THE RADIO AND ELECTRONIC ENGINEER

The Journal of the British Institution of Radio Engineers

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

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

VOLUME 25 FEBRUARY 1963 NUMBER 2

NEW LIGHT ON AN OLD SUBJECT

THE most recent toy of the electronic engineer is the optical maser or laser. It provides a number of morals which are applicable to research workers, engineers, publishers and even politicians. It is interesting to see why this is so and, one hopes, what the morals are.

In the first place the laser is interesting, not for itself, but because it is a scientific advance of the first magnitude which has been discovered sufficiently recently to make possible an exhaustive investigation of its literature. The first published account of the device is generally ascribed to Schawlow and Townes and occurred in December 1958. This was followed, in the same month, by a popular exposition in *Scientific American* and the subsequent history of laser publications is shown in the following table:

Year	Fundamental Papers	Engineering Papers	Applications Papers	Popular Expositions	Total
1958	1	—	—	1	2
1959	7	1	—	2	10
1960	7	3	—	10	20
1961	37	25	27	74	163
1962	48	53	34	171	306

The interesting lesson for the electronics research worker, and for the electronic engineer, is that all of the initial scientific publications in 1958 and 1959 were made in physical journals and that, similarly, no early popular exposition was printed in an electronic, or even electrical "glossy" magazine.

One cannot help wondering if the earlier descriptions of this new physical device, in a place and forum which would have been readily available to the engineering profession, might have shortened the "incubation period" 1958–61. Is it possible that the upturned noses which tend to greet the popular exposition are, in fact, the forces of reaction which slow down scientific progress?

The second moral is one for the publisher and for the student of information storage and retrieval. It is revealed by a comparison of the ratios of the various types of publication

to the totals; from this it is seen that popular expositions form a steady proportion of about one-half whereas fundamental papers (as might be expected) have decreased steadily in proportion so that they now form only about one-sixth of the total. The trends in engineering and in applications are also interesting, although predictable; engineering study came after the basic scientific work and applications followed the reliable groundwork provided by the engineer.

Perhaps the most important feature is one not revealed by the mere statistics, it is that enormous duplication of effort and of publication occurs. Not only do the same people republish their material in several journals, but the same work is duplicated in several laboratories.

All of this means that the problem of literature search is becoming more and more impossible for old established subjects and that not only must something be done to bring important new facts rapidly to the attention of the engineer, but also mechanical retrieval systems are becoming a necessity.

These two morals lead automatically to the third: for our politicians. The considerable publicity recently given to the National Lending Library for Science and Technology may suggest that all is well. This, however, is not so. The new library, whilst using some automation, is almost entirely mechanical (as distinct from electronic) in its approach to information storage. The philosophy seems to be to store everything in its original form and to retrieve it by hand. It is true that ample space, or waste land, is available to make this possible at the present time, but it is equally clear, from the overall growth of publication revealed by the totals in our table, that this method is certainly not efficient even at present, and that it will be completely impracticable in the fairly near future.

Research in progress may lead to a computer-orientated method of information storage in which ideas, rather than texts, are preserved and unoriginal material is rejected, but this leads to the difficulty of deciding what is original. A more practical and immediate scheme is to ask authors to describe, in a single short sentence, each idea contained in their papers and which they conceive to be worthy of preservation for posterity. A fairly simple code of writing could be imposed on authors for these summaries, and this code could be so designed to permit machine reading and collation with the existing corpus of knowledge.

No doubt this, rather deflating, self-analysis would be unpopular with authors and publishers alike but, to avoid the sudden reduction of journal size which would result, an element of gamesmanship could be introduced. Thus the author need not submit his description of originality until after the paper had been accepted. Furthermore, the nature of the condensation then submitted would have no influence on the previously made decision to publish.

It should be clear that, not only would this system improve storage and retrieval, but it would also have a salutary effect upon the refereeing of papers intended for journals of professional standing.

A. D. BOOTH

U.D.C. 681.32

A Computer Fixed Store using Light Pulses for Read-out

By

G. R. HOFFMAN, Ph.D.†

AND

D. C. JEFFREYS, M.Sc.
(*Associate Member*)†

Summary: The store has an access time of 20 ns and can be interrogated every 100 ns. Information is stored on a punched card, or photographic matrix, and interrogation is by means of light pulses transmitted along optical systems composed of bunches of fine transparent optical fibres. Calculations show that a particular case is capable of storing $>10^8$ bits but practical considerations reduce this to $2{\cdot}5 \times 10^6$ bits in a storage area of 40×40 cm.

List of Symbols

- A area of fibre end (mm^2).
- B surface brightness of source (lumens/mm^2).
- C length constant of the fibre (mm).
- D signal to interference ratio.
- E total light flux generated by a source.
- H length of the sides of a square storage plane (mm).
- K fraction of the incident light, reflected at a detector or source.
- L attenuation through the optical system.
- M number of detectors.
- N number of sources.
- S sensitivity of a detector (amperes/lumen).
- T required output pulse amplitude (ampere).
- Z storage capacity.
- μ refractive index of glass.

1. The Fibre Optical System

Fibre optics[1] enables the light from a source to be propagated along any flexible path to a detector. The fibre must have a refractive index greater than that of the surrounding medium so that any rays of light entering its end will travel along its length by total internal reflection (Fig. 1).

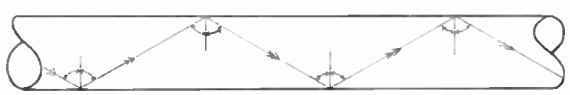


Fig. 1. Total internal reflection of a ray of light in an optical fibre.

If a number of such fibres is assembled together so that the inputs to the fibres are distributed over a circular area then the outputs may be distributed over a rectangular area. Furthermore, the rectangular area may be so drawn out that it approximates to a line distribution of the fibre ends. Light incident on the circular section will now be uniformly distributed over a linear section, the path between being flexible. A suitable bundle of fibres is shown in Fig. 2.

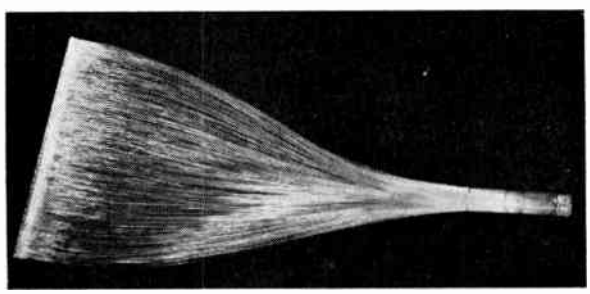


Fig. 2. Photograph of a suitable bundle of optical fibres.

Such a bundle would enable the light from a circular source to illuminate the holes along a row in a punched card. A number of similar bundles each with its source, assembled as shown in Fig. 3 would enable any of the rows of holes in the punched card to be illuminated whilst a similar set of bundles behind the card, lying along the columns enables the light from holes in each column to fall on a particular detector.

Thus by the selection of one light source, and observing the output of one detector it is possible to tell whether or not there is a hole in the card at the intersection of the two bundles, corresponding to a "1" or a "0". As the fibres may be small in diameter, high packing densities may be achieved and for a total storage of MN bits of information, M detectors and N light sources are required.

2. Row and Column Selection

2.1. *Signal Attenuation in the Selection System*

In common with other types of computer fixed store the energy produced in one source is distributed

† Electrical Engineering Department, Computing Machine Laboratory, University of Manchester.

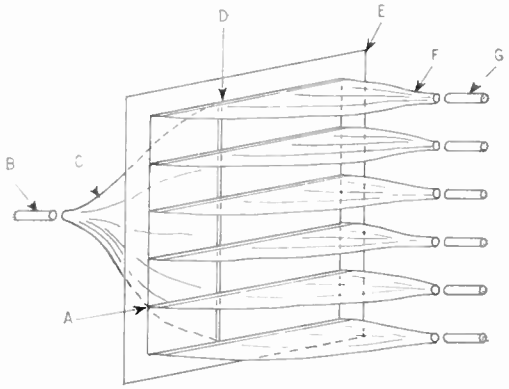


Fig. 3. The selection system

- A position of a row of holes.
- B detector.
- C bundle of fibres for the detector.
- D position of a column of holes.
- E storage plane.
- F bundle of fibres for the source.
- G source.

by the selection system so that the maximum energy which is available in any one detector from any one source is $1/M$ of the energy produced by the source. This is because the total light flux produced from the source is distributed along a line of M possible digit positions each with its associated detector so that the light flux incident on each digit position is only $1/M$ of the total flux transmitted. Thus in the absence of any other attenuation the total flux transmitted must be M times greater than that required to produce a readable pulse from the detector. This extra energy is available for causing interference signals.

The fibres themselves due to imperfections will also have light losses giving rise to further energy losses. A certain percentage will be lost at the entrance and exit from a fibre by surface reflection effects, but in general over a long fibre a greater percentage will be lost due to absorption in the bulk material of the fibre, scattering from occluded particles in the fibre, or total internal reflection losses due to scratches or imperfections in the cylindrical surface of the fibre.

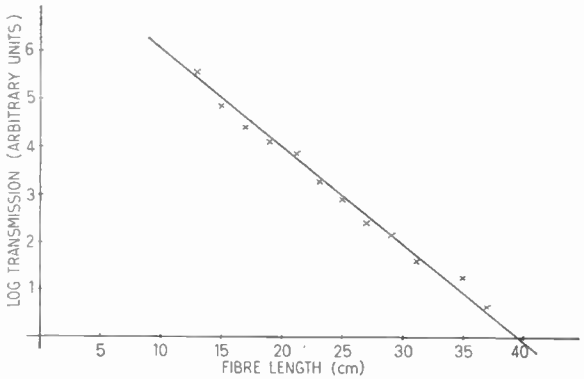


Fig. 4. The transmission of a polystyrene optical fibre.

If the fibre is homogeneous then there will be a uniform concentration of these imperfections throughout its length. This will give rise to a constant percentage loss per unit length of the fibre, and so the total loss, or attenuation of the fibre will be an exponential function of its length plus a small factor due to the reflection at the entrance and the exit from it (Fig. 4), that is

$$L = \exp\frac{x}{C}$$

where L is the attenuation of the fibre,

x is the fibre length,

C is the length of fibre which reduces the transmitted flux to 1/e and will be called the "length constant".

If the light flux which ultimately emerges from a bundle is only a small fraction of the total flux incident on the bundle, then the light which is scattered by the bundle may be picked up by other bundles giving rise to unwanted light signals in the optical system. This mechanism of interference can readily be eliminated by surrounding each fibre bundle by an opaque material thus absorbing this major source of interference light.

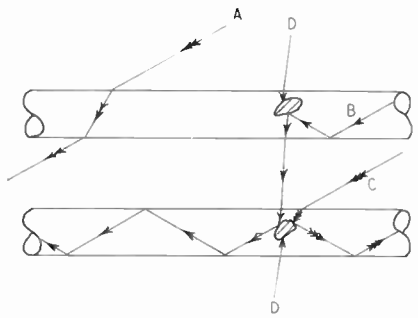


Fig. 5. Cross talk between fibres due to fibre imperfections

- A ambient light ray normally refracts into and out of the fibre.
- B light ray propagating in one fibre is transferred into the next at two imperfections.
- C ambient light ray is propagated in a fibre at an imperfection.
- D imperfections in the fibres.

Cross talk between individual fibres in a bundle can generally be ignored. For example all the fibres in a bundle associated with a particular source will carry the same light flux. A random process of exchange of flux between fibres will not disturb this condition. Fibres on the outside of a bundle will however tend to lose more flux to the opaque covering on the bundle than they receive from surrounding fibres and the uniformity of distribution of flux will be slightly disturbed.

Only a few $(1/M)$ of the fibres in a bundle of fibres associated with a detector will be carrying light flux. A fraction of this flux being lost due to imperfections in the fibre may be picked up by other fibres at their own imperfections and propagated in them (Fig. 5). If the direction of propagation of this cross-talk light is towards the detector then this will continue to produce the signal and there is no mechanism for interference, but if the direction is towards the source then there is a mechanism for interference.

This interference light will proceed back down the fibre system towards the mask and the sources. If it passes through holes in the mask then it will arrive back at a source. Out of the N sources available only one is illuminated for selection. By this mechanism light from the selected source can fall on others as if they were also illuminated, but at a much lower level. This could give rise to false signals in the detectors. A detector which should be receiving no light due to there being no hole in the mask can receive light by this mechanism of alternative paths.

Normally, however, ambient light incident on the cylindrical walls of a fibre will be refracted out again, (Fig. 5) so this interference effect is only of concern where there are adjacent imperfections in each of two fibres. In a practical system this mechanism of interference has proved impossible to detect under the worst conditions, using the maximum flux illumination, and the most sensitive available detector.

2.2. *The Interference Problem*

When a light pulse falls on a detector then there is a possibility that light will be reflected from the surface of the detector into fibres adjacent to the fibre carrying the light signal (Fig. 6). These fibres will also ultimately carry the light through holes in the mask to their corresponding sources, where further reflection will occur giving rise to a more serious source of interference signals.

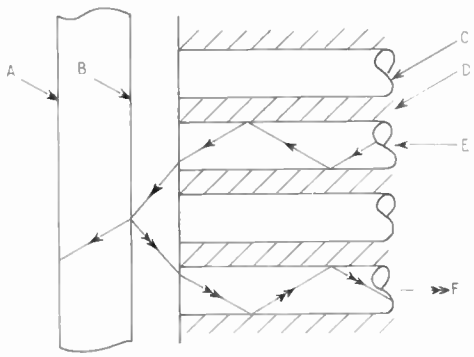


Fig. 6. The mechanism of reflection interference

- A photo-cathode of a photo-multiplier.
- B glass surface of a photo-multiplier.
- C optical fibre.
- D low refractive index binder.
- E light signal passing through a fibre to the photo-cathode.
- F interference signal reflected at the photo-multiplier surface.

Consider the worst case for this mechanism of interference, which is the reading of an $N \times M$ mask with only one bit position being opaque (x, y). When this bit is selected, there should be no signal from the detector. However, there are $(M-1)$ other detectors each fully illuminated. A certain fraction of the light incident on these will be reflected back into the fibre system, say K.

If the total incident light flux is E, and the attenuation of a fibre through the system is L, then the light flux incident on each detector is

$$\frac{E}{LM} \qquad \text{.....(1)}$$

The total reflected light flux is therefore

$$\frac{EK}{L} \cdot \frac{M-1}{M}$$

or if M is large, $\dfrac{EK}{L}$

When this flux passes back down the fibre system through the $(MN-1)$ holes in the mask towards the sources, it suffers further attenuation $L\dfrac{MN}{MN-1}$

The interference light flux returning to the sources is therefore $\dfrac{EK}{L^2} \cdot \dfrac{MN-1}{MN} = \dfrac{EK}{L^2}$ if M and N are large.

If a fraction K is now reflected at the sources back into the fibre system we have an incident interference light signal $\dfrac{EK^2}{L^2}$. This light signal will now pass back towards the detectors through all the holes in the mask suffering a further attenuation L. Assuming that all the light falling on unselected detectors is now perfectly reflected, and that there are no further fibre attenuation losses on the interference light then $\dfrac{1}{M-1}$ of this interference light will ultimately fall on the selected detector.

The worst possible condition for this interference light therefore gives a total interference flux of

$$\frac{EK^2}{L^3(M-1)} \text{ or if } M \text{ is large } \frac{E}{ML} \cdot \frac{K^2}{L^2}.$$

Equation (1) gives the normal signal incident on a detector, so that signal to interference ratio is

$$D = \frac{L^2}{K^2} \qquad \text{.....(2)}$$

Examination of this expression shows that D must be greater than 1. The optical fibres are bound to

have some attenuation so $L > 1$ and the system is similar to resistor-path fixed storage systems. However in this case, the most important factor is K, the reflection fraction.

Light energy which is incident on an efficient detector will be almost entirely absorbed to be converted to electrical energy. The fraction K will therefore be very small, and will be due almost entirely to the air-glass boundary at the surface of the detector. As 85% of the transmitted flux lies within 15 deg from normal in a typical fibre (Fig. 7), then for purposes of calculating K, normal incidence will be assumed.

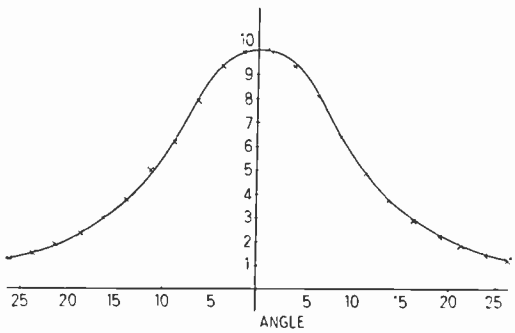


Fig. 7. The angular distribution of flux emerging from a typical fibre.

In the case of normal incidence[2]

$$K = \left(\frac{\mu - 1}{\mu + 1}\right)^2$$

where μ is the refractive index of the glass = 1·5 so $K = 4\%$.

Even if K is as high as 0·1, then the signal-to-interference ratio is still 100 in the absence of fibre attenuation. This almost total absorption of light at the surface of the detectors effectively produces a "virtual earth" at the termination of the light path, hence allowing negligible interference.

Three possible sources of optical interference have been considered:—

1. Cross talk between bundles.
2. Cross talk between fibres.
3. Reflection at input and output.

All these mechanisms have been shown to give negligible interference, the interference being eliminated primarily by absorption in the opaque covering around each fibre bundle and secondarily by the inherent absorption at each detector. Although attenuation due to the fibres assists in the elimination of interference it is not essential. It will be shown that this attenuation is the fundamental limitation to the size of the store, and is therefore undesirable.

3. Light Sources

3.1. *Discharge Devices*

In order to achieve maximum reading speed with the maximum digit packing density, the light sources should be capable of producing rapid high intensity luminous output pulses. At the same time, the drive requirements of the source should be compatible with normal computing machine pulse circuitry to allow economy of equipment.

Suitable sources may be found broadly in two groups:

1. Devices producing light from excited ions in a gas.
2. Devices producing light from phosphors.

The devices in the first group generally obtain their energy from charge stored in a capacitor at a high voltage. When the correct conditions are established a short, high intensity discharge occurs, the duration being determined by the parameters of the circuit. There will be a time delay involved for the storage of charge on the capacitor, which will govern the cycle time, and triggering with conventional computer circuits is difficult.[3] A thyratron type of discharge tube may be triggered more easily but its light output is low compared with devices in the second group.[4]

3.2. *Cathodo-luminescent Sources*

Cathodo-luminescent devices can be constructed[5] with phosphors having rise times better than 20 ns, and afterglow times better than 100 ns (Fig. 8). The rise time determines the read-out time, and the afterglow time determines the cycle time between sequential digits.

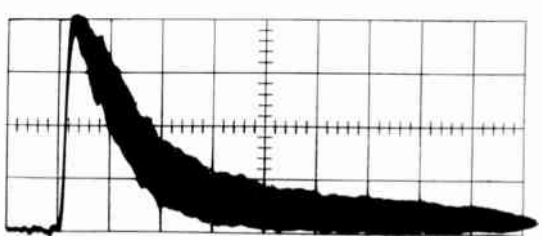


Fig. 8. Light output pulse from a high-speed cathodo-luminescent phosphor.
1 horizontal division = 50 ns.

Two cathode-ray tube systems may be used:

1. Individual cathode-ray tubes for each fibre optical bundle. This involves only a standard decoding system for the driving pulses to reach the tube corresponding to the correct address.

2. A large cathode-ray tube where individual areas on the screen are associated with separate fibre optical bundles. This system involves the use not only of

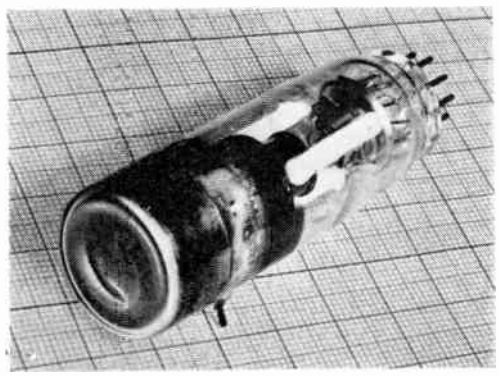


Fig. 9. Photograph of a small cathode-ray flash tube.

conventional digital selection between cathode-ray tubes, but also digital-to-analogue conversion to generate the deflection voltages to select the correct area on the screen. Although the system may be designed to have a high tolerance on deflection accuracy the time involved in setting up the deflection voltages would slow down the access time.

Thus it is better to use individual cathode-ray tubes for each selection optical system. A typical tube is shown in Fig. 9. This tube has a peak beam current of 100 mA, with an anode voltage of 15 kV giving a peak anode dissipation of 1·5 kW. Even with low efficiency phosphors the light output is very intense. Forced cooling of these cathode-ray tubes is necessary if high duty ratios are to be used. In order to achieve the maximum packing density in the store it would be necessary to construct similar flash tubes only 4 mm diameter. This would also have the effect of reducing the peak power dissipation, eliminating the necessity for cooling even with high duty ratios.

It is important to note that the power available in the light pulse does not have to be derived from the signal source as in many other forms of computer store. The power comes from the e.h.t. power supply which is a common unit to all cathode-ray tubes in the store. Expensive driving amplifiers are therefore eliminated. The drive pulses from conventional circuits may be applied to the control grid.

3.3. *Electroluminescent Sources*

Fixed storage systems using a punched card or photographic mask as the storage plane, and an electroluminescent matrix for scanning have already been described[6,7] but efficient electroluminescent phosphors have very slow light decay times (Fig. 10). A simplified scanning system may be achieved using a specially fabricated configuration of electroluminescent lines (Fig. 11). This system, whilst having slow operating times, would be very economical to construct, requiring only one electroluminescent scanning panel and only one set of fibre optical systems into the detectors. This electroluminescent scanning system would unfortunately derive its power from the selection voltage sources, but the number of such sources may be reduced by a sub-matrix selection system of transformers.[8]

3.4. *Avalanche Breakdown*

The avalanche breakdown of a *p–n* junction produces a light output for the duration of the breakdown. Very short light pulses may therefore be generated (Fig. 12) by high speed circuitry techniques. Unfortunately this class of device is extremely inefficient because the *p–n* junction must necessarily be a few microns inside the semiconductor crystal. Considerable absorption of light occurs during its passage from the junction to the surface of the crystal, so that the light output power is small.

4. Detectors

There are four categories of light detectors capable of giving a suitable electrical response for use in this storage system.

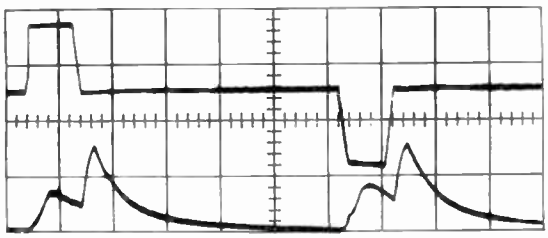


Fig. 10. Light output pulse of an electroluminescent phosphor.
1 horizontal division = 2 μs.
(*a*) excitation voltage.
(*b*) light output.

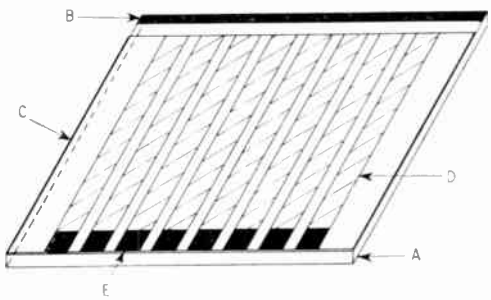


Fig. 11. Electroluminescent panel for a scanning system
A glass base plate coated with a conducting transparent layer.
B contact to the conducting transparent layer.
C electroluminescent layer.
D strip electrode of high conductivity copper.
E contact to the copper strip.

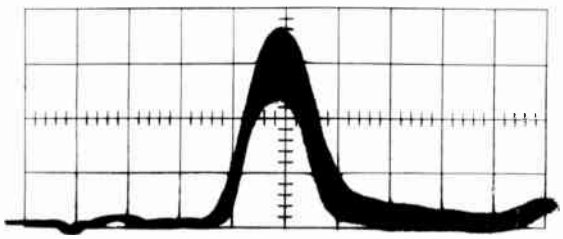


Fig. 12. Light output from an avalanche device.
1 horizontal division = 10 ns.

4.1. *Photo-conductive Devices*

Photo-conductive devices are inherently slow in their response to light signals, generally taking milliseconds to operate. Their spectral response does not normally match the emission of either efficient electro-luminescent phosphors or high speed cathodo-luminescent phosphors. The first point alone is sufficient to bar them from any further consideration.

4.2. *Photo-voltaic Devices*

Modern photo-voltaic devices are the most efficient means of converting light signals into electrical signals, and response times of less than a micro-second can be achieved. The limit on the response time is set by the time-constant of the device itself. However for maximum sensitivity, amplification of the signals is necessary, the limit of sensitivity being set by the thermal noise produced in the cell. Under conditions of low intensity illumination the cell has a high resistance producing a large thermal noise voltage. This together with the amplifier noise gives a low ultimate sensitivity for these devices.

4.3. *Photo-transistors*

Photo-transistors have a fairly high sensitivity due to the current gain of the transistor. However, the sensitivity of the device is considerably limited by light absorption in the semi-conductor crystal before the light reaches the junction, the only useful light being that which enters the emitter-base region of the transistor. Modern photo-transistors having lenses to focus the maximum amount of light on to a junction grown very close to the surface of the crystal still require an illumination of several foot candles for reasonable output pulses.[9]

4.4. *Photo-electric Devices*

Photo-electric devices normally have a sensitivity less than that of photo-voltaic devices. However, electron multiplication may be used to amplify the photo-current, reducing the amplification noise to a negligible level. The main noise source is now the thermal emission of electrons from the photo-cathode.

The photo-multiplier is capable of high gain, and fast rise times. Unfortunately the thermal electrons emitted randomly from the photo-cathode are subject to the same gain as the photo-electrons. If a typical case is considered, where the anode rise time is limited to 4 ns and the electron gain is 10^8, one electron leaving the cathode will produce an anode current pulse of 4 mA. It is therefore necessary to work with light signals on the photo-cathode which are capable of producing several photo-electrons almost coincidently giving peak signal currents at the anode of several times the noise current. Using a typical photo-cathode having a quantum efficiency of 10% this represents a very low level of illumination.

The photo-multiplier has a further advantage for use in this store because it contains its own built-in wide-band amplifier deriving its power from a supply potential which may be common to all the photo-multipliers in the store. The output current pulses, which must have a level of at least 10 mA to be greater than the noise pulses, are compatible, with direct connection to conventional high speed computer circuitry, and a suitable buffer shaping circuit is shown in Fig. 13.

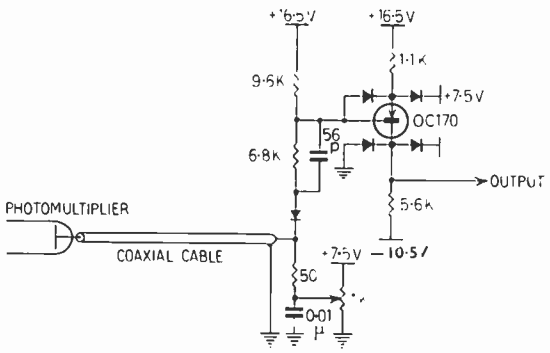


Fig. 13. Photo-multiplier pulse shaping circuitry.

5. Storage Capacity and Packing Density

5.1. *Conditions for Maximum Storage Capacity*

In order to discover the maximum store size and packing density, a uniform diffuse source of surface brightness B lumens/mm^2 will be assumed. If a fibre of area A mm is placed close to this surface then the incident flux on the fibre is AB lumens and the flux emergent from the fibre will be AB/L lumens. This flux is incident on a detector of sensitivity S amp/lumen so if the required output pulse is T amperes then

$$\frac{AB}{L} = \frac{T}{S} \qquad \text{.....(3)}$$

This condition establishes that each fibre can carry sufficient flux to give an output pulse from the detector. However, to safeguard against faulty fibres or damage to a fibre it is advisable to have more than one fibre per hole in the mask, say 4 fibres per hole,

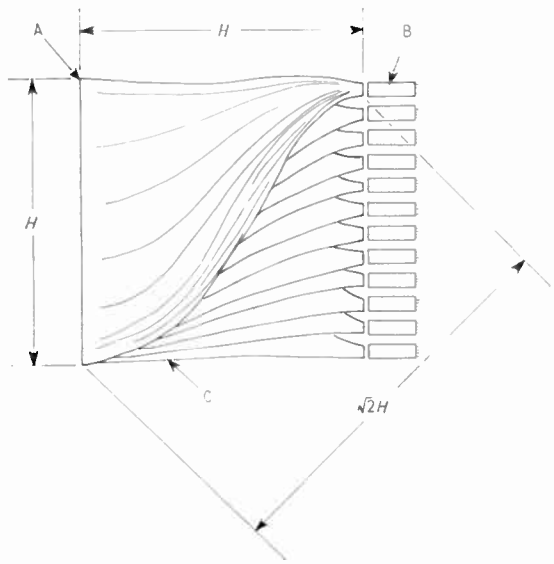


Fig. 14. Layout of sources adjacent to the storage plane

A storage plane.
B sources.
C fibre optical bundles.
H dimension of storage plane.

arranged in a square cross-section. As the diameter of a fibre is $\sqrt{\frac{4A}{\pi}}$, the size of the holes in the mask will be $4\left(\frac{4A}{\pi}\right)$ mm^2.

In order to separate a hole from its neighbours it is necessary to surround each hole by an opaque section. If this is made equal to half the hole width, giving a hole width-to-space ratio of 2 : 1, then the total area of each bit position on the mask is $9\left(\frac{4A}{\pi}\right)$ mm^2. If the size of the matrix is H^2 mm^2 then the total number of bits in the store is

$$Z = \frac{H^2\pi}{36A} \qquad \text{.....(4)}$$

As the store is made larger to contain more bits of information, then the length of the fibres will increase. The attenuation of a fibre is an exponential function of its length, so the attenuation of the fibres will increase rapidly as the size of the store is increased, setting an upper limit on the storage capacity.

The layout of the sources and detectors in the store will also affect the length of the fibres so this must be given careful attention. Whilst the input end of a fibre bundle from a source is circular, the output end is long and narrow. Thus in order to accommodate all the sources (and detectors) as near as possible to the storage plane an arrangement as shown in Fig. 14 may be adopted. The longest fibre is thus $\sqrt{2}H$ mm long.

Allowing a similar arrangement to the detectors, the total length of the longest fibre will be $2\sqrt{2}H$ mm. The attenuation of the longest fibre therefore will be

$$L = \exp\frac{2\sqrt{2}H}{C} \qquad \text{.....(5)}$$

Combining equations (3) and (5)

$$A = \frac{T}{SB} \cdot \exp\frac{2\sqrt{2}H}{C}$$

so the total storage capacity becomes

$$Z = \frac{H^2\pi}{36} \cdot \frac{SB}{T} \cdot \exp\frac{-2\sqrt{2}H}{C}$$

The maximum storage capacity will be given by

$$\frac{dZ}{dH} = 0$$

giving

$$H = \frac{C}{\sqrt{2}} \qquad \text{.....(6)}$$

Equation (6) now gives the condition for maximum storage capacity

$$Z_{(H=C/\sqrt{2})} = \frac{C^2SB}{T} \times 5{\cdot}9 \times 10^{-3}$$

and the fibre diameter must be $\sqrt{\frac{9{\cdot}43T}{SB}}$

5.2. *The Conditions in Practice*

Typical figures for the sensitivity of a photo-multiplier and the brightness of a high-speed cathodo-luminescent phosphor are $B = \frac{1}{\pi}$ lumens/mm^2 and $S = 2000$ amp/lumen.

Designing for a 10 mA output signal from a photo-multiplier the above expressions yield that a store can be constructed to contain 10^8 bits at a packing density of 75×10^3 bits/cm^2 provided that C is greater than 53 cm with a fibre diameter of 12 μ. The total matrix size would be approximately $(40)^2$ cm^2.

Whilst fibres may easily be produced from glass, having this standard of performance, and from plastics with a slightly inferior performance, the design is clearly impractical because of the difficulty of arranging the detectors and sources within the space available. In order to achieve these figures it is necessary to arrange 10^4 sources and 10^4 detectors each within an area of $(40)^2$ cm^2, i.e. 16 mm^2 per unit. This is possible for the sources, where it is conceivable to manufacture small flash tubes within this size, but the smallest photo-multiplier available is 25 mm in diameter. Only 250 of these may be distributed over the 40 cm square area thus reducing the storage capacity to $2\frac{1}{2}$ million bits. This capacity which is achievable

in practice, whilst being useful does not exploit the system of its full advantages. New components must be designed to take benefit of these.

If the attenuation of the fibres is less, then the optical path length may be extended and the storage capacity correspondingly increased. Alternatively advantage may be taken of the potentially high storage capacity by spreading the photo-multipliers over a large area, but ensuring that the conditions of optical attenuation and geometry discussed in Section 5·1 are observed.

Considering the previous example, if all the conditions are the same except that $C = 100$ cm, then the area available for the photo-multipliers becomes $(70)^2$ cm^2 which will accommodate 10^3 photo-multipliers giving a usable capacity of 10^7 bits.

6. Conclusions

A fixed storage system has been discussed which has the advantages of rapid access, and high storage capacity. The physical dimensions of the unit must of necessity be small. The currently available sources and detectors suggested will enable the calculated figures to be achieved in practice, whilst at the same time their inherent power amplification allows direct connection to standard computer logic elements without expensive buffer circuitry. The problem of unwanted signal pick-up, common to most fixed storage systems has been shown to be negligible.

It is suggested that this storage system should have application in many types of computing installation.

Note added by the Authors in proof

Since the preparation of the manuscript for this paper gallium arsenide devices have been introduced. These have the property of emitting intense light output signals from a forward-biased *p-n* junction and are inherently small in size. The light output response to the electrical signals is very fast (< 10 ns) and so these devices should prove useful in this storage system by reducing the operating times for the store and possibly increasing the store capacity.

7. References

1. G. V. Novotny, "Fibre optics for electronics engineers", *Electronics*, **35**, No. 22, p. 37, 1st June 1962.
2. P. Moon, "The Scientific Basis of Illuminating Engineering", p. 254. (McGraw-Hill, London, 1936).
3. G. Porter and E. R. Wooding, "Nanosecond light sources", *J. Photographic Science*, **9**, No. 3, p. 165, May–June 1961.
4. Zs. Naray and P. Varga, "Production of light pulses of nanosecond rise time and duration by means of gas-discharge tubes", *J. Sci. Instrum.*, **38**, No. 9, p. 352, September 1961.
5. Y. Koechlin, "Measuring the speed of response of the 56 AVP photo-multiplier", *Electronic Applications*, **19**, No. 3, p. 121, 1958–59.
6. G. R. Hoffman, D. H. Smith and D. C. Jeffreys, "High speed light output signals from electroluminescent storage systems", *Proc. Instn Elect. Engrs*, **107**, Part B, No. 36, p. 599 November 1960. (I.E.E. Paper No. 3217 M.)
7. G. R. Hoffman and P. L. Jones, "An electroluminescent fixed store for a digital computer", *Proc. Instn Elect. Engrs*, **109**, Part B, No. 44, p. 177, March 1962. (I.E.E. Paper No. 3821 M.)
8. D. C. Jeffreys and G. R. Hoffman, "The electroluminescent matrix for pattern display and recording", *J. Brit.I.R.E.*, **22**, No. 1, p. 53, July 1961.
9. "L9000 Transistor", *Solid-State Design*, p. 20, April 1962.

Manuscript first received by the Institution on 27th August 1962 (Paper No. 785/C49).

U.D.C. 621.375.4: 512.831

Matrix Analysis Applied to Transistor Amplifier Design

By

G. ZELINGER (*Member*) †

Summary: The amplifier is treated in terms of *h*-parameters and the generalized matrix combines passive input and output terminating networks. The exact input and output impedances, current, voltage and power gain are derived and can be applied to transistor amplifiers of any configuration and arbitrary complexity of terminating impedances.

1. Introduction

The transistor amplifier design problem may be looked upon as essentially a procedure of optimizing transistor performance by assigning suitable parameters to the input and output matching networks. In this process the starting point is usually the reduction of the transistor to a reasonable approximation of a linear, active-four-terminal network model. A large variety of possible configurations of the transistor model have been suggested and extensively covered in the literature.[1-4]

This paper deals with an amplifier design technique which utilizes the transistor hybrid *h* parameters, those which are frequently listed in manufacturer's data sheets or may be established by measurements. Matrix analysis will be used for manipulating the transistor and external circuit parameters as best suited to the problem at hand. The concept will be exploited whereby the transistor is considered as forming a link in the chain between generator and output networks. The passive four-terminal networks themselves are conveniently described by the general parameters of the transmission matrix. An analytical approach based on this philosophy, it is believed, yields a better "feel" of the transistor's linear model and its inherent limitations. Furthermore, the compact yet rigorous matrix analysis will contribute to a systematic and elegant mathematical procedure for solving the majority of amplifier design problems.

In terms of basic principles, single stage amplifier with generalized terminal conditions will be studied. However, the method of analysis is valid for transistor amplifiers of any configuration and arbitrary complexity of the terminating networks. It is assumed that the reader has some familiarity with transistors and matrix algebra. For a review references 1 to 12 inclusive may be consulted.

2. General Consideration of Design Target

When designing a linear a.f. or r.f. amplifier one is usually concerned with the establishment of quantitative data in respect of:

(1) Current gain with specified generator and load terminations.

(2) Voltage gain with specified generator and load terminations.

(3) Input impedance of the transistor with specified load terminations.

(4) Output impedance of the transistor with specified generator source impedance.

(5) Device dissipation and d.c. biasing conditions.

This paper will deal with items 1 to 4 only and it is tacitly assumed that the correct d.c. bias conditions exist.

3. Interconnecting the Transistor and Terminating Networks, Definition of Parameter Matrices

It is well known that the transistor is a non-unilateral device. Therefore amplifier gain characteristics and terminal conditions are strongly inter-related. These will also be demonstrated later with mathematical reasoning.

Consider a transistor amplifier in the most general form in Fig. 1 (*a*). A block diagram which conveys also the required basic information is shown in Fig. 1 (*b*). Of course, the transistor may operate in any of the conventional modes, that is, common emitter, common base or common collector. Note furthermore that the transistor and terminating networks have been designated in Fig. 1 (*b*) with internal parameter matrices. The transistor appears to be "embedded" between the two "black boxes" containing the terminating networks. It forms a link between these terminations. It has been established[5, 8, 10] that the passive terminating networks are most effectively described in mathematical terms by the general *A*, *B*, *C*, *D* parameters of the transmission matrix.

If taken separately with due consideration of the customary sign convention, then either of the terminating four-terminal networks may be redrawn as shown in Fig. 2.

The pair of equilibrium equations corresponding to this network model have been defined as follows:[5, 8, 10]

† Canadian Marconi Company, 2442 Trenton Avenue, Montreal 16.

$$V_1 = AV_2 + BI_2 \qquad \text{.....(1)}$$

$$I_1 = CV_2 + DI_2 \qquad \text{.....(2)}$$

Expressing (1) and (2) in matrix form:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \times \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \qquad \text{.....(3)}$$

Reverting now to the remaining link in the chain, the middle "black box" which stands for the transistor with the hybrid h parameters. A block diagram of this network with the customary sign convention is shown in Fig. 3. Here the dependent variables are the input voltage V_1 and output current I_2.

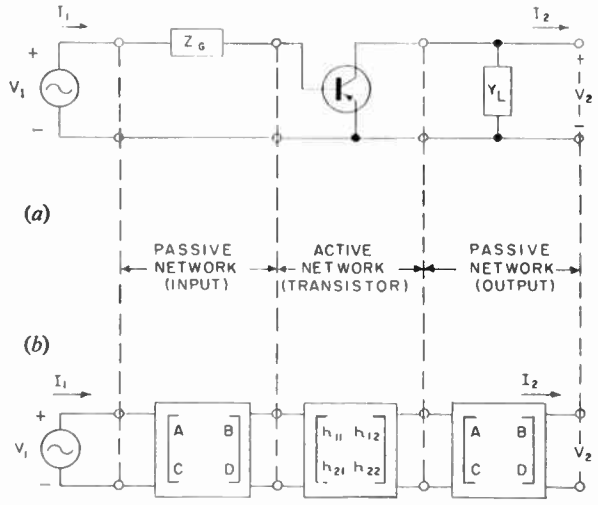


Fig. 1. The single-stage transistor amplifier, general configuration.

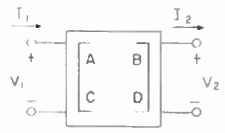


Fig. 2. The generalized four-terminal network.

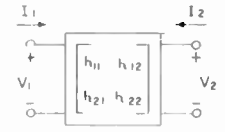


Fig. 3. Four-terminal network with hybrid h parameters.

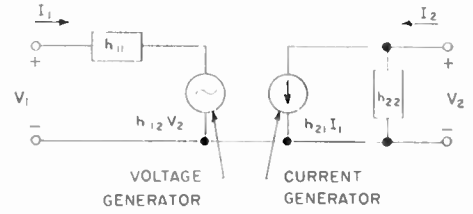


Fig. 4. Linear model of transistor with hybrid parameters.

Writing down the corresponding equilibrium equations:[1, 2, 3, 10]

$$V_1 = h_{11} I_1 + h_{12} V_2 \qquad \text{.....(4)}$$

$$I_2 = h_{21} I_1 + h_{22} V_2 \qquad \text{.....(5)}$$

From these equations a two-generator linear model of the transistor may be synthesized as shown in Fig. 4.

From eqns. (4), (5) and with reference to Fig. 4, the h parameters will be obtained by the use of simple algebra:

$$\left.\frac{V_1}{I_1}\right|_{I_2=0} = h_{11}$$ Has the dimensions of impedance

$$\left.\frac{V_2}{I_2}\right|_{I_1=0} = h_{22}$$ Dimensions of admittance

$$\left.\frac{V_1}{V_2}\right|_{I_1=0} = h_{12}$$ A constant of proportionality, reverse voltage transfer ratio

$$\left.\frac{I_2}{I_1}\right|_{V_2=0} = h_{21}$$ A constant of proportionality, forward current transfer ratio

Equations (4) and (5) in their present form are not particularly attractive for analysis of the type of amplifier interconnections as suggested in Fig. 1. However, the shortcoming can be easily corrected by simple algebraic manipulations. This will be carried out in the next section.

4. Derivation of the Transmission Matrix in Terms of the Hybrid h Parameters

At the outset it has been postulated that the transistor will be treated as an active-linear four-terminal coupling network linking the input and output terminations as shown in Fig. 1. In order to establish the most convenient mathematical description of the complete amplifier, it will be necessary that the assumed direction of current flow in Figs. 3 and 4 are modified. If I_2 is reversed, then the desired conditions will exist. This is shown in Fig. 5.

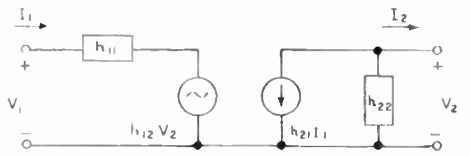


Fig. 5. Linear model of transistor with hybrid parameters after direction of output current I_2 reversed.

Rewriting now eqns. (4) and (5), with the appropriate change in sign of I_2:

$$V_1 = h_{11} I_1 + h_{12} V_2 \qquad \text{.....(6)}$$

$$-I_2 = h_{21} I_1 + h_{22} V_2 \qquad \text{.....(7)}$$

It is further required now that eqns. (6) and (7) are transformed. The variables will be then related as dictated by the structure of the transmission matrix in eqn (3).

From (7) by algebraic transposition:

$$-h_{21} I_1 = I_2 + h_{22} V_2 \qquad \text{.....(8)}$$

$$I_1 = -\frac{1}{h_{21}} I_2 - \frac{h_{22}}{h_{21}} V_2 \qquad \text{.....(9)}$$

Substituting now eqn. (9) for I_1 into eqn. (6):

$$V_1 = h_{11}\left(-\frac{1}{h_{21}} I_2 - \frac{h_{22}}{h_{21}} V_2\right) + h_{12} V_2 \qquad \text{...(10)}$$

$$V_1 = -\frac{h_{11}}{h_{21}} I_2 - \frac{h_{11} h_{22}}{h_{21}} V_2 + h_{12} V_2 \qquad \text{.....(11)}$$

Rearranging terms:

$$V_1 = \left(h_{12} - \frac{h_{11} h_{22}}{h_{21}}\right) V_2 - \frac{h_{11}}{h_{21}} I_2 \qquad \text{.....(12)}$$

$$V_1 = \frac{h_{12} h_{21} - h_{11} h_{22}}{h_{21}} V_2 - \frac{h_{11}}{h_{21}} I_2 \qquad \text{.....(13)}$$

Multiplying by -1 the numerator and denominator of the first term on the right-hand side:

$$V_1 = -\left(\frac{h_{11} h_{22} - h_{12} h_{21}}{h_{21}}\right) V_2 - \frac{h_{11}}{h_{21}} I_2 \qquad \text{...(14)}$$

It will be now recognized that the numerator of the first term represents the determinant of the h matrix, that is,

$$(h_{11} h_{22} - h_{12} h_{21}) \equiv \Delta_h \qquad \text{.....(15)}$$

The identity of (15) enables us to rewrite eqn. (14) in a more compact form:

$$V_1 = -\frac{\Delta_h}{h_{21}} V_2 - \frac{h_{11}}{h_{21}} I_2 \qquad \text{.....(16)}$$

Note that the coefficients in eqns. (9) and (16) correspond to the elements A, B, C, and D in the transmission matrix. Rewriting (9) and (16) as a pair of equilibrium equations which will then characterize the transistor as an active-four-terminal network, suitable to further processing.

$$V_1 = -\frac{\overbrace{\Delta_h}^{A}}{h_{21}} V_2 - \frac{\overbrace{h_{11}}^{B}}{h_{21}} I_2 \qquad \text{.....(17)}$$

$$I_1 = -\underbrace{\frac{h_{22}}{h_{21}}}_{C} V_2 - \underbrace{\frac{1}{h_{21}}}_{D} I_2 \qquad \text{.....(18)}$$

Expressing (17) and (18) in matrix form:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = -\frac{1}{h_{21}} \begin{bmatrix} \Delta_h & h_{11} \\ h_{22} & 1 \end{bmatrix} \times \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} \qquad \text{.....(19)}$$

From (17), (18), and (19), it is evident that the transmission matrix of the transistor has been defined in terms of the hybrid or h parameters:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}_h \equiv -\frac{1}{h_{21}} \begin{bmatrix} \Delta_h & h_{11} \\ h_{22} & 1 \end{bmatrix} \qquad \text{.....(20)}$$

The elements in (20) are identified as follows:

$$A \equiv -\frac{\Delta_h}{h_{21}}$$

$$B \equiv -\frac{h_{11}}{h_{21}}$$

$$C \equiv -\frac{h_{22}}{h_{21}}$$

$$D \equiv -\frac{1}{h_{21}}$$

Reverting now to eqn. (19) it appears to be self-evident that the transistor is a non-unilateral device because generally

$$\Delta_h \neq 0$$

5. Input Impedance of the Transistor Amplifier

In the simplest form the transistor and the cascaded load termination may be represented either as in Fig. 6, or as in Fig. 7.

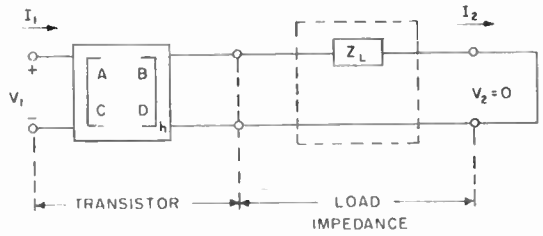


Fig. 6. Transistor amplifier with generalized load impedance.

Consider first Fig. 6; the transmission matrix of this composite network is formed from the product of the individual transmission matrices.

$$\underbrace{\begin{bmatrix} A & B \\ C & D \end{bmatrix}_h}_{\text{Transistor}} \times \underbrace{\begin{bmatrix} 1 & Z_L \\ 0 & 1 \end{bmatrix}}_{\text{Load impedance}} = \underbrace{\begin{bmatrix} A & (AZ_L + B) \\ C & (CZ_L + D) \end{bmatrix}}_{\text{Matrix product}} \qquad \text{...(21)}$$

The equilibrium equation corresponding to the network configuration in Fig. 6, may be written down at once using (1), (2), and (21):

$$V_1 = AV_2 + (AZ_L + B) I_2 \qquad \text{.....(22)}$$

$$I_1 = CV_2 + (CZ_L + D) I_2 \qquad \text{.....(23)}$$

From the ratio of eqns. (22) and (23), the input impedance may be readily obtained.

$$Z_{in} = \frac{V_1}{I_1}\bigg|_{V_2=0} = \frac{AZ_L+B}{CZ_L+D} \qquad \text{.....(24)}$$

Substituting now from the transistor matrix (20), the applicable h parameter into eqn. (24):

$$Z_{in} = \frac{\Delta_h Z_L+h_{11}}{h_{22}Z_L+1} \qquad \text{.....(25)}$$

It is often convenient to consider the output load network as an admittance parameter. This will yield the type of block diagram shown in Fig. 7:

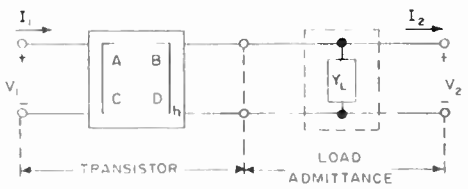


Fig. 7. Transistor amplifier with generalized load admittance.

The transmission matrix of this cascaded network is similarly obtained from the product of the transmission matrices of transistor and load admittance Y_L:

$$\underbrace{\begin{bmatrix}A & B\\ C & D\end{bmatrix}_h}_{\text{Transistor}} \times \underbrace{\begin{bmatrix}1 & 0\\ Y_L & 1\end{bmatrix}}_{\text{Load admittance}} = \underbrace{\begin{bmatrix}(A+BY_L) & B\\ (C+DY_L) & D\end{bmatrix}}_{\text{Matrix product}} \qquad \text{...(26)}$$

Using (1), (2) and the matrix (26), a pair of equilibrium equations are readily obtained:

$$V_1 = (A+BY_L)V_2+BI_2 \qquad \text{.....(27)}$$

$$I_1 = (C+DY_L)V_2+DI_2 \qquad \text{.....(28)}$$

The input impedance is defined as the ratio of eqns. (27) and (28):

$$Z_{in} = \frac{V_1}{I_1}\bigg|_{I_2=0} = \frac{A+BY_L}{C+DY_L} \qquad \text{.....(29)}$$

Substituting into eqn. (29) the approximate parameters from the transistor matrix (20):

$$Z_{in} = \frac{\Delta_h+h_{11}Y_L}{h_{22}+Y_L} \qquad \text{.....(30)}$$

6. Output Impedance of the Transistor Amplifier

It has been pointed out already that the transistor is a non-unilateral device. Consequently, the output impedance will also depend on the input termination. Consider the cascaded network structure in Fig. 8, consisting of a generator V_1 with an internal source impedance of Z_G and the transistor.

By definition, the output impedance is equal to the ratio of V_2/I_2. Therefore when setting up the equilibrium equations for this configuration, it will be necessary to choose V_2 and I_2 as dependent variables. It will be also required that the conventional direction of current flow be reversed as shown in Fig. 8. In terms of matrices; the input and output quantities are now related by the inverse of the transmission matrix of the transistor. Hence from eqn. (3) and Fig. 8:

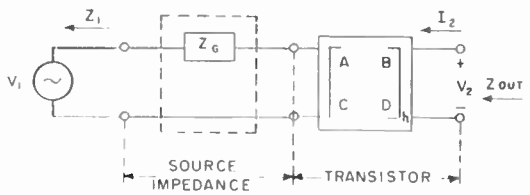


Fig. 8. Reversed current flow for the output impedance derivation.

$$\underbrace{\begin{bmatrix}V_2\\ I_2\end{bmatrix}}_{\text{Output}} = \underbrace{\begin{bmatrix}A & B\\ C & D\end{bmatrix}_h^{-1}}_{\substack{\text{Inverse of}\\ \text{transistor}\\ \text{matrix}}} \times \underbrace{\begin{bmatrix}1 & Z_G\\ 0 & 1\end{bmatrix}}_{\substack{\text{Source}\\ \text{impedance}}} \times \underbrace{\begin{bmatrix}V_1\\ I_1\end{bmatrix}}_{\text{Input}} \qquad \text{...(31)}$$

Performing the inversion of the transistor matrix, from eqn. (31):

$$\begin{bmatrix}V_2\\ I_2\end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix}D & B\\ C & A\end{bmatrix}_h \times \begin{bmatrix}1 & Z_G\\ 0 & 1\end{bmatrix} \times \begin{bmatrix}V_1\\ I_1\end{bmatrix} \qquad \text{...(32)}$$

where: $\Delta \equiv (AD-BC)$ = determinant of the transmission matrix.

When multiplying out the triple matrix product, the desired pair of equilibrium equations will be obtained:

$$V_2 = \frac{D}{\Delta}V_1 + \frac{DZ_G+B}{\Delta}I_1 \qquad \text{.....(33)}$$

$$I_2 = \frac{C}{\Delta}V_1 + \frac{CZ_G+A}{\Delta}I_1 \qquad \text{.....(34)}$$

Taking the ratio of eqns. (33) and (34):

$$Z_{out} = \frac{V_2}{I_2}\bigg|_{V_1=0} = \frac{DZ_G+B}{CZ_G+A} \qquad \text{.....(35)}$$

It is now an easy matter to substitute into eqn. (35) the applicable h parameters from the transistor matrix (20):

$$Z_{out} = \frac{Z_G+h_{11}}{h_{22}Z_G+\Delta_h} \qquad \text{.....(36)}$$

If $h_{22}Z_G \ll \Delta$, then for the purpose of engineering approximations eqn. (36) may be simplified:

$$Z_{out} \simeq \frac{Z_G+h_{11}}{\Delta_h} \qquad \text{.....(37)}$$

7. Current Gain

For the network model of the transistor amplifier in Fig. 6, the equilibrium eqns. (22) and (23) hold. Taking the ratio of I_2/I_1, from eqn. (23):

$$\left.\frac{I_2}{I_1}\right|_{V_2=0} = \frac{1}{CZ_L+D} \qquad \text{.....(38)}$$

Substituting into this equation the applicable h parameters from the transistor matrix (20):

$$\frac{I_2}{I_1} = -\frac{h_{21}}{h_{22}Z_L+1} \qquad \text{.....(39)}$$

The negative sign in (39) is due to the assumed direction current flow. If $h_{22}Z_L \ll 1$, then for the purpose of engineering approximations eqn. (39) simplifies to:

$$\frac{I_2}{I_1} \simeq -h_{21} \qquad \text{.....(40)}$$

which is by earlier definitions the short-circuit current gain.

8. Voltage Gain

For derivation of the voltage gain formula it is convenient to revert to Fig. 7 and the related equilibrium eqn. (28). Note that in this case the load is represented by the admittance parameter Y_L.

From (28), by algebraic transposition,

$$\left.\frac{V_2}{V_1}\right|_{I_2=0} = \frac{1}{A+BY_L} \qquad \text{.....(41)}$$

Repeating here again the routine of substituting into eqn. (41), the appropriate h parameters from the transmission matrix (20) of the transistor,

$$\frac{V_2}{V_1} = -\frac{h_{21}}{\Delta_h+h_{11}Y_L} \qquad \text{.....(42)}$$

The negative sign is attributed again to the adopted sign convention.

Equation (42) may be put into a different form by remembering that

$$Y_L = \frac{1}{Z_L} \qquad \text{.....(43)}$$

$$\frac{V_2}{V_1} = -\frac{h_{21}}{\Delta_h+h_{11}\dfrac{1}{Z_L}}$$

Multiplying numerator and denominator by Z_L

$$\frac{V_2}{V_1} = -\frac{h_{21}Z_L}{\Delta_h Z_L+h_{11}} \qquad \text{.....(44)}$$

If $\Delta_h Z_L \ll h_{11}$ then for the purpose of engineering approximations eqn. (44) may be simplified to

$$\frac{V_2}{V_1} \simeq -\frac{h_{21}}{h_{11}}Z_L = -h_{21}\frac{Z_L}{h_{11}} \qquad \text{.....(45)}$$

9. Power Gain

The operating power gain, PG, of an amplifier is defined as:

$$PG = I_G V_G \qquad \text{.....(46)}$$

where I_G = current gain as defined in eqn. (39),

V_G = voltage gain as defined by eqns. (42) or (44),

hence the product of eqns. (39) and (42) will satisfy (46). Thus

$$PG = \left[-\frac{h_{21}}{h_{22}Z_L+1}\right]\left[-\frac{h_{21}}{\Delta_h+h_{11}Y_L}\right] \qquad \text{...(47}$$

$$PG = \frac{h_{21}^2}{(h_{22}Z_L+1)(\Delta_h+h_{11}Y_L)} \qquad \text{.....(48)}$$

Alternatively the power gain is obtained from the product of eqns. (39) and (44):

$$PG = \left[-\frac{h_{21}}{h_{22}Z_L+1}\right]\left[-\frac{h_{21}Z_L}{\Delta_h Z_L+h_{11}}\right] \qquad \text{...(49)}$$

$$PG = \frac{h_{21}^2 Z_L}{(h_{22}Z_L+1)(\Delta_h Z_L+h_{11})} \qquad \text{.....(50)}$$

If for a practical transistor amplifier one substitutes numerical constants into eqn. (50), then one will find that generally $h_{11} \gg \Delta_h Z_L$. Therefore, for the majority of applications eqn. (50) may be simplified to:

$$PG \simeq \frac{h_{21}^2 Z_L}{h_{11}(1+h_{22}Z_L)} \qquad \text{.....(51)}$$

Further simplification will be possible if $h_{22}Z_L \ll 1$. In such cases eqn. (51) reduces to:

$$PG \simeq h_{21}^2 \frac{Z_L}{h_{11}} \qquad \text{.....(52)}$$

Note that in this paper the transistor and terminating networks have been considered in a generalized form. Therefore if they are complex, then the real parts must be taken in the power gain eqns. (48) to (52) inclusive.

10. Conclusions

A generalized matrix analysis of the transistor amplifier design problem has been demonstrated. The transistor has been reduced to a network model and characterized by the transmission matrix in terms of h parameters. The transistor matrix so formed has been readily combined with the passive input and output terminating networks.

By applying elementary rules of matrix algebra the exact input and output impedances, as well as formulae for current, voltage and power gain have been derived. The results so obtained are completely general and applicable to transistor amplifiers of any configuration and arbitrary complexity of terminating impedances.

11. Acknowledgments

This paper has been the outgrowth of a broad transistorized equipment development programme at Canadian Marconi Company in Montreal. Initially, it has been issued to the staff of the Engineering Department as a Canadian Marconi Technical

Report, No. L212, dated October 1961. The author wishes to express his thanks to K. C. M. Glegg, Chief Engineer and B. N. Sherman, Engineer-in-Charge, Product Development, Commercial Products Division for permission to publish this work.

12. References

1. W. G. Gartner, "Transistors—Principles, Design and Applications". (Van Nostrand, Princeton, 1960.)
2. A. W. Lo and R. O. Endres, "Transistor Electronics". (Prentice Hall, Englewood Cliffs, 1955.)
3. R. F. Shea, "Principles of Transistor Circuits". (John Wiley, New York, 1953.)
4. D. E. Thomas, "Some design considerations for high-frequency transistor amplifiers", *Bell Syst. Tech. J.*, **38**, pp. 1551–80, November 1959.
5. E. A. Guillemin, "Communication Networks", Vol. II, Chapter IV. (John Wiley, New York, 1935.)
6. E. A. Guillemin, "Mathematics of Circuit Analysis", Chapters I and II. (John Wiley, New York, 1949.)
7. S. J. Mason and H. T. Zimmerman, "Electronic Circuits, Signals and Systems", Chapter 2. (John Wiley, New York, 1960.)
8. E. Weber, "Linear Transient Analysis", Vol. II, Chapters 2 and 5. (John Wiley, New York, 1956.)
9. G. Zelinger, "Basic Matrix Algebra and Transistor Circuits". (Pergamon Press, Oxford, to be published.)
10. G. Zelinger, "Introduction to Matrix Analysis of Transistor Circuits", Canadian Marconi Co. Technical Report No. L211, December 1959.
11. K. G. Nichols, "A matrix representation of linear amplifiers", *J. Brit.I.R.E.*, **21**, pp. 517–33, June 1961.
12. L. B. Johnson, "A simple explanation of transistor *h* parameters", *Mullard Tech. Commun.*, **2**, No. 13, pp. 58–63, July 1955.

Manuscript received by the Institution on 22nd November 1961. (Paper No. 786.)

U.D.C. 621.396.677

Multiplicative Arrays in Radio-Astronomy and Sonar Systems

By

Professor
D. G. TUCKER, D.Sc.
(*Member*) †

Presented at the Symposium on Sonar Systems in Birmingham on 9th–11th July 1962.

Summary: The paper is concerned with gathering together, and extending, some ideas on multiplicative receiving arrays as applied to radio-astronomy and sonar. The main emphasis is on systems using coherent-tone signals and designed for direction-determination. It is shown that, using arrays in which the outputs from the two halves are multiplied together, the accuracy of determination of target bearing may be increased by a factor of $\sqrt{2}$ up to 2 as compared with ordinary arrays of the same size; split-beam systems have no theoretical advantage over full-beam systems. Using arrays in which the two portions have the same centre-point, some new and attractive directional patterns are developed (including pencil beams) in which secondary responses may be almost eliminated over a considerable range of bearing.

1. Introduction

The idea of multiplying together the signals received on two spaced receiving elements may well have originated in the idea of correlation as a means of detecting a directional signal in a background of isotropic noise.[1] The well-known phase-switching receiving system used by Ryle[2] and others[3, 4] in radio-astronomy appears to have been introduced as a correlator giving the spatial correlation function, although later used more widely as a multiplier. The use of correlator arrays for detection and bearing-determination of signals with a continuous spectrum, and the relationship between effective directional pattern and bandwidth, are too well-known to warrant further mention here.

Later work in radio-astronomy[3, 4] has used signal-multiplication more flexibly as a means of obtaining new types of directional pattern. An example giving a pencil beam is the well-known Mills' cross array. The important development appears to the present author to be the multiplication of signals from two partial-arrays which are not spaced apart, but may have coincident centre points so that no concept of spatial correlation function is involved; moreover, the two arrays may have very different configurations. One objective may then be to obtain a directional pattern which, while based on a line of spaced elements, nevertheless avoids some or all of the repetitive "diffraction-secondary" peaks which normally recur at angular intervals dependent on the length of the array. An example of this will be discussed later, together with some other systems concerned with pencil beams using partial arrays at right angles. Such methods are essentially applicable to narrow-band signals, and especially to coherent-tone signals.

In the radio-astronomy applications, discrimination against background or ambient noise does not appear to be always of prime importance. For maximum output signal/noise ratio in a multiplicative receiver, the total effective apertures of the two partial arrays should be approximately equal (or, expressed differently, the signal/noise ratio should be the same on both inputs to the multiplier); this is readily seen from Appendix 1 of ref. 5.‡ But the system shown in Fig. 4 of ref. 3 multiplies the output of 2 elements against that of 8 elements—all of equal size—and so fails to obtain a signal/noise ratio better than that of 2 elements alone. Probably the availability of low-noise amplifiers (e.g. parametric amplifiers and masers) and the very low dissipation in the aerial, together with the very cold sky as background, and—perhaps dominantly—the possibility of very long post-detector integration times, all justify the sacrifice of signal/noise performance for angular narrowness of the directional pattern.

It is also important to observe that a multiplicative receiver does not inherently give as good a signal/noise ratio as the ordinary linear array-plus-detector,[7] but has practical advantages due to the fact that its d.c. output is zero in the absence of correlated signal.

If now we turn to sonar applications, we find the principles of multiplicative reception have been used for both continuous-spectrum and coherent-tone

† Electrical Engineering Department, University of Birmingham.

‡ It should be observed that the proportionality of signal/noise ratio to aperture fails for dimensions smaller than one wavelength,[6] and does to some extent depend on the shape of the array.

signals,[5, 8] the latter, of course, being of primary importance in active pulsed sonar. One difficulty in multiplicative systems is that if two targets occur at the same range on different bearings, there is some interaction between them; but experience so far suggests that this may not be a serious objection, and in some circumstances may even be an advantage in clearly distinguishing two separate targets from one larger one.[8] In any case, the ordinary additive array gives a somewhat similar effect due to the interactions in the rectifier which must follow it, and results so far available indicate that the multiplicative arrangements may often be superior in this respect. The multiplicative arrangements so far used in active sonars have used two partial-arrays with spaced centres.

Throughout the remainder of this paper, coherent-tone signals will be assumed. Since in sonar systems the question of signal/noise ratio is often vital, and since reverberation—which is a common limiting factor—behaves in many circumstances exactly like noise, it will generally be assumed that the two partial arrays involved in a multiplicative system have the same effective area or aperture.

The ideas discussed in this paper are applicable primarily to direction-determining systems, and seem unlikely to have any worthwhile applications in ordinary communication systems. The usual assumption will be made that inter-element or inter-section coupling in the arrays may be neglected.

2. Determination of Direction of a Signal Source

If we assume that initial detection of the signal source (or "target") is not the prime objective of the system, but rather the determination of its direction, then signal/noise ratio has a somewhat different significance to that most usually considered. We are now concerned with the influence of noise on the accuracy with which the direction of the target can be determined. It has already been shown by the author[7]

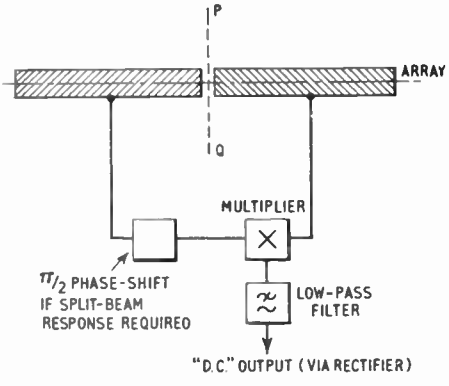


Fig. 1. Multiplicative system with spaced centres.
Beam arrangement gives peak on direction PQ
Split-beam arrangement gives null on direction PQ.

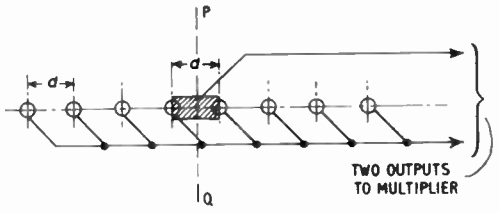


Fig. 2. Multiplicative system with $(\sin x)/x$ output.

that the "split-beam" systems, in which a null is obtained on the direction of the target, the multiplicative arrangement gives $\sqrt{2}$ times better performance in respect of directional accuracy in noise than the linear subtractive arrangement for the same total size of array. (The multiplicative arrangement is the one where the signals from two halves of a line array are multiplied after suitable phase-shifting—see Fig. 1; the subtractive arrangement is the one where the signals from the two halves are subtracted.) Federici[9] has shown that, on the assumption that the two portions of the array have little individual directivity (i.e. they form a simple interferometer—but this assumption can have little effect on the comparison), the ordinary "beam", or peak of the directional pattern, of a linear additive system gives potentially the same accuracy of bearing as the linear subtractive split-beam system using the same array. Applying Federici's method to the directional pattern and noise factor of the multiplicative "beam" as given in ref. 7—which multiplies the signals from the two halves of an array, and gives half the beamwidth but a 3 dB worse noise factor†—we see that the multiplicative beam once more gives an accuracy $\sqrt{2}$ times better than the linear additive system.

We thus conclude that the multiplicative system is superior whether used with a beam or with a null.

The basis of the above results is that the accuracy of bearing determination is dictated by the fact that the peak or the null can be identified only as lying between the bearings at which the change of signal output exceeds a certain proportion of the r.m.s. noise amplitude. The proportion is usually taken as unity, and on this basis the probable angular error for the linear systems corresponds to $\sqrt{2}/R$ radians in electrical phase angle between the centres of the two parts of the array, where R is the signal/noise (r.m.s.) ratio. If, however, the signal/noise ratio is very high, then in many circumstances, e.g. with an ordinary intensity-modulated display, in which special adjustments of gain and contrast for individual targets are not possible, the noise hardly affects the issue. The accuracy is then determined by the smallness of amplitude change which the eye can detect. In this case, the fact that the

† This applies for signal/noise ratios which are fairly high above unity.

multiplying system gives half the beamwidth means that it gives twice the angular accuracy of the linear additive system.

3. Multiplicative Systems with Common Centres

There is a great deal of scope for obtaining a wide variety of the directional patterns‡ by multiplying the signals from two partial-arrays having a common centre point.

3.1. *Multiplicative Directional Pattern of (sin x)/x Type*

One arrangement, which is mentioned by Jennison,[4] is shown in Fig. 2. A line array of n small elements has its total output multiplied by that of a continuous strip array of length, d, equal to the spacing of the small elements. The output of the line array, for received signal cos pt, is

$$A_1 \frac{\sin ny}{n \sin y} \cos pt \qquad \text{.....(1)}$$

where A_1 is the amplitude and $y = \frac{\pi d}{\lambda} \sin\theta$, λ being the wavelength, and θ the direction of a point signal source relative to the normal axis PQ. The output of the strip array is

$$A_2 \frac{\sin y}{y} \cos pt \qquad \text{.....(2)}$$

so that the d.c. output from the multiplier is

$$\tfrac{1}{2} A_1 A_2 \frac{\sin x}{x} \qquad \text{.....(3)}$$

where $x = ny$. Thus the directional pattern is $(\sin x)/x$, as shown in Fig. 3(*a*).

It is, of course, well known that the directional pattern of a continuous strip array of length nd is $(\sin x)/x$, but this is then the *envelope* of a signal of frequency p. On rectification, which is essential to the use of the signal, the polarity reversals of alternate lobes of the pattern are lost, so that the effective pattern is as shown in Fig. 3(*b*).

Now when the multiplicative system is being used, there is rarely any difficulty in connecting a rectifier in its output, so that the lobes of Fig. 3(*a*) of negative polarity are removed, leaving the directional pattern of Fig. 3(*c*). (If a cathode-ray-tube or chemical-recorder display is used, this rectification occurs automatically, since the display responds only to one direction of voltage or current.) We thus finally obtain a directional pattern much better than that of the ordinary $(\sin x)/x$

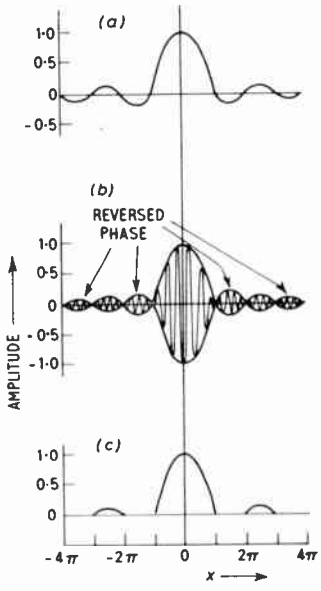


Fig. 3.

(*a*) Directional pattern of arrangement of Fig. 2.

(*b*) Directional pattern of ordinary continuous array.

(*c*) As (*a*), but with negative lobes removed by rectifier.

type from an array whose total area is only $2/n$ of that required for the normal pattern. Of course, a price is paid for this improvement: the smaller area means the signal/noise performance is poorer, and, as previously mentioned, the multiplicative system may get into difficulty with multiple targets.

3.2. *Elimination of Secondary Responses with Positive Polarity*

Another interesting type of multiplicative array can be developed from the ideas of ref. 10 by a quite simple extension of the reasoning. It is there shown that a continuous strip array which has the centre one-fifth of its length omitted has all its secondary-lobe responses of reversed polarity within the range of $|x|$ from just below π to just above 4π, where

$$x = \frac{l\pi}{\lambda} \sin\theta,$$

l being the length of the array. Multiplication of the output of this array with a signal having negligible dependence on x, (e.g. with the output of a point element at the centre of the array) then gives a d.c. output whose dependence on x shows no positive response except in the main centre beam, and outside the range of $|x| \leqslant 4\pi$. If this range of $|x|$ includes all real angles (i.e. $|\theta| \leqslant \pi/2$), then a rectifier following the multiplier and low-pass filter will ensure that *there are no secondary responses at all* in the range of real angles, and the central beam is, moreover, slightly narrowed, as shown in Fig. 4.

If the second partial-array of the multiplicative arrangement is in fact the whole of the centre one-fifth omitted from the first, then the directional pattern of the second part alone has well-placed polarity reversals to compensate for the positive lobes of the

‡ By directional pattern is meant the graph of variation of output voltage against an abscissa representing direction as the array is rotated while a signal is being received from a single distant point source. A discussion of directional pattern, directional discrimination, and noise factor (which are not closely related in multiplicative systems) is given in ref. 5.

phase-reversing properties; but it is probably still a suitable array for keeping the maximum secondary responses small. It should be noted that in the system of Fig. 9, the second multiplier has to operate down to zero-frequency inputs, and the detector is preferably of square-law type.

It should be observed that if the outputs of two multiplicative arrays are themselves multiplied together, then rectifiers may be used *before* the second multiplication in order to prevent a pair of negative secondary responses producing a positive response.

4. References

1. H. Nodtvedt, "The correlation function in the analysis of directive wave propagation", *Phil. Mag.*, **42**, p. 1022, 1951.
2. M. Ryle, "A new radio interferometer and its application to the observation of weak radio stars", *Proc. Royal Soc.*, A, **211**, p. 351, 1952.
3. R. N. Bracewell, G. Swarup and C. L. Seeger, "Future large radio telescopes", *Nature*, **193**, p. 412, 3rd February 1962.
4. R. C. Jennison, "Some radio astronomy techniques", *J. Brit.I.R.E.*, **23**, p. 121, February 1962.
5. V. G. Welsby and D. G. Tucker, "Multiplicative receiving arrays", *J. Brit.I.R.E.*, **19**, p. 369, June 1959.
6. V. G. Welsby, "The signal/noise gain of ideal receiving arrays", *Proc. Instn Elect. Engrs*, **109**, Part C, p. 108, 1962, (I.E.E. Monograph 470 E, September 1961).
7. D. G. Tucker, "The signal/noise performance of electro-acoustic strip arrays", *Acustica*, **8**, p. 53, 1958.
8. V. G. Welsby, "Multiplicative receiving arrays: the angular resolution of targets in a sonar system with electronic scanning", *J. Brit.I.R.E.*, **22**, p. 5, July 1961.
9. M. Federici, "La precisione di misura della direzione di una sorgente sonora con sistemi riceventi direttivi", *La Ricerca Scientifica*, **29**, p. 2301, November 1959.
10. D. G. Tucker, "Improved directivity using synchronous demodulation", *Acustica*, **11**, p. 45, 1961.

Manuscript received by the Institution on 6th June 1962. (Paper No. 787/SS25.)

POINTS FROM THE DISCUSSION

Mr. P. G. Redgment: This paper uses the "polar diagram" as a description of the directive properties of the array; I suggest that this, whilst common practice, can be misleading in the case of certain multiplicative arrays. All applications of the polar diagram known to me tacitly assume that linear superposition of the output responses is possible, and I believe that the diagram is meaningful only in those cases where this process is applicable.

As mentioned by Dr. Welsby, where phase can be neglected (incoherent sources) and the statistics are Gaussian, *power* can be added linearly and the processing system should incorporate *one* multiplication, either a squaring process or a cross-multiplication as requisite for the purpose in view. The polar diagram expressed in terms of this power output then adequately describes the performance of the array.

The arrangement illustrated in Fig. 9, however, incorporates a further stage of multiplication and is thus analogous to the time average product arrays of Berman and Clay† and linear superposition no longer applies. It has been shown by Fakley‡ that the polar diagram in systems of this type gives little guide to their performance in any of the usual applications.

I wish to emphasize that the optimum quality of a 2nd order system assumes Gaussian statistics, and higher order processing may be required for other statistics, but whether the polar diagrams of such systems will be useful in their description is not at present clear to me.

Professor D. G. Tucker (*in reply*): I am afraid I do not yet fully understand all aspects of the use of multiplicative arrays in practice, but I very much wonder whether the theoretical considerations referred to by Mr. Redgment really take account of how a system might be used operationally. I am fully in agreement with him that the polar diagram—or directional pattern—must be used with care. I gave a careful definition of it in the footnote to Section 3 of my paper, and I would hesitate to use it in any other sense. I think my paper contains a sufficient number of warnings that there are various difficulties in using multiplicative arrays, but I would like to assure Mr. Redgment that the improved directional pattern of the multiplicative systems (compared with an additive system) has been shown experimentally to correspond to a markedly improved performance in an active sonar (see my ref. 8).

It is worth noting that a "linear" additive array requires a rectifier to process the output signal, and this rectifier introduces into the additive system some of the adverse features of the multiplicative system (see my ref. 7), as well as the loss of polarity information in the secondary responses. In our experimental use of multiplicative systems we have used a rectification process (in effect) to remove negative secondary responses after the multiplier, and many of the proposals in my paper depend on this; but I do not think that this should be regarded as having in any way the significance of the rectifier used in an additive system. It would thus seem that our success with systems using one stage of multiplication is consistent (even if fortuitously) with the theory quoted by Mr. Redgment. No experimental work has been done on the double-multiplication arrangement shown in Fig. 9, and it is, therefore, possible that it would not be successful in practice.

It seems to me likely, however, that excessive emphasis is being put on multiple-target performance; in practice one would not expect the frequent occurrence of multiple targets within an area of one beamwidth times one pulse-length. The performance with respect to large, extended targets may be more important, but has not yet been investigated.

† *J. Acoust. Soc. Amer.*, **29**, p. 805, July 1957.

‡ *J. Acoust. Soc. Amer.*, **31**, p. 1307, October 1959.

U.D.C. 534.88:621.391.883.2

The Detection of Sonar Echoes in Reverberation and Noise

By

J. O. ACKROYD,
M.A., B.Sc. (Eng.)†

Presented at the Symposium on "Sonar Systems" in Birmingham on 9th–11th July 1962.

Summary: A method is presented for calculating the frequency spectrum of reverberation in a shipborne pulse sonar and hence its effect in masking an echo which exhibits doppler. Signal processing is considered, as affecting the masking due to reverberation and noise combined. This enables the relative merits of sonars, or of proposed modifications, to be assessed more realistically than by considering noise alone.

1. Introduction

The performance of a sonar is determined by its ability to separate the desired echoes from a background of unwanted noise. In a shipborne set this background is made up of two components, (a) noise due to the sea and the ship and (b) reverberation due to the presence of back-scattering of the transmission from surface, bottom or from the water itself.

The first part of this paper is devoted to an examination of the power spectrum of reverberation. This enables reverberation and noise to be considered together in the second part of the paper which deduces their masking effect on signal detection. Finally a hypothetical sonar is considered to illustrate the use of the method in assessing its performance.

2. Reverberation

Consider a pulse sonar transmitting a spreading beam of sound in the sea, with reverberation, noise and any echoes received through a band pass filter. The sonar is in a moving ship whose speed has, in general, a component in the direction of the sound rays radiated and this speed component varies across the width of the beam. It may be assumed that the scatterers which throw back the reverberation have a velocity distribution among themselves. The transmission is a square pulse of continuous wave.

The frequencies present in the reverberation are those in the transmitted pulse, with doppler shifts due to the velocity of the scatterers and to the resolved component of the ship velocity. The output of the receiving filter is determined by the nature of the reverberation and by the characteristics of the filter itself.

Let $W_a(f)$ be a function describing the power response of the filter in terms of frequency

$W_b(f)$ describes the distribution of doppler shifts across the sonar beam

$W_d(f)$ describes the velocity distribution of scatterers among themselves

$W_p(f)$ is the power spectrum of the pulse.

Finally, let

$W_r(f)$ be the actual power spectrum of the reverberation

and $W_m(f)$ the measured output of the receiving filter as a function of its mid-band frequency.‡

It can be shown that the measured output $W_m(f)$ is given by the successive convolution of $W_a(f)$, $W_b(f)$, $W_d(f)$ and $W_p(f)$. Writing $\psi(\tau)$ for the inverse Fourier transform of $W(f)$ then

$$\psi_r(\tau) = \psi_b(\tau) \cdot \psi_d(\tau) \cdot \psi_p(\tau)$$

and

$$\psi_m(\tau) = \psi_a(\tau) \cdot \psi_b(\tau) \cdot \psi_d(\tau) \cdot \psi_p(\tau) \quad \text{.....(1)}$$

While it would be possible to compute a numerical value for any given case it is more instructive (and easier) to use an approximation.

The characteristic of a single resonant circuit is

$$W_a(f) = \frac{1}{f^2 + a^2} \times \text{constant} \quad \text{.....(2)}$$

where f is the frequency difference from the midband frequency f_0.

This function is symmetrical about f_0, its sidebands are inversely proportional to f^2 for values of $f \gg a$. Its inverse Fourier transform is

$$\psi_a(\tau) = e^{-a|\tau|} \times \text{constant}$$

The power spectrum of a pulse of duration t is

$$W_p(f) = \frac{\sin^2 \pi f t}{(\pi f t)^2} \times \text{constant}$$

† Royal Naval Scientific Service.

‡ This is the total power output of the filter as a function of its midband frequency; it is *not* the spectrum of the filter output which would be $W_r(f) \times W_a(f)$.

While this goes through a series of maximum and zero values the trend of its sidebands falls with increasing f as did eqn. (2) so the expression

$$W_p(f) = \frac{1}{f^2+p^2} \times \text{constant}$$

is a useful, albeit smoothed, approximation for the pulse spectrum. The value $p = 1/\pi t$ gives the correct "noise bandwidth" and sidebands which pass through the successive maxima of the rigorous expression. The corresponding inverse Fourier transform is

$$\psi_p(\tau) = e^{-p|\tau|} \times \text{constant}$$

The horizontal beam pattern of a rectangular transducer, expressed in terms of power, is of the form

$$W_b(f) = \frac{\sin^2\phi}{\phi^2}$$

where $$\phi = \frac{\pi f_0 l \sin\theta}{c} \qquad \text{.....(3)}$$

f_0 = frequency
l = width of transducer
c = velocity of sound
θ = angle from the beam axis.

If the transducer is in a ship having speed V and is pointed at right angles to the direction of motion, the relative doppler shift f in the direction θ is given by

$$f = \frac{2f_0}{c} V \sin\theta$$

Substituting for sin θ, eqn. (3) becomes

$$\phi = \frac{\pi l f}{2V}$$

(This breaks down, of course, at values of θ greater than $\pi/2$.)

We have, therefore,

$$W_b(f) = \frac{\sin^2 f/b}{(f/b)^2} \times \text{constant}$$

which, as before, can be approximated by

$$W_b(f) = \frac{1}{f^2+b^2} \times \text{constant}$$

where $$b = \frac{2V}{\pi l}$$

giving $$\psi_b(f) = e^{-b|\tau|} \times \text{constant}$$

(When the transducer is trained in other directions the value of b may be corrected accordingly, but the approximation holds over a narrower range of θ.)

We will assume a similar expression for $\psi_d(\tau)$ due to the velocity distribution of the scatterers.

This leads to

$$\psi_m(\tau) = e^{-a|\tau|} \cdot e^{-b|\tau|} \cdot e^{-d|\tau|} \cdot e^{-p|\tau|} \times \text{constant}$$

Whence $$\psi_m(\tau) = e^{-m|\tau|} \times \text{constant}$$

where $$m = a+b+d+p \qquad \text{.....(4)}$$

This delightfully simple result is supported by the fact that an equation of the form

$$W_m(f) = \frac{1}{m^2+f^2} \times \text{constant}$$

can be fitted with an error of less than 1 dB to observed experimental results. The experiments were for a particular pulse length, but a range of values of b, the beamwidth parameter was used. The fit was adequately good with a value of zero assigned to the quantity d defining the relative velocities of the scatterers.

The author does not intend to discuss these experiments further save to say they would not have shown up a value of d corresponding to, say, $\frac{1}{4}$ knot as being significant (that is $d \simeq 3$ c/s at a carrier frequency of 20 kc/s).

The useful result given above can be stated broadly: the spectrum width, as observed by an analyser filter, is given by the sum of the bandwidths of the pulse spectrum, the doppler spread of the sonar beam and of the filter, bandwidths being measured to the 3 dB points. Figure 1 shows the spectrum shape $W_r(f)$ for a pulse length of 100 ms, (a) for ship speed zero and (b) for ship speed 9 knots, the transducer width l being 2 ft (60 cm approximately).

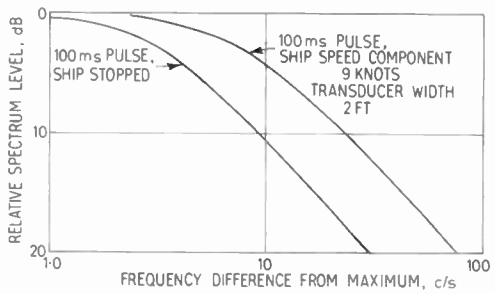


Fig. 1. The spectrum of sonar reverberation.

If the transducer is trained at a different angle the value of V becomes approximately the ship speed resolved in the direction of the axis. Thus the second curve in Fig. 1 will serve for speed 18 knots and training 30 deg on the bow.

3. Masking Effect of Background

The concept of recognition differential is well known; the ratio of signal power to noise background power for a signal to be perceived, with a false alarm

rate either explicitly stated or implicit in the habits of the operator.

The term "equivalent masking level" is used here to define the ability of the background to interfere with detection and to cause false alarms. As a starting point consider the detection of a single pulse of duration t in the presence of noise. It is well known[1] that the optimum i.f. bandwidth B is given approximately by $Bt = 1$. The author will fix arbitrarily the equivalent masking level in this case as the level of noise power in the bandwidth $B = 1/t$.

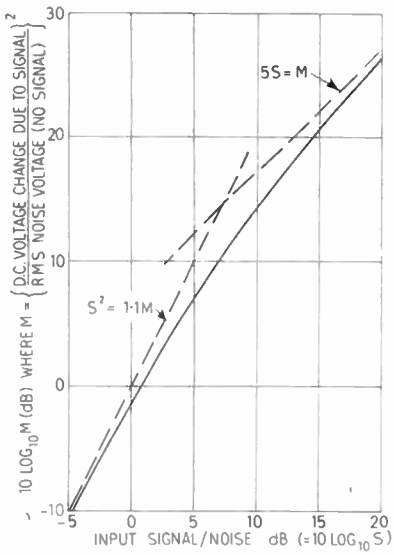


Fig. 2. Input and output signal/noise ratio for a linear detector.

If the i.f. bandwidth is greater than $1/t$ it can be shown that the same signal detection threshold is obtained by using a video frequency bandwidth $b = 1/t$, *provided* that the signal rectifier is linear in operation: however even a so-called linear rectifier only satisfies this requirement if the signal/noise ratio at its input is sufficiently large. Any improvement of signal recognition by video frequency filtering causes a reduction of the ratio threshold signal/noise at the input of the rectifier. A mathematical expression for the law relating these quantities is given in Reference 2 (Equation 3.10–11). It is shown as a curve in Fig. 2. This curve is closely approximated by the equation

$$S = \sqrt{1{\cdot}1M} + 0{\cdot}2M \qquad \text{.....(5)}$$

where S = input signal/noise power ratio

$$\sqrt{M} = \frac{\text{change of d.c. output due to signal}}{\text{r.m.s. of a.c. component of noise, no signal}}$$

It can be seen that at low signal/noise ratios the performance of the rectifier becomes square-law. In consequence there is a "detection loss" determined by the amount of post-detector processing. If this improves the output signal/noise power ratio by a factor n then the threshold signal/noise ratio D at the final output becomes $M = D/n$ at the detector output.

Then, from eqn. (5),

$$S_n = \left(\frac{D}{n} + 5{\cdot}25\sqrt{\frac{D}{n}}\right) \times \text{constant}$$

where S_n = input signal/noise power ratio for threshold detection.

The detection loss is then given by

$$L = \frac{(D/n + 5{\cdot}25\sqrt{D/n})n}{D + 5{\cdot}25\sqrt{D}}$$

$$= \frac{5{\cdot}25\sqrt{n/D} + 1}{5{\cdot}25\sqrt{1/D} + 1}$$

For a noise background with a Gaussian distribution, $D = 28$ (or 10 log $D = 14{\cdot}5$ dB) is a realistic detection threshold for 50% detection probability with false alarms 1 in 10^6. Then the detection loss factor is

$$L = \frac{\sqrt{n} + 1}{2} \qquad \text{.....(6)}$$

We shall call 10 $\log_{10}$ n the gross processing gain and 10 $\log_{10}$ n/L the nett processing gain, the difference being the detection loss as shown in Fig. 3.

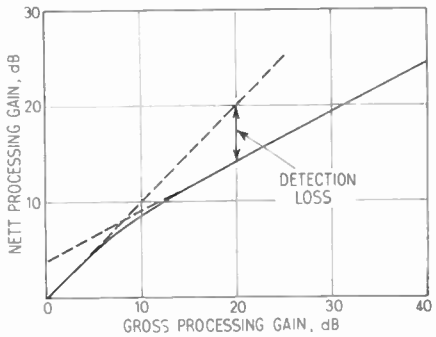


Fig. 3. Detection loss as a function of gross processing gain.

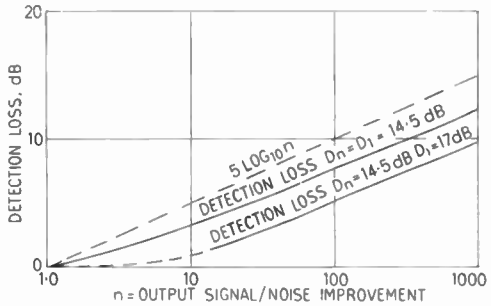


Fig. 4. Nett processing gain due to post-detector integration.

In all the foregoing the output noise was assumed to be Gaussian; this is the case for large values of n, but at lower values the form of the distribution curve changes until at $n = 1$ the output noise has a Rayleigh distribution for which the detection threshold is some 2½ dB higher, i.e. 17 dB rather than 14·5 dB. (A value of 17 dB at the output corresponds to an input signal/noise ratio of 12 dB—see Fig. 2.) Equation (6) therefore over-estimates the detection loss for post-detector filtering at the higher values of n. To allow for this the bottom curve of Fig. 4 has been drawn 2½ dB lower than that representing eqn. (6) for values of n greater than 10; the part between $n = 1$ and $n = 10$ has been sketched in.

If there were no v.f. filtering, the detection threshold of a pulse-signal would be related to the i.f. bandwidth as shown by the uppermost curve of Fig. 5, with a minimum in the region of $Bt = 1$. We can use the relationship between nett processing gain and n as given in Fig. 4 to determine the effect of optimum post-detector filtering. The result is shown by the curve labelled "$bt = 1$ $N = 1$ pulse", in Fig. 5 from which it will be observed that even for $Bt = 100$, the detection loss is only 4·5 dB.

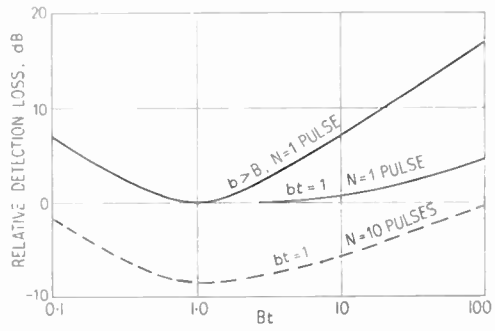


Fig. 5. Detection threshold as affected by signal processing.

A physical v.f. filter of bandwidth $1/t$ is not essential since, subject to certain provisos, the eye can provide this function in visual systems.[1] In radars with high scanning rates it can also integrate from pulse to pulse.[3] If N is the number of pulses integrated the gross processing gain is increased by 10 log N. As a matter of interest, the lowest curve of Fig. 5 has been drawn for the case $N = 10$. With this degree of pulse-to-pulse integration the detection threshold is considerably improved.

The noise arriving at the detector has a bandwidth B determined by the i.f. filter. Reverberation, however, as was seen in Section 2, has a bandwidth often little more than that of the signal. While any pulse-to-pulse integration is still beneficial, the masking level of reverberation is therefore little reduced by v.f. filtering.

How far is the assumption of a "white noise" justified in a shipborne sonar? The combined effect of ship and sea produces high peaks in the filtered output, say 20 dB above the mean, far more frequently than the theory predicts.[4] Reverberation normally has a Rayleigh distribution,[4] but it too can provide excessively high and frequent peaks, particularly in shallow water. The signals, too, have a variance of their own. These are probably the reasons why a recognition differential of at least 15 dB is more realistic in a shipborne sonar than the value of 12 dB obtained for white noise for 50% detection probability and a false alarm rate of 1 in 10^6.

4. Calculation of Sonar Performance

With the aid of the foregoing the detection performance of a sonar can be deduced, given such parameters as the source level, directivity index, receiving bandwidth etc. The ordinates of Fig. 5 are used to obtain the masking level of noise and reverberation.

For present purposes it is convenient to assume a simple propagation law, neglecting refraction effects, namely

$$H = 20 \log R + \frac{2R}{1000}$$

where H = attenuation, one way, dB

R = range in yards.

Reverberation power is assumed to be inversely proportional to the cube of distance: this is a reasonable approximation over the ranges to be considered[4] for a horizontal-looking sonar. A typical level for the reverberation is 120 dB below the source level for a 1-second pulse at 1000-yard range, presumably for a typical transducer directivity.[4] An isotropic noise spectrum level of −40 dB will be assumed.

A hypothetical sonar will be assumed having the following parameters

Transducer width	2 ft
Operating frequency	15 kc/s
Directivity index	−25 dB
Source level referred to 1 yard	130 dB
Pulse length, t	100 ms
Target strength	15 dB
Reverberation level, 1000 yd range,	0 dB
10 000 yd range,	−30 dB
Noise spectrum level	−65 dB
Recognition differential (Signal/noise in bandwidth $1/t$)	15 dB
I.f. bandwidth, B	500 c/s
V.f. bandwidth, b	10 c/s

No ping-to-ping integration is assumed, i.e. $N = 1$.

The noise power level in a band $1/t$ is -55 dB. With the i.f. bandwidth given above, $n = 50$, so, referring directly to Fig. 5 the detection loss is 3 dB whence the masking level of the noise is -52 dB. Taking the target strength and the recognition differential as both equal to 15 dB we have left, for attenuation both ways, a margin of 182 dB. In the absence of reverberation the range would be 7000 yards and as the detection loss is a mere 3 dB it would not appear to be very profitable to use a narrower i.f. band if we were concerned with noise alone.

The reverberation spectrum is narrow compared with the predetector bandwidth of the system, so in the present case the masking level of reverberation is nearly equal to its power level and is far higher than that of noise at all relevant ranges: in fact the range is limited to 2000 yards compared with 7000 yards if noise were the limiting factor.

Consider now the effect of tuning a filter of bandwidth 10 c/s to the echo frequency. The echo strength will be hardly affected, but, depending on the amount of target doppler, the reverberation will be attenuated considerably. For simplicity, assume that the filter characteristic is rectangular. From the analysis in Section 2 it can be shown that, if the target doppler is 90 c/s, the reverberation level has been reduced 21 dB for a target on the beam and over 29 dB if the target is ahead of the ship. Using the former figure, the detection range is now extended from 2000 yards to 6000 yards.

5. Conclusion

A method of treating signal detection has been developed for a background of reverberation, or noise, or both. A simple example has been given to show that it is misleading to calculate detection ranges by considering noise alone. It showed incidentally the value of a filter tuned to the target echo frequency: for the example chosen the effect of this was to increase the range (as limited by reverberation) threefold.

6. Acknowledgments

The author thanks the Admiralty for permission to publish this paper and his colleagues at the Admiralty Underwater Weapons Establishment, Portland, for encouragement and helpful comment, particularly Mr. P. R. Wallis whose early work provided the basis of Section 3.

7. References

1. J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals", (McGraw-Hill, New York, 1950).
2. S. O. Rice, "Mathematical analysis of random noise", *Bell Syst. Tech. J.*, **23**, No. 3, pp. 282–332, July 1944 and **24**, No. 1, pp. 46–156, January 1945.
3. D. Middleton, "Statistical Methods for the Detection of Pulsed Radar in Noise", Paper read at the London Symposium on Information Theory, 1952.
4. "Principles of Underwater Sound", National Defense Research Committee, Washington, D.C., 1946.

Manuscript first received by the Institution on 22nd June 1962 (Paper No. 788/SS21).

The 1963 Convention

"ELECTRONICS AND PRODUCTIVITY"—SOUTHAMPTON, 16th–20th APRIL

The Institution's Contribution to National Productivity Year

Synopses of Papers to be presented at the Convention

This is the first selection from papers which have so far been accepted for presentation during the Convention. The allocation of papers to particular sessions should be regarded as provisional and will be confirmed later. Further synopses of papers will be published in the March issue of *The Radio and Electronic Engineer.*

Tuesday afternoon 16th April INTRODUCTORY SESSION

Optimization Techniques and Process Control

PROFESSOR A. D. BOOTH, D.SC. (*Member*). (*University of Saskatchewan.*)

1. The problem of process optimization considered as a minimization in "n" dimensions.
2. An introduction to possible mathematical techniques.
3. Gradient methods—steepest descents.
4. Direct search by pointwise sampling.
5. Axial transformations as an aid to efficiency in dynamic situations.
6. Constraints.
7. Conditional probability analogues vs. the general purpose computer.

Wednesday 17th April SESSION ON "SENSING DEVICES, MEASUREMENT AND TELEMETRY"

Mass Flow Measurement with Turbine-type Flowmeters

I. C. HUTCHEON, M.A., and L. S. DUFFY. (*George Kent Limited.*)

The use of an electrical pick-off to detect rotor speed greatly extends the scope of the simple turbine-type liquid flowmeter by providing a signal which can be used for transmission, computation, or control. If the liquid temperature is measured as well, further possibilities are opened up.

A system of this type is described which measures the temperature and volume flow rate of the fuel delivered to an aircraft engine, and computes the mass flow rate and the total mass of fuel consumed. The principles apply equally to industrial mass flow measurement.

Determination of Sulphur Content of Hydrocarbons by Bremsstrahlung Absorption Measurement

T. B. ROWLEY. (*Isotope Developments.*)

The paper discusses the usefulness of beta-excited sources of electro-magnetic radiation (bremsstrahlung) for incorporation into a radiation absorption measuring equipment, designed primarily to indicate continuously the sulphur content of refinery hydrocarbon streams. A conventional Geiger counter/rate-meter system is used to give simplicity of design and reliability of operation of the detector/indicator equipment. An accuracy of better than 0·03% weight sulphur over long periods has been obtained under refinery operating conditions. A second equipment, based on the same principle of operation but designed for laboratory use on a wide range of base stocks, is also described, where the sulphur determination is independent of density, and hydrogen/carbon ratio error effects are much reduced.

The Development of Eddy Current Testing Techniques for Tube Inspection

D. TERRY. (*Accles and Pollock.*)

The paper outlines the need for a high-speed method of examining tubes for defects. A short history is given of the introduction of eddy current testing used for the examination of solid and hollow cylinders. Eddy current testing principles are explained, together with methods of measuring the effects of variations in a test specimen, with the aid of block circuit diagrams and complex impedance curves.

The merits of a.c. and d.c. saturation techniques are dealt with and a description is given of practical testing equipment which employs phase and modulation analysis. The application of phase sensitive devices and modulation analysis techniques are explained and their limitations assessed. The uses of internal probes and surface probes are briefly described.

A Frequency Meter with Continuous Digital Presentation

P. WOOD. (*The Plessey Company.*)

Many transducers produce an output whose frequency is proportional to the transduced quantity, e.g. some types of a.c. tacho-generator, vibrating cylinder pressure transducers etc. The normal method of converting the output into a suitable digital form is by counting the number of cycles that occur during a fixed time period. This paper describes an alternative approach in which the binary number corresponding to the input frequency is continuously available.

The system is based on the use of a reversible binary counter. The input frequency, converted into a suitable train of pulses, is applied to the "add" input of the counter, and a second train of pulses produced by a binary rate multiplier to the "subtract" input. The binary input to the rate multiplier is taken from the counter and therefore its output is proportional to the number stored in the counter and equilibrium will finally be reached. Circuits are included to prevent jitter of the display count, which could arise from a number of causes.

The main advantages of the system are that the count is available for continuous display and transmission, and there is less delay in responding to changes in the transduced parameter.

Leak Detection and the use of the Halogen Detector

T. S. WORTHINGTON (*Associate Member*).

When a system or component is required to remain pressure-tight, the question of leak testing, followed generally by the detection of leaks, has to be considered.

Equipment to be tested is manifold and new applications constantly arise. Heat exchangers, aircraft pneumatic controls, fuel tanks, refrigeration systems, gas-filled cables, pipelines laid for water, town gas or special "services", represent a cross-section from many industries, which are interested in these problems.

In the manufacture of components, such products as diaphragms, bellows or anything with a hermetic seal or closed system, e.g. transistors and various sensing devices, all require special tests.

Productivity is served by finding the leak at the inspection stages, with a degree of certainty.

The paper attempts to outline some of the parameters involved. Several different applications are covered, with reference to the methods adopted. A detailed description of the principles upon which the Halogen Detector operates is given since these are not commonly known.

Thursday 18th April SESSION ON "CONTROL AND INFORMATION PROCESSING"

Design Features of a Digital Computer for Industrial Process Control

J. A. FREER, M.A. (*International Systems Control.*)

The criteria which apply in designing a computer for use in controlling industrial processes differ radically from those which are taken into account in the design of a business or scientific data processing computer. This paper discusses these factors and their bearing on the economics, productivity and reliability of a computer controlled process. Among the topics considered are the following:

1. Functions required of a digital computer which is to be economically applied to control of industrial processes.

2. The flexibility of input/output arrangements, with special reference to

(*a*) Checking and calibrating the instruments which take the primary measurements.

(*b*) Providing fail-safe arrangements to minimize the effects of a computer error or failure.

(*c*) Allowing effective communication between the computer and the process operator.

3. Provision of special facilities for real-time operation, including a priority interrupt system.

4. System reliability and maintenance procedures and their effects on the overall productivity of the plant.

5. Programming features required to reduce the time required to code the problem.

These general topics are related to the particular designs adopted in the TRW-330 Digital Process Control Computers. Operating experience gained in the field with this and an earlier computer is cited to show how careful consideration of these design criteria leads to significant reductions in the operational and maintenance costs of process operation.

The Economic Justification of On-Line Computers for Process Control

W. T. LEE, B.SC. (*International Systems Control.*)

Digital computer systems for process control provide an opportunity for greater productivity by increasing plant throughout, increasing yield, giving closer control of product quality, reducing operating cost, and sometimes reducing expenditure in plant hardware. When considering the installation of a process computer system it is necessary to establish the potential for improvement through advanced control, how much this improvement would be worth in relation to the cost of the system, and whether means can be devised for controlling the process better with a computer. Each of these considerations is discussed in some detail. Finally, a number of examples are given of the gains possible from computer control, drawn largely from the author's own experience.

Friday 19th April SESSION ON "INDUSTRIAL APPLICATIONS OF ELECTRONIC SYSTEMS"

A High-Efficiency Low-Frequency Vibration Generator

B. H. VENNING, B.SC.(ENG.), A.C.G.I. (*Associate Member*) (*University of Southampton.*)

The production of continuous sinusoidal vibrations of low frequency or transients of a comparably long period is beyond the scope of existing electro-dynamic force generators where displacement is relatively limited. Moreover, the transformer-coupled amplifiers employed to drive the units have a limited low-frequency bandwidth unable to cater for the pulse shapes required.

This paper describes a system which has been devised to overcome these disadvantages, and also to demonstrate a means of high-efficiency amplification which can also be employed with the existing pattern of vibration generators. The principle components of interest are:—

(i) An electro-dynamic force generator with a travel of 6 in., employing a sliding coil to eliminate low-frequency resonances.

(ii) The power amplifier, employing a pulse-width modulated signal and output transistors operating as saturated switches, capable of an overall efficiency of 95–98%. Frequency bandwidth is d.c. to about 150 c/s, and the output power, at present 800 W, can be increased considerably.

(iii) Force, displacement and acceleration transducers, all working as variable-capacitance devices.

(iv) High input-impedance pre-amplifiers (transistorized) for use with barium-titanate transducers, with input impedance values in the region 200–1000 megohms.

Data Logging in Power Generating Stations

W. E. WILLISON (*Associate Member*) (*Elliott Brothers (London).*)

The applications of alarm scanning and data logging techniques in the operation of boiler-turbine units in conventional power generating stations are discussed. Consideration is given to the operational requirements, with emphasis on the importance of careful selection of the data to be logged.

A generalized equipment specification is developed covering such points as: environmental conditions, availability and reliability, scanning speeds, periodical logging intervals, measurement accuracy, print-out and display formats, etc. Problems encountered at the transducer interface and the methods of dealing with a wide range of different types of input signal are discussed. These include such things as the rejection of interference pick-up from "noise" sources, conversion of input signals to a common-language and the linearization of transducer characteristics.

A range of modular transistor units which can be used for the construction of "custom-built" systems is described. A description is given of a 220-point alarm-scanning and data-logging system, several of which have been installed at the Central Electricity Generating Board's Northfleet power station. An account is given of the practical experience gained as a result of this project.

The Footprint-and-Wheel Measuring System for Steel Tubes or Bars

R. ASHFORD AND P. HUGGINS (*Member*). (*T.I. Steel Tube Division.*)

The paper describes a method of measurement of throughput of steel from a rolling mill or on a conveyor. Magnetic spots are created at one foot intervals along the extensive length: these are counted, thereby recording the aggregate footage. A wheel of one foot circumference is also held in contact with the material and this is used to measure that fraction of a foot not measured by the foot counter. The combination of footprints and wheel makes for an accurate system of measurement at high speeds where the surface conditions of the steel are such that a simple wheel and counter would be highly inaccurate due to slip.

Laboratory experiments are described showing the potential accuracy of the system. A proposed application for the measurement of steel tubes is described and its probable performance evaluated.

U.D.C. 621.375.4

An Engineering Approach to the Design of Transistor Feedback Amplifiers

By

E. M. CHERRY, M.Sc., Ph.D.†

Based on a paper first published in The Proceedings of the Institution of Radio Engineers, Australia. The original paper was nominated by the Brit.I.R.E. for the award of the Norman W. V. Hayes Memorial Medal for 1962.

Summary: A technique is developed for the design of transistor feedback amplifiers, based on the use of impedance mismatches between stages. Expressions are derived for the transmittance of the four basic building blocks—the series and shunt single stage feedback amplifiers and the current and voltage feedback pairs—and methods of interconnection which achieve the mismatch are considered. The expressions for the transmittance are both simple and highly accurate, yet they involve no quantitative information about the transistors at all. It is concluded that transistor circuitry is far more designable than valve circuitry.

Three examples in the use of the design philosophy are given, covering the audio and video frequency ranges up to 20 Mc/s.

1. Introduction

There are a number of engineers whose formal training in electronics was confined to valve circuits, and who have an instinctive mistrust of transistors. This attitude probably springs from the general human mistrust of anything new, but the situation has been aggravated by some published papers which purport to give analyses of quite simple circuits. Particularly is this the case for transistor feedback amplifiers.

The source of the trouble is in the fact that there is a much greater tolerance on the parameter β_N of a transistor than there is on any parameter of a valve; β_N is not usually specified with a tolerance closer than -50%, $+100\%$. This has the unfortunate consequence that the gain of a transistor amplifier (without feedback) is a relatively unknown quantity, so that a transistor amplifier whose gain is to be known with any precision must be a feedback amplifier. Now the transistor is a bilateral device even at low frequencies; the rigorous analysis of a feedback amplifier with bilateral active devices is a mathematical exercise of the most extreme complication. The heart of the trouble is that the vast majority of writers on the subject of transistor feedback amplifiers do not appear to have realized what approximations can and should be made early in the analysis, nor have they appreciated the physical implications of these approximations.

The aim of this paper is two-fold. First, it is intended to fill a serious gap in the literature on transistor feedback amplifiers. Up to the present, the majority of papers on this subject have fallen into one of two classes. On the one hand there is the purely descriptive paper which gives little more than the statement that a particular circuit happens to work. On the other hand there is the formal mathematical approach, usually in terms of circuit parameters, which results in a most impressive array of algebra. For the following reasons, this latter approach is of little use in practical design work.

First, a number of approximations must be made to simplify the circuit parameter equations to a usable form, but there is little indication of which approximations are valid in any particular set of circumstances. Secondly, the transistor data to be substituted into circuit parameter equations is in a most inaccurate form; the dependence on the unknown β_N is concealed in the algebra.

A circuit parameter approach in its exact form may have application to the most precise calculations of gain in cases where the transistor parameters have actually been measured. For more normal purposes, a simple, direct approach can give just as accurate an expression in a fraction of the time. This paper outlines such an approach, which is, in addition, particularly suited to practical design work.

The second aim of this paper is to show that transistor circuits are far more designable than valve circuits. For the vast majority of applications there is no need for any quantitative transistor data at all. The current amplification factor β_N, the parameter with the greatest variation from type to type and from transistor to transistor of a type, occurs only in second-order expressions of an intelligent design philosophy; a simple statement that a transistor is a "high" or

† Department of Electrical Engineering, The University of Melbourne, Victoria; now with Bell Telephone Laboratories, Murray Hill, New Jersey, U.S.A.

"low" gain type is sufficient. Further, the frequency response of a transistor is adequately specified for most purposes if it is known that the type is "audio" "r.f." or "drift". Very occasionally some information regarding the causes of the high frequency performance may be required, e.g. the relative sizes of the two components of emitter capacitance; such information does not involve complicated algebra, but rather a knowledge of the physical principles of transistor action.

The design philosophy outlined in the following pages has resulted in some quite efficient and designable circuits, as the examples given show. The philosophy is directly applicable to both audio amplifiers and to video amplifiers with bandwidths as great as 20 Mc/s. It is hoped that this paper may persuade even a few that the transistor is not the mysterious black box surrounded by a mass of algebra that some would have it.

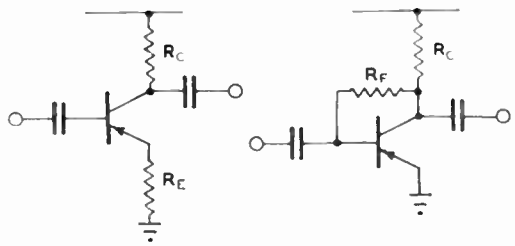


Fig. 1. Single-stage feedback amplifier circuits; (*a*) series feedback, (*b*) shunt feedback. The biasing circuits have been omitted for simplicity. Note that the output current in the series circuit is taken as the collector current. In both circuits, the load resistance R_L is the parallel combination of R_C and any external load.

2. Single Stage Feedback Amplifiers

There are two basic forms of single-stage feedback amplifier, shown in Fig. 1. These circuits are duals; the series feedback circuit, Fig. 1 (*a*), is a voltage-in-to-current-out amplifier, while the shunt feedback circuit, Fig. 1 (*b*), is a current-in-to-voltage-out amplifier. If the feedback is increased so that the loop gain is very large, the former has a transconductance of G_E (i.e., $1/R_E$) and the latter has a transresistance of R_F:

series: $$\left[\frac{\partial I_o}{\partial V_i}\right]_{V_o} = G_E = 1/R_E \qquad \text{.....(1a)}$$

shunt: $$\left[\frac{\partial V_o}{\partial I_i}\right]_{I_o} = -R_F \qquad \text{.....(1b)}$$

With the majority of practical circuits, there is sufficient feedback for the above expressions to apply quite accurately. Further, the requirement of constant V_o or I_o is not at all stringent, so the transmittance is given fairly accurately by:

series: $$\frac{dI_o}{dV_i} = 1/R_E \qquad \text{.....(2a)}$$

shunt: $$\frac{dV_o}{dI_i} = -R_F \qquad \text{.....(2b)}$$

In principle, the calculation of the overall gain of a multistage amplifier involves nothing more than the multiplication of the individual stage transmittances, as above. If a certain amount of discretion is used in making second-order corrections, it is not too much to expect the answer obtained to be well within $\pm 0{\cdot}5$ dB per stage. In order to see how much discretion is required, it is necessary to derive an accurate expression for the transmittance and a few other approximate expressions for the orders of magnitude of the input and output impedances, and then see what approximations can be made. The emphasis in the following analyses is on the physical mode of operation of the circuit, as this shows best what the approximations imply. The analyses are directed towards justifying the approximate expressions quoted above.

3. Series Feedback Amplifier

For junction transistors operating at low frequencies, the current gain α_N from emitter to collector is very closely unity and real; α_N is rarely less than 0·95, so the emitter and collector currents are closely equal. Therefore, in the series feedback circuit there appears across R_E a voltage which is directly proportional to the collector current, i.e. the output current. This feedback voltage is added in series with the input voltage to give the signal applied between the base and emitter of the transistor.

Thus, the feedback is series at both the input and output, so that both the input and output impedances are increased. Further, since a voltage proportional to the output current is fed back into the input circuit, the transmittance defined as i_o/v_i is reduced by some factor F, the return difference, and distortion in this transmittance is reduced by the same factor F. It is not immediately obvious what the effect on transmittance defined as i_o/i_i, v_o/i_i or v_o/v_i will be.

3.1. *Operation of the Circuit*

The mutual conductance of a transistor operating in the common emitter configuration is

$$g_m \equiv \left[\frac{\partial I_C}{\partial V_B}\right]_{V_C}$$

If the emitter-to-base voltage of a transistor is changed by δV_{EB}, the emitter current changes by

$$\delta I_E = \delta V_{EB}/r_E$$

where r_E is the emitter resistance, and is given by

$$r_E = (kT/q)/I_E$$
$$\simeq 25/I_E \text{ ohms} \qquad \text{.....(3)}$$

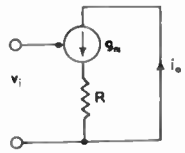


Fig. 2. Location of feedback resistor R for controlling the mutual conductance of a generator.

at 16° C for I_E in milliamperes.

Therefore

$$\delta I_C = \alpha_N \delta I_E = (\alpha_N/r_E)\delta V_{EB}$$

Now δV_{EB} can be produced by holding the emitter voltage fixed and changing the base voltage, so that in the common emitter configuration

$$\delta V_{EB} \equiv \delta V_B$$

Hence $$\delta I_C = (\alpha_N/r_E)\delta V_B$$

and therefore $$g_m \equiv \left[\frac{\partial I_C}{\partial V_B}\right]_{V_C} = \frac{\alpha_N}{r_E}$$

i.e. $$g_m \simeq 1/r_E \qquad \text{.....(4)}$$

Again, $$g_m = \alpha_N/r_E = \alpha_N(I_E/25) \quad \text{A/V}$$
$$= 40\alpha_N I_E \quad \text{mA/V}$$

i.e. $$g_m \simeq 40 I_C \quad \text{mA/V}$$

for I_C in milliamperes. Thus, the mutual conductance of all transistors is the same at the same collector current.

If a resistance R is added to a generator of mutual conductance g_m, as shown in Fig. 2, the combination has a transconductance G_t given by

$$1/G_t = 1/g_m + R$$

If the generator is a transistor, and R is an unbypassed emitter resistor R_E, then

$$1/G_t \simeq r_E + R_E$$

Therefore $$G_t \simeq 1/(r_E + R_E)$$

Thus, the ratio of the mutual conductance of a transistor with and without series feedback, that is, the return difference is

$$F \equiv (g_m/G_t) = (r_E + R_E)/r_E$$
$$= 1 + R_E/r_E$$

and for heavy feedback, i.e. large F, it is sufficient that

$$R_E \gg r_E$$

so that $$G_t \equiv i_o/v_i \simeq 1/R_E \qquad \text{.....(5)}$$

3.2. *Network Equations*

Figure 3 shows the complete equivalent circuit† for a series feedback amplifier. Note that R_L is the total load resistance which, in general, has three components; these are the collector supply resistor for the transistor, the base biasing resistors for the following transistors, and the input resistance of the following transistor. The following results are derived in any of the standard texts:

$$G_t = \frac{i_o}{v_i} = \frac{\alpha_N}{r_B/\beta_N + r_E + R_E} \simeq \frac{1}{R_E} \qquad \text{.....(6a)}$$

$$A_v = \frac{v_o}{v_i} = -\left(\frac{i_o}{v_i}\right) \times R_L \qquad \text{.....(6b)}$$

$$A_i = \frac{i_o}{i_i} = \beta_N \qquad \text{.....(6c)}$$

$$R_t = \frac{v_o}{i_i} = -\left(\frac{i_o}{i_i}\right) \times R_L \qquad \text{.....(6d)}$$

$$r_i = \frac{v_i}{i_i} = r_B + \beta_N(r_E + R_E) \simeq \beta_N R_E \qquad \text{.....(6e)}$$

and, if R_S tends to zero

$$r_o = r_C(1 + R_E/r_E) \qquad \text{.....(6f)}$$

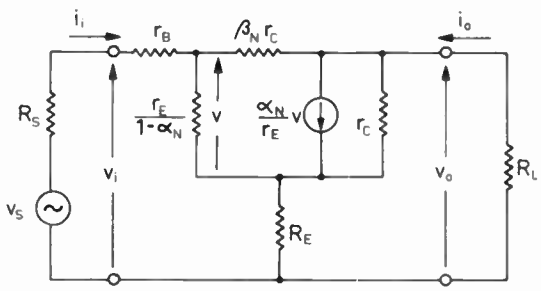


Fig. 3. Equivalent circuit for a series feedback amplifier.

From these expressions, it is seen that to a first approximation the inclusion of R_E has no effect on transmittance defined in terms of a current input; *there is no feedback when a series feedback amplifier is fed from a current source.*

In order to study the effect of R_E on transmittance as defined in terms of a voltage input, it is necessary to derive an exact expression for this transmittance.

$$G_t = \frac{i_o}{v_i} = \frac{\alpha_N}{R_E} \cdot \frac{1}{1 + (\alpha_N/R_E G_{to})} \qquad \text{.....(7)}$$

where G_{to} is the transfer conductance of the transistor without feedback, given by

$$G_{to} = \frac{\alpha_N}{r_B/K\beta_N + r_E} \cdot \frac{r_C}{r_C + R'_L} \qquad \text{.....(8a)}$$

where $$K = \frac{r_C + R'_L}{r_C + 2R'_L} \qquad \text{.....(8b)}$$

and $$R'_L = R_L + R_E \text{ (total load)} \qquad \text{.....(8c)}$$

The basis of approach to the design of a series feedback amplifier is to make the term $(\alpha_N/R_E G_{to})$ in the denominator of eqn (7) small, so that if there is an error in the value of G_{to} due to an uncertainty in

† In the hybrid π equivalent circuit used in Fig. 3, the parameters r_B, r_C and r_E are not the parameters r_b, r_c and r_e given in the T equivalent circuit data by many manufacturers.

transistor parameters, etc., the value of G_t is not significantly affected. For example, if a circuit is designed so that $(\alpha_N/R_E G_{to})$ is expected to be 0·1, then, even if G_{to} is in error by a factor of two, the error in G_t is less than 10%. The same procedure allows an approximate expression to be used for G_{to}, since errors due to the approximation are insignificant. An approximate expression for G_{to} which is unlikely to be in error by anything like as much as 50% is

$$G_{to} = \frac{\alpha_N}{r_B/\beta_N + r_E} \qquad \text{.....(9)}$$

If this expression is substituted into eqn (7), the value of G_t is given to an excellent approximation by

$$G_t = \frac{i_o}{v_i} = \frac{\alpha_N}{r_B/\beta_N + r_E + R_E} \qquad \text{.....(10)}$$

3.3. *Sources of Error in Eqn* (10)

The variation of r_E with temperature is not great and follows a known law, eqn (3). Therefore, if R_E is made ten times the expected value of r_E, $(r_E + R_E)$ will be very close to $1{\cdot}1R_E$.

Also, r_B is largely determined by the manufacturing techniques used in the production of a transistor of a particular general type; r_B is large for low frequency transistors and small for high frequency types. The values

$r_B = 300\,\Omega$ for audio types (excluding power transistors) and

$r_B = 50\,\Omega$ for r.f. and drift types

are of the right order of magnitude. This means that r_B/β_N is within the range zero to 10 Ω, which is probably less than r_E at a typical operating current—200 μA to 10 mA; particularly is this so because audio transistors with the higher values of r_B/β_N are unlikely to be operated at the higher currents for which r_E is small. Thus r_B/β_N in eqn (10) is a correcting term which can generally be neglected, or at the most be given an order-of-magnitude value.

The current gain α_N may be taken as unity for many purposes. A good working rule for the most precise calculation is:

$\alpha_N = 0{\cdot}99$ for high gain types

$\alpha_N = 0{\cdot}97$ for medium gain types

$\alpha_N = 0{\cdot}94$ for low gain types

The corresponding values of β_N are 100, 30, 15.

Finally, there is a theoretical restriction on the value of R_L, for the error in eqn (9) reaches 30% when

$$R'_L \geqslant r_C$$

and therefore, if R_E is ten times r_E, the error in eqn (10) is 3%. This restriction is of no practical consequence. The value of r_C at 1 mA emitter current is unlikely to be less than 20 kΩ. Therefore, with any normal supply voltage and operating current, the collector supply resistor R_C will be large enough to cause bottoming long before it has reached a value equal to the critical value of R_L. But R_L cannot exceed R_C. Therefore, R_L cannot even approach the critical value.

Thus, eqn (10), together with a knowledge of the physical principles of transistors, enables the gain of a series feedback amplifier to be designed to better than 5% with nothing more than a statement of the general type of a transistor. Without even this information, the gain can probably be estimated to better than 10%.

3.4. *Output Resistance*

It is worth pointing out that the output resistance of a series feedback amplifier such as that in Fig. 1 (*a*) measured at the output terminals is not a large value, of the order of r_C; it is the output resistance of the transistor itself which has this value. The output resistance of the amplifier is the parallel combination of R_C and the output resistance of the transistor, which for all practical purposes is equal to R_C.

4. Shunt Feedback Amplifier

In the shunt feedback circuit, Fig. 1 (*b*), a current flows from the output to the input circuit via R_F. This current is (approximately) proportional to the output voltage, and is added in shunt to the input current. The feedback is therefore shunt at both the input and output, so that the input and output impedances are reduced. Further, since a current proportional to the output voltage is fed back into the input, the transmittance defined as v_o/i_i will be reduced by a factor F, and distortion in this transmittance will also be reduced by F. It is not immediately obvious what the effect on transmittance defined as v_o/v_i, i_o/v_i and i_o/i_i will be.

4.1. *Operation of the Circuit*

The transconductance of a transistor is given by eqn (4) so that the voltage gain of a shunt feedback amplifier is given by

$$A_v = v_o/v_i = -g_m R'_L \simeq -R'_L/r_E$$

where R'_L is the effective load—approximately the parallel combination of R_L, R_F and the output resistance of the transistor—and R_L is the parallel combination of R_C and the external load.

Now, an impedance Z_F connected between the output and input of an amplifier with voltage gain A_v is equivalent (so far as the input circuit is concerned) to an impedance $Z_F/(1-A_v)$ connected to earth, e.g. the Miller input capacitance of a triode valve. The total input resistance r_i of a shunt feedback amplifier is therefore r_1 in parallel with $R_F/(1-A_v)$ where r_1 is the input resistance of the transistor itself.

The criterion for high loop gain in a shunt feedback amplifier is

$$R_F/(1-A_v) \ll r_1$$

so that $$r_i \simeq R_F/(1-A_v)$$

An input current i_i then develops an input voltage

$$v_i = i_i r_i = i_i R_F/(1-A_v)$$

and the output voltage is

$$v_o = A_v v_i = i_i R_F A_v/(1-A_v)$$

i.e. $$v_o \simeq -i_i R_F$$

Therefore $$R_t \equiv v_o/i_i \simeq -R_F \qquad \text{.....(11)}$$

where R_t is the transresistance of the amplifier, and R_t is independent of A_v if A_v is large.

Physically, if A_v changes for any reason, the input resistance is changed so that a different input voltage is developed by the input current, and this change in input voltage tends to restore the output at the new value of A_v.

4.2. *Network Equations*

A reasonably complete equivalent circuit for a shunt feedback amplifier stage is shown in Fig. 4. The following approximate results are derived in any of the standard texts:

$$G_t = \frac{i_o}{v_i} = \frac{\alpha_N}{r_B/\beta_N + r_E} \cdot \frac{R_F}{R_F+R_L} \qquad \text{.....(12a)}$$

$$A_v = \frac{v_o}{v_i} = -\left(\frac{i_o}{v_i}\right) \times R_L \qquad \text{.....(12b)}$$

$$A_i = \frac{i_o}{i_i} = -\left(\frac{v_o}{i_i}\right)\Big/ R_L \qquad \text{.....(12c)}$$

$$R_t = \frac{v_o}{i_i} = -\frac{R_F}{1+(R_F+R_L)/\beta_N R_L} \simeq -R_F \qquad \text{.....(12d)}$$

$$r_1 = \frac{v_i}{i_B} = r_B + \beta_N r_E \qquad \text{.....(12e)}$$

$$r_i = \frac{v_i}{i_i} = (r_B/\beta_N + r_E)(1+R_F/R_L) \qquad \text{.....(12f)}$$

and, if R_S tends to infinity

$$r_o = R_F/\beta_N \qquad \text{.....(12g)}$$

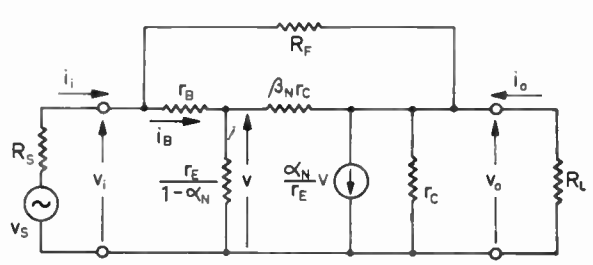


Fig. 4. Equivalent circuit for a shunt feedback amplifier.

From these equations it is seen that the effect of R_F on transmittance defined in terms of an input voltage is second order, namely, the shunting effect R_F on R_L; the load for the transistor is R_F and R_L in parallel. *There is no feedback when a shunt feedback stage is fed from a voltage source.*

In order to study the effect of R_F on transmittance defined in terms of a current input, it is necessary to derive an exact expression for the transmittance. Such an expression is derived in Appendix 1:

$$R_t = \frac{v_o}{i_i} = -\frac{R_F}{1-(R_F+r_1)/A_v r_1}$$

where r_1 and A_v are the *exact* input resistance of the transistor and the voltage gain.

4.3. *Approximations*

For high loop gain it is sufficient that

$$\{(r_1+R_F)/A_v r_1\} \ll 1$$

Therefore, small errors in r_1 and A_v are corrections within a correcting term, so that the approximate expressions of eqn (12) may be substituted with negligible loss of accuracy, giving

$$R_t = \frac{v_o}{i_i} = -\frac{R_F}{1+\dfrac{(R_F+r_B+\beta_N r_E)(R_F+R_L)}{\beta_N R_F R_L}} \qquad \text{.....(13a)}$$

With the majority of practical shunt feedback circuits,

$$R_F \gg (r_B+\beta_N r_E)$$

Therefore, to a further approximation

$$R_t = \frac{v_o}{i_i} \simeq -\frac{R_F}{1+(R_F+R_L)/\beta_N R_L} \qquad \text{.....(13b)}$$

If the expression K/β_{NA} is given the value 0·05, the expression K/β_N will be within the range zero to 0·1 for β_N in the range

$$\tfrac{1}{2}\beta_{NA} \leqslant \beta_N \leqslant \infty$$

where β_{NA} is the value of β_N expected for an "average" transistor. Therefore, if the expression $(R_F+R_L)/\beta_{NA}R_L$ is given the value 0·05, the transmittance of a shunt feedback amplifier will be within $\pm 5\%$ of the value

$$v_o/i_i = -R_F/1{\cdot}05$$

for β_N in the above range. This expression applies reasonably accurately even if R_F is only two or three times greater than $(r_B+\beta_N r_E)$.

Thus the gain of a shunt feedback amplifier is known to within $\pm 5\%$ if

$$(R_F+R_L)/\beta_{NA}R_L = 0{\cdot}05$$

i.e. $$R_F/R_L + 1 = 0{\cdot}05\beta_{NA}$$

i.e. $$R_F/R_L = 0{\cdot}05\beta_{NA} - 1$$

for β_N in the above range.

By performing a series of calculations similar to that above, it is possible to construct Table 1 which gives the value of R_F/R_L in terms of β_{NA} for any gain accuracy requirement. Table 1 should be used with a certain amount of caution; it has been constructed on the assumption that $R_F \gg (r_B+\beta_N r_E)$. If this is not the case, it is relatively simple to calculate a new value for R_F/R_L.

Table 1

Range of β_N	Gain accuracy requirement ± dB	R_F/R_L	$-v_0/i_i$
$\frac{1}{2}\beta_{NA}$ to ∞	0·5	$0{\cdot}05\beta_{NA}-1$	$0{\cdot}95R_F$
	1	$0{\cdot}1\beta_{NA}-1$	$0{\cdot}90R_F$
	3	$0{\cdot}4\beta_{NA}-1$	$0{\cdot}71R_F$
$\frac{2}{3}\beta_{NA}$ to $2\beta_{NA}$	0·5	$0{\cdot}1\beta_{NA}-1$	$0{\cdot}90R_F$
	1	$0{\cdot}3\beta_{NA}-1$	$0{\cdot}77R_F$
	3	$0{\cdot}6\beta_{NA}-1$	$0{\cdot}38R_F$

In general, a variation of β_N on the high side of β_{NA} has less effect on gain than a variation on the low side. Therefore, it is better to base the design of a shunt feedback amplifier on a rather low value of β_{NA}. A good working rule is:

β_{NA} = 30 for medium gain transistors

β_{NA} = 100 for high gain transistors.

5. Cascades of Single-Stage Feedback Amplifiers

It has been shown above, that a series feedback amplifier

(i) is a voltage-in-to-current-out amplifier,

(ii) has a high input resistance, and should be fed from low resistance source,

(iii) has a high output resistance, and should work into a low load resistance.

and conversely a shunt feedback amplifier,

(i) is a current-in-to-voltage-out amplifier,

(ii) has a low input resistance and should be fed from a high resistance source,

(iii) has a low output resistance, and should work into a high load resistance.

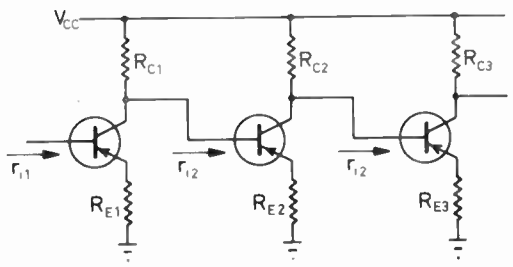


Fig. 5. Elemental diagram for cascaded series feedback stages.

Intuitively, therefore, it is expected that a cascade of feedback stages in which series and shunt amplifiers alternate would be the most satisfactory. This is in fact the case; all possible methods for cascading feedback stages are analysed in the following sections, and it is shown that the alternate cascade is best by about a factor of three per stage.

5.1. *Series Feedback Stages*

In a cascade of series feedback stages, Fig. 5, the load resistance for the first stage is the input resistance of the second, neglecting, for the moment, the collector supply resistor R_C. If the input voltage to the first stage is known, the output current from the first stage can be defined to any degree of precision by the application of sufficient series feedback to the first transistor. Equally, it is possible to define the output current from the second transistor if its input voltage is known; this input voltage is given by

$$v_{i2} = v_{o1} = i_{o1} \times R_{L1}$$

Now, R_{L1} is the input resistance of the second transistor:

$$R_{L1} = r_{i2} \simeq \beta_{N2} R_{E2}$$

Therefore, the voltage at the junction of the stages is not known even though i_{o1} is known, because it is directly proportional to the relatively unknown β_{N2}. Thus, it is not sufficient to apply series feedback to stabilize the transfer conductance of the individual stages; it is necessary also to stabilize the load resistance, and this can be done by shunting down the large but unknown input resistance of the following stage with a small, known resistance R_C. Fundamentally, the problem is that there is interaction between stages; the voltage gain of one stage depends on β_N for the following stage.

If β_{NA} represents the value of β_N expected for an "average" transistor, a typical production tolerance is

$$\tfrac{2}{3}\beta_{NA} \leqslant \beta_N \leqslant 2\beta_{NA} \qquad \text{.....(14)}$$

Making this numerical assumption so as to simplify the algebra, it follows that the total load in a cascade of identical stages is known to within ± 5% if R_C is set equal to

$$R_C = 0{\cdot}11 r_i = 0{\cdot}11\{r_B+\beta_{NA}(r_E+R_E)\}$$

so that R_L is

$$R_L = 0{\cdot}1\{r_B+\beta_{NA}(r_E+R_E)\} \pm 5\%$$

The voltage gain per stage is given by

$$A_v = G_t \times R_L$$

and if R_E is made large enough to stabilize the transfer conductance G_t, A_v is also known to within ± 5%. Substitution from eqn (10) therefore gives the maximum gain available with ± 5% accuracy as

$$A_{v\,\max} = 0{\cdot}1\beta_{NA}$$

Over two stages, the maximum gain available with $\pm 10\%$ uncertainty is

$$A_{v\,max} = (0{\cdot}1\beta_{NA})^2 \qquad \text{.....(15)}$$

In practice, the shunting resistor R_C is made up of the collector supply resistor for the first transistor in parallel with the biasing potentiometer of the second transistor.

5.2. *Shunt Feedback Stages*

In a cascade of shunt feedback stages, Fig. 6, it is necessary to interpose a series resistor R_S between the stages; this resistor swamps out the small but uncertain input resistance of the second stage, and thereby converts the voltage output from the first stage into a current drive for the second.

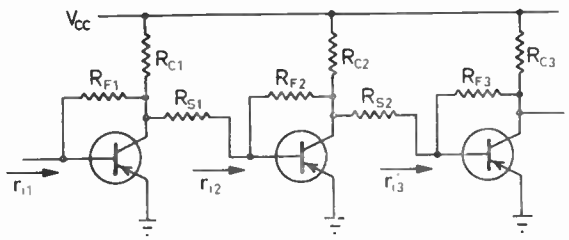


Fig. 6. Elemental diagram for cascaded shunt feedback stages.

The effective load resistance R_L for the first stage is given by

$$R_{L1} = R_S + r_{i2}$$

neglecting, for the moment, the collector supply resistor R_c, which serves no useful purpose at signal frequencies. Table 1 shows that, if the spread in β_N is as given by eqn (14), the transfer resistance is known to within $\pm 5\%$ if

$$R_F/R_L \geqslant 0{\cdot}1\beta_{NA} - 1$$

The current gain per stage is given by eqn (12c) as

$$A_i = R_t/R_L$$

and substituting from eqn (12d) gives

$$A_i \simeq R_F/R_L$$

Thus, the maximum gain available with $\pm 5\%$ accuracy is

$$A_{i\,max} = 0{\cdot}1\beta_{NA} - 1$$

and over two stages, the gain available with $\pm 10\%$ accuracy is

$$A_{i\,max} = (0{\cdot}1\beta_{NA} - 1)^2 \qquad \text{.....(16)}$$

This maximum gain is achieved when R_S is set equal to

$$R_S \simeq R_S + r_i = R_L = (0{\cdot}1\beta_{NA} - 1)R_F$$

Essentially, the limit to the gain per stage arises from the interaction between stages; the input resistance of one stage (including R_S) loads the preceding stage.

In practice, when the collector supply resistor R_C is taken into account, the loading on a shunt feedback stage is increased, and less stable gain is available. It is apparent that R_C should be made as large as possible; half the available gain is wasted when R_C is as small as R_S.

Finally, whilst the limiting gains for cascades of series or shunt feedback have been derived in terms of A_v and A_i respectively, either method for cascading stages stabilizes all possible transfer functions between all stages in the amplifier. Figure 7 shows this for the cascade of series feedback stages; in all cases, the limiting gain over two stages is given by eqn (15). A cascade of shunt stages may be treated similarly, and the limit is always given by eqn (16).

5.3. *Alternation of Stages*

Both the above methods for cascading single-stage feedback amplifiers have been in use for some years, but, as is shown above, neither method is particularly satisfactory. A third method for cascading stages—the use of alternate series and shunt feedback stages—is much more satisfactory. The low input resistance of a shunt stage forms an ideal load for a series stage, and the high input resistance of a series stage forms an ideal load for a shunt stage. Similarly, the high output resistance of a series stage forms an ideal source for a shunt stage, and the low output resistance of a shunt stage forms an ideal source for a series stage.

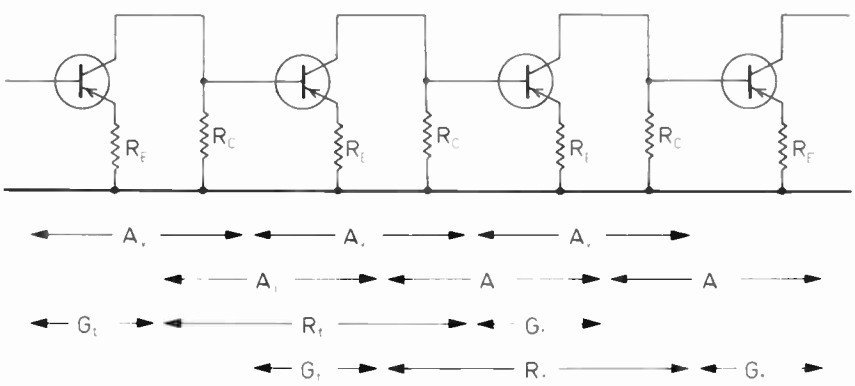


Fig. 7. Cascaded series feedback stages, showing how all transfer functions of all transistors are stabilized.

Basically what is done is to introduce a gross impedance mismatch between the output of one stage and the input of the following stage. In this way the individual stages are effectively isolated from each other; the transmittance of each stage depends only on that stage, so a much higher gain per stage can be realized for the same uncertainty—cf. cascades of series and shunt stages, as above. The whole concept of impedance mismatching is fundamental in the production of designable circuits; it is so fundamental that the performance of the feedback pairs, section 6, is discussed only for the mismatched case.

Figure 8 shows an elemental circuit diagram. As suggested in Sect. 2, the gain is found by multiplying the individual stage transmittances, given by eqns (5) or (10), and (11) or (13) respectively. Thus, when a series stage is followed by a shunt stage, the approximate voltage gain is

$$A_v \simeq G_{t1} \times R_{t2} \simeq \frac{1}{R_{E1}} \times R_{F2} \quad \text{.....(17a)}$$

whilst for a shunt stage followed by a series stage, the exact current gain is

$$A_i = R_{t2} \times G_{t3} \simeq R_{F2} \times \frac{1}{R_{E3}} \quad \text{.....(17b)}$$

A small correction is required in a precise calculation of gain when a shunt stage follows a series stage. Whilst the input resistance of the shunt stage is small, it is not zero. Therefore, a small fraction of the output current from the series stage flows into its collector supply resistor R_{C1}. The coupling efficiency η is given by

$$\eta = \frac{R_C}{R_C + r_{i\,(\text{shunt})}} \quad \text{.....(18)}$$

It is rare that η is less than 0·95, so the approximate expression for r_i, eqn (12f), is adequate even in a precise calculation of gain. It is apparent that R_C should be as large as possible. No correction is required when a series stage follows a shunt stage, since the output voltage of the shunt stage is exactly equal to the input voltage of the series stage.

There is an interaction between stages when a series stage follows a shunt stage. If the collector supply resistor R_{C2} in Fig. 8 is neglected, the load on

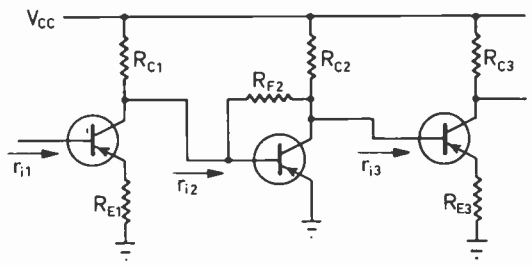


Fig. 8. Elemental circuit diagram for alternate series and shunt feedback stages.

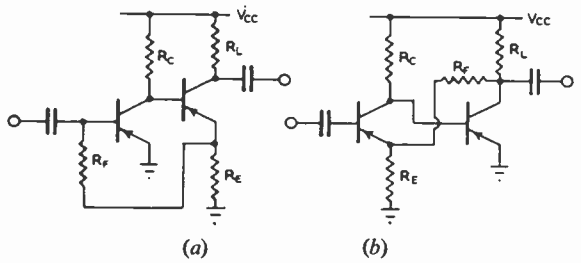


Fig. 9. Circuit diagrams for the feedback pairs; (*a*) current feedback, (*b*) voltage feedback. The biasing components have been omitted for simplicity. Note that the output current in the current feedback pair is taken as the collector current of T_2. In both circuits, R_L includes any external load.

the shunt stage is the input resistance of the series stage. Table 1 shows that the transfer resistance of a shunt feedback stage is known to within ±5% if

$$R_F/R_L \geqslant 0{\cdot}1\beta_{NA} - 1$$

and substituting

$$R_{L\,(\text{shunt})} = r_{i\,(\text{series})} \simeq \beta_N R_E$$

gives

$$\frac{R_F}{\beta_N R_E} \geqslant 0{\cdot}1\beta_{NA} - 1$$

But the current gain of the two stages is given by eqn (17b). It follows that the limiting gain for ±10% accuracy over the two stages is

$$A_{\max} = \beta_{NA}(0{\cdot}1\beta_{NA} - 1) \quad \text{.....(19)}$$

Comparison of eqn (19) with either eqns (15) or (16) shows that the advantage of the alternate cascade is about a factor of three per stage.

In contrast with the cascades of similar stages, a cascade of alternate stages stabilizes only one transfer function between any two points in the amplifier. Equation (19) gives the limiting voltage or current gain of two stages, as appropriate.

In a practical case, the collector supply resistor R_{C2} in Fig. 8 increases the loading on the shunt feedback stage, and reduces the available stable gain. As when a shunt stage follows a series stage, it is apparent that R_C should be made as large as possible. It is most important in the design of an alternate cascade that the loading effect of both R_C and $r_{i\,(\text{series})}$ on a shunt stage should be taken into account.

6. Feedback Pairs

There are two basic forms of the two-stage feedback amplifier, the current and voltage feedback pairs (Figs. 9 (*a*) and 9 (*b*)). The former is a current amplifier; it should work into a low (ideally zero) load resistance, it has a low input resistance and a high output resistance, and there is no feedback when it is fed from a voltage source. Conversely, the latter is a

voltage amplifier, it should work into a high (ideally infinite) load, it has a high input resistance and a low output resistance, and there is no feedback when it is fed from a current source. As with the single-stage feedback circuits, the requirement of low or high load resistance is not very stringent, and the gains are given quite accurately by

current feedback pair: $A_i \simeq R_F/R_E$ .....(20a)

voltage feedback pair: $A_v \simeq R_F/R_E$ . .....(20b)

Because there are so many more parameters in a two-stage feedback amplifier than in a single-stage, it is not practicable to derive an expression for the gain of such an amplifier when it is operating under non-ideal conditions, that is, when it is not operating under mismatched conditions. It will therefore be assumed that in a cascade of feedback pairs, all pairs are of the same type; if it is necessary to change from current feedback pairs to voltage feedback pairs or vice versa, it is necessary to interpose a single shunt or series stage, as appropriate to preserve the impedance mismatch. It may be shown that in such circumstances, the maximum gain available with $\pm 10\%$ uncertainty from either pair is

$$A_{max} = 0{\cdot}1(\beta_{NA})^2 \qquad .....(21)$$

As with the single-stage circuits, the emphasis in the following analyses is on the mode of operation of the circuit. The consequences of any approximations in a practical design may therefore be better appreciated. For the same reason, several minus signs, signifying phase reversal, are omitted.

6.1. *Current Feedback Pair*

The current feedback pair is in many ways similar to a combination of the two single-stage circuits.

Suppose initially that R_L is negligible. This is a good approximation; R_L is very much smaller than the output resistance of T2 particularly as T2 has an unbypassed emitter resistor—see Sect. 3. The circuit consists then of T1 as a voltage amplifier with T2 as an emitter follower output stage to drive the feedback resistor R_F. As in the shunt feedback circuit, the overall input resistance is r_1 of T1 in parallel with $R_F/(A_v+1)$ where A_v is the voltage gain to the emitter of T2. The transmittance defined as v_{E2}/i_i is

$$\frac{v_{E2}}{i_i} = \frac{R_F}{1+(R_F+r_1)/A_v r_1}$$

(cf. shunt feedback amplifier, Appendix 1).

Now, the emitter current of T2 is

$$i_{E2} = v_{E2}/R'_{E2}$$

where R'_{E2} is the effective resistance from the emitter of T2 to ground, i.e. R_E and R_F in parallel (very closely). Therefore, current feedback stabilizes the transmittance defined as i_{E2}/i_i (cf. series feedback which stabilizes i_E/v_i). The collector current of T2 is $\alpha_{N2} i_{E2}$, and since α_N is stable and closely unity for any transistor, the overall current gain is stable:

$$A_i = i_{C2}/i_i$$
$$= \left(\frac{R_F}{1+(R_F+r_1)/A_v r_1}\right) \cdot \left(\frac{R_F+R_E}{R_F R_E}\right) \cdot \alpha_{N2}$$

Therefore

$$A_i \simeq \frac{R_F+R_E}{R_E} \cdot \frac{\alpha_{N2}}{1+R_F/A_v r_1} \qquad .....(22)$$

i.e. $A_i \simeq R_F/R_E$ .....(23)

All terms in eqn (22) are known or may be evaluated approximately.

The voltage gain A_{v1} of T1 is approximately

$$A_{v1} = \frac{\alpha_N}{r_B/\beta_N + r_E} R_{L\,eff} \simeq R_{L\,eff}(\beta_{N1}/r_1)$$

where R_{Leff} is the effective load resistance made up of the parallel combination of R_C and the input resistance r_2 of T2:

$$r_2 \simeq \beta_{N2} R_E$$

Note that R_C includes the biasing potentiometer for T2, if any.

The voltage gain A_{v2} of T2 as an emitter follower is closely unity if $R_E \gg r_{E2}$.

Therefore $A_v = A_{v1} A_{v2}$

$$= \frac{\beta_{N1}}{r_1} \frac{R_C \beta_{N2} R_E}{R_C + \beta_{N2} R_E}$$

and substitution in eqn (22) gives

$$A_i = \frac{R_F+R_E}{R_E} \cdot \frac{\alpha_{N2}}{1+(R_F/R_E)\{(R_C+\beta_{N2} R_E)/\beta_{N1}\beta_{N2} R_C\}} \qquad .....(24)$$

For a given value of R_F/R_E, i.e. for a given value of expected current gain, eqn (23), the loop gain is maximum for

$$(R_C+\beta_{N2} R_E)/\beta_{N1}\beta_{N2} R_C$$

as small as possible. Therefore, R_C should be as large as possible, and desirably should be much larger than $\beta_{N2} R_E$ if this is consistent with $R_E \gg r_{E2}$. If the transistors have unequal values of β_N, the higher gain type should be used for T1.

The input resistance of a current feedback pair is

$$r_i \simeq R_F/A_v \qquad .....(25$$

6.2. *Voltage Feedback Pair*

The voltage feedback pair is analysed as a conventional voltage feedback amplifier, the steps being:

(i) Break the feedback loop by returning R_F to earth rather than R_E.

(ii) Calculate the open voltage gain A_o between input and output.

(iii) The closed loop voltage gain A_v is approximately

$$A_v \simeq \frac{R_F+R_E}{R_E} \cdot \frac{1}{1+R_F/A_o R_E} \quad \text{.....(26)}$$

i.e. $$A_v \simeq R_F/R_E \quad \text{.....(27)}$$

The voltage gain of T1 is approximately

$$A_1 = \frac{\alpha_{N1}}{(r_{B1}/\beta_{N1})+r_{E1}+R_E} \cdot \frac{R_C r_2}{R_C+r_2}$$

where r_2 is the input resistance of T2, approximately given by

$$r_2 = r_{B2}+\beta_{N2} r_{E2}$$

and where R_C includes the biasing potentiometer for T2, if any. The voltage gain of T2 is approximately

$$A_2 = \frac{\alpha_{N2}}{(r_{B2}/\beta_{N2})+r_{E2}} \cdot \frac{R_F R_L}{R_F+R_L}$$

Therefore, the overall open gain is

$$A_o = A_1 A_2$$

i.e.

$$A_o = \frac{\alpha_{N1}\alpha_{N2}\beta_{N2} R_C R_L}{R_C+r_{B2}+\beta_{N2} r_{E2}} \cdot \frac{1}{(r_{B1}/\beta_{N1})+r_{E1}+R_E} \cdot \frac{R_F}{R_F+R_L} \quad \text{.....(28)}$$

By differentiation, A_o is maximum and the loop gain is maximum for

$$R_E R_F = (r_{B1}/\beta_{N1}+r_{E1})R_L \quad \text{.....(29)}$$

Thus, in a voltage feedback pair, the ratio R_F/R_E is chosen to give approximately the required closed-loop gain (eqn (27)), the absolute values of R_F and R_E are then found from eqn (29), and finally the accurate closed loop gain is found from eqns (26) and (28). In general, R_C and R_L should be as large as possible, and if the transistors have unequal values of β_N, the higher gain type should be used for T2.

The design equations for the voltage feedback pair are, unfortunately, much more complicated than those for the other three basic structures. If a rather less accurate expression for the gain can be tolerated, it is possible to simplify the equations considerably. From eqn (28), neglecting a number of small corrections,

$$A_o \simeq \frac{\beta_{N2}}{r_{E1}+R_E} \cdot \frac{R_F R_L}{R_F+R_L}$$

and this expression is likely to be as much as 50% too large. Substitution in eqn (26) gives

$$A_v \simeq \frac{(R_F+R_E)/R_E}{1+(r_{E1}+R_E)(R_F+R_L)/\beta_{N2} R_E R_L} \quad \text{...(30)}$$

and $$R_E R_F \simeq r_{E1} R_L \quad \text{.....(31)}$$

for maximum loop gain. With typical values, A_v as given by eqn (30) will be in error by 10%.

The input resistance of a voltage feedback pair is given by

$$r_i = \beta_{N1}(r_{E1}+R_E)A_o R_E/R_F$$

i.e. $$r_i \simeq \beta_{N1}\beta_{N2} R_E R_L/(R_F+R_L) \quad \text{.....(32)}$$

7. Overall Amplifier Design

7.1. *Cascades of Single Stages*

From the point of view of obtaining the highest gain from a given number of cascaded single-stage feedback amplifiers with a given sensitivity to changes in the active elements, a cascade of alternate stages is by far the best and a cascade of shunt feedback stages is the worst. Table 2 lists the maximum gains available with ± 10% uncertainty.

Unless circumstances are such as to compel the use of similar stages (and it is difficult to conceive of such a set of circumstances) a cascade of alternate stages is the only cascade of single stage feedback amplifiers worth considering.

Table 2

Amplifier Type	Theoretical Limit	Value for Typical β_{NA}		
		$\beta_{NA}=30$	$\beta_{NA}=60$	$\beta_{NA}=100$
Two series stages	$(0{\cdot}1\ \beta_{NA})^2$	9	36	100
Two shunt stages	$(0{\cdot}1\ \beta_{NA}-1)^2$	4	25	81
Two alternate stages	$\beta_{NA}\times(0{\cdot}1\beta_{NA}-1)$	60	300	900
Feedback pair	$0{\cdot}1\ (\beta_{NA})^2$	90	360	1000

7.2. *Feedback Pairs and Triples*

In general, for a given overall gain and a given number of stages, the sensitivity of a multi-stage feedback amplifier to changes in the active elements is reduced as the number of stages within each feedback loop is increased. Thus, a cascade of feedback pairs has a lower sensitivity than a cascade of single stage feedback amplifiers, and a cascade of feedback triples is still less sensitive than a cascade of feedback pairs.

Experience has shown that feedback groups of more than two transistor stages are sufficiently undesignable as to more than offset their theoretical advantages of lower sensitivity. The prime trouble is a tendency towards high-frequency instability if the loop gain is other than small. A second, related source of trouble is the wide variation of open gain; a three-to-one variation over three stages is likely with the normal variation of β_N for transistors. If a feedback triple has sufficient feedback to counteract changes of β_N, it is most likely to become unstable if the transistors all have gains higher than "average". Thus, for normal purposes, a multi-stage amplifier is best designed as a cascade of feedback pairs; the outside broadcast amplifier described in Section 8 is an example.

However, the sensitivity of a cascade of alternate series and shunt stages is very nearly as small as that

of a cascade of feedback pairs. The added flexibility introduced into the design by the use of single stages sometimes offsets the slight rise in sensitivity. One example is the gramophone preamplifier described in Section 8; in this design, the feedback impedances of the individual stages are used to control the frequency response. Again, video amplifiers are best designed as cascades of single stages.

7.3. *Overall Feedback*

The aim in this paper has been to derive a set of simple expressions for the transmittance of various feedback amplifier types. The expressions produced are quite accurate—the predicted and measured gains of the outside broadcast amplifier (Section 8) differ by less than 0·5 dB in a total gain of more than 90 dB—but they are approximate only. It is therefore unreasonable to attempt to design an amplifier of very close gain accuracy requirement by substituting highly accurate data in the expressions derived in this paper.

A more reasonable approach is the use of overall feedback around an amplifier whose gain has been stabilized to (say) $\pm 5\%$ by the methods discussed in this paper. Quite a small gain around the overall loop will then result in a most precise overall gain. No trouble with instability should be encountered, as the frequency response of the forward path can be accurately controlled by the local feedback loops.

7.4. *Distortion*

This paper is primarily concerned with the gain of transistor feedback amplifiers. Space does not at present allow a detailed analysis of the distortion.

It was pointed out in the preliminary discussion of the single stage circuits that the distortion in the transmittance (as appropriately defined) is reduced by a factor equal to the return difference F; this applies also to the feedback pairs. With the mismatched interconnection, the individual feedback blocks operate under very closely the ideal conditions for which their transmittance is defined, so their distortion is reduced by the factor F; the order of magnitude of F is about 10 for the circuit arrangements considered in this paper. This distortion is about as low as can be achieved with simple circuits using the same number of transistors.

75. *Biasing System*

The biasing system of a multi-stage feedback amplifier should divide the amplifier into a number of single stages and d.c. feedback pairs, and this break-up of the amplifier into biasing blocks need not necessarily bear any relation to the break-up into signal feedback blocks. The division of the amplifier into biasing blocks should be so organized that biasing resistors are connected only to the bases of transistors which have a low input impedance. In this way, shunting of the signal by the biasing networks is reduced to a minimum; the effective load for shunt or voltage feedback amplifiers may be kept as large as possible.

Desirably, bias resistors should be connected only to the bases of transistors which are:

(a) single stage shunt feedback signal amplifiers.
(b) the first stage of a signal current feedback pair.
(c) the second stage of a signal voltage feedback pair.

There will be occasions when this is not possible; the input stages of the outside broadcast amplifier (Section 8), are an example.

8. Examples of Audio Feedback Amplifiers

To illustrate the principles developed in this paper, the designs are given for two audio feedback amplifiers. The first is an outside broadcast amplifier (Fig. 10); this example shows how accurate even the most approximate expressions for transmittance can be when applied to a well-designed circuit, and how simple it is to calculate the gain to a high degree of accuracy. The second is a gramophone preamplifier (Fig. 11), which shows how easily the principles developed for the design of amplifiers of flat frequency response may be adapted to the design of amplifiers of controlled frequency response.

8.1. *Outside Broadcast Amplifier*

8.1.1. General specifications

(i) gain: more than 90 dB
(ii) output: more than 1 W
(iii) distortion: less than 1% at 1 W output, 1 kc/s
(iv) power consumption: $\frac{1}{2}$ A at 12 V, i.e. 6 W
(v) two inputs with independent gain.controls, and one master gain control.

8.1.2. Input stage

The input stages T1 and T2 are single-stage series feedback amplifiers, driven from the voltage output of the input transformers; the input transformers should have 600 Ω secondaries. Emitter feedback biasing is used, and the quiescent conditions are

$$I_E = 0{\cdot}5 \text{ mA}, \qquad V_{CE} = 2{\cdot}5 \text{ V}.$$

The stages overload at about 100 mV input.

The approximate transconductance is, from eqn (5), $1/R_E = 1/330\,\Omega$ (neglecting 3·9 kΩ in parallel) i.e. $G_t \simeq 3$ mA/V. The accurate transconductance is given by eqn (10); substitution of the data

$$\alpha_N = 0{\cdot}98, \qquad \beta_N = 60, \qquad r_B = 50\ \Omega,$$
$$r_E = 50\ \Omega, \qquad R_E = 300\ \Omega,$$

gives

$$G_t = 2{\cdot}76 \text{ mA/V}$$

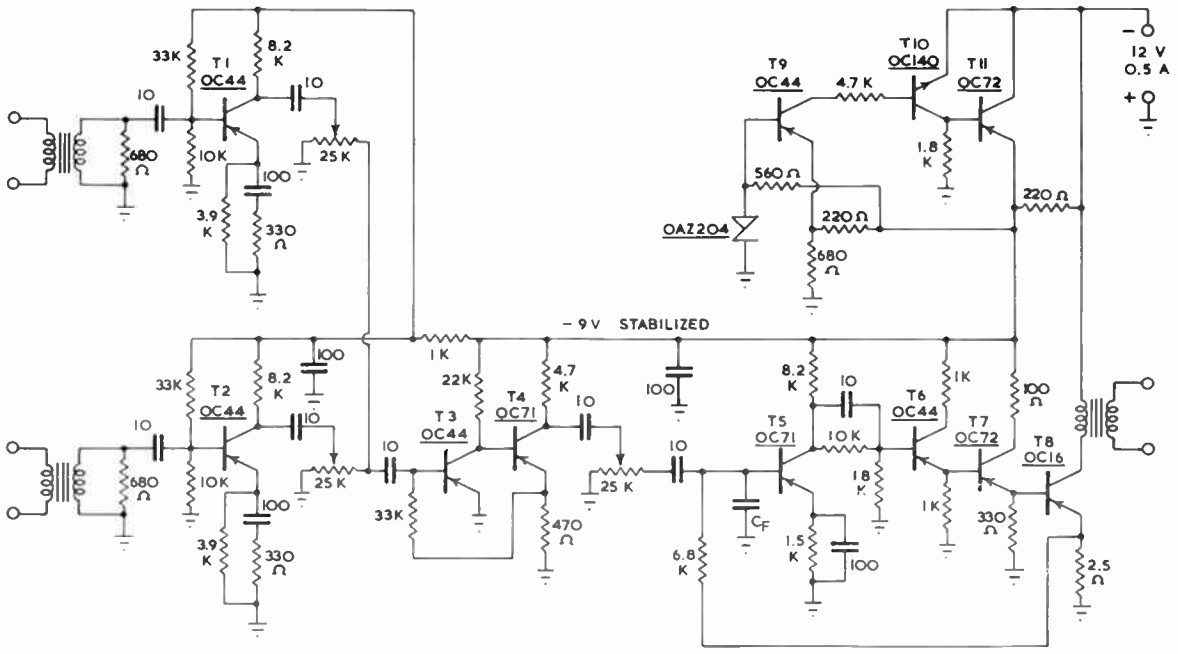


Fig. 10. Circuit diagram for a 90 dB outside broadcast amplifier.

T1 and T2 are chosen for low noise. The r.f. type OC44 appears to be quieter than the audio type OC71; either may be used without change of gain.

8.1.3. Second stage

The second pair of transistors, T3 and T4 are a current feedback pair. T3 is run under "starvation" conditions so that R_C may be large:

$$I_{E3} = 0{\cdot}4 \text{ mA}, \qquad V_{CE3} = 0{\cdot}7 \text{ V}$$

T4 is run at 1 mA to give a small value of r_{E4} so that a small R_E may be used for maximum feedback. D.c. feedback biasing is used over the pair; for bias stability, T3 should be a type with low I_{CO}.

The parameters are:

$$R_E = 470\ \Omega, \qquad R_F = 33\ \text{k}\Omega, \qquad R_C = 22\ \text{k}\Omega,$$
$$\alpha_{N2} = 0{\cdot}98, \qquad \beta_{N1} = 60, \qquad \beta_{N2} = 40.$$

The approximate gain is given by eqn (23) as 70·3, and eqn (24) gives the accurate gain as 66·3.

The pair overloads at rather more than 500 μA r.m.s. output. Since the output stage requires 100 μA to drive it to full output, the amplifier should not be operated with the master gain control set at more than 15 dB attenuation.

8.1.4. Output stage

The output stage is another current feedback pair; T5 is the first transistor and T6, T7 and T8 form a composite second transistor. D.c. feedback biasing is used; T8 runs at 0·4 A emitter current.

The output transformer should present a 20 Ω load to the OC16, and have a primary inductance of 0·1 H at 0·5 A d.c. The 1 kΩ and 100 Ω resistors in the collectors of T6 and T7 protect the transistors from over-driving if the load is accidentally removed. The 1 kΩ and 330 Ω resistors in the emitter circuits of T6 and T7 improve the high frequency performance of the composite transistor at high outputs.

The feedback phasing capacitor C_F is required to stabilize the output stage at high frequencies. With four transistors in the feedback loop, the phase shift is sufficient to cause oscillation. C_F was 0·1 μF in the prototype; it was selected for the best square wave response to the emitter of T8.

The approximate gain is given by eqn (23)

$$A_i \simeq 6800/2{\cdot}5 = 2700$$

The accurate gain may be calculated as 2400.

8.1.5. Power supply

T9, T10 and T11 form a regulated power supply which stabilizes the supply voltage to all parts of the amplifier except the collector of T8. There is no trace of low frequency instability when the amplifier is operated from a battery of 1 Ω internal resistance. The OC72 series transistor T11 is operated rather close to maximum ratings; it may be replaced by the larger type OC30 if the 220 Ω parallel resistor is removed.

8.1.6. Overall gain

The overall voltage gain is

$$G_{t1} \times A_{i2} \times A_{i3} \times R_{L3} \times \text{turns ratio}.$$

R_{L3} is 20 Ω, and the transformer turns ratio is 5·48 for a 600 Ω secondary.

The approximate gain is

$$3 \times 70{\cdot}3 \times 2700 \times 20 \times 5{\cdot}48 = 6{\cdot}23 \times 10^4 = 95{\cdot}8 \text{ dB}$$

The accurate gain is

$$2{\cdot}76 \times 66{\cdot}3 \times 2400 \times 20 \times 5{\cdot}48 = 4{\cdot}34 \times 10^4 = 93{\cdot}7 \text{ dB}$$

less 0·5 dB loss at the interstage couplings

$$= 93{\cdot}2 \text{ dB}$$

The approximate gain obtained by making no corrections at all is less than 3 dB from the theoretical gain in a total gain of more than 90 dB. The measured gain, using

5% resistors in the feedback circuits

10% resistors elsewhere

unselected transistors

was 92·8 dB, excluding transformer losses.

8.2. *Gramophone Preamplifier*

The gramophone preamplifier has sufficient gain to operate directly from the very low output of an Ortofon type C pick-up head (sensitivity 300 μV per cm s^{-1}) without a transformer. Three positions of equalization are provided which, in conjunction with the tone controls, enable a curve within 2 dB of any recording characteristic at all to be synthesized between 100 c/s and 10 kc/s. Input circuits for tape, with C.C.I.R. equalization, and radio are provided. There are switched base and treble controls with five positions each, and a sharp cut-off low-pass filter. All switches are wired so as to be entirely free from "clicks" in all positions of rotation. The noise level of the complete gramophone is more than 80 dB below 20 W. Critical listening tests do not reveal any distortion. The 3 dB bandwidth in the "radio" position with the tone controls "flat" is 1·8 c/s—480 kc/s.

The circuit given, Fig. 11, is not exactly that of the unit constructed. The original was designed to operate from the positive supply of the main amplifier.

8.2.1. Pick-up head-amplifier

The pick-up head-amplifier is a single stage series feedback amplifier, which requires no comment. Its transconductance is about 2 mA/V. The noise at the input is 0·4 μV.

8.2.2. Tape input stage

Tape playback heads should work into a high load resistance. Up till the present there have been two common methods† for obtaining a high input resistance transistor amplifier. Neither is particularly satisfactory. The approach adopted in this preamplifier is to con-

† J. H. Giddy, "A note on high input resistance single transistor stages", *Proc. Instn Radio Engrs, Aust.*, **22**, p. 174, March, 1961.

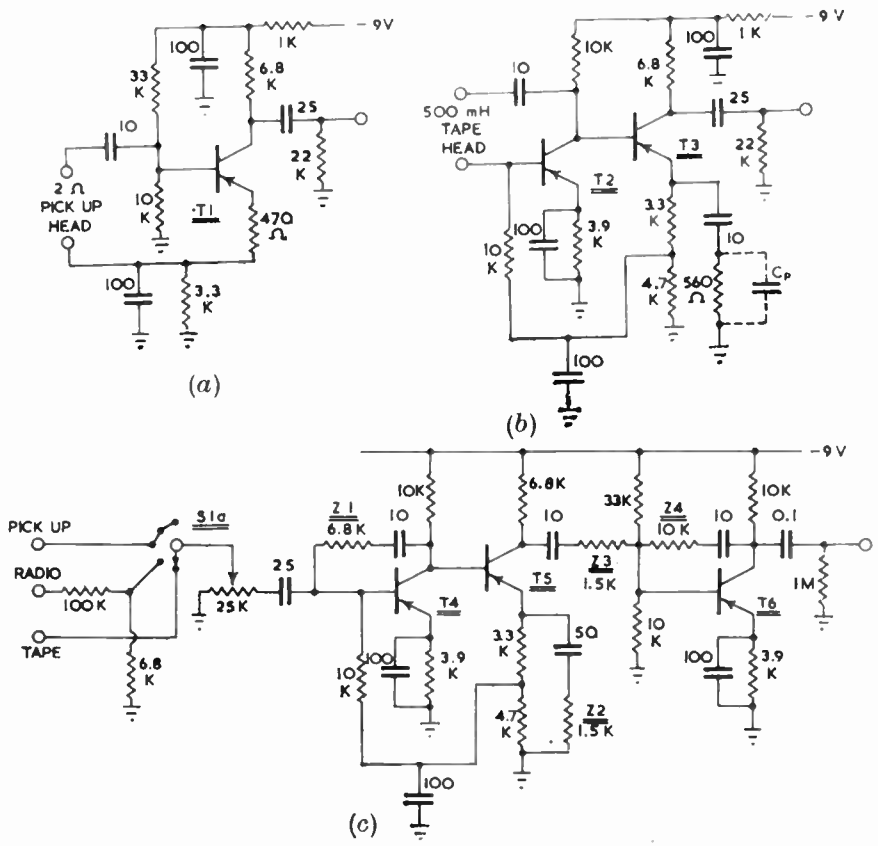


Fig. 11. Circuit diagram for a gramophone preamplifier.
(*a*) pick-up head amplifier
(*b*) tape head-amplifier
(*c*) main unit

All transistors in the prototype were type OC44. Any medium gain low noise transistors may be substituted. Details of Z1–4 are given in Fig. 12.

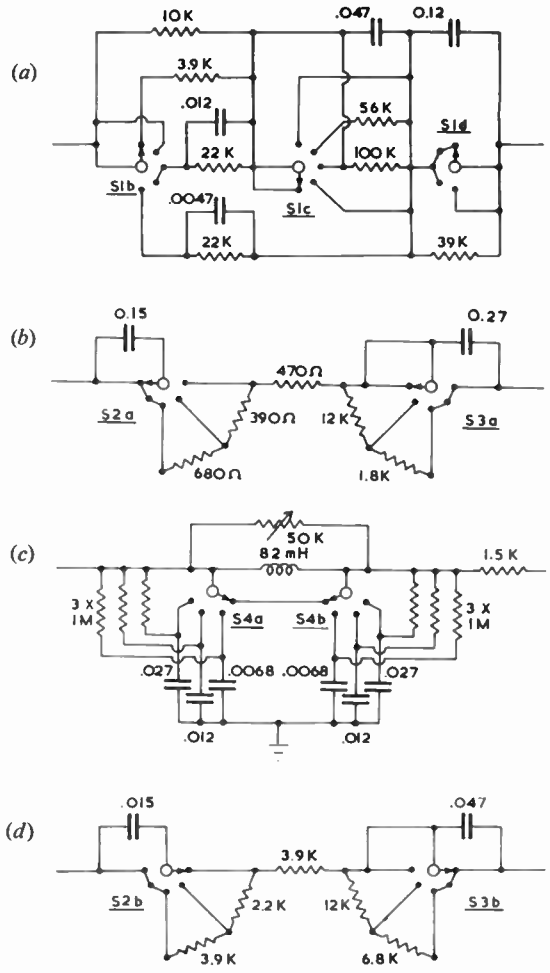


Fig. 12. Detail of switch wiring for gramophone preamplifier.
(*a*) compensating impedance Z_1
(*b*) bass cut and treble boost Z_2
(*c*) low-pass filter Z_3
(*d*) bass boost and treble cut Z_4

The switch positions are:

Input selector S_1 : (i) tape, (ii) radio, (iii) RIAA, (iv) EMI, (v) 78.

Treble S_2 : (i) − 8 dB, (ii) − 4 dB, (iii) flat, (iv) + 4 dB, (v) + 8 dB.

Bass S_3 : (i) − 8 dB, (ii) − 4 dB, (iii) flat, (iv) + 4 dB, (v) + 8 dB.

Filter S_4 : (i) flat, (ii) 10 kc/s, (iii) 7 kc/s, (iv) 5 kc/s

The slope is variable from 6 to 18 dB/octave.

All switches are shown at position (i).

nect the tape head between the collector and base of a common emitter stage, where the input impedance is very high. This circuit has a number of advantages:

(i) the impedance from either side of the input to ground is low, so that hum pick-up is minimized.

(ii) there is very little noise, as the feedback increases so as to compensate almost exactly for the $1/f$ noise of a transistor.

In the circuit diagram, T2 is the matching stage, and T3 is a series feedback amplifier whose transconductance is about 2 mA/V. C_P is a high frequency peaking capacitor which may be required to compensate for a fall in the response of the tape head. The noise at the input is 1·5 μV.

8.2.3. Equalizing stage

T4 is a shunt feedback stage which provides the equalization for the pick-up and tape recording characteristics. The feedback impedance is a switched combination of resistors and capacitors whose impedance/frequency law is the inverse of the recording characteristic. Figure 12 shows the detail of the switch wiring, and Fig. 13 shows the impedance effective at each position of the switch, with its frequency response asymptotes. The resistors controlling the bass roll-off points at +14 dB and more especially +20 dB and +28 dB are calculated for a finite loop gain, and for a finite emitter by-pass capacitor; if the loop gain had been large enough for eqn (11) to apply, their values would have been:

$$39\ \text{k}\Omega \rightarrow 27\ \text{k}\Omega$$
$$100\ \text{k}\Omega \rightarrow 56\ \text{k}\Omega$$

8.2.4. Tone compensation

T5 and T6 are a cascaded pair of series and shunt feedback stages. The emitter impedance of T5 is switched to provide bass cut and treble boost, while the feedback impedance of T6 is switched to provide bass boost and treble cut. A single-section π low-pass filter, terminated in its characteristic impedance 1500 Ω, is placed between the stages. Figure 12 shows the detail of the switch wiring.

9. Transistors at Video Frequencies †

The bandwidth of an amplifier is limited by the stray capacitances which shunt the input and output of the active devices. The input capacitance of a transistor is very much larger than that of a valve, but the transconductance is correspondingly large so the gain × bandwidth product is of the same order of magnitude.

† Since the original printing of this paper, a new paper has been prepared, covering transistor video amplifiers fully. It is shown that the mismatched design technique of alternate series and shunt feedback stages gives results superior to any other technique known at present. A further paper describes a 120-Mc/s transistor oscilloscope.

E. M. Cherry and D. E. Hooper, "The design of wideband, transistor feedback amplifiers", *Proc. Instn Elect. Engrs*, **110**, p. 375, February 1963.

G. A. Rigby, "Wideband transistor amplifiers for oscilloscopes", *Proc. Instn Radio Engrs, Aust.*, **24**, 1963 (*To be published*).

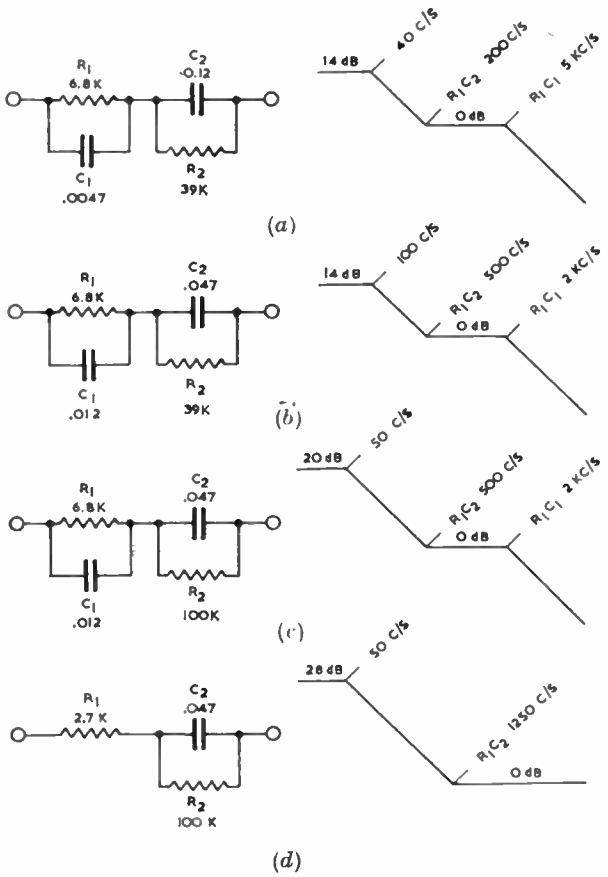


Fig. 13. Effective compensating impedance Z_1:
(*a*) nearly Decca f.f.r.r. (*c*) R.I.A.A.
(*b*) E.M.I.-Columbia (*d*) C.C.I.R. tape (7½ in/s)
In the radio position (not shown) the impedance is 10 kΩ.

Table 3 compares the relevant parameters of a typical valve and transistor.

Since the gain × bandwidth product for a valve and transistor are of the same order of magnitude, it might be expected that the design of a satisfactory transistor video amplifier could be based on the design of a valve amplifier, with all impedances scaled down by about 100. This is not the case, because the input capacitance of a transistor is inaccessible. The ohmic base resistance r_B isolates the input capacitance from the base terminal of the transistor, and r_B is of the same order of magnitude as the load resistance required by a transistor video amplifier. It is therefore impossible to place the peaking chokes of a transistor video amplifier at the points at which they are required.

Table 3

Device	Operating current mA	g_m mA/V	C pF	Gain × Bandwidth $= g_m/2\pi C$ Mc/s
6AK5	$I_K = 10$	5	10	80
OC170	$I_E = 10$	400	600	100

The most satisfactory approach to the design of transistor video amplifiers appears to be the use of negative feedback. Broadly, negative feedback reduces the effective input capacitance; the detailed operation is discussed in the following sections. In addition to increasing the bandwidth of a video amplifier, feedback stabilizes its gain.

Single-stage feedback circuits appear preferable to two-stage circuits. The phase shift in a single transistor can exceed 90 deg at high frequencies, so, in principle, a two-stage feedback circuit can become unstable. Two-stage circuits, even without peaking, appear prone to ringing or overshoot on transients.

9.1. *Operating Point*

The operating point of the transistors in the audio amplifiers discussed in the previous sections of this paper was of the order of

$$V_{CE} = 3 \text{ V}, I_E = 1 \text{ mA}$$

This operating point was selected on the basis of power supply economy and not for maximum gain; slightly more gain could have been obtained by running the transistors at higher currents from a higher supply voltage.

Video amplifiers with bandwidths of the order of megacycles will use drift transistors rather than the diffusion types assumed for audio applications. Drift transistors must be operated at high voltages and currents if the high frequency performance is to be anywhere near the maximum attainable. There are two effects to be considered.

9.1.1. Emitter-base circuit time-constant

In the h.f. equivalent circuit (Fig. 14), there are two components of input capacitance

C_T is the junction transition capacitance

C_B is the base charging capacitance, also known as the junction diffusion capacitance.

The operation of a *p-n* junction is such that:

C_T is approximately constant over the voltage range encountered at an emitter junction.

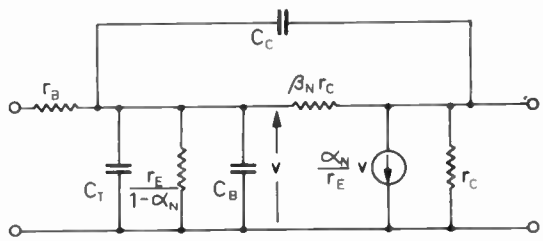


Fig. 14. The h.f. equivalent circuit for a transistor.

C_B is directly proportional to emitter current:

$$C_B = (Kq/kT)b^2 I_E$$

where K is a constant, depending on the impurity concentration in the base region.

r_E is inversely proportional to emitter current:

$$r_E = (kT/q)(1/I_E)$$

The overall junction time constant is

$$\begin{aligned}\tau &= r_E(C_T + C_B)\\ &= (kT/qI_E)\{C_T + (Kq/kT)b^2 I_E\}\\ &= kTC_T/qI_E + Kb^2\end{aligned}$$

and τ should be least for the best high frequency performance. Therefore, the emitter current should be as large as possible. In the interests of economy, I_E is normally made only so large that Kb^2 is the dominating term in the expression for τ. Diffusion transistors have a small transition capacitance, and $I_E = 250$ μA is adequate. Drift transistors have much larger transition capacitances at the emitter junction, so I_E should be larger. The optimum emitter current is specified in the manufacturer's data for a particular transistor type; 5 mA is a typical value.

9.1.2. Transit time

The cut-off frequency of a transistor depends on the transit time of carriers across the base region; the transit time should be as short as possible. Now the transit time is approximately proportional to the square of the base width; therefore, the base should be as narrow as possible.

When the collector voltage of a transistor is increased, the collector depletion layer widens and the residual base width is reduced. The resistivity of the base wafer of a diffusion transistor is quite low, so the change in depletion layer width (and hence base width) for a reasonable change in collector voltage is quite small. However, the base wafer near the collector junction of a drift transistor has a very high resistivity; the first few volts applied to the collector of a drift transistor therefore reduce the base width considerably. Data sheets on transistors specify the minimum desirable collector voltage; 5 V is adequate for most types, but some require as much as 20 V.

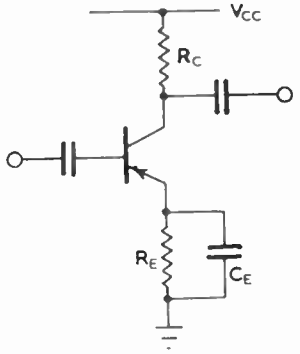


Fig. 15. Series feedback stage with peaking.

9.2. *Series Feedback Circuit*

The mathematical analysis† of the series feedback amplifier with peaking (Fig. 15) is beyond the scope of a paper such as this. The physical principles of peaking circuits are outlined in the following sections; the selection of the optimum value of the peaking capacitor in any particular case is a matter of systematic trial and error. The operation of the series feedback peaking circuit is as follows:

(1) The intrinsic transistor (i.e., without r_B) has a mutual conductance $g_m = \alpha_N/r_E$, with cut-off frequency well above f_1:

$$2\pi f_1 = \omega_1 = 1/r_E C_B$$

f_1 for drift transistors is of the order of 100 Mc/s—well above any operating frequency considered in this paper. The intrinsic transistor also has an input resistance $r_E/(1-\alpha_N)$ and an input capacitance $C_B + C_T$.

(2) The base resistance r_B forms a potential divider with the input impedance of the intrinsic transistor. Since the input impedance has a capacitive component, high frequency signals at the internal base are attenuated more than low frequency signals. The cut-off frequency of the transconductance of a practical transistor is therefore much less than f_1.

(3) If R_E is added, the input impedance of the internal transistor is increased by the feedback; in particular, the input capacitance is reduced. The cut-off frequency of the input potential divider formed by z_i and r_B is raised, so the cut-off frequency of the effective transconductance of the transistor with feedback is raised to more nearly f_1. The low frequency transconductance is reduced.

(4) If a small capacitor is connected in parallel with R_E, two effects occur:

(a) the effective transconductance of the intrinsic transistor with feedback is raised at high frequencies since $|Z_E|$ falls:

$$G_{t\,\mathrm{eff}} = \alpha_N/(r_E + Z_E)$$

(b) the input capacitance of the intrinsic transistor is increased, so the cut-off frequency of the input divider is lowered.

These two effects tend to counteract each other.

A very approximate physical analysis shows that the transient response of a series feedback stage with peaking is optimum, i.e. critically damped, when

$$R_E C_E = r_E(C_B + C_T) \simeq 1/\omega_1 \qquad \ldots\ldots(33)$$

† An attempt at an analysis, with which the author does not entirely agree, is: G. Bruun, "Common emitter transistor video amplifiers", *Proc. Inst. Radio Engrs*, **44**, p. 1561, 1956.

The approximations are such that C_E as given by this expression is rather too small.

9.3. *Shunt Feedback Circuit*

Thomas† gives a circuit (Fig. 16) for peaking shunt feedback video amplifiers. The transmittance of a shunt feedback amplifier is (approximately) Z_F. If an inductance is connected in series with R_F, $|Z_F|$ increases at high frequencies. The rise in transmittance due to this effect may be used to offset the normal high frequency droop.

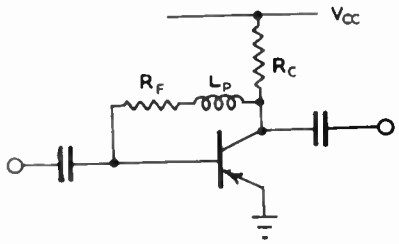


Fig. 16. Shunt feedback stage with peaking.

Experience has shown that this form of peaking is unsatisfactory; as the peaking inductance is increased from zero, the transient response develops an overshoot and then rings without any preliminary stage of reduced rise time. A complete analysis has not been performed, but it appears probable that the phase shift in a transistor can reach 180 deg at high frequencies. Thomas's explanation is therefore at fault; the frequency response of a system with feedback lies below the response without feedback at all frequencies only if the response is due to a single time constant. The frequency response is therefore as shown in Fig. 17; a shunt feedback stage is self-peaking. Further evidence that the above explanation is correct is that a shunt feedback stage without peaking overshoots if the load is made capacitive.

9.4. *Cascaded Stages*

As with audio amplifiers, the only cascade of single stages worth considering is a cascade of alternate series and shunt stages. Each pair of stages should be regarded as a unit, and the emitter peaking capacitor of the series feedback stage is adjusted for the best response over the pair.

In general, video amplifiers are required to have a good low-frequency transient response, that is, the top of a low-frequency square wave should be free from tilt. It is desirable, therefore, to remove as many low-frequency time-constants as possible from a video amplifier, and to add compensating elements to correct the distortion produced by those low-frequency time-constants which cannot be removed. Direct coupling of stages is used where possible, and d.c. feedback biasing is used so that low-frequency emitter by-pass capacitors are not required for at least some of the transistors. Low frequency compensation is best effected by a capacitor in series with the feedback resistor of a shunt feedback stage.

† An excellent reference on the application of transistors to high-frequency circuits is: D. E. Thomas, "Some design considerations for high-frequency transistor amplifiers", *Bell Syst. Tech. J.*, **38**, p. 1551, 1959.

Thomas's paper is open to criticism in a few points, but his general engineering design philosophy is thoroughly to be recommended.

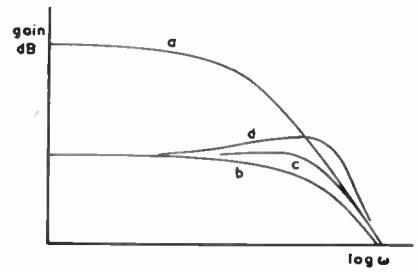


Fig. 17. Frequency response of shunt feedback stage:
(*a*) without feedback
(*b*) Thomas's curve for feedback
(*c*) Thomas's curve for peaking
(*d*) probable true curve for feedback (exaggerated).

9.5. *Example*

Figure 18 is the circuit diagram for a 40 dB voltage amplifier with a bandwidth in excess of 10 c/s-20 Mc/s. It illustrates most of the principles outlined in the preceding sections.

The operating point of both transistors is approximately

$$V_{CE} = 5 \text{ V}; \qquad I_E = 5 \text{ mA}.$$

The two stages are direct coupled, so eliminating one low-frequency time-constant; d.c. feedback biasing is used so that no low-frequency emitter by-pass capacitor is required for the first transistor. The low-frequency response is probably adequate without compensation, but a compensating capacitor could be connected in series with the 4·7 kΩ feedback resistor of the shunt feedback stage.

The shunt feedback stage has no high-frequency peaking. The series feedback stage has a peaking capacitor across its emitter resistor, this capacitor being selected for the best overall transient response of the amplifier. Its value is found experimentally to be about twice that suggested by eqn (33). This is quite a satisfactory agreement considering the approximations used in deriving eqn (33), and also that the capacitor is selected for the best performance over two stages.

10. Conclusions

Simple expressions for the transmittance of transistor feedback amplifiers have been derived in terms of the physical mode of operation of these circuits, and their

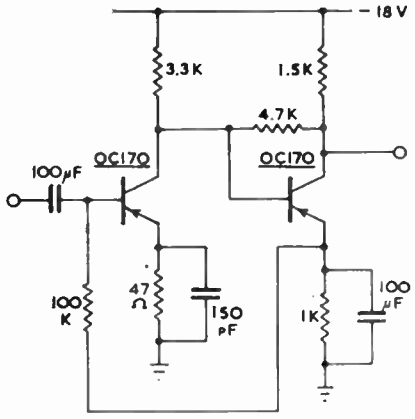


Fig. 18. Circuit diagram of a 40 dB video amplifier, 10 c/s to 20 Mc/s.

accuracy has been demonstrated in the examples. These expressions require no quantitive transistor data at all; a simple statement of the general type of the transistor is sufficient.

It is concluded from the accuracy of this physical approach that there is no requirement for an abstract mathematical design procedure in terms of circuit parameters. It is further suggested that such an approach is undesirable, as it requires more precise transistor data, gives little indication of which approximations are justified, and therefore involves much more labour.

As a general conclusion, it appears that transistor circuitry with a physical design procedure is more designable than valve circuitry can ever be, because all transistors have substantially the same properties.

11. Acknowledgments

This paper is based on work carried out in the Department of Electrical Engineering, University of Melbourne. This work was made possible by a Research Fellowship awarded by the Imperial Chemical Industries of Australia and New Zealand, Ltd., which the author gratefully acknowledges.

In addition, the author wishes to thank his numerous colleagues for many hours of helpful discussion and for their constant encouragement, making particular mention of Mr. D. E. Hooper. Finally, the author acknowledges the help of Messrs. L. J. Harding and E. M. Hooper in the development of the o.b. amplifier.

12. Appendix 1. Shunt feedback amplifier

In Fig. 19

$$i_i = v_i\{(1/r_1)+(1/R_F)\}+v_o(-1/R_F)$$

$$v_o = A_v v_i$$

Therefore

$$i_i = (v_o/A_v)\{(1/r_1)+(1/R_F)\}+v_o(-1/R_F)$$

Thus

$$(v_o/i_i) = [(1/A_v)\{(1/r_1)+(1/R_F)\}-(1/R_F)]^{-1}$$

i.e. $$(-v_o/i_i) = R_F/\{1-(r_1+R_F)/A_v r_1\}$$

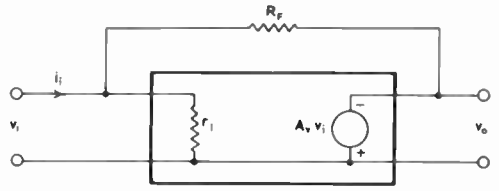


Fig. 19. Equivalent circuit for a shunt feedback amplifier.

13. Appendix 2. Manufacturers' data

If a design procedure similar to that outlined in the preceding pages should become widely adopted, much of the data at present supplied by transistor manufacturers will be of little use. The following data should be adequate:

General description

(1) *n-p-n* or *p-n-p*.

(2) Structure type, i.e. grown, alloy junction, surface barrier, etc.

(3) High, medium, or low gain.

(4) Audio, r.f. or drift.

(5) Maximum collector dissipation for normal mounting and ambient temperature.

Thermal data

(1) Maximum junction temperature.

(2) Thermal resistance from junction to case.

(3) Maximum I_{CO} at a particular temperature.

Low frequency parameters

(1) Maximum and minimum β_N.

(2) Average and maximum r_B; this might be more appropriately considered under "high frequency parameters", but would be useful for the most precise low frequency calculations.

(3) Collector–base breakdown voltage.

High frequency parameters

(1) Average and minimum $\omega_1 = 1/r_E C_B$.

(2) Collector transition capacitance at a particular voltage. $C_C \propto V_C^{-n}$ where n varies from 0·3 to 0·5 depending on the junction type.

(3) Emitter transition capacitance.

(4) Minimum desirable collector voltage for drift transistors.

First published in the Proceedings of the Institution of Radio Engineers Australia, 22, No. 5, May 1961. Manuscript of the new version received by the British Institution of Radio Engineers on 5th November 1962 (Paper No. 789).

U.D.C. 621.373.521.114

A New Analysis of the Transistor Phase-Shift Oscillator

By

Lt. WILLIAM M. LOCKE, U.S.N., M.S., †

Lt. VIRGIL W. MOORE, Jr., U.S.N., M.S. †

AND

WILLIAM W. HAPP, Ph.D. ‡

Summary: The relationship between transistor parameters and circuit characteristics of an RC phase-shift oscillator is analysed by flow-graph and root-locus techniques. Exact formulae and approximations for the starting frequencies of three- and four-section RC phase-shift oscillators are derived. Prediction of the operating frequency for a given current gain is presented graphically and compared with experimental results.

1. Introduction

In the course of an investigation of radiation effects on a transistor phase-shift oscillator, it was necessary to ascertain the function of the transistor in the oscillator operation.[1] It is known that the transistor is more sensitive to gamma radiation than resistors or capacitors[2] and that components other than the transistor can be selected which will be essentially unaffected by radiation levels which exceed transistor destruction levels. Since data are available on the effects of gamma radiation on many transistors, it is possible to predict parameter variation.[3] However, frequency variation of the phase-shift oscillator can be determined only if the relationship of transistor parameters to the rest of the circuit is known.

A literature search indicated that transistor phase-shift oscillators[4-8] have not been as thoroughly analysed as vacuum tube phase-shift oscillators.[9-13] It is the objective of this investigation to examine the operation of a phase-shift oscillator with variation in the transistor parameters. The methods developed in this investigation should be useful for a wide range of problems involving transistors with variable parameters.

2. Theory and Procedure

A phase-shift oscillator is essentially a single-stage amplifier with positive feedback through a passive network as illustrated in Fig. 1. At some frequency this passive circuit produces a phase-shift of 180 deg and oscillation may occur. This passive network consisting of three or four sections or steps with capacitors as the series elements and resistors as the shunt elements or vice versa. The principal advantage of RC circuits is their smaller physical size and higher Q in the audio frequency range than inductors.

The passive network most suitable for tubes for an RC phase-shift oscillator is the generalized network shown in Fig. 2,[13] since a tube is equivalent to an ideal

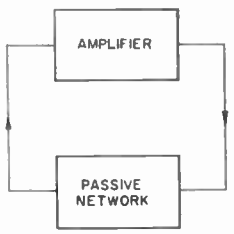


Fig. 1. Block diagram of a phase-shift oscillator.

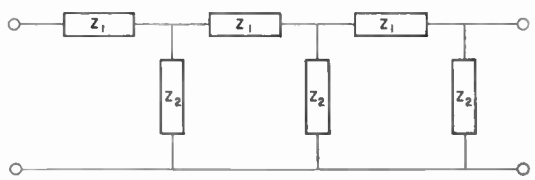


Fig. 2. Generalized phase-shifting network for voltage amplifiers.

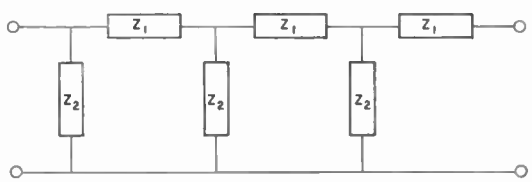


Fig. 3. Generalized phase-shifting network for current amplifiers.

voltage amplifier with low output and high input impedance. The transistor in the grounded emitter configuration approaches an ideal current amplifier with high output and low input impedance. Duality, therefore, dictates that the passive network should be of the

† U.S. Naval Postgraduate School, Monterey, California.

‡ Research Professor of Engineering, Arizona State University Tempe, Arizona, U.S.A.

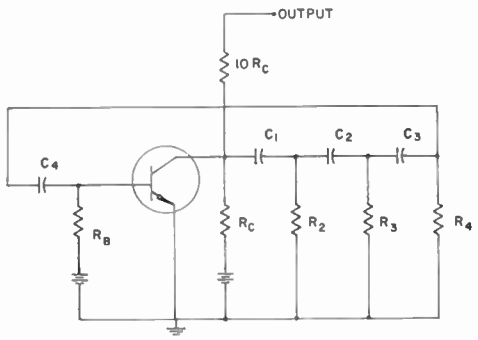


Fig. 4. Four-section transistor RC phase-shift oscillator.

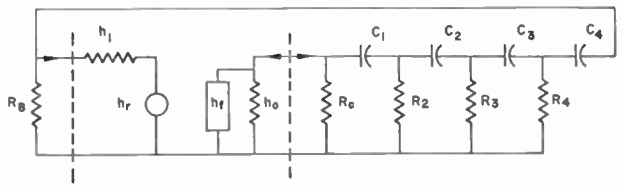


Fig. 5. Four-section transistor phase-shift oscillator.

form shown in Fig. 3.[7] If the wrong network is inadvertently used, either no oscillations result or excessively high gain is required to produce oscillations.

For convenience of design in the circuit shown in Fig. 4, the shunt elements were chosen as resistors so that the normal load resistor of the amplifier stage forms part of the passive network. By neglecting the non-linearities of the transistor except as a limit on the amplitude, a small-signal sinusoidal equivalent may be drawn as in Fig. 5 which has the following properties:[14, 15]

Natural oscillations can be set up in a circuit by the introduction of a random impulse.

The natural oscillations are observed in all locations of the circuit simultaneously.

Random pulse to start oscillation can originate at any point in the circuit.

The natural oscillations of a linear system are independent of the initial energy storage in the reactive elements.

The natural oscillations of a linear system are determined by the topology equation (or the characteristic equation) of the system.[16]

The characteristic polynomial of the linear incremental model of the transistor phase-shift oscillator can be found by systematic circuit analysis, i.e. mesh, nodal, or matrix methods. In this paper, flow-graph analysis will be used since it offers the following advantages:[17–21]

Quantities can be rearranged without loss of their interrelationships.

Assumptions can be readily justified.

The effect of neglecting any parameter can be easily observed.

The form of the topology equation is convenient for root locus plotting.[18, 22]

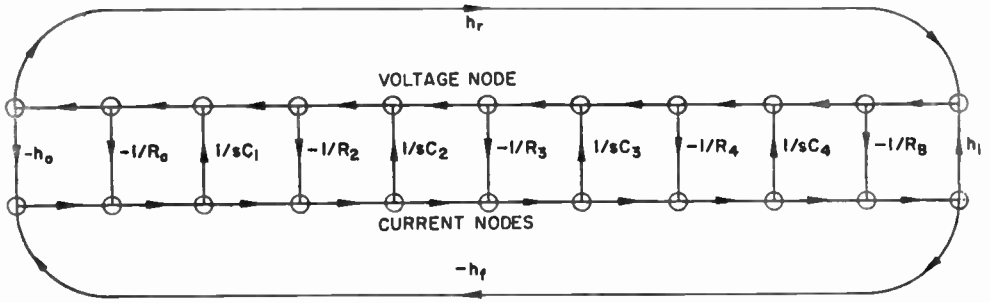


Fig. 6. Flow graph of transistor RC phase-shift oscillator.

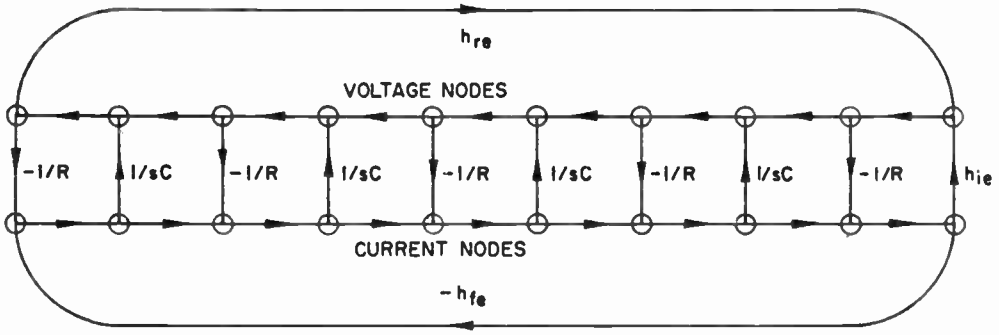


Fig. 7. Simplified flow-graph.

Table 1

Typical Circuit Parameter Values

h_{ie}	h_{re}	h_{fe}	h_{oe}	R_c	R_B	C	R	h_{ie}/R	h_{ie}/R_B
2·0 kΩ	$8{\cdot}0\times10^{-4}$	35	20 $\mu\Omega^{-1}$	6·8 kΩ	330 kΩ	0·010 μF	6·0 kΩ	0·33	$1{\cdot}8\times10^{-2}$

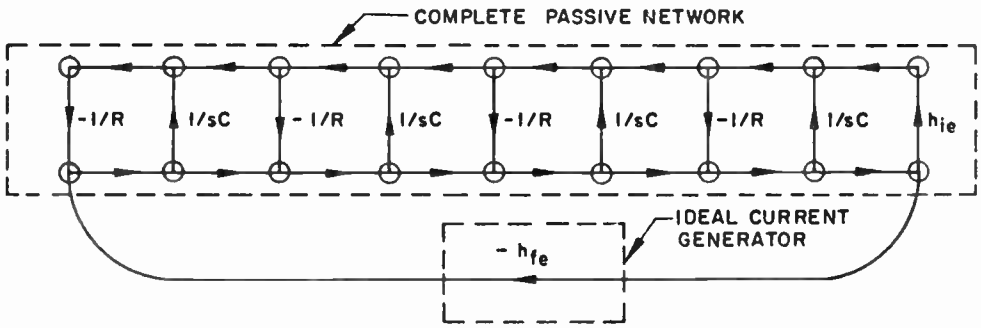


Fig. 8. Approximate flow-graph.

3. The Topology Equation of the Oscillator

The flow graph shown in Fig. 6 was constructed from the linear incremental model in Fig. 5. Since it was anticipated that the change in current gain h_{fe} would have the greatest effect on the oscillator characteristics, the flow graph was arranged so that the transistor current gain formed an exterior loop.

The simplifications shown in Fig. 7 were made to focus interest on the variation of transistor parameters, not on the variation of the other parameters of the other components, and were as follows:

$$C_1 = C_2 = C_3 = C_4 = C \qquad \text{.....(1)}$$

$$\frac{1}{R_1} = \frac{1}{R_2} = \frac{1}{R_3} = \frac{1}{R_4} = \frac{1}{R} \qquad \text{.....(2)}$$

where by design

$$\frac{1}{R_1} = \left(h_o + \frac{1}{R_c}\right) \qquad \text{.....(3)}$$

Application of the topology equation to the circuit of Fig. 7 gives the characteristic equation as:

$$\left(1 + \frac{h_{ie}}{R_B}\right)F_4 + \frac{h_{ie}}{R}G_4 + h_{fe}(1-h_{re}) - h_{re} = 0 \qquad \text{...(4)}$$

where F_4 and G_4 are functions of sCR listed in Table 2.

Typical circuit parameter values are listed in Table 1. The transistor used was a Texas Instruments type 904.

Since

$$h_{ie} \ll R_B \text{ and } h_{re} \ll 1 \qquad \text{.....(5)}$$

Equation (4) can be reduced to

$$F_4 + \left(\frac{h_{ie}}{R}\right)G_4 + h_{fe} = 0 \qquad \text{.....(6)}$$

and may be represented by the flow graph shown in Fig. 8. The dotted partitions emphasize the correspondence with Fig. 1.

4. Derivation of the Generalized Equation for *n* Sections

By changing the number of sections of the transistor RC phase-shift oscillator, the flow graph corresponding to Fig. 7 yields a topology equation of the type

$$F_n + \left(\frac{h_{ie}}{R}\right)G_n + h_{fe} = 0 \qquad \text{.....(7)}$$

where F_n and G_n are functions of sCR as shown in Table 2. F_n and G_n correspond to n sections or steps in the ladder network and obey the following rules:

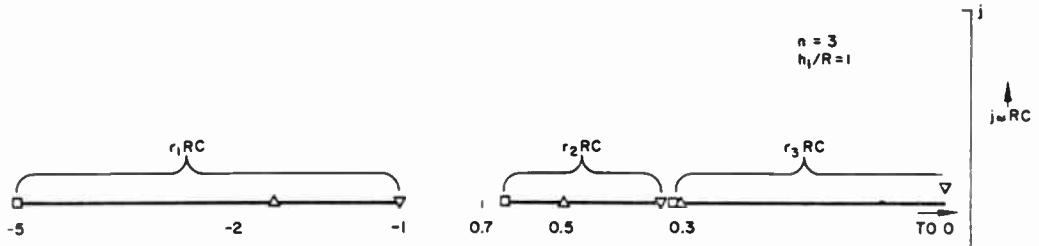
$$F_{n-1} + G_{n-1} = G_n$$

$$F_{n-1} + (sCR)^{-1}G_n = F_n$$

The binomial coefficients in Table 2 may be obtained from these recurrence formulas, thus, the topology equation for an n-section flow graph similar to Fig. 8 becomes

$$\sum_{k=0}^{n}\left[\binom{n+k}{n-k} + \binom{n+k}{n-k-1}\left(\frac{h_{ie}}{R}\right) + h_{fe}(sCR)^k\right] = 0 \qquad \text{.....(8)}$$

Note that only when n is greater than two does a phase-shift oscillator result. For one- and two-section networks there is no frequency for which the phase shift is 180 deg. For n greater than six there are multiple frequencies at which this criterion is met. However, the gain in all except one frequency is so large as to preclude, for practical purposes, the possibility of multiple oscillations.



h_{ie}/R =	LARGE	SMALL	rCR
SYMBOL	□	▽	△
r_1RC =	5.0	1.0	1.7
r_2RC =	0.64	0.33	0.50
r_3RC =	0.31	0	0.30

Fig. 9. Range of roots of a three-section passive network of a phase-shift oscillator.

5. Root Locus of the Circuit

The location of the roots of the topology equation in the s plane are indicative of the time response of a given system. Since the roots of the topology equation for the phase-shift oscillator will be affected by variations of transistor parameters, it is instructive to plot the root loci topology equations generated by this network.

The topology equation for any active network such as Fig. 4 with n sections was stated in equation (7)

$$F_n + \left(\frac{h_{ie}}{R}\right) G_n + h_{fe} = 0$$

which, when rearranged, yields

$$\frac{\dfrac{h_{fe}}{F_n}}{1 + \dfrac{h_{ie} G_n}{R F_n}} = -1 \qquad \text{.....(9)}$$

It is always possible to expand F_n and G_n in terms of roots f_n and g_n of these polynomials:

$$s^n F_n = \left(s - \frac{f_1}{CR}\right)\left(s - \frac{f_2}{CR}\right) \cdots \left(s - \frac{f_n}{CR}\right)$$

$$s^n G_n = ns\left(s - \frac{g_1}{CR}\right)\left(s - \frac{g_2}{CR}\right) \cdots \left(s - \frac{g_{n-1}}{CR}\right) \qquad \text{.....(10)}$$

Table 2

General Functions for Phase-Shift Oscillator with n Sections

n	$F_n = F_n(sCR)$	$G_n = G_n(sCR)$
1	$1 + \dfrac{1}{sCR}$	1
2	$1 + \dfrac{3}{sCR} + \dfrac{1}{(sCR)^2}$	$2 + \dfrac{1}{sCR}$
3	$1 + \dfrac{6}{sCR} + \dfrac{5}{(sCR)^2} + \dfrac{1}{(sCR)^3}$	$3 + \dfrac{4}{sCR} + \dfrac{1}{(sCR)^2}$
4	$1 + \dfrac{10}{sCR} + \dfrac{15}{(sCR)^2} + \dfrac{7}{(sCR)^3} + \dfrac{1}{(sCR)^4}$	$4 + \dfrac{10}{sCR} + \dfrac{6}{(sCR)^2} + \dfrac{1}{(sCR)^3}$
5	$1 + \dfrac{15}{sCR} + \dfrac{35}{(sCR)^2} + \dfrac{28}{(sCR)^3} + \dfrac{9}{(sCR)^4} + \dfrac{1}{(sCR)^5}$	$5 + \dfrac{20}{sCR} + \dfrac{21}{(sCR)^2} + \dfrac{8}{(sCR)^3} + \dfrac{1}{(sCR)^4}$
n	$\displaystyle\sum_{k=0}^{n} \binom{n+k}{n-k}\left(\frac{1}{sCR}\right)^k$	$\displaystyle\sum_{k=0}^{n=1} \binom{n+k}{n-k-1}\left(\frac{1}{sCR}\right)^k$

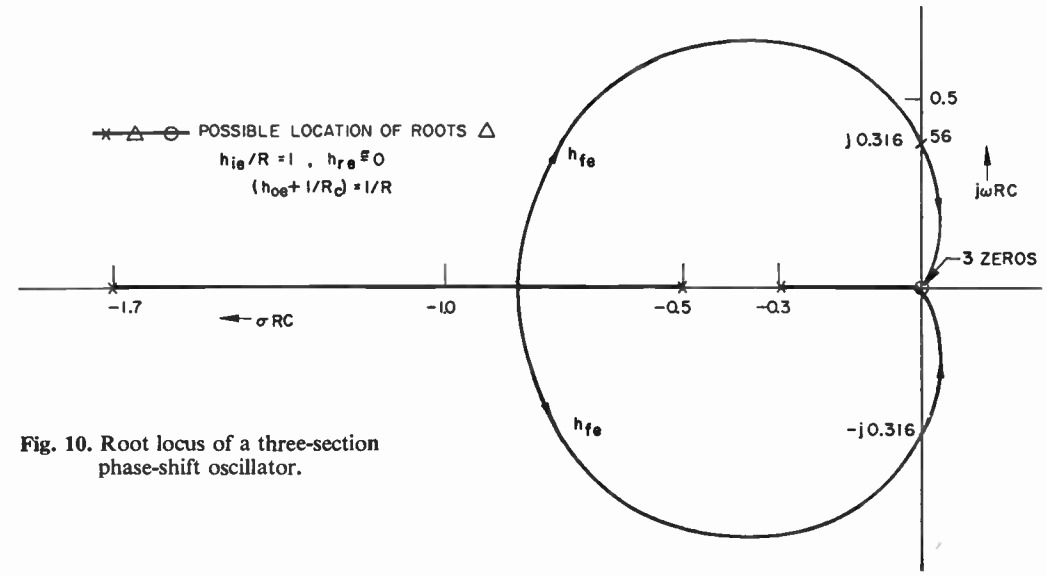
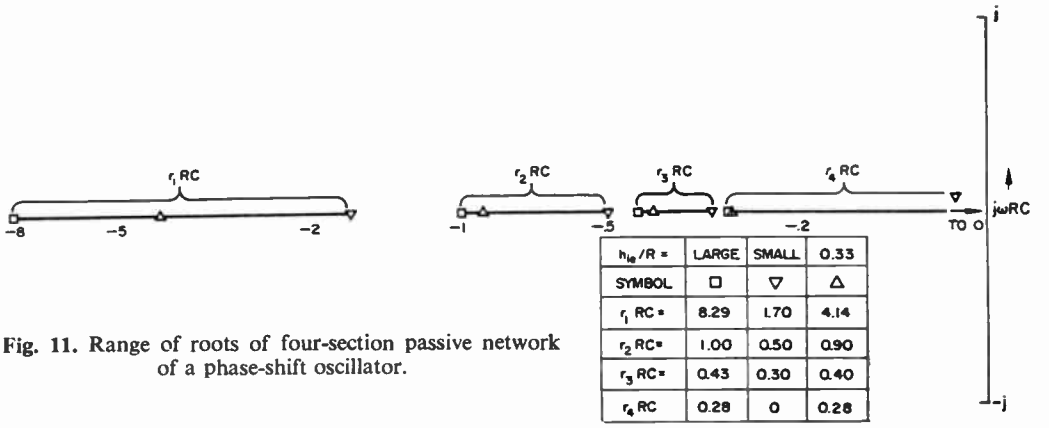


Fig. 10. Root locus of a three-section phase-shift oscillator.



h_{ie}/R =	LARGE	SMALL	0.33
SYMBOL	□	▽	△
r_1 RC =	8.29	1.70	4.14
r_2 RC=	1.00	0.50	0.90
r_3 RC=	0.43	0.30	0.40
r_4 RC	0.28	0	0.28

Fig. 11. Range of roots of four-section passive network of a phase-shift oscillator.

The root locus for the passive network in the form used in Evans' method[8, 22] is then obtained from:

$$\frac{h_{ie}G_n}{RF_n} = \frac{n(h_{ie}/R)s(s+g_1/CR)\ldots}{(s+f_1/CR)(s+f_2/CR)\ldots} = -1 = \exp(\pm jm\pi) \qquad \ldots\ldots(11)$$

where

$$m = 1,\ 3,\ 5\ldots$$

The roots ($r_i = f_i/CR$) of this equation for a three-section oscillator with a typical value (unity) of h_{ie}/R are plotted in Fig. 9. Note that the root locus for a passive network of only resistance and capacitance can be only on the negative real axis. Limits of r_iRC for very large and very small values of h_{ie}/R are shown in Fig. 9. Having thus determined the roots for the passive network by assuming reasonable values of h_{ie}/R, these roots now become the poles of the active network.

$$\frac{ps^n}{(s+r_1)(s+r_2)\ldots(s+r_n)} = -1 \qquad \ldots\ldots(12)$$

where

$$p = h_{fe}/F_n$$

The root locus of equation (12) may now be plotted using h_{fe} as a parameter for a three-section oscillator with a given value of h_{ie}/R as shown in Fig. 10. When h_{fe} exceeds 56, two of the roots have positive real components, thus sustained oscillations are possible when this critical value of h_{fe} is exceeded.

Similar graphs may now be obtained for a four-section oscillator. Figure 11 gives the possible network and shows the range of the four roots with h_{ie}/R as a parameter. The root locus of the active network for two typical values of h_{ie}/R is presented in Figs. 12 and 13, where h_{fe} is the parametric variable.

6. Oscillation Frequencies

6.1. *Time Response*

The loci of the roots are significant design criteria since they predict from h_{fe} and h_{ie}/R the type of

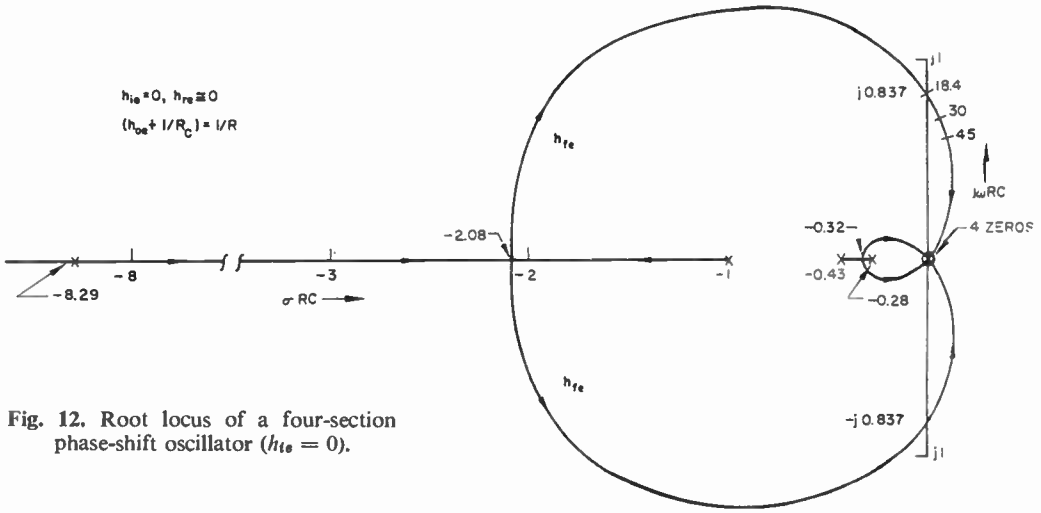


Fig. 12. Root locus of a four-section phase-shift oscillator ($h_{fe} = 0$).

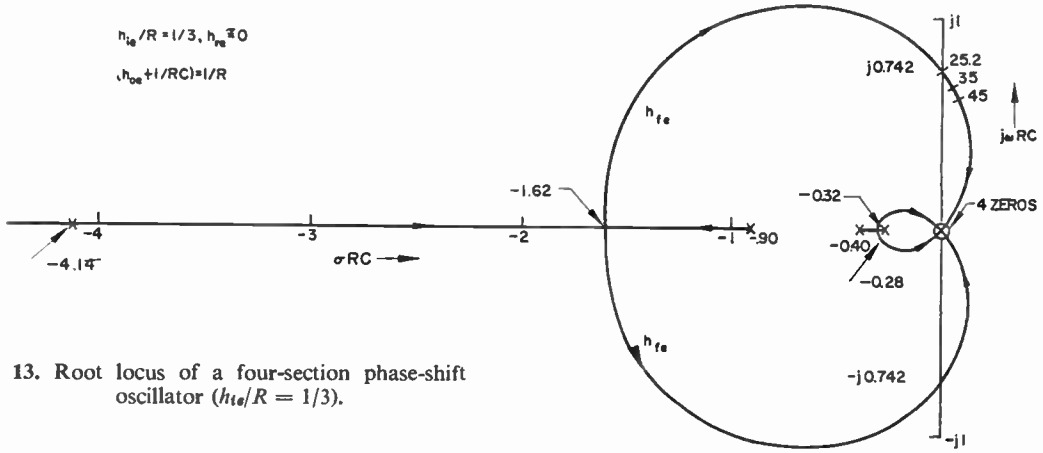


Fig. 13. Root locus of a four-section phase-shift oscillator ($h_{fe}/R = 1/3$).

oscillation:

If the roots are located in the left-hand s-plane, damping will result.

If the roots are exactly on the imaginary axis sustained oscillations will result, at least in theory.

If the roots are in the right-hand plane, the oscillations will be ever increasing for a linear system. In practice, the non-linear action of the transistor serves to limit the divergence of the amplitude.

It will be assumed that the frequency of operation will be very closely indicated by the location of the roots in the right-hand plane. For poles lying close to the imaginary axis, the error introduced by this assumption is negligible.

For example, from Fig. 12 for a given value of h_{fe}, say 35, the roots take the form

$$s = (-\sigma_1 \pm j\omega_1)(RC) \quad \text{in the left-hand plane}$$

and

$$s = (+\sigma_2 \pm j\omega_2)(RC) \quad \text{in the right-hand plane}$$

Roots in the left-hand plane produce decaying oscillations and are ignored.

For $h_{fe} = 35$, typical values for the roots are:

$$s = (+0{\cdot}5 \pm j0{\cdot}66)RC$$

which corresponds to an increasing oscillation. However, the non-linear action of the transistor will cause the roots to move to the left until they lie on the imaginary axis and constant oscillation occurs. Thus we have:

$$s = (0 \pm j0{\cdot}66)RC \text{ for non-linear limiting}$$

and by the residue theorem we may find the time response as:

$$f(t) = \left[\sum \text{residues} \frac{e^{st}}{(s+\sigma-j\omega)(s+\sigma+j\omega)}\right] u(t) \text{ at roots} \quad \ldots\ldots(13)$$

$$= (1/\omega)\sin(\omega t) \quad \ldots\ldots(14)$$

where $\omega = (0{\cdot}66)RC$ for this example.

6.2. *Minimum Gain Requirements*

Theoretically, oscillations start as the locus crosses the imaginary axis. This starting point is of interest and can be easily found.

On the jω axis

$$\sigma = 0 \quad \text{and} \quad s = \pm j\omega \qquad \text{.....(15)}$$

With $n = 4$ and $h_{ie} = 0$, the topology equation (9) becomes in this example

$$F_4(j\omega CR) + h_{fe} = 0 \qquad \text{.....(16)}$$

or

$$[\omega RC^4(1+h_{fe}) - 15(\omega RC)^2 + 1] + \\ + j[-10(\omega RC)^3 + 7\omega RC] = 0 \qquad \text{.....(17)}$$

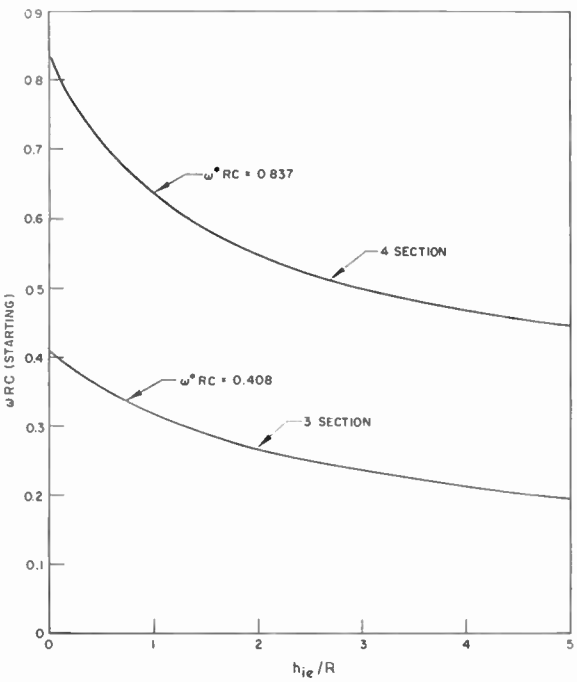


Fig. 14. Starting frequencies for three- and four-section phase-shift oscillators as a function of transistor input impedance.

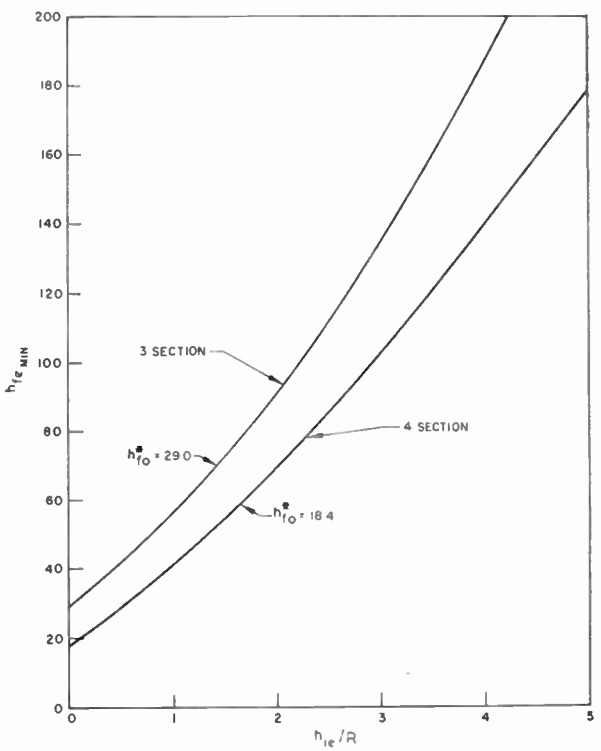


Fig. 15. Minimum current amplification for three- and four-section phase-shift oscillators as a function of transistor input impedance.

The imaginary part gives the starting frequency of oscillation f_0^*

$$\omega = (1/RC)\sqrt{7/10} = 2\pi f_0^* \qquad \text{.....(18)}$$

Minimum current gain h_{fo}^* follows in the same manner. Values of starting frequency and minimum current gain are summarized in Table 3 and plotted in Figs. 14 and 15.

It is often convenient to plot the graphs of Figs. 14 and 15 such that ωRC is a function of h_{fe}, the current gain, with h_{ie}/R as a parameter as shown in

Table 3

Starting Parameters for Three- and Four-Section Phase-Shift Oscillators as a Function of $x = h_{ie}/R$

		3 Sections	4 Sections	Refs.
Starting frequency f_0	$2\pi f_0 RC$	$(6+4x)^{-\frac{1}{2}}$	$(7+x)^{\frac{1}{2}}(10+10x)^{-\frac{1}{2}}$	Fig. 14
Minimum current gain for starting oscillation	h_{fo}	$4x^2+23x+29$	$\dfrac{56x^3+473x^2+1210x+901}{x^2+14x+49}$	Fig. 15
Limiting value of h_{fo} as $x \to 0$	h_{fo}^*	29·0	18·4	Fig. 15
Limiting value of f^* as $x \to 0$	$2\pi f_0^* RC$	0·408	0·837	Fig. 14

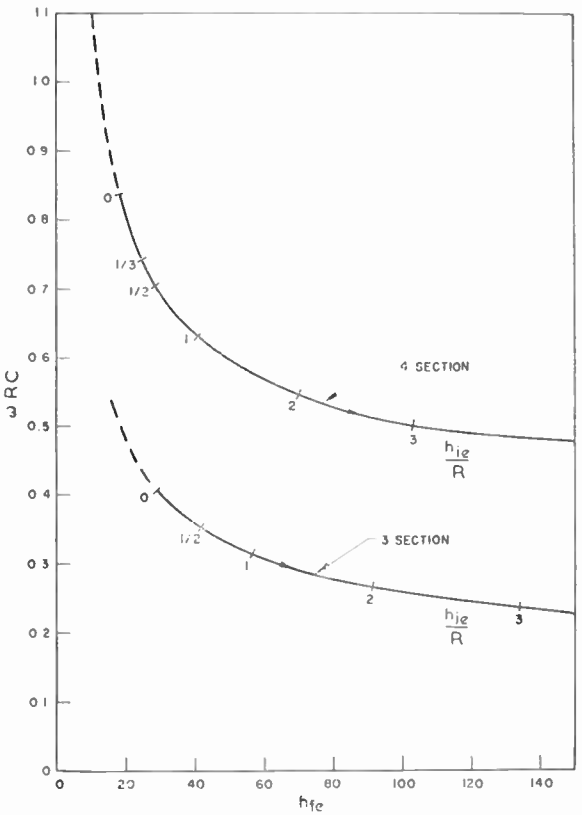


Fig. 16. Starting frequency of oscillation of a phase-shift oscillator as a function of current gain and maximum value of input impedance of the transistor.

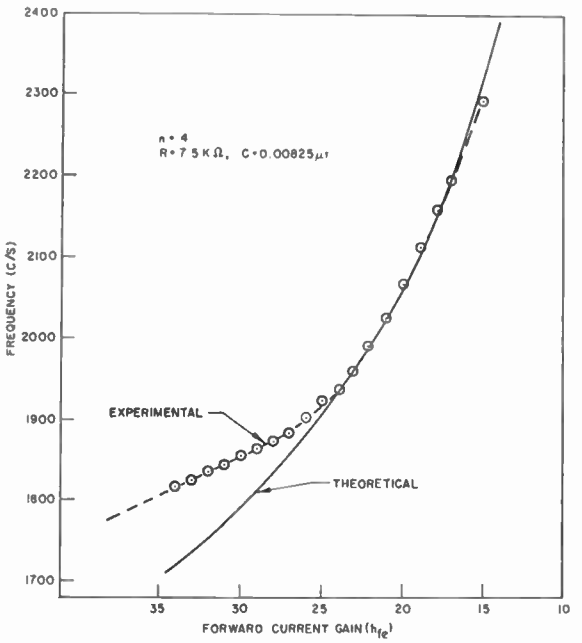


Fig. 17. Comparison of experimental and calculated starting frequency of a four-section phase-shift oscillator as a function of transistor current gain.

Fig. 16. Note that increasing values of h_{fe} indicate increasing values of frequency stability. This graph is of practical importance to the circuit designer in determining the current gain required for a prescribed frequency stability.

7. Predicting the Derating of the Oscillator

The locus of roots possible for a four-section transistor phase-shift oscillator were plotted on an expanded frequency scale for comparison with experimental results in Fig. 17.

The variation of the experimental results from the theoretical prediction may be explained by one or all of the following reasons:

The capacitors were not equal.

The action of the transistor was non-linear.

Graphical approximations were made.

Figure 16 shows that the frequency stability of a transistor phase-shift oscillator can be increased by designing the circuit to use a high-gain transistor. Considerable improvement can be obtained by using an h_f above 50. Beyond this value, further increase in h_{fe} does not significantly improve frequency stability. Since h_{fe} is the parameter most susceptible to radiation damage, attention to choice of h_{fe} is important if frequency stability is desired.

The minimum value of h_{fe} at which oscillations occur is 18·4. With h_{ie} as 2·0 kΩ h_{fe} cannot be less than 24. In view of experimental uncertainty in the measurements of h_{fe}, only relative values of h_{fe} are presented. The scale shown is adjusted to make the lowest measured value of h_{fe} coincide with the lowest permissible calculated value of h_{fe}. Agreement between experimental and theoretical values in Fig. 17 is good for low values of h_{fe}, but is not quite as good at higher values of h_{fe} due to non-linearities of circuit response at high h_{fe}.

8. Acknowledgments

This paper is a condensation of a thesis submitted jointly by W. M. Locke and V. W. Moore, Jr., for the M.S. degree at the United States Naval Postgraduate School, Monterey, California. The experimental work was conducted at Lockheed Missiles and Space Division, Palo Alto, California.

The authors wish to thank Dr. C. H. Rothauge of the United States Naval Postgraduate School, for his encouragement and guidance as faculty advisor; and T. R. Nisbet, N. K. Marshall, and W. E. Price at Lockheed, for their suggestions and assistance.

9. References

1. V. W. Moore, Jr., W. M. Locke, and W. W. Happ, "Effects of Gamma Radiation on a Transistor RC Phase-Shift Oscillator", LMSD-238123, Lockheed Missiles and Space Division, Sunnyvale, Calif., May 1960.

   Also presented at the Winter Meeting of the Professional Group on Military Electronics of the Institute of Radio Engineers at Los Angeles. California, on 2nd February, 1961.
2. J. F. Hanson, S. E. Harrison, and W. L. Hood, "The Effect of Nuclear Radiation on Electronic Components and Systems", Radiation Effects Information Center, Battelle Memorial Institute, Columbus, Ohio, 1957.
3. G. L. Keister and H. V. Stewart, "Preliminary Report of an Investigation of the Effects of Nuclear Radiation on Selected Transistors and Diodes", Report D5-1183, Boeing Airplane Company, 22 August 1956.
4. W. Hicks, "Transistor phase-shift oscillator", *Electronic Ind. Tele-Tech.*, pp. 55–6, July 1956.
5. R. P. Turner, "Transistorized phase-shift oscillator", *Radio and Television News*, p. 108, April 1956.
6. R. F. Shea, *et al.*, "Transistor Circuit Engineering", (McGraw-Hill, New York, 1957).
7. E. Wolfendale, "The Junction Transistor and its Applications", (Heywood, London, 1958).
8. V. M. Luibin, "Transistor RC oscillators with phase reversal", *Radio Engineering* (translated by Pergamon Institute, New York, 1958), **13**, No. 2, pp. 60–9.
9. E. L. Ginzton and L. M. Hollingsworth, "Phase-shift oscillators", *Proc. Inst. Radio Engrs*, **29**, pp. 43–9, February 1941.
10. M. Artzt, "Frequency modulation of resistance-capacitance oscillators", *Proc. Inst. Radio Engrs*, **32**, pp. 409–14, July 1944.
11. W. C. Vaughan, "Phase-shift oscillator", *Wireless Engr*, **26**, pp. 391–9, December 1949.
12. W. R. Hinton, "The design of RC oscillator phase-shifting networks", *Electronic Engng*, **22**, pp. 13–17, January 1950.
13. S. Sherr, "Generalized equations for RC phase-shift oscillators", *Proc. Inst. Radio Engrs*, **42**, pp. 1169–72, July 1954.
14. W. Edson, "Vacuum Tube Oscillators", (John Wiley, New York, 1953).
15. R. B. Hurley, "Junction Transistor Electronics", (John Wiley, New York, 1958).
16. T. R. Nisbet and W. W. Happ, "Visual engineering mathematics", *Electronics Digest*, 7, No. 25, December 1959.
17. S. J. Mason, "Feedback theory—some properties of signal flow graphs", *Proc. Inst. Radio Engrs*, **41**, p. 1144, September 1953.
18. J. G. Truxal, "Automatic Feedback Control System Synthesis", (McGraw-Hill, New York, 1955).
19. S. J. Mason, "Feedback theory—further properties of signal flow graphs", *Proc. Inst. Radio Engrs*, **44**, p. 920, July 1956.
20. O. Wing, "Ladder network analysis by signal flow graphs", *Trans. Inst. Radio Engrs* (*Circuit Theory*), **CT-3**, p. 289, 1956.
21. W. W. Happ, "Signal flow graphs", *Proc. Inst. Radio Engrs*, **45**, p. 1293, September 1957.
22. W. R. Evans, "Control System Dynamics", (McGraw-Hill, New York, 1954).

Manuscript received by the Institution on 27th January 1962 (Paper No. 790).

INSTITUTION NOTICES

The 1963 Convention

"ELECTRONICS AND PRODUCTIVITY"

University of Southampton—April 16th–20th

Synopses of some of the papers which have been selected for inclusion in the Convention on "Electronics and Productivity" are printed on pages 124–126 of this issue. Further synopses and the final programme will be published in the March issue of *The Radio and Electronic Engineer*.

The outline arrangements for the Convention are as follows:

Tuesday, 16th April	**Introductory session:** Opening address, Survey papers on economic and theoretical aspects.
Wednesday, 17th April	**Sensing Devices, Measurement and Telemetry.**
Thursday, 18th April	**Control and Information Processing.**
Friday, 19th April	**Industrial Applications of Electronic Systems.**
Saturday, 20th April	**Visits.**

In addition to the visit to the Esso Oil Refinery at Fawley, near Southampton, on Saturday, 20th April, there will be a limited number of conducted visits during the Convention to the Departments of Electronics, Mechanical Engineering, Aeronautical Engineering and Electrical Engineering by kind permission of the Heads of the Departments concerned. Further information about visits will be circulated in due course to all who have registered to attend the Convention.

Registration Forms for the 1963 Convention

Members in Great Britain have already received copies of the registration form for the Convention in the *Proceedings* for January–February. The forms enclosed with this issue of *The Radio and Electronic Engineer* are therefore particularly for use by members overseas and by non-member subscribers. No doubt, however, members in Great Britain will be able to use this second form to pass to colleagues.

Early return of the completed forms is advisable, particularly if it is desired to apply for resident accommodation in one of the University's Halls of Residence.

Circulation of the *Journal*

The Audit Bureau of Circulation has announced that the average monthly circulation of the Institution's *Journal* for the last six months of 1962 was 9915 copies per issue. This represents an increase of 335 compared with the previous six months and 1004 compared with a year ago.

The Norman W. V. Hayes Memorial Medal

The British Institution of Radio Engineers acts in alternate years with the Institute of Radio Engineers of America, as adjudicators for the award of the Norman W. V. Hayes Memorial Medal of the Institution of Radio Engineers, Australia. The Brit.I.R.E. recommendation for the award of the Medal for 1962 was that it should be presented to Dr. E. M. Cherry for his paper on "An Engineering Approach to the Design of Transistorized Feedback Amplifiers" which was published in the *Proceedings of the I.R.E. Australia* for May 1961. Dr. Cherry who was a lecturer in the Department of Electrical Engineering at the University of Melbourne has recently taken up an appointment at the Bell Laboratories.

There is a long standing mutual arrangement between the Brit.I.R.E. and the I.R.E. of Australia for the reprinting of outstanding papers from each others journals, and the Papers Committee recommended unanimously to the Council that Dr. Cherry's paper should be published in this way. The author has considerably revised the manuscript of the original paper which is printed on pages 127–144.

Canadian Report on Engineers' Salaries

The Provincial Associations of Professional Engineers in Canada have recently carried out a survey of salaries paid to professional engineers in British Columbia, Ontario and Quebec. The report divides engineers into six salary levels from a median of 5,400 dollars per annum to a median of 13,200 dollars p.a.

The classification guide to engineering responsibility levels is interesting in that it endeavours to lay down the duties, responsibilities, supervision—received and exercised—and entrance qualifications for each level and the normal experience requirement. The salary figures do not include pension, insurance and other employee benefits.

Copies of the report may be obtained from the Canadian Council for Professional Engineers or through the Provincial Associations, namely, Association of Professional Engineers of the Province of Ontario, 236 Avenue Road, Toronto 5, Ont.; Corporation of Professional Engineers of Quebec, 1600 Pine Avenue West, Montreal 25, P.Q.

Index to Volume 24

The index for Volume 24 of the *Journal* (July–December 1962) is enclosed with this issue. Members who wish to have their issues bound should send them to the Institution with a remittance of 16s. 6d. per volume and postage (U.K., 3s.; Overseas, 4s.).

U.D.C. 621.394.3: 681.38

Spectral Density Distributions of Signals for Binary Data Transmission

By

H. J. PUSHMAN, B.Sc., A.R.C.S.†

Summary: The analysis of spectral density distributions of transmissions is of particular interest in the assessment and comparison of high speed digital data transmission systems. Where information rates are of the order of magnitude of the bandwidth available, the particular waveforms used become significant with regard to the bandwidth utilization of the link. On the assumption that all messages are equally likely, it would appear reasonable to use the average spectral density distribution of a random message as one basis for formulating the bandwidth requirement of a particular transmission system.

In this paper, a general method of determining the average spectral density distribution directly in the frequency domain is developed. The emphasis is, however, on types of waveform which are of interest in high speed transmission, and various particular cases are discussed in detail. It can be argued in the case of f.m. that to avoid discontinuities in the waveform, the signal can be generated by switching between locked oscillators, and for a smooth transition—a condition for minimum high frequency components—a digit of the higher frequency must contain an integral number of cycles more than a digit of the lower frequency. Waveforms with these restrictions are discussed for a wide variety of frequencies. The distributions for relevant p.m. systems are derived also. The spectral density distribution for f.m. at 600 bauds with frequencies of 1200 c/s and 1900 c/s (of interest with regard to G.P.O. transmissions) are also determined and results presented. Finally, the analysis of a part-random message is presented as an example of this type of problem.

1. Introduction

Bandwidth requirements of a binary data transmission link can be specified for a particular message or group of messages and an assessment made on the efficiency of the bandwidth utilization of the particular link. In addition, estimates of the signal to noise level in which the system will function satisfactorily can be made. However, the difficulty is to determine what is a fully representative message or group of messages on which to base these assessments. It is here that the concept of random messages is of use. If it be assumed that over a transmission link, all possible messages will eventually in time be sent, and that at any arbitrary time chosen, it is equally likely that any one of all the possible messages will be sent, then we can define therefore a frequency distribution which is based on an average over all these possible messages. It must be appreciated that any one message may deviate considerably from this average but that when viewed *in toto*, it would appear reasonable to use this as one basis for formulating the bandwidth requirement of any particular transmission system.

In this paper, a general method of determining the average spectral density distribution (i.e. the average power distribution as a function of frequency) is developed. It is completely general in that once the waveform for transmitting a digit has been specified, an expression for the average spectral density distribution can be formulated. However, the evaluation of this expression in certain cases can be extremely tedious. An introduction to this approach has been given by Lawson and Uhlenbeck.‡

This approach originally arose in conjunction with work on the assessment of f.m. and p.m. binary data high-speed transmission over voice links and consequently concentrated on waveforms which are of interest in this context.§ Several examples are worked out in detail and discussed.

2. General Theory

The transmission of binary information over a transmission link in the case of frequency modulation can be viewed as the switching at a predetermined rate between the two frequencies, the frequency state being determined by the information (or message)

† The Plessey Company Ltd., Roke Manor, near Romsey, Hants.

‡ J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals", p. 43 (McGraw-Hill, New York, 1950).

§ F. G. Jenks and D. C. Hannon, "A comparison of the merits of phase and frequency modulation for medium-speed serial binary digital transmission over telephone lines", *J. Brit.I.R.E.*, 24, No. 1, pp. 21–36, July 1962.

being sent. (Phase modulation will be treated as a particular case of f.m., in that here the switching is between two locked waveforms of the same frequency but 180 deg out of phase.)

At any arbitrary time t_1, the waveforms will be either (depending on the information)

$$f(t_1) = \sin(2\pi f_1 t_1 + \varphi_1(t_1))$$

or $$\sin(2\pi f_2 t_1 + \varphi_2(t_1))$$

where f_1 and f_2 are the two carrier frequencies and φ_1 and φ_2 are the associated phasings.

If T is the digit period, then t_1 can be expressed in terms of T, and the waveform of the rth digit can be written as

$$f(rT+t) = \sin(2\pi f_1 t + \varphi_{1,r})$$

or $$\sin(2\pi f_2 t + \varphi_{2,r})$$

where $$t_1 = rT + t \text{ for } 0 \leqslant t < T$$

and $\varphi_{1,r}$ and $\varphi_{2,r}$ are the phasings of the waveform for the rth digit. In the case of a continuous waveform $\varphi_{1,r} = \varphi_{2,r}$. We can then write this as

$$f(rT+t) = a_r \sin(2\pi f_1 t + \varphi_{1,r}) + (1-a_r)\sin(2\pi f_2 t + \varphi_{2,r}) \quad \ldots\ldots(1)$$

where a_r takes the value of 0 or 1 during the period $rT \leqslant t_1 < (r+1)T$ depending on the information being transmitted. The assumption will be made later that either value will be equally likely for a random message.

The average spectral density distribution is defined as

$$\Phi(f) = \left\{ \frac{2}{T} \lim_{N\to\infty} \frac{1}{N+1} |A(f)|^2 \right\}_{\substack{\text{averaged over all} \\ \text{possible messages}}} \quad \ldots\ldots(4)$$

It is therefore required first to evaluate $|A(f)|^2$ and then to average this over all messages.

$$|A(f)|^2 = \sum_{r=0}^{N} \sum_{s=0}^{N} \{[a_r F_{1,r}(f) + (1-a_r)F_{2,r}(f)] [a_s F^*_{1,s}(f) + (1-a_s)F^*_{2,s}(f)]\, e^{-2\pi j f(r-s)T}\}$$

where s is a dummy suffix for the purposes of the double summation and the asterisk denotes the complex conjugate.

Before proceeding to the averaging process, it should be noticed that since a_r takes the values of 0 or 1, with equal likelihood, then (the bar denoting an average)

$$\overline{a_r^2} = \tfrac{1}{2} \qquad \overline{(a_r - a_s)} = 0 \qquad \bar{a}_r = \bar{a}_s = \tfrac{1}{2}$$

$$\overline{a_r a_s} = \tfrac{1}{4} \qquad \overline{(\bar{a} - a_s)} = \overline{(\bar{a} - a_r)} = 0$$

and $$\overline{(\bar{a} - a_s)(\bar{a} - a_r)} = \tfrac{1}{4} \text{ for } r = s$$

$$= 0 \text{ for } r \neq s$$

Rewriting the coefficients a_r about their mean, then

$$|A(f)|^2 = \sum_{r=0}^{N} \sum_{s=0}^{N} \{[(\bar{a}-a_r)(F_{2,r}(f) - F_{1,r}(f)) + \bar{a}(F_{2,r}(f) + F_{1,r}(f))] \times [(\bar{a}-a_s)(F^*_{2,s}(f) - F^*_{1,s}(f)) + \bar{a}(F^*_{2,s}(f) + F^*_{1,s}(f))]\, e^{-2\pi j f(r-s)T}\} \quad \ldots\ldots(5)$$

Expanding this expression and averaging over the values of a_r (in which case the cross-product terms disappear using $\overline{(\bar{a}-a_r)} = 0$) then

$$\overline{|A(f)|^2} = \sum_{r=0}^{N} \overline{(\bar{a}-a_r)^2} |F_{2,r}(f) - F_{1,r}(f)|^2 + \sum_{r=0}^{N} \sum_{s=0}^{N} \bar{a}^2 (F_{2,r}(f) + F_{1,r}(f))(F^*_{2,s}(f) + F^*_{1,s}(f))\, e^{-2\pi j f(r-s)T} \quad \ldots\ldots(6)$$

Suppose now that a particular message has been sent with specified values for a_r for all r, then the Fourier Transform of this message can be written as

$$A(f) = \sum_{r=0}^{N} a_r F_{1,r}(f)\, e^{-2\pi j f r T} + \sum_{r=0}^{N} (1-a_r) F_{2,r}(f)\, e^{-2\pi j f r T} \quad \ldots\ldots(2)$$

where $$F_{1,r}(f) = \int_0^T \sin(2\pi f_1 t + \varphi_{1,r})\, e^{-2\pi j f t}\, dt$$

and $$F_{2,r}(f) = \int_0^T \sin(2\pi f_2 t + \varphi_{2,r})\, e^{-2\pi j f t}\, dt \quad \ldots(3)$$

It will be noted that it is possible to treat the averaging process of terms of the form

$$|(a-a_r)(F_{2,r}(f) - F_{1,r}(f))|^2$$

as separable providing $F_{1,r}(f)$ or $F_{2,r}(f)$ are independent of a_r. In general, $F_{1,r}(f)$ and $F_{2,r}(f)$ are dependent only on the past history of the information (i.e. on $a_{r-1}, a_{r-2}, \ldots a_0$) and are independent of the present value a_r. Under these conditions it is permissible to write

$$\overline{|(\bar{a}-a_r)(F_{2,r}(f) - F_{1,r}(f))|^2} = \overline{(\bar{a}-a_r)^2} |F_{2,r}(f) - F_{1,r}(f)|^2$$

but this point must be remembered when evaluating particular examples.

Forming the average spectral density function from eqn. (4)

$$\Phi(f) = \lim_{N\to\infty} \frac{2}{T}\frac{1}{N+1}\left\{\sum_{r=0}^{N} \tfrac{1}{4}\left|F_{2,r}(f)-F_{1,r}(f)\right|^2 + \sum_{r=0}^{N}\sum_{s=0}^{N} \tfrac{1}{4}(F_{2,r}(f)+F_{1,r}(f))(F^*_{2,s}(f)+F^*_{1,s}(f))\, e^{2\pi jf(r-s)T}\right\} \quad \ldots\ldots(7)$$

It will be noticed that the first term in the brackets of eqn. (7) should approach a definite limit for all values of f. In the limit as $N\to\infty$ the second term involving the double summation will be infinite for $f = n/T$ where n is an integer. For other values of f, the sum will be oscillatory and for $N\to\infty$, the limiting value will be zero. The limit of this second term has the character of a series of peaks, or delta functions at the frequencies $f = n/T$.

It is worthwhile here to determine the significance of these two terms, the first of which is a continuous distribution covering the whole range of f while the second is a series of delta functions spaced at fixed frequency intervals of $1/T$.

Going back to eqn. (2) and rewriting about $\bar{a}$

$$A(f) = \sum_{r=0}^{N} [(\bar{a}-a_r)(F_{2,r}(f)-F_{1,r}(f)) + \bar{a}(F_{2,r}(f)+F_{1,r}(f))]\, e^{-2\pi jfrT} \quad \ldots\ldots(2a)$$

The second term is independent of any of the information contained in the values of a_r for any message, and it is this term that contributes to the delta functions in the average spectral density function. It can be inferred that the delta functions do not contribute to the information bearing part of the distribution and that all the information is carried in the continuous distribution. Possibly this conclusion could have been deduced intuitively in that delta functions, being transforms of pure sinusoidal waveforms, cannot be expected to convey any information.

The average power P_c in the continuous spectrum is given by

$$P_c = \int_0^{\infty} \lim_{N\to\infty} \frac{2}{T}\cdot\frac{1}{N+1}\left\{\sum_{r=0}^{N} \tfrac{1}{4}\left|F_{2,r}(f)-F_{1,r}(f)\right|^2\right\} df$$

which can be rewritten, using the "energy" relationship as

$$P_c = \lim_{N\to\infty} \frac{1}{T}\cdot\frac{1}{N+1}\int_0^T \sum_{r=0}^{N} \tfrac{1}{4}\left|f_{2,r}(t)-f_{1,r}(t)\right|^2 dt \quad \ldots\ldots(8)$$

where $f_{1,r}(t) = \sin(2\pi f_1 t+\varphi_{1,r})$

and $f_{2,r}(t) = \sin(2\pi f_2 t+\varphi_{2,r})$

as defined in eqn. (1).

Expanding the modulus term

$$P_c = \lim_{N\to\infty} \frac{1}{T}\cdot\frac{1}{N+1}\sum_{r=0}^{N}\frac{1}{4}\int_0^T [f^2_{2,r}(t)+f^2_{1,r}(t)-2f_{1,r}(t)f_{2,r}(t)]\, dt \quad \ldots\ldots(9)$$

Now the total power in the original f.m. waveform is given by

$$P = \lim_{N\to\infty} \frac{1}{T}\cdot\frac{1}{N+1}\sum_{r=0}^{N}\frac{1}{2}\int_0^T [f^2_{2,r}(t)+f^2_{1,r}(t)]\, dt \quad \ldots\ldots(10)$$

so that the average power transmitted in the delta functions is (from eqns. (9) and (10))

$$P-P_c = \lim_{N\to\infty} \frac{1}{T}\cdot\frac{1}{N+1}\sum_{r=0}^{N}\frac{1}{4}\int_0^T [f^2_{2,r}(t)+f^2_{1,r}(t)+2f_{2,r}(t)f_{1,r}(t)]\, dt \quad \ldots\ldots(11)$$

In the case of a continuous waveform in which $\varphi_{1,r} = \varphi_{2,r}$ ($= \varphi_r$ say) in eqn. (1)

$$\int_0^T f_{1,r}(t)f_{2,r}(t)\, dt = \int_0^T \sin(2\pi f_1 t+\varphi_r)\sin(2\pi f_2 t+\varphi_r)\, dt$$

$$= \frac{T}{2}\left\{\frac{\sin 2\pi(f_1-f_2)T}{2\pi(f_1-f_2)T} - \frac{\sin(2\pi(f_1+f_2)T+2\varphi_r)}{2\pi(f_1+f_2)T} + \frac{\sin 2\varphi_r}{2\pi(f_1+f_2)T}\right\}$$

Similarly $$\int_0^T f_{1,r}^2(t)\,dt = \frac{T}{2}\left\{1 - \frac{\sin(4\pi f_1 T + 2\varphi_r) - \sin 2\varphi_r}{4\pi f_1 T}\right\}$$

and $$\int_0^T f_{2,r}^2(t)\,dt = \frac{T}{2}\left\{1 - \frac{\sin(4\pi f_2 T + 2\varphi_r) - \sin 2\varphi_r}{4\pi f_2 T}\right\} \quad \text{.....(12)}$$

If we take the limit as $N \to \infty$ of the sum over r in eqn. (11), the terms containing φ_r in equations (12) will average out and tend to zero for both f_1T and f_2T other than integral values. In the event that f_1T and f_2T are both integral, then the terms containing φ_r are identically zero. Equation (11) then reduces to

$$P - P_c = \tfrac{1}{4} + \tfrac{1}{4} \cdot \frac{\sin 2\pi(f_1 - f_2)T}{2\pi(f_1 - f_2)T} \quad \text{.....(13)}$$

The function $\dfrac{\sin 2\pi(f_1 - f_2)T}{2\pi(f_1 - f_2)T}$ is a measure of the cross-correlation between the two waveforms. It can take values varying between 1 and $-0{\cdot}217$ depending on the relationships between f_1 and f_2. In the vicinity of the value 1, f_1 is approximately equal to f_2 and is obviously not of interest here. As the difference between f_1 and f_2 increases, so the function tends to zero, i.e. the correlation between the waveforms gets increasingly smaller as is to be expected. In most practical cases, therefore, the cross-correlation can be taken as approximately zero implying from eqn. (13) that about half the power of an f.m. system is contained in the continuous part of the spectral density distribution while the other half appears as a series of delta functions.

In the light of the preceding analysis it is possible to deduce equivalent waveforms which contribute only to the continuous distribution.

Define $$G_{1,r}(f) = \tfrac{1}{2}\{F_{2,r}(f) - F_{1,r}(f)\}$$

and $$G_{2,r}(f) = \tfrac{1}{2}\{F_{1,r}(f) - F_{2,r}(f)\}$$

Then

$$\{G_{1,r}(f) - G_{2,r}(f)\} = \{F_{2,r}(f) - F_{1,r}(f)\}$$

and $$\{G_{1,r}(f) + G_{2,r}(f)\} = 0$$

$G_{1,r}(f)$ and $G_{2,r}(f)$ now represent a pair of waveforms π radians out of phase, and hence define an equivalent phase modulation system with complex waveforms given by $\frac{1}{2}(f)_{1,r}(t) - f_{2,r}(t))$ and $-\frac{1}{2}(f_{1,r}(t) - f_{2,r}(t))$. The average power required in this case to transmit the same continuous part of the distribution by reason of the previous argument, is approximately half that required by the equivalent f.m. system. It should be noted that a p.m. system corresponds to a cross-correlation factor of -1.

A general expression for the average spectral density distribution is derived in eqn. (7). This expression is completely general and can be evaluated for any particular waveforms. The solutions for some examples of general interest are outlined below.

3. Application to Specific Examples

3.1. *Locked Oscillators*

One class of f.m. of considerable application is where the waveform is generated by switching between two locked oscillators. The assumption is made here that the two frequencies can be expressed in the form

$$f_1 T = n_1/m_1 \quad \text{and} \quad f_2 T = n_2/m_2$$

where n_1, n_2, m_1 and m_2 all take integral values. This implies that, if m is the lowest common multiple of m_1 and m_2 the waveforms at both frequencies will eventually be repeated every m digit periods, i.e.

$$F_{1,r}(f) = F_{1,r+m}(f) = F_{1,r+2m}(f) = \ldots$$

and similarly for $F_{2,r}(f)$.

Using this periodicity property of the waveforms in the continuous part of the spectrum, eqn. (7) can now be written as

$$\Phi(f) = \frac{2}{T} \cdot \frac{1}{m}\left\{\sum_{r=0}^{m-1} \tfrac{1}{4}\left|F_{2,r}(f) - F_{1,r}(f)\right|^2\right\} +$$

$$+ \frac{2}{T} \lim_{N\to\infty} \frac{1}{N+1} \sum_{r=0}^{N} \sum_{s=0}^{N} \tfrac{1}{4}(F_{2,r}(f) + F_{1,r}(f))(F_{2,s}^*(f) + F_{1,s}^*(f))\, e^{-2\pi j f(r-s)T} \quad \text{.....(14)}$$

where $$F_{1,r}(f) = \int_0^T \sin(2\pi f_1 t + \varphi_{1,r})\, e^{-2\pi j f t}\, dt$$

and $$\varphi_{1,r} = \frac{2\pi n_1}{m_1} \cdot r$$

Similarly for $F_{2,r}(f)$.

Evaluation of this integral gives

$$F_{1,r}(f) = \frac{1}{4\pi}\left\{(\cos\varphi_{1,r} - j\sin\varphi_{1,r})\left[\frac{1-\cos 2\pi(f_1+f)T + j\sin 2\pi(f_1+f)T}{(f_1+f)}\right] + \right.$$
$$\left. + (\cos\varphi_{1,r} + j\sin\varphi_{1,r})\left[\frac{1-\cos 2\pi(f_1-f)T - j\sin 2\pi(f_1-f)T}{(f_1-f)}\right]\right\}$$

and similarly for $F_{2,r}(f)$.

Substitution of values of f_1, f_2 and T in these expressions will readily yield the continuous part of the distribution. However, the process, particularly for large values of m, can be very tedious.

A particular and important sub-class of the above will now be investigated more fully. A restriction is placed on the two frequencies such that a digit at the higher frequency must contain an integral number of cycles more than a digit at the lower frequency. This type of f.m. is important in that the waveform is continuous and satisfies a condition for minimum high frequency components in that discontinuities can be avoided.

Cases of interest dealt with here are for $m = 1, 2, 4$ and these are discussed in detail below. Interest is mainly centred on the continuous part of the distribution, although in some cases, the delta functions are identified. It will be noted that in eqn. (13), the cross-correlation function $\frac{\sin 2\pi(f_1-f_2)T}{2\pi(f_1-f_2)T}$ is identically zero for all f_1 and f_2, $(f_1 \neq f_2)$, with the above restriction. Consequently, in these cases considered, only half the transmitted power is present in the continuous part of the distribution, the remainder being in the delta functions.

3.1.1. Frequency modulation ($m = 1$)

Here there are an integral number of cycles per digit period and it can be shown that

$$|F_{1,1}(f) - F_{2,1}(f)|^2 = \frac{\sin^2\pi fT}{4\pi^2}\left|(\cos\varphi - j\sin\varphi)\left(\frac{1}{f_1+f} - \frac{1}{f_2+f}\right) + (\cos\varphi + j\sin\varphi)\left(\frac{1}{f_1-f} - \frac{1}{f_2-f}\right)\right|^2$$

The delta functions can be determined as follows: $\left|\sum_r e^{-2\pi jfrT}\right|^2$ has delta functions at $f = n/T$ for all integral values of n.

However, $|F_{1,1}(f) + F_{2,1}(f)|^2$ has double zeros for all $f = n/T$ except at $f = f_1$ and $f = f_2$ where its value is finite.

The average spectral density distribution is given therefore by

$$\Phi(f) = \frac{2}{T}\cdot\frac{1}{4}\cdot\frac{\sin^2\pi fT}{4\pi^2}\left|(\cos\varphi - j\sin\varphi)\left(\frac{1}{f_1+f} - \frac{1}{f_2+f}\right) + (\cos\varphi + j\sin\varphi)\left(\frac{1}{f_1-f} - \frac{1}{f_2-f}\right)\right|^2 + \tfrac{1}{8}[\delta(f-f_1) + \delta(f-f_2)]$$

A case of interest with $n_1 = 1$, $n_2 = 2$ and $n_1 = 1$, $n_2 = 3$ is plotted in Figs. 1 and 2 for both $\varphi = 0$ and $\varphi = \pi/2$.

3.1.2. Phase modulation ($m = 1$)

As a special case, phase modulation can be represented by

$$n_1 = n_2 \quad \text{and} \quad F_{1,1}(f) = -F_{2,1}(f)$$

Then $$|F_{2,1}(f) - F_{1,1}(f)|^2 = 4|F_{1,1}(f)|^2$$

and $$|F_{2,1}(f) + F_{1,1}(f)|^2 \equiv 0$$

The spectral density distribution is therefore

$$\Phi(f) = \frac{T}{2}\cdot\frac{\sin^2\pi fT}{\pi^2T^2}\left|(\cos\varphi - j\sin\varphi)\frac{1}{(f_1+f)} + (\cos\varphi + j\sin\varphi)\frac{1}{(f_1-f)}\right|^2$$

The curve for $n_1 = n_2 = 1$ is plotted in Fig. 3 for both $\varphi = 0$ and $\varphi = \pi/2$. There are no delta functions.

3.1.3. Frequency modulation ($m = 2$)

$$f_1 = (2n_1+1)/2T \quad \text{and} \quad f_2 = (2n_2+1)/2T$$

i.e. an odd number of half cycles per digit period.

$$F_{1,1}(f) = -F_{1,2}(f)$$
$$= \frac{1}{4\pi}(1+\cos 2\pi fT - j\sin 2\pi fT)\left(\frac{\cos\varphi - j\sin\varphi}{(f_1+f)} + \frac{\cos\varphi + j\sin\varphi}{(f_1-f)}\right)$$

and similarly for $F_{2,1}(f)$ and $F_{2,2}(f)$.

Once again delta functions occur only at $f = f_1$ and $f = f_2$.

Substituting in the spectral density distribution gives

$$\Phi(f) = \frac{2}{T}\cdot\frac{1}{4}\cdot\frac{\cos^2\pi fT}{4\pi^2}\left|(\cos\varphi - j\sin\varphi)\left(\frac{1}{f_1+f}-\frac{1}{f_2+f}\right)+(\cos\varphi + j\sin\varphi)\left(\frac{1}{f_1-f}-\frac{1}{f_2-f}\right)\right|^2$$
$$+\tfrac{1}{8}[\delta(f-f_1)+\delta(f-f_2)]$$

Examples for $n_1 = 0$, $n_2 = 1$ and $n_1 = 1$, $n_2 = 2$ and $n_1 = 0$, $n_2 = 2$ are plotted in Figs. 4, 5 and 6 respectively.

3.1.4. Phase modulation ($m = 2$)

Once again, as a special case take $n_1 = n_2$ and

$$F_{1,1}(f) = -F_{1,2}(f) = -F_{2,1}(f) = F_{2,2}(f)$$

giving

$$\Phi(f) = \frac{T}{2}\frac{\cos^2\pi fT}{\pi^2T^2}\left|\frac{(\cos\varphi - j\sin\varphi)}{(f_1+f)} + \frac{(\cos\varphi + j\sin\varphi)}{(f_1-f)}\right|^2$$

The case for $n = 1$ is plotted in Fig. 7 for both $\varphi = 0$ and $\varphi = \pi/2$.

3.1.5. Frequency modulation ($m = 4$)

The cases not previously covered are

$$f_1 = (4n_1 \pm 1)/4T \quad \text{and} \quad f_2 = (4n_2 \pm 1)/4T$$

The Fourier transforms of the waveforms are

$$F_{1,1}(f) = \frac{1}{4\pi}\left\{\frac{(1\pm\sin 2\pi fT \pm j\cos 2\pi fT)}{(f_1+f)}(\cos\varphi - j\sin\varphi) + \frac{(1\mp\sin 2\pi fT \mp j\cos 2\pi fT)}{(f_1-f)}(\cos\varphi + j\sin\varphi)\right\}$$

and $F_{1,r}(f)$ can be deduced from $F_{1,r}(f)$ by changing φ through multiples of $\pi/2$ and similarly for $F_{2,r}(f)$.

Summing over r, then $\sum_{r=1}^{4}|F_{1,r}(f)-F_{2,r}(f)|^2 = 4|F_{1,1}(f)-F_{2,1}(f)|^2$

After a certain amount of algebraic manipulation, the average spectral density distribution is obtained as

$$\Phi(f) = \frac{1}{T}\cdot\frac{1}{16\pi^2}\left\{(1\pm\sin 2\pi fT)\left(\frac{1}{f_1+f}-\frac{1}{f_2+f}\right)^2+(1\mp\sin 2\pi fT)\left(\frac{1}{f_1-f}-\frac{1}{f_2-f}\right)^2\right\}+\tfrac{1}{8}[\delta(f-f_1)+\delta(f_2-f)]$$

It will be noted that the final distribution is independent of the phasing φ.

Examples for $f_1T = \frac{3}{4}$, $f_2T = \frac{7}{4}$ and $f_1T = \frac{5}{4}$, $f_2T = \frac{9}{4}$ and $f_1T = \frac{3}{4}$, $f_2T = \frac{11}{4}$ are presented in Fig. 8.

3.2. *Unlocked Oscillators*

In this type of problem, the waveform is assumed continuous and generated by switching between two frequencies, the frequency state being determined by the information being transmitted. A general expression is derived and then applied to frequencies of 1200 c/s and 1900 c/s at 600 bauds, this form of f.m. being of interest with regard to G.P.O. transmissions.

In eqn. (1), since the waveform is continuous, then $\varphi_{1,r} = \varphi_{2,r}$ ($= \varphi_r$ say). The phasing φ_r is given by

$$\varphi_r = \varphi + f_1T\sum_{k=0}^{r-1}a_k + f_2T\sum_{k=0}^{r-1}(1-a_k)$$

i.e. φ_r is determined by an arbitrary constant phasing φ and the phase changes that have occurred due to f_1 and f_2 during the preceding part of the message.

Now $$F_{1,r}(f) = \int_0^T \sin(2\pi f_1 t + \varphi_r)\, e^{-2\pi jft}\, dt$$

which will be written as

$$F_{1,r}(f) = \cos\varphi_r F_{1,\sin}(f) + \sin\varphi_r F_{1,\cos}(f)$$

where $$F_{1,\sin}(f) = \int_0^T \sin 2\pi f_1 t\, e^{-2\pi jft}\, dt$$

and $$F_{1,\cos}(f) = \int_0^T \cos 2\pi f_1 t\, e^{-2\pi jft}\, dt$$

and similarly for $F_{2,r}(f)$.

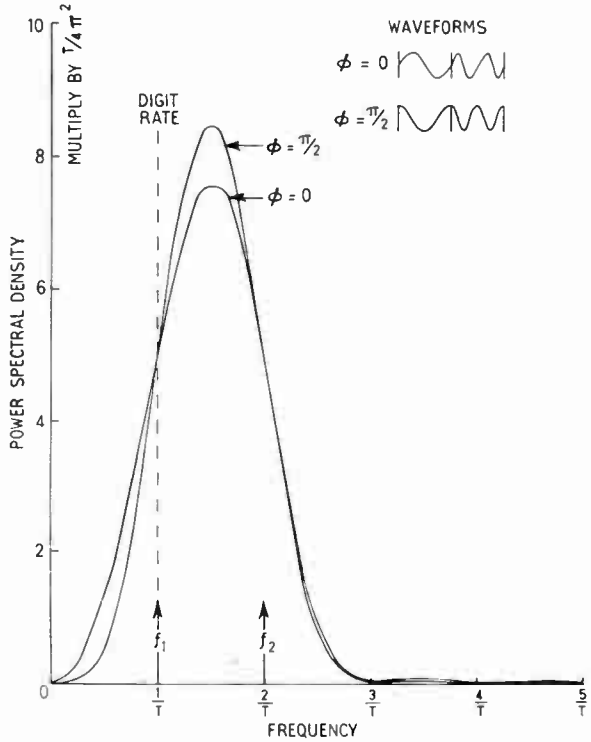


Fig. 1. Power spectral density distribution. Frequency modulation: $f_1T = 1$, $f_2T = 2$.

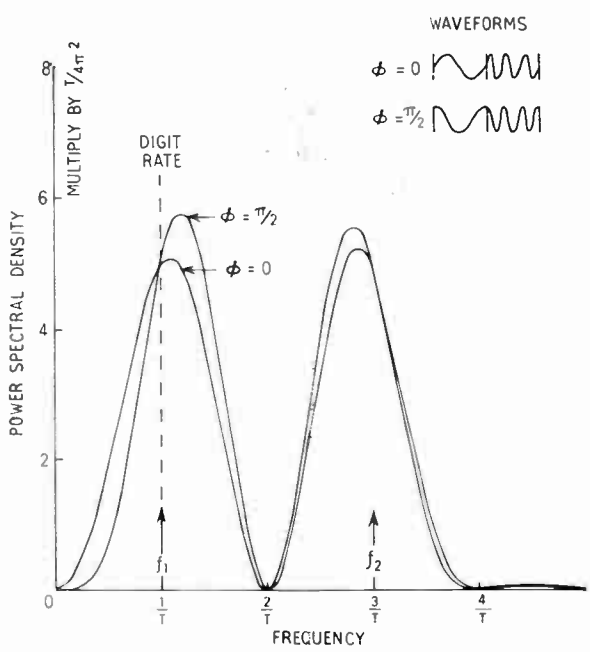


Fig. 2. Power spectral density distribution. Frequency modulation: $f_1T = 1$, $f_2T = 3$.

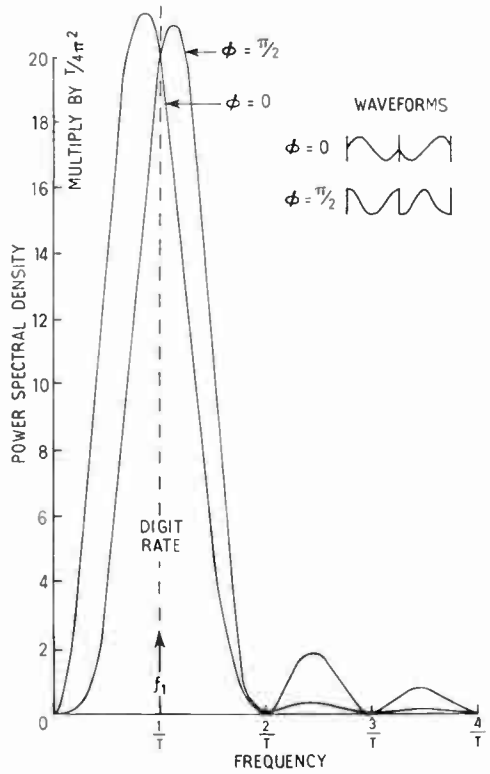


Fig. 3. Power spectral density distribution. Phase modulation: $f_1T = 1$.

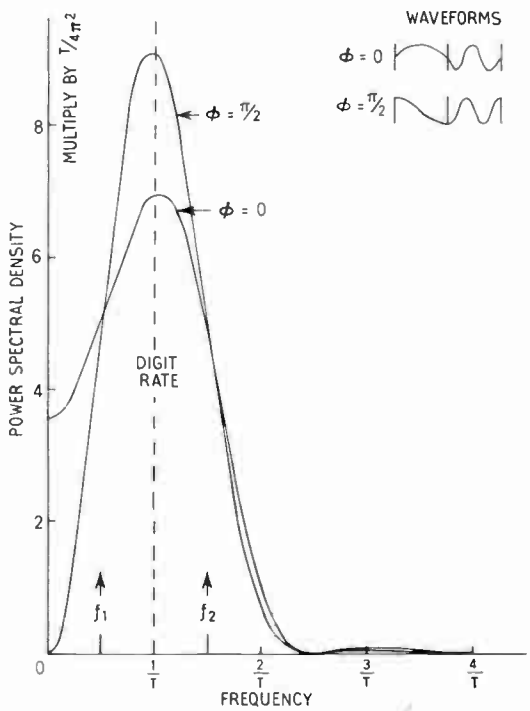


Fig. 4. Power spectral density distribution. Frequency modulation: $f_1T = 1/2$, $f_2T = 3/2$.

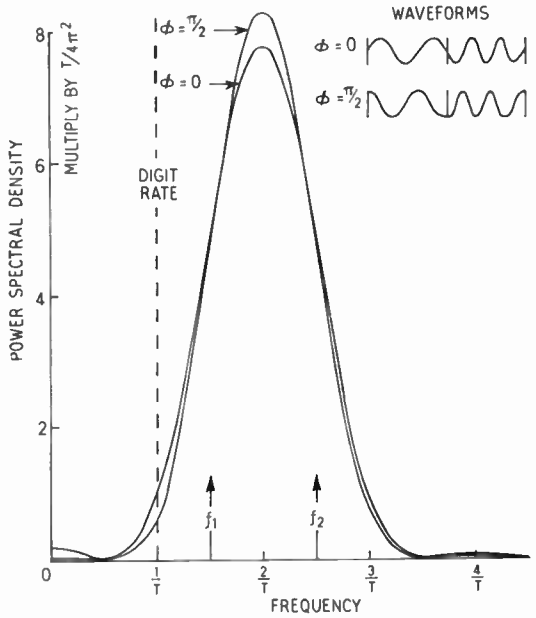


Fig. 5. Power spectral density distribution. Frequency modulation: $f_1T = 3/2$, $f_2T = 5/2$.

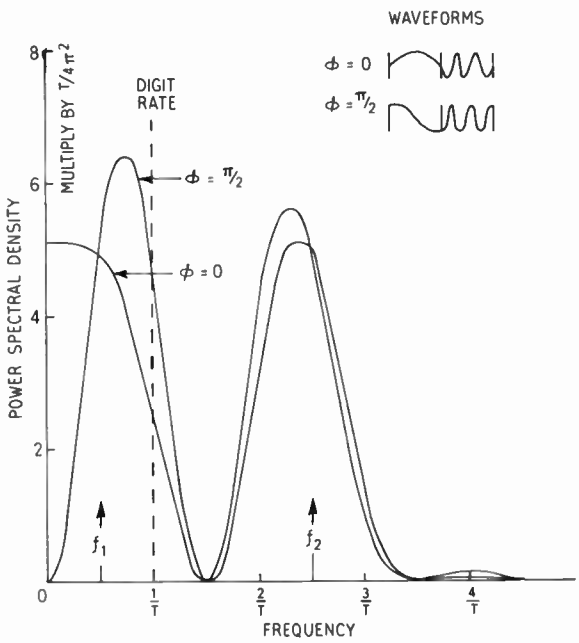


Fig. 6. Power spectral density distribution. Frequency modulation: $f_1T = 1/2$, $f_2T = 5/2$.

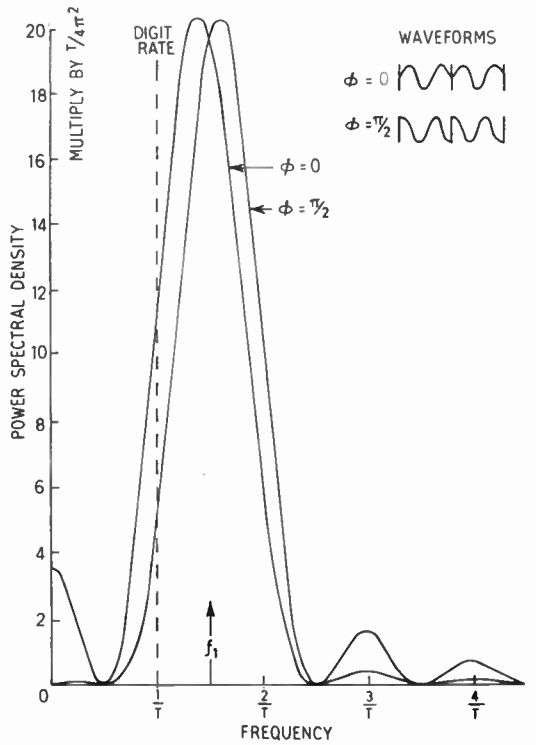


Fig. 7. Power spectral density distribution. Phase modulation: $f_1T = 3/2$.

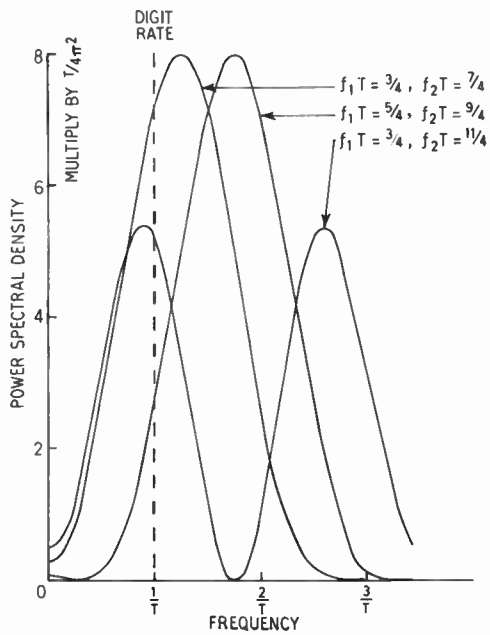


Fig. 8. Power spectral density distribution. Frequency modulation:

$f_1T = 3/4$, $f_2T = 7/4$

$f_1T = 5/4$, $f_2T = 9/4$

$f_1T = 3/4$, $f_2T = 11/4$.

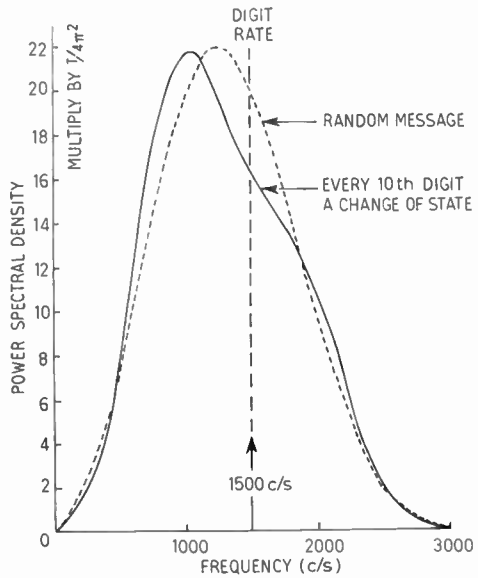


Fig. 10. Power spectral density distribution for part random message: phase modulation.

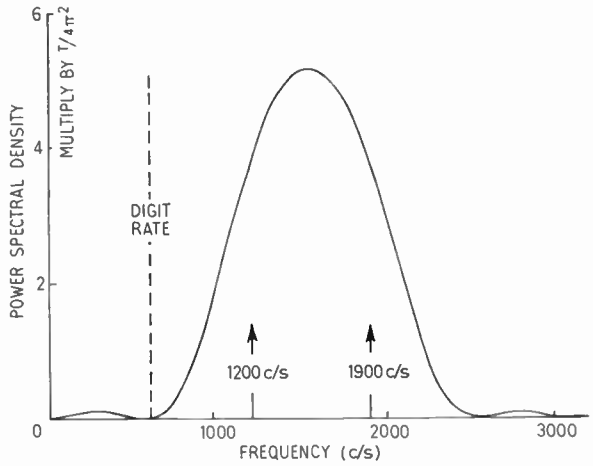


Fig. 9. Power spectral density distribution. Frequency modulation: $f_1 = 1200$ c/s, $f_2 = 1900$ c/s, 600 bauds.

Substituting these values into eqn. (7) and noting that in general

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{r=0}^{N}\cos^2\varphi_r = \lim_{N\to\infty}\frac{1}{N+1}\sum_{r=0}^{N}\sin^2\varphi_r = \tfrac{1}{2}$$

and

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{r=0}^{N}\sin\varphi_r\cos\varphi_r = 0$$

then

$$\Phi(f) = \frac{1}{4T}\{|F_{1,\sin}(f)-F_{2,\sin}(f)|^2 + |F_{1,\cos}(f)-F_{2,\cos}(f)|^2\} + \text{delta function terms} \quad \ldots\ldots(15)$$

It will be noted that the continuous part of the spectrum is independent of the arbitrary phasing φ.

We are primarily interested in the continuous part of the distribution as the information-bearing part and will not consider the delta functions further.

Substituting in the above expression the appropriate transforms, the continuous part of the distribution can be expressed in the form

$$[\Phi(f)]_{\text{cont}} = \frac{T}{8}\left\{\frac{\sin^2\pi(f_1-f)T}{\pi^2(f_1-f)^2T^2} + \frac{\sin^2\pi(f_1+f)T}{\pi^2(f_1+f)^2T^2} + \frac{\sin^2\pi(f_2-f)T}{\pi^2(f_2-f)^2T^2} + \frac{\sin^2\pi(f_2+f)T}{\pi^2(f_2+f)^2T^2} - \right.$$
$$- \frac{1}{2}\frac{(1+\cos 2\pi(f_1-f_2)T - \cos 2\pi(f_1-f)T - \cos 2\pi(f_2-f)T)}{\pi^2(f_1-f)(f_2-f)T^2} -$$
$$\left. - \frac{1}{2}\frac{(1+\cos 2\pi(f_1-f_2)T - \cos 2\pi(f_1+f)T - \cos 2\pi(f_2+f)T)}{\pi^2(f_1+f)(f_2+f)T^2}\right\}$$

Hence, for the particular example in mind, namely, $f_1 = 1200$ c/s, $f_2 = 1900$ c/s and $T = 1/600$, the expression for $\Phi(f)$ can be calculated.

The resulting curve for the continuous part of the distribution is presented in Fig. 9.

3.3. *Part Random Messages*

Sometimes it is necessary to analyse messages where part only is random, the remainder being systematic in a sense and may be used for synchronization or carrier regeneration purposes, etc. One type of problem already investigated by the author is one in which every kth digit is characterized by a change of state from the previous digit, and does not carry any of the message information.

Two cases considered here are

(1) Phase modulation with $f = 1/T$, i.e. one carrier cycle per digit period

(2) Frequency modulation with $f_1T = \frac{3}{4}$ and $f_2T = \frac{7}{4}$.

These two cases are treated in detail here and it is hoped that they will serve as heuristic examples for any similar types of problem.

3.3.1. Phase modulation (carrier frequency $1/T$)

With every kth digit representing a change of state from the previous digit, eqn. (2) must be rewritten as

$$A(f) = \sum_{r=0}^{N} \left\{ \sum_{s=0}^{k-1} [a_{0,r} + (1-a_{0,r})e^{-2\pi jfT} + \ldots a_{s,r} e^{-2\pi jfsT} + \ldots a_{k-1,r} e^{-2\pi jf(k-1)T}]F_1(f) + \right.$$
$$\left. + [(1-a_{0,r}) + a_{0,r} e^{-2\pi jfT} + \ldots (1-a_{s,r}) e^{-2\pi jfsT} + \ldots (1-a_{k-1,r}) e^{-2\pi jf(k-1)T}]F_2(f) \right\} \times e^{-2\pi jfkrT}$$

where
$$F_1(f) = \int_0^T \sin(2\pi f_1 t + \varphi) e^{-2\pi jft} dt$$

and
$$F_2(f) = -F_1(f)$$

and $a_{s,r}$ takes values of 0 or 1, depending on the information being transmitted.

Taking the modulus squared and averaging over the values of a gives

$$\overline{|A(f)|^2} = \tfrac{1}{4}|F_1(f)|^2 \left\{ \frac{(k-2) + 4\sin^2 \pi fT}{k} \right\}$$
$$= \tfrac{1}{4}|F_1(f)|^2 \left\{ 1 - \frac{2\cos 2\pi fT}{k} \right\}$$

Here
$$|F_1(f)|^2 = \frac{\sin^2 \pi fT}{\pi^2} \left| \frac{\sin\varphi + j\cos\varphi}{f_1 + f} - \frac{\sin\varphi - j\cos\varphi}{f_1 - f} \right|^2$$

A case of interest is when $\varphi = 0$

i.e.
$$[\Phi(f)]_{\phi=0} = \frac{2}{T} \cdot \frac{1}{4} \cdot \frac{\sin^2 \pi fT}{\pi^2} \left(\frac{1}{f_1 + f} - \frac{1}{f_1 - f} \right)^2 \left(1 - \frac{2\cos 2\pi fT}{k} \right)$$

It will be noted that the spectral density distribution when every kth digit has a change of state, by comparison with the results of section 3.1.2, is the same as that for a completely random message modified by the factor $\left(1 - \frac{2\cos 2\pi fT}{k}\right)$

The distribution with $k = 10$ and $\varphi = 0$ is presented in Fig. 10, together with the curve of a completely random message for comparison purposes.

3.3.2. Frequency modulation

Rewriting eqn. (2) making every kth digit a change of state gives

$$A(f) = \sum_{r=0}^{N} \left\{ \sum_{s=0}^{k-1} a_{0,r} F_{1,p}(f) + (1-a_{0,r}) F_{1,p+1}(f) e^{-2\pi jfT} + \ldots a_{s,r} F_{1,p+s}(f) e^{-2\pi jfsT} + \right.$$
$$+ \ldots a_{k-1,r} F_{1,p+k-1}(f) e^{-2\pi jf(k-1)T} +$$
$$+ \sum_{s=0}^{k-1} (1-a_{0,r}) F_{2,p}(f) + a_{0,r} F_{2,p+1}(f) e^{-2\pi jfT} + \ldots (1-a_{s,r}) F_{2,p+s}(f) e^{-2\pi jfsT} +$$
$$\left. + \ldots (1-a_{k-1,r}) F_{2,p+k-1}(f) e^{-2\pi jf(k-1)T} \right\} e^{-2\pi jfkrT}$$

where
$$p = \text{remainder of } \frac{rk}{4}$$

Writing

$$F_{1,p+s}(f) = F_{1,q} \text{ where } q = \text{remainder of } \frac{p+s}{4}$$

then
$$F_{1,0}(f) = \int_0^T \sin(2\pi f_1 t + \varphi) e^{-2\pi jft} dt$$

$$F_{1,1}(f) = \int_0^T \cos(2\pi f_1 t + \varphi) e^{-2\pi jft} dt$$

$$F_{1,2}(f) = -F_{1,0}(f)$$

$$F_{1,3}(f) = -F_{1,1}(f)$$

Similarly for $F_{2,p+s}(f)$.

On taking the modulus squared and averaging over the values of a after considerable algebraic manipulation, it can be shown that, for the particular value $k = 10$

$$[\Phi(f)]_{k=10} = \frac{2}{T} \cdot \frac{1}{32\pi^2} \{\tfrac{4}{5}[(1-\sin 2\pi fT)A^2 + (1+\sin 2\pi fT)B^2] + \tfrac{2}{5}\cos^2 2\pi fT(A^2+B^2+2AB\cos 2\varphi)\}$$

where $$A = \left(\frac{1}{f_1+f} - \frac{1}{f_2+f}\right) \quad \text{and} \quad B = \left(\frac{1}{f_1-f} - \frac{1}{f_2-f}\right)$$

This function is plotted in Fig. 11 for values of $\varphi = 0$ and $\varphi = \pi/2$ showing that the phasing φ of the carrier waveforms both to the digit period and to the change of state every 10th digit have very little effect on the distribution. For comparison purposes, the distribution for a completely random message derived in Section 3.1.5 is superimposed.

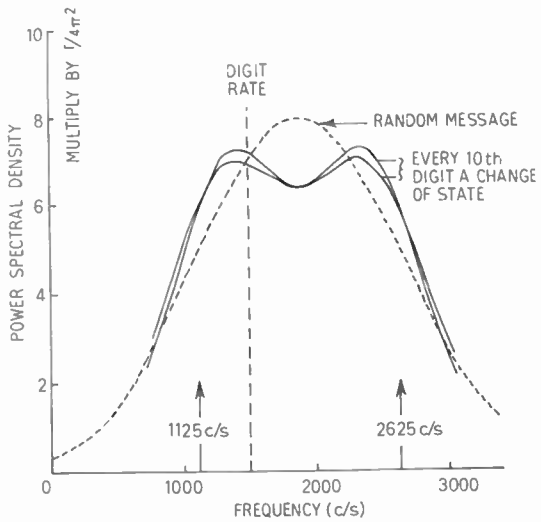


Fig. 11. Power spectral density distribution for part random message: frequency modulation.

4. Conclusions

The average spectral density distributions of a random message is considered a useful measure of the bandwidth utilization of a particular transmission system. A general expression for the average spectral density distribution is derived in eqn. (7). The evaluation of this expression in certain cases can be lengthy and tedious; however, by considering the waveforms generated by locked and unlocked oscillators separately, a certain amount of simplification in the expression is achieved as shown by eqns. (14) and (15). Fortunately, in most practical cases where the two frequencies for f.m. are simply related this complexity is not too severe.

Phase modulation can be treated as a singular form of frequency modulation in that the frequencies of the two possible states are the same, but the phasing differs by π radians.

The analysis demonstrates that the amount of power in the continuous part of the spectrum is directly related to the cross-correlation coefficient of the waveforms. On this basis, phase modulation with a cross-correlation coefficient of -1 has no associated delta functions and all the power is contained in the continuous part of the spectrum. For frequency modulation systems it is shown that in most practical cases, the cross-correlation coefficient is in the region of zero and that approximately half the power contributes to the continuous part of the distribution, the remainder appearing as delta functions.

Finally, the method of analysis can be applied to part random messages. It is felt that it is more convenient to return to the basic equations when analysing this type of problem, and no attempt has been made to generalize the result.

5. Acknowledgments

The author is indebted to Mr. F. G. Jenks of The Plessey Co. for proposing the original problems for which this analysis was developed and for stimulating discussions on the subject matter. He would also like to thank Mr. S. M. Cobb of The Plessey Co. for his careful reading and checking of the original draft.

Finally, the author is grateful to The Plessey Company for permission to present this paper.

Manuscript received by the Institution on 3rd February 1962. (Paper No. 791/C50.)

ULTRASONIC AID FOR BLIND PERSONS

A sonic aid for blind persons, produced by Ultra Electronics on the basis of a design by Dr. Leslie Kay of the Electrical Engineering Department of Birmingham, has now undergone a series of tests. This device was described in some detail in a paper by Dr. Kay—"Auditory Perception and its Relation to Ultrasonic Blind Guidance Aids"—published in the Brit.I.R.E. *Journal* in October 1962. Its development was sponsored by the National Research Development Council in association with St. Dunstan's who are naturally fully alive to any new devices which can be produced for helping blind persons.

The aid consists of a transistorized transmitter-receiver which operates with a hand-held "torch", emitting a frequency-modulated ultrasonic beam. Any energy received back by the torch differs in frequency from that leaving it, at that instant, by an amount proportional to the time taken for the energy to travel out and be reflected back. Thus, by receiving the differing signals in an earphone, the blind person is able to judge the distance of an obstacle ahead of him.

One user of the ten evaluation models so far produced, Mr. Walter Thornton, a Birmingham youth club organizer, has already made considerable progress with the use of the device, finding it extremely helpful, particularly with thick snow on the ground. He had always found it difficult previously to avoid walking into people standing immediately in front with their backs to him, but with the aid good warning is given.

On unfamiliar routes he has found it much easier to get around without personal risk, and although he has used the device over a period of three months without any training, he is quite sure that with instruction beforehand, blind persons would learn to use it more easily, and would get more benefit from it. Mr. Thornton has said that one has to develop the speed of reaction to the impulse and recognize the signal. He pointed out:

"Pitch indicates distance, and higher the pitch the further away the obstacle. Some surfaces reflect the sound back better. From a surface like glass, a smooth surface, I would expect a 'harder' type of signal. From a hedge there would be a variety of signals, and if one is pointing the torch at a person wearing a loose woollen garment, the reaction comes back in a more muffled and softer shape—although this, of course, has to be related to distance."

Sound spectra from various objects were shown in Dr. Kay's paper and subjective impressions such as Mr. Thornton describes are of considerable interest.

In addition to being used by Mr. Thornton, the aid has undergone field and evaluation trials at the Worcester College for Blind Boys and at St. Dunstan's training centre, at Ovingdean, near Brighton.

As a result of the trials so far carried out, St. Dunstan's believe the device to be capable of further development, but even at its present stage Mr. Richard Dufton, Research Director, says that it is in advance of anything yet produced in the United States or in Europe. While recognizing this, St. Dunstan's are anxious to avoid giving the impression that the aid, in its present form, is the complete answer to the mobility of the blind.

Further development to implement the lessons learned in the evaluation trials will almost certainly result in reduction of size and weight of the equipment, even though, at present, the transmitter-receiver weighs only just over two pounds (1 kg), and the torch 12¼ oz (0.35 kg). In production quantities, it is anticipated that the device can be produced for less than £100, with a further reduction if the demand were great enough.

The blind guidance aid in use. The hand-held "torch" contains two piezo-electric transducers. A hearing-aid type earphone is used.

U.D.C. 621.397.22 : 621.397.332

Effects of Manufacturing Tolerances on a Typical Non-Stabilized Line Deflection Circuit

By

D. E. LAWTON†

Summary: A typical circuit in which correct amplitude and linearity is obtained with nominal components is considered. The conditions studied are those which will cause a reduction in deflection amplitude and/or cause a deterioration in linearity. The items considered effecting width and/or linearity are: line transformer, deflection coils, drive shaping network, h.t. feed choke, tuning capacitor, feedback capacitor, c.r.t., vertical stabilizing choke, "S" correction, h.t. and heater variations, and the earth's magnetic field.

1. Introduction

By far the most critical and highest stressed part of the domestic television receiver is the horizontal deflection output circuit; it is a fast power switch, which also generates the extra high voltage for the picture tube final anode during the rapid retrace, and is particularly vulnerable to parasitic effects like spurious velocity and amplitude modulation.

Technical literature concerns itself almost entirely with basic operation principles and theoretical aspects of design. The object of this paper is to investigate the practical effect of tolerances of active and passive component elements on width, horizontal linearity and e.h.t. A typical circuit where correct amplitude and linearity is obtained with nominal components is considered. The conditions studied are those which will cause a reduction in deflection amplitude and/or cause a deterioration in linearity.

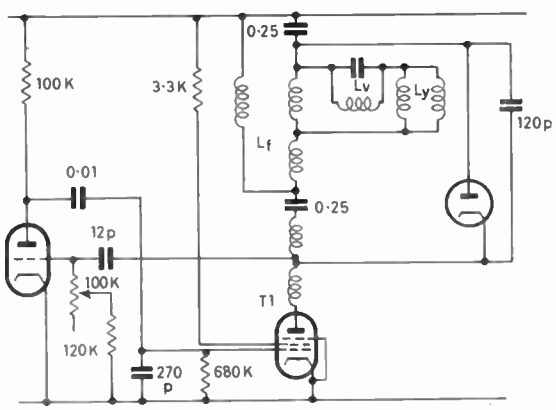


Fig. 1. The horizontal deflection time-base circuit considered in the paper.

The circuit is shown in Fig. 1. It is shown in this investigation that a reduction in scan amplitude of 6·5% and a non-linearity of 4·5% due to tolerance of components can be expected. The overscan required with nominal components to compensate for these tolerances is 11%.

The items considered effecting width and/or linearity are:

line transformer, deflection coils, drive shaping network, h.t. feed choke, tuning capacitor, feedback capacitor, c.r.t., vertical stabilizing choke, "S" correction, h.t. and heater variations, and the earth's magnetic field.

2. Picture Shift due to Non-linearity and Flyback Time

Screen width	L mm
Modulation time	T μs
Line blanking	l μs
Scan linearity	p %
Front porch	b μs
Flyback time	a μs
Sync. delay	c μs

The scan non-linearity is defined as

$$P = \frac{d-e}{d+e} \times 100\% \qquad \text{.....(1)}$$

where d and e are the largest and smallest distances respectively, at unit intervals of time.

The delay in sync. can be measured by measuring the front porch in microseconds at the grid of the sync. separator and then measuring the distance on the tube that the front porch occupies. Thus by measuring the distance the sync. delay occupies on the tube, the delay in microseconds can be calculated.

2.1. *Picture Shift due to Flyback Time*

Only the portion of scan containing picture information is displayed on the screen, the rest of the scan $(l-a)$ μs is masked off. Depending where the modulation is situated on the raster, the picture may have to be shifted either to the left or the right.

† The Plessey Company Ltd., Ilford, Essex.

The required shift is $\frac{1}{2}\{(l-a)-2(b+c)\}$ μs to the left if positive and to the right if negative.

In terms of scan length the required shift Df is:

$$Df = \frac{1}{2}\frac{L}{T}\{(l-a)-2(b+c)\} \text{ mm} \qquad \ldots\ldots(2)$$

2.2. *Picture Shift due to Non-linearity*

Is is shown in Appendix 1 that for non-linearity which is uniform along the scan the required shift to the right is $Ds = \frac{L}{600} \cdot P$ mm.

Total picture shift is the arithmetic difference between equations (2) and (7), plus of course any d.c. shift and the shift can be either to the left, to the right or zero.

2.3. *Non-linearity due to Picture Shift*

If the picture is non-linear then the modulation requires shifting to the right as can be seen from Fig. 2 (*b*). The amount of non-linearity will determine the degree of shift required, which is the shift of the deflection centre. With the new deflection centre, the deflection angle to the left side of the screen is greater than the deflection to the right. This means that the negative excursion of deflection current will be greater than the positive excursion.

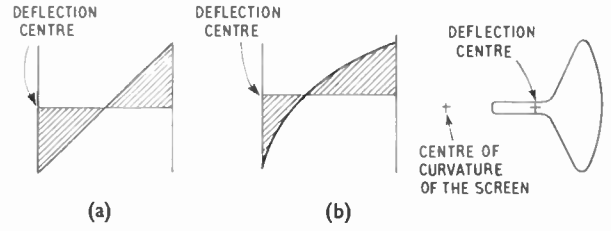


Fig. 2. Deflecting currents for linear and non-linear scanning. (*a*) Linear scan (*b*) Non-linear scan

The non-linearity is defined as the ratio of the largest to smallest distances at unit intervals of time. These are on the left and right of the scan respectively if the non-linearity varies uniformly.

The centre of deflection shift can be calculated for a particular non-linearity. The deflection angle will then be $\frac{110°}{2}+\theta$ for the deflection to the left and $\frac{110°}{2}-\theta$ for deflection to the right of the screen.

Taking a case of non-linearity on an AW43–88 tube where the largest and smallest distances are 50 mm and 42·5 mm.

$$P = \frac{7{\cdot}5}{92{\cdot}5} \times 100 = 8{\cdot}1\%$$

$$Ds = \frac{374{\cdot}5}{600} \times 8{\cdot}1 = 5{\cdot}05 \text{ mm}.$$

If this is now substituted in eqn. (8) (Appendix 1) for deflection on the theoretically flat screen:

$$1{\cdot}55 \text{ deg} = 5{\cdot}05 \text{ mm}.$$

This is in effect adding 1·55 deg on to the deflection angle to the left and subtracting 1·55 deg from the right. Thus the deflection angles are 56·55 deg and 53·45 deg.

If the deflection centre and radius of curvature of the tube face were coincident, *then* shifting the deflection centre would *not* affect the linearity but the deflection centre is much closer to the screen.

Taking equal angles of deflection of 11 deg on either side of the tube, i.e. 56·55 deg to 45·55 deg and 53·45 deg to 42·45 deg and substituting in eqn. (12) derived for deflection on the curved screen:

When $\theta = 56{\cdot}55°$; $D = 206{\cdot}9$ mm
„ $\theta = 45{\cdot}55°$; $D = 160{\cdot}3$ mm
„ $\theta = 53{\cdot}45°$; $D = 193{\cdot}3$ mm
„ $\theta = 42{\cdot}45°$; $D = 148{\cdot}3$ mm

Thus the largest and smallest distances are 46·6 mm and 45·05 mm and the non-linearity is

$$P = \frac{46{\cdot}6-45{\cdot}05}{46{\cdot}6+45{\cdot}05} \times 100$$

This is the non-linearity due to shifting the deflection centre 5·05 mm.

3. Picture Shift due to the Earth's Magnetic Field

An electron with velocity v in a magnetic field strength B will describe an arc the radius of which is r.†

$$r = \frac{mv}{eB}$$

where m = electron mass
e = electron charge
V = final anode voltage.

The velocity

$$v = \sqrt{\frac{2Ve}{m}}$$

So

$$r = \frac{m\sqrt{\frac{2eV}{m}}}{eB}$$

$$= \frac{1}{B}\sqrt{\frac{2Vm}{e}}$$

† D. G. Fink (Ed.), "Television Engineering Handbook", pp. 5–11 (McGraw-Hill, New York, 1957).

$$\frac{e}{m} = 1{\cdot}77 \times 10^{11} \text{ coulombs/kg.}$$

The vertical component of the earth's magnetic field measured at Greenwich is $0{\cdot}43 \times 10^{-4}$ Wb/m^2.

The final anode voltage $V = 16\,000$ volts.

$$r = \frac{10^4}{0{\cdot}43}\sqrt{\frac{2 \times 16\,000}{1{\cdot}77 \times 10^{11}}}$$

$$= 9{\cdot}78 \text{ metres.}$$

The distance from the c.r.t. gun aperture l on the AW43–88 is 0·244 metres.

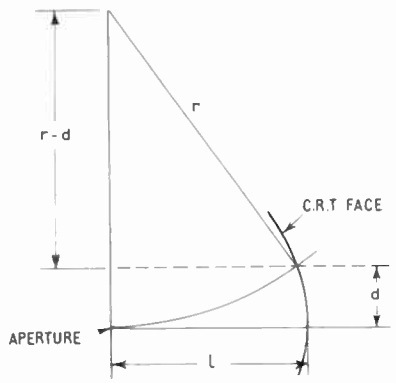


Fig. 3. Deflection of spot on c.r.t. screen.

The spot will be shifted on the screen by a distance d. From Fig. 3

$$r^2 = l^2 + (r-d)^2$$

Hence $$d = r - \sqrt{r^2 - l^2}$$

$$= 9{\cdot}78 - \sqrt{9{\cdot}78^2 - 0{\cdot}244^2} = 3 \text{ mm.}$$

The scanning coils, however, offer an amount of screening to the electron beam from the earth's magnetic field, and with present deflection yokes this shift is reduced by approximately 12%. In the British Isles a total shift to the left of 2·6 mm will result, causing a non-linearity of 0·87%.

4. Picture Shift due to the C.R.T. Gun Assembly

The electron beam may be off-centre due to the misalignment of the gun assembly. The limit set by the manufacturers is ±3% of the overall length. For the tube under consideration this is 9·6 mm, causing a non-linearity of 3·2%.

5. Non-linearity of Screen Deflection and Corresponding Non-linearity of Deflecting Current

Component tolerances affecting the linearity of the deflecting current will produce a larger screen deflection non-linearity because of the flat-faced nature of the tube.

For a linear scan the deflection angle will be divided into ten equal parts of 11 deg. As was stated above for a non-linearity of 8·1% screen deflection, the deflection to the left and right of the screen will be 55°+1·55° and 55°−1·55° respectively. The largest and smallest angles will therefore be:

$$\frac{55° + 1{\cdot}55°}{5} = 11{\cdot}31°$$

$$\frac{55° - 1{\cdot}55°}{5} = 10{\cdot}68°$$

The non-linearity of the deflection will be:

$$\frac{11{\cdot}31° - 10{\cdot}68°}{11{\cdot}31° + 10{\cdot}68°} \times 100 = 2{\cdot}86\%$$

As the deflection angle is proportional to the current through the scanning coils a non-linearity of scanning current of 2·86% will result in a screen deflection non-linearity of 8·1%.

6. The Necessity for Over-scanning 110-degree Tubes

The aspect ratio of the AW43–88 is not exactly 4 : 3 and hence to maintain the picture aspect ratio at 4 : 3 the line must be over-scanned.

Height of tube is	295 mm
Thus the scan will be	$\frac{295}{3} \times 4 = 393{\cdot}32$ mm
The tube width is	374·5 mm
Required overscan is	$\frac{393{\cdot}32 - 374{\cdot}5}{3{\cdot}745} = 5\%$

7. Relationship between E.H.T. and Deflection Angle

It is shown in the literature† that

$$\sin\theta = 0{\cdot}30\frac{BL}{\sqrt{V}}$$

where θ = deflection angle of the electron beam
L = length of deflecting field in cm
V = final anode voltage in volts
B = flux in gauss.

If we put $L = 5$ cm; $\theta = 45°$ and $V = 16\,000$

then $$B = \frac{\sqrt{16\,000} \times \sin 45°}{5 \times 0{\cdot}3}$$

$$= 60 \text{ gauss}$$

If H now remains constant and the final anode voltage is varied, the variation in deflection angle can be calculated.

When $V = 15\,000$ V; $\theta = 46{\cdot}9°$
$V = 16\,000$; $\theta = 45°$
$V = 17\,000$; $\theta = 43{\cdot}3°$

† T. Soller, M. A. Starr and G. E. Valley, "Cathode Ray Tube Displays", page 304 (McGraw-Hill, New York, 1948).

Thus θ will be reduced by 1·8 deg per kV that is 1 deg per 555·6 volts. Additionally, it can be shown that when calculating the effects of increase in tube angle that an increase of 1 deg will require 1·5% increase in scan. Thus an increase of 1·1 kV will reduce the scan by 3%.

8. Cathode-ray Tube

The maximum to minimum variation in angle for the AW43–80 is stated in Mullard Report No. 99 to be 4 deg. It is therefore assumed that the same variation in angle can be expected for the AW43–88. Thus an increase of 2 deg can be expected over a nominal tube.

It is shown in Appendix 2 that a top limit tube having an increase in angle of 2 deg would require an increase in scan of approximately 3%.

An additional affect was experienced with the introduction of aluminized tubes. This caused a reduction in deflection coil Q, and so a reduction in scan amplitude. The tube manufacturers now take precautions to mask the tube neck and area near to the scanning coils, so that a negligible reduction in scan amplitude is experienced.

9. Line Transformer

From experience and production measurements the estimated tolerance of the line transformer self-capacitance will not exceed ±3% and the linearity 1·5%. The tolerance of 3% in scan due to the line transformer variation is readily obtainable. This includes variations due to e.h.t. winding, i.e. the jig for testing the line transformers has the display tube fed from the line transformer e.h.t., so the total variation is ±3%.

The capacitance necessary to tune the transformer to the desired flyback time consists of the self-capacitance of the windings and externally added capacitances. Both of these tolerances will affect the flyback time and therefore the final anode voltage.

If the added capacitor is 50% of the total capacitance and 10% tolerance, the total capacitance is, say, 240 pF. Then the total variation of capacitance is 6·5%. When the added capacitance is 10% of the total the variation will be 2·8% of the total, but when the added capacitance is 90% of the total the variation becomes 9·3%.

The self-capacitance of the transformer is taken from the anode of the output valve to earth, and the external capacitance is usually added from the efficiency diode tap to earth. Thus the self-capacitance should be transformed to the diode tap.

Let $\frac{t}{2} = T_f$ = flyback time

$$f = \frac{1}{2\pi\sqrt{LC}}$$

$$\frac{1}{f} = 2\pi\sqrt{LC} = t$$

$$\frac{t}{2} \propto \sqrt{C}$$

$$T_f^2 \propto C$$

A 10% increase in C will result in a 5% increase in T_f.

The feedback capacitor is connected from the efficiency diode to the discharge valve, and in this circuit is 12 pF. When the discharge valve is conducting the grid to earth impedance due to grid current flowing is about 1 kΩ, and the impedance of the feedback capacitor is in the region of 400 kΩ. This is effectively adding 12 pF from efficiency diode tap to earth, which is in parallel with the tuning capacitor. In the transformer considered the total capacitance is 240 pF, thus the feedback capacitance will be 5% of the total.

It is current practice in line transformer design to employ "third harmonic" tuning, and de-tuning the transformer will result in a reduction of the e.h.t. that would have occurred under the same conditions if third harmonic tuning had not been employed.

If now three cases are considered where the added capacitance is 90%, 50%, 10% of the total capacitance, and the tolerance of the tuning capacitor, and feedback capacitor is 10%.

Case 1. When the added capacitance is 50% of the total. Taking all negative tolerances the capacitance will be reduced by 7% and the flyback time by 3·5%.

As T_f is the only variable the e.h.t. would increase by 3·5%, but a reduction will occur due to the third harmonic de-tuning. Taking a nominal flyback time of 15·5 μs, a reduction of 3·5% is 14·9 μs. The maximum voltage will occur at $\frac{T_f}{2}$ = 7·4 μs.

The self-capacitance of the e.h.t. winding is, say, 10 pF ±3%, and the C_{a-k} of the e.h.t. rectifier is 1·7 pF. A reduction of 3% winding capacitance will produce a reduction in the third harmonic frequency of 1·27%, so that the positive peak will occur at 7·6 μs.

The amplitude of the third harmonic at $\frac{T_f}{2}$ is:

$$0° \text{ to } 90° = 7{\cdot}6\ \mu s - 5{\cdot}1\ \mu s = 2{\cdot}5\ \mu s$$

Therefore

$$1° = \frac{2{\cdot}5}{90} = 0{\cdot}028\ \mu s$$

$$0° \text{ to } X° = 7{\cdot}4\ \mu s - 5{\cdot}1\ \mu s = 2{\cdot}3\ \mu s$$

Therefore

$$X^\circ = \frac{2{\cdot}3}{0{\cdot}028} = 83{\cdot}7^\circ$$

and $\sin 83{\cdot}7^\circ = 0{\cdot}994$

Thus the third harmonic content is reduced by 99·4%. Taking a third harmonic content of 1/6th of the fundamental amplitude then,

$$E = \hat{E} \sin \theta + \frac{\hat{E}}{6} \sin 3(\pi + \theta)$$

The nominal e.h.t. is taken as 16 000 volts. Thus the fundamental is 13 714 volts and third harmonic 2285 volts.

The e.h.t. is reduced by 0·094% with 99·4% of the third harmonic.

The total increase in e.h.t. will be 3·5% − 0·094% ≃ 3·4% which is 0·54 kV. As was shown when deriving the relationship between e.h.t. and deflection angle, an increase of 1·1 kV in e.h.t. would result in 3% reduction in scan. So 0·54 kV will result in 1·48% reduction in scan.

Case 2. When the added capacitance is 10% of the total, the total capacitance is reduced by 4% and so T_f by 2%. In this case, third harmonic tuning can be obtained and so the e.h.t. will be increased by 2%, which is 0·32 kV, resulting in 1·12% reduction in scan.

Case 3. When the added capacitance is 90% of the total capacitance, the total capacitance will be reduced by 13%, and T_f will increase by 6·5%. By the same approach as in Case 1 the reduction in e.h.t. due to third harmonic detuning will be 0·4%. Thus total increase in e.h.t. will be 6·1% which is 0·98 kV resulting in 2·9% reduction in scan.

10. Effect of H.T. Feed Choke and Scanning Coil Inductance

The effect of low limit inductance scanning coils and h.t. feed choke will be to reduce the flyback time and so increase e.h.t. The tolerance on the components being considered is ±5% which are readily obtainable in production, but only bottom limit components need to be considered as it is these which will reduce the scan.

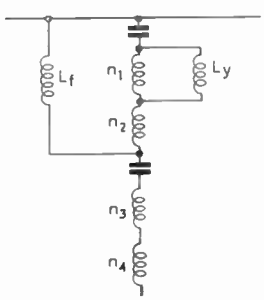


Fig. 4. Scanning coils and H.T. feed choke.

The scanning coils and h.t. choke can be transformed to the anode of the output valve and considered as three parallel inductances. The measured inductance of the line transformer is 2·8 H. Scanning coils are 5·2 mH ± 5% and the h.t. feed choke 200 mH ± 5%.

Referring to Fig. 4,

$$n_1 = 195, \quad n_2 = 45, \quad n_3 = 520, \quad n_4 = 240$$
$$N_1 = n_1 + n_2 + n_3 + n_4; \quad N_2 = n_1 + n_2$$

L_y then becomes

$$\left[\frac{N_1}{n_1}\right]^2 L_y = 136{\cdot}7 \text{ mH}$$

and L_f becomes

$$\left[\frac{N_1}{N_2}\right]^2 L_f = 3{\cdot}47 \text{ H}$$

Now $\frac{1}{L_t} = \frac{1}{L_p} + \frac{1}{L_y} + \frac{1}{L_f} = 125{\cdot}6$ mH for nominal components. When the scanning coil inductance is −5%

$$L_t = 119{\cdot}8 \text{ mH}$$

This is 4·6% reduction in inductance, thus the flyback time is reduced by 2·3% which will give an increase of 0·36 kV e.h.t. and reduce the scan by 0·98%.

Likewise if the feed choke is −5%

$$L_t = 123{\cdot}68 \text{ mH}$$

This is 1·55% reduction in inductance, thus the flyback time is reduced by 0·77%, the e.h.t. increased by 0·12 kV, and the scan reduced 0·35%.

11. Screen Feed Resistor

This should be chosen such that at the bottom limit the screen dissipation is not exceeded, and at the top limit the output valve remains operating below the knee, which will ensure minimum changes in width due to valve tolerance.†

12. Effects of Heater and H.T. Dropper Tolerance

The worst condition is when most resistance is in circuit (i.e. 250 V tap). Top limit h.t. and heater droppers were fitted (each section +5%) and the scan reduction compared with nominal resistors.

Heater dropper top limit reduced the scan by 0·5%.

H.t. dropper top limit reduced the scan by 1%.

The heater dropper consists of 5 sections each +5%, 10 Ω, 160 Ω, 66 Ω, 25 Ω. The reduction in scan

† A. Ciuciura, "Peak anode current nomograms for line output valves", *Mullard Tech. Commun.*, 3, No. 26, pp. 162–8, October 1957.

produced by each one is 0·016%, 0·105%, 0·105%, 0·016%. The h.t. dropper consists of 4 sections, each +5%, 25 Ω, 20 Ω, 5 Ω, 10 Ω. The reduction in scan produced by each one is 0·412%, 0·33%, 0·082%, 0·16%. It has been the experience in production for the h.t. line to vary ±5 volts due to the h.t. rectifier and this reduced the scan by 1%.

13. "S" Correction Capacitor

When "S" correction capacitor is on top limit less correction will be applied to the deflecting current. A top limit capacitor was measured against a nominal and the following results noted.

The scan is reduced by 1·5% and the centre of the picture cramped by 1·15%.

14. Shaping Network

The shape of the grid waveform of the line output valve will be determined mainly by the anode load of the oscillator, the grid leak, and the grid timing capacitor of the line output valve. The coupling capacitor has negligible effect at 10 kc/s. Measurements were taken and compared with nominal components.

680 kΩ − 10% width reduced 0·8% right-hand side cramped 0·48%.

100 kΩ − 10% width reduced 0·8% right-hand side cramped 0·48%.

270 pF − 10% width reduced 1·07% right-hand side cramped 0·97%.

15. Vertical Stabilizing Choke

When the choke is −5% the capacitor in series with the scanning coils will not be so effective in making the scanning current "S" shaped and will result in a reduction in scan.

Measurements were taken on a set and the scan was reduced by 0·26%.

16. Scanning Coils

The tolerance of inductance has already been taken into account when considering its effect on flyback time. A bottom limit inductance scanning coil was obtained with nominal resistance. Resistance was added in series with the coil until top limit resistance was obtained. The scan was reduced by 0·4%, the linearity not being affected.

A third tolerance on the scanning coils which has not yet been considered is the deflection sensitivity. This is ±3% and will result in ±3% variation in the scan.

17. Non-Linearity due to Flyback Time

The picture can only be linear for one particular flyback time. Modifications of the flyback time will result in shift of the modulation and hence shift of the deflection centre. A shorted-turn linearity sleeve is arranged to take care of this and other variations in linearity, but the variations in linearity for one setting of the linearity sleeve only will be considered.

If the picture is arranged to be linear with a nominal flyback time of 15·5 μs, the variation due to capacitance, scanning coils and h.t. feed choke will be 6·57%, which is 1·02 μs. Measured sync. delay was 0·857 μs.

With nominal flyback time

$$D_f = \frac{1}{2}\frac{L}{T}\{(l-a)-2(b+c)\}$$

$$= \frac{1}{2}\cdot\frac{374{\cdot}5}{80}\{(18-15{\cdot}5)-2(1{\cdot}5+0{\cdot}857)\}$$

$$= \frac{1}{2}\cdot\frac{374{\cdot}5}{80}(-2{\cdot}214)$$

$$= 5{\cdot}183 \text{ mm required shift to the right.}$$

With fastest flyback time

$$D_f = \frac{1}{2}\cdot\frac{374{\cdot}5}{80}\{(18-14{\cdot}48)-(1{\cdot}5+0{\cdot}857)2\}$$

$$= \frac{1}{2}\cdot\frac{374{\cdot}5}{80}(-1{\cdot}19)$$

$$= 2{\cdot}79 \text{ mm required shift to the right.}$$

It is stated that nominally a linear picture was obtained with 5·183 mm shift to the right. With the faster flyback time this is the same as a shift of 3·38 mm to the left.

In Section 2.3 it was shown that 5·05 mm of shift caused 1·69% non-linearity; thus 5·38 mm shift will cause 1·1% non-linearity, but this will cause the right-hand side to open out, whereas the other things affecting linearity will cause the left-hand side to open out. When the total non-linearity is calculated this must be added as a negative quantity.

The tolerance of the components effecting flyback time should, of course, be added statistically in order to obtain a practical figure. Thus the flyback time will vary 4·18% which is 0·65 μs.

Thus

$$D_f = \frac{1}{2}\cdot\frac{374{\cdot}5}{80}\{(18-14{\cdot}85)-(1{\cdot}5+0{\cdot}857)2\}$$

$$= \frac{1}{2}\cdot\frac{374{\cdot}5}{80}(-1{\cdot}564)$$

$$= 3{\cdot}66 \text{ mm shift to the right.}$$

This is the same as 1·56 mm shift to the left, and will cause a non-linearity of 0·45% to be added as a negative quantity when total non-linearity is calculated.

18. Tolerance of Components Affecting Width

Component	*Reduction in Scan*
C.R.T. + 2°	3%
Scan coil inductance − 5%	0·98%
Scan coil resistance + 6%	0·4%
Scan coil deflection sensitivity − 3%	3%
"S" correction capacitor is + 10%	1·5%
H.T. feed choke inductance − 5%	0·35%
The line transformer considered has 50% of the total tuning capacitance added	
(i) Self-capacitance − 3% (ii) Added capacitance − 10% (iii) Feedback capacitance − 10%	1·48%
Vertical stabilizing choke inductance − 5%	0·26%
Line transformer scan − 3%	3%
Shaping network components − 10%	2·67%
H.T. rectifier is − 5 volts	1%
H.T. dropper is + 5%	1%
Heater dropper is + 5%	0·5%
	19·2%

In order that the correct aspect ratio may be obtained with nominal components the line must be overscanned by 5%. Thus the total overscan required for the tolerance of these components is 24·2%. This is only theoretically correct because the receiver is designed for a practical minimum, rather than zero rejects. Therefore the root mean square of the tolerances should be taken. The reduction in width is then 6·5%, and adding to this the 5% overscan necessary to give the correct aspect ratio, the overscan will be approximately 11%.

19. Items Affecting Linearity

Component	*Degree of non-linearity*
Line transformer	1·5%
Shaping network	
(i) Anode load of oscillator	0·48%
(ii) Grid leak of line output valve	0·48%
(iii) Grid timing capacitor	0·97%
C.R.T. spot "non-centrality"	3·20%
Earth's magnetic field	0·87%
	7·5%

When these are added statistically the non-linearity is 3·8% but 0·45% must be subtracted, because of the shift due to the reduction in flyback time, resulting in a non-linearity of 3·4%.

20. Non-linearity due to Picture Shift, Originated by Non-linearity of the Picture

The shift due to a non-linearity of 3·4% will be

$$D_s = \frac{374{\cdot}5}{600} \times 3{\cdot}4\%$$

$$= 2{\cdot}1 \text{ mm}$$

A shift of 2·1 mm will cause a non-linearity of 0·7%.

Thus the total non-linearity will be 4·5%.

21. Conclusion and Acknowledgments

In conclusion it may be said that if a linear picture is obtained with nominal components and an overscan of 11%, then a minimum of rejects due to lack of width and non-linearity will be experienced.

The author wishes to express his thanks to The Plessey Company Ltd., for permission to publish this paper, and to his colleagues for helpful discussions and assistance in this investigation.

22. Appendix 1
Picture Shift due to Non-linearity

It is assumed for these calculations that the non-linearity is uniform along the scan. (This is a reasonable assumption as if the "S" correction is chosen properly then d will be on the left and e on the right of the scan.)

$$\frac{di}{dt} = e + \frac{t}{T}(d-e)$$

where t is the time at any instant.

Let deflection sensitivity = K amps/mm

Then $$\frac{di}{dt} = K\frac{ds}{dt} = K\left\{e + \frac{t}{T}(d-e)\right\}$$

Therefore $$i = K\left\{et + \frac{t^2}{2T}(d-e)\right\} + q \qquad \text{.....(3)}$$

As $$\frac{di}{dt} = K\frac{ds}{dt}$$

then $$\int_0^T i\,dt = 0$$

Therefore

$$\int_0^T K\left\{et + \frac{t^2}{2T}(d-e)\right\} + q\,dt = 0$$

$$= K\left\{\frac{et^2}{2} + \frac{t^3}{6T}(d-e)\right\} + qt$$

$$= K\left\{\frac{eT^2}{2} + \frac{T^3}{6T}(d-e)\right\} + qT$$

$$= K\left\{\frac{eT}{2} + \frac{T^2}{6T}(d-e)\right\} + q$$

If $$T \neq 0$$

then $$K\left\{\frac{eT}{2} + \frac{T}{6}(d-e)\right\} + q = 0$$

Therefore $$q = -\left\{\frac{eT}{2} + \frac{T}{6}(d-e)\right\}$$

Substituting this in eqn. (3)

$$i = K\left[\left\{et + \frac{t^2}{2T}(d-e)\right\} - \left\{\frac{eT}{2} + \frac{T}{6}(d-e)\right\}\right]$$

At $t = 0$

let $i = I'$

$$I' = -K\left\{\frac{eT}{2} + \frac{T}{6}(d-e)\right\}$$

At $t = T$

let $i = I''$

$$I'' = K\left\{eT + \frac{T}{2}(d-e) - \frac{eT}{2} - \frac{T}{6}(d-e)\right\}$$

$$= K\left\{\frac{eT}{2} + \frac{T}{3}(d-e)\right\}$$

The current at the centre of the tube will be

$$\frac{I'+I''}{2} = \frac{K}{2}\left\{\frac{eT}{2} + \frac{T}{3}(d-e) - \frac{eT}{2} - \frac{T}{6}(d-e)\right\}$$

$$i_c = \frac{K}{2}\left(\frac{T}{6}(d-e)\right)$$

As $i_c = KS_c$

$$Ds = \frac{T}{12}(d-e) \text{ required shift to the right in mm} \qquad \text{.....(4)}$$

when $t = T, \quad S = L$

$$I''-I' = K\left\{eT + \frac{T}{2}(d-e)\right\}$$

$$= K\left\{\frac{T}{2}(d+e)\right\}$$

and

$$\frac{I''-I'}{K} = S = L = \frac{T}{2}(d+e)$$

The physical centre of the screen is at

$$\tfrac{1}{2}L = \frac{T}{4}(d+e) \qquad \text{.....(5)}$$

Substituting eqn. (5) in (1)

$$P = \frac{(d-e)}{2L}T \times 100\% \qquad \text{.....(6)}$$

Substituting eqn. (6) in (4)

$$Ds = \frac{L}{600} \cdot P \text{ mm shift to the right} \qquad \text{.....(7)}$$

23. Appendix 2

Deflection Angle and Corresponding Scan

In order to determine the screen deflection for increasing deflection angles it is first necessary to ascertain the apparent centre of deflection. It will be shown that when the deflection θ is zero then the apparent centre of deflection will be in the mid-position of the deflecting field, and as θ increases to the theoretical 90 deg the apparent centre of deflection will move towards the screen by a distance which is half the length of the deflecting field.

Let L be the length of the deflecting field.

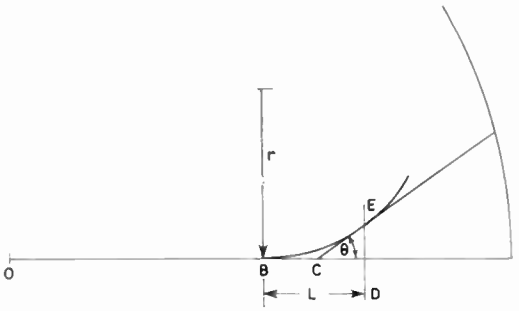


Fig. 5. Geometry of deflecting field.

From Fig. 5

$$\text{BD} = \text{DC} + \text{BC}$$

$$\text{and CD} = \text{EC} \cos\theta$$

$$\text{Now CE} = \text{BC}$$

$$\text{Therefore BD} = \text{BC} + \text{BC} \cos\theta$$

$$\text{So BC} = \frac{\text{BD}}{1+\cos\theta}$$

When θ is zero the apparent centre of deflection is $\frac{\text{BD}}{2}$.

When θ is 90 deg the apparent centre of deflection will be from point D.

A maximum deflection of 110 deg may be obtained across the diagonal of the tube. Thus two dimensions may be obtained from the tube data, the radius of curvature of the screen and the diagonal. From the radius and half the diagonal a third dimension may be obtained. (Fig. 6.)

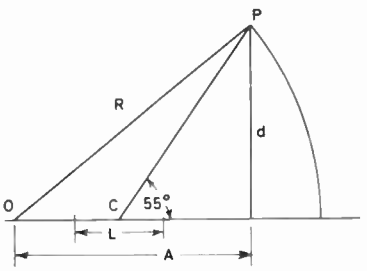


Fig. 6.

R is the radius of curvature; D is the diagonal, and $\frac{D}{2} = d$. Thus $A = \sqrt{R^2-d^2}$

The apparent centre of deflection from the radius of curvature when $\theta = 55°$ will be $A - \frac{d}{\tan 55°}$ (Fig. 7).

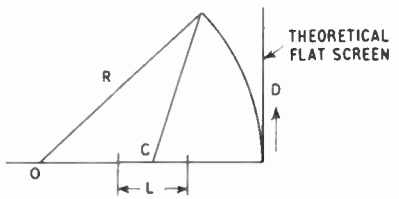


Fig. 7.

The apparent centre of deflection in the deflecting field $= \frac{L}{1+\cos\theta}$.

and $$PC = PQ - CQ \qquad \text{.....(10)}$$

$$\cos\theta = \frac{QC}{OC}$$

Thus $$OC\cos\theta = QC \qquad \text{.....(11)}$$

Substituting (10) in (9)

$$PC = PQ - OC\cos\theta$$

Equation (8) then becomes:

$$D = (PQ - OC\cos\theta)\sin\theta$$

and $$PQ = \sqrt{R^2 - (OQ)^2}$$

But $OQ = OC\sin\theta$

Thus $D = [\sqrt{R^2 - (OC\sin\theta)^2} - OC\cos\theta]\sin\theta$

Substituting for OC the apparent centre of deflection which was calculated earlier.

$$D = \left[\sqrt{R^2 - \left(A - \frac{d}{\tan 55°} - \frac{L}{1+\cos 55°} + \frac{L}{1+\cos\theta}\right)^2 \sin^2\theta} - \left(A - \frac{d}{\tan 55°} - \frac{L}{1+\cos 55°} + \frac{L}{1+\cos\theta}\right)\cos\theta\right]\sin\theta \qquad \text{.....(12)}$$

The apparent centre of deflection for any angle will then be

$$OC = A - \frac{d}{\tan 55°} - \frac{L}{1+\cos 55°} + \frac{L}{1+\cos\theta}$$

Considering first the theoretically flat screen. The distance of the apparent centre of deflection from the screen is:—

$$R - \left[A - \frac{d}{\tan 55°} - \frac{L}{1+\cos 55°} + \frac{L}{1+\cos\theta}\right]$$

Thus the screen deflection D is

$$D = \left[R - \left(A - \frac{d}{\tan 55°} - \frac{L}{1+\cos 55°} + \frac{L}{1+\cos\theta}\right)\right]\tan\theta \qquad \text{.....(8)}$$

The screen is however curved and the true screen deflection can be calculated with reference to Fig. 8.

$$D = PC\sin\theta \qquad \text{.....(9)}$$

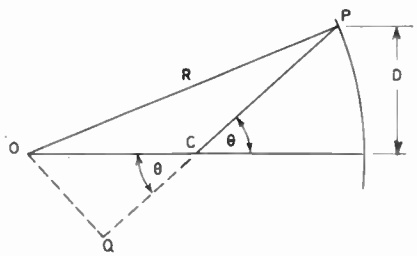


Fig. 8.

Substituting dimensions for AW43–88:

$R = 508$ mm

$d = 200$ mm

Calculated value of $A = 466{\cdot}92$ mm

Taking a value for L as 50 mm.

When $\theta = 2°$, $D = 6{\cdot}554$ mm

When $\theta = 4°$, $D = 13{\cdot}143$ mm

Thus D from 2° to 4° $= 13{\cdot}143 - 6{\cdot}554 = 6{\cdot}589$ mm

When $\theta = 53°$; $D = 190{\cdot}99$ mm

When $\theta = 55°$; $D = 200{\cdot}03$ mm

When $\theta = 57°$; $D = 208{\cdot}98$ mm

So D from 53° to 55° $= 200{\cdot}03 - 190{\cdot}99 = 9{\cdot}04$ mm

and D from 55° to 57° $= 208{\cdot}98 - 200{\cdot}03 = 8{\cdot}95$ mm

The ratio of D for largest angles to smallest angles of deflection is $\frac{9}{6{\cdot}5} = 1{\cdot}38 \simeq 1{\cdot}4$.

Thus it can be seen from these calculations that a top limit tube having an increase in angle of 2 deg would require an increase in scan of approximately 3%.

Manuscript first received by the Institution on 17th May 1961 and in revised form on 20th March 1962. (Paper No. 792/T19).

Recording the Effects of Weather on a Welsh Lake

An unusual and interesting scientific investigation, which will last for seven years, is being undertaken by the Research and Development Department of the North Western Region of the Central Electricity Generating Board at the site of Trawsfynydd Nuclear Power Station, in North Wales.

The Power Station is being built on the banks of a large lake and an opportunity is provided to study the effect of meteorological influences upon the cooling of warm water discharged from Power Stations. For this purpose a standard, transistorized Elliott 250-point Data Logger is being supplied, which, for the next seven years, will automatically collect and collate information from 250 electrical transducers situated in and around the lake and will record it on punched paper tape. The information will consist of wind speeds and direction, air and water temperatures, solar radiation, the hours of sunshine, the level of the lake, rainfall, pressure and flow measurements and many other such factors.

The resulting records will be analysed at regular intervals by an electronic digital computer, so that there will be built up an extremely accurate picture of the natural effect of the weather upon the conditions in the lake. When the Power Station goes into operation it will then be possible to record what changes, if any, are caused by the discharge of cooling water into the lake. In addition to the perforated paper tape record, a duplicate record will be printed at preset intervals by the automatic typewriter in the Data Logger. This printed record can also be obtained whenever required by merely pressing a button.

The Central Electricity Generating Board makes wide use of data logging equipment, particularly in monitoring power generating station operation. One of the papers to be presented at the 1963 Convention—"Data Logging in Power Generating Stations" by W. E. Willison—will describe an installation of this type. A synopsis of Mr. Willison's paper appears on page 126 of this issue.

Bathythermograph Slug for Measurement of Ocean Temperatures

A device for determining temperature variations at different depths of water has been developed and produced by EMI-Cossor Electronics Ltd, of Dartmouth, Nova Scotia. The bathythermograph slug, as the device is known, was originally designed for helping to locate enemy submarines, but it is also of considerable value in gathering oceanographic data.

The conventional method of employing ships to obtain temperature readings is both slow and expensive. By dropping bathythermograph slugs from aircraft, temperature readings over large areas of ocean can quickly be taken. The bathythermograph package is dropped from the aircraft and, upon hitting the water, the resultant impact releases the slug which, after a delay of one minute, sinks through the water at five feet per second. A temperature-sensitive device within the slug detects changes in sea-water temperature and transmits an acoustic signal.

A standard sonobuoy, which may have been dropped for the purpose or is already floating on the water detects the signal which modulates the radio transmitter of the sonobuoy. After demodulation by the aircraft's receiver, the signal is applied to a translator unit. The output from this unit drives a pen recorder which provides a graph of temperature versus depth. The depth factor is introduced as elapsed time from the release of the slug.

The temperature measuring range of the slug is from $-3{\cdot}9^{\circ}$ C to 35° C and its accuracy is between 1 and 2 deg C. The maximum working depth is 305 m and its sink rate is 1·52 m per second—the depth is known to an accuracy of about 3·5 m at maximum depth. The acoustic power output from the slug, which weighs only 3·2 lb (1·45 kg) is about 100 mW.

Power for the electronic circuit is provided by a sea-water-activated battery which remains inert until it makes contact with the water when the slug is released. Use of this type of battery enables the slug to be stored for several years before use.

The translator unit in the aircraft operates from a 28 V d.c. supply. Transistor circuits minimize the effects of sea noise and multipath interference on the recorded trace. Self-calibration facilities are provided.

One great advantage of using the slug for submarine detection is that most of the methods currently employed are based on measurements of primary or reflected sound energy from the enemy vessel, and such data must be modified by compensation for temperature variations at different depths which influence the propagation speed and path of the sound waves. The bathythermograph slug provides a reliable means of obtaining information regarding these temperature variations, so enabling the target's position to be accurately determined.

Variable Depth Sonar Equipment

A new device known as Variable Depth Sonar has been developed and produced for the Royal Canadian Navy by EMI-Cossor Electronics Ltd, of Dartmouth, Nova Scotia, Canada and is also to be used by the Royal Navy. This consists of a sonar transducer—transmitter and receiver—towed astern of a ship. By varying the length of tow, the depth of the transducer can be controlled, and it is possible to lower it beneath any temperature layers below which may be lurking submarines which are undetectable by normal sonar.

The problem of temperature layers, which refract and reflect the normal sonar beam, has long been an acute one, and many a submarine escaped during the war by hiding beneath these layers. This new device will make the submarine's task a much more difficult one.

One of the papers presented at last year's Symposium on Sonar Systems—"Absorption of sound in sea-water" by M. Schulkin and H. W. Marsh—discussed in some detail the basic physical problems associated with sound propagation under different conditions. Other papers studied theoretically the modes of sound wave propagation in stratified layers of water having varying refractive indexes (e.g. "The use of ray tracing in the study of underwater acoustic propagation" by M. J. Daintith).

U.D.C. 621.375.132

A Twin-T Amplifier with Stable Gain and Variable Selectivity

By

H. C. BERTOYA,

(*Associate Member*) †

Summary: The principles of operation of low-frequency "twin-T" selective amplifiers are discussed and information is given regarding the design of such amplifiers in order to obtain a stable and predictable performance. The paper opens with the analysis of a simple system and its shortcomings. The twin-T network and the operation of its associated amplifier are then considered in detail. The complete system is analysed and design parameters are given for optimum performance. Details are given of the construction and performance of both valve and transistor amplifiers which were designed according to the principles outlined.

1. Introduction

Many measuring systems operate with a low-frequency carrier which is modulated in amplitude in accordance with the magnitude of the input signal. In such cases it is often desirable to employ a filter which will transmit the carrier but reject harmonics of it, as well as any induced hum or other spurious signals. One such filter is the well-known "twin-T" selective amplifier of which the principle of operation is recapitulated below.

2. Principle of Operation

If we plot $|\beta|$ (that is, the ratio of E_{ac}/E_{ab} for a twin-T filter) versus radian frequency (ω) we obtain the curve shown in Fig. 1. Note that at one radian frequency (ω_∞), $\beta = 0$.

Now let us suppose that the network is connected in the feedback loop of an amplifier as in Fig. 2. The equations for such a system are $E_{ec} = A_{bc}E_{eb}$ (where A_{bc} is the gain of the amplifier), $E_{eb} = E_{ea} - E_{ef}$, and $E_{ef} = \beta E_{ec}$, where $\beta = f(\omega)$. Solving for E_{ec} in terms of E_{ea} gives the overall gain of the system A_{ac}. This is

$$A_{ac} = E_{ec}/E_{ea} = \frac{1}{\frac{1}{A_{bc}} + \beta} \qquad \ldots\ldots(1)$$

The characteristic of such a system is shown in Fig. 3.

From eqn. (1) it will be seen that when

$$\beta = 0 \text{ (at } \omega_\infty) \text{ then } A_{ac} = A_{bc},$$

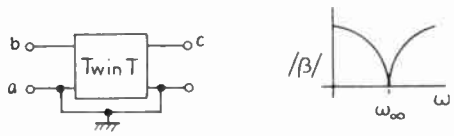
and when

$$\beta \gg \frac{1}{A_{bc}} \text{ then } A_{ac} \simeq \frac{1}{\beta}.$$

Suppose we now define a variable S, the ratio of the gain at ω_∞ to the gain at any other radian frequency. If A'_{ac} is the value of A_{ac} when $\beta = 0$, then

$$S = \frac{A'_{ac}}{A_{ac}} = 1 + \beta A_{bc} \qquad \ldots\ldots(2)$$

Thus $S = f(\beta, A_{bc})$. The selectivity of the system for a given β depends on the value of A_{bc}.



$$\beta = \frac{E_{ac}}{E_{ab}} \qquad \omega_\infty = \text{frequency at which } |\beta| = 0$$

Fig. 1. Characteristic of twin-T circuit.

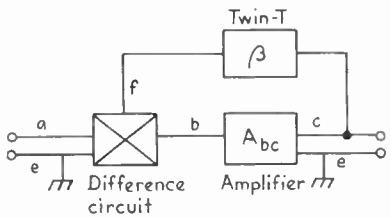


Fig. 2. The twin-T selective amplifier.

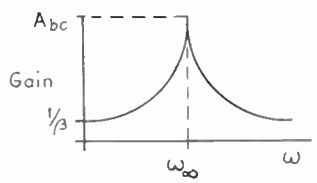


Fig. 3. Characteristic of selective amplifier.

3. Limitations of the Simple Amplifier

Simple twin-T selective amplifiers normally employ a single active element for the amplifier, either a valve or transistor. When $\beta = 0$, therefore, the gain

† British Scientific Instrument Research Association, Chislehurst, Kent.

is that of a simple element without feedback. Because of this the gain may vary considerably with changes of supply voltage, temperature etc., and—as has been shown—such changes will also alter the selectivity. Furthermore, large changes in gain and selectivity may be obtained with identical circuits due to valve or transistor tolerances—especially the latter. A batch of four simple transistor amplifiers[1] constructed in the laboratory gave gains varying between 30 and 90, with the selectivity varying between wide limits. Further investigation showed that some positive feedback was occurring in these amplifiers. It was therefore decided to investigate the mode of operation of such amplifiers and to attempt a design which could be generally applicable in the laboratory. The amplifiers would be required to operate in experimental closed and open loop systems with various carrier frequencies, and the design was undertaken to the following specification:

(i) Stable and predictable gain,

(ii) Stable and predictable selectivity,

(iii) Minimum phase shift,

(iv) Selectivity to be variable without alteration of gain,

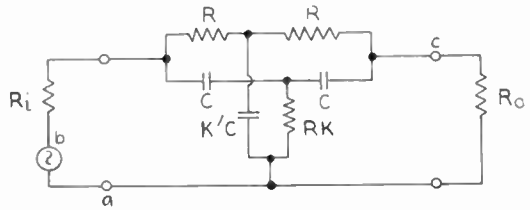
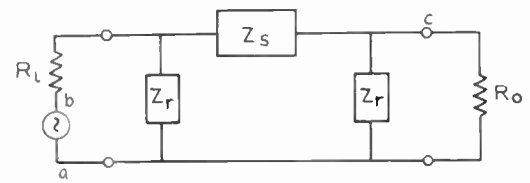
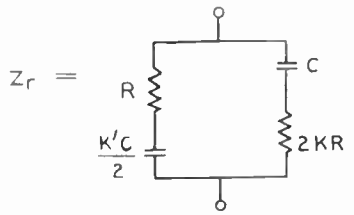


Fig. 4. Twin-T network.





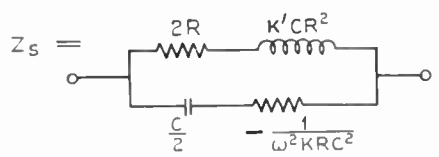


Fig. 5. Equivalent π network of Fig. 4.

(v) Performance to be repeatable from model to model,

(vi) Frequency of operation to be determined (over as wide a range as possible) merely by the insertion of the correct twin-T network into a standard amplifier.

4. The Twin-T Network: Theoretical Considerations

Before considering the complete system we must consider the twin-T network itself. Such a network is shown in Fig. 4. The transmission ratio of this network is defined by the relationship $\beta = E_{ac}/E_{ab}$. The equivalent π-network is given in Fig. 5.

From Fig. 5 it can be shown[2] that

$$\beta = \frac{1}{1 + \frac{Z_s}{Z_p}\left(1 + \frac{R_i}{R_o}\right) + \frac{R_i}{Z_p}\left(2 + \frac{Z_s}{Z_p}\right) + \frac{R_i}{R_o} + \frac{Z_s}{R_o}} \quad \ldots\ldots(3)$$

where $Z_p = R_p + jX_p$ and $Z_s = R_s + jX_s$. R_p, X_p, R_s and X_s are all functions of frequency and of the parameters R, k and k'. They may be evaluated from the circuits of Fig. 5. The evaluation is extremely tedious and it is preferable to fix certain parameters[2] for optimum performance and then to observe the variation of β with frequency.

The parameters are fixed by the requirement that the network should have zero transmission ($\beta = 0$) at the null radian frequency ($\omega = \omega_\infty$), the highest value of $d\beta/d\omega$ around the null point, that is, the highest selectivity, and minimum phase shift.

4.1. *Optimum Design Parameters*[2-6]

The manner in which the requirements are satisfied is detailed below. In the graphs which follow, β is shown as a function of the ratio $\omega/\omega_\infty = n$.

4.1.1. Condition for zero output

This is achieved by making $k' = 4k$. If $k' \neq 4k$, the output will fall to a minimum at one frequency, but not to zero. This is shown in Fig. 6.

4.1.2. Condition for maximum $d\beta/d\omega$

The value of $d\beta/d\omega$ is affected by the value of k, the relationship of R, R_i and R_o, and the ratio R_o/R_i. It can be shown that for $R_i/R \to 0$, optimum k lies between 0 and 0·5. For $R_o/R \to \infty$, k lies between 0·5 and 1·0. A convenient value of k is, therefore, 0·5. If k is fixed then the steepest gradient is achieved by putting $R^2 = \left(\frac{2k+1}{2k}\right) R_o R_i$. Figure 7 shows the variation of gradient with respect to R^2. In addition the gradient is affected by the ratio R_o/R_i. With $k = 0{\cdot}5$ and $R^2 = 2R_oR_i$, the gradient at the null point for $R_o = 50R_i$ is approximately 2·5 times the gradient for $R_o = R_i$.

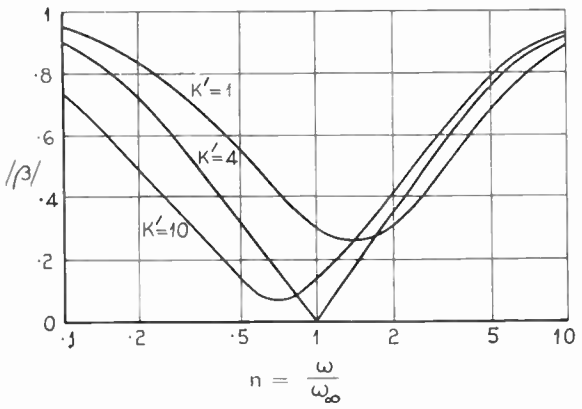


Fig. 6. Variation of minimum value of $|\beta|$ with variation of k'. $R_t = 0$, $R_o = 100\,R$, $k = 1$.

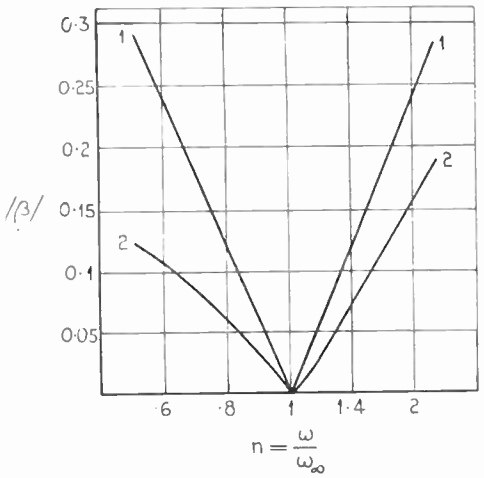


Fig. 7. Transmission ratio $|\beta|$ showing effect of variation of R^2. $k' = 4k$, $k = 0{\cdot}5$, $R_o/R_t = 50$.
Curves: (1) $R^2 = R_tR_o$. (2) $R^2 = 200\,R_tR_o$.

4.1.3. Condition for minimum phase shift

For a network in which $k' = 4k$ the phase will shift by 180 deg around the null point. If

$$R^2 = \frac{(2k+1)R_oR_i}{2k}$$

then the phase change will be symmetrical about zero i.e. ± 90 deg. If $R^2 \neq \frac{(2k+1)R_oR_i}{2k}$ then the phase will still change by 180 deg but not symmetrically, e.g. from $-60°$ to $+120°$. (See Fig. 8.) If $k' \neq 4k$ larger shifts can take place; with $k' = 10k$ for example, the phase will change by 360 deg around the null. (See Fig. 9.) The effects of phase are important in two respects. Firstly, there is the consideration of stability of the system and, secondly, there is the requirement of minimum phase change through the system.

4.1.4. Summary of design parameters

The desirable parameters for the twin-T network may thus be summarized by stating the relationships $\omega_\infty = 1/RC$, $k' = 4k$, $k = 0{\cdot}5$, $R^2 = 2R_oR_i$, and R_o/R_i as high as possible consistent with the limitations imposed by the external circuits. When these requirements are satisfied the network of Fig. 10 is obtained.

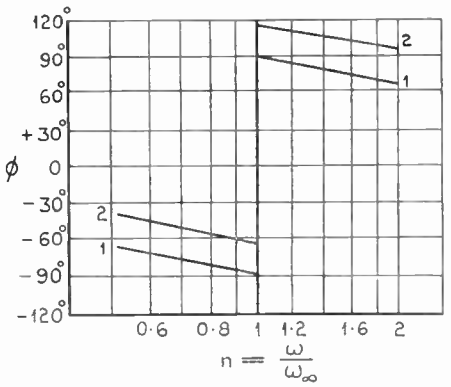


Fig. 8. Variation of phase angle with frequency showing effect of R^2
Curves: (1) $R^2 = 2R_tR_o$. (2) $R^2 = 200R_tR_o$, $k = 0{\cdot}5$, $R_o/R_t = 50$, $k' = 4k$.

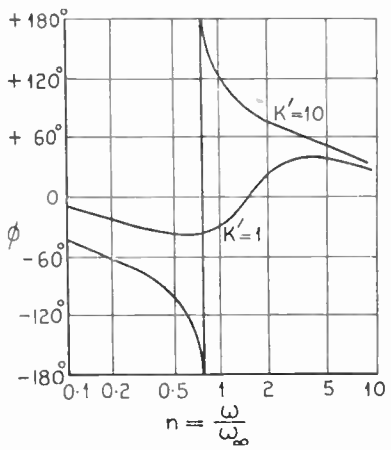


Fig. 9. Variation of phase angle with frequency showing effect of k'. $R_t = 0$, $R_o = 100\,R$, $k = 1$.

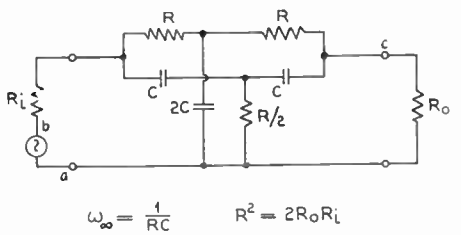


Fig. 10. Twin-T network for optimum performance.

Therefore $$S = 1 + \frac{A_{bc}\beta\gamma\mu_1}{\mu_4} \qquad \text{.....(13)}$$

Now, since β is complex, S will be complex. Thus put $\beta = a+jb$, then

$$|S| = \left[\left(1 + \frac{A_{bc}\gamma\mu_1 a}{\mu_4}\right)^2 + \left(\frac{A_{bc}\gamma\mu_1 b}{\mu_4}\right)^2\right]^{\frac{1}{2}} \qquad \text{.....(14)}$$

If, for the given twin-T, the values of a and b are known, this last equation enables the ratio of the gain at ω_∞ to the gain at any other frequency ω to be calculated. Considering eqn. (13) we have that

$$S = \mathrm{f}(A_{bc},\ \gamma,\ \mu_1,\ \mu_4)$$

Generally A_{bc}, μ_1, μ_4 will be fixed.

If, however, γ is varied it will be possible to alter the circuit selectivity without affecting the gain at ω_∞, since A'_{ac} is a function only of A_{bc} and μ_4.

5.3. *Phase Shift through the Complete System*

From eqn. (11) we have that

$$E_{ec} = \frac{E_{ea}}{\mu_4/A_{bc} + a\gamma\mu_1 + \mathrm{j}b\gamma\mu_1};$$

therefore $$E_{ec} = \frac{E_{ea} \cdot \mathrm{e}^{\mathrm{j}\theta}}{G},$$

where $$G = [(\mu_4/A_{bc} + a\gamma\mu_1)^2 + (b\gamma\mu_1)^2]^{\frac{1}{2}} \qquad \text{.....(15)}$$

$$\theta = -\tan^{-1}\left(\frac{b\gamma\mu_1}{\mu_4/A_{bc} + a\gamma\mu_1}\right) \qquad \text{.....(16)}$$

6. The Twin-T Network: Practical Considerations

The network must be constructed carefully and high-quality components should be used. After calculation of R and C the following procedure was adopted.

The two series capacitors were selected by bridge measurements to give identical values. The capacitors were then placed in parallel and the bridge reading noted. The shunt capacitor was selected to give the same reading. The two series resistors were selected and compared on a bridge. The circuit was then constructed with a value of $R/2$ less than the value required but with the provision for inserting a small padding resistor in series. The ratio $|\beta|$ was then measured. The padding resistor was then varied to give the best minimum at ω_∞. It was also necessary on occasions to insert small capacitors of value ΔC, $2\Delta C$ across the series and shunt capacitors. By this means it was possible without undue difficulty to obtain a minimum β of less than 0·001 at a frequency within 0·5% of ω_∞.

Notice must be taken in this measurement of the harmonics present in the generator E_{ab} if a valve voltmeter is used to measure E_{ac}; harmonics of ω_∞ will suffer less attenuation in the network and will still be indicated by the voltmeter. It may be necessary to measure E_{ac} with a device such as a wave analyser if a very low value of $|\beta|$ is obtained at the null frequency. If phase measuring equipment is not available it is still possible to ascertain whether the phase shift is symmetrical. Reference to Fig. 7 shows that when $R^2 = 2R_i R_o$ the transmission $|\beta|$ is symmetrical about ω_∞. This is easily checked by altering the input generator frequency about ω_∞ and observing that the output is the same at $\omega/\omega_\infty = n$ and $1/n$.

6.1. *Performance of a Typical Example*

As an indication of the accuracy with which a twin-T network can be constructed, such a network was built for use with the valve amplifier. The design

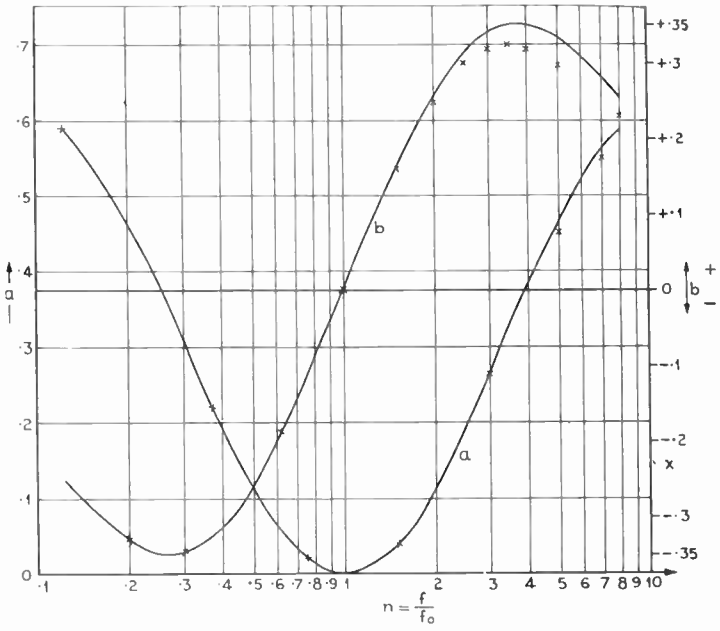


Fig. 16. Plot of $a + jb$ for ideal and practical twin-T filter.
× = measured points of practical filter.

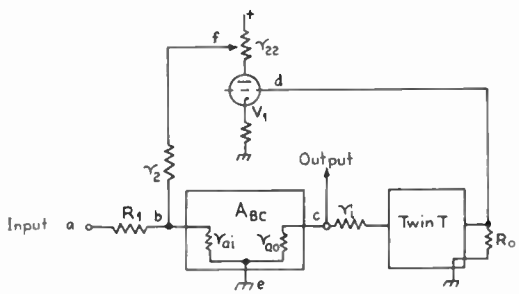


Fig. 17. Schematic diagram of practical amplifier.

parameters were: $\omega_\infty = 1/RC = 2\pi \times 400$ c/s; $R_o/R_i = 50$; $R_i = 9{\cdot}7$ kΩ; $R_o = 485$ kΩ; $R = 97$ kΩ; and $C = 0{\cdot}0041$ μF.

Figure 16 shows the response $\beta = a+jb$ for the actual circuit superimposed on the theoretical curves obtained from eqn. (8); ω_∞ occurs at 401·5 c/s at which point $|\beta| = 0{\cdot}00065$. The response of the network was measured with a Solartron Resolved Component Indicator which has a stated accuracy of ∓2% of f.s.d. The generator was a Wayne Kerr Type 5121 which has a frequency accuracy of ∓1%.

7. The Complete Amplifier: Practical Considerations

Two types of amplifier were constructed, one employing valves and the other transistors.

7.1. *The Valve Amplifier*

The complete circuit is shown in Fig. 18. The salient design features are explained with reference to the schematic diagram of Fig. 17.

(i) The amplifier A_{bc} is a conventional two-stage feedback amplifier. The input impedance is about 10 MΩ and the output impedance is 1·62 kΩ.

(ii) The difference network consists of R_1, $R_2 = r_2+r_{22}$, $R_3 = r_{ai}$ where r_{ai} is the input impedance of the amplifier. Since valve V1 is operating with current feedback the effective r_a is high so that the output impedance is virtually the anode load r_{22}.

(iii) The function of valve V1 is threefold; it enables the twin-T to work into a high impedance, it transmits the feedback voltage in the correct phase, and it presents a known low output impedance (r_{22}).

The setting of the potentiometer r_{22} may be varied in order to alter γ and hence the selectivity. The stage must be capable of handling a voltage of magnitude $E_{ec} \times \beta_{max}$, or approximately E_{ec}.

(iv) The twin-T has the constants $\omega_\infty = 2\pi \times 401{\cdot}5$ c/s, $R_o/R_i = 50$, $R^2 = 2R_iR_o$, $R_i = r_i+r_{ao}$, where r_{ao} is the output impedance of the amplifier. The resistor r_i was inserted so that it could be varied to compensate for small changes in r_{ao} from model to model and also to reduce the effect of external loads across the amplifier output.

7.1.1. System constants and calculated performance for the valve amplifier

The various system constants were $R_1 = R_2 = 100$ kΩ, $R_3 = 10$ MΩ, $\mu_1 = 1$, $\mu_3 = 0{\cdot}01$, $\mu_4 = 2{\cdot}01$.

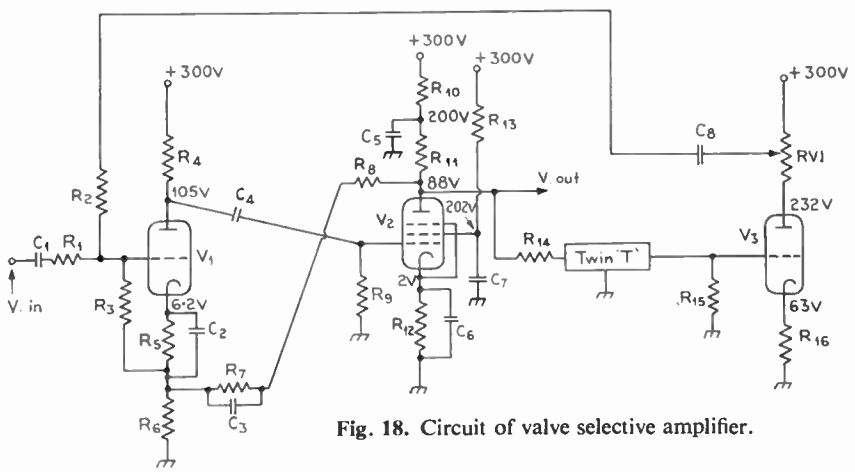


Fig. 18. Circuit of valve selective amplifier.

R1 —100 kΩ (H.S. ½ W, 1%)
R2 —90 kΩ (H.S. ½ W, 1%)
R3 —1 MΩ
R4, R9—100 kΩ
R5 —680Ω
R6 —2·2 kΩ (H.S. ½ W, 1%)
R7 —320 kΩ (H.S. ½ W, 1%)
R8 —12 kΩ (H.S. ½ W, 1%)
R10 —18 kΩ (1 W, 10%)
R11 —22 kΩ (1 W, 10%)
R12 —330Ω
R13 —68 kΩ
R14 —8·2 kΩ
R15 —470 kΩ (H.S. ½ W, 1%)
R16 —8·2 kΩ (1 W, 10%)
RV1 —10 kΩ w.w. lin. pot. 3 W
All resistors ½ W 10% unless otherwise stated.

C1, C8—1·0 μF 350 V
C2, C6—25 μF 25 V electrolytic
C3 —47 pF
C4 —0·1 μF 350 V
C5 —32 μF 350 V electrolytic
C7 —16 μF 350 V electrolytic
V1, V3—ECC 81
V2 —EF 91

Table 1

Values of $\beta = a + jb$, $|\beta|$ and $\phi = \arg \beta$ for $m = 0{\cdot}2$

n	a	b	$\lvert\beta\rvert$	ϕ degrees
0	0·70422	0	0·70422	
·125	·59157	— ·25816	·6454	— 23·58
·25	·38276	— ·35078	·5192	— 42·50
·375	·21676	— ·32506	·3907	— 56·30
·5	·11269	— ·25819	·2817	— 66·42
·625	·05246	— ·18491	·1922	— 74·16
·75	·01972	— ·11619	·1178	— 80·37
·875	·00425	— ·05456	·0547	— 85·55
1	0	0	0	0
1·25	0·01187	0·09066	0·0914	82·54
1·5	·03911	·16128	·1656	76·37
2·0	·11269	·25819	·2817	66·42
2·5	·19147	·31333	·3672	58·56
3	·26466	·34108	·4317	52·19
3·5	·32860	·35133	·4810	46·91
4	·38276	·35078	·5192	42·50
5	·46557	·33333	·5726	35·60
6	·52278	·30799	·6067	30·50
7	·56285	·28209	·6296	26·62
8	·59157	·25816	·6454	23·58

With the potentiometer at maximum

$$\gamma = E_{ef}/E_{ed} = 1{\cdot}06 \quad \text{and} \quad A_{bc} = 135$$

thus
$$\frac{A_{bc}\mu_1\gamma}{\mu_4} = \frac{135 \times 1{\cdot}06}{2{\cdot}01} = 71{\cdot}2$$

Suppose we now refer to Table 1 of $\beta = a + jb$ for $n = 2$, $a = 0{\cdot}113$ and $b = 0{\cdot}258$ and insert these values in eqn. (14)

Then

$$|S| = [(1 + 71{\cdot}2 \times 0{\cdot}113)^2 + (71{\cdot}2 \times 0{\cdot}258)^2]^{\frac{1}{2}} = 20{\cdot}47$$

and $\quad 20 \log_{10} 20{\cdot}47 = 26{\cdot}2$

therefore $\quad |S| = 26{\cdot}2\,\text{dB}$

The following table compares the predicted results with measurements made with a valve voltmeter. The last result was obtained by inserting the value of $\beta = 1/1{\cdot}42$ from eqn. (9), since $n \gg 1$.

Table 2

n	Calculated $\lvert S\rvert$ dB	Measured $\lvert S\rvert$ dB
1	0	0
1·25	16·48	16·2
2	26·2	25·7
25	34·17	34

7.1.2. Performance data for valve amplifier

(Measurements taken with 10 V r.m.s. output.)

A'_{ac} (overall gain when $n = 1$) $= 67{\cdot}5$

$\omega_\infty = 2\pi \times 401{\cdot}5$ c/s

Input impedance at $\omega_\infty = 190\,\text{k}\Omega$

Output impedance $= 1{\cdot}6\,\text{k}\Omega$

Maximum $Q = 14{\cdot}3$

($Q = \omega_\infty/(\omega_1 - \omega_2)$ where ω_1, ω_2 are the points where the response is 3 dB down)

Bandwidth (when $\gamma = 0$) = 3 dB down at 25 kc/s
$1\frac{1}{2}$ dB down at 20 c/s

Change of gain from minimum to maximum selectivity $= 0{\cdot}2$ dB

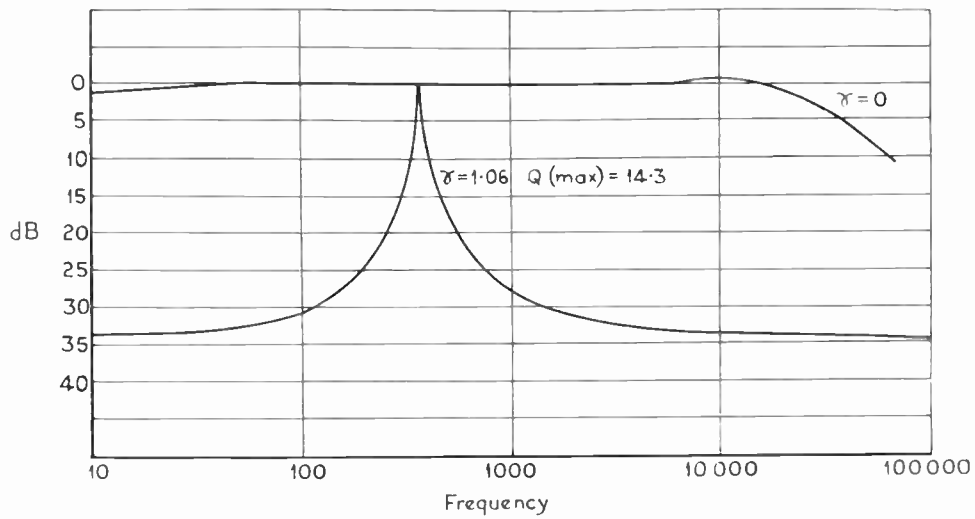


Fig. 19. Frequency response of valve selective amplifier.

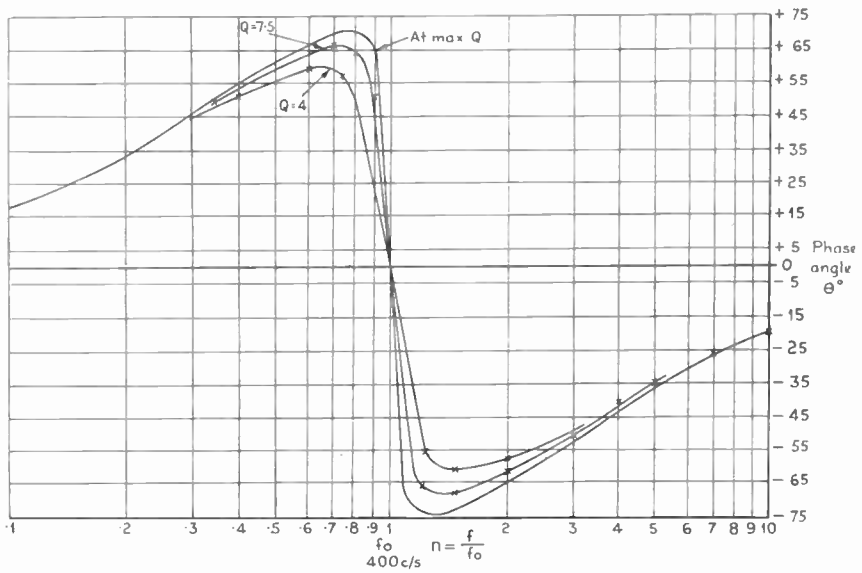


Fig. 20. Phase shift in selective valve amplifier with variation in frequency.

Change of gain at ω_∞ for 20% change of heater and h.t. voltages	negligible
Maximum output signal at ω_∞	= 30 V r.m.s.

The source impedance supplying the amplifier should be less than 10 kΩ. The load should be greater than 100 kΩ. Gain versus frequency is plotted in Fig. 19 for $\gamma = 0$ and 1·06. The phase shift with various values of selectivity is plotted in Fig. 20.

7.2. *The Transistor Amplifier*

The full circuit is shown in Fig. 21. While the design follows that of the valve amplifier, there are a number of additional considerations which arise because of the characteristics peculiar to transistors. These are:

(i) The cost of silicon transistors has dropped appreciably, so it was decided to use this type and to include no temperature stabilizing networks.

(ii) The range of current gain in silicon transistors tends to be wide, so that for best performance some degree of selection is required. If a lower standard of performance than the maximum obtainable is acceptable, the degree of feedback in the forward amplifier A_{bc} enables a reasonably wide range of current gain to be tolerated. It was found necessary to use three transistors in the forward amplifier in order to obtain sufficient gain with the correct input and output impedances.

(iii) The twin-T network was designed to work between much lower impedances. The constants are $R_i = 300\ \Omega$, $R_o = 15\ \text{k}\Omega$, $\omega_\infty = 2\pi \times 400$ c/s, $R = 3\ \text{k}\Omega$ and $C = 0{\cdot}1325\ \mu\text{F}$. R_i is formed by $R12$ and the output impedance of VT3 in series. R_o is formed by $R13$ in parallel with the input impedance of VT4.

(iv) The component $C2$ removes a rise in the response above 50 kc/s due to fall off in gain of the main

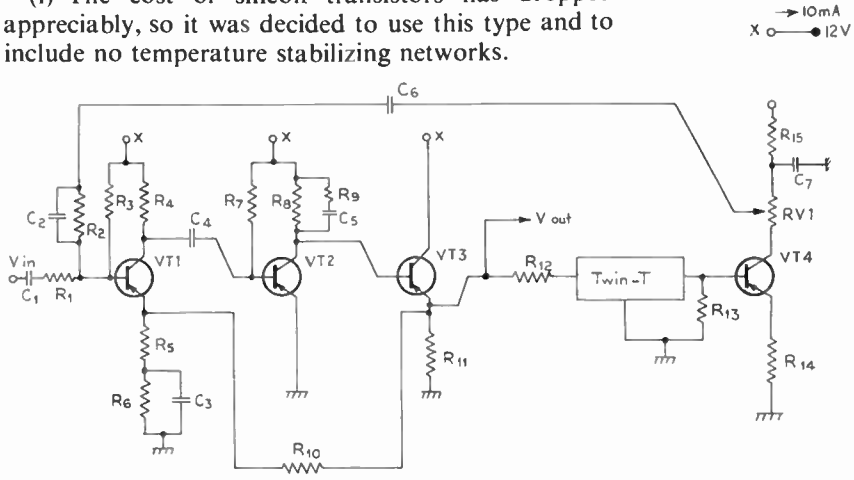


Fig. 21. Circuit of transistor selective amplifier.

All resistors ½ W 10% unless otherwise stated.

R1	—10 kΩ (H.S. ½ W, 1%)
R2	—8·2 kΩ (H.S. ½ W, 1%)
R3	—270 kΩ
R4, R6, R8	—4·7 kΩ
R5	—56 Ω (H.S. ½ W, 1%)
R7	—270 kΩ
R9	—100 Ω
R10	—6·8 kΩ (H.S. ½ W, 1%)
R11, R14	—1 kΩ
R12	—270 Ω (H.S. ½ W, 1%)
R13	—27 kΩ (H.S. ½ W, 1%)
R15	—330 Ω
RV1	—1 kΩ pot. w.w.

All capacitors 100 μF 25 V electrolytic except:

C2—47 pF mica
C5—0·02 μF paper
C7—1000 μF 25 V electrolytic

Transistors: All BCZ 11

amplifier. The value of $C2$ depends on the characteristic of the amplifier at high frequencies and is in the range 50 pF to 500 pF. Too great a value of $C2$ will cause oscillation.

(v) The filter ($R15$ and $C7$) prevents power supply ripples from being amplified by VT1.

(vi) The correct d.c. conditions in the amplifier are arrived at by selection of $R3$ and $R7$—in the 10% range—so that the voltage at the emitter of VT1 and VT3 is as close as possible to 5 V.

(vii) The biasing of VT4 will not be affected by a change of twin-T operating frequency. To maintain a given R_o and R_i the series R is unaltered and the capacitors only are changed, as was shown in section 4.

(viii) No trouble has been experienced with the single resistor biasing circuits ($R7$ and $R3$ in Fig. 21) because of electrolytic capacitor leakage. (Hunt type L37/1, 25 V working capacitors are used. The leakage current has always been less than that quoted in their catalogue.) A batch of 15 used and unused capacitors was placed in an oven and the leakage current was measured with 9 V applied. The equivalent resistance was always greater than 45 MΩ at 20° C and greater than 20 MΩ at 50° C. These resistances do not affect the 270 kΩ bias resistors appreciably.

7.2.1. System constants and calculated performance of transistor amplifier

The system constants are $R_2 = 9{\cdot}2$ kΩ, $r_{22} = 1$ kΩ, $r_2 = 8{\cdot}2$ kΩ, $R_3 = r_{ai} = 46$ kΩ, and $R_1 = 10$ kΩ. Therefore $\mu_1 = 1{\cdot}09$ and $\mu_4 = 2{\cdot}308$. $\gamma_{max} = 1$. $A_{bc} = 112$.

Entering values of a and b from the tables and calculating eqn. 14 gives the following results.

Table 3

n	Calculated $\lvert S\rvert$ dB	Measured $\lvert S\rvert$ dB
1·25	14·09	14·1
2	23·54	23·8
25	31·65	32·4
1/25	31·65	31·6

7.2.2. Performance data for transistor amplifier

A'_{ac} (overall gain when $n = 1$)	= 48
$\omega_\infty/2\pi$	= 400 c/s
Input impedance at ω_∞	= 18 kΩ approx.
Output impedance	= 30 ohms approx.
Maximum Q	= 11·5
Bandwidth (when $\gamma = 0$) with an output of 1 V r.m.s.	= 3 dB down at 23 c/s and 12 kc/s
Change of gain from maximum to minimum selectivity	= 0·7 dB
Change of gain for ∓16% change in h.t.	= ∓1·5%
Maximum output at ω_∞	= 2·0 V r.m.s.

The amplifier was placed in an oven with the following results:

Temperature °C	Gain A'_{ac}	Maximum Q
20	48	11·5
30	49	11·7
40	49	12·1
50	44·5	10·3

The twin-T network in the amplifier was changed to one with the same values of R_i and R_o but with $\omega_\infty/2\pi = 750$ c/s. The performance was unchanged except that the maximum Q was 11·4 and the change of gain from maximum to minimum selectivity was 0·5 dB. Two further amplifiers operating at 750 c/s were constructed using the same type of transistor but with differing values of current gain. In amplifier 1 the values of current gain for VT1, VT2 and VT3 were 41, 43 and 47 respectively. In amplifier 2 they were 29, 29 and 47.

The principal parameters versus temperature are given in Table 4.

Table 4

Temperature °C	Amplifier 1		Amplifier 2	
	Gain (A'_{ac})	Max Q	Gain (A'_{ac})	Max Q
22	44	8·7	38·5	8
30	44·5	8·8	39	7·9
40	42·5	8·6	37	7·5
50	39·5	8·1	34·5	7·1

8. Other Characteristics of the Amplifiers

8.1. *Stability*

The phase shift in the amplifier is negligible at ω_∞. The phase shift in the twin-T networks does not exceed ∓90 deg, so that the system is inherently stable at ω_∞.

Since both versions use three or more stages in the complete loop, high-frequency oscillations are possible. The valve version oscillated at 300 kc/s but was stabilized by the 47 pF capacitor in the internal feedback loop. The transistor version oscillated at about 100 kc/s and was stabilized by the series 100 Ω resistor and 0·02 μF capacitor across the load of VT2.

Care must be taken to ensure that such stabilizing networks do not introduce phase shift at ω_∞. If this occurs, the maximum output will not occur at ω_∞.

The denominator of eqn. (12) will contain the term $\mu_4/A_{bc} = (c+jd)$ as well as $\gamma\mu_1\beta = \gamma\mu_1\ (a+jb)$, and the maximum gain will occur at a frequency at which the whole denominator is a minimum. This will be in the vicinity of the point given by $+jd = -jb\gamma\mu_1$. A phase lag in the amplifier will cause maximum gain to occur at $\omega < \omega_\infty$.

8.2. *Increase in Selectivity*

It appears possible to increase the maximum selectivity by making $\gamma_{max} > 1$. However, if this is done the residual signal through the twin-T will be amplified so that at ω_∞, $A'_{ac} \neq A_{bc}/\mu_4$.

The problems of loop stability and the input impedance of transistor VT4 become more difficult if gain is to be obtained in this stage.

As is shown in eqn. (14), the selectivity depends primarily on the gain A_{bc} which has a practical limit. The selectivity may only be increased further by the application of positive feedback. If this is done the maximum output no longer occurs at ω_∞ and the frequency of maximum output will vary as γ is varied.

Positive feedback can occur because of the values of the twin-T terminating impedances. This partially accounts for the variable performance of the amplifiers noted under ref. 1.

8.3. *Variation of Gain with γ*

The alteration of gain as the potentiometer (and hence γ) is varied is due to two causes. The first is the change in the effective value of R_2. In the circuits given r_{22} was made as small as possible in comparison with r_2. Second, there is the residual signal through the twin-T at ω_∞. These two effects will tend to cancel each other.

8.4. *Frequency Limits for Satisfactory Operation*

The limit of operation as regards frequency is set principally by the requirement (see Sect. 8.1) that the phase shift through the network external to the twin-T should be zero at ω_∞.

8.5. *Input Impedance*

This varies from approximately $R_1+r_2+r_{22}$ at ω_∞ (where there is no overall feedback) to approximately R_1 at a frequency where $\beta\gamma\mu_1$ is sufficiently large.

9. Conclusions

The design of stable selective amplifiers, predictable in performance, may therefore be summarized as:

(i) Construct a stable conventional amplifier having a low output impedance and a high input impedance.

(ii) Construct a suitable twin-T network from the formulae in section 4.4. It has been found that a practical value for β at ω_∞ is in the range 0·001 to 0·0005 if adjustment is made according to section 6. Two networks constructed from 1% resistors and 0·5% capacitors gave a β at ω_∞ of less than 0·001, at a frequency within 1% of the theoretical value. This was without adjustment; the sample is too small to indicate whether the limits could be maintained.

(iii) Plot eqn. (7) for the value of m employed.

(iv) Predict the response from eqn. (14).

10. Acknowledgments

The author wishes to thank the Director and Council of the British Scientific Instrument Research Association for permission to publish this paper and Mr. K. Gristwood and Mr. A. Ramsay for their assistance in this work.

11. References

1. W. D. Ryan, "A tuned transistor audio amplifier", *Electronic Engng*, **31**, No. 372, p. 103, February 1959.
2. D. H. Smith, "The characteristics of parallel-T R-C networks", *Electronic Engng*, **29**, No. 348, p. 71, February 1957.
3. L. G. Cowles, "The parallel-T resistance capacitance network", *Proc. Inst. Radio Engrs*, **40**, p. 1712, 1952.
4. A. E. Hastings, "Analysis of R-C parallel-T networks and applications", *Proc. Inst. Radio Engrs*, **34**, p. 126, 1946.
5. A. Wolf, "A note on parallel-T R-C networks", *Proc. Inst. Radio Engrs*, **34**, p. 659, 1946.
6. A. P. Bolle, "Theory of twin-T R-C networks and their application to oscillators", *J. Brit.I.R.E.*, **13**, p. 571, December 1953.
7. L. Stanton, "Theory and application of parallel-T R-C frequency selective networks", *Proc. Inst. Radio Engrs*, **34**, p. 447, 1946.

12. Bibliography

R. Hutchins, "Selective R-C amplifiers using transistors", *Electronic Engng*, **33**, No. 396, pp. 84–7, February 1961.

R. A. Hall, "A 1 kc/s transistor high gain tuned amplifier", *Electronic Engng*, **30**, No. 362, pp. 192–5, April 1958.

J. J. Ward and P. V. Landshoff, "Parallel-T RC selective amplifiers", *Electronic Radio Engineer*, **35**, pp. 120–4, April 1958.

F. J. Hyde, "Selective amplification at sub-audio frequencies", *Electronic Engng*, **29**, No. 352, pp. 260–5, June 1957.

J. R. Beattie and G. K. T. Conn, "A simple low frequency amplifier", *Electronic Engng*, **25**, pp. 299–301, July 1953.

F. J. Hyde, "Highly selective amplification at low frequencies", *Wireless Engineer*, **10**, pp. 39–59, February 1959.

S. W. Punnett, "Audio frequency selective amplifiers", *J. Brit. I.R.E.*, **10**, pp. 39–59, February 1950. Correction: R. Hutchins, *J. Brit.I.R.E.*, **19**, p. 436, July 1959.

G. May and J. James, "Tuned transistor audio amplifier" (Letter), *Electronic Engng*, **33**, No. 404, pp. 674–5, October 1961.

E. Dell'Oro, "Transistorized selective amplifiers using R-C networks", *Electronic Engng*, **33**, No. 405, p. 734, November 1961.

Manuscript first received by the Institution on 24th January 1962 and in final form on 10th May 1962. (Paper No. 793.)

U.D.C. 621.37/8:362.41

Some Experiments on Reading Aids for the Blind

By

K. ELLIS, B.Sc.†

Based on a contribution to the Discussion at the Symposium on "Practical Electronic Aids for the Handicapped" in London on 28th March 1962.

Summary: Discrimination tests were carried out using scanning devices to produce aural and tactile outputs. The use of stereophonic techniques to display sounds from different parts of a character at corresponding points in the inter-aural space was investigated. A two-dimensional display is considered to be a worthwhile improvement.

Reading aids for the blind in use at the moment are unsatisfactory mainly because it is difficult to learn to use them and also the reading speeds obtainable with them are low. The highest reading speed attained with these devices is about 30 words/minute, this being obtained by one person after a considerable period of use. However, the generally obtainable speed is much lower than this and varies between 6 and 12 words/minute. The work carried out at the National Physical Laboratory and which is described in this paper is taken from "Research Study on Reading Aids for the Blind", Progress Reports Nos. 1 & 2, produced while the author was connected with the project. The idea behind this research was to produce something better than existing devices in the way of ease of learning and reading speeds, but not necessitating the use of print recognition; a device embodying this principle would be excessively expensive and large.

The Battelle optophone,‡ with which reading speeds of up to about 12 words/minute have been obtained, was used for comparison purposes. This device has a hand-held reading probe containing a vertical column of 9 photocells spanning the ascender/descender height of the character. Each cell is connected to an oscillator the outputs of which are mixed, amplified and fed to earphones. The oscillator frequencies vary from 400–4000 c/s. A vertical slice of the character is imaged on to the column of photocells and the reading probe is scanned horizontally along the line of print. When a particular cell "sees" black the associated oscillator is switched on. Thus a character produces a serially presented combination of single notes and chords. For the purposes of the experiment printed material was driven beneath the reading probe mechanically.

Two main lines of approach to the problem were adopted based on an aural output and a tactile output. Taking the first of these—an aural output—a device basically the same as that invented by P. E. Argyle § in Canada was used. This particular system was chosen for a number of reasons: a considerable amount of work was already being carried out on other devices such as the Battelle optophone, it is non-sensitive to vertical alignment of the character and also it is capable of lending itself to further modifications such as stereophony. In the Argyle device (Fig. 1) the character C is imaged on to the rectangular aperture S and the rotating disc D. The aperture

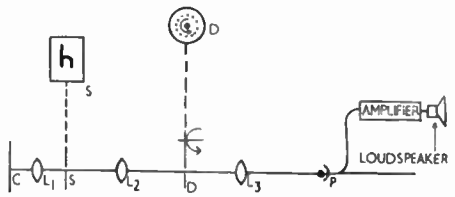


Fig. 1. Argyle reader (not to scale).

defines the vertical extent of the field of view of the system. This is scanned vertically by holes in the disc, evenly spaced along a constant radius. The entrance pupil of the system (lens L1) is imaged on to the photomultiplier P. With the height of the aperture S equal to the inter-hole distance, and a uniform field of view, the light reaching P is constant. Interruptions caused by sections of the character upset this "balance" condition and give rise to a characteristic spectrum based on the fundamental frequency of 200 c/s. The output of the photomultiplier is fed to

† Communication from the National Physical Laboratory; Mr Ellis is now at Admiralty Research Laboratory, Teddington, Middlesex.

‡ J. H. Davis, "Print recognition apparatus for blind readers", *J. Brit.I.R.E.*, **24**, No. 2, pp. 103–10, August 1962.

H. Freiberger and E. F. Murphy, "Reading machines for the blind", *Trans. Inst. Radio Engrs (Human Factors in Electronics)*, **HFE-2**, No. 1, pp. 8–19, March 1961.

§ P. E. Argyle, Private communication, September 1959.

an amplifier and then to earphones. A long vertical produces a sound best described as a "wump" whereas a horizontal line produces a "beep", with slightly varying sounds for other configurations in the character. Test material was transported using the same mechanism as for the Battelle optophone.

The second line of approach—a tactile output—was chosen because the display is inherently two-dimensional and also because little work has been carried out in this modality. Initially an actual device to produce a tactile display from print was not constructed, but instead letters embossed in heavy paper were used. The size of these letters was such that the ascender height was 0·5 in. The letters were driven beneath a rectangular aperture 1 in. square, where they were sensed by the subject's forefinger.

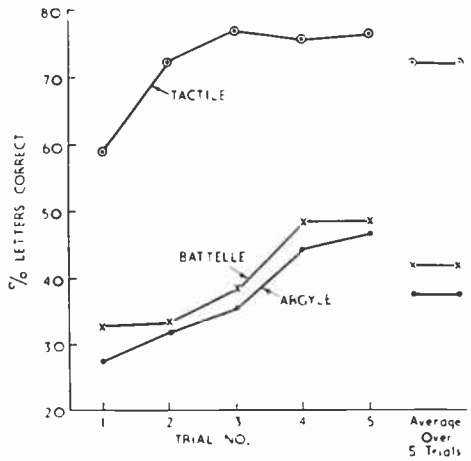


Fig. 2. Discrimination tests on Argyle, Battelle and embossed systems.

Simple discrimination tests were then carried out using the Argyle device, embossed type, and—for comparison purposes—the Battelle optophone. In each case the ten most frequent letters in the English language were used, namely e t a o n r i s h d . These were lower case letters and for the Argyle and Battelle devices were produced on an IBM Executive typewriter, this giving a very clear type. In all three cases it was arranged that the letters passed at the same speed and in order to simplify testing arrangements tape recordings were made of the outputs of the two aural devices. Five sets of 10 letters in random order were presented to the subject at 10 second intervals. Between signals he responded aurally. Before the actual test a teaching period was given in which the letters were presented three times each and then in groups depending on expected confusions, the subject being informed of the identity of each stimulus. Five subjects were used for each device and the tests were given five times each. The overall results are shown in Fig. 2. It will be seen that the two aural systems were very similar, both in accuracy and learning rate, but that with the tactile presentation a high degree of accuracy was attained on the third trial.

A tactile device to produce a display from actual print was then constructed. This consists of two portions, a scanner and a stimulator. In the former a lens images the character on to a vertical column of seven photo-voltaic cells. The output of each is fed to a Schmidt trigger which in turn operates a relay. The stimulator consists of a vertical array of seven pins at intervals of $\frac{1}{8}$ in. and covering the ascender/descender height of the character. The pins are set in a brass block and are operated by solenoids which raise them $\frac{1}{16}$ in. They are sensed by the subject's forefinger. Power is switched to a solenoid when the corresponding photocell "sees" black. In this device test material was driven beneath the scanner by means of a hand-wheel operated by the subject, thus he was able to control the speed of presentation and repeat letters where necessary. The Argyle device was modified in the same way, as also was the mechanism for presenting the embossed type.

Self-paced tests were then carried out using the tactile stimulator, the Argyle device and embossed type. The same letters were used as before: 50 symbols in all, but this time they were arranged in 5 letter nonsense words. Also the same subjects as before took part in these tests and so no teaching period was involved. These results are shown in Fig. 3. It can be seen that the tactile stimulator produced results better in speed and accuracy than the Argyle device, but not as good as with the embossed type. This is consistent with the work of Muratov, Verbuk and Fishelev † who have reported an improvement of up to $2\frac{1}{2}$ times in reading speed over an optophone system. In a further experiment on embossed type, but this time with more meaningful material, a reading speed of about 8 words/minute was achieved after only a short amount of practice.

Two methods of improving upon the Argyle device were then investigated. Firstly, an attempt was made to resolve confusions between long and short verticals (as in "h" and "n") and between ascenders and descenders (as in "b" and "p"). The defining aperture was oscillated (at about 50 c/s) in such a way that the extreme ends of ascenders and descenders were masked, thus these components produced a modulated tone. By suitable design it was possible to

† P. C. Muratov, M. A. Verbuk and Ya. R. Fishelev, "Visual Training Apparatus and Equipment for Schools for Blind and Visually Defective Children". (Institute of Defectology, Sverdlovsk, 1959.)

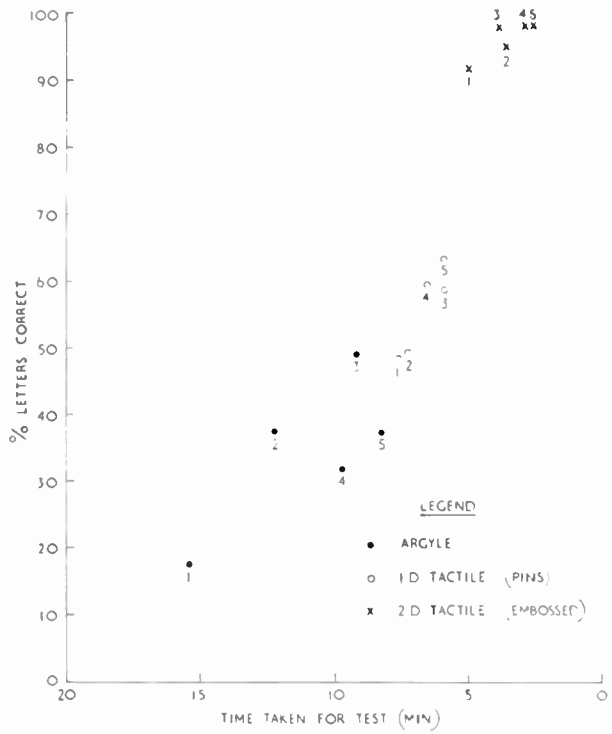


Fig. 3. Self-paced tests on Argyle and tactile (embossed and pins) systems. Plotted points represent average over all subjects for each session. Order of sessions is indicated numerically.

modulate ascenders and descenders at a different frequency. Secondly, an attempt was made to display sounds from different parts of a letter at corresponding points in the inter-aural space, i.e. produce a stereophonic display. Using both phase and amplitude stereo good localization of a single black bar in the field was obtained, but fusion occurred with two or more bars. Attempts were made to make dissimilar the two sources, but the results were not very encouraging, and other methods are being sought.

Limited success has been achieved with serial presentation in both aural and tactile modalities. Further modifications to the one-dimensional tactile device (i.e. by improving the stimulator) would improve reading speeds somewhat, but it is felt that the main improvement here would be the extension to a two-dimensional display. This could be achieved with a belt matrix where pins are pushed up in a belt as it passes over a column of actuators.† In this way, a number of letters can be sensed at the same time. The main problems are of an economic or technical nature. With regard to the aural modality it is hoped that further work on stereophonic displays may prove fruitful. Some pointers from existing techniques may be of value in guiding research though it is felt that further *ad hoc* experiments on devices should not be attempted. Rather it would be better to investigate the basic psychological parameters involved in the recognition of auditory patterns before further construction of devices is carried out.

Acknowledgments

The work described in this paper has been written as part of the research programme sponsored by St. Dunstan's at the National Physical Laboratory, and is published by permission of the Director of the Laboratory.

† In a private communication, Dr. Clowes indicated that this is now being pursued with a tape-embossing mechanism.

Manuscript first received by the Institution on 6th July 1962, and in final form on 22nd September 1962
(Contribution No. 59/MBE 9).

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e. title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied. Information on translating services will be found in the Institution publication "Library Services and Technical Information".

NEW MICROWAVE POSITION FIXING SYSTEM

An angular measurement system using microwave and digital techniques is described in a recent paper by two engineers with the National Research Council, Ottawa. It is primarily designed as a control for inshore hydrographic surveying, but has application wherever position may be found by measuring the included angles between three known points. Accuracy is better than 3 minutes of arc.

"Microwave position-fixing system uses digital display", K. Ayukawa and R. I. Mott. *Canadian Electronics Engineering*, 6, pp. 32–7, July 1962.

INTERFERENCE IN COLOUR TELEVISION

The signal-to-noise ratio for the limit of perceptibility of statistical interference (noise) as a function of frequency has been determined by viewer tests in colour-television using the N.T.S.C. system in the same way as for monochrome television. A recent German paper describes how by means of a bandpass filter with a variable centre frequency, a uniform level noise band with a width of approximately 0·5 Mc/s was filtered out of a uniform noise spectrum and added to the video signal in order to determine the perceptibility limit for this particular frequency range. Noise evaluation curves for the saturated primary colours and their complementary colours as well as for some ordinary colour slides were determined.

"The perceptibility of statistical interference in colour television pictures using the N.T.S.C. system", J. Müller and G. Wengenroth. *Nachrichtentechnische Zeitschrift*, **15**, pp. 438–41, September 1962.

BREAKDOWN IN DIODES

When the reverse voltage applied to a silicon power diode through a high series resistance exceeds a critical value, the silicon diode will break down. In many cases, however, the breakdown does not mean permanent damage because the reverse current is limited. It was found from these non-destructive tests that the breakdown in actual silicon power diode was induced by localized thermal runaway in one or a few small local parts which were about $10^{-2} \sim 10^{-4}$ mm² area.

After the localized thermal runaway occurs, the temperature of that local spot is determined by the value of the series resistance. The v–i characteristics of the diode are determined from the thermal-equilibrium relation.

"Breakdown in silicon power diodes", H. Oka and S. Oshima. *Mitsubishi Denki Laboratory Reports*, **3**, No. 2, pp. 165–81, October 1962.

AGEING OF SELENIUM RECTIFIERS

In a recent Czech paper the technical consequences of ageing in selenium rectifiers in the inverse as well as the forward direction is treated. The static response and the customary ageing characteristics are used as an illustration. The physical reasons for ageing seem to lie in the structure changes caused by diffusion and are accelerated by enhanced working temperature, causing in turn a resistance rise in the forward direction and resulting in an undesirable life-time limitation of the selenium plates. Another physical reason is the out-diffusion of activators from the selenium layer and the diffusion of undesired deactivators into it. The problem of ageing may be approached using the classical diffusion laws together with the quantitative requirements. The deactivator mentioned above is oxygen from the surrounding air and chemically active elements used in the manufacture of the selenium plates, e.g. potassium in the form of potassium permanganate. To prevent ageing, elements with a small diffusion coefficient are recommended; the influence of atmospheric oxygen—which cannot be totally avoided—on semi-finished plates should not be prolonged. The finished plates should be preserved by varnishing and by immersing the stacks in oil ("oil-immersed rectifiers"). Under these circumstances even selenium rectifiers may be considered to be stable and practically non-ageing components.

"The ageing of semi-conductor rectifiers with special reference to selenium", J. Kroczek and J. B. Slavik. *Slaboproudy Obzor*, 23, No. 7, pp. 369–73, July 1962.

WAVEGUIDE MEASUREMENTS

Resonant-cavity measuring equipment operating in the frequency region of 50 Gc/s has been designed in Japan to study the characteristics of solid copper and helix waveguides of 51 mm inner diameter. The attenuation constant of the waveguide has been obtained by measuring the Q-factors of the cavity which was constructed by short-circuiting both ends of the waveguide. Flat frequency characteristics of the cavity Q-factors have been obtained for the helix waveguides, but fluctuations of the frequency characteristics have been observed for the solid copper waveguide. Mode conversion losses at imperfect joints of the waveguides have been measured by the same method and the results were in good agreement with the theoretical values.

"Measurements on circular electric waveguide characteristics by the resonant-cavity method", M. Shimba. *Review of the Electrical Communications Laboratory of the Nippon Telegraphs and Telephone Company*, **10**, pp. 248–60, May-June 1961.

COLOUR TELEVISION GRADATION EQUALIZATION

In colour television transmission of colour films or colour dispositives it is frequently a desirable feature to add to the gradation input equalizer usually provided in scanners for gradation compensation of the picture tube, a further gradation equalizer for matching the contrast in a transmission system to the contrast on the film or slide.

The gradation equalizer described in a recent German paper operates on the known principle of multiplicative gradation equalization in which the three chrominance signals are multiplied with a signal which is equal for all three channels and is derived from the luminance value of the colour signals. By suitable control of this signal a continuous remote control of the gamma value becomes possible. For the purpose of simplifying the level adjustments, the gradation equalizer contains a control circuit which permits a continuous supervision of the equality of the gradation curves and of the adjustment of compensation for the correcting signal in the three channels.

"A gradation equalizer for colour television scanners", H. Strauss. *Nachrichtentechnische Zeitschrift*, **15**, pp. 371–9, August 1962.

TRANSISTOR BARRIER-LAYER TEMPERATURE

In transistor circuit engineering it is important to know the maximum barrier-layer temperature occurring in the case of dissipations that vary with time, as encountered in pulse and amplifier operation. It is also valuable to have guidance with respect to the conditions under which the arithmetic mean of the dissipated power can be used in circuit design.

Under these aspects the calculation of the barrier-layer temperature is shown and discussed for two cases: cyclic square-wave variation with time of the dissipated power with superimposed transitional losses encountered in switching the transistor; and constant dissipated power with a sinusoidal superimposition of power variations.

"Variation of transistor barrier-layer temperature with varying dissipation", F. Weitzsch. *Archiv der Elektrischen Übertragung*, **16**, No. 7, pp. 335–42, July 1962.

SECAM COLOUR TELEVISION ERRORS

After summarizing the transmission errors permissible with the N.T.S.C. method, a recent German paper discusses the improvements on amplitude and phase influences on the subcarrier by using the SECAM method with frequency-modulated subcarrier (SECAM-f.m.). The problem of signal distortion introduced by the relatively strong band limiting of the frequency modulated colour signal and the influence of the delay errors is dealt with by reference to experimental investigations. The effects of reflection interference with the SECAM-f.m. method and the residual influence of differential phase errors are studied. The paper concludes by comparing colour subcarrier interference and internal interference with the brightness signal ("cross-colour") in the case of the N.T.S.C. and SECAM-f.m. methods.

"The influence of system and transmission errors on a colour television system using the SECAM methods", von Helmut Schonfelder. *Archiv der Elektrischen Übertragung*, **16**, pp. 385–99, August 1962.

TESTING STEREOPHONIC SYSTEMS

The tests described in a recent German paper show that a stereophonic transmission of square wave pulses having a transmission bandwidth slightly larger than in the case of signal channel transmission, gives satisfactory results when the directional information is available in the form of differences in delay. However, the measured direction depends on the type of signal when delay differences only are used. For a certain method of reproduction a geometrical disposition of the microphones can be given, which produces the most accurate directional effect. A transmission arrangement in which two closely positioned microphones with different directional patterns are used requires approximately twice the channel capacity of signal channel transmission when test pulses are to be transmitted with high quality.

"Stereophonic double-channel transmission with reduced bandwidth", B. Cramer and K. Lohmann. *Nachrichtentechnische Zeitschrift*, **15**, pp. 485–94, October 1962.

A SIMPLE STATISTICAL INSTRUMENT

An instrument which is capable of charting the resistance distribution diagram of some hundreds of specimens of fixed resistors without the operator noting any readings, making clerical entries or performing calculations is described in a recent Australian paper. The instrument is quick and accurate, and the data are presented in a form which is readily interpreted visually.

The basic method is the use of small steel spheres, one taking up its place in the transparent chart board for each specimen observed. The spheres build up a profile which is essentially the distribution curve or histogram for the lot under study.

Whilst the method is illustrated in relation to its use in the process control of resistor manufacture, the principle is applicable to the charting of any variables that can be indicated on a dial either of an electrical instrument or a mechanical dial gauge.

"A simple instrument for statistical process control", M. M. Lusby. *Proceedings of the Institution of Radio Engineers, Australia*, **23**, pp. 580–2, October 1962.

DIAPHRAGM LOADED WAVEGUIDES

An engineer in the High Frequency Institute at Aachen has investigated whether diaphragm-loaded waveguides can serve as structural elements of low-power travelling-wave tubes at very high frequencies, and what advantages they offer over delay lines of conventional design. It turns out that, when properly dimensioned, diaphragm-loaded waveguides present an almost constant phase velocity of the (plus) first partial mode over relatively wide frequency bands. This characteristic, their simple structural design, and their high resistance to heat hold promise for their possible applications in travelling-wave tubes for highest microwave frequencies. From the diaphragm-loaded waveguide a multiple delay line can be derived which allows an increase in the power output by using a number of electron beams guided in parallel.

"Diaphragm-loaded waveguides as delay lines", A. Fiebig. *Archiv der Elektrischen Übertragung*, **16**, pp. 283–90. June 1962.