WIRELESS ENGINEER

Vol. 32 SEPTEMBER 1955 No. 9

Radio-Astronomy

As a direct outcome of the spectacular advances made in radio technique during the 1939-45 war, there has been created and developed in the past decade a new science which has now received the well-established name of radio-astronomy. In this science, radio-frequency reception technique is used to explore radiation emanating from various sources in the universe. While it was regarded initially, perhaps, as a supplementary aid to astronomy by normal optical means, the advances and discoveries which have already been made are so spectacular as to leave no doubt of the fact that radio astronomy is now firmly established as a science in its own right.

It has long been an accepted fact that the electromagnetic radiation from a so-called 'black' body would extend over a very wide range of frequencies or wavelengths, the amount and distribution of radiation depending upon the temperature of the source and the absorbing characteristics of the medium through which it travels towards the receiver. More than fifty years ago, Professor S. P. Langley measured the energy of the sun's radiation through the visible spectrum and to a considerable distance on either side. After various unsuccessful attempts to extend this measurement into the radio-frequency spectrum, the reception of 'cosmic noise' on a wavelength of 187 cm (frequency 160 Mc/s) was reported by G. Reber in 1940, and later established as emanating from the sun; while some nine years earlier K. G. Jansky had discovered radiation coming from the direction of the Milky Way on frequencies in the range 10–20 Mc/s. Also British radio amateurs, notably D. W. Heightman, had noticed the existence of a curious hiss in their receivers on wavelengths of about 10 metres; they found this to occur in the daytime only and to be associated with periods of solar activity. In 1945, G. C. Southworth described the measurement of radiation from the sun on wavelengths between 1 and 10 cm (frequencies between 3 and 30 Gc/s*); and he showed that the power flux received, which was of the order of 10^{-20} watts per square metre per cycle per second, was consistent with the Rayleigh-Jeans relationship for the distribution of radiation from a sun at a temperature of about 20,000° K, or about three times the optically-observed temperature.

During the war years, radiation from the sun was observed as a source of interference in the operation of British Army radar sets on wavelengths between 4 and 6 metres; and Sir Edward Appleton and J. S. Hey described, in 1946, the considerable increase in solar 'noise' on these wavelengths which accompanied the passage of sunspots over the sun's visible hemisphere. At about the same time, similar observations on the sun were being made by J. L. Pawsey and his colleagues at Sydney, Australia, using a wavelength of 1·5 metres (200 Mc/s).

From this starting point centres of research in radio-astronomy have been established at Cambridge and Manchester Universities in this country, in Australia, and in the United States of America, Canada, France and Belgium. The results of the work already achieved have not only added considerably to our knowledge of the universe and the various sources of radiation in it, but they are also of considerable practical importance

* 1 Gc/s = 1 Gigacycle per second = 1 kMc/s = 1000 Mc/s.

in many instances. The radio astronomer has, for example, provided a means of exploring the characteristics of the ionosphere by using radiation coming to us from outside the earth's atmosphere, in contrast with the usual method of sounding the ionosphere by emitting and receiving pulses of radio waves at the earth's surface. A tremendous advance has also been made in the study of meteors which can now be detected by day or night under all weather conditions, and on a scale which could never be approached in optical observations. Such discoveries, together with the advances in apparatus technique, which have made them possible, are of considerable importance in the future of radio communications.

The Radio Telescope

The first major advance made in this new science was the development of the 'radio telescope' as an instrument which could be used to determine the direction of arrival of these radio waves which originate outside the terrestrial atmosphere. The simplest type of telescope consists of a concave mirror with a receiving aerial at its focus; but bearing in mind that the wavelength of light is of the order of 5×10^{-5} cm, it will be appreciated that to have the same directional resolving power as an astronomical telescope, the corresponding radio instrument would need to be a million times the diameter, for a wavelength of 50 cm. In spite of this limitation, a great deal of useful work has been done already in various countries with the steerable paraboloid type of aerial with diameters up to 75 ft. The greatest advance in this direction has been achieved at the Jodrell Bank Experimental Station of the University of Manchester, by a team of scientists under the leadership of Professor A. C. B. Lovell. After gaining experience with a semi-fixed telescope of 220-ft aperture, which is normally directed vertically upwards, but can be displaced $\pm 14°$ from the central position, a completely steerable radio telescope, with a paraboloidal reflector 250 ft in diameter, is now in an advanced stage of construction.

An alternative technique has been adopted, notably at Cambridge in this country and at Sydney in Australia, in which an array of aerials with appropriate reflectors is installed along a fixed base-line forming a linear type of interferometer with a narrow scanning beam in either vertical or horizontal direction as may be required. The rotation of the earth is used to sweep this beam across the sky while the radiation from active sources is received and recorded. Such interferometers have been constructed and used by M. Ryle at Cambridge for certain frequencies in the range 80 to 200 Mc/s, while W. N. Christiansen in Australia has developed and constructed an array of 32 parabolic reflectors along a base-line of nearly 720 feet, to provide a resolution of about one minute of arc on a wavelength of 21 centimetres (1,420 Mc/s). More recently, B. Y. Mills and his associates have constructed a combination of two arrays, of an overall length of 1,500 ft and which cross at right angles, giving a cone of reception of about 45 minutes of arc at a wavelength of 3·5 m (86 Mc/s).

Concurrently with this development of aerial systems, various novel receiving devices have been produced which enable the wanted radiation to be detected and recorded in the presence of a general background of noise radiation from extra-terrestrial sources. It has so far usually been possible by a judicious choice of frequencies and direction of observation to avoid serious interference from man-made radio sources, including those used for communication and navigation purposes: but a constant watch is necessary in order to secure the preservation of suitable spaces in the radio-frequency spectrum, for the radio astronomer to continue the good work that he has already accomplished.

Radio Observations on the Sun

The amount of radiation received from the sun at radio frequencies has already been found to vary markedly with the actual frequency or wavelength used and also with the state of quietness or disturbed conditions on the sun itself. At centimetre wavelengths, the radio energy received from the sun in a quiet condition is found to correspond to an apparent temperature of two or three times that corresponding to the value for optical measurements. As the wavelength is increased, however, the radiation corresponds to a rapidly increasing apparent temperature of the sun's disk. For example, at wavelengths of one metre or above (300 Mc/s or below) the flux of energy received corresponds to an apparent solar temperature of about one million degrees absolute; furthermore, it is found that the 'radio' sun is somewhat larger than the 'optical' sun and that, unlike the optical radiation, the emission at radio frequencies falls off gradually near the edge. These observations are consistent with the assumption that the main source of radiation at the lower radio frequencies is the solar corona which, from other reasoning, is supposed to have a temperature of the order of one million degrees and to be completely ionized. Radio-astronomy has thus provided us with a new and powerful tool for studying solar corona; and it has already demonstrated that even under undisturbed conditions the temperature over the area of the sun's disk is by no means uniform.

While we thus have a reasonably good theory

for the emission of radio waves from the quiet sun, it is to be noted that there is no correspondingly simple explanation of the greatly enhanced radiation that accompanies the presence of sunspots travelling across the sun's disk. Increases in the 'noise' level of the sun's radiation by a factor of one thousand over that of the quiet sun have already been observed. More than one type of such increase in radiation has been recognized, and while these outbursts are generally associated with visible sunspots, the presence of the latter is not always accompanied by solar noise storms.

This phase of radio-astronomy is of great importance to the radio scientist and engineer, in view of the established close connection between sunspots and the ionospheric conditions which determine the efficiency of radio communication over great distances. It is possible that future research may establish a radio-noise index of solar activity, which will be found more useful and reliable in the long-term forecasting of high-frequency radio transmission conditions involving the ionosphere.

Radio Stars

In a different application, the development of radio-astronomy has already indicated the existence of a large number of discrete sources of radiation distributed throughout the sky, and the emission of which is superimposed on a relatively continuous background radiation. The distribution of intensity in this background follows closely the contours of the galaxy or Milky Way. In this field, the interferometer has proved invaluable in locating the discrete sources or radio stars as they are termed; and the most striking fact discovered so far is that very few of the conspicuous stars and nebulae of our Milky Way system are among the recognized objects in the radio sky. Conversely many of the radio stars, of which nearly 2,000 have now been identified and catalogued, have no obvious and easily-recognized counterpart in the optical astronomers' observations. The strongest observed radio star, in the constellation of Cassiopeia, has a position in the sky which is far from conspicuous on the best optical astronomical photographs; although the intensity at metre wavelengths corresponds to a flux of about 10^{-22} watts/m^2 per cycle per second, or about one per cent of that received from the quiet sun at centimetre wavelengths.

This example serves to illustrate the difficulty in accounting for the existence and nature of these radio stars. It has been estimated, for example, that if all the visible stars emitted radio waves like the quiet sun, the total radiation would only be about 10^{-8} of that observed from the galaxy. If they all emitted like the sun when it is most disturbed, the total radiation would still be less than one per cent of that recorded by the radio astronomer. This experience of the intense emission received from the radio stars forms a surprising inversion of the familiar ratio of sunlight to starlight, and emphasizes our lack of understanding of the fundamental mechanism of the emission of radiation from the galaxy itself. While no firm theory has yet been advanced to explain the observations, this work is of direct practical application to radio communication. For example, the second most intense source in the sky is the radio star in Cygnus, which at times gives a very steady radiation and at other times displays fluctuations of intensity or 'twinkling'. This twinkling is due to irregular diffraction of the incoming radiation in passing through the upper regions of the ionosphere. By the simultaneous observation of these irregularities at three receivers, the direction and speed of movements or 'winds' in the F region of the ionosphere have been determined.

The 21-cm Line of Neutral Hydrogen

This review may be concluded by a brief account of what is probably the most spectacular achievement of radio-astronomy—the discovery in the radio spectrum of the neutral hydrogen line at a wavelength of 21 cm (about 1,420 Mc/s). At a wartime astronomical colloquium held in Holland in 1944, H. C. van der Hulst predicted that it might be possible to detect the 21-cm radiation originating in the clouds of neutral hydrogen in our galaxy. Seven years later—in March 1951—H. E. Ewen and E. M. Purcell of Harvard University announced their discovery of this radiation, and this was quickly confirmed by radio groups in Holland and Australia. The radiation results from the transition of the neutral hydrogen atom between two energy levels and provides an entirely new means of research on the nature and structure of the Milky Way and of other galaxies outside our own. One great and fortunate circumstance is that the dense interstellar dust clouds in our own galaxy which obscure our optical view of the more distant elements of its spiral structure, transmit unabsorbed the 21-cm radiation reaching us from great distances. This is one example of the manner in which radio-astronomy has provided a means of observing farther into the universe than has hitherto been possible by the largest and most expensive of optical telescopes. The outlook for the future is vast indeed.

R.L.S-R.

MEASUREMENTS IN TRAVELLING-WAVE STRUCTURES

Use of Resonant-Cavity and Perturbation Methods

By A. W. Aikin, M.A., Ph.D., A.M.I.E.E.

(Research Department, Metropolitan-Vickers Electrical Co. Ltd.)

SUMMARY.—The measurements of resonant frequency and Q of a resonant cavity are well-tried methods needing little introduction. Less well known, but extremely useful, is the method of measuring the field distribution within a resonant cavity by introducing a small perturbing object. A combination of these methods is of great value in determining the characteristics of a travelling-wave structure, and in particular has been used to obtain information and design data on the corrugated waveguide structures used in linear electron accelerators.

1. Perturbation Methods

THE principle of perturbation methods depends upon the fact that if a small perturbing object, e.g., a small metallic sphere, is introduced into a resonant cavity, a change in resonant frequency takes place which may be used to evaluate the field strength at the point of perturbation. Slater[1] has given a general theorem which states that if the walls of a cavity are pushed in at a point of high magnetic field, the resonant frequency is increased, while if pushed in at a point of high electric field the resonant frequency is decreased. He also derives the quantitative relationship given below. A rigorous proof of the general perturbation theory is both lengthy and tedious and has been attempted by several authors[2-4]. The most satisfactory proof known to the author is that of Casimir[2]. No attempt will be made to reproduce such a proof here but the following derivation due to J. Brown (unpublished) and based upon lumped circuit concepts is a novel and instructive illustration of the concepts involved.

We can consider our cavity in terms of a series-tuned circuit with elements L, C such that $1/LC = \omega_0^2$, where ω_0 is the angular resonant frequency of the cavity. One other condition must be satisfied, namely, that the average stored magnetic and electric energies in the inductance L and capacitance C shall be the same as the corresponding magnetic and electric energies in the cavity. This condition is

$$\frac{1}{2}LI^2 = \frac{1}{2}\int_V \mu H^2 \, d\tau = F_M \quad .. \quad .. \quad (1)$$

and

$$\frac{1}{2}\cdot\frac{Q^2}{C} = \frac{1}{2}\int_V \epsilon E^2 \, d\tau = F_E \quad .. \quad .. \quad (2)$$

where I, Q are the r.m.s. current and charge in the tuned circuit, E, H the r.m.s. electric and magnetic field strengths in the cavity, ϵ, μ the dielectric constant and magnetic permeability, and F_E, F_M the average stored electric and magnetic energies, the integrals being evaluated over elements $d\tau$ of the volume V of the cavity. Also $Q = I/\omega_0$ and $F_E = F_M = F$ at resonance.

If the cavity is assumed to be divided into two regions V', V'', where the region V' is very much smaller than V'', containing stored energies δF_E, δF_M, then we can define an inductance L' and capacitance C' such that

$$\frac{1}{2}L'I^2 = \delta F_M \quad .. \quad .. \quad .. \quad (3)$$

and

$$\frac{1}{2}\cdot\frac{Q^2}{C'} = \delta F_E \quad .. \quad .. \quad .. \quad (4)$$

The amended circuit may be drawn as in Fig. 1, where the elements L', C' correspond to the volume V' of the cavity, and L'', C'' correspond to the volume V''.

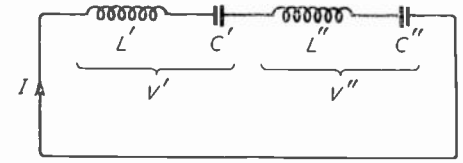


Fig. 1. *Equivalent circuit of resonator.*

This circuit must be the same as the simple tuned circuit with L, C and hence

$$L' + L'' = L \quad .. \quad .. \quad .. \quad .. \quad (5)$$

$$\frac{1}{C'} + \frac{1}{C''} = \frac{1}{C} \quad .. \quad .. \quad .. \quad .. \quad (6)$$

If now a perturbation is introduced into the cavity which removes the small volume V', then L', C' are small, and the fields in V'' remain substantially unchanged. The perturbed cavity resonates at a frequency ω given by $\omega^2 = 1/L''C''$, which, from equations (1) to (6) gives

$$\frac{\omega^2}{\omega_0^2} = \frac{LC}{L''C''} = \frac{1 - C/C'}{1 - L'/L} = \frac{1 - \delta F_E/F_E}{1 - \delta F_M/F_M}$$

MS accepted by the Editor, November 1954

Since δF_E, δF_M, and $\delta\omega = \omega - \omega_0$ are all small, and since $F = F_E = F_M$ we may write this as

$$\frac{2\delta\omega}{\omega_0} = \frac{\delta F_M - \delta F_E}{F} \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

This is the form of the general perturbation formula which is usually written as

$$\frac{\delta\omega}{\omega_0} = \frac{1}{2}\,\frac{\delta\int_V (\mu H^2 - \epsilon E^2)\,d\tau}{\int_V \mu H^2\,d\tau} \quad \ldots \quad \ldots \quad (8)$$

The significance of this formula may be seen as follows:

When a perturbation is introduced into a cavity, changes occur in the current and charge distributions in the walls, and currents and charges are also induced in the perturbing object. Changes thus occur in the stored magnetic and electric energies associated with these currents and charges, and it is these energy changes, expressed in terms of the changes in field distribution, which are represented by the quantities $\frac{1}{2}\,\delta\int_V \mu H^2\,d\tau$ and $\frac{1}{2}\,\delta\int_V \epsilon E^2\,d\tau$. The difference between these quantities represents an unbalance of the average stored magnetic and electric energies, and since at resonance these must be equal, a change of resonant frequency occurs which restores the equality, the amount of the change $\delta\omega$ being related to the energy unbalance introduced by the perturbation by formula (8).

2. Types of Perturbation

If the perturbing element takes the form of a small metallic sphere, introduced in a region of dielectric constant ϵ, the boundary conditions for the electric field in the electromagnetic case are the same as those in the electrostatic case provided the sphere is sufficiently small, and we may use the well-known solution[5] of the problem of a perfectly conducting sphere of radius r_0 in a uniform field E_0. The charges induced on the sphere produce outside it an additional field which is the same as that of a dipole of moment $4\pi\epsilon E_0 r_0^3$ at the centre of the sphere. The energy involved in separating these induced charges is loosely known as the 'energy of the sphere' in the field. Using the formula[5] for the energy of a dipole in a uniform field this energy is $4\pi\epsilon E_0^2 r_0^3 = 3\epsilon E_0^2 V$, where V is the volume of the sphere. This is identified as the quantity $\delta\int_V \epsilon E^2\,d\tau$ of formula (8), where E_0 now represents the peak and not the r.m.s. value of field strength at the point where the sphere is introduced.

A similar approach may be used to obtain the magnetic energy, but in this case it is necessary to substitute the boundary conditions for the electromagnetic case since these differ from those in the magnetostatic case because of induced currents. Consider a potential

$$\phi = -H_0 r\cos\theta + \sum_0^\infty \frac{A_m P_m(\theta)}{r^{m+1}}$$

where P_m is the Legendre Polynomial of order m. Then

$$H_r = -\frac{\partial\phi}{\partial r} = H_0\cos\theta - \frac{A_0}{r^2} - \frac{2A_1\cos\theta}{r^3} + \ldots - \frac{A_m P_m(\theta)\,.\,(m+1)}{r^{m+2}}$$

$$H_\theta = \frac{1}{r}\left(\frac{\partial\phi}{\partial\theta}\right) = -H_0\sin\theta + \frac{A_1}{r^3}\sin\theta + \ldots$$

The boundary equations in this case are obtained from Maxwell's equations, viz:
$H_r = 0$ when $r = r_0$, whence $A_0 = 0$, $A_1 = H_0 r_0^3/2$, $A_m = 0\ (m \neq 1)$.

The form of the induced potential in this case is the same as that of a dipole of moment $4\pi\mu H_0 r_0^3/2$ at the centre of the sphere.

Thus the currents induced in the sphere by the electromagnetic field have associated with them a stored energy $3/2\,\mu H_0^2 V$.

Substituting $\delta\int_V \mu H^2\,d\tau = 3/2\,\mu H_0^2 V$ and $\delta\int_V \epsilon E^2\,d\tau = 3\epsilon E_0^2 V$, equation (8) becomes

$$\frac{\delta\omega}{\omega_0} = \frac{3V}{2}\,\frac{(\frac{1}{2}\mu H_0^2 - \epsilon E_0^2)}{\int_V \mu H^2\,d\tau} \quad \ldots \quad \ldots \quad (9)$$

where H_0, E_0 are the unperturbed peak fields at the point where the sphere was introduced.

In the case of a dielectric sphere of dielectric constant ϵ_1, inserted in a region of dielectric constant ϵ_2, the electrostatic solution gives the energy in a uniform field E_0 to be $\frac{3(\epsilon_1 - \epsilon_2)}{\epsilon_1 + 2\epsilon_2} E_0^2 V$ while in this case the magnetic field is continuous across the boundary and so there is no first-order contribution to the magnetic energy. The perturbation formula in this case becomes

$$\frac{\delta\omega}{\omega_0} = -\frac{3V}{2}\,\frac{\epsilon_1 - \epsilon_2}{\epsilon_1 + 2\epsilon_2}\,\frac{E_0^2}{\int_V \mu H^2\,d\tau} \quad \ldots \quad \ldots \quad (10)$$

Other types of perturbation have also been used, e.g., needle-shaped objects which are sensitive largely to that component of electric field which is oriented in the direction of the needle[6], and button type perturbations[7] which are particularly suited to measurements at the walls on an axis of symmetry of the cavity.

3. Travelling-Wave Problem

The above formulae relate the perturbation of the resonant frequency of the cavity to the peak field strengths at the point of perturbation and the total stored energy in the cavity. In a travelling-wave structure we wish to relate the peak field strength to the power flow down the structure. To do this we first take a section of the structure and convert it into a resonant cavity by the introduction of suitable reflecting terminations. We may thus consider the oscillations in this cavity as composed of two waves, each of amplitude E, travelling in opposite directions.

Let S be the energy stored per unit length for each of the two waves and W the associated power flow. The cavity will have a length $n\lambda_g/2$ at the resonant frequency, where λ_g is the guide wavelength in the structure at that frequency and n is an integer. The total stored energy will then be $2S \cdot n\lambda_g/2 = Sn\lambda_g$.

Now suppose that a small perturbing object, e.g., a small dielectric sphere, has been introduced at a point of maximum E field in the cavity and the change in resonance frequency measured. Then, using formula (10) a quantity can be calculated which we shall term κ and which is defined by the equation

$$\kappa = \frac{\epsilon E_0^2}{\int_V \mu H^2 \, d\tau} \qquad \ldots \quad \ldots \quad \ldots \quad (11)$$

Then $E_0 = 2E$ and since $\int_V \mu H^2 \, d\tau =$ twice the stored energy in the cavity $= 2Sn\lambda_g$ we have

$$Sn\lambda_g = 2\epsilon E^2/\kappa \qquad \ldots \quad \ldots \quad \ldots \quad (12)$$

Now by definition, the energy velocity v_e in a structure is given by $v_e = W/S$ and it can be shown[8,9] that the energy velocity and group velocity v_g are equal for a lossless structure.

$$\therefore \frac{W}{v_g} = S = \frac{2\epsilon E^2}{\kappa n \lambda_g}$$

whence

$$\frac{E^2}{W} = \frac{\kappa n \lambda_g}{2\epsilon v_g} \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (13)$$

Thus, provided the group velocity v_g can be measured, the field strength E in the travelling-wave structure may be evaluated in terms of the power flow W by a perturbation method. The fields in the other parts of the structure may be evaluated by comparison methods; e.g., by a plot of variation of resonant frequency with position of the perturbing object.

4. Measurement of Group Velocity

The length of the resonant cavity is always some integral multiple n of half the guide wavelength λ_g. There will be a number of different resonant frequencies corresponding to different values of n and λ_g for any given length of cavity. By measuring a number of such resonances a plot of several points on the curve of λ_g against λ_0 (the free-space wavelength) can be made. Also by choosing several different cavity lengths the number of points on this curve may be increased so as to enable a complete dispersion curve to be plotted.

The group velocity at any point may then be evaluated from the formula[10]

$$\frac{v_g}{c} = \left(\frac{\lambda_g}{\lambda_0}\right)^2 \frac{d\lambda_0}{d\lambda_g} \qquad \ldots \quad \ldots \quad \ldots \quad (14)$$

where $d\lambda_0/d\lambda_g$ is derived from the slope of the dispersion curve, and c is the velocity of light.

The phase velocity v of the wave at any point is of course given by $v/c = \lambda_g/\lambda_0$.

5. Measurement of Attenuation

It is also possible to obtain the attenuation constant α of the travelling-wave structure, by measuring the Q of such a cavity. The derivation only applies when the attenuation is sufficiently small that the group and energy velocities may be regarded as equal, but this is substantially true in most practical cases.

Neglecting for the moment any loss in the reflecting planes inserted to form the cavity, the loss in the structure per unit length for a power flow W is $2\alpha W$ per unit time, and so for two equal and opposite travelling-waves the total loss will be $4\alpha W$ per unit time.

$$\text{Now } Q = \frac{2\pi \times \text{energy stored in the cavity}}{\text{energy dissipated in the walls in one cycle}}$$

$$= \frac{2\pi \cdot 2SL}{L \cdot 4\alpha W/f} = \frac{\pi f}{\alpha v_g} \text{ since } \frac{W}{S} = v_g$$

where f is the resonant frequency and L the length of the cavity. Hence

$$\alpha = \frac{\pi f}{Q v_g} = \frac{\omega}{2Q v_g} \qquad \ldots \quad \ldots \quad \ldots \quad (15)$$

gives the attenuation constant in terms of quantities which can be obtained by cavity measurements.

This formula only applies when the loss in the reflecting planes may be neglected and this is not always the case. If the loss is reasonably small, however, we can make two cavities of lengths L and $2L$ using the same shorting planes. If T is the loss in the reflecting planes in one cycle for a stored energy S per unit length and R the loss per unit length in the walls in one cycle, and if the cavity of length L has a Q-value Q_1 and that of length $2L$ a value Q_2, then

$$Q_1 = \frac{2\pi SL}{RL + T} = \frac{2\pi S}{R}\left(1 - \frac{T}{RL}\right)$$

if T is small compared with RL.
Similarly

$$Q_2 = \frac{2\pi S}{R}\left(1 - \frac{T}{2RL}\right),$$

$$\text{and } Q = \frac{2\pi S}{R} = 2Q_2 - Q_1 \quad .. \quad .. \quad (16)$$

is the value which would have been measured had the reflecting planes been entirely lossless.

6. Application to Linear Accelerator Structures

The corrugated waveguide structures used in linear accelerators have been described by several authors[11-13]. The type of most interest is that in which a circular waveguide is loaded by equally spaced discs having a central hole. With such a waveguide it is possible to propagate a wave which is axially symmetrical and has a strong axial component of electric field, and which has a phase velocity lower than the velocity of light. Such a wave may be used for accelerating electrons by a technique somewhat analogous to 'surf-riding'.

Fig. 2 shows the way in which a section of such corrugated waveguide may be turned into a resonant cavity. In such a cavity the axial component of electric field is a maximum at the ends of the cavity and has n nodes, where $n\lambda_g/2$ is the length of the cavity, and to a first approximation, the field may be considered to vary sinusoidally along the length of the cavity. In order that the reflecting end planes shall not disturb the field configuration in the cavity, they must be placed at a plane about which the waveguide would be symmetrical. This means that we must place conducting planes either in the centre of an iris, or in the middle of a corrugation. The cavity length must therefore be chosen as an integral multiple of half the corrugation pitch D.

7. Cavity Construction

The cavities were formed from a number of interlocking sections each corresponding to one corrugation together with an adjacent iris. Several different types of end sections were constructed so as to terminate the cavity either in the middle of a corrugation, or in the middle of an iris. Fig. 2 shows a typical cavity with a termination of each type at the ends. Any cavity whose length was an integral multiple of half the corrugation pitch could be constructed by clamping together an appropriate number of sections.

8. Measurements

A feed-through method of measuring cavity response was used. The cavity was fed by means of a small loop from a low-power oscillator and the response measured by a second loop and crystal detector. The coupling of the loops was so small as not to perturb the cavity frequency nor appreciably affect the cavity Q. Dispersion measurements were made on a guide of 2-cm pitch using a series of cavities varying in length from 5 cm to 13 cm in steps of 1 cm. Taking several values of the mode number n for each cavity a dispersion curve could be drawn. Fig. 3 shows a typical curve and the group-velocity curve deduced from it. The results of these measurements agree with those made by standing-wave methods[11] on long sections of guide, to better than 1 part in 2,000 in the worst case compared.

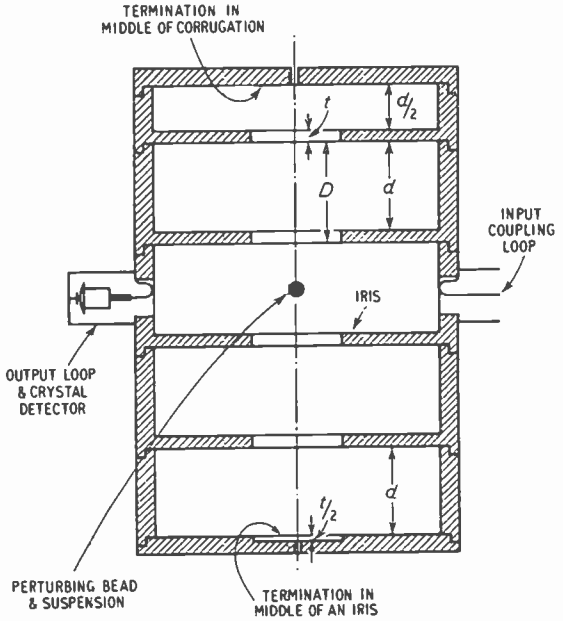


Fig. 2. *Typical cavity showing: 1. Two different types of end-plate terminations; 2. Method of suspension of perturbing bead; and 3. Methods of coupling and detection.*

For the perturbation measurements of field strength a small metallic sphere was used. This consisted of a small phosphor-bronze ball bearing, drilled and threaded on to a fine strand of silk which passed through small holes in the reflecting end planes (see Fig. 2). The bead could be adjusted externally to any position on the axis of the cavity. Here the magnetic field component is zero and so for a sufficiently small bead the term $\frac{1}{2}\mu H_0^2$ of formula (9) is negligible. By comparing the results obtained using both $\frac{1}{4}$-in. diameter and $\frac{3}{8}$-in. diameter spheres, it was shown that this term could in fact be neglected. The field strength was only evaluated at its maximum (or maxima) in the cavity so as to obtain the maximum E-field in the travelling-wave case. Elsewhere the field has approximately a sinusoidal

variation which would complicate the measurement. Using a sphere, no measurements can be made near the reflecting walls since the energy of the sphere in the field is modified by its reflection in these walls, and the perturbation effect is approximately doubled. At these walls, however, a button-type piston method[7] was used which has not this defect, and the results obtained agreed to within 2% with those obtained by the bead method.

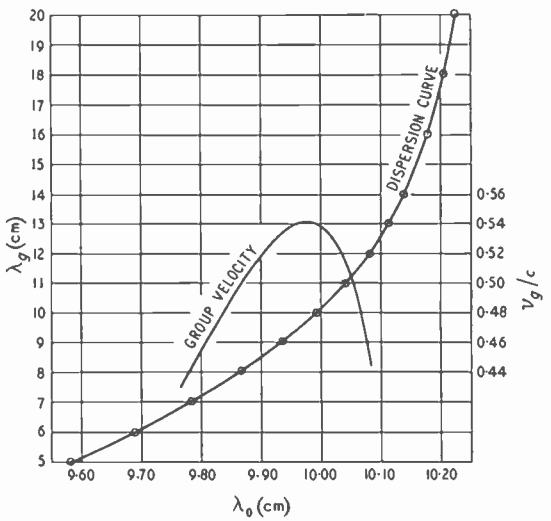


Fig. 3. *Typical dispersion and group-velocity curves.*

One complication which arises in these measurements is that where the structure is periodic rather than uniform the axial field distribution is also non-uniform and varies across a corrugation. As a result, the field measured in a cavity in which the maximum occurs in the plane of an iris gives a value slightly different from that measured in a cavity in which the maximum occurs in the middle of a corrugation. This discrepancy becomes more marked as the hole diameter in the iris is decreased. By using both sorts of cavity and taking a mean value, however, the mean axial field in the travelling-wave case could be found, and it is this value which is of interest for the design of linear accelerators. In all, it is estimated that an error not exceeding 3% is obtained in most cases, rising possibly to 5% in the case of small hole diameters.

As a result of these measurements it was shown experimentally that the formula[13] $E = \lambda/\pi a^2 \sqrt{(480W)}$ for the axial field strength in a guide of iris radius a with a power flow W is not sufficiently accurate for most purposes. The error involved may be as high as 30% and for accurate calculation the more detailed theory[14] must be used, taking into account the power flow associated with the higher-order space harmonics.

Measurements of attenuation have also been made by the Q-method described and the results obtained agree quite well with measurements made on long sections of guide. The measured values of Q also agreed with theoretical values, provided that account was taken of the increase in surface resistivity due to surface finish.

9. Conclusion

Cavity methods have been used extensively to obtain design data on a wide range of corrugated waveguides for use in linear accelerators and, in particular, perturbation methods have been used to measure field strengths. The detailed results of these measurements are to be published separately.

Acknowledgments

The author wishes to thank Mr. J. Brown of Imperial College for permission to reproduce his derivation of the perturbation formula and Dr. Willis Jackson, F.R.S., M.I.E.E., Director of Research and Education, and Mr. B. G. Churcher, M.Sc., M.I.E.E., Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

[1] J. C. Slater, "Microwave Electronics", p. 80 (Van Nostrand), 1950.
[2] H. B. G. Casimir, *Philips Research Reports*, 1951, Vol. 6, p. 162.
[3] A. Cunliffe & L. E. S. Mathias, *Proc. Instn elect. Engrs*, Pt. III, 1950, Vol. 97, p. 367.
[4] L. C. Maier, Jr. & J. C. Slater, *J. Appl. Phys*, 1952, Vol. 23, p. 78.
[5] J. A. Stratton, "Electromagnetic Theory" (McGraw-Hill), 1941.
[6] L. C. Maier, Jr., M.I.T.Tech.Rep. 143, 1949.
[7] W. W. Hansen & R. F. Post, *J. Appl. Phys.*, 1948, Vol. 19, p. 1059.
[8] J. S. Bell, A.E.R.E. Report T/R, 858, 1952.
[9] R. B. R. Shersby-Harvie, *Proc. Phys. Soc*, 1948, Vol. 61, p. 255.
[10] H. R. L. Lamont, "Waveguides" (Methuen) 1942.
[11] L. B. Mullett & B. G. Loach, *Proc. Phys. Soc*, 1948, Vol. 61, p. 271.
[12] J. C. Slater, *Rev. Mod. Phys.*, 1948, Vol. 20, p. 473.
[13] W. Walkinshaw, *Proc. Phys. Soc*, 1948, Vol. 61, p. 246.
[14] W. Walkinshaw & J. S. Bell, A.E.R.E. Report G/R. 675.

LOSS AND PHASE OF SIMPLE EQUALIZERS

By H. J. Orchard, M.Sc., A.M.I.E.E.

SUMMARY.—An approximate method of computing the loss and phase of certain simple equalizers is obtained by using the first one or two terms of a rapidly-convergent power-series expansion. The saving in computation which is achieved is greatest when the amount of equalization is small and when many points on the characteristic are required.

1. Introduction

THIS paper is concerned with the loss and phase characteristics which can be obtained from the constant-impedance networks shown in Fig. 1. The principal use of these networks is in simulating or equalizing smoothly-varying loss characteristics such as are given, for example, by unloaded cables. In simple cases one network alone may suffice, but where a more precise match is needed over a frequency band many octaves wide, it is usually necessary to employ a tandem connection of several networks, each with different characteristics.

Although the constant-impedance form is most convenient when several networks have to be connected in tandem, the characteristics of a single network can be obtained from one or other of the circuits shown in Figs. 2 and 3, and when operation between valve stages is called for it is often possible to use these circuits by combining them with the coupling components.

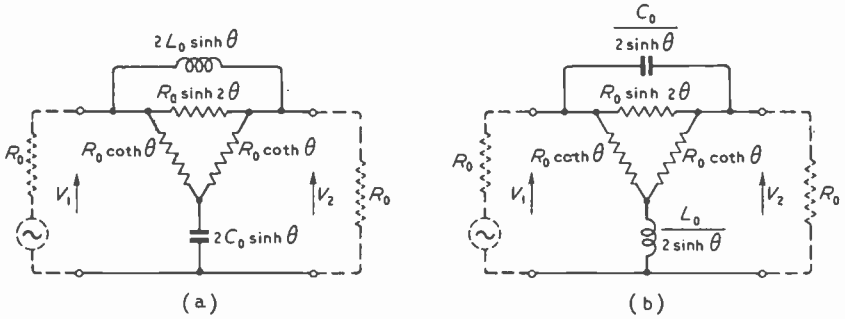


Fig. 1. *Constant-impedance networks with image impedance R_0 and in which $\omega_0 L_0 = R_0 = (\omega_0 C_0)^{-1}$.*

Taking the loss α, in nepers, and the phase β, in radians, as given by

$$\alpha + j\beta = \log_e(V_1/V_2) \quad .. \quad .. \quad .. \quad (1)$$

with V_1 and V_2 defined as in the figures, then, for the network of Fig. 1(a), we have

$$\alpha + j\beta = \log_e \frac{\omega_0 + pe^{\theta}}{\omega_0 + pe^{-\theta}} \quad .. \quad .. \quad (2)$$

where $p = j\omega = j2\pi f$.

2θ is the image attenuation, in nepers, of the pad formed by the three resistances and is the limiting value of the loss of the network as ω tends to infinity, while ω_0 $(= 2\pi f_0)$ is the angular frequency at which the network loss is equal to θ nepers; i.e., one half the limiting value. Plotted on a logarithmic scale of frequency the loss and phase have the general form shown in Fig. 4; around $f = f_0$ the loss is an odd function and the phase an even function.

The network of Fig. 1(b) has a loss and phase given by

$$\alpha + j\beta = \log_e \frac{p + \omega_0 e^{\theta}}{p + \omega_0 e^{-\theta}} \quad .. \quad .. \quad (3)$$

with ω_0 and θ having the same significance as before except that 2θ is now the limiting value of the loss as ω tends to zero. A tandem connection of this network and one of the previous type, both having the same values for f_0, θ and R_0 would give a constant loss of 2θ nepers and no phase shift. This can be seen by adding equations (2) and (3) and cancelling identical factors.

Hence, if we calculate the loss and phase for a network of Fig. 1(a) having specific values of f_0 and θ, then we can obtain the loss and phase for the corresponding network of Fig. 1(b) simply by subtracting the loss from 2θ and by reversing the sign of the phase.

In the practical design of equalizers of this kind it is customary[1] to prepare curves of the loss for a variety of values of θ and then, by curve-matching techniques, to determine approximately the correct values of f_0 and θ to use. Following this the loss of the chosen equalizer is computed accurately and from a study of the residual error in the equalization (or simulation as the case may be) a decision is made as to whether or not a small change in the parameters

MS accepted by the Editor, November 1954

will improve the situation. Then follows a certain amount of 'cut and try' to make the equalizing as precise as possible. Finally, when the loss is correctly adjusted, the associated phase characteristic may be required.

All this involves a fair quantity of computing, particularly if the complete equalizer comprises many sections. It is the purpose of the paper to show how this computing may sometimes be simplified and speeded up. The simplification is achieved by expanding the loss and phase into a rapidly-convergent power series of which the first one or two terms provide an adequate approximation to the sum when the pad loss is not too large.

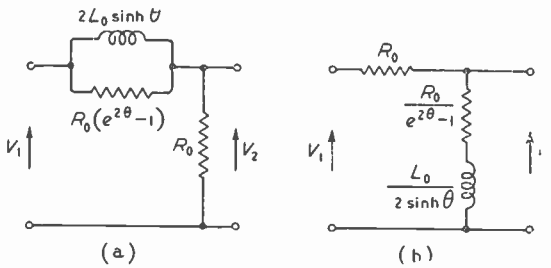


Fig. 2. *Networks having the same loss and phase characteristics as those of Fig.* 1.

2. Power Series Expansion

Consider first the right-hand-side of equation (2). It is an odd function of θ which, for real values of ω, is analytic in the complex θ-plane inside a circle round the origin whose radius is at least $\frac{1}{2}\pi$. Thus we may expand it into a power series as follows

$$\log_e \frac{\omega_0 + pe^{\theta}}{\omega_0 + pe^{-\theta}} = c_1\theta + c_3\frac{\theta^3}{3!} + c_5\frac{\theta^5}{5!} + \ldots . \quad .. \quad .. \quad (4)$$

which will converge for $|\theta| \leqslant \frac{1}{2}\pi$ nepers (i.e., 13·6 dB). To find the c_s as functions of p and ω_0 we differentiate equation (4) with respect to θ; a little rearrangement then gives

$$2p^2 + 2\omega_0 p\cosh\theta = (\omega_0^2 + p^2 + 2\omega_0 p\cosh\theta)\left(c_1 + c_3\frac{\theta^2}{2!} + c_5\frac{\theta^4}{4!} + \ldots\right) \quad .. \quad .. \quad (5)$$

By equating the coefficients of corresponding powers of θ we derive the recurrence relation

$$c_{2n+1} = \frac{2\omega_0 p}{(\omega_0+p)^2}\left[\frac{\omega_0 - p}{\omega_0 + p} - \binom{2n}{2}c_3 - \binom{2n}{4}c_5 - \ldots - \binom{2n}{2n-2}c_{2n-1}\right] \quad .. \quad .. \quad (6)$$

together with the value of c_1. Explicitly, the first three terms are

$$c_1 = \frac{2p}{(\omega_0+p)} \quad .. \quad .. \quad .. \quad (7.1)$$

$$c_3 = \frac{2\omega_0 p(\omega_0 - p)}{(\omega_0+p)^3} \quad .. \quad .. \quad .. \quad (7.2)$$

$$c_5 = \frac{2\omega_0 p(\omega_0 - p)(\omega_0^2 - 10\omega_0 p + p^2)}{(\omega_0+p)^5} \quad (7.3)$$

The importance of this power-series expansion lies in the rapidity with which it converges. For values of θ less than about 0·5 neper (i.e., 4·3 dB) the contributions from the terms in the series die away so quickly that the error due to truncating the series is given fairly accurately by the value of the first term which is neglected.

When this condition holds, the first term of the series by itself provides quite a useful approximation to the loss and phase. The maximum value of $|c_3|$ is unity so that the maximum error due to using only the first term will be of the order of $\theta^3/3!$ nepers or radians; at most frequencies it will be considerably less than this. If greater accuracy is required then the first two terms can be used. As $|c_5| \leqslant 5$ the error will then not exceed $5\theta^5/5! = \theta^5/4!$ nepers or radians.

Slightly better estimates of the errors involved are given by using the maximum moduli of the real and imaginary parts of c_3 and c_5: the derivation is straightforward although laborious. The results have been summarized for reference in the curves of Fig. 5 which show the maximum contributions made to the loss and phase by the second and third terms, expressed as a function of the pad loss 2θ.

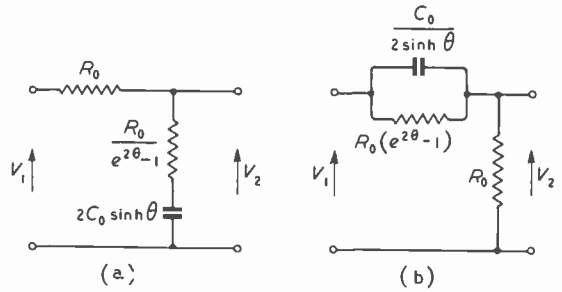


Fig. 3. *Alternative networks having the same loss and phase characteristics as those of Fig.* 1.

3. Computation

Explicit formulae suitable for computing are found by splitting equations (7.1) and (7.2) into their real and imaginary parts and substituting into equation (4)

$$\alpha = a_1 2\theta + a_3\frac{\theta^3}{3!} + \ldots \text{ nepers} \quad .. \quad (8.1)$$

$$\beta = b_1 2\theta + b_3\frac{\theta^3}{3!} + \ldots \text{ radians} \quad .. \quad (8.2)$$

where

$$a_1 = \frac{x^2}{1 + x^2} \qquad b_1 = \frac{x}{1 + x^2}$$

$$a_3 = \frac{8x^2(1 - x^2)}{(1 + x^2)^3} \qquad b_3 = \frac{2x(1 - 6x^2 + x^4)}{(1 + x^2)^3}$$

$$x = \frac{f}{f_0}$$

Plotted on a logarithmic scale of x the coefficients in (8.1) and (8.2) are odd and even functions respectively about the point $x = 1$.

Practical computation could conveniently proceed as follows. One will be given θ, f_0 and a series of values of f at which the loss and phase are required. The first step is to find the values of $x = f/f_0$ corresponding to the prescribed frequencies; these can be read from a slide rule. Next, one must decide whether one or two terms in the series are to be used, basing the decision upon estimates, found from Fig. 5 (a) and (b), for the maximum error committed in each case. If one term is adequate then it may be sufficient to read off the values of a_1 and b_1 from accurately drawn curves and to multiply them on a slide rule by 2θ. Alternatively, if greater precision is required, a slide rule or desk calculating machine can be used to find the entire term. When two terms are to be used, the first term cannot be found accurately enough from a single curve and the use of slide rule or calculating machine is essential. The contribution of the second term, however, is so small in comparison that it need not be known to the same percentage accuracy and can be found quite easily via a curve of a_3 or b_3.

Table 1 gives the values to four decimal places of a_1, b_1, a_3 and b_3 for selected values of x; these are adequate for the purpose of drawing the required curves.

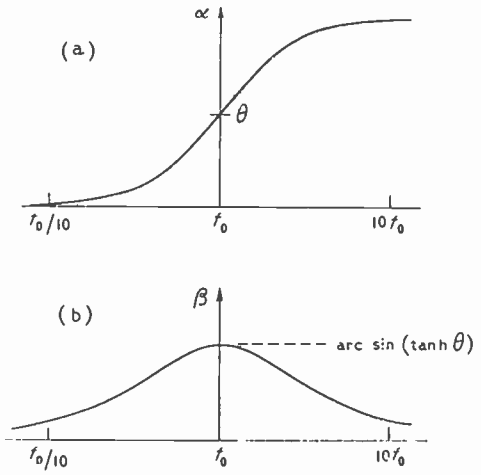


Fig. 4. *General form of the loss and phase characteristics of the networks of Figs.* 1(*a*), 2(*a*) *and* 3(*a*).

TABLE 1

x	a_1	b_1	a_3	b_3	x	a_1	b_1	a_3	b_3
	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$		$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$
0·05	24	498	198	977	1·0	5000	5000	0	−10000
0·07	48	696	384	1339	1·1	5475	4977	−1883	−9774
0·10	99	990	768	1824	1·2	5901	4918	−3489	−9196
0·12	141	1182	1087	2101	1·3	6282	4832	−4792	−8393
0·14	192	1373	1450	2331	1·4	6621	4729	−5804	−7469
0·16	249	1560	1849	2512	1·5	6923	4615	−6554	−6499
0·18	313	1743	2279	2639	1·6	7191	4494	−7081	−5536
0·20	384	1923	2730	2708	1·7	7429	4370	−7423	−4613
0·22	461	2098	3197	2718	1·8	7641	4245	−7617	−3751
0·24	544	2269	3670	2668	1·9	7830	4121	−7693	−2958
0·26	633	2435	4143	2559	2·0	8000	4000	−7680	−2240
0·28	727	2596	4609	2392	2·2	8287	3767	−7464	−1019
0·30	825	2752	5059	2168	2·4	8520	3550	−7100	−59
0·35	1091	3118	6080	1385	2·6	8711	3350	−6666	682
0·40	1379	3448	6888	336	2·8	8868	3167	−6210	1250
0·45	1683	3742	7430	−900	3·0	9000	3000	−5760	1680
0·50	2000	4000	7680	−2240	3·5	9245	2641	−4739	2334
0·55	2322	4222	7638	−3601	4·0	9411	2352	−3907	2621
0·60	2647	4411	7327	−4915	4·5	9529	2117	−3249	2715
0·65	2970	4569	6781	−6126	5·0	9615	1923	−2730	2708
0·70	3288	4697	6043	−7194	6·0	9729	1621	−1990	2560
0·75	3600	4800	5160	−8094	7·0	9800	1400	−1505	2360
0·80	3902	4878	4178	−8815	8·0	9846	1230	−1174	2163
0·85	4194	4934	3138	−9357	9·0	9878	1097	−940	1983
0·90	4475	4972	2076	−9725	10·0	9900	990	−768	1824
0·95	4743	4993	1022	−9934	15·0	9955	663	−349	1280
1·00	5000	5000	0	−10000	20·0	9975	498	−198	977

If θ is specified in decibels rather than nepers, and if α and β are required in decibels and degrees, then equation (8) can be rearranged into

$$\alpha = a_1.P + a_3.\,0{\cdot}276(P/10)^3 + \ldots \text{ decibels} \quad (9.1)$$

$$\beta = b_1.\,6{\cdot}596P + b_3\,1{\cdot}82\,(P/10)^3 + \ldots \text{ degrees} \quad (9.2)$$

where P is the value of 2θ expressed in decibels.

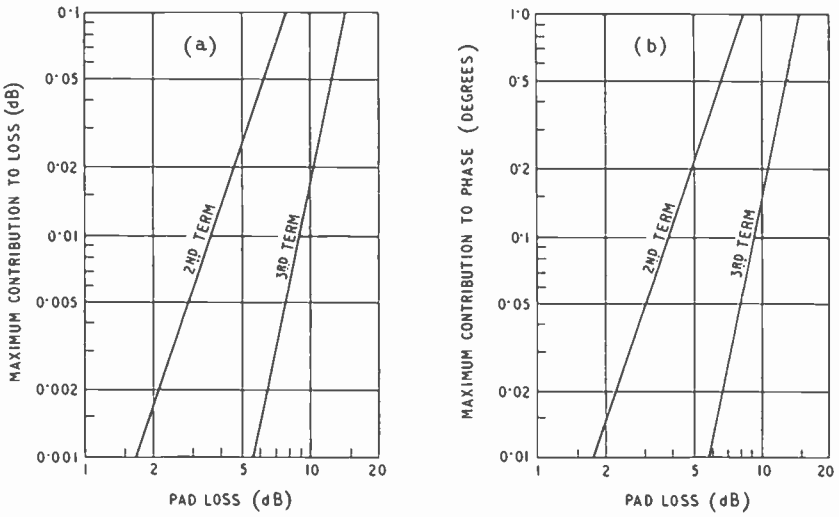


Fig. 5. *Maximum contribution of the second and third terms of the series in equation (4) to the loss (a), and the phase (b).*

Equations (8) and (9) are applicable as they stand to the network of Fig. 1(a). The corresponding formulae for the other network are found simply by reversing the sign associated with a_3, b_1 and b_3 and by replacing a_1 with $(1 - a_1) = (1 + x^2)^{-1}$.

4. Discussion on Application

The advantage of using the formulae is naturally greatest when only the first term of the series is necessary to give the required accuracy. This will hold in practical cases for pad losses up to about 3 or 4 dB. In such cases the computation of loss and phase is particularly simple and quick; with an automatic desk calculating machine each point can be evaluated in a few tens of seconds.

When the contribution of the second term must be included, computing time increases, but as the correction involved is small it is sufficient to obtain it graphically and it can be added mentally to the value found for the first term.

The error due to neglecting the third and higher terms increases, initially, with the fifth power of the pad loss, and this sets a fairly sharp barrier, around 10 dB, to the range over which the first two terms are a useful approximation. Any attempt to use more than two terms, so as to increase this range, would probably take longer than evaluating the original exact expression and hence would defeat the object. This sets a limit to the practical application of the expansion.

Acknowledgment

The writer is much indebted to Dr. J. M. Linke for his many helpful comments regarding the composition of the paper. Acknowledgment is also made to the Engineer-in-Chief of the G.P.O. for permission to make use of the information contained in this paper.

REFERENCE

[1] H. Kimball, "Motion Picture Sound Engineering", D. Van Nostrand, New York, 1938, Chaps. 16 and 17.

SOME ASPECTS OF STANDING-WAVE PATTERNS

By C. P. Allen, B.Sc.(Eng.), A.C.G.I., A.M.I.E.E. and P. A. Lindsay, Ph.D., D.I.C., A.C.G.I.

(*Communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England*)

STANDING waves occur in many engineering problems and are met in most branches of physical sciences. Since oblique projection of a three-dimensional representation of mathematical functions is often helpful in providing a visual aid to the understanding of their behaviour[1], it is thought that some useful purpose might be served in presenting such a picture of standing waves, in spite of the fact that the basic equations are already established.

As is well known, a standing wave is given by the sum of two progressive waves travelling in opposite directions; the one being a forward-travelling wave, and the other a backward-travelling one, for example, arising from the reflection of the forward-travelling wave at some discontinuity in the system.

It is obvious that a function of n independent variables will require $n + 1$ dimensions to represent it pictorially. Since, in general, a wave function depends upon three space co-ordinates and one time co-ordinate, it is only possible to represent in the form of a three-dimensional model a wave function which depends upon one space co-ordinate (distance), the other co-ordinate being time. An example of such a wave function is given by the displacement on a stretched string, or voltage or current variations along a transmission line.

In a loss-free medium, a one-dimensional wave disturbance is fully described by the second-order partial differential equation, generally known as the wave equation,

$$\frac{\partial^2 w(t, z)}{\partial z^2} = \frac{1}{k^2}\frac{\partial^2 w(t, z)}{\partial t^2} \quad .. \quad .. \quad (1)$$

in which $w(t, z)$ = a function of the independent variables, t and z, describing the disturbance in the loss-free medium,

k = constant coefficient having the dimensions of velocity.

Equation (1) is one of the few partial differential equations for which a general solution is known, its form is

$$w(t, z) = g\left(t - \frac{z}{k}\right) + h\left(t + \frac{z}{k}\right) \quad .. \quad (2)$$

where g and h are arbitrary functions of the independent variables t and z.

Since the phase velocity of a wave is defined as the velocity of a constant phase, $t - z/k$, relative to a stationary frame of reference, one gets from Equ. (2)

$$\frac{d}{dt}\left(t - \frac{z}{k}\right) = 0$$

$$\left(\frac{\partial}{\partial t} + v_z\frac{\partial}{\partial z}\right)\left(t - \frac{z}{k}\right) = 0$$

$$1 - \frac{v_z}{k} = 0$$

$$\therefore v_z = k \quad .. \quad (3)$$

Similarly, for the point $t + z/k = \text{const.}$ the velocity of a fixed point on the wave is given by

$$v_z = -k \quad .. \quad .. \quad .. \quad (4)$$

which is equal but opposite to the velocity for $t - z/k = \text{const.}$ To stress this fact it is now convenient to put $k = v$, where v is the magnitude of the phase velocity of the waves g and h for non-dispersive media.

Further, since both g and h will satisfy Equ. (2), it is clear that the function $w(t, z)$ describes either a forward-travelling wave, or a backward-travelling wave or the sum of both, neither of which need, in general, be periodic in t or z. However, in practice any wave can be expressed as the sum of a series of trigonometrical functions by applying the methods of Fourier analysis to it and, furthermore, most physical and engineering problems concerning wave motion are usually formulated in terms of trigonometrical functions of space and time co-ordinates. It will be more convenient, therefore, without any loss of generality (except for the boundary conditions) to express in the usual exponential form both arbitrary functions g and h and in what follows to consider only the fundamental Fourier component of the wave.

Equation (2) then takes the form

$$\begin{aligned}\mathbf{W}(t, z) &= \mathbf{A}\exp. j\omega(t - z/v) + \mathbf{B}\exp. j\omega(t + z/v)\\ &= \mathbf{W}_1(t, z) + \mathbf{W}_2(t, z)\\ &= W_1\exp. j\chi_1(t, z) + W_2\exp. j\chi_2(t, z)\end{aligned} \quad .. \quad .. \quad .. \quad (5)$$

which is itself a solution of Equ. (1) and where $\mathbf{A}$ and $\mathbf{B}$ are arbitrary complex constants to be determined by the boundary conditions, and ω

[1] E. Jahnke and F. Emde: "Tables of Functions", Teubner, 1938.

MS accepted by the Editor, August 1954

is the fundamental angular frequency of oscillation of the disturbance, w of Equ. (2) being either the real or the imaginary part of W, whichever happens to be more convenient.

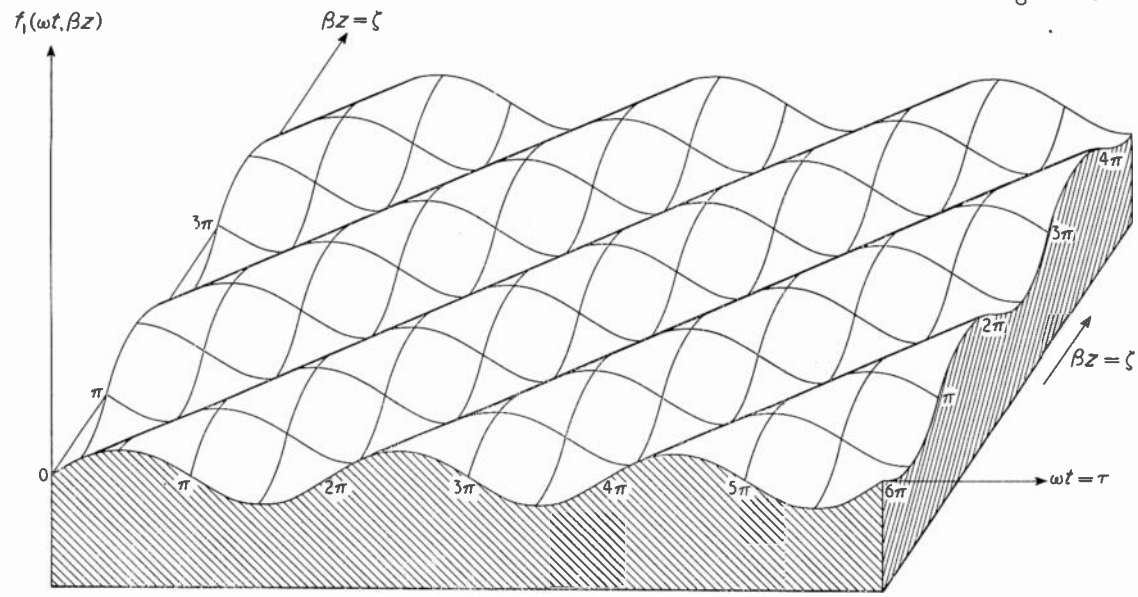


Fig. 1. *Three-dimensional representation of the function* $f_1(\omega t, \beta z) = \sin(\omega t - \beta z)$.

In Equ. (5) W_1 and W_2 are the magnitudes of $\mathbf{W}_1$ and $\mathbf{W}_2$ and are constant, whereas χ_1 and χ_2, the arguments of $\mathbf{W}_1$ and $\mathbf{W}_2$, are functions of t and z.

Since the second term $\mathbf{W}_2$ on the right-hand side of Equ. (5) is the backward-travelling wave arising, for example, from the reflection at some discontinuity met by the forward-travelling one in its progression along the system, it is convenient to introduce a new complex quantity, $\boldsymbol{\rho} = \rho \exp.(j\phi) = \mathbf{B}/\mathbf{A}$, which is called the reflection coefficient of magnitude ρ and phase angle ϕ, and is simply the ratio of the complex amplitude of the backward to that of the forward-travelling wave. Hence Equ. (5) can be written as

$$\begin{aligned}\mathbf{W}(t, z) &= \mathbf{A}\{\exp. j\omega(t - z/v) + \boldsymbol{\rho} \exp. j\omega(t + z/v)\} \\ &= A\{\exp. j[\omega(t - z/v) + \psi] + \\ &\quad \rho \exp. j[\omega(t + z/v) + \phi + \psi]\} \qquad (6)\end{aligned}$$

where A is the modulus of $\mathbf{A}$, and ψ its phase angle. In terms of magnitude and phase angle of W one can derive from Equ. (6)

$$\begin{aligned}\mathbf{W}(t, z) &= A\{1 + \rho^2 + 2\rho\cos(2\omega z/v + \phi)\}^{\frac{1}{2}} \\ &\exp. j \tan^{-1} \frac{\sin\{\omega(t - z/v) + \psi\} + \rho \sin\{\omega(t + z/v) + \phi + \psi\}}{\cos\{\omega(t - z/v) + \psi\} + \rho\cos\{\omega(t + z/v) + \phi + \psi\}} \\ &= W(z) \exp. j\chi(t, z) \qquad .. \qquad .. \qquad (7)\end{aligned}$$

It is to be noted here that contrary to $\mathbf{W}_1$ and $\mathbf{W}_2$ the magnitude of W is no longer constant but depends on the distance z.

It is now necessary to determine the value of the complex constant of integration A. Suppose that the initial conditions define magnitude and phase angle of the standing-wave function W of Equ. (7) at a point $t = t_0$, $z = z_0$, thus

$$\mathbf{W}(t_0, z_0) = W(z_0) \exp. j\chi(t_0, z_0) \qquad .. \qquad (8)$$

Then for the particular case of a matched system, that is when $\boldsymbol{\rho} = 0$, one can show from Equ. (7) that

$$A = W \text{ (now independent of } z_0) \qquad .. \qquad (9a)$$

and

$$\tan\psi = \frac{\tan\chi\cos\omega(t_0 - z_0/v) - \sin\omega(t_0 - z_0/v)}{\tan\chi\sin\omega(t_0 - z_0/v) + \cos\omega(t_0 - z_0/v)} \qquad .. \qquad .. \qquad (9b)$$

Equ. (9b) simplifies to $\tan\psi = \tan\chi$ for $t_0 = 0$, $z_0 = 0$; that is, when the point of observation (the point at which the boundary conditions are noted) coincides with the origin $t = 0$, $z = 0$. It is possible to write Equ. (9b) also in the form

$$\tan\{\omega(t_0 - z_0/v) + \psi\} = \tan\chi \qquad .. \qquad (10)$$

This shows that for $\boldsymbol{\rho} = 0$ the same value ψ (for a given χ) will be obtained for any point of observation along the lines

$$\omega(t_0 - z_0/v) = (\chi - \psi) \pm n\pi$$

in the t,z plane. This is entirely the consequence of the functional properties (existence of periodicity and phase velocity) of a progressive undamped sinusoidal wave as shown in Fig. 1.

For $\boldsymbol{\rho} \neq 0$, Equs. (9) are no longer true. In this case, the magnitude and phase angle of the

constant of integration, **A**, can be derived from different expressions, again in terms of the magnitude and phase constant of the wave function $\mathbf{W}(t,z)$ at the point of observation, that is, in terms of the boundary conditions,

$$A = W(z_0)\{1 + \rho^2 + 2\rho\cos(2\omega z_0/v + \phi)\}^{-\frac{1}{2}} \quad .. \quad .. \quad .. \quad (11a)$$

$$\tan\psi = \frac{\tan\chi\{\cos\omega(t_0 - z_0/v) + \rho\cos[\omega(t_0 + z_0/v) + \phi]\} - \sin\omega(t_0 - z_0/v) - \rho\sin[\omega(t_0 + z_0/v) + \phi]}{\tan\chi\{\sin\omega(t_0 - z_0/v) + \rho\sin[\omega(t_0 + z_0/v) + \phi]\} + \cos\omega(t_0 - z_0/v) + \rho\cos[\omega(t_0 + z_0/v) + \phi]} \quad .. \quad .. \quad .. \quad (11b)$$

They are valid for the most general case of $\rho \neq 0$ and $t_0 \neq 0$, $z_0 \neq 0$. For the more common case of $t_0 = 0$, $z_0 = 0$ Equs. (11) simplify to

$$A = W(0)\{1 + \rho^2 + 2\rho\cos\phi\}^{-\frac{1}{2}} \quad (12a)$$

$$\tan\psi = \frac{\tan\chi\{1 - \rho\cos\phi\} - \rho\sin\phi}{\{1 + \rho\cos\phi\} + \rho\tan\chi\sin\phi} \quad (12b)$$

It is worth noting at this point that even in the most general case A does not depend on the phase angle χ, which the standing-wave function $\mathbf{W}(t,z)$ will have at the point of observation t_0, z_0. This is also clear from Equ. (7), where it can be seen that any change in χ will merely require a new value of ψ, leaving A unaltered. Conversely, changes in W do not affect ψ. Furthermore, once the integration constant $A \exp. j\psi$ has been calculated from the boundary conditions $W(z_0) \exp. j\chi\,(t_0, z_0)$ at a point $t_0 \neq 0$, $z_0 \neq 0$, it is always possible to express these boundary conditions in terms of the boundary conditions $W(0) \exp. j\chi\ (0, 0)$ which hold at the origin and lead to the same value of the integration constant A.

The main object of interest of this note is the standing-wave pattern given by taking the real or imaginary parts of Equs. (6) and (7). The imaginary part of Equ. (6) divided by A is

$$A^{-1}\,\mathrm{Im}\,\mathbf{W}(t,z) = \sin\{\omega t - \omega z/v + \psi\} + \rho\sin\{\omega t + \omega z/v + \phi + \psi\} \quad .. \quad .. \quad (13)$$

One can see from Equ. (13) that the phase angle ψ of the constant of integration merely displaces the whole function along the ωt axis. It is therefore always possible to express the function of Equ. (13) in terms of suitably chosen co-ordinates which each time would make $\psi = 0$, without any loss of generality.

Now for $\psi = 0$ and $A \neq 0$, $\rho \neq 0$, and putting $\beta = \omega/v$, the following set of functions is obtained from Equs. (5) and (13),

$$f_1(\omega t, \beta z) = A^{-1}\,\mathrm{Im}\,\mathbf{W}_1(t,z) = \sin(\omega t - \beta z) \quad .. \quad .. \quad (14a)$$

$$f_2(\omega t, \beta z) = A^{-1}\,\mathrm{Im}\,\mathbf{W}_2(t,z) = \sin(\omega t + \beta z) \quad .. \quad .. \quad (14b)$$

$$f_3(\omega t, \beta z;\ \rho, \phi) = A^{-1}\,\mathrm{Im}\,\mathbf{W}(t,z) = \sin(\omega t - \beta z) + \rho\sin(\omega t + \beta z + \phi) \quad .. \quad (14c)$$

f_1, f_2 and f_3 being normalized with respect to amplitude. In what follows ωt and βz will quite often be put equal to τ and ζ respectively, the functions being thus valid for any angular frequency ω.

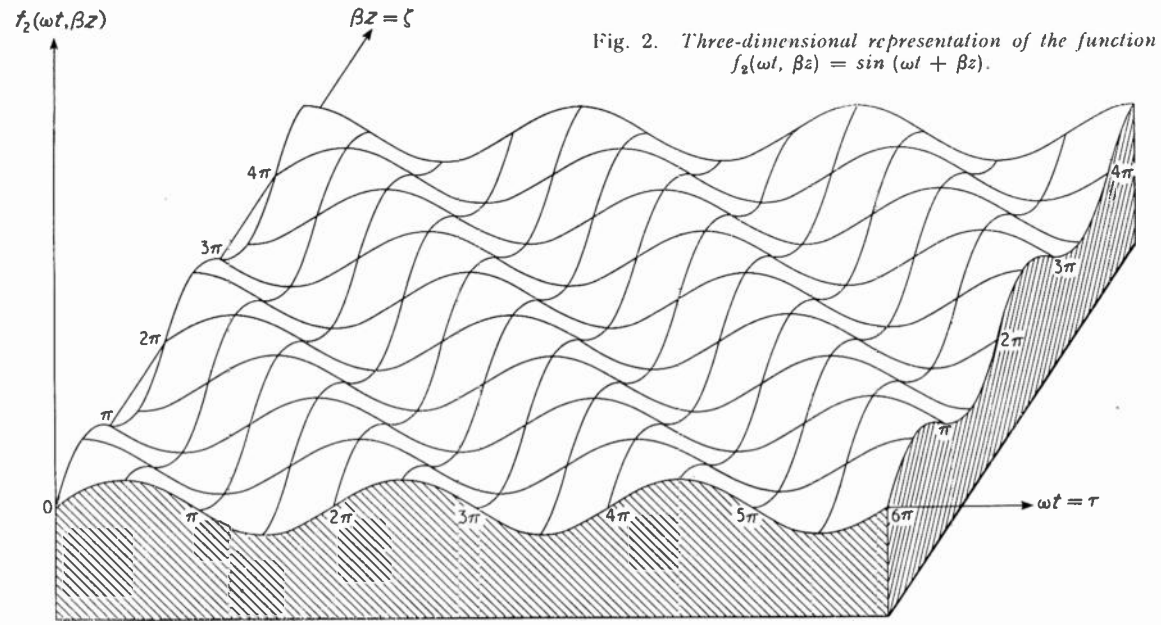


Fig. 2. *Three-dimensional representation of the function* $f_2(\omega t, \beta z) = \sin(\omega t + \beta z)$.

Expressing functions f_1, f_2 and f_3 in terms of the two independent variables $\tau = \omega t$ and $\zeta = \beta z$ one obtains in three dimensions a surface for each function f. Their perspective drawings are shown in the figures.

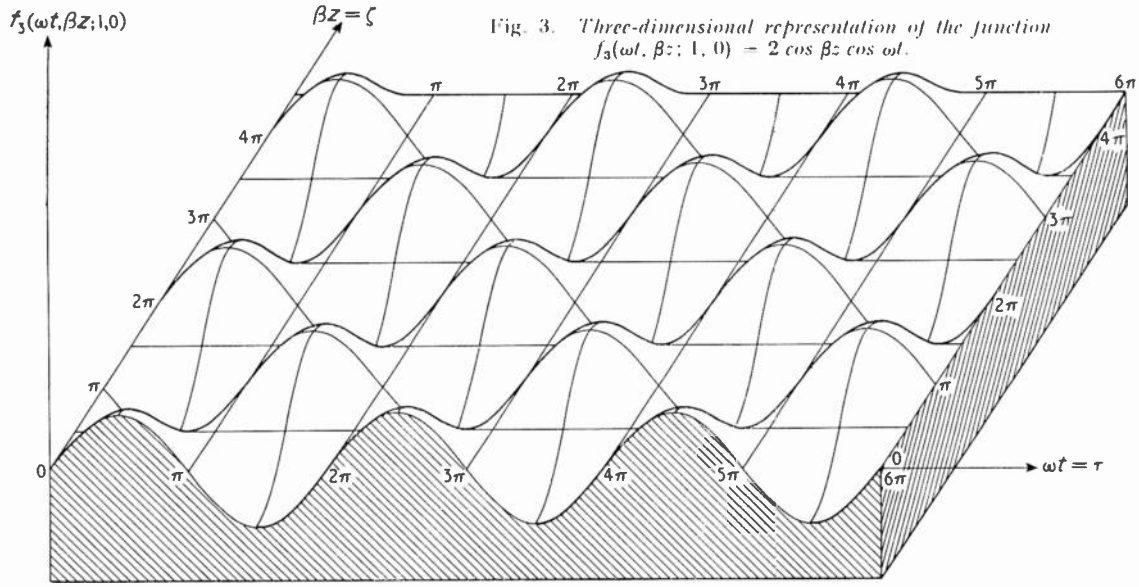


Fig. 3. *Three-dimensional representation of the function* $f_3(\omega t, \beta z; 1, 0) = 2 \cos \beta z \cos \omega t$.

In Fig. 1 is shown the forward-travelling wave $\sin(\omega t - \beta z)$. The figure represents a sample of an infinite wave train set up in the system at a point $z \to -\infty$, $t \to -\infty$ (to avoid reflections and transients). From the figure one can see that the curves of constant displacement (height) are straight lines parallel to the crests and troughs of the surface f_1. By definition they are the lines of constant phase velocity v of the wave. Thus, for example, proceeding along a crest of the surface from the bottom left-hand corner to the top right-hand corner of the diagram is equivalent to riding with the wave at velocity v along the whole length of the system.

It is perhaps of some interest at this juncture to point out that, if the co-ordinate axes τ and ζ are rotated by an arbitrary angle θ, Equ. (14a) changes to

$$f_1(\tau', \zeta') = \sin(\tau'\Theta_1 - \zeta'\Theta_2) \quad .. \quad .. \quad (15)$$

where $\Theta_1 = \cos\theta - \sin\theta$, $\Theta_2 = \cos\theta + \sin\theta$. Thus it can be deduced from the form of Equ. (15) that the cross-section by a vertical plane of the surface shown in Fig. 1 is always a sine wave (of zero amplitude in the limiting case) irrespective of the direction of the plane. (This is fairly obvious in the case of Figs. 1 and 2, but is of interest in connection with Figs. 3 and 4, to which it also applies on the strength of Equ. (14c). For $\theta = \pi/4$, Equ. (15) reduces to

$$f_1(\tau', \zeta') = -\sin\sqrt{2}\,\zeta' \quad .. \quad .. \quad (16)$$

For this particular angle of rotation the function f_1 becomes independent of one of the independent variables; it also acquires this property for $\theta = \pm n\pi/4$.

Fig. 2 shows a one-dimensional backward-travelling wave given by the function $f_2(\omega t, \beta z)$. As before, the figure is a representation of a sample of an infinite wave train set up in the system at a point $z \to \infty$ and at a time $t \to -\infty$. The lines of constant phase velocity are seen from the figure to lie again along the lines of crests and troughs. The slope of these lines is negative in the t,z plane, since the figure represents a backward-travelling wave. Thus, proceeding along a crest of the surface from the top left-hand corner down towards the bottom right-hand corner is equivalent to riding on the wave along the system with a velocity equal to $-v$.

The function $f_3(\omega t, \beta z; \rho, \phi)$, Equ. (14c), presents a standing wave for various values of ρ, and is more complicated than either f_1 or f_2. First of all, it can be found from writing f_3 in the form

$$f_3(\tau, \zeta; \rho, \phi) = \sin(\tau - \zeta) + \rho \sin(\tau + \zeta + \phi) \quad .. \quad .. \quad .. \quad (17)$$

that the effect of the phase angle of the reflection coefficient, ϕ, is merely to shift the whole pattern in the direction of v. Thus by putting $\tau'' = \tau + \phi/2$, $\zeta' = \zeta'' + \phi/2$, Equ. (17) now becomes

$$f_3(\tau'' \zeta''; \rho, \phi) = \sin(\tau'' - \zeta'') + \rho \sin(\tau'' + \zeta'') \quad .. \quad .. \quad .. \quad (18)$$

Hence, it is seen that, in general, by a suitable choice of the origin, it is always possible to eliminate the phase angle ϕ, in discussing the form

of the standing-wave pattern f_3. Thus without any loss of generality Figs. 3–8 have been drawn for the sake of simplicity for the case of $\phi = 0$.

Putting $\phi = 0$, f_3 reduces to f_1 for $\rho = 0$, which of course it must do, since this is the case of an infinite system or in practice a reflectionless load at the end of a transmission line.

For $\rho = 1$, Equ. (17) gives

$$f_3(\tau, \zeta; 1, 0) = 2 \sin \tau \cos \zeta \quad . . \qquad . . \qquad (19)$$

which is shown in Fig. 3. It can be seen from the figure that in this case, the nodal lines given by $f_3 = 0$ (lines of cross-section of the surface and the τ, ζ plane) form a square grid. This means physically that zero displacement occurs either (1) for any value of t at points $z = 1/\beta\,(2n + 1)\,\pi/2$ or (2) at all points z (that is along the whole of the system) for given values of $t = (1/\omega)n\pi$. Furthermore, it is seen from Fig. 3 that if one travels along the system with an arbitrary velocity $0 < |\,dz/dt\,| < \infty$, one can never experience a constant displacement as is the case for an observer travelling with the phase velocity v of a progressive wave shown in Figs. 1 and 2.

In Fig. 4. is shown the case $0 < \rho < 1$, which reduces to Figs. 1 and 3 respectively in the limiting cases of $\rho = 0$ and $\rho = 1$. The case $0 < \rho < 1$ is characterized by the nodal pattern of Fig. 5 which is intermediate between the straight lines of Fig. 1 and the square grid of Fig. 3. The expression for nodal curves can be obtained from Equ. (14c) by equating it to zero. This then gives

$$\zeta = \tan^{-1}\left\{\frac{1 + \rho}{1 - \rho} \tan \tau\right\} \quad . . \qquad . . \qquad (20a)$$

or

$$\rho\,(\tau, \zeta) = \frac{\sin\,(\zeta - \tau)}{\sin\,(\zeta + \tau)} \quad . . \qquad . . \qquad (20b)$$

or in dashed co-ordinates of Equ. (16)

$$\zeta' = \frac{1}{\sqrt{2}} \sin^{-1}\left\{\rho \sin \frac{1}{\sqrt{2}} \tau'\right\} \quad . . \qquad (21a)$$

or

$$\rho\,(\tau', \zeta') = \frac{\sin \sqrt{2}\,\zeta'}{\sin \sqrt{2}\,\tau'} \quad . . \qquad . . \qquad (21b)$$

As can be seen from Fig. 5 for $0 < \rho < 1$ the curves of Equ. (20a) take the form of wavy lines sloping in general at $\pi/4$. In Fig. 6 is shown the whole family of the nodal curves as expressed by Equ. (20b) for different values of the parameter ρ. They form a surface with τ and ζ placed horizontally and ρ vertically. As a visual help the $\rho = \rho\,(\tau, \zeta)$ surface has drawn upon it the horizontal contours for $\rho = 0{\cdot}0$, $0{\cdot}1$, $0{\cdot}2$, . . . $1{\cdot}0$, and the lines of steepest descent for every 30°.

It should be added that Fig. 6 represents only a slice $0 \leqslant \rho \leqslant 1$ of the surface given by Equ. (20b), which in spite of its simple algebraic form is fairly complicated. Fig. 8 gives the function already shown in Fig. 6 for the complete range of values $-\infty < \rho < \infty$. Since ρ is the magnitude of the coefficient of reflection, only positive values of the function have a physical significance. The function shown in Fig. 8 has saddle points, lines of

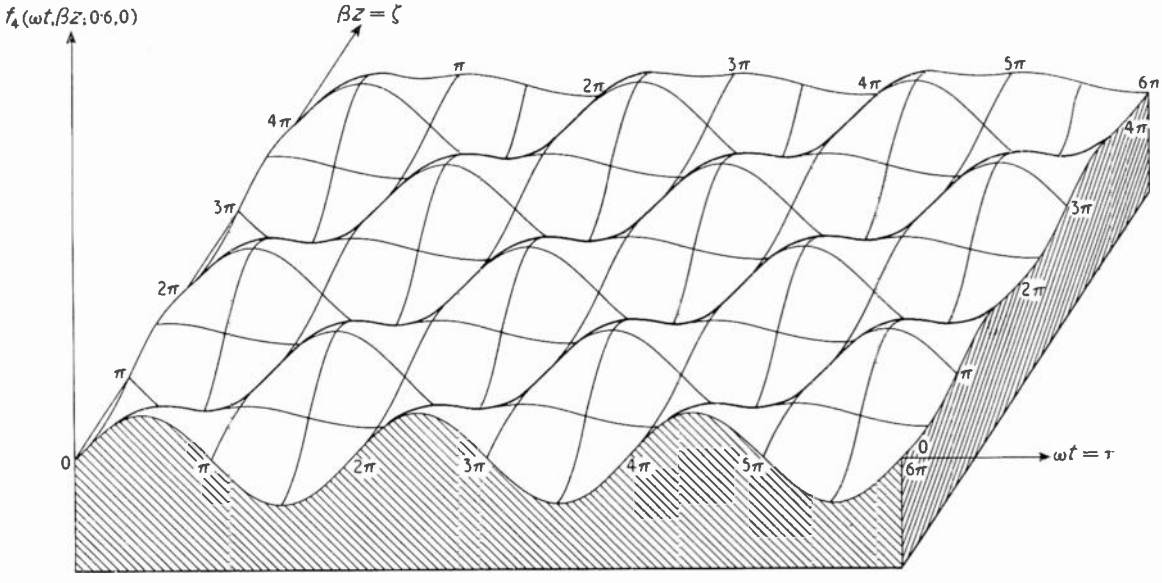


Fig. 4. *Three-dimensional representation of the function* $f_4(\omega t, \beta z;\ 0{\cdot}6, 0) = 1{\cdot}6 \sin \omega t \cos \beta z - 0{\cdot}4 \cos \omega t \sin \beta z$.

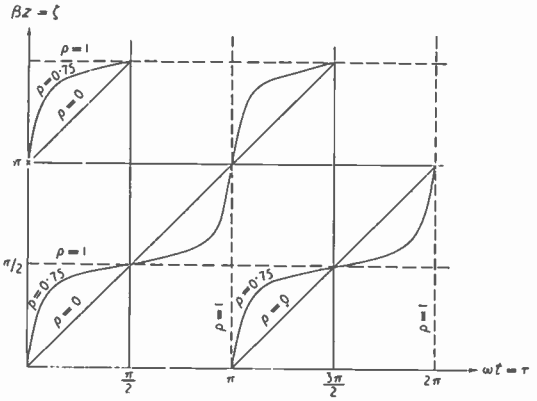


Fig. 5. *Nodal pattern for* $0 < \rho < 1$, *the intermediate condition between Figs.* 1 *and* 3; $\zeta = \tan^{-1}\left\{\frac{1+\rho}{1-\rho}\tan\tau\right\}$.

since by definition ρ can only be infinite for $A = 0$ $B \neq 0$.

Returning to Fig. 4 ($0 < \rho < 1$), one can see that the surface shown lies between the limiting cases of Fig. 1 ($\rho = 0$) and Fig. 3 ($\rho = 1$). Thus from the figures and Equs. (14a), (17) (for $\phi = 0$) and (19), for a fixed position along the system, ζ constant, τ variable, one experiences: (a) for $\rho = 0$, sinusoidal displacement always between 1 and -1 for any fixed ζ, (b) for $\rho = 1$, at the worst sinusoidal displacement between 2 and -2, for $\zeta = n\pi$, at the best zero displacement at the nodes of the standing-wave pattern $\zeta = (2n + 1)\pi/2$, (c) for $0 < \rho < 1$, at the worst displacement between $1 + \rho$ and $-(1 + \rho)$ (maximum of the standing-wave envelope $\zeta = n\pi$), at the best variations between $1 - \rho$ and

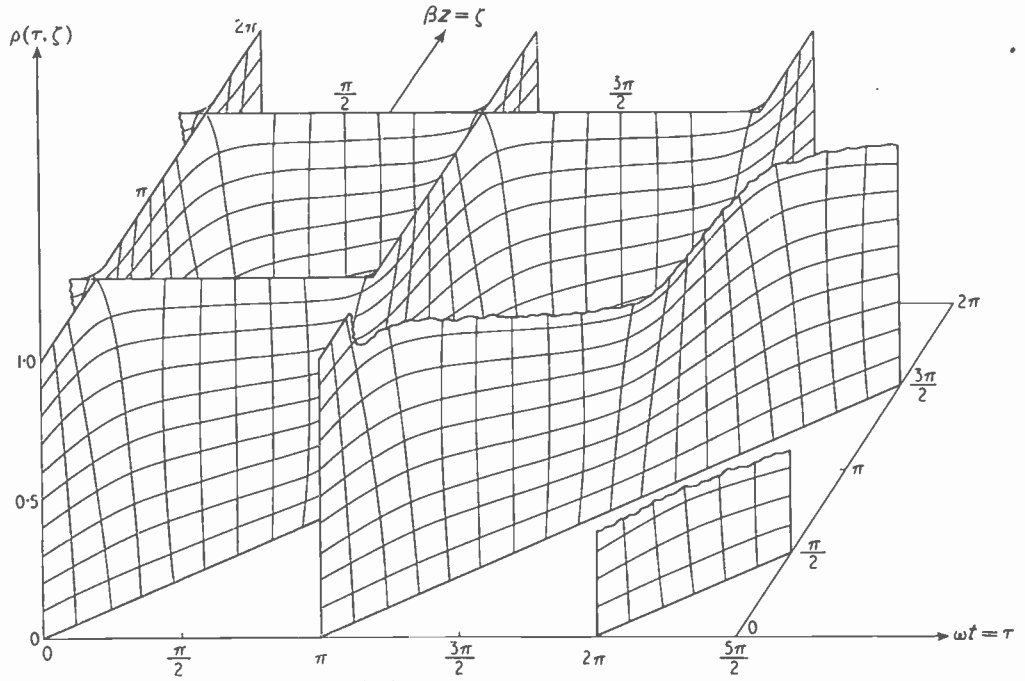


Fig. 6 (above). *Nodal pattern for* $\rho(\tau, \zeta) = \dfrac{\sin\zeta - \tau}{\sin\zeta + \tau}$.

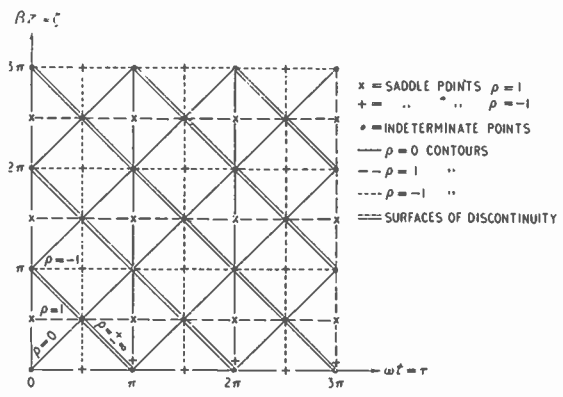


Fig. 7 (right). *Saddle points, lines of discontinuity and lines cutting the* τ, ζ-*plane corresponding to Fig.* 8.

discontinuity and lines along which it cuts the τ, ζ-plane. For convenience they are shown separately in Fig. 7. For the sake of clarity, the $\rho = \pm 1$ contours are shown in Figs. 7 and 8. It is of some interest to note that the planes of discontinuity of ρ (see Fig. 8) cut the τ, ζ-plane along the nodal lines $f_2 = 0$ of the backward-travelling wave of Fig. 2. This corresponds physically to a system feeding power in the reverse direction with no forward-travelling wave,

$-(1-\rho)$ (minimum of the standing-wave envelope) $\zeta = (2n+1)\pi/2$.

Alternatively, using the same equations and keeping time fixed, one has for the displacement at any point along the system, τ constant, same point of the wave (that is, one experiences the same displacement) all the time, (b) for $\rho = 1$, one varies one's position following a sinusoidal path of amplitude 1, whatever the value of ζ', (c) for $0 < \rho < 1$, one varies one's position again,

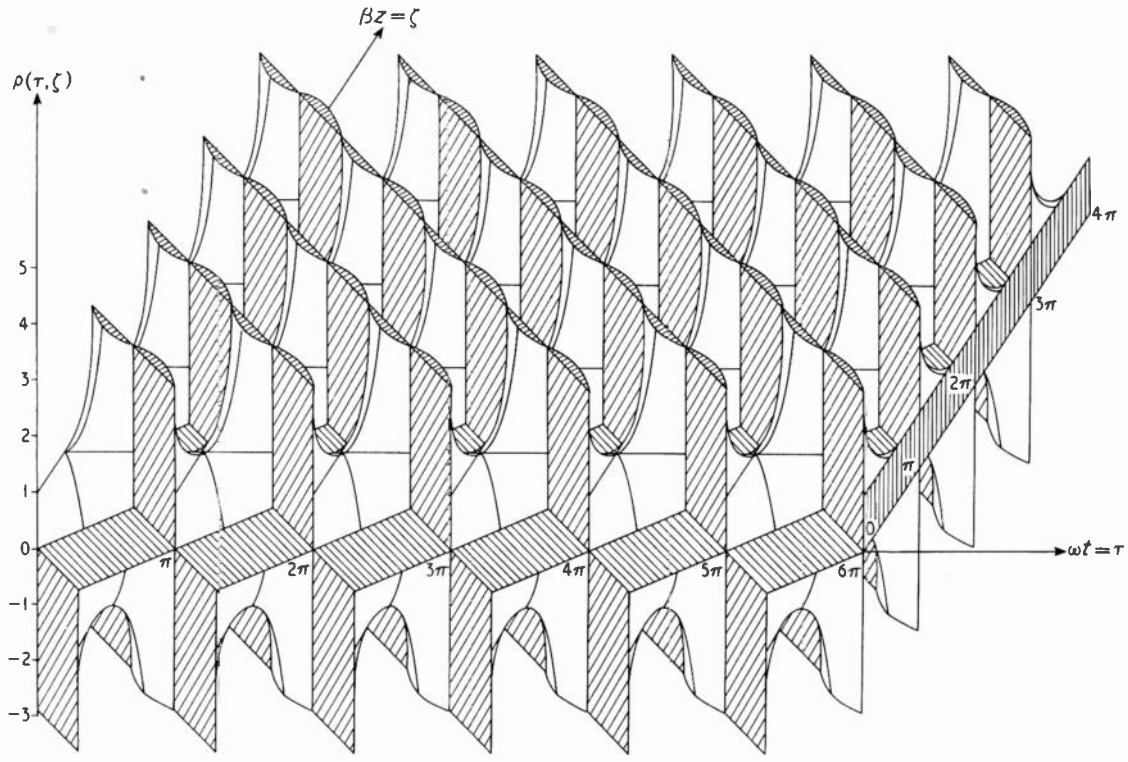


Fig. 8. *Nodal pattern for* $\rho = \dfrac{\sin(\zeta - \tau)}{\sin(\zeta - \tau)}$.

ζ variable: (a) for $\rho = 0$, always a sinusoidal wave of amplitude 1, (b) for $\rho = 1$, at the worst a sinusoidal wave of amplitude 2 for $\tau = (2n+1)\pi/2$, at the best no displacement anywhere along the system for $\tau = n\pi$, (c) for $0 < \rho < 1$, at the worst a sinusoidal wave of amplitude $1 + \rho$ for $\tau = (2n+1)\pi/2$, at the best a sinusoidal wave of amplitude $1 - \rho$ for $\tau = n\pi$.

In order to consider what happens if one travels along the system with the phase velocity v, it is most convenient to rotate the co-ordinates by $\pi/4$ and express Equ. (17) in terms of the dashed co-ordinates τ', ζ', hence obtaining

$$f_3(\tau', \zeta';\ \rho, 0) = -\sin \sqrt{2}\,\zeta' + \rho \sin \sqrt{2}\,\tau' \quad (22)$$

Thus if one travels in the new co-ordinate system with the phase velocity v (ζ' constant, τ' variable), then (a) for $\rho = 0$, one stays at the following a sine curve but of amplitude ρ, whatever the value of ζ'.

Thus there are two extreme cases of (1) constant displacement when travelling along the system and (2) no displacement, either at point ζ during the whole time interval considered, or along the whole system for a given value of τ. They can only occur for the limiting values of $\rho = 0$ and $\rho = 1$ respectively. In the general case of $0 < \rho < 1$ the cross-sections of f_3 with $\tau =$ constant or $\zeta =$ constant planes always give according to Equ. (17) sine waves, which vary in amplitude and phase but whose mean value remains zero. On the other hand, it can be seen from Equ. (22), that the cross-sections of f_3 with $\tau' =$ constant, $\zeta' =$ constant planes give sine waves of constant amplitude for a given ρ, but whose mean value varies.

STABILITY OF OSCILLATION IN VALVE GENERATORS

By A. S. Gladwin, Ph.D., D.Sc.
(*University of Sheffield*)

(*Continued from p. 214, August issue*)

7. Criteria for Stability

THE oscillation is stable if all the roots of the equation (6.5) have negative real parts, for this ensures that the amplitude of any transient decreases with time. In general, a direct solution of the equation is not feasible: only in certain special cases, some of which are considered later, can the expression for $D(p)$ be sufficiently simplified to permit the direct calculation of p. However, a complete solution is unnecessary since all that need be known is whether or not the real parts of all the roots are negative, and this information can be obtained without actually solving the equation. Two methods will be discussed. The first is an application of the Routh-Hurwitz stability rules.

It is known from electric circuit theory[22] that any impedance function $Z(p)$ associated with a linear network having lumped parameters is a real rational function of p; i.e., the quotient of two polynomials with real coefficients. Functions like $Z(p + jn\omega_0)$ will also be rational functions of p but with complex coefficients. Since $D(p)$ consists of sums, products, and quotients of such functions it is also a rational function. It remains to find the nature of the coefficients.

Now for any real impedance[24] $Z(p^* + jn\omega_0) = Z^*(p - jn\omega_0)$. Inspection of the expressions for a_1, etc., shows that if p^* is substituted for p the following transformations take place.

$$\left.\begin{array}{lll} a_1 \rightarrow c_{-1}^*, & b_1 \rightarrow b_{-1}^*, & c_1 \rightarrow a_{-1}^* \\ a_0 \rightarrow c_0^*, & b_0 \rightarrow b_0^*, & c_0 \rightarrow a_0^* \\ a_{-1} \rightarrow c_1^*, & b_{-1} \rightarrow b_1^* & c_{-1} \rightarrow a_1^* \end{array}\right\} \quad \ldots \quad (7.1)$$

Then from (6.4), $D(p^*) = D^*(p)$ and from this it follows that the coefficients of the polynomials are real. It can also be shown by using (4.11), (4.12), (4.15), and (4.17) that when $p = 0$

$$a_1 = c_1, \quad a_0 = c_0, \quad a_{-1} = c_{-1} \quad \ldots \quad (7.2)$$

Hence $D(0) = 0$, and $D(p)$ can therefore be written as $D(p) = pP_1(p)/P_2(p)$, in which $P_1(p)$ and $P_2(p)$ are real polynomials. The root $p = 0$ corresponds to an oscillation of constant amplitude and frequency ω_0; i.e., the steady state. All other roots of $D(p) = 0$ are the same as the roots of $P_1(p) = 0$. Routh[26], and later Hurwitz[28], investigated the conditions for all the roots of such equations to have negative real parts. The Routh-Hurwitz stability criteria take the form of a number of inequalities between the coefficients of the polynomial. The details are given in later Sections where the method is applied to particular problems.

The second method originates also in the work of Routh[27], and is based on a theorem of Cauchy[29] (Routh also made use of the theorem in deriving the first method). One way of stating the theorem is as follows: "If $D(p) = u + jv$ is analytic except for a finite number of poles inside and on a closed contour, then the number of times which the locus of $D(p)$ encircles the origin when p moves once round the contour is $N - P$, where N is the number of zeros and P the number of poles of $D(p)$ inside the contour."

In Routh's application the contour of p was the imaginary axis from $-j\infty$ to $j\infty$ and a semi-circle of infinite radius, centred on the origin, lying in the right-hand half-plane. This contour encloses all values of p having positive real parts. As the functions considered by Routh had no poles (except at ∞) the stability criterion was that the locus of $D(p)$ should not enclose the origin. What Routh in fact considered was the number of times which the ratio u/v passed through 0 and changed in sign from positive to negative and vice versa, but this is simply another way of specifying the number of encirclements. Bode[23] reached the same conclusion also by way of Cauchy's theorem, and Nyquist[30] had previously obtained a similar result by another method.

The Routh-Nyquist criterion, as it may properly be named, is more general than the Routh-Hurwitz in so far as it applies to any analytical function and not merely to polynomials with real coefficients, but the criterion fails if the function has any poles within the contour. This difficulty can be overcome by using a slightly more sophisticated definition[31] of 'encirclement', but for the present purpose this refinement is unnecessary. The impedances Z_i, Z_0, and Z_t are those of a passive network and so can have no poles in the right-hand half-plane, (for otherwise an exponentially-increasing voltage would appear spontaneously across the network terminals), but Z_t, being a transfer function, may have zeros there. Since Z_t occurs in the denominators of some terms in the expressions for a_1, etc., these terms will have poles where Z_t has zeros. It can be shown, however, that all such

terms cancel in the final result. Hence $D(p)$ has no poles in the right-hand half-plane.

Two difficulties prevent the Routh-Nyquist criterion from being applied directly to oscillators. First $D(0) = 0$; i.e., the locus of $D(p)$ passes through the origin. This can be avoided by indenting the p-plane contour with a small semi-circle about the origin in the right-hand half-plane. The second point, which was discussed in Section 3, is that the domain of p must be restricted so that its imaginary part lies between $-\frac{1}{2}\omega_0$ and $\frac{1}{2}\omega_0$. Instead of the Routh-Nyquist contour, the contour shown in Fig. 7 must therefore be used. For convenience the contour is described in the clockwise direction. The stability criterion can then be stated as:

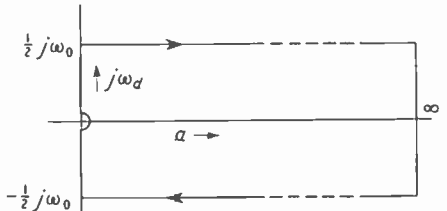
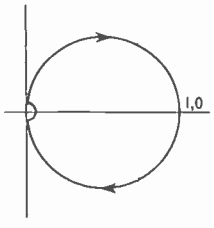


Fig. 7. *p-plane contour.*

"The oscillation is stable if the locus of $D(p)$ does not enclose the origin when p describes the contour of Fig. 7." Any encirclement—indicating instability—will be in the clockwise direction.

The Routh-Nyquist criterion is more usually expressed in terms of encirclement of the point 1,0. With this convention the oscillation is stable if the locus of $1 - D(p)$ does not encircle the point 1,0. In the language of feedback-amplifier theory $D(p)$ corresponds to the 'return difference' and $1 - D(p)$ to the 'loop transmission', although there is no physical loop in the oscillator corresponding to this function. In the present application it is more convenient to adhere to the expression of the criterion in terms of $D(p)$.



Although stability can be discussed completely in terms of either the Routh-Hurwitz or the Routh-Nyquist criteria it is advantageous to use both. The Routh-Nyquist locus diagram is valuable in illustrating points which are not immediately obvious from the Routh-Hurwitz criteria, and in deriving numerical relations for the simpler types of instability. With the more complicated forms of instability the Routh-Hurwitz rules are the only practicable method of obtaining numerical results.

Some general features of the locus are now considered. Earlier it was shown that $D(p^*) = D^*(p)$. This means that the locus has mirror symmetry with respect to the real axis. Also, for very large values of p all the network impedances vanish because of the shunting effect of stray capacitances. Then $a_1 = b_0 = c_{-1} = 1$, and all the other coefficients are 0. Hence $D(p) \rightarrow 1$ as $p \rightarrow \infty \pm \frac{1}{2}j\omega_0$. In most oscillators $D(p) = 1$ when $p = \pm \frac{1}{2}j\omega_0$, so only imaginary values of p (i.e., real frequencies) need be considered.

Fig. 8 shows loci corresponding to (a) stability (b) instability—one zero, and (c) instability—two zeros. When p is small $D(p) = D(0) + pD'(0) = pD'(0)$ and because of symmetry $D'(0)$ is a real number. As p traverses the small semicircle near the origin $D(p)$ also describes a semicircle, and this lies in the right-hand half-plane if $D'(0)$ is positive, and in the left-hand half-plane if $D'(0)$ is negative. These remarks, together with obvious topological considerations, lead to the following conclusions:

(1) If $D'(0)$ is positive the locus of $D(p)$ encircles the origin an even number of times or not at all.

(2) If $D'(0)$ is negative $D(p)$ makes an odd number of encirclements.

It follows that

(3) The oscillation can be stable only if $D'(0) > 0$.

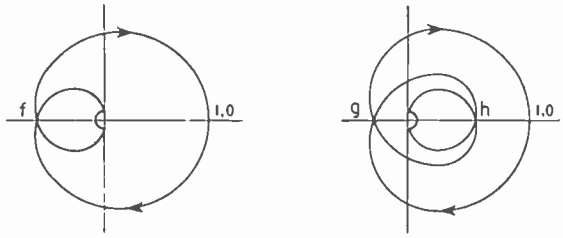


Fig. 8. *Loci of $D(p)$ or $F(p)$.*

The possibility that $D'(0) = 0$ is excluded, for this would imply that $D(p)$ had a double zero at the origin and this would correspond to a transient of frequency ω_0 with linearly increasing amplitude. Condition (3) therefore covers all possibilities. Double or multiple zeros can exist only in theory. They represent a critical adjustment of the network and amplifier parameters which cannot be achieved in practice. This remark applies also to any zero, simple or multiple, occurring at any other point on the contour.

The stability criteria derived here indicate only whether a given possible steady state is stable or unstable with respect to small disturbances. They cannot be used to predict whether

the given state, if stable, will in fact be realized, nor can they be made to reveal what course the oscillation will take when it departs from an unstable steady state.

Corresponding to a given set of parameters, an oscillator may have two or more stable steady states, and which of these is realized depends on the history of the oscillator. A disturbance of sufficient magnitude may change the oscillation from one stable mode to another. Each possible steady state has associated with it a different set of transient normal modes, and although the constants of the network and amplifier may remain unchanged, yet each steady state represents in effect a different system, and requires its own characteristic equation to describe stability.

8. Symmetrical Networks

Symmetrical networks are special cases of general (asymmetrical) networks and the reason for considering the particular before the general is one of convenience. Because of the great simplification resulting from symmetry, the analysis of symmetrical networks can be carried out more easily and in greater detail than is feasible for the general type. Also the feedback networks of many practical oscillators are symmetrical.

An impedance is symmetrical with respect to a frequency ω_0 if the real part has even symmetry, and the imaginary part odd symmetry, about the line $p = j\omega_0$ in the p plane. Thus

$$Z(p^* + j\omega_0) = Z^*(p + j\omega_0)$$

No real impedance can be symmetrical about a frequency other than 0, but some networks such as high-Q resonant circuits have approximate symmetry over a limited range of p. For practical purposes a symmetrical impedance is therefore defined as one for which $|Z(p^* + j\omega_0) - Z^*(p + j\omega_0)|$ is negligible compared with the maximum value of $|Z(p + j\omega_0)|$ over the appropriate range of p. Since for all real impedances, $Z(p^* + j\omega_0) = Z^*(p - j\omega_0)$, it follows that for symmetrical impedances $|Z(p + j\omega_0) - Z(p - j\omega_0)|$ is also negligible.

If the impedances of an oscillator feedback network are symmetrical with respect to ω_0 then $Z_i^- = Z_i(p - j\omega_0) = Z_i(p + j\omega_0) = Z_i^+$, with similar expressions for $Z_0^{\pm}$, etc. It follows that the modified impedances Z_I^+, etc., are also symmetrical. When $p = 0$, Z_I, etc., are resistances, and so from (4.17) $\theta = 0$ and $q = \exp. j\theta = 1$.

Inspection of expressions (6.3) then shows that $a_1 = c_{-1}$, $b_1 = b_{-1}$, $c_1 = a_{-1}$, $a_0 = c_0$. The determinant for $D(p)$ becomes, from (6.4)

$$D(p) = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_0 & b_0 & a_0 \\ c_1 & b_1 & a_1 \end{vmatrix}$$

This can be factorized as follows:

$$D(p) = F(p).A(p) \quad .. \quad .. \quad (8.1)$$

$$\left.\begin{aligned} \text{where} \quad F(p) &= a_1 - c_1 \\ A(p) &= b_0(a_1 + c_1) - 2a_0b_1 \end{aligned}\right\} \quad (8.2)$$

$D(p)$ is zero if either $F(p)$ or $A(p)$ is zero and the number of encirclements of the origin made by the locus of $D(p)$ is the sum of the encirclements by $F(p)$ and $A(p)$. Since neither of these factors has a pole within the contour of p any encirclement must be clockwise. Hence the Routh-Nyquist stability criterion is that the locus of neither $F(p)$ nor $A(p)$ should enclose the origin.

To see the physical significance of these two conditions, the transient grid voltage, which is the real part of (5.2), is added to the steady state voltage (4.1). The high-frequency part of the total voltage is

$$V_{g1}\{1 + m_1 \exp. at \cos(\omega_d t + \phi_1)\} \times \cos\{\omega_0 t + m_2 \exp. at \cos(\omega_d t + \phi_2)\} \quad .. \quad (8.3)$$

$$\left.\begin{aligned} \text{where } m_1 \exp. j\phi_1 &= (u_1 + u_{-1})\, V_d/V_{g1} \\ m_2 \exp. j\phi_2 &= -j(u_1 - u_{-1})\, V_d/V_{g1} \end{aligned}\right\} \quad (8.4)$$

(8.3) is the expression for a wave modulated in both amplitude and phase (or frequency). The two modulations have the same complex frequency $p = a + j\omega_d$, and the coefficients of modulation may be taken as the complex amplitudes (8.4).

Using expression (8.2), the equations (6.2) for u_n can be transformed to the equivalent set

$$\left.\begin{aligned} (u_1 - u_{-1})F(p) &= 0 \\ (u_1 + u_{-1})A(p) &= 0 \\ (u_1 + u_{-1})a_0 &= -b_0u_0 \end{aligned}\right\} \quad .. \quad .. \quad (8.5)$$

Suppose that for some particular value, p_1, $F(p_1) = 0$ but $A(p_1) \neq 0$. Then $u_1 + u_{-1} = 0$ and from (8.4) $m_1 = 0$. Also $u_0 = 0$, so there are no voltages or currents of frequency p. (8.3) shows that the transient disturbance takes the form of a frequency modulation of the steady-state oscillation. Hence if $F(p)$ encircles the origin, p_1 has a positive real part and the oscillation frequency is unstable.

Similarly if $A(p_2) = 0$ but $F(p_2) \neq 0$ then $u_1 - u_{-1} = 0$ and $m_2 = 0$. This corresponds to amplitude modulation of the steady-state oscillation, and if $A(p)$ encircles the origin the amplitude is unstable. The third equation in (8.5) shows that there is now a low-frequency component in the total voltage; i.e., the grid-bias voltage is also modulated.

Thus in a symmetrical-network oscillator the transient takes the general form of independent modulations of frequency and amplitude, the characteristic modulation frequencies being the roots of $F(p) = 0$ and $A(p) = 0$ respectively. This independence holds only for small disturbances. If the frequency is unstable the disturbance

will eventually become large enough to produce a sensible change of amplitude. On the other hand large changes of amplitude can take place without affecting the frequency.

In the previous Section it was pointed out that $a_1 = c_1$ when $p = 0$. Hence and from (8.2), $F(0) = 0$. It can also be shown that $F(p) \to 1$

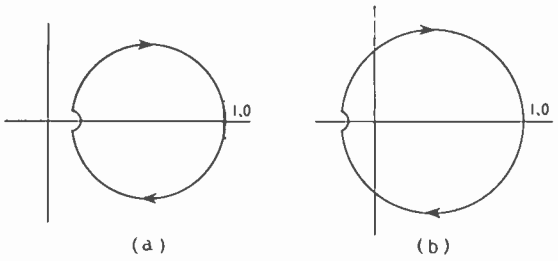


Fig. 9. *Loci of* $A(p)$.

and $A(p) \to 1$ when $p \to \infty \pm \frac{1}{2}j\omega_0$. The loci of $F(p)$ have the same general form as those of Fig. 8 and, using the same argument as for $D(p)$, the frequency can be stable only if $F'(0) > 0$. Typical loci for $A(p)$ are shown in Fig. 9. $A(0)$ is either positive or negative, for the condition $A(0) = 0$ would require a critical adjustment of parameters. Inspection of Fig. 9 shows that the amplitude can be stable only if $A(0) > 0$. The locus then encircles the origin an even number of times or not at all. Similarly if $A(0) < 0$ $A(p)$ makes an odd number of encirclements.

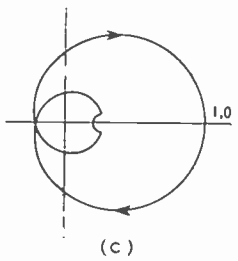
In terms of $D(p)$, $D'(0) = F'(0)\ A(0)$. $D'(0)$ is positive if $F'(0)$ and $A(0)$ have the same sign and it is clear that this must be positive, for otherwise both amplitude and frequency would be unstable and $D(p)$ would make at least two encirclements.

The various forms of instability may conveniently be classified according to the signs of $F'(0)$ and $A(0)$. Starting with $A(0)$, the simplest type of instability occurs when $A(0) < 0$ and the locus makes a single encirclement. This means that the equation $A(p) = 0$ has one real positive root and the disturbance therefore takes the form of a unidirectional movement of the amplitude away from the steady-state value. Instability of this type will be described as 'aperiodic'.

The next simplest kind of instability occurs when $A(0) > 0$ and the locus makes two encirclements. The equation $A(p) = 0$ has now two real positive roots or two complex conjugate roots with positive real parts. In the latter case the amplitude is modulated by an exponentially-expanding sine wave, and in the former the amplitude changes have the highly non-sinusoidal form typical of relaxation oscillations. In both cases the instability is 'periodic'. A triple encirclement would indicate a combination of periodic and aperiodic instabilities, and so on for more complicated loci.

Similarly the frequency has aperiodic instability if $F'(0) < 0$, and periodic instability if $F(p)$ makes two encirclements.

Aperiodic instability results in the various hysteresis effects some of which were studied by van der Pol and others. The various forms of frequency and amplitude instability which may occur in symmetrical-network oscillators are now studied in some detail.



9. Frequency Stability

The frequency stability function $F(p) = a_1 - c_1$ may be written in terms of the network and amplifier parameters by substituting in expressions (6.3) for $S_0 - S_2$ according to (5.5), for $Z_t{}^+$, etc., after the manner of (4.11), and for $G_0 - G_2$ according to (5.14). This gives

$$F(p) = \{1 + (Z_0{}^+ - kZ_t{}^+)/r_g\}(1 - Z_E{}^+/R_E) \quad (9.1)$$

where $Z_E{}^+ = (Z_T{}^+ + Z_I{}^+/\mu)/(1 - kZ_T{}^+/r_g)$

The low-frequency impedances $Z_t{}^0$, etc., do not appear. This is to be expected since if only the frequency is modulated there are no voltages or currents of frequency p. Also the only parameters of the amplifier which appear are k, μ and the constants which help to determine r_g. It would seem that frequency stability is independent of the non-linear amplifier characteristic. However, in deriving these results it was assumed that the amplifier currents were single-valued functions of the voltages, and the conclusions will therefore be valid only for such types of amplifier.

From (9.1), $F'(0) = -\{1 + (R_0 - kR_t)/r_g\}$ Z_E'/R_E,

and since R_t and R_E are negative the criterion for aperiodic stability is $Z_E' > 0$. This can be expressed in terms of real frequencies as follows: $Z_E{}^+ = R_E{}^+ + jX_E{}^+$, but the components are not themselves functions of p. However, as $Z_E{}^+$ is analytic it follows from the Cauchy-Riemann definition that at any point on the real frequency axis,

$(d/dp)\ Z_E\ (p + j\omega_0) = (d/dj\omega_d)\ R_E\ (j\omega_0 + j\omega_d)$
$+ (d/dj\omega_d)\ jX_E\ (j\omega_0 + j\omega_d)$

Since $Z_E{}^+$ is symmetrical, $R_E{}^+$ is an even function of ω_d and its derivative at $\omega_d = 0$ is therefore 0. Hence $Z_E' = X_E'$ and

$$F'(0) = -\{1 + (R_0 - kR_t)/r_g\}\ X_E'/R_E \quad (9.2)$$

where X_E' denotes the derivative of X_E with

respect to ω_d at $\omega_d = 0$. The criterion for aperiodic stability is then

$$-X_E'/R_E > 0 \text{ or } X_E' > 0 \quad \ldots \quad (9.3)$$

Another useful form is $F'(0) = \lim_{p \to 0} F(p)/p \quad (9.4)$

Inspection of Fig. 8(b) shows that the simplest type of network showing aperiodic instability has a transmission characteristic with two peaks of maximum response and a minimum at the oscillation frequency. It can be shown that there are three possible steady states, the frequencies of the other two corresponding to the point 'f' and that these are stable. (See Section 14.) When the oscillation frequency departs from the unstable value ω_0 it finally settles down at one of these points.

The double encirclement of Fig. 8(c) represents a new type of instability. Assuming the corresponding roots of $F(p) = 0$ to be complex conjugate, the oscillation frequency is modulated by an exponentially-expanding sine wave. The transmission characteristic has three maxima the smallest being at ω_0. There are five possible steady-state frequencies but only the stability of the central frequency can be discussed here, since it is only with respect to this frequency that the network is symmetrical. Of the other frequencies it can be shown that the two corresponding to the point 'h' are unstable, and the two corresponding to 'g' are stable.

Under suitable conditions a sustained periodic frequency modulation of small magnitude can be produced. When the parameters are adjusted to well beyond the critical values the inevitable slight asymmetry favours one or other of the component frequencies $\omega_0 \pm \omega_d$, and the oscillation eventually settles down at one of the stable points 'g'.

The network of Fig. 10, though not representing any practical oscillator, is the simplest in which the two forms of instability can be demonstrated. The central circuit resonates at ω_0 and has a shunt resistance nR; the other two resonate at $\omega_0 \pm \omega_m$ and have a resistance R. For simplicity it is assumed that the coils have the same Q-factor, that the coefficient of coupling between the two coils in each of the three pairs is 1, and that the voltage transformation ratio r is the same for each pair. No grid bias arrangement is shown as the result is independent of the particular way in which the bias voltage is obtained. Then

$$Z_t^+ = Z_i^+/r,\ Z_0^+ = Z_i^+/r^2,\ Z_n^+ = 0$$

Let $T = 2Q/\omega_0$ and $c = T\omega_m \quad \ldots \quad (9.5)$

Then

$$\frac{Z_i^+}{R} = \frac{(2+n)(1+pT)^2 + nc^2}{(1+pT)\{(1+pT)^2 + c^2\}} \quad \ldots \quad (9.6)$$

This is an approximation valid for large values of Q. Substituting for Z_i^+, etc., according to (9.5) and (9.6)

$$F(p) = 1 - Z_i^+/R_i = \frac{pT\{p^2T^2 + 2pT(1-bc^2) + 1 + (1-4b)c^2\}}{(1+pT)\{(1+pT)^2 + c^2\}} \quad \ldots \quad (9.7)$$

where $1/b = 2 + n + nc^2$.

The simplest oscillator in which aperiodic instability can exist has a two-circuit feedback network obtained by eliminating the central coil in Fig. 10. This is equivalent to putting $n = 0$. Then from (9.4),

$$F'(0) = T(1 - c^2)(1 + c^2).$$

For aperiodic stability

$$c^2 < 1;\ \text{i.e.,}\ 2Q\omega_m/\omega_0 < 1 \quad \ldots \quad (9.8)$$

Similarly for the triple-circuit network the criterion for aperiodic stability is $1 + (1 - 4b)c^2 > 0$. This is more conveniently expressed as a criterion for instability by substituting for b, thus

$$2(c^2 - 1)/(c^2 + 1)^2 > n > -2/(c^2 + 1) \quad (9.9)$$

A negative value of n could be realized by reversing the leads to the central coil but, when this is done, the equations remain valid only if μ and r_g are very large.

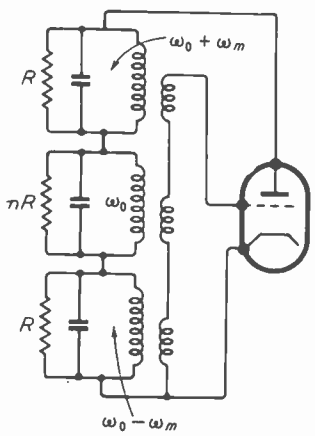


Fig. 10. *Circuit for frequency instability.*

Since the numerator of (9.7) is a quadratic it is not difficult to find the conditions for the roots of $F(p) = 0$ to have negative real parts. More directly, the Routh-Hurwitz stability criteria are: $1 + (1 - 4b)c^2 > 0$ and $1 - bc^2 > 0$. The first is simply the condition for aperiodic stability, the second is the condition for periodic stability. Substituting for b and using (9.9) this can be more easily written as a criterion for periodic instability:

$$(c^2 - 2)/(c^2 + 1) > n > 2(c^2 - 1)/(c^2 + 1)^2 \quad \ldots \quad (9.10)$$

Periodic instability can exist only if the left-hand

side of this inequality is greater than the right-hand side, and this requires that $c > \sqrt{3}$.

Further examination of (9.7) reveals the nature of these unstable states. If $n > (c^2 + 1)^{-\frac{1}{2}} - (c^2 + 1)^{-1}$ the roots of $F(p) = 0$ are complex conjugates, and for smaller values of n both roots are real.

When $n < -2/(c^2 + 1)$ it can be shown that R_t is positive and no oscillation is then possible. From (9.9) and (9.10) the stability criterion is therefore

$$n > (c^2 - 2)/(c^2 + 1) \text{ if } c > \sqrt{3} \quad .. \quad (9.11)$$
$$\text{and } n > 2(c^2 - 1)/(c^2 + 1)^2 \quad c < \sqrt{3}$$

All other possible values of n correspond to either periodic or aperiodic instability.

10. Amplitude Stability—Hysteresis

In Section 8 it was shown that the condition for the amplitude to have aperiodic stability is $A(0) > 0$. The behaviour of the oscillator is investigated by writing $A(0)$ in terms of the network and amplifier parameters. It is first supposed that the grid-bias voltage is derived entirely from the flow of grid current, so that S_1, etc., may be substituted according to (5.5). Using also (4.11), (5.14), (6.3) and (8.2), $A(0)$ becomes

$$\begin{aligned} A(0) = {} & 2(1 + S_0R_g)(1 + R_0/r_g)[(R_T + R_l/\mu) \\ & \{G_2 + (G_0G_2 - G_1^2)R_a/\mu\} \\ & + G_1R_T(1 - kR_a/\mu R_g)(dV_g/dV_{g1})] \\ & - (1 + S_0R_g)(1 + R_0/r_g)(V_{g1}/r_g)(dr_g/dV_{g1}) \\ & \times[(1 + G_0R_a/\mu)(R_\theta - kR_T)/r_g + \\ & \{(G_0 + G_2)(1 + G_0R_a/\mu) - 2G_1^2R_a/\mu\}R_NR_T/\mu r_g] \\ & .. \quad .. \quad .. \quad .. \quad (10.1) \end{aligned}$$

The network elements appearing in this expression are all resistances. Time constants have no influence on aperiodic stability: their sole effect is to limit the speed with which the amplitude moves away from an unstable value.

$A(0)$ has been written in this form in order to show that an important factor in stability is the manner in which the grid-bias voltage and grid input resistance vary with oscillation amplitude. The derivatives dV_g/dV_{g1} and dr_g/dV_{g1} are strictly defined only when V_{g1} changes infinitely slowly, but this does not mean that the criterion is valid only for slow changes. The derivatives appear because they are related to S_0, S_1, and S_2 through the equations defining the steady-state values of V_g and r_g, and it is these relations which have been used in (10.1).

In many oscillators the grid-current characteristic is such that when V_{g1} is moderately large, V_g is almost proportional to V_{g1}, and r_g is almost constant. Then $dV_g/dV_{g1} = V_g/V_{g1}$ and $dr_g/dV_{g1} = 0$.

Substituting for K according to (4.18), and for R_E and then R_T from (4.14) and (4.11) gives

$$\begin{aligned} A(0) = {} & -2(1 + S_0R_g)\{1 + (R_0 - kR_l)/r_g\}R_E \\ & \times\{KG_1 - G_2 + (G_1^2 - G_0G_2 - kKG_1/R_g)R_a/\mu\} \\ & .. \quad .. \quad .. \quad .. \quad (10.2) \end{aligned}$$

Since R_E is negative the criterion for stability is

$$KG_1 - G_2 + (G_1^2 - G_0G_2 - kKG_1/R_g)R_a/\mu > 0 \quad .. \quad .. \quad .. \quad .. \quad (10.3)$$

If R_a is small or μ large this reduces to

$$KG_1 - G_2 > 0 \quad .. \quad .. \quad .. \quad (10.4)$$

This criterion can be interpreted in terms of the slope of the graph of Y plotted against N of which the curves of Fig. 3 are particular examples. By differentiating equations (4.8) and (4.13) with respect to Y, keeping K and V_{ca} constant, and using also (5.10) and (5.14), it can be shown that

$$\partial N/\partial Y = 2N^2(KG_1 - G_2)/gY$$

Since g is positive, $\partial N/\partial Y$ has the same sign as $KG_1 - G_2$, and so for stability, $\partial N/\partial Y > 0$. In deriving this result it was assumed that V_{ca} is independent if V_{g1}, but this is true only if $R_a = 0$ or $\mu = \infty$, for otherwise a change in V_{g1} would change the mean anode current which would change the mean anode voltage which in turn would change V_{ca}. A straightforward, but tedious calculation shows that when V_{ca} varies, $\partial N/\partial Y$ has the same sign as the l.h.s. of the more general criterion (10.3).

These results are independent of the form of the amplifier characteristic provided only V_g/V_{g1} is constant. For a three-halves-law amplifier G_1 and G_2 can be expressed in terms of a parameter $H = 1 - K + K/Y$. As G_1 is now positive (10.4) can be written $G_2/G_1 < K$, and the corresponding value of H obtained from Fig. 6. A more direct method is to substitute for G_2/G_1 according to (5.17) and to reverse the series thus obtained. (10.4) then becomes

$$1/Y > \{1 + (1 - K)/9 + ..\}(1 - K)/3K \quad (10.5)$$

Since Y cannot be negative the amplitude is stable for all values of Y when $K > 1$. When $K < 1$ the stable values of Y are less than the critical value given by turning the inequality (10.5) into an equation.

In terms of the graphs of Y against N, the greatest stable amplitude occurs when $\partial N/\partial Y = 0$; i.e., where N is also a maximum. It can be shown that this value is

$$N_{max} = (27/32)^{\frac{1}{2}}(1 - K)^{-3/2}\{1 + (1 - K)/6 + ..\} \quad .. \quad .. \quad .. \quad .. \quad (10.6)$$

These tendencies are suggested in the graphs of Fig. 3 and shown clearly in Fig. 11(a). The full line is the theoretical value of Y for $K = 0{\cdot}6$. Once Y has passed the critical point 'b' there is nothing in the theory to prevent its increasing indefinitely. Actually the amplitude is then limited by the diversion of anode current to the

grid or screen. The broken line shows a possible form of Y due to this effect. In most oscillators this limitation occurs before the critical value of Y is reached.

A hysteresis effect exists in the region 'abcd'. As N is varied Y moves round the boundary in the direction shown, the portions 'bc' and 'da' being irreversible. The second critical point 'd' does not lie on the theoretical curve and the calculation of the second critical amplitude is beyond the scope of the present analysis. Although points on both branches 'ab' and 'cd' represent stable states it is obvious that an impressed force of sufficient magnitude could shift the operating point from one branch to another.

A second kind of hysteresis effect can exist when r_g and V_g/V_{g1} vary considerably with change of V_{g1}. This variation is most marked for small values of V_{g1}; i.e., near to the threshold of oscillation. Fig. 11(b) shows the effect of varying a parameter of the feedback network; e.g., the mutual inductance M between anode and grid circuits. Oscillation begins when M is increased to the critical point 'b', but small amplitudes are unstable and V_{g1} immediately moves to the point 'c'. When M is reduced a second critical point 'd' is reached at which V_{g1} suddenly falls to zero. The effect is due mainly to the variation of r_g. If this increases with V_{g1} the loop gain of the amplifier and feedback network, measured at the oscillation frequency, may also increase and become sufficient to maintain an oscillation of large amplitude for a value of M less than that needed to initiate oscillation.

The condition for avoiding this kind of hysteresis effect is that vanishingly small amplitudes should be stable. It is sufficient to consider the particular case of a three-halves-law amplifier (4.21) with an exponential grid-current characteristic (4.23). The values of G_n for small values of V_{g1} are given by (5.16), and the derivatives of V_g and r_g by (5.8). Substituting these into (10.1), using also (4.18) and (5.14), and neglecting powers of V_{g1} higher than the square gives

$$A(0) = (b + c + d)f \quad .. \quad .. \quad .. \quad (10.7)$$

where $b = \dfrac{(1 - 3R_a/\mu R_E)\,(1 + R_I/\mu R_T)^2}{32\,(1/Y - 1)^2}$

$$c = n(1 - kR_a/\mu R_g)/4(1/Y - 1)$$

$$d = \tfrac{1}{2}m(1 - R_a/\mu R_E)\{R_E(k - R_\theta/R_T) + R_N/\mu\}/(1 + R_I/\mu R_T)r_g$$

$$f = 2(V_{g1}/V_g)^2(1 + S_0R_g)\{1 + (R_0 - kR_I)/r_g\}$$

Since $f > 0$ the stability criterion is $b + c + d > 0$. Substituting for m and n from (5.8) this becomes

$$1 - kR_a/\mu R_g + (1 - 3R_a/\mu R_E)(1 + R_I/\mu R_T)^2 (1 - V_g/V_0)(V_0/V_g)^2/4(1/Y - 1) - (1 + V_g/V_0)(1/Y - 1)(1 - R_a/\mu R_E)\{R_E (k - R_\theta/R_T) + R_N/\mu\}/(1 + R_I/\mu R_T)r_g > 0 \quad (10.8)$$

The values of the parameters are those at the threshold of oscillation. In many oscillators R_N is negligible. If also μ is large, then on substituting for R_E according to (5.16) the criterion simplifies to

$$1 + (1 - V_g/V_0)(V_0/V_g)^2\,4(1/Y - 1) + (k - R_\theta/R_T)(1 + V_g/V_0)(1 - Y)^{1/2}/gYr_g > 0 \quad .. \quad .. \quad .. \quad .. \quad (10.9)$$

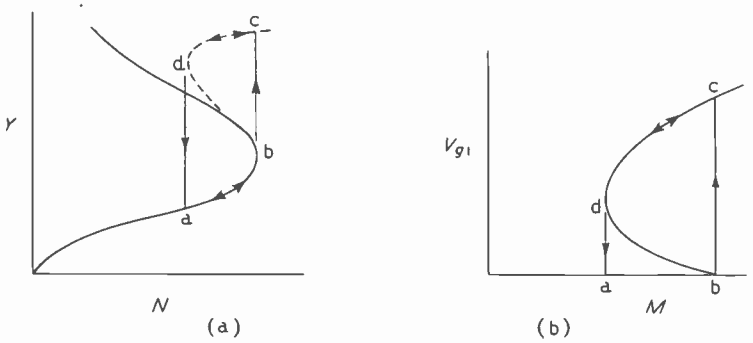


Fig. 11. *Amplitude hysteresis.*

In this expression R_T is negative, Y is small and positive, and g and $1 - V_g/V_0$ are also positive. It follows that instability can exist only if $1 + V_g/V_0 < 0$. Referring to (5.8) it is seen that this is also the condition that r_g should increase with V_{g1}.

In (10.9), k, V_0 and g are constants of the valve, and V_g, Y and r_g, which are given by (4.23) and Fig. 4, depend only on R_g and the valve constants. Hence the only two independently adjustable parameters are R_g and the ratio R_θ/R_T. In most practical situations $1 + V_g/V_0 < 0$. Stability is then assisted by making R_θ/R_T small. This can be achieved by placing the oscillatory circuit in the anode lead and using a small grid-coupling coil. The effect of varying R_g is less easy to follow but it can be seen that stability is obtained with all sufficiently large values of R_g, for V_g changes much less rapidly than r_g and so the l.h.s. of (10.9) can be made to approach 1. The damping due to grid current is then a negligible fraction of the total. Stability is also assured for values of R_g small enough to make $1 + V_g/V_0 > 0$ (Fig. 4). This form of instability has been studied experimentally by Zepler[32].

In the next type of oscillator to be considered the grid-bias voltage is fixed and large enough to stop grid current. The stability criterion follows

at once from (10.1) by putting $dV_g/dV_{g1} = dr_g/dV_{g1} = 0$, and is (since R_E is negative)

$$-G_2 + (G_1^2 - G_0G_2)R_a/\mu > 0 \quad .. \quad (10.10)$$

If R_a is small or μ large this reduces to

$$G_2 < 0 \quad .. \quad .. \quad .. \quad .. \quad (10.11)$$

Now for a three-halves-law amplifier $G_1^2 - G_0G_2 > 0$ (from (5.18)). Hence whatever the value of R_a, stability is assured if (10.11) is satisfied, and whatever the value of G_2 (10.10) can be satisfied by choosing a sufficiently large value of R_a.

Criterion (10.11) can be interpreted in terms of the graphs of N plotted as a function of Y and K. Since $R_a/\mu = 0$, V_{ca} is constant, and as V_g has been assumed constant Y is also constant. By differentiating (4.8) and using (5.10) and (5.14) it can be shown that $\partial N/\partial K = 2N^2G_2/gK$. Since g and K are positive $\partial N/\partial K$ has the same sign as G_2. If $R_a/\mu \neq 0$, V_{ca} varies with V_{g1} (i.e., with K) and a more lengthy calculation then shows that $-\partial N/\partial K$ has the same sign as the l.h.s. of (10.10).

From Fig. 3 it would seem that for the three-halves-law amplifier $\partial N/\partial K$ is always positive and all amplitudes therefore unstable. However, for small values of Y and N the graphs cross one another and $\partial N/\partial K$ becomes negative. This cannot be shown in Fig. 3 because the graphs would be too close to be distinguished.

Returning to (10.11), Fig. 6 shows that $G_2 < 0$ when $H = 1 - K + K/Y > 1{\cdot}42$. Since K is positive this inequality can be satisfied only if $Y < 1$; i.e., the grid-bias voltage must not exceed the cut-off value V_{ca}. Hence for stability $K > 0{\cdot}42Y/(1 - Y)$, and the corresponding stable amplitudes are

$$V_{g1} < (V_g - V_{ca})/0{\cdot}42\,(1 + R_I/\mu R_T) \quad (10.12)$$

All oscillations of stable amplitude will also be self-starting, for if this inequality is satisfied for the steady-state value it is satisfied for all smaller values of V_{g1}. A second restriction is imposed by the requirement that no grid current should flow. With the semi-linear form of grid characteristic (4.19) this means that

$$V_{g1} < V_{cg} - V_g \quad .. \quad .. \quad .. \quad (10.13)$$

Finally, the amplifier may operate with a fixed grid-bias voltage insufficient to prevent the flow of grid current. This mode of operation, like the previous one, is little used in practice but it is of theoretical importance in connection with periodic instability. If the bias voltage is to be independent of V_{g1} then $R_g = 0$. $A(0)$ can be obtained from (10.1) by substituting for dV_g/dV_{g1} and dr_g/dV_{g1} according to (5.5) and then letting $R_g \to 0$. For the present purpose it is sufficient to consider the simplified case where $R_a = R_N = 0$. Then

$$A(0) = 2\{G_2\,(R_I + R_I/u) + S_2\,(R_0 - kR_I)\} \quad .. \quad .. \quad .. \quad .. \quad (10.14)$$

Using (5.14) the stability criterion can be expressed as

$$G_2R_E + (S_2r_g + G_2R_E)(R_0 - kR_I)\,r_g > 0 \quad (10.15)$$

(To be continued)

CORRESPONDENCE

Letters to the Editor on technical subjects are always welcome. In publishing such communications the Editors do not necessarily endorse any technical or general statements which they may contain.

Multiloop Feedback Amplifiers

SIR,—I was very interested to see the correspondence following my papers "Nyquist's Criterion" and "Multiloop Feedback Amplifiers" in *Wireless Engineer* for October and November, respectively, of last year. Perhaps I should explain that my object in writing these papers was:—

(i) To develop a proof of the Nyquist stability criterion for single-loop amplifiers *from first principles* and without using contour integration in the complex plane.

(ii) To point out that an alternative formulation of Nyquist's criterion exists; viz., the 'minimum-phase criterion'.

(iii) To show that the same reasoning can be applied to derive an unambiguous stability criterion for multiloop amplifiers. (This criterion can, of course, also be formulated as a 'minimum-phase criterion'.)

With regard to (i), I still feel that the simplest derivation is provided by working from the gain-with-feedback expressions, $A/(1 - A\beta)$, etc., rather than by introducing the characteristic equation and then having to link up the zeros of the denominators of these expressions with the roots of the characteristic equation. Of course, any criterion for the stability of a linear system can be linked up with the zeros of Δ or Δ' (the mesh or nodal determinants of the system) since the absence of these zeros from the right-half p-plane is the *necessary and sufficient* condition for stability. I agree with most of B. D. Rakovich's remarks regarding the Routh-Hurwitz determinant method, but it was not the purpose of my papers to compare the Nyquist and Routh-Hurwitz criteria. However, it should be pointed out that criteria based directly on the characteristic equation, such as the Routh-Hurwitz, cannot be used with experimentally-determined loop-gain characteristics.

I cannot agree with B. D. Rakovich that his function (1) appropriate to my Fig. 1(c) (November 1954, *Wireless Engineer*), viz., $1 - A_1\beta_1 - A_1A_2\beta_3 - A_2\beta_2 + A_1A_2\beta_1\beta_2$, is exactly the same as the characteristic function of the system, Δ, since function (1) is in general a ratio of two polynomials in p whereas, excluding a possible pole at the origin, Δ is a polynomial in p. For a full evaluation of the significance of a Nyquist plot of function (1) the

possible positions of its poles must first be determined, as was done in my paper.

The connection between the functions obtained by the analysis of feedback amplifiers on a mesh or nodal basis on the one hand and by the block diagram method with A and β terms on the other may be found by comparing expressions for the same network function in the two systems. Thus, for example, the return difference* for a valve in the A_2 section of the amplifier of my Fig. 1(c) is

$$\frac{1 - A_1\beta_1 - A_1A_2\beta_3 - A_2\beta_2 + A_1A_2\beta_1\beta_2}{1 - A_1\beta_1} \quad \text{(a)}$$

in terms of the A and β terms and Δ/Δ° in terms of the determinant of the mesh equations Δ. Both Δ and Δ° are polynomials in p (ignoring for brevity a possible pole at the origin) and both the numerator and denominator in expression (a) are each ratios of two polynomials in p. However, as the poles of $1 - A_1\beta_1$ are also poles of $1 - A_1\beta_1 - A_1A_2\beta_3 - A_2\beta_2 + A_1A_2\beta_1\beta_2$ the zeros of $1 - A_1\beta_1 - A_1A_2\beta_3 - A_2\beta_2 + A_1A_2\beta_1\beta_2$ and the zeros of Δ coincide.

O. P. D. Cutteridge

Department of Electrical Engineering,
Faculty of Technology,
University of Manchester,
Manchester.

4th July 1955.

* H. W. Bode, "Network Analysis and Feedback Amplifier Design", pp. 46-49 (D. Van Nostrand, 1945).

High-Stability Oscillators

Sir,—In a letter to *Wireless Engineer* (May 1955), W. B. Bernard criticizes the statement in Gouriet's and in Clapp's papers concerning the influence of the L/C ratio upon the frequency instability caused by nonlinear effects in oscillators; the oscillator under consideration has large grid and anode capacitances C_g and C_a and a series-resonance circuit $L_0C_0R_0$ between grid and anode.

As I am very interested in high-stability oscillators and as my last paper (now in printing) deals with the above-mentioned series-resonance oscillator in a nonlinear state of operation, I would like to make one or two remarks on this subject.

With respect to the incorrectness of the division of the circuit elements between the maintaining circuit and the frequency-determining circuit, I am not far from sharing Mr. Bernard's doubts.

However, I cannot agree with his further reasoning as to the equivalence of the series-resonance oscillator and the ordinary Colpitts' oscillator with regard to the nonlinearity effects caused by the driving valve. I have no doubt that the former oscillator has distinct advantages over the latter, not only because of the more suitable values of components used, but owing to its essential superiority in respect of frequency stability: the attenuation effect for the higher harmonic currents, occurring in the $L_0C_0C_g$ limb of the series-resonance oscillator, is considerably stronger than that in the LC_g limb of the Colpitts' oscillator. The ratio of the grid voltage of harmonic V_{gk} to the anode voltage of harmonic V_{ak} (k being the order of harmonics for both circuits) can be expressed as follows:—

For L_0C_0 series-resonance circuit, Fig. 1,

$$\frac{V_{gk}}{V_{ak}} = \frac{-j\frac{1}{k\omega C_g}}{j\left(k\omega L_0 - \frac{1}{k\omega C_0} - \frac{1}{k\omega C_g}\right)} = -\frac{C_0}{C_g}\cdot\frac{1}{k^2\omega^2 L_0C_0 - \frac{C_0}{C_g} - 1}.$$

If $C_0 \ll C_g$, $\omega^2L_0C_0 \approx 1$, and for sufficiently high k we obtain

$$\left|\frac{V_{gk}}{V_{ak}}\right| \approx \frac{C_0}{C_g}\cdot\frac{1}{k^2}.$$

For Colpitts' circuit, Fig. 2,

$$\frac{V_{gk}}{V_{ak}} = \frac{-j\frac{1}{k\omega C_g}}{j\left(k\omega L - \frac{1}{k\omega C_g}\right)} = -\frac{1}{k^2\omega^2 LC_g - 1}.$$

Since $\omega^2LC_g \sim 1$, again for sufficiently high k we have

$$\left|\frac{V_{gk}}{V_{ak}}\right| \approx \frac{1}{k^2}.$$

When comparing both results, it is seen that the reduction factor of grid voltage harmonics for the series-resonance circuit is C_0/C_g.

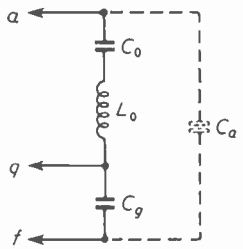


Fig. 1

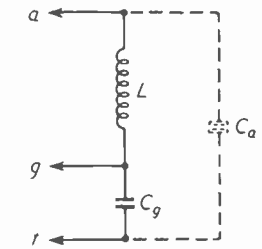


Fig. 2

The analysis based on the principle of the reactive power balance of harmonics enables one to establish in a series-resonance circuit oscillator an accurate formula for the frequency instability due to various factors. The partial instability caused by the nonlinearity of the anode-current characteristic is given by the expression

$$\left(\frac{\Delta\omega}{\omega}\right)_{nk} = \frac{1}{1 + \frac{C_0}{C_g}\left(\frac{C_0}{C_a}\cdot\frac{R_0^2\omega^2C_0C_g}{\mu - \frac{C_g}{C_a}}\right)} \times \sum_{k=2}^{\infty}\left(1 - \mu\frac{C_0}{C_g}\cdot\frac{n_k^2}{k^2-1}\right)\left(\frac{\Delta n_k}{n_k}\right).$$

The nonlinearity effect is expressed here by the content of anode-current harmonics $n_k = I_k/I_1$, (μ = amplification factor of the valve, R_0 = resistance of series $L_0C_0R_0$ circuit, $\omega/2\pi$ = fundamental frequency).

On the assumption that $C_a \approx C_g = C$, $\mu \gg 1$, $C_0 \ll C$, $4 < k^2 < \mu\, C_0/C_g$ the formula can be simplified:

$$\left(\frac{\Delta\omega}{\omega}\right)_{nk} \approx -R_0^2\omega^2C_0^2\frac{C_0}{C}\sum_{k=2}^{k'}\frac{n_k^2}{k^2}\left(\frac{\Delta n_k}{n_k}\right) = -\frac{1}{Q_0^2}\frac{C_0}{C}\sum_{k=2}^{k'}\frac{n_k^2}{k^2}\left(\frac{\Delta n_k}{n_k}\right),$$

Q_0 being the quality factor of the $L_0C_0R_0$ circuit.

The analogical expression for an ordinary Colpitts' oscillator is

$$\left(\frac{\Delta\omega}{\omega}\right)_{nk} \approx -\frac{1}{Q^2}\sum_{k=2}^{k'}\frac{n_k^2}{k^2}\left(\frac{\Delta n_k}{n_k}\right),$$

Q being the quality factor of the LC_aC_g circuit.

It is seen from these formulae that the instability of the series-resonance circuit can be made smaller because $Q_0 \gg Q$ and $C_0 \ll C$; it diminishes with the third power of C_0 (for R_0, ω and C constant).

Radio Institute, Janusz Groszkowski
Warsaw Technical University, Poland.

22nd May 1955.

SIR,—Mr. W. B. Bernard's letter in your May 1955 issue is unconvincing. I submit that the true reason for the dependence of frequency stability on L/C ratio is as follows:—

When in a condition of stable oscillation, the sum of the reactances and resistances shown in Fig. 1(b) of Mr. Bernard's letter must be zero. Both C_g and R_g are dependent on the frequency as well as on the amplitude of oscillation, the amplitude dependence arising from valve nonlinearity. L_0, R_0, and C_0 are completely independent of amplitude and to a first order of approximation, of frequency also.

At any given amplitude and frequency ($\omega/2\pi$), C_g and R_g are specified by the values of C_1, C_2 [Fig. 1 (a)] and the valve parameters. For operation at the frequency $\omega/2\pi$, the series combination of L_0 and C_0 must present a total reactance $\omega L_0 - 1/\omega C_0 = 1/\omega C_g$.

If, due to a change in operating conditions, a change occurs in the effective value of C_g, the reactance balance around the circuit must be restored by a change in frequency. A change in ω of $d\omega$ produces a change in reactance of the L_0 C_0 combination of $[L_0 + 1/\omega_0^2 C_0] \times d\omega$ so that the larger the value of L_0, and hence of the L_0/C_0 ratio, the smaller is the frequency change necessary to restore the reactance balance. Hence the greater frequency stability.

Farnborough, Hants. D. G. REID

12th July 1955.

Compression and Expansion of Programme Time

SIR,—The Editorial in the July issue of *Wireless Engineer* reminded me that some years ago I suggested* an electronic analogue of Gabor's film-and-slit method of frequency compression or expansion. Briefly, the idea was that the signal should be sent down a delay line, the velocity in which corresponds to the velocity of the tape or film in the mechanical methods. Attached to this line there would be a number of sampling devices, e.g., valves, at regular intervals and these would be activated by a scanning signal which progressed at a rate differing from that of the signal. While each sampling device was actuated it would reproduce unchanged the section of signal passing it, but transfer of the action to the next sampling unit would either repeat or omit part of the signal, according to the sense of the relative velocity of signal and sampling control. Since the device is electronic and contains no mechanical parts, the durations of successive samples could be made as short as desired. The sampling devices could also be made to fade in and out gradually if desired (cf. the use of graded slits in the optical method), so as to minimize discontinuities in the output.

The University, Birmingham. D. A. BELL

13th July 1955.

*British Patent No. 640015

V.H.F.–F.M. Broadcasting

SIR,—Your Editorial in the June issue says that "good f.m. reception is possible with lower field strengths than are necessary for good television pictures and so the need for an efficient aerial system is less".

This statement needs some qualification. It is probably true that a lower signal input to the receiver is necessary for the satisfactory reception of sound broadcasting in Band II using frequency modulation than for television broadcasting in Band I; many listeners will, therefore, be content with a simpler aerial for Band II than they would use for television in Band I. On the other hand, experience has shown that a field strength of 100 μV/m corresponding with peak white is generally acceptable for vision in Band I, whereas the B.B.C. regards 250 μV/m as the limit of the normal service area for Band II. This does not seem to be in accord with your statement as quoted above and some explanation may therefore be necessary.

Various factors enter into the situation, some technical, some psychological. The television service offers to the public a means of entertainment additional to the well-established sound broadcasting service and this no doubt accounts for the willingness of large numbers of viewers to purchase and install television receivers in areas in which the field strength is comparatively low and where appreciable interference and, possibly, fading are experienced. In fact, about three per cent of television licence holders in this country are receiving a field strength of less than 100 μV/m.

V.H.F. sound in Band II is, however, intended by the B.B.C. to provide a higher standard of reception, with almost complete freedom from interference, than that now provided by the existing long- and medium-wave sound-broadcasting services. A high standard is necessary since experience shows that ignition interference, for example, is much more disturbing in sound alone than when the viewing of a picture serves to distract the listener's attention from the sound. It is also intended that listeners, whenever practicable, shall be able to obtain good reception using simple indoor aerials.

Hence it is considered that, notwithstanding the considerable advantage of frequency modulation over amplitude modulation, 250 μV/m is necessary for a fully satisfactory sound service in Band II.

Perhaps a further qualification is necessary. Although good v.h.f. sound reception is possible with indoor aerials (or even with aerials built into the receiver) in areas where the field strength is reasonably high, efficient outdoor aerials are strongly to be recommended near the fringes of the service area and in built-up areas where ignition interference and multi-path effects can be troublesome.

E. L. E. PAWLEY

British Broadcasting Corporation,
Broadcasting House,
London, W.1.
14th July 1955.

[While it is true that the picture does distract the listener's attention from defects in television sound, ignition interference usually has a much greater nuisance value on vision than on sound. We still feel, therefore, that television calls for a good deal higher field strength than f.m. sound for an equally good performance in the two cases.

It may well be that a field strength of 250 μV/m is necessary for f.m. sound in order that a worth-while improvement over the medium- and long-wave transmissions may be obtained. Undoubtedly, the results are then good, but we should not class television reception with a field strength of only 100 μV/m as the same order of goodness, even if it is "generally acceptable".—W.T.C.]

Linear Phase Modulator

SIR,—A simple method of phase modulation has been devised by using the variation of phase of the output voltage of a feedback amplifier which occurs when the mutual conductance g_m of the valve is varied. It is found that when a large capacitance is connected between the grid and anode of an amplifier valve, the output phase varies linearly with g_m over a considerable range, thus giving a suitable means of phase modulating a signal.

If Z_1 be the anode–grid impedance and Z_2 be the anode load, then the output voltage is given by (Fig. 1)

$$E_0 = E_g \frac{(1 - g_m Z_1)Z_2}{Z_1 + Z_2}$$

provided $Z_1 Z_2 \ll (Z_1 + Z_2)r_a$ which is always the case in a multigrid valve.

We shall consider the simplest case, when $Z_1 = 1/j\omega C_1$ and $Z_2 = R_2$. The phase of the output voltage is given by

$$\tan \theta = \frac{g_m + 1/R_2}{\omega C_1 - g_m/\omega C_1 R_2} \quad .. \quad .. \quad (1)$$

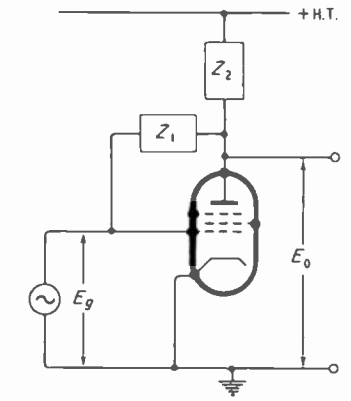


Fig. 1

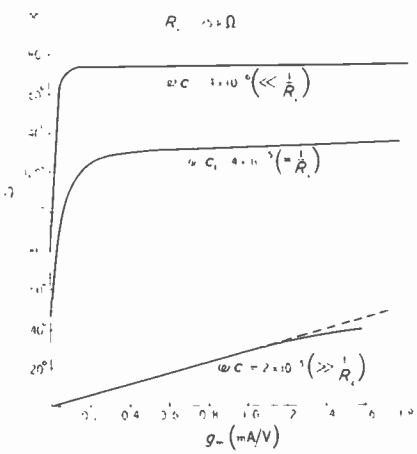


Fig. 2

Fig. 2 shows the variation of θ with g_m for the three cases, viz. (i) $\omega C_1 \gg 1/R_2$, (ii) $\omega C_1 = 1/R_2$, (iii) $\omega C_1 \ll 1/R_2$ at a carrier frequency of 160 kc/s. It is seen that in case (i), θ varies linearly with g_m for about 30° (about 0·5 radian). As it is possible to vary the g_m of a pentagrid converter valve (e.g., 6SA7) linearly with the oscillator grid bias, a linear phase modulation of 0·25 radian on either side may be easily produced by this method. Incidentally, it is half the amount of modulation produced in a conventional Armstrong modulator.

The method is much simpler than that of Armstrong and, employing two stages in cascade, it will give the same amount of modulation with crystal frequency stability.

The magnitude of the gain A of the amplifier is given by

$$|A| = \sqrt{\frac{g_m^2 + \omega^2 C_1^2}{1/R_2^2 + \omega^2 C_1^2}} \quad .. \quad .. \quad (2)$$

and is plotted in Fig. 3.

It is seen that for case (i) the change in $|A|$ with g_m is negligibly small and the consequent amplitude modulation may be eliminated easily.

The reflected input impedance due to the grid–anode coupling is given by $Z_g = \frac{1 - j/R_2\omega C_1}{g_m + 1/R_2}$, the magnitude of which is plotted in Fig. 4 for all the three cases. In order to prevent the variation of input loading with g_m and the consequent production of amplitude modulation, the source impedance should be made low, for example, by using a cathode follower. The phase of

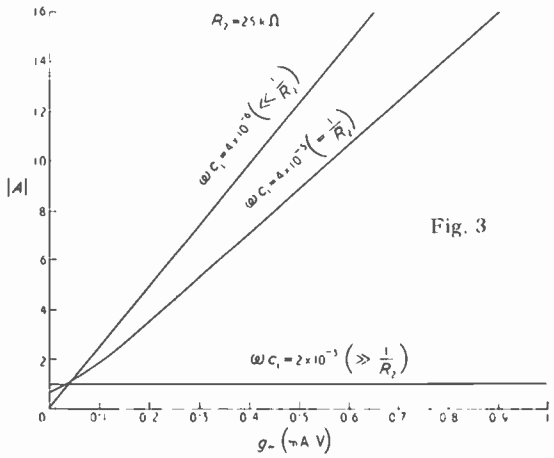


Fig. 3

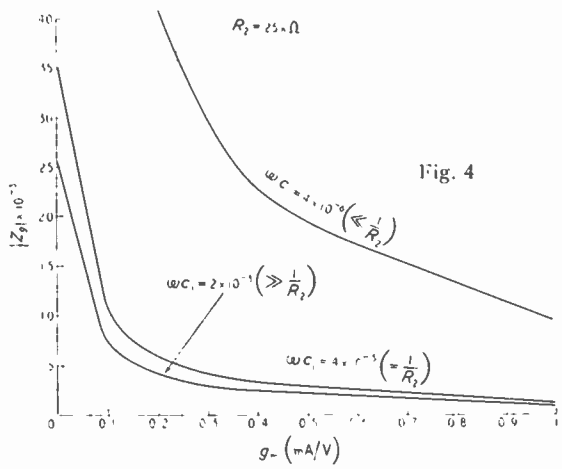


Fig. 4

the input remains practically constant with g_m. For $R_2 = 25$ kΩ, $\omega C_1 = 2 \times 10^{-3}$ and the source impedance = 1 kΩ, the input phase varies from zero to − 34′ only, for a variation in g_m from 0 to 1 mA/V.

It has been found that the addition of a few tens of pF of capacitance in parallel with the anode load has practically no effect on θ or $|A|$. The change in input phase is reduced still further. Thus the use of a capacitive resistance in the plate load (which happens always in practice) will not alter the situation as shown above.

G. S. Sanyal
B. Chatterjee

Indian Institute of Technology,
Kharagpur, India.
26th May 1955.

Rectifier-Filter Characteristics

SIR,—I should like to add a few remarks to the excellent article on rectifiers by Mr. Heymann in your June issue. Mr. Heymann states correctly that a constant direct voltage across the filter capacitor is a reasonable assumption for $2\pi fCR > 25$ (half-wave rectifier) or $2\pi fCR > 10$ (full-wave rectifier). It seems, however, that the use of this qualification in his approximate solution is over-restrictive in many cases and theoretically is not the most convenient one to use.

From Fig. 3 in Schade* the approximate solution is good in the following examples:

r/R	$2\pi fCR$	
1	2	r/R is the ratio of effective
1/2	4	rectifier resistance
1/5	10	to load resistance

From theoretical considerations the condition for the source voltage V to be in phase with the diode current i, as shown in Mr. Heymann's Fig. 2, is that the reactance of the capacitor be negligible in comparison to the diode resistance in series with it. Thus, a restriction on the approximate solution is:

$2\pi fCr \gg 1$ (half-wave)
$4\pi fCr \gg 1$ (full-wave).

where the lowest frequency is $2f$ in the full-wave case. Since normally $R > r$, the condition $X_c/R \gg 1$ will then also be satisfied (constant direct voltage).

From Schade, Figs. 3 and 4, a very mild inequality is required for accuracy consistent with the aims of this article:

$2\pi fCr > 2$ (half-wave)
$2\pi fCr > 1$ (full-wave).

This requires that $2\pi fCr$ vary from > 2 for $r/R = 1$ to > 20 for $r/R = 1/10$ in the half-wave case.

I should also like to point out that Mr. Heymann's methods of approximation may be applied to the case of valve-voltmeter rectifier efficiency considered by Mr. Scroggie in your February 1955 issue. An approximate solution to Mr. Scroggie's equation (1) is

$$\theta = \left(3\pi \frac{R_c}{R}\right)^{1/3}$$

and

$$\eta = \left|\frac{V_r}{E_{max}}\right| = 1 - 2{\cdot}22\left(\frac{R_c}{R}\right)^{2/3}$$

for the series-diode circuit. With R_S replacing R_c the above is a solution to his equation (4) for the shunt diode circuit, where θ is the conduction angle in radians, R_c the diode plus source resistance, R_S the source resistance, V_r the direct output voltage, and E_{max} the peak source voltage, all in Mr. Scroggie's notation.

This approximate equation for rectification efficiency, η, agrees with Mr. Scroggie's Fig. 7 to as closely as can be read on his graph. It agrees with his Fig. 6 to within 1% of η for $R_c/R \leqslant 0{\cdot}01$ and deteriorates in accuracy as θ and R_c/R increase.

MAURICE V. JOYCE

Polytechnic Institute of Brooklyn,
Brooklyn, New York, U.S.A.
12th July 1955.

* O. H. Schade, *Proc. Inst. Radio Engrs*, 1943, Vol. 31, pp. 341-361.

Differential-Amplifier Design

SIR,—I should like to reply to the letter by Mr. J. Ross Macdonald which appeared in your issue for July 1955, commenting on my article in the March 1955 issue.

It can be shown that the type of circuit described in reference 1 of Mr. Macdonald's letter (E.M.I. Laboratories, 1946) cannot be relied upon to have a transmission factor (as defined by Parnum, and in my article) greater than $\mu^2/\Delta\mu$, where $\Delta\mu$ is the difference between the amplification factors of the upper two valves in the circuit. Also, Mr. Macdonald is incorrect in supposing that the rejection ratio of two amplifier stages in cascade is the product of their separate rejection ratios. The calculation of the rejection ratio of a multi-stage amplifier, from the properties of the separate stages, has been treated by Parnum[1].

Mr. Macdonald and I have now exchanged some correspondence on this subject, and have agreed that his comments would be appropriate if rejection ratio were taken, as he took it, to mean the ratio of in-phase input signal to in-phase output signal. This ratio is important when a balanced output is required, and an accurately balanced output was the chief aim in the work described in Mr. Macdonald's references 2 and 3. The purpose of the circuits of my article was different, however.

A. M. ANDREW

Massachusetts Institute of Technology,
Cambridge,
Mass., U.S.A.
8th August 1955.

REFERENCE

[1] D. H. Parnum, "Transmission Factor of Differential Amplifiers", *Wireless Engineer*, 1950, Vol. 27, p. 125.

NEW BOOKS

Magnetic Amplifiers

By H. F. STORM. Pp. 545 + xix. Chapman & Hall Ltd., 37 Essex Street, London, W.C.2. Price 108s.

Mathematics of Engineering Systems

By DEREK F. LAWDEN, M.A. Pp. 380 + viii. Methuen & Co. Ltd., 36 Essex Street, London, W.C.2. Price 30s.

P.H. Brans Radio-Tube Vade Mecum. 12th Edition.

Pp. 381 + xxvi. P. H. Brans, Antwerp. Agents: Bailey Bros. & Swinfen Ltd., 46 St. Giles High Street, London, W.C.2. Price 27s. 6d.

Photo-Electric Handbook

By G. A. G. Ive. Pp. 152 + vii. George Newnes Ltd., Southampton Street, Strand, London, W.C.2. Price 17s. 6d.

Eine Anlage für Impuls-Code-Modulation

By Dr. Camillo Margna. Pp. 83. Verlag Leeman, Zurich. Price Fr. 8.30.

Second Thoughts on Radio Theory

By "CATHODE RAY". Pp. 409. Published for *Wireless World* by Iliffe & Sons Ltd., Dorset House, Stamford Street, London, S.E.1. Price 25s.

***Wireless World* F. M. Tuner**

By S. W. AMOS, B.Sc.(Hons.), A.M.I.E.E. and G. G. JOHNSTONE, B.Sc.(Hons.). Pp. 14. Published for *Wireless World* by Iliffe & Sons Ltd., Dorset House, Stamford Street, London, S.E.1. Price 2s.

Thermionic Valves 1904—1954: The First Fifty Years

Pp. 69. Institution of Electrical Engineers, Savoy Place, London, W.C.2. Price 4s. (members), 9s. (non-members).

The A.R.R.L. Antenna Book

Pp. 411. American Radio Relay League, West Hartford 7, Connecticut, U.S.A. Price $2.25.

Television, A World Survey

Pp. 51. Published for UNESCO by H.M. Stationery Office, P.O. Box 569, London, S.E.1. Price 3s.

The Suppressed Frame System of Telerecording

By C. B. B. Wood, E. R. Rout, A. V. Lord, B.Sc. (Tech.), A.M.I.E.E. and R. F. Vigurs.

B.B.C. Engineering Division Monograph No. 1. Pp. 15. B.B.C. Publications, 35 Marylebone High Street, London, W.1. Price 15s.

Questions and Answers on Radio and Television (5th Edition)

By E. Molloy. Pp. 148 + iv. George Newnes Ltd., Tower House, Southampton Street, London, W.C.2. Price 6s.

Practical Wireless Servicing Manual (10th Edition)

By F. J. Camm. Pp. 284. George Newnes Ltd., Tower House, Southampton Street, London, W.C.2. Price 17s. 6d.

Électronique Industrielle

By G. Goudet. Pp. 635. Editions Eyrolles, 61 Boulevard Saint-Germain, Paris Ve. Price 5,500 francs.

BRITISH STANDARD

Gramophone Records Transcription Disk Recordings and Disk Reproducing Equipment

B.S.1928:1955. Price 6s.

British Standards Institution, 2 Park Street, London, W.1.

NATIONAL BUREAU OF STANDARDS

Specification for Dry Cells and Batteries

National Bureau of Standards Circular 559. Pp. 17. Price 25 cents.

Experiments in the Computation of Conformal Maps

Edited by John Todd. National Bureau of Standards Applied Mathematics Series 42. Pp. 61. Price 40 cents.

Computer Development (SEAC and DYSEAC) at the National Bureau of Standards

Circular 551. Pp. 146. Price $2.

Obtainable from the Government Printing Office, U.S. Department of Commerce, Washington 25, D.C.

TECHNICAL LITERATURE

Audio Frequency Power Measurements

National Physical Laboratory. Notes on Applied Science No. 8. Pp. 16. H.M. Stationery Office, York House, Kingsway, London, W.C.2. Price 1s.

Joint Commission on Radio-Meteorology

Proceedings of the Third Meeting of the International Council of Scientific Unions. Pp. 30. Secrétariat-Général, U.R.S.I., 42 Rue des Minimes, Bruxelles, Belgium. Price 4s. (including postage).

Radio Research 1954

Report of the Radio Research Board. Pp. 47 + iv. H.M. Stationery Office, York House, Kingsway, London, W.C.2. Price 2s. 6d.

Nickel and Nickel Alloys in Electronic Valves

Pp. 22. Describes and illustrates some uses of nickel in electronic valves and includes tabular data on the physical and electrical properties of nickel. Henry Wiggin & Co. Ltd., Wiggin Street, Birmingham 16.

Patent Abstracts Journal

This journal is published weekly and contains abstracts of all new British patents. They are classified into three groups: general and mechanical, chemical, electrical. Each group has a subject index and there are also name and class indexes.

The journal is published by the Technical Information Co., Newton House, Mount Street, Liverpool 1. The annual subscription is £26. The individual groups are available separately each with its own index, at £15.

CORRECTION

In "Radiation from Aerials" in the August issue, an error occurred in Equ. (23) on p. 225. This equation should read $P = P_s^*$. In Equ. (21) the symbol I_m was omitted from the second term on the right-hand side.

BRIT.I.R.E. MEETING

28th September. "Extending the Limits of Resistance Measurement using Electronic Techniques", by G. I. Hitchcox, to be held at 6.30 at the London School of Hygiene and Tropical Medicine, Keppel Street, Gower Street, London, W.C.1.

STANDARD-FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Values for July 1955

Date 1955 July	Frequency deviation from nominal: parts in 10^8		Lead of MSF impulses on GBR 1000 G.M.T. time signal in milliseconds
	MSF 60 kc/s 1429-1530 G.M.T.	Droitwich 200 kc/s 1030 G.M.T.	
1	−0·1	+2	+10·1
2	NM	+2	NM
3	NM	−2	NM
4	0·0	+2	+9·7
5	+0.1	+2	+9·6
6	−0.0	+1	+8·4
7	0·0	−1	+8·1
8	0·0	−1	NM
9	NM	−2	NM
10	NM	+3	NM
11	+0·1	+3	+8·5
12	+0·1	+2	+8·3
13	+0·1	−2	+8·2
14	+0·1	−2	+8·2
15	+0·1	+2	+8·6
16	NM	+1	NM
17	NM	NM	NM
18	+0·1	+3	+8·8
19	+0·2	−2	+8·9
20	+0·2	−3	+9·2
21	+0·2	−3	NM
22	+0·2	+3	+9·4
23	−0·3	−3	NM
24	−0·3	−3	NM
25	+0·2	+3	+10·4
26	+0·2	+4	NM
27	+0·2	+3	NM
28	+0·2	+4	+10·0
29	+0·3	+4	+12·2
30	+0·3	+3	NM
31	+0·3	+4	NM

The values are based on astronomical data available on 1st August 1955.
NM = Not Measured.