewart and E. C. Westerfield, "A theory of active sonar detection", *Proc. Inst. Radio Engrs*, **47**, pp. 872–81, May 1959.
11. J. K. Parks and J. J. Downing, "System Considerations for Random-signal Sonars", Technical Report No. LMSD 895060, Lockheed Missiles and Space Division, December 1960.
12. A. N. Lowan, P. M. Morse, H. Feshbach and M. Lax, O.S.R.D., AMP Report No. 62 I.R., Sectn. 6.1–sr 1046–2032, U.S. Navy Department, Office of Research and Invention, Washington, U.S.A., February 1945.
13. R. W. G. Haslett, "Acoustic back-scattering from short translucent cylinders". (Paper in preparation.)
14. D. O. Sproule, U.K. Patent 604,429, 9th June, 1944.
15. W. Halliday, U.K. Patent 734,101, 6th March, 1953.
16. R. W. G. Haslett, "A high-speed echo-sounder recorder, having seabed lock", Sonar Systems Symposium paper, 1962.
17. R. W. G. Haslett, "Determination of the acoustic back-scattering patterns and cross sections of fish", *Brit. J. Appl. Phys.*, **13**, pp. 349–57, July 1962.

Summaries of Papers presented at the Symposium on Sonar Systems

PROPAGATION

Propagation of sound in shallow water

D. E. Weston (Admiralty Research Laboratory)

The way the medium controls sound reception in shallow water is discussed generally, and illustrated by some North Sea experiments with explosion sources. The ray and mode theories are introduced, and it is shown where it is convenient to change from one to the other. Transmission loss is the most important parameter, and is least around 200 c/s. Time dispersion is mainly due to the dispersion in the vertical arrival angle. Large fluctuations are observed when using a continuous wave source. The propagation is dominated by the characteristics of the two boundaries.

Absorption of sound in sea-water

M. Schulkin and H. W. Marsh (Avco Corporation, U.S.A.)

Laboratory and field data on the absorption of sound in sea water are reviewed in the light of modern theory, including effects of pressure, temperature and salinity. The effects of temperature and $MgSO_4$ salt concentration appear to be compatible from one investigator to another. There is some dispute, however, about the absolute value of the coefficient. The pressure effect has been measured for pure water, but has not been applied to the ocean as far as the authors know. The importance of this pressure effect on bottom loss measurements is shown to be sizeable. An expression is presented for sound absorption in sea water as a function of frequency, temperature, pressure and salt concentration.

Directional distribution of ambient sea noise

B. A. Becken, P. Rudnick and V. C. Anderson (Scripps Institution of Oceanography)

Observations obtained with a 32-element spherical array yielding 32 simultaneous beams covering all directions in three dimensions for depths of 100 to 1000 feet, in the octave 0·75 to 1·5 kc/s, are reviewed. A description is given of the DIMUS beam-forming system (clipping and shift-register delays) which is employed.

Some researches at O.R.L. on propagation using the techniques of ray acoustics†

R. E. Zindler (Pennsylvania State University)

Within the bounds of ray acoustics, exact equations governing intensity loss are derived for the sound velocity as a function of depth or depth and horizontal range. The solution is well behaved mathematically, if the velocity function is of class C_2, and can be extended indefinitely, provided the ray is not horizontal and the velocity gradient is not zero simultaneously. Exact solutions, specific examples, and approximate numerical solutions are used for interpretive purposes. It is concluded that consideration of gradients' gradient is required for "1 dB" computation of intensity and that the equations offer the potential of producing results as observed both from velocity microstructure and macrostructure, if appropriate oceanographic data are available.

The use of ray tracing in the study of underwater acoustic propagation†

M. J. Daintith (Admiralty Underwater Weapons Establishment)

A new formulation of the basic ray-tracing equations avoids some of the shortcomings of the standard methods. Its application to the usual situation of horizontally stratified refractive index is discussed and the design of an analogue calculator is described, with some practical examples. Extension to the general case where the refractive index varies throughout the medium is not difficult.

Prediction methods for sonar systems

R. J. Urick (U.S. Naval Ordnance Laboratory)

Performance prediction in sonar is centred around the so-called sonar equation relating various parameters defined by the medium, the target and the equipment. In this paper an

† These two papers will be published as a joint work.

active sonar equation valid for both short transient and long-pulse sonars is presented, and a discussion is given of each of the sonar parameters. Of particular interest is a generalized source level for comparing explosive and pulsed sonars, and prediction expressions for reverberation and detection threshold. An illustration of the use of these concepts is a recent determination of the back-scattering coefficient of the deep seabed using explosive sound sources.

Motion-induced seabed echo amplitude fluctuations

B. K. GAZEY (University of Birmingham)

Seabed echo-amplitude fluctuations result from a displacement of the transducers of a shipborne echo sounder relative to the sea-bottom. The origins of these fluctuations are discussed and analysis is given which determines the magnitude of the various components of a vessel's motion necessary to ensure that successive echoes are uncorrelated. It is shown that horizontal motion is of more importance than vertical motion and that randomizing displacements are small compared with the dimensions of the insonified area of the seabed. Finally, experimental evidence is presented in support of the theoretical conclusions, this being obtained using a scale model of an echo-sounding environment.

The effect of ice on long range underwater sound propagation

J. D. MACPHERSON (Naval Research Establishment, Canada)

A study has been made of the effect of ice cover on long range (100 n miles) acoustic propagation at low frequencies in an area having an approximate constant water depth of 30 fathoms. 1·8 lb explosive charges dropped from an aircraft were used as acoustic sources and the acoustic pulses were detected on a bottomed barium titanate hydrophone. Comparative measurements have been made with identical water velocity structures with no ice and with a variety of different ice conditions. Large variations of attenuation have been observed depending upon the type of ice prevalent in the area.

SURVEY EQUIPMENT

Profile and arca echograph for surveying and location of obstacles in waterways

S. FAHRENTHOLZ (Echo Sounder Factory, Kiel, Germany)

The survey ship carries outriggers floating athwartships which are fitted with sets of ultrasonic transducers (up to 50 receivers). The transducers are switched to avoid picking up spurious echoes. The record is made on a wide paper chart.

Reflection seismic methods for exploring the sediments and crust of the earth beneath the ocean

J. B. HERSEY, S. T. KNOTT, D. D. CAULFIELD, H. E. EDGERTON and E. E. HAYS (Woods Hole Oceanographic Institution)

The continuous seismic profiler has been developed at Woods Hole to measure the shape and acoustic properties of geological structures in water-covered areas by the reflection and refraction methods. Its operation at sea and the recording techniques are similar to those of echo sounding; its source and receiving equipment are more similar to those of commercial seismic exploration. Several variants of this method of seismic observation have been used by different workers in submarine geophysics.

The "thumper" system used by N.I.O.

R. BOWERS (National Institute of Oceanography)

In order to obtain a clean half cycle pressure pulse from the thumper transducer it has been mounted in a fibre glass dinghy which is towed 75 yards behind the ship. A hydrophone is towed between the ship and the dinghy. This hydrophone is directional and has a null in the direction of the ship and the dinghy. This reduced background noise to such a level that surveys can be carried out at 8 knots. The record is displayed on a Mufax recorder. With this system sub-bottoms separated by 16 feet can be discriminated.

SONAR FOR FISHERIES

The fisheries application of sonar

R. E. CRAIG (Department of Agriculture and Fisheries for Scotland)

Sonar has proved invaluable in pelagic fisheries as a direct detector of fish. In demersal fisheries the value of sonar except in Arctic cod fisheries is still mainly indirect—its chief value is for navigation and determination of the type of seabed. Real advance seems to depend on methods which will improve definition and interpretability really substantially—to the stage of giving some sort of recognizable picture of the undersea scene.

Measurements of the target strength of fish

D. H. CUSHING, F. R. HARDEN JONES and R. B. MITSON (Fisheries Laboratory, Lowestoft), and G. H. ELLIS and G. PEARCE (Kelvin Hughes Division of S. Smith & Sons)

Measurements have been made at 30 kc/s of the target strength of dead cod, herring, plaice and perch. Four independent series of measurements have been made between 1955 and 1959. In three of the series, false Onazote swimbladders were placed in the body of the fish to simulate the acoustic properties of the normal swimbladder. The results are tabulated and presented in a graph, the reference level being the target strength of a sphere of 2 m radius. The results are discussed in relation to the frequencies which are most likely to be suitable for echo-sounding on fish targets when quantitative data are required.

A new sonar system for marine research purposes

T. S. GERHARDSEN (Simonsen & Mustad, Norway)

A complete "search installation" comprises two combined sounder/sonar sets working on 11 kc/s and 30 kc/s and two additional sounders on 38 kc/s and 18 kc/s. The remote-controlled sonar transducers are housed in a retractable, 18 knots dome. The transmitters have variable pulse length and an output power of 8·5 kW. The receiving equipment is calibrated and comprises special circuits for time-variable gain, automatic gain control and "white line" features. Display and recording equipment includes range/depth recorders, tape recorder, loudspeaker and an oscilloscope for detailed echo studies and playback of tape recordings. The whole installation, including the two additional sounders, is manned by a single operator.

Cathode-ray tube displays for fish detection on trawlers

P. R. Hopkin (Kelvin Hughes Division of S. Smith & Sons)

Current echo sounder recorder and cathode-ray tube display techniques are reviewed and their shortcomings stated. An improved cathode-ray tube display is described in which the seabed echo always occurs in the same position on the screen irrespective of changes in range due to the contour of the seabed or vertical motion of the trawler. A magnetic recording drum with out-of-contact heads stores the echoes from the range adjacent to the seabed and enables them to be displayed repeatedly during each sounding period.

A high-speed echo-sounder recorder having seabed lock

R. W. G. Haslett (Kelvin Hughes Division of S. Smith & Sons)

Following the associated paper by Hopkin (concerning an expanded seabed lock c.r.t. display for trawlers), a complementary electro-mechanical recorder giving an intensity-modulated record on dry paper, is described. Both displays indicate the presence of fish within $1\frac{1}{2}$ fathoms of the seabed (i.e. under the headline), but the recorder gives a permanent record and improved trace-to-trace correlation.

Experiments were made with several possible types of triggered recorder to determine the best design, also with various signal delays required for seabed lock. Details are given of the moving-iron recorder, finally adopted, which displays a range of 4 fathoms with high accuracy across paper 6 in. wide. The magnetic-drum storage system, described in the associated paper, is common to both displays. These displays form part of a comprehensive fish-detection equipment of improved acoustic performance for use on trawlers.

A high-resolution electronic sector-scanning sonar

V. G. Welsby and J. R. Dunn (University of Birmingham)

The success of sea-trials of an experimental pulsed sonar system using within-pulse electronic scanning has led to the development of a new version using a considerably larger number of channels and a shorter wavelength in the water. The new system, which has 30 channels and works at a carrier frequency of 500 kc/s in the water, is designed for a maximum range of 100 yards and a scanned sector of $\pm$ 15 deg. Pulse lengths down to 100 μs (i.e. scanning speeds up to 10 000 per second) can be used, corresponding to a range resolution of about 6 inches. Facilities are provided to permit a rapid change-over from normal additive signal processing to a multiplicative process which is capable of roughly doubling the angular resolution of the system, at the expense of some reduction in effective maximum range. One of the objects in building the set was to confirm certain theoretical predictions about the maximum range resolution, bearing resolution and scanning rate which can be obtained with this kind of system. The equipment is nevertheless destined for practical work in connection with the study of fish movements in a hydro-electric reservoir and for other biological investigations.

ARRAYS AND SIGNAL PROCESSING

Space-time sampling and likelihood ratio processing

P. L. Stocklin (Office of Naval Research, U.S.A.)

Space-time sampling plans based on a three-dimensional Whittaker sampling function are reviewed for single frequency and series band-limited time limited acoustic pressure fields. Following a brief discussion of probabilistic wave fields, likelihood ratio measurement procedure is outlined. Two examples are treated: the one-dimensional resolution problem for Gaussian noise, and the space-time likelihood ratio processor for signal known exactly in Gauss and Markov-Gauss noise.

Visual detection in intensity-modulated displays

J. W. R. Griffiths and N. S. Nagaraja (University of Birmingham)

In a two-dimensional intensity-modulated display such as the plan position indicator used in radar systems, the observer is looking for and recognizing the signal as an area, or pattern, of brightness differing from that of the surround. By using a closed circuit television system it was possible to reproduce the essential features of this type of display whilst having complete control over the important parameters—signal area, background noise, target presentation time etc. In particular, the background can be changed from being uniformly illuminated, i.e. the situation studied by many psycho-physiologists, to the more realistic situation appertaining to radar displays, i.e. when the background is completely perturbated by noise. Threshold signal/noise ratios have been measured for a number of such conditions and compared with those of an equivalent theoretical model. The results suggest that the visual detection system is a sub-optimum one and its efficiency is dependent, among other things, on the area of the signal. This point goes some way to explain some previously observed discrepancies between experimental and theoretical rates of improvement with increase of area.

Investigations of an interaction anomaly between sound projectors mounted in an array

J. S. M. Rusby (Admiralty Research Laboratory)

Measurements on acoustically small, low loss projectors in arrays in water have shown that near resonance violent changes in their electrical admittance take place. This behaviour has been investigated by computing theoretical admittance diagrams for a cruciform array and comparing these with the anomalous measured diagrams. The shapes of the two sets of diagrams are in agreement so that it can be shown from the analysis that the behaviour is caused by the dominant mutual radiation impedance terms. Near resonance, where the mechanical impedance of the projectors is low, these terms control the phase and amplitude of motion through the water coupling.

Optimum processing for acoustic arrays

W. Vanderkulk (Scripps Institution of Oceanography)

Bryn's optimum detector is briefly described. Its performance is compared with the conventional detector for background noise consisting of self-noise and isotropic sea-noise.

The detection of sonar echoes in reverberation and noise

J. O. Ackroyd (Admiralty Underwater Weapons Establishment)

A method is presented for calculating the frequency spectrum of reverberation in a shipborne pulse sonar and hence its effect in masking an echo which exhibits doppler. Signal processing is considered, as affecting the masking due to reverberation and noise combined. This enables the relative merits of sonars, or of proposed modifications, to be assessed more realistically than by considering noise alone.

The directional discrimination of volume arrays

J. W. Horton (U.S. Navy Underwater Sound Laboratory)

The directivity factor of an array of sonar transducers distributed throughout a cubical volume has been found to vary inversely as the square of the element spacing, as does the factor for the elements occupying one face of the cube. Arrays occupying the volume of a sphere, its surface, and its plane projection show directivity factors which are not inversely proportional to the number of elements but which are nearly equal. At spacings for which these arrays have useful directivity patterns there is negligible advantage in using a volume array in an isotropic noise field.

The angular resolution of a receiving aperture in the absence of noise

V. G. Welsby (University of Birmingham)

Much has been published on the performance of particular directional receiving arrays and their associated signal processing arrangements. This paper is concerned with a more general approach to the problem and discusses certain physical limitations to the kind of information which can be obtained about a completely unknown far-field distribution, even if noise is assumed to be absent. It is shown, for example, that the "fineness of detail" which can be discerned by means of a given aperture is determined solely by the dimensions of the aperture in terms of the half-wavelength at a frequency corresponding to half of the total frequency band-width of the signal waveform. Additive processing of the outputs of the elements of a spatial array can be shown to give the best result when the field is time-stationary but unknown. More complicated processes, such as multiplication (time-average-product or space-average-product techniques), have advantages when the field is either non-stationary in the time domain or when prior information about its form is available (e.g. knowledge that it is due to a limited number of "point" sources rather than a continuous source distribution. It is shown that, for a field at a single frequency W c/s, a multiplicative ("correlation") process provides the same fineness of details of the far-field as that obtainable with an additive array of twice the length, provided the field has a time-stationary amplitude distribution and a relatively rapidly time-varying phase distribution. Furthermore, provided the field is of this type, comparable performance can be obtained by retaining only two receiving elements, situated at the extremities of the aperture, and using a "multi-frequency" signal waveform whose spectrum occupies a total bandwidth 2 W c/s.

Synthesis of multi-element directional patterns with a two-element single-frequency receiving array

E. D. R. Shearman (University of Birmingham)

The principle of space-frequency equivalence has been shown to permit synthesis of multi-element directional patterns with only two physical elements by the use of a multi-frequency carrier. Disadvantages of the system are the use of more spectrum space and the need for a special transmission. It is shown that multi-element patterns can be synthesized from the outputs of two elements at a single frequency. By the use of frequency multipliers at the element output, multi-frequency carriers are produced synthetically at the receiver. Penalties are paid in signal/noise ratio and multi-source resolution relative to a physical multi-element array.

Multiplicative arrays in radio-astronomy and sonar systems

D. G. Tucker (University of Birmingham)

The paper is concerned with gathering together, and extending, some ideas on multiplicative receiving arrays as applied to radioastronomy and sonar. The main emphasis is on systems using coherent-tone signals and designed for direction-determination. It is shown that, using arrays in which the outputs from the two halves are multiplied together, the accuracy of determination of target bearing may be increased by a factor of $\sqrt{2}$ up to 2 as compared with ordinary arrays of the same size; split-beam systems have no theoretical advantage over full-beam systems. Using arrays in which the two portions have the same centre-point, some new and attractive directional patterns are developed (including pencil beams) in which secondary responses may be almost eliminated over a considerable range of bearing.

The improvement of detection and precision capabilities of sonar systems

Maurizio Federici (Politecnico di Milano)

The paper surveys recent advances in signal processing techniques in sonar, indicating trends in the use of diversified types of pulses, correlation and multiplication techniques and instrumental evaluation of signals.

Visual displays of integrated video waveforms

D. C. Cooper (University of Birmingham)

The operation of single loop and double loop video integrators is discussed briefly and a qualitative measure of integrator performance is given in terms of a simple video signal/noise ratio. A description is given of a set of pulse detectability experiments which have been conducted using a single trace display of the integrator output. A "peak level" criterion of detection was used in these experiments and the results obtained show that the simple video signal/noise ratio does not give a direct measure of integrator performance. It is concluded that the performance of a video integrator is somewhat better than would be predicted by the use of the simple theory, providing that the pre-detector signal/noise ratio is greater than -5 dB. Finally a set of theoretical results are given for comparison with the experimental ones.

The portrayal of body shape by a sonar or radar system

A. Freedman (Admiralty Underwater Weapons Establishment)

When acoustic waves are incident upon a rigid body or electromagnetic waves are incident upon a perfectly conducting body, the amplitude of the scattered radiation usually varies in a very complicated manner with direction of incidence, with direction of observation, and with frequency. Even if consideration is restricted to back-scattering, the resulting scattering directivity curves appear to offer no clue to the shape of the scattering body. Yet the visual identification of objects by scattered radiation is an everyday experience. This paper endeavours to clarify the mechanism whereby the eye perceives shape and to see how this is applicable to methods of display in sonar and radar equipment. The analysis, which is qualitative, is based upon a recent theory of the mechanism of scattering of short wavelength acoustic radiation by a rigid body immersed in a fluid medium, and its electromagnetic counterpart.

Publication of papers read at the Symposium on Sonar Systems will begin in the December issue of J.Brit.I.R.E. *All the papers will be collected together and issued as soon as possible as "Proceedings of the Symposium on Sonar Systems" irrespective of whether the paper has already appeared in the* Journal.

Trends in Aviation Electronics

The 1962 Exhibition and Flying Display of the Society of British Aircraft Constructors, held at Farnborough, Hampshire, between 3rd and 9th September was announced as being the last for two years. As far as the introduction of new aircraft was concerned, the Exhibition certainly achieved enough to satisfy the visitor until 1964. In the field of radio and electronics, however, the picture was mainly of steady development of equipment already presented in previous years and comparatively few innovations were apparent. This is not necessarily a sign that the British radio and electronics industry is resting on its laurels; the introduction of such important new electronic techniques as automatic landing and improved navigation and air traffic control equipment into practical form must necessarily be a slow business.

The trend towards increasing automation has become more evident. The objective is to reduce the load on the human operator, whether he is the pilot of an aircraft or an air traffic control officer, by relieving him of routine actions which can be performed as well, if not better, by electronics. Automatic all-weather landing is probably the most familiar—and most difficult—of the problems being worked on. Automation also means that test equipment must be designed for rapid system checks on the ground, and in-flight "confidence" check facilities be provided.

An outstanding event of the display was the public appearance of the Vickers *V.C.10* and de Havilland *Trident* aircraft; these are the first civilian aircraft in the world designed specifically to utilize fully automatic landing equipment and certification for its use is expected by 1966. It was incidentally stated that the *V.C.10*, which contains several tons of flight test instruments, also carries three closed circuit television cameras. One is mounted on the tailplane to view the engines, the second is buried in an engine pod to observe the tail fin, and a third, fitted with a steerable periscope and remote focus, inspects the under surfaces of the aircraft. The pictures are viewed on a monitor installed in the cabin.

The Ferranti moving map display presents a pilot with a continuous information of present position and track on an optically-projected moving map. As the aircraft moves over the earth's surface, a map moving under a fixed marker which indicates present position, is displayed on a special glass screen. Aircraft track is shown by a line on the screen. Airways maps are reproduced on 35 mm colour film which is projected on to the glass screen at the front of the unit. The film strip is mechanically driven from an external navigation computer in such a way that the aircraft position corresponds to the point on the map under the present position marker. The moving map display can be used with any system of navigation which provides outputs in a suitable form such as doppler or inertial navigators.

A new instrument system was shown for the first time at this year's Show, namely, the Smith's taxi-speed indicating system which has been developed to give a reliable indication of ground speed to pilots of large aircraft, where an accurate estimation of speed, particularly under poor visibility conditions, is difficult. A Hall-effect detector is mounted on the fixed part of the aircraft undercarriage wheel assembly and a pair of small permanent magnets attached at diametrically opposite points on the wheel rim. The nominal gap between the detector and the magnets is about 0·1 in. Rotation of the wheel generates pulses of current in the detector and these pulses are fed to a pulse shaping and integrating circuit. The output current is proportional to the pulse recurrence frequency, and therefore to the aircraft wheel rotational speed. The indicator is a standard moving coil instrument scaled 10 to 50 knots, based on the mean effective rolling radius of the aircraft wheel.

The Marconi transistorized tabular display finds application in a wide variety of data-handling problems, including the display of flight data to air traffic controllers as well as industrial and commercial applications. The complete tabular display equipment is designed for use in data processing systems where the information is derived from a digital computer or processer, or from a keyboard. Typically, this information is stored and then scanned sequentially to control an electronic character generator which provides the waveforms required to produce alpha-numeric symbols on the display unit. As the equipment is capable of accepting a modification of a single character without affecting the remainder of the characters, the minimum of computer loading is achieved.

The display consists of a number of alpha-numeric characters arranged in a series of horizontal lines across a cathode-ray tube face. The individual character is written by continuous movement of the electron beam spot through the intersections of a 4 unit × 4 unit grid or matrix and legibility is extremely good. The equipment can write 50 000 characters per second. The cathode-ray tube used has an after-glow characteristic which retains a steady picture if repeated 10 times per second; the actual rate provided by the timing circuits is approximately 12 times per second, and the store capacity of 4096 words is arranged to match this rate. This storage capacity is sufficient to feed a number of individual operator positions (e.g. 10 displays each using 400 characters may be used, all controlled from one central position). Any one display can have up to 16 lines of information, with up to 60 characters on a line; under these conditions the presentation is clearly readable at a distance up to 3 ft.