THE RADIO AND ELECTRONIC ENGINEER

The Journal of the Institution of Electronic and 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 27 MARCH 1964 NUMBER 3

CHANGE OF THE INSTITUTION'S NAME

ADVICE has been received from the Clerk of the Privy Council that The Queen was pleased, at a Privy Council held by Her Majesty on 26th February 1964, to make an Order in Council allowing amendments to the Charter of The British Institution of Radio Engineers changing its name to THE INSTITUTION OF ELECTRONIC AND RADIO ENGINEERS.

When the Institution was formed on 31st October 1925 there was considerable debate as to the most apt title and many founder members favoured inclusion of the term "electronics". Discussion on this point was revived in 1944 when the Institution's *Post-War Development Report* stressed the future importance of electronics, which was defined as "the radio valve or kindred devices at work in ways other than direct aural or visual communication". The development of electronics, as distinct from communication science and technology, was even then showing increasing prominence in the learned society activities of the Institution. This was also reflected in the professional occupations of the members.

This trend was constitutionally recognized when, in 1945, an Extraordinary General Meeting of members resolved to amend the Memorandum of Association of the Institution so as to include in its objects ". . . the theory, science, practice and engineering of electronics and all kindred subjects and their applications".

Subsequently, and even more important, the Royal Charter of Incorporation granted on 2nd August 1961, refers to the objects of the Institution as including advancement of the science and practice of radio engineering and the theory, science, practice and engineering of electronics.

A Resolution of the Council of the Institution was finally circulated in order that Corporate Members should have an opportunity to comment on the acceptability of change of name. As a result, a Resolution was passed at a Special General Meeting of Corporate Members held in November 1963 which sought permission to amend the Charter so that the Institution might change its name to THE INSTITUTION OF ELECTRONIC AND RADIO ENGINEERS.

A corollary to the change of name might well have been rearrangement of the title of the *Journal* which, since January 1963, has been called *The Radio and Electronic Engineer*. If, however, the title of the *Journal* were to be "The Electronic and Radio Engineer", some bibliographical and reference difficulty might arise, inasmuch as the technical journal founded as "Wireless Engineer" and subsequently known as "Electronic Technology" bore the title of "Electronic and Radio Engineer" from 1957 to 1959. To avoid any ambiguity of references, therefore, it has been decided that we must retain as the title of the Institution's Journal *The Radio and Electronic Engineer*.

G. D. C.

INSTITUTION NOTICES

Proceedings of the Symposium on "Sonar Systems"

The papers presented at the Symposium on "Sonar Systems" in Birmingham from 9th to 11th July 1962 have now been published in collected form with reports of the associated discussions. The volume may be purchased from the Publications Department of the I.E.R.E., 8–9 Bedford Square, London, W.C.1, at a charge of £3 including postage. A list of papers which were presented at this Symposium was published in the Brit.I.R.E. *Journal*, for June 1962 (page 507).

The Prince Philip Medal

The Prince Philip Medal of the City and Guilds of London Institute for 1963 has been awarded to Mr. R. O. R. Chisholm (Associate Member). The medal is awarded annually 'in recognition of outstanding promise or achievement in the promotion theory or practice of science and technology' and is restricted to those who have received qualifications by examination by the Institute.

Mr. Chisholm is a chief projects engineer with a guided weapons manufacturer and is a past chairman of the Institution's South Western Section. The medal was presented to him by H.R.H. The Duke of Edinburgh at Buckingham Palace on 25th March.

Colour Television in Europe

Delegations from the Broadcasting Administrations of 19 countries met in London from 14th–25th February under the auspices of the C.C.I.R. (Comité Consultatif International des Radio-Communications) to discuss the problems involved in the choice of standards for public colour television series in the European Broadcasting Area. During these ten days in London, the delegations were shown a wide variety of colour television installations and techniques, and much background information was provided in the form of results of extensive field trials carried out in Great Britain, France, Holland, Italy, Switzerland and Western Germany.

It had been widely hoped that the technical evidence of the work undertaken to date on the N.T.C.S., SECAM and PAL systems—coupled with the operating experience of ten years of the N.T.C.S. system in the U.S.A.—would have allowed a decision to be given at this point in time. Most of the delegations, however, felt it preferable to wait until the next meeting of C.C.I.R. Study Group XI to be held in Vienna in the spring of 1965 to make a final review and to reach a final recommendation.

Abstracts of J. Brit.I.R.E. Papers 1952–63

A new edition—the seventh—of "Abstracts of papers published in the *Journal* of the British Institution of Radio Engineers" has been prepared and copies are available from the Institution, price 10s. 6d. including postage (price to members of the Institution is 7s. 6d.). Covering the period from 1952 to 1963 the new edition of "Abstracts" completes the listing of papers published by the "Brit.I.R.E." and altogether some 850 papers and reports are shown with details of content, publication date and present availability.

Papers on "Cold Cathode Tubes"

Copies of papers presented at the Symposium on "Cold Cathode Tubes and their Applications", held in Cambridge from 17th to 19th March 1964, are still available price 2s. 6d. each from the Publications Department of the Institution. The complete list of papers and synopses was published on pp. 117–124 of the February issue of *The Radio and Electronic Engineer*.

Institution Meetings in France

In the course of a visit to the International Components Exhibition in Paris in February, Mr. F. W. Sharp (Assistant Editor of the *Journal*) met informally a number of members who are resident in France and discussed the possibility of arranging occasional technical meetings. The majority of these members live in the Paris area and considerable interest was expressed in the proposals. A further announcement will be made in due course and details will be sent to members.

Undelivered Mail

A number of postal items, including copies of the *Journal* and *Proceedings*, have been returned to the Institution as undelivered mail. In order to receive correspondence and publications promptly from the Institution, members are requested to notify the Secretary prior to or *immediately* after changing their address. This will also help in keeping the records correct for the publication of "List of Members".

Change of Address

Following the change of name of the Institution members are asked to note that the correct address of the Institution is now:

The Institution of Electronic and Radio Engineers,
8–9 Bedford Square,
London, W.C.1.

U.D.C. 621.372.542

A Comparison of Five Methods of Low-pass Passive Filter Design

By

A. G. J. HOLT, Ph.D.

(*Associate Member*)†

Summary: The phase and magnitude characteristics of five network functions, all three-pole approximations to an ideal low-pass filter characteristic, are compared. The functions considered are the Butterworth, Chebyshev, Papoulis Class L, 'maximally flat delay' and 'in-line pole'; all of these can be realized with a simple passive circuit.

The equations for the pole locations are quoted. Plots of the pole locations and the magnitude and phase responses are given for each of the functions normalized for unit resistance and a pass band from $\omega = 0$ to $\omega = 1$. The suitability of the different functions for use in audio circuits and when the output waveform is to be observed on a c.r.o. is also discussed. A plot of the pole locations for the five functions is given and the effect of changes in the pole positions is noted.

Finally, experimental results are given showing the magnitude, phase and pulse responses of practical filters designed to realize the network functions discussed.

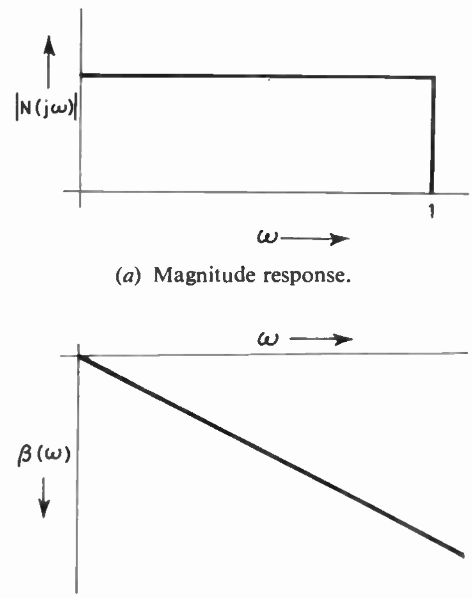
1. Introduction

Since the development of the synthesis of linear passive networks using the insertion loss method by Darlington[1] and others, many different forms of network response have been devised.

When a network is used as a low-pass filter the ideal shape for magnitude response is the rectangular form shown in Fig. 1(*a*). The ideal phase response is a linear function of frequency as shown in Fig. 1(*b*). Intuition leads one to expect, and it can be shown, that this ideal response cannot be obtained using a finite number of components. Hence one is left with the problem of obtaining an approximation to the ideal which will be a satisfactory solution for the problem in hand.

1.1. *Approximations to the Ideal*

Owing to the difficulties experienced in approximating to both ideal magnitude and phase characteristics at the same time, a simpler solution is often sought by concentrating the attention upon one characteristic or the other. Such approximations to the magnitude response have resulted in the 'maximally flat', or Butterworth, characteristic[2] and the 'equal ripple', or Chebyshev characteristic. These forms of response are frequently employed in filters intended for use in speech circuits, for which the phase response is unimportant because of the insensitivity of the human ear to the phase relations between the components in the sound waves impinging upon it. When it is expected that the output waveform from a filter will be applied to a cathode-ray tube the phase response cannot be ignored and attempts are made to approximate to the linear phase characteristic. These have resulted in the 'maximally flat delay' and the 'in-line pole'[3] characteristics.



(*a*) Magnitude response.

(*b*) Phase response.

Fig. 1. Ideal low-pass filter.

An intermediate case has been obtained by Papoulis.[4] This was derived from a consideration of the magnitude characteristic but was intended to give a good transient response, for which the phase characteristic is also important.

† Department of Electrical Engineering, University of Newcastle upon Tyne.

All these characteristics can be obtained with simple ladder networks consisting of inductors and capacitors and terminated at each end in equal resistors.

Although the characteristics are chosen to be an approximation to the ideal magnitude or phase responses, these two responses are associated with one another so that when, say a magnitude response is selected, a phase response has also been selected with it. The different characteristics have responses which differ from one another in and out of the pass band of the filter, and in this paper the phase and magnitude responses for the five characteristics mentioned above are compared. Practical filters have been constructed to the normalized designs given and the phase, magnitude and pulse responses of these circuits are shown.

Only all pole approximations to the ideal will be considered. The general form for the network function $N(p)$ for an all pole network is:

$$N(p) = \frac{\text{response}}{\text{excitation}} = \frac{H}{a_n p^n + a_{n-1} p^{n-1} + \ldots + a_0} \qquad \ldots\ldots(1)$$

Here H and the a's are real constants.

For simplicity the three-pole approximation ($n = 3$) will be used, this means that all the characteristics can be obtained from a simple network of the form in Fig. 2.

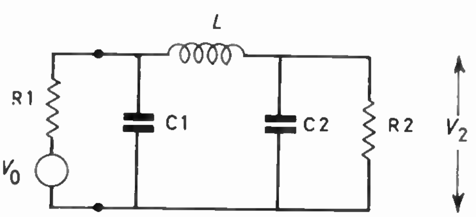


Fig. 2. Simple filter network.

The network function for this circuit has the form:

$$N(p) = \frac{V_2}{V_0} = \frac{H}{p^3 + a_2 p^2 + a_1 p + a_0} \qquad \ldots\ldots(2)$$

This can be factorized as follows:

$$N(p) = \frac{H}{(p-p_1)(p-p_2)(p-p_3)} \qquad \ldots\ldots(3)$$

In equation (3) the p's are the pole locations which, when plotted, must lie in the left half of the complex plane ($p = \sigma + j\omega$). This ensures that the network is stable. In fact, one of the poles will lie on the real ($-\sigma$) axis and the other pair will be complex conjugates in the left half plane.

Changing the pole locations in equation (3) results in different values for the a's in equation (2). Analysis shows that the transfer function for the network in Fig. 2 when $R_1 = R_2 = 1$ is

$$N(p) = \frac{\dfrac{1}{LC_1C_2}}{p^3 + \left(\dfrac{C_1+C_2}{C_1C_2}\right)p^2 + \left(\dfrac{1}{C_1C_2} + \dfrac{C_1+C_2}{LC_1C_2}\right)p + \dfrac{2}{LC_1C_2}} \qquad \ldots\ldots(4)$$

Comparison of the coefficients of p in the denominators of equations (2) and (4) shows that changes in the pole locations in equation (3) result in changes to the element values C_1, C_2 and L of the network Fig. 2. Hence, when selecting a characteristic one is selecting a network function having specified pole locations and a network having specified element values.

The characteristics listed in the Introduction will now be considered in turn.

2. Butterworth or 'Maximally Flat Magnitude' Characteristic

In order to obtain a magnitude response which is almost a constant at the low-frequency end of the filter pass-band one finds the magnitude of equation (2) as a function of ω and expands this as a Taylor series about $\omega = 0$. Setting the first $(2n - 1)$ derivatives with respect to ω of the magnitude response equal to zero gives the maximally flat magnitude characteristic of the form

$$|N(j\omega)| = \frac{H}{\sqrt{1+\omega^{2n}}} \qquad \ldots\ldots(5)$$

or

$$|N(j\omega)| = \frac{\frac{1}{2}}{\sqrt{1+\omega^6}} \qquad \ldots\ldots(6)$$

when the network in Fig. 2 is used.

Setting the derivatives equal to zero ensures that the difference between the ideal and equation (5) is zero at $\omega = 0$ and increases as ω is increased.

Applying analytic continuation to equation (6) gives the response as a function of the complex variable p:

$$N(p) \,.\, N(-p) = \frac{(\frac{1}{2})^2}{(1-p^6)} \qquad \ldots\ldots(7)$$

2.1. *Pole Locations*

Clearly, the poles of $N(p) \,.\, N(-p)$ may be obtained from the zeros of the denominator in equation (7).

In general, when the number of poles is not limited to three but is equal to n, the pole locations are given by the following equations:

$$\left.\begin{aligned} & p_K = \sigma_K + j\omega_K \\ \text{where} \quad & \sigma_K = \sin\left(\frac{2K-1}{2n}\right)\pi \\ & \omega_K = \cos\left(\frac{2K-1}{2n}\right)\pi \\ \text{and} \quad & K = 1, 2, \ldots 2n \end{aligned}\right\} \quad \ldots\ldots(8)$$

The poles lie on a circle of unit radius. In order to ensure that the network is stable the three poles in the left half plane must be selected as the poles of the network function $N(p)$.

This gives the pole locations shown in Fig. 3.

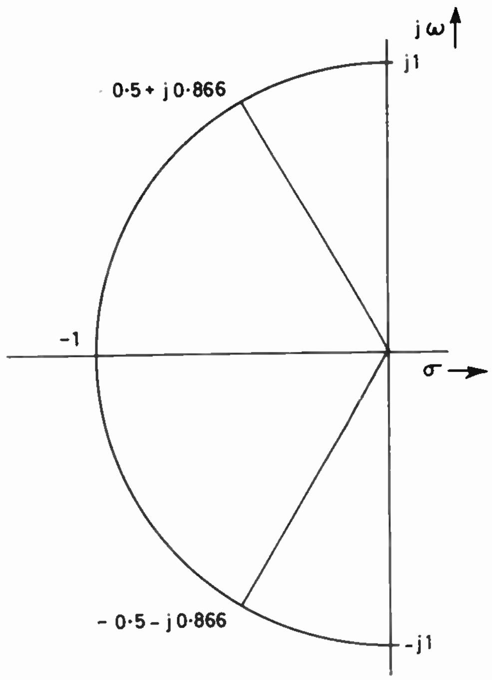


Fig. 3. Pole locations for 3-pole Butterworth function.

The calculated pole locations are:

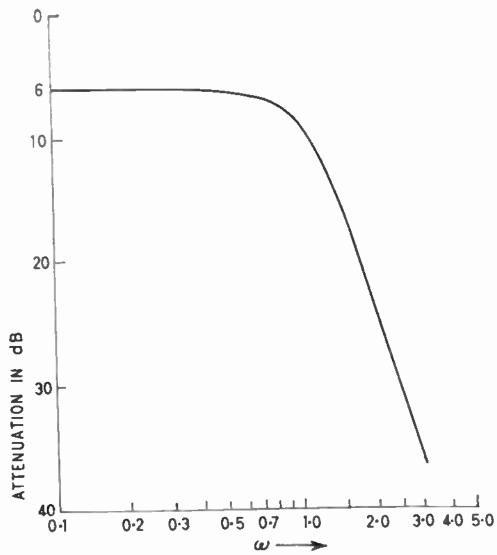
$$\left.\begin{aligned} p_1 &= -1 \\ p_2 &= -0{\cdot}5 + j0{\cdot}866 \\ p_3 &= -0{\cdot}5 - j0{\cdot}866 \end{aligned}\right\} \quad \ldots\ldots(9)$$

Hence the network function for the three-pole maximally flat magnitude characteristic is

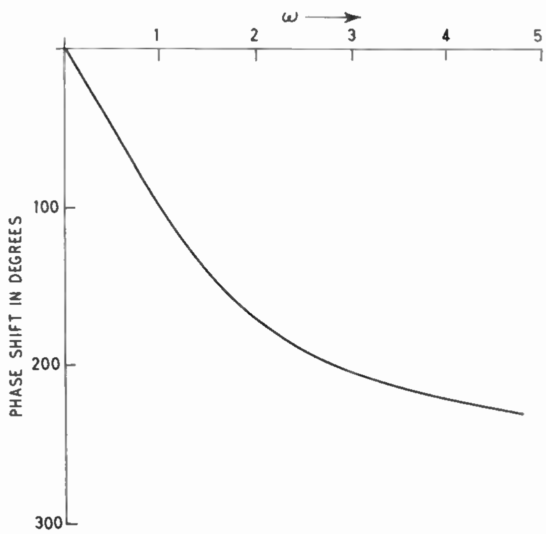
$$N(p) = \frac{\frac{1}{2}}{p^3 + 2p^2 + 2p + 1} \quad \ldots\ldots(10)$$

2.2. *Magnitude and Phase Responses*

A plot of the magnitude and phase responses obtained from a pole plot of equation (10) is shown in Fig. 4(*a*) and 4(*b*).



(*a*) Magnitude response.



(*b*) Phase response.

Fig. 4. Normalized Butterworth function.

3. Chebyshev or 'Equal Ripple Magnitude' Characteristic

This characteristic is also derived by concentrating upon the magnitude response. The magnitude of the network function can be written

$$|N(j\omega)| = \frac{\frac{1}{2}}{\sqrt{1 + \varepsilon^2 C_n^2(\omega)}} \quad \ldots\ldots(11)$$

where ε is a constant and $C_n(\omega)$ is the Chebyshev polynomial of order n.

The recursion formula for these polynomials is

$$C_{n+1}(\omega) = 2.\omega.C_n(\omega) - C_{n-1}(\omega) \qquad \text{......(12)}$$

where $\quad C_0 = 1 \quad \text{and} \quad C_1 = \omega$

For $\quad n = 3,$

$$C_3(\omega) = 4\omega^3 - 3\omega$$

The Chebyshev characteristic differs from the ideal by a ripple having an amplitude which can be controlled in the design. Hence the difference between the Chebyshev and the ideal magnitude responses can be made zero at a number of frequencies in the pass band. The number of frequencies for which response is equal to the ideal depends upon the value of n in equation (11), that is, upon the complexity of the network.

In the stop band ($\omega \geqslant 1$) the rate at which the attenuation increases with frequency depends upon n and ε.

3.1. *Pole Locations for Chebyshev Design*

Using analytic continuation on equation (11) gives

$$N(p).N(-p) = \frac{(\frac{1}{2})^2}{1+\varepsilon^2.C_n^2\left(\frac{p}{\mathrm{j}}\right)} \qquad \text{......(13)}$$

It follows that the pole locations for equation (13) are the zeros of the denominator. It can be shown that the poles p_K are given by

$$\left.\begin{aligned} p_K &= \sigma_K + \mathrm{j}\omega_K \\ \text{where} \quad \sigma_K &= \sin u_K.\sinh v \\ \omega_K &= \cos u_K.\cosh v \\ K &= 1, 2, \ldots 2n \\ \text{and} \quad \cos^{-1} p/\mathrm{j} &= u + \mathrm{j}v \end{aligned}\right\} \qquad \text{......(14)}$$

These poles lie on an ellipse in the p plane, the dimensions of which are affected by both n and ε.

3.2. *Bandwidth*

The upper limit of the pass band occurs at $\omega = 1$ where the magnitude response becomes

$$|N(\mathrm{j}1)| = \frac{\frac{1}{2}}{\sqrt{1+\varepsilon^2}}$$

Note that the magnitude response at $\omega = 1$ is only 3 dB below its maximum value if ε is almost equal to unity. For the Chebyshev response the bandwidth from $\omega = 0$ to $\omega = 1$ is known as the tolerance band.

3.3. *Pole Locations for Three-pole Case*

Setting $n = 3$ and designing for a ripple amplitude of 3 dB in the pass-band gives $\varepsilon = 0{\cdot}9976$. This value of ripple is fairly large but is acceptable in applications for which the large attenuation obtained in the stop band is very desirable.

Equations (14) gives the pole locations for $n = 3$; selecting the poles in the left half-plane to give a stable network results in the pole locations shown in Fig. 5 being those of the required network function.

The pole locations are:

$$\left.\begin{aligned} p_1 &= -\,0{\cdot}2986 \\ p_2 &= -\,0{\cdot}1493 + \mathrm{j}0{\cdot}9038 \\ p_3 &= -\,0{\cdot}1493 - \mathrm{j}0{\cdot}9038 \end{aligned}\right\} \qquad \text{......(15)}$$

Hence the network function $N(p)$ giving the 3-pole Chebyshev response with a ripple of 3 dB with the network shown in Fig. 2 is

$$N(p) = \frac{0{\cdot}1252}{p^3 + 0{\cdot}5972p^2 + 0{\cdot}9283p + 0{\cdot}2506} \qquad \text{......(16)}$$

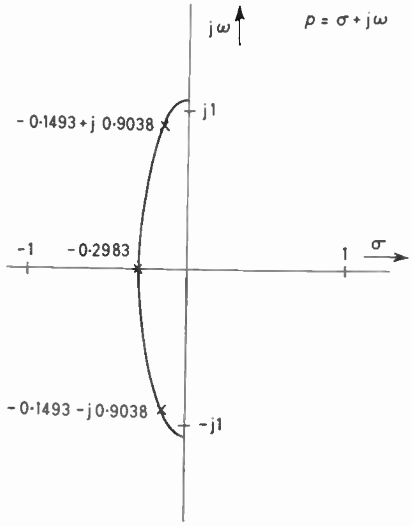


Fig. 5. Locations for 3-pole Chebyshev function.

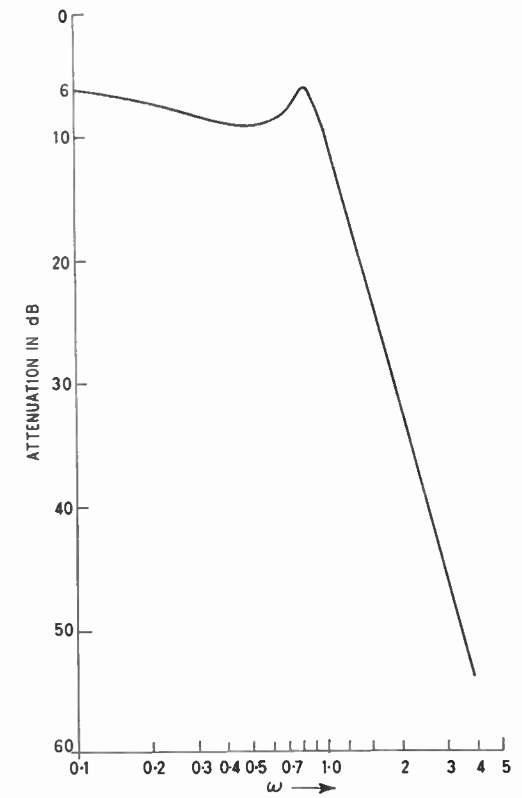
Note that the poles are located much closer to the $\mathrm{j}\omega$ axis in Fig. 5 than are the poles for the maximally flat magnitude design Fig. 3. This is especially so for the pair of complex poles. Because of this fact small changes in the positions of the complex poles can make a considerable difference to the magnitude response of the Chebyshev design, and care must be taken to ensure that the pole locations are correct if the required response is to be obtained.

3.4. *Magnitude and Phase Characteristics*

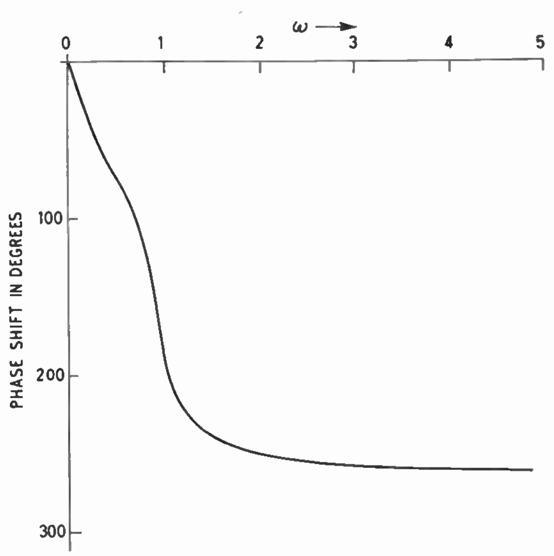
Plotting the magnitude and phase responses from the pole locations in Fig. 5 results in the curves shown in Figs. 6(*a*) and 6(*b*).

4. Papoulis Class L Characteristic

An intermediate case between the Butterworth and the 3 dB ripple Chebyshev characteristics is that derived by Papoulis[4] and named the Class L response by him. The magnitude response of the Class L design has no ripples, although a close examination in the pass band range reveals a rounded stepped response of the staircase pattern.



(*a*) Magnitude response.



(*b*) Phase response.

Fig. 6. Normalized Chebyshev function.

At $\omega = 1$, the limit of the pass band, the attenuation is 3 dB greater than the minimum value at $\omega = 0$. The magnitude response of the Class L characteristic is designed to have the greatest slope that can be obtained at $\omega = 1$ with a characteristic which is free from ripples in its pass band. The resulting magnitude response has a stop-band attenuation which is greater than that given by a Butterworth design having the same number of poles n, that is, for example, by the circuit form shown in Fig. 2. A greater stop-band attenuation may be obtained using the Chebyshev design but this is accompanied by ripples in the pass band which may be undesirable.

Papoulis has shown that a monotonic magnitude response having maximum slope at $\omega = 1$ is obtained from the expression

$$|N(j\omega)| = \frac{H}{\sqrt{1+L_n(\omega^2)}} \qquad \ldots\ldots(17)$$

Here $$L_n(\omega^2) = \int_{-1}^{2\omega-1} \gamma^2(x)\,dx$$

and $$\gamma(x) = b_0 + b_1 P_1(x) + \ldots + b_K . P_K(x)$$

$$b_0 = \frac{b_1}{3} = \ldots \frac{b_K}{2K+1} = \frac{1}{\sqrt{2(K+1)}}$$

The P_K are the tabulated Legendre polynomials of the first kind.[5]

Fukada has provided a table of the $L_n(\omega^2)$ which also shows the slope of the magnitude characteristic at the limit of the pass band for different values of n.

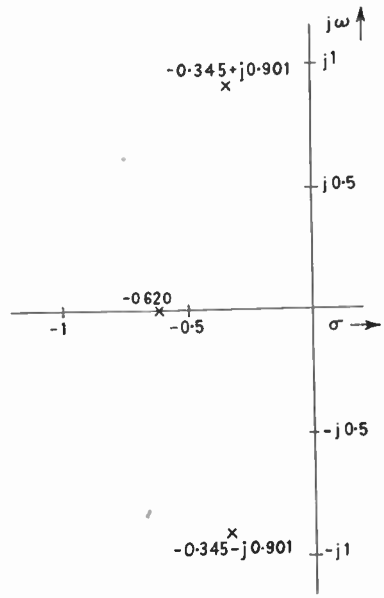


Fig. 7. Pole locations for 3-pole Papoulis class-L network function.

4.1. *Pole Locations*

For the three-pole case ($n = 3$), which we are using as an example, $L_n(\omega^2)$ becomes

$$L_3(\omega^2) = 3\omega^6 - 3\omega^4 + \omega^2$$

Writing $p/\mathrm{j} = \omega$ and using analytic continuation gives

$$N(p).N(-p) = \frac{H}{1 - p^2 - 3p^4 - 3p^6} \qquad \ldots\ldots(18)$$

When equation (18) is factorized and the poles in the left half-plane are retained, the network function $N(p)$ is given by equation (19).

$$N(p) = \frac{0{\cdot}289}{(p+0{\cdot}620)(p+0{\cdot}345+\mathrm{j}0{\cdot}901)(p+0{\cdot}345-\mathrm{j}0{\cdot}901)} \qquad \ldots\ldots(19)$$

Multiplying out the denominator gives

$$N(p) = \frac{0{\cdot}289}{p^3 + 1{\cdot}310p^2 + 1{\cdot}359p + 0{\cdot}577} \qquad \ldots\ldots(20)$$

The poles are shown plotted in the complex plane in Fig. 7.

Note that the poles for the Papoulis design lie between those for the Butterworth and the Chebyshev designs.

4.2. *Magnitude and Phase Characteristics*

Plotting the magnitude and phase responses for the three-pole case from Fig. 7 the curves shown in Figs. 8(*a*) and (*b*).

5. Summary on Designs for Magnitude Characteristics

The three designs considered so far were derived from consideration of the magnitude characteristic. In each case the phase response was treated as a matter of secondary importance which would have to be accepted when the required magnitude response was obtained.

This has resulted in responses for which the stop-band attenuation increases fairly rapidly and phase responses which show appreciable departure from the ideal straight line form.

Let us now consider designs in which obtaining linearity of the phase response is treated as a matter of first importance.

6. Maximally Flat Delay Characteristic

This is an approximation to the linear phase response. The delay $\tau(\omega)$ introduced by a network is defined as follows:

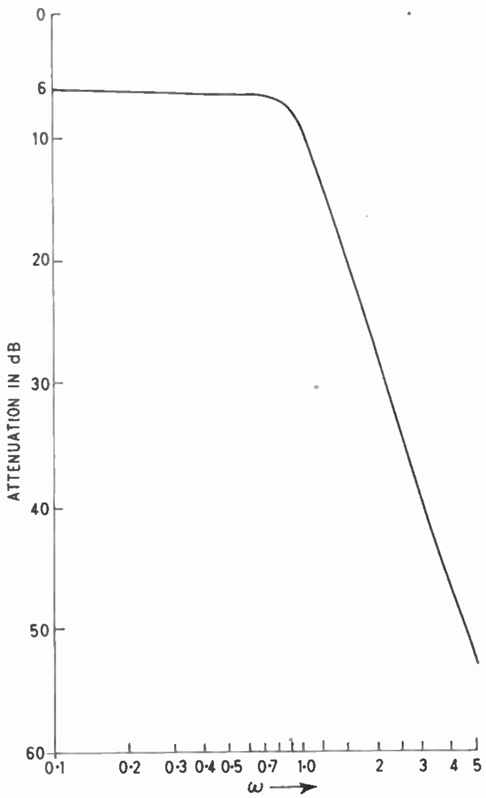
$$\tau(\omega) = -\frac{\mathrm{d}\beta(\omega)}{\mathrm{d}\omega} \qquad \ldots\ldots(21)$$

where $\beta(\omega)$ is the phase response of the network.

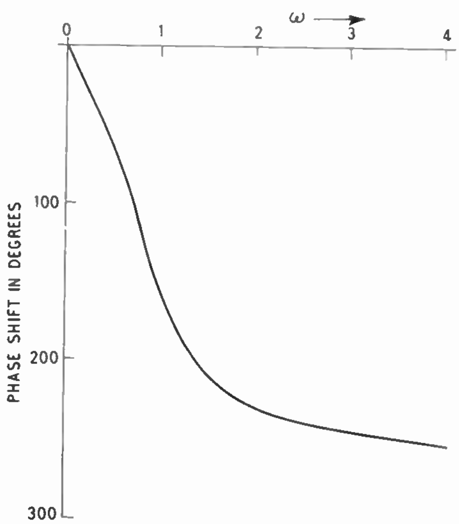
For the ideal phase response shown in Fig. 1(*b*)

$$\beta(\omega) = -\omega t_d \qquad \ldots\ldots(22)$$

where t_d is a constant.



(*a*) Magnitude response.



(*b*) Phase response.

Fig. 8. Normalized Papoulis class-L function.

Hence the ideal delay function is a constant equal to t_d.

By the application of the Fourier or Laplace transformations to the network function it can be shown that, when the magnitude and delay characteristics are constants, the output voltage V_2 is of the same form as the input voltage V_0 but delayed in time by t_d.

It is clear from equation (22) that in order to obtain an approximation to the ideal delay it is necessary to approximate to the linear phase characteristic. This can be done by the same method as was used to obtain the Butterworth characteristic. One approximates by expanding the delay function for the network and the ideal delay as Taylor series at one frequency $\omega = 0$. The coefficients of the terms in the two series are then equated; this means that all terms in the series for the delay of the network except the first are set equal to zero and the first term is equated to t_d.

Consider the three-pole case where the network function $N(p)$ has the form of equation (2).

The delay function for the three-pole network function can be found by differentiating $\beta(\omega)$ with respect to ω.

In the ideal case the delay is given by equation (23)

$$-\frac{\mathrm{d}\beta(\omega)}{\mathrm{d}\omega} = t_d \qquad \ldots\ldots(23)$$

Expanding as a series and equating terms gives the coefficients a_0, a_1 and a_2 in terms of t_d, which is usually normalized to unity.

These coefficients can be found more simply from the Bessell polynomials, the subject of a paper by L. Storch.[6]

It is known from the Laplace transform that the network function required to give pure time delay is

$$N(p) = \mathrm{e}^{-pt_d} \qquad \ldots\ldots(24)$$

If the delay is normalized to be one second ($t_d = 1$) then

$$N(p) = \mathrm{e}^{-p} = \frac{1}{\mathrm{e}^p} = \frac{1}{\cosh p + \sinh p} \qquad \ldots\ldots(25)$$

The terms $\cosh p$ and $\sinh p$ can be expressed as infinite series and the ratio $\dfrac{\cosh p}{\sinh p} = \cosh p$ as an infinite continued fraction

$$\cosh p = \frac{1}{p} + \cfrac{1}{\cfrac{3}{p} + \cfrac{1}{\cfrac{5}{p} + \cfrac{1}{\ldots}}} \qquad \ldots\ldots(26)$$

If this continued fraction is terminated at the $(2n-1)/p$ term one has an approximation to the network function e^{-p}. The denominator in the resulting network function is given by the Bessell polynomials as follows:

$$\text{denominator} = \sum_{K=0}^{n} \frac{(n+K)!\, p^n}{(n-K)!\, K!}\left(\frac{1}{2p}\right)^K \qquad \ldots\ldots(27)$$

where n is the number of degrees of freedom and $K = 0, 1, \ldots n$.

6.1. *The Three-pole Case*

When n is set equal to 3 the network function for the circuit shown in Fig. 2 becomes

$$N(p) = \frac{7{\cdot}5}{p^3 + 6p^2 + 15p + 15} \qquad \ldots\ldots(28)$$

6.2. *Bandwidth*

The equation (28) is normalized for unit delay and when p is set equal to $\mathrm{j}\omega$ the frequency for which the attenuation is 3 dB greater than its minimum value is $\omega = 1{\cdot}754$. In order to facilitate comparison with the other responses discussed in this article it is convenient to change the frequency variable to make $\omega = 1$ the frequency for 3 dB increases in attenuation.

When this change is carried out and the coefficient of p^3 set equal to unity, in the form of equation (2), the resulting network function is

$$N(p) = \frac{1{\cdot}399}{p^3 + 3{\cdot}429p^2 + 4{\cdot}898p + 2{\cdot}799} \qquad \ldots\ldots(29)$$

6.3. *Pole Locations*

The poles may be located by finding the roots of the denominator of equation (29) or, more simply, by dividing the tabulated poles of equation (28) by 1·754.

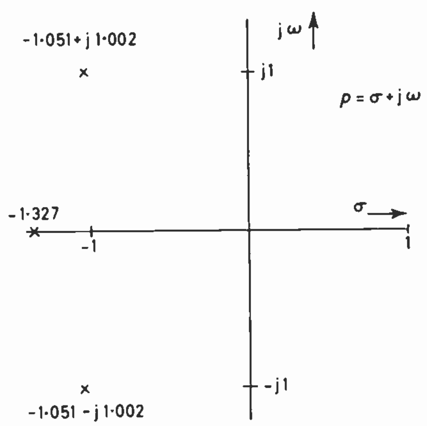
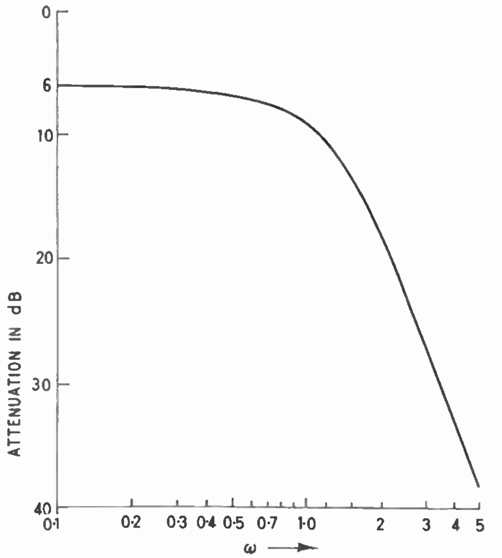
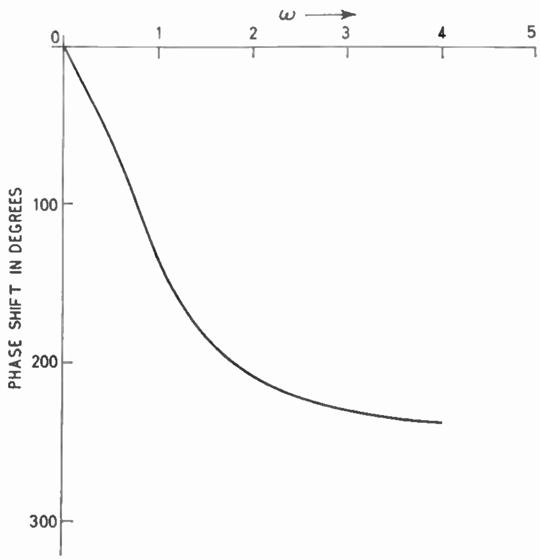


Fig. 9. Pole locations for maximally flat delay function.



(*a*) Magnitude response.



(*b*) Phase response.

Fig. 10. Normalized maximally flat delay function.

The resulting poles are:

$$\left.\begin{aligned} p_1 &= -\,1{\cdot}327 \\ p_2 &= -\,1{\cdot}051 + \mathrm{j}1{\cdot}002 \\ p_3 &= -\,1{\cdot}051 - \mathrm{j}1{\cdot}002 \end{aligned}\right\} \qquad \ldots\ldots(30)$$

These are shown plotted in the complex plane in Fig. 9.

6.4. *Magnitude and Phase Characteristics*

The magnitude and phase responses obtained from the pole plot in Fig. 9 are shown in Figs. 10(*a*) and (*b*).

7. In Line Pole Distribution

It has been shown by Guillemin[3] that if the poles of a network function are equally spaced in a linear array in the complex plane as shown in Fig. 11, the phase response differs from the ideal linear response in an equal ripple manner.

If the number of poles is very large and the interpole spacing $\Delta\omega$ is much greater than the real part σ, the average slope of the phase response is almost independent of σ and is approximately equal to $\pi/\Delta\omega$. Once the average slope is determined the ripple deviation can be controlled by choice of σ. A large value of σ reduces the ripple except at the limit of the pole array also increasing the frequency range for which the deviation from equal ripple becomes large.

7.1. *Pole Locations for the Three-pole Case*

It is evident that the three-pole example used in this paper will not provide a very close approximation to an infinite linear array of poles. However, it is instructive and interesting to arrange the three poles in a linear manner and to plot the magnitude and phase responses obtained.

It was arbitrarily decided to choose $\sigma = -1$ as one pole; this is the same as one of the pole locations for the maximally flat magnitude characteristic. The positions of the pair of complex poles was chosen to make the magnitude response 3 dB below the maximum value when $\omega = 1$.

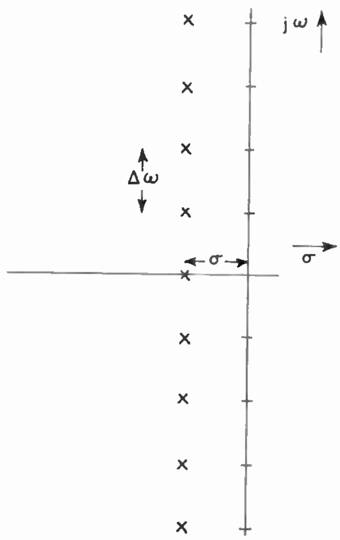


Fig. 11. Pole locations for in-line pole network function.

The three pole locations are:

$$\left.\begin{aligned} p_1 &= -1 \\ p_2 &= -1+j1{\cdot}239 \\ p_3 &= -1-j1{\cdot}239 \end{aligned}\right\} \quad \ldots\ldots(31)$$

Figure 12 shows these three poles plotted in the complex plane.

Writing $N(p)$ in the form of equation (2) gives

$$N(p) = \frac{1{\cdot}268}{p^3+3p^2+4{\cdot}535p+2{\cdot}535} \quad \ldots\ldots(32)$$

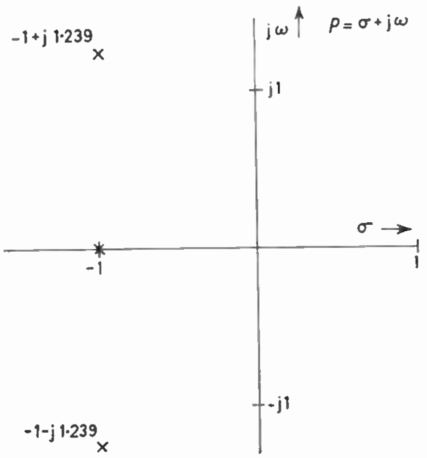


Fig. 12. Pole locations for 3-pole in-line pole network function.

7.2. *The Magnitude and Phase Characteristics*

The magnitude and phase responses plotted from Fig. 12 are shown in Figs. 13(*a*) and (*b*).

8. Conclusions from Network Function Responses

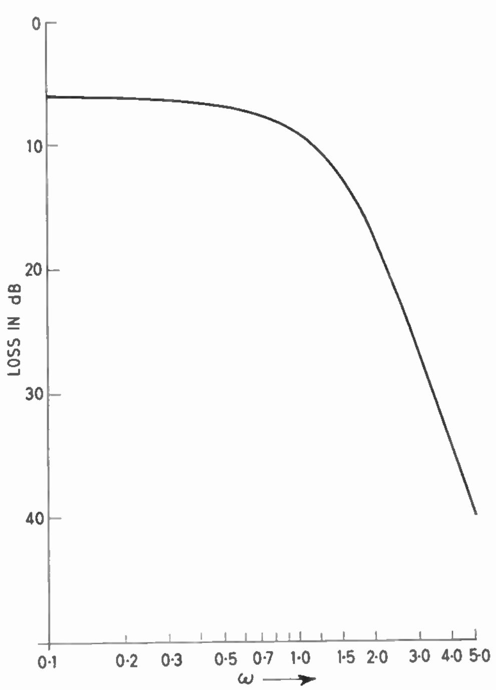
Figures 14 and 15 show the magnitude and phase responses for the five network functions considered. All the pole locations are plotted in Fig. 16 in order that the effects of changes in pole location on the responses may be seen.

The normalized element values required to give the five responses from the network in Fig. 2 are listed in Table 1.

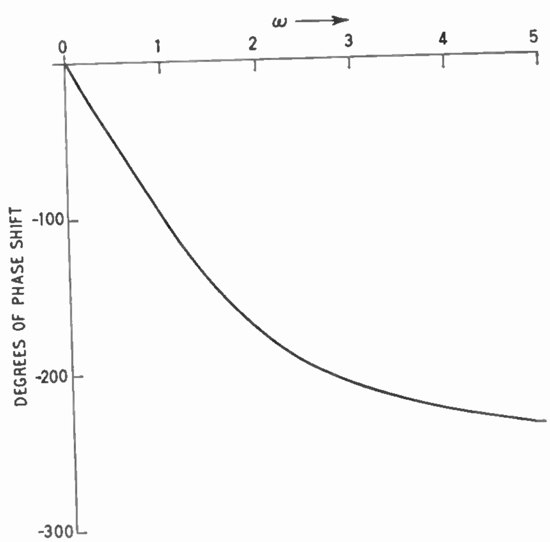
Table 1

Characteristic	L	C_1	C_2
Maximally Flat Magnitude	2	1	1
Chebyshev (3 dB ripple)	0·7117	3·3487	3·3487
Papoulis Class L	1·353	2·18	1·175
Maximally Flat Delay	0·5528	0·1922	1·255
In Line Poles	0·876	0·3895	2·312

Consider the changes in magnitude response first. It is evident from examination of Fig. 14 that the increase in attenuation for frequencies just outside the pass band is much greater for those network functions which were designed by concentrating upon the magnitude characteristic than for those which were designed to approximate to a linear phase response. This is as one would expect. When the poles are moved further away from the $j\omega$ axis in Fig. 16 the attenuation for frequencies just outside the pass



(*a*) Magnitude response.



(*b*) Phase response.

Fig. 13. Normalized in-line pole function.

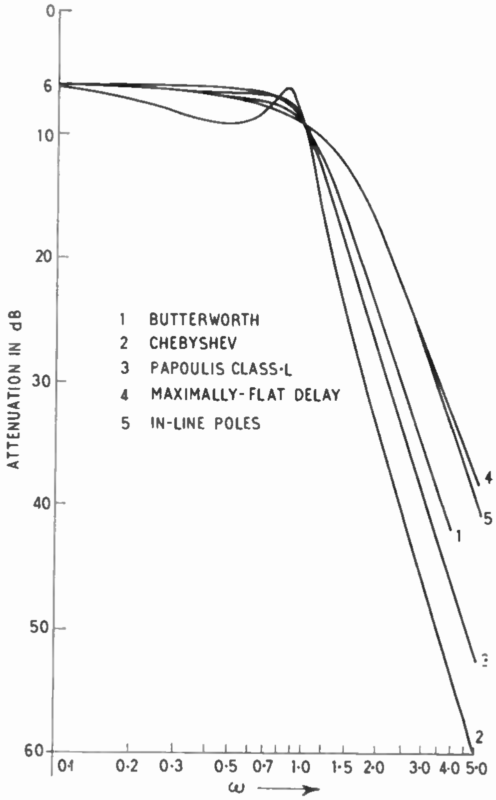


Fig. 14. Magnitude responses.

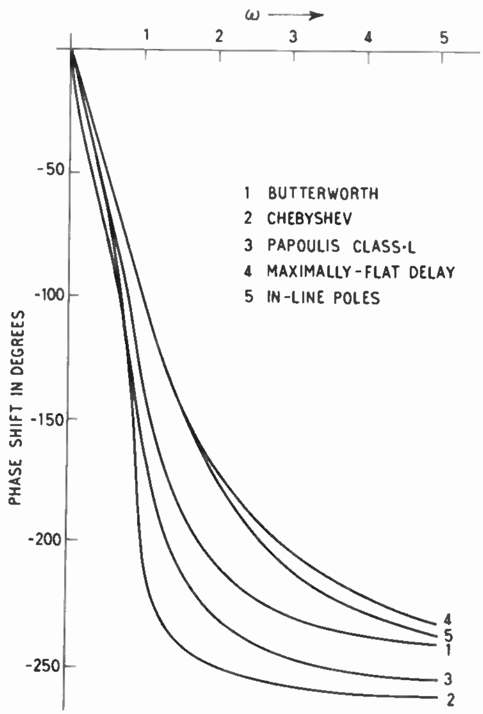


Fig. 15. Phase responses.

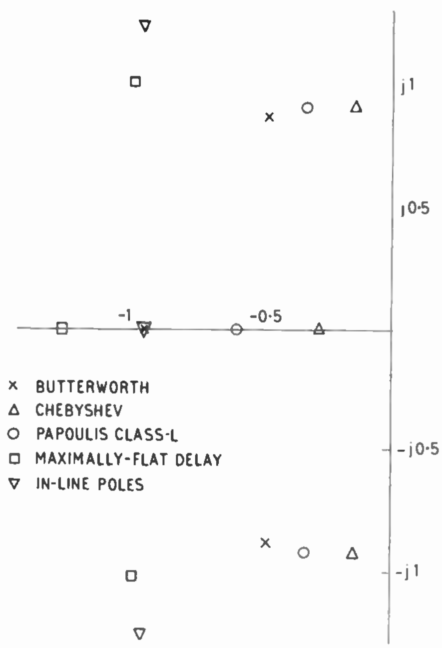


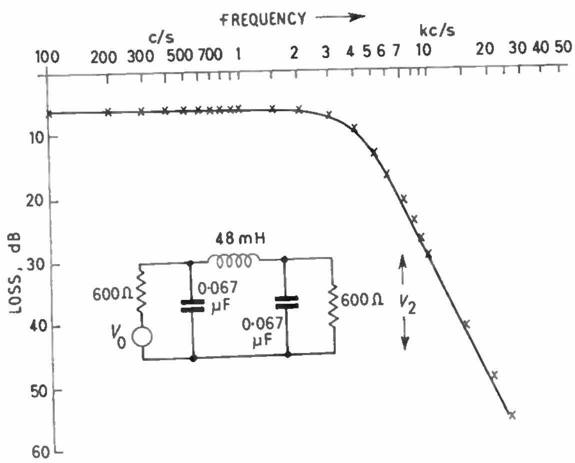
Fig. 16. Pole locations for the five designs.

band decreases. Again this is a predictable result because when the poles are close to the jω axis both the length and phase angle of the vector from pole to the jω axis vary rapidly when the frequency is changed. For frequencies far removed from the pass band the attenuation increases at 18 dB/octave, as is to be expected from a three-pole network function. It is to be noted, however, that the presence of the multiplying factor H in equation (2), which has different values in the numerators of each of the network functions equations (10), (16), (20), (29) and (32), introduces an additional term into the expressions for the attenuation produced by the network. The result of this term is for the spacing between the curves in Fig. 14 to differ from that which would appear if the constants were all equal to unity. This is the case often quoted. If this were so for the responses in Fig. 14 the spacing between the maximally flat magnitude and the Chebyshev curves for $\omega = 4$ would be 6 dB instead of approximately 12 dB as is plotted.

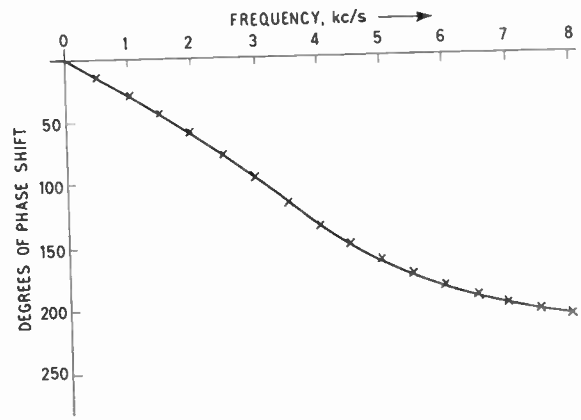
It is evident from the curves that if a magnitude characteristic having the greatest increase in attenuation is required and the effects of a ripple in the pass band can be tolerated, the Chebyshev network function with ϵ approximately equal to unity is the most desirable characteristic of those plotted. The 3 dB ripple amplitude is greater than can be tolerated in some applications and smaller ripple amplitudes may be employed, say 1/2 or 1/10 dB, but these are accompanied by a much smaller increase in attenuation for frequencies just outside the pass band.

Note that the network function giving the Chebyshev response has poles closer to the $j\omega$ axis than any of the other characteristics. Small differences between the desired and the actual positions of the poles have a greater effect upon the resulting magnitude and phase responses when the pole to $j\omega$ axis spacing is small than when it is large. Hence if it is important to realize a designed magnitude and phase characteristic exactly more care must be taken in finding the component values for the Chebyshev than for the other designs. It may be necessary to make some allowance for the shift in the desired pole locations due to the effects of dissipation in the circuit elements, particularly in the inductor. Details of a method for making this allowance are given by Guillemin.[3]

When the network in Fig. 2 is to have pulses applied to it and the output waveform is to be observed on a cathode ray tube, the phase response becomes important and one of the network functions giving a close approximation to a linear phase shift may be chosen. These will produce less overshoot in their response when a step function is applied at the input than would be obtained if the poles were moved closer to the $j\omega$ axis to give, for example, the Chebyshev response.

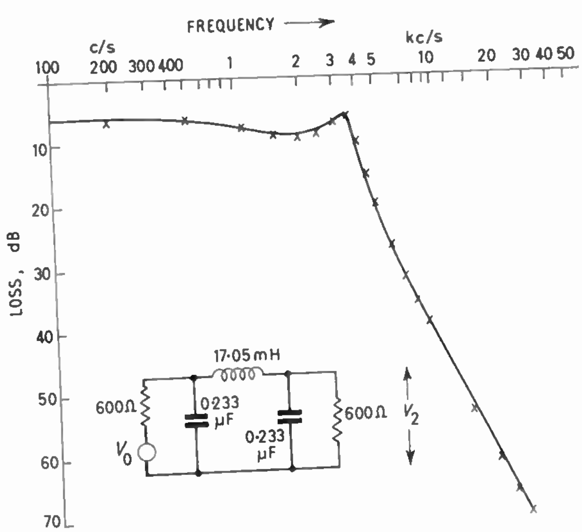


(a) Magnitude response.

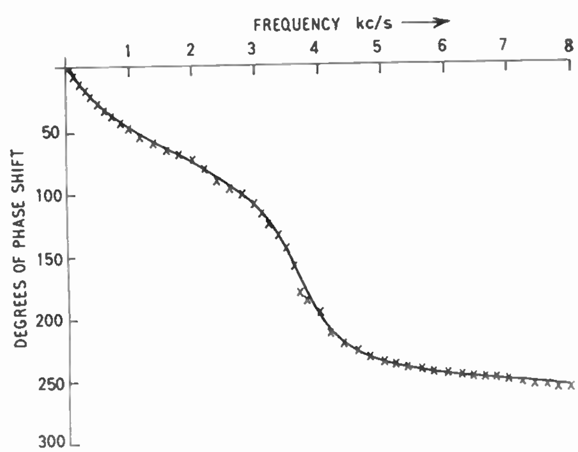


(b) Phase response.

Fig. 17. Butterworth function.

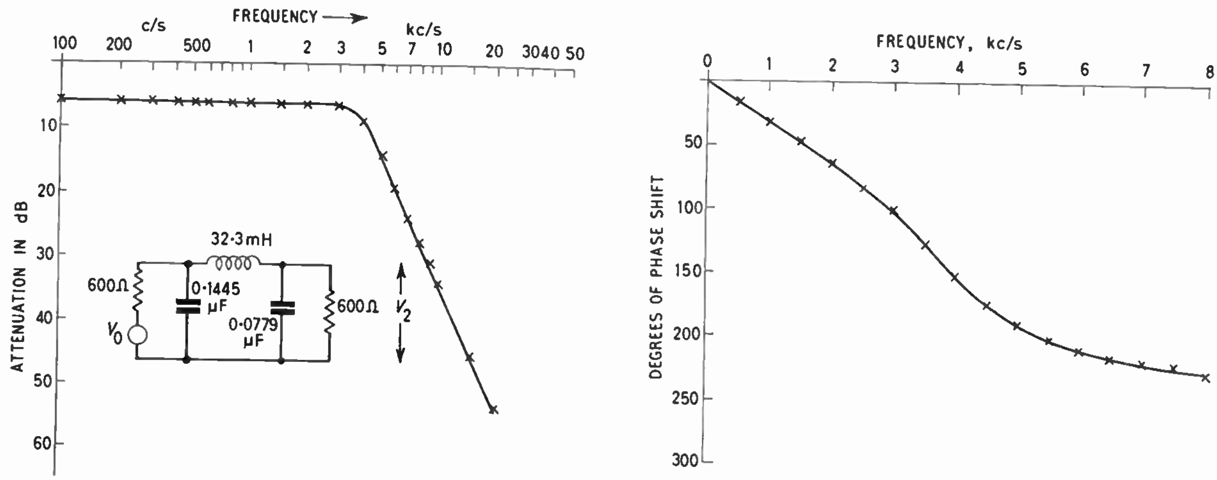


(a) Magnitude response.



(b) Phase response.

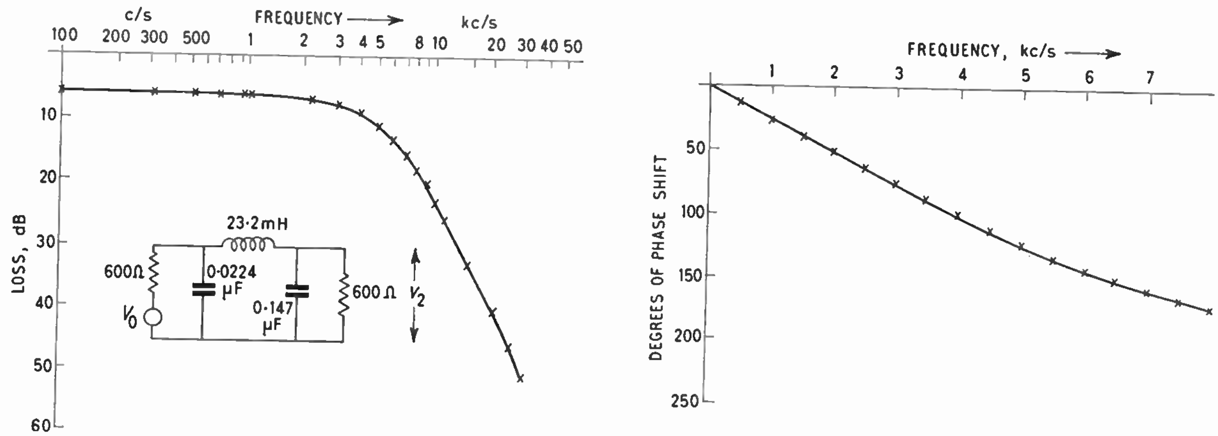
Fig. 18. Chebyshev function.



(*a*) Magnitude response.

(*b*) Phase response.

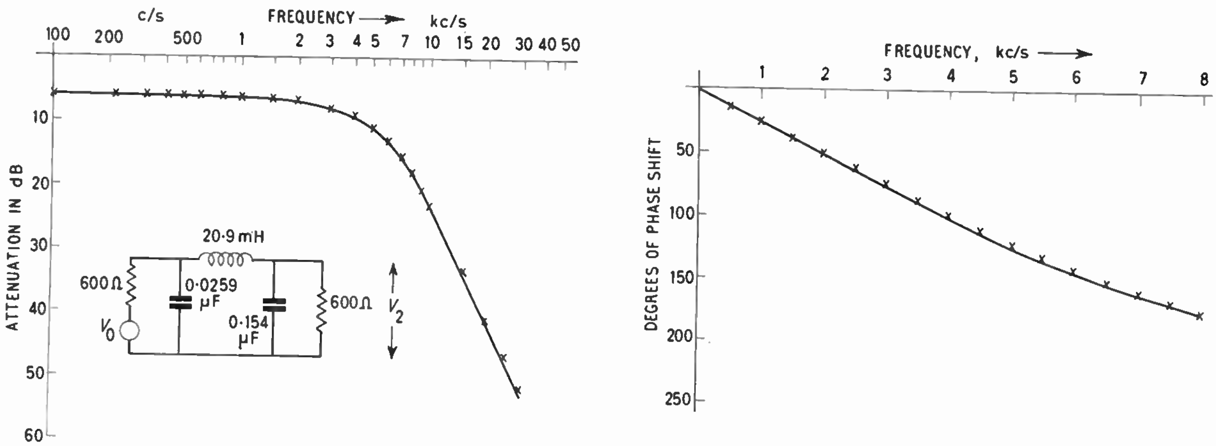
Fig. 19. Papoulis class-L function.



(*a*) Magnitude response.

(*b*) Phase response.

Fig. 20. Maximally flat delay.



(*a*) Magnitude response.

(*b*) Phase response.

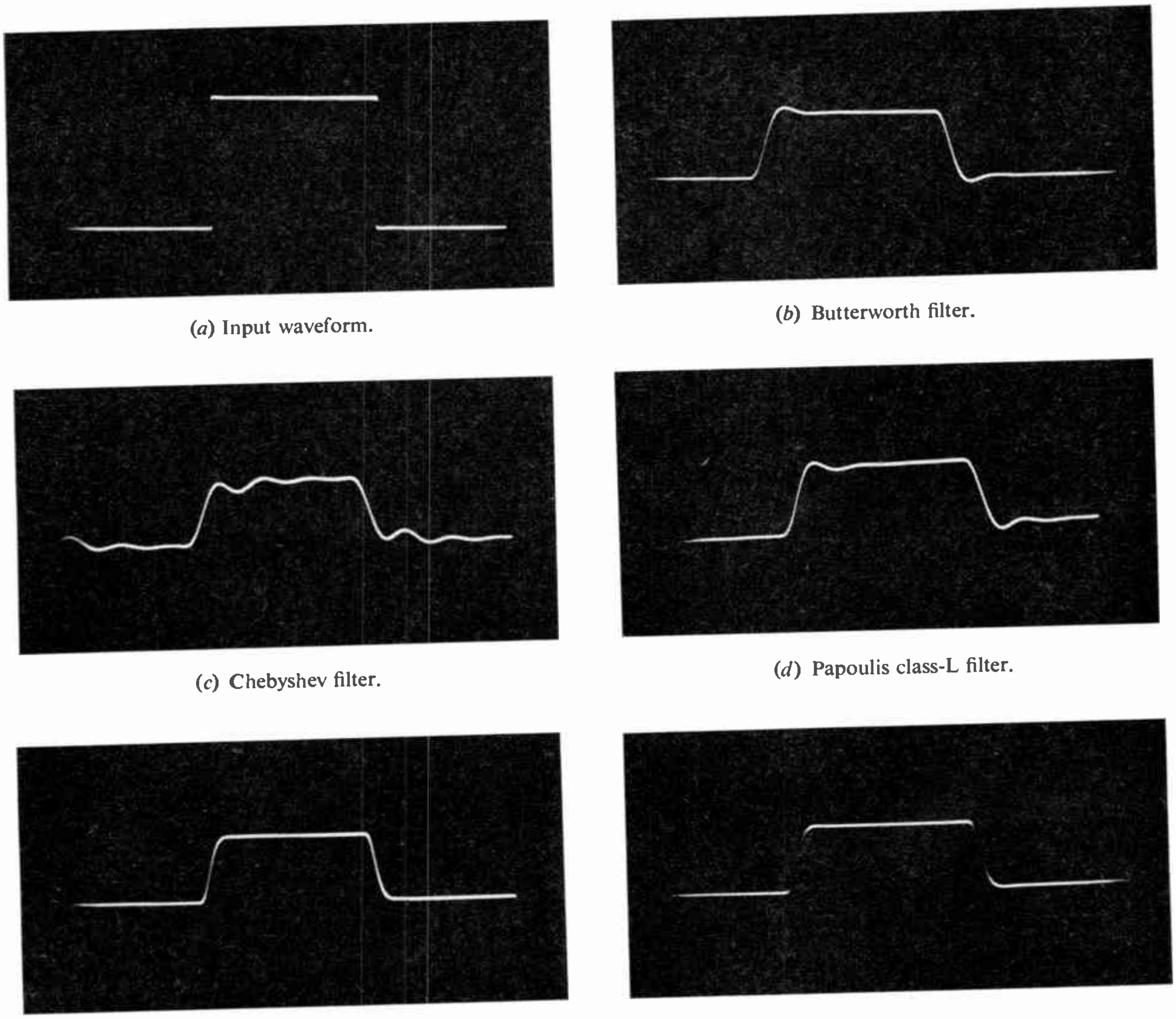
Fig. 21. In-line poles.

The Chebyshev, Butterworth and the Papoulis Class L characteristics show appreciable deviation from the linear phase response in the pass band and may be expected to produce more overshoot when a step function is applied than would be obtained from the 'maximally flat delay' or 'in line pole' characteristics.

The Papoulis Class L characteristic gives the greatest slope of attenuation at the limit of the pass band of all the magnitude responses not having any ripples in the pass band. As is to be expected from the pole locations shown in Fig. 9, the Class L response is an intermediate case between the 3 dB ripple Chebyshev and the 'maximally flat magnitude' characteristics.

The magnitude responses for the 'maximally flat delay' and 'in line pole' network functions are so close to being identical that they can be discussed together. Both have a ripple free form and for both the rate of increase of attenuation just outside the pass band is small, much less than is offered by the Papoulis Class L characteristic. It is clear that if a magnitude response having large attenuation outside the pass band is required neither of these characteristics would be suitable. For ($\omega = 4$) four times the frequency marking the limit to the pass band they offer some 20 dB of attenuation less than can be obtained with the 3 dB ripple Chebyshev design.

It is evident from the phase responses in Fig. 15 that the two characteristics designed to give a linear



(*a*) Input waveform.

(*b*) Butterworth filter.

(*c*) Chebyshev filter.

(*d*) Papoulis class-L filter.

(*e*) Maximally flat delay filter.

(*f*) In-line pole filter.

Fig. 22 (*b*) *to* (*f*). Output waveforms.

phase response do give a much closer approximation to the ideal straight line than can be obtained from any of the other responses. It is interesting to note that both of the linear phase approximations have their poles positioned considerably further from the $j\omega$ axis than any of the magnitude approximations. Figure 15 shows that as the poles move closer to the $j\omega$ axis the deviation from linearity increases until the Chebyshev response displays a ripple of considerable magnitude in the pass band. All the phase responses tend to the limiting value of $-\frac{\pi}{2}n$ or -270 deg, as to be expected from a three-pole network function.

9. Experimental Results

Five filter networks were constructed, all of the form of the simple circuit in Fig. 2 and having the normalized element values given in Table 1. The impedance level of the networks was raised from 1Ω to 600Ω, this was done by multiplying all impedances by a factor of 600; the frequency at which the attenuation (defined as 20 log V_0/V_2) was 9 dB was raised from $\omega = 1$ to $\omega = 2\pi \cdot 4000$ by dividing the L and C element values by $2\pi \cdot 4000$.

Measurement of the magnitude and phase responses of these networks were carried out; the results are shown in Figs. 17 to 21. When these responses are compared with those predicted from the equations given in the earlier sections of the paper it is evident that the agreement between theory and experiment is very close.

In order to emphasize the effect of different network designs on the pulse response of the filters square waves were applied to the filter inputs, care being taken to ensure that the impedance presented to the filter by the source was of the correct value. Figure 22 shows the input and output waveforms for an input frequency of 500 c/s. These figures show clearly how the degree of distortion of the pulse passing through the filter is changed, and particularly how it is reduced, when networks designed to have a good approximation to the linear phase response are used.

10. Acknowledgments

The author has pleasure in thanking Professor R. L. Russell, Professor of Electrical Engineering at the University of Newcastle upon Tyne, for his encouragement and for the use of the facilities of the Electrical Engineering Department.

11. References

1. S. Darlington, "Synthesis of reactance 4-poles which produce prescribed insertion loss characteristics including special applications to filter design", *J. Maths. Phys.*, **18**, pp. 257–353, September 1939.
2. S. Butterworth, "On the theory of filter amplifiers", *Exptl Wireless & Wireless Engr*, **7**, pp. 536–41, October 1930.
3. E. Guillemin, "Synthesis of Passive Networks", (John Wiley, New York, 1957).
4. A. Papoulis, "Optimum filters with monotonic response", *Proc. Inst. Radio Engrs*, **46**, pp. 606–9, March 1958.
5. E. Jahnke and F. Emde, "Tables of Functions with Formulae and Curves", 4th Edn. (Dover Publications, New York, 1945).
6. L. Storch, "Synthesis of constant-time-delay ladder networks using Bessel polynomials", *Proc. Inst. Radio Engrs*, **42**, pp. 1666–75, November 1954.

Manuscript first received by the Institution on 11th February 1963, in revised form on 11th August and in final form on 2nd November 1963. (Paper No. 886.)

A Phase Modulation Data Transmission System for use over the Telephone Network

By

A. P. CLARK, M.A.†

Summary: A synchronous system is described which transmits binary-coded digital-information at 750 or 1500 bauds over normal telephone circuits. It uses a phase-modulated carrier of 1500 c/s and both sidebands are transmitted. The receiver employs differentially-coherent detection to overcome the one serious weakness of coherent detection, which is the low tolerance of the latter to the frequency-modulation effects sometimes experienced over telephone-carrier circuits.

The results obtained after using the system for more than two years over the switched telephone-network, together with the results of many laboratory tests, show a very good performance under all the various conditions of noise and distortion likely to be experienced over telephone circuits.

A careful comparison between this system and equivalent f.m. systems of optimum design, has shown clearly that in applications over telephone circuits, p.m. systems of this type are superior to the equivalent f.m. systems in every important respect, both technically and economically.

List of Symbols and Definitions

SWIFT = switched phase transmission

f.m. = frequency modulation

p.m. = phase modulation

A 'signal element' is a unit signal pulse.

A 'synchronous system' is a system in which the sending and receiving equipments are operating continuously at substantially the same number of signal elements per second and are maintained, by correction, in the desired phase relationship.

A 'baud' is the unit of modulation rate.

In a synchronous system the modulation rate in bauds is equal to the number of signal elements per second, which is the signalling speed.

1. Introduction

An experimental study of the noise and distortion characteristics of telephone circuits,[1] followed by a theoretical study of the tolerance to these characteristics of different modulation-methods, showed that the medium-speed data-transmission system giving the best overall performance over telephone circuits for a given degree of complexity in the design, should be a synchronous system using a single phase-modulated carrier with a binary double-sideband signal and employing differentially-coherent detection at the receiver. These conclusions are in general fully supported by the much more detailed theoretical and practical investigations which have since been carried out by other workers in this field.[2]

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

A phase-modulation data-transmission system, which is known as SWIFT, was designed in accordance with these conclusions. SWIFT has been in use for commercial applications since the middle of 1961,[3, 4] and considerable experience has been gained in its performance, not only as a result of its commercial use over the switched telephone-network but also as a result of many tests both in the laboratory and over various switched and private telephone-circuits. These tests have fully confirmed the original conclusions on which the design of the system was based.

2. Method of Operation

2.1. *Basic Outline of the Complete System*

Figure 1 shows a complete signalling-channel using SWIFT data-transmission equipment. This is a synchronous system in which the rate of transmission is determined by the SWIFT transmitter and is either 750 or 1500 bauds, using a binary-coded signal.

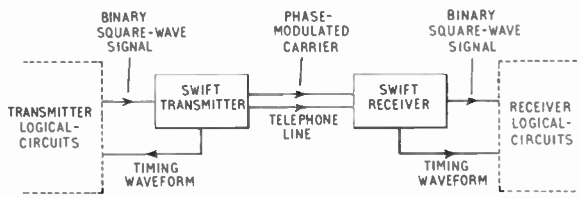


Fig. 1. A SWIFT data-transmission-channel.

A timing waveform synchronized to the signal carrier is fed from the transmitter to the associated logical-circuits, where it is used to synchronize the binary-square-wave signal which is fed from this equipment to the SWIFT transmitter. The transmitter converts the binary-square-wave signal into the corresponding phase-modulated carrier. In the latter signal a '1' is transmitted as a 180 deg phase shift in the carrier between two adjacent elements and a '0' as no phase shift in the carrier between these elements, the carrier being synchronized to the modulating waveform. A SWIFT transmitter operating at 1500 bauds uses exactly one complete cycle of the 1500 c/s carrier per signal element, and a SWIFT transmitter operating at 750 bauds uses exactly two complete cycles.

For operation over switched telephone-circuits a signalling speed of 750 bauds is used, occupying a frequency band from about 1000 to 2000 c/s, whereas over private circuits a speed of 1500 bauds may be used, occupying a frequency band from about 500 to 2500 c/s. Each of these frequency bands is ideally placed in the available bandwidth.

The SWIFT receiver converts the received phase-modulated carrier back again into an undistorted binary-square-wave signal, identical to that fed to the transmitter at the other end of the line. This square-wave signal is fed together with the associated timing-waveform to the receiver logical-circuits.

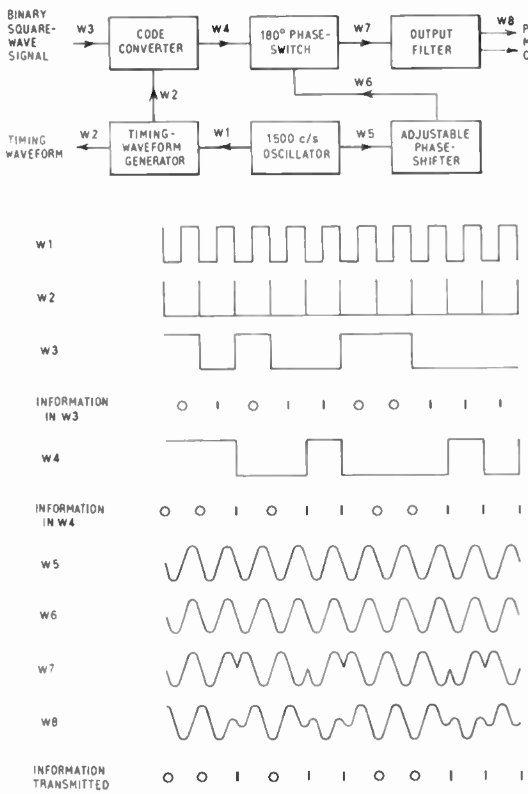


Fig. 2. SWIFT transmitter.

2.2. SWIFT *Transmitter*

Figure 2 shows the block diagram and associated waveforms for the 1500 baud transmitter (type SFT2). All the waveforms in the transmitter are originally derived from a 1500 c/s oscillator which gives both a sine-wave output W5 and a square-wave output W1, these waveforms being accurately in phase. The timing-waveform W2 is derived from W1 and the transmitted signal-carrier from W5. The binary-square-wave signal W3 is the signal fed to the SWIFT transmitter from the associated logical-circuits. W3 is synchronized to W2 and represents a '1' as a negative level and a '0' as a positive level. The code converter changes a '1' in the waveform W3 into a change in level and a '0' into no change in level, giving the waveform W4 which acts as the modulating waveform in the 180 deg phase-switch. The latter circuit passes without change the sine-wave carrier W6 when W4 is positive and inverts the sine-wave carrier when W4 is negative. In the waveform W7 a '1' is thus represented as a 180 deg phase shift between two adjacent elements and a '0' as no phase shift between these elements. W7 is filtered to restrict somewhat the width of the transmitted frequency-spectrum, giving the waveform W8 which is fed to line.

Because the SWIFT transmitter uses a single oscillator both to generate the signal carrier and to determine the modulation rate, not only is the complexity of a second stable oscillator avoided but also the transmitted signal has the advantage of the carrier being synchronized or phase-locked to the modulating waveform. Furthermore, the square-wave signal W4 at the output of the code converter, is used directly as the modulating waveform. Thus, not only is a low-pass filter not needed here to remove the high-frequency components of the modulating waveform, but also a simple balanced-gate-circuit (here called the 180 deg phase-switch) can be used for the modulator instead of the more complex balanced-linear-modulator which must always be used with a band-limited and therefore rounded modulating-waveform.

The fact that in the SWIFT transmitter the modulating waveform has frequency components higher than the carrier frequency, causes foldover of the frequency spectrum upon modulation. If there were no phase coherence between the carrier and modulating waveform, this would inevitably result in a much wider frequency-spectrum at the modulator output, thus requiring more expensive filtering to limit the transmitted frequency-spectrum to an acceptable bandwidth. However, with the phase coherence between the carrier and modulating waveform obtained in the SWIFT transmitter, there is a marked tendency for the energy to be concentrated into certain parts of the frequency spectrum and reduced in other parts. This is particularly evident at the higher modulation-rate where the degree of foldover of the frequency spectrum upon modulation is correspondingly greater. Furthermore, the portions of the frequency spectrum where the energy is increased and the portions where it is reduced, are dependent on the phase relationship between the carrier and modulating waveform. Thus by providing a means for adjusting the carrier phase, a very versatile arrangement has been produced which has the properties of a variable filter with a linear phase-characteristic.

The additional filtering required at the transmitter output, because the modulating waveform is rectangular instead of rounded in shape, is certainly no greater than that which would alternatively be required to remove the high-frequency components from the modulating waveform. In the latter arrangement, however, because there would no longer be any foldover of the frequency spectrum upon modulation, the frequency spectrum of the modulator output-signal would be independent of the carrier phase. Adjustments in the carrier phase could therefore no longer be used to optimize the energy distribution in the output signal-spectrum.

The adjustable phase-shifter enables the phase of the signal carrier W6 to be set either for no phase shift as shown in Fig. 2, for which the larger part of the energy in W7 lies below the 1500 c/s carrier frequency, or the carrier phase may be shifted by 90 deg so that the transition between one signal element and the next always occurs at a peak of the carrier, for which arrangement the larger part of the energy in W7 lies above the 1500 c/s carrier, or again the carrier phase may be shifted by 45 deg for which the frequency spectrum of W7 is approximately symmetrical about the 1500 c/s carrier. Thus using a very simple phase-shifting network an appreciable filtering action is obtained which not only introduces no signal distortion but is readily adjustable so that the bulk of the transmitted energy can be located in the best part of the available frequency-band.

The 750-baud transmitter SFT4, which is designed for use over switched telephone-circuits, is the same as the 1500-baud transmitter, SFT2, except that the timing-waveform W2 has a positive-going pulse at every second negative-going edge of W1, giving two cycles of the carrier per signal element in W7 and W8. Also the adjustable phase-shifter is always set for a 90 deg phase-shift, thus enabling an adequate reduction of signal power to be achieved in the frequency band 420 to 900 c/s with only the minimum additional attenuation by the output filter. Over switched telephone-circuits in Great Britain there is a considerable restriction on the power which may be transmitted in the band 420 to 900 c/s.

The transmitter SFT3 can be set to give either 750 or 1500-baud operation, depending on whether or not a bistable circuit arranged as a divide-by-two stage is switched into the signal path at the input of the timing-waveform generator.

Two terminals are provided at each SWIFT transmitter which when shorted together cause the suppression of the transmitted signal. This facility enables no-signal to be used if required as a stand-by condition.

2.3. SWIFT *Receiver*

Figure 3 shows the block diagram and associated waveforms for the 1500-baud receiver SFR2. The received signal-waveform W8 is first filtered to remove as far as reasonably possible any noise frequencies outside the signal frequency-band. It is then demodulated in a differentially-coherent detector which comprises the delay network, the two amplifier-limiters and the modulo-2-adder circuit.

The delay network is a seven-section all-pass network which produces an accurately constant delay over the whole of the signal frequency-band and therefore introduces negligible signal distortion. Each of the two high-gain amplifier-limiters slices the input waveform along a very narrow and accurately centred horizontal-section, to produce an output square-wave signal clamped between two fixed voltage-levels. Correct slicing of the received signal is maintained not only over a very wide range of signal levels but also during the period including and following any variation in the level within this range. The various disadvantages associated with automatic gain-control[2] are thus avoided and the receiver will tolerate large and sudden signal-level-changes.

The receiver is designed to give correct operation for signal levels down to -50 dBm with only a small change in the acceptable signal/noise ratio, correct operation being maintained in the absence of noise and distortion for received signal levels down to about -70 dBm. This value is raised to -65 dBm for the models of the SWIFT receiver which are designed for use over private telephone-circuits.

The delay used in the delay network is exactly equal to the duration of one signal element. The original and delayed signal waveforms, after each has been sliced in an amplifier-limiter, are compared in a standard modulo-2-adder circuit whose output-waveform W12 is negative whenever the two inputs are different and positive whenever these are the same. W12, although ideally as shown in Fig. 3, will often have a number of fine irregularities due to the effects of signal distortion. The low-pass filter, which has a linear phase-characteristic, performs an approximate process of integration over the duration of each signal element in W12 and also removes any fine irregularities. Its output signal is the rounded waveform W13. The waveforms W12 and W13 as shown in

Fig. 3 are inverted with respect to the corresponding waveforms in a practical SWIFT receiver. This change has been made to clarify the description and it in no way affects the fundamental method of operation. The waveform W13 is fed to the slicer which amplifies a very narrow horizontal-section of the waveform to give the square-wave signal W14. The vertical level at which W13 is sliced, determines the threshold level below which W13 represents '1' and above which it represents '0'.

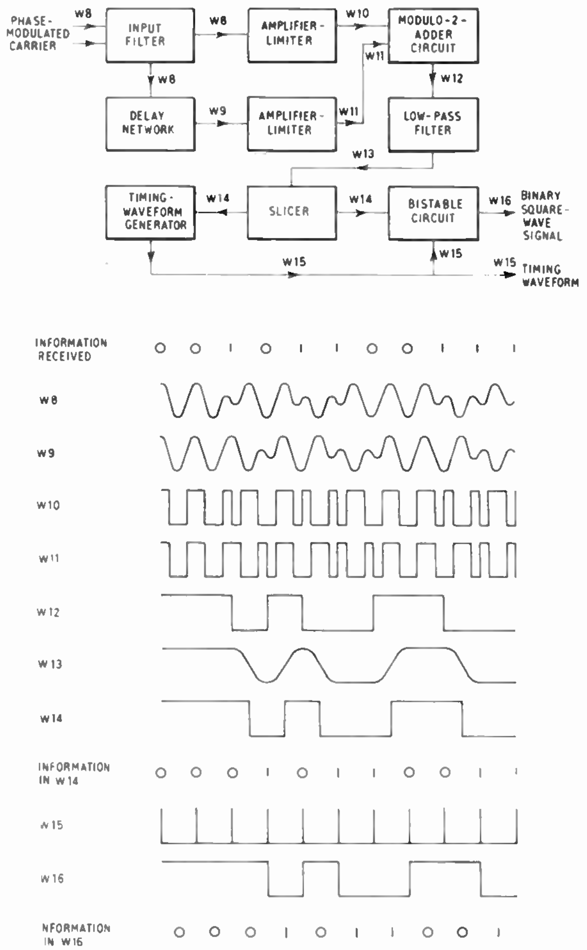


Fig. 3. SWIFT receiver.

Each signal element is therefore directly compared with that immediately preceding it. If at the end of this comparison of any pair of signal elements it is decided that the two were on the average the same, then at this instant in time a '0' is indicated in the waveform W14 at the slicer output. If it is decided that they were on the average different, a '1' is indicated here. Since a '1' is transmitted as a phase shift between two adjacent signal-elements and a '0' as no phase shift between these elements, the received signal is thus not only detected but also reconverted to the original code in which a '1' is represented by a negative level and a '0' by a positive level.

The advantage of this method of detection is that it overcomes the one serious weakness of the more conventional coherent-detection, which is the low tolerance of the latter to the frequency-modulation effects sometimes experienced over telephone-carrier circuits.[2] In coherent detection a continuous sine-wave is extracted from the received signal and used as a phase reference for this signal, whereas in differentially-coherent detection the received signal is used as its own phase reference, as described above. A typical receiver using coherent detection for phase-modulated signals will tolerate a frequency translation of the transmitted signal-spectrum not exceeding ± 10 or 20 c/s. Practical tests, however, have shown that the 1500-baud SWIFT receiver will tolerate a frequency translation of the transmitted signal-spectrum of ± 100 c/s with only a small reduction in tolerance to noise and distortion, and in the absence of noise and distortion it will tolerate a frequency translation of a little more than ± 200 c/s. In return for this, when there are no frequency-translation effects on the telephone circuit, the SWIFT receiver has a tolerance to white noise which is nearly 1 dB lower than that of the equivalent receiver using coherent detection. It also has a lower tolerance to certain types of attenuation and delay distortions. Neither of these effects however is serious enough to be of much importance in practice.

By comparing the original and delayed signal-waveforms, after these have been sliced and amplified to give the rectangular-waveforms W10 and W11, it has been possible to use a very simple modulo-2-adder circuit to act as the phase-sensitive detector or product demodulator. This is a well known logical-gate circuit which is used here in the conventional manner. The two binary input signals, W10 and W11, each have at any time one of two fixed voltages, the periods spent in transitions between one voltage and the other being normally of extremely short duration in comparison with the periods spent at either voltage. Whenever the two input signals are either both positive or both negative, the output signal W12 from the modulo-2-adder is positive. Whenever one of the two input signals is positive and the other negative, the output signal W12 is negative. Thus, as in the 180 deg phase-switch in the transmitter, the use of a linear-modulator circuit has been avoided here. This eliminates the use of linear amplifiers which must have stable characteristics and a low level of harmonic distortion. It also provides a design whose practical performance can be made to approximate closely to the theoretical ideal while using at the same time a relatively simple arrangement with standard circuit components.

The timing waveform in the receiver is extracted from the detected-signal waveform W14 at the slicer output, no additional signal being transmitted for the timing information. This arrangement gives the greatest tolerance of the data-transmission system to noise and distortion, together with the most economical design.[2] The timing-waveform generator converts the detected-signal waveform W14 into a series of approximately triangular d.c. pulses whose fast leading edges are coincident with the positive and negative transitions in W14. It then uses a high-Q resonant circuit to extract from the triangular pulses the frequency component corresponding to the signal modulation-rate. The sine-wave signal obtained is fed to a high-gain amplifier-limiter similar to those used in the differentially-coherent detector. The output square-wave from this circuit is accurately symmetrical and correctly phased over a very wide range of input levels and it is used to generate the timing-waveform W15. This arrangement has a high effective-Q during normal transmission, so that before any appreciable jitter can be produced in the timing waveform by noise, the received information will have been almost completely destroyed. It also has the great advantage in that after first receiving a signal following a no-signal condition on the line, it only requires a very short period, containing no more than three or four signal elements, for correctly phasing the timing waveform. In other words, for a very short period after first receiving a signal from line, the effective Q of the resonant circuit is greatly reduced, thus enabling the phase of the timing waveform to be rapidly adjusted. The one limitation of this design is that the correct phase of the timing waveform will only be maintained in every case, over the full ambient-temperature range of 0° to +55° C, if the maximum number of consecutive '1's' or '0's' in any signal pattern is limited to between 20 and 40, depending on the particular component tolerances used. This requirement can however easily be satisfied with probably the majority of practical signal-codes.

At the input of the bistable circuit, the timing-waveform W15 is used to sample the detected-signal waveform W14 at the instants corresponding to the end of each comparison of a pair of adjacent elements. The output signal from the bistable circuit is the undistorted binary-square-wave signal W16 which is fed together with the timing-waveform W15 to the receiver logical-circuits.

There are two different 750-baud SWIFT-receivers which are known as SFR4 and SFR5. SFR4 is similar in all respects to the 1500-baud receiver, except that it is designed for signal elements each containing two cycles of the 1500 c/s carrier. The delay network here therefore produces twice the delay and there are corresponding changes in the low-pass filter and in the resonant circuit of the timing-waveform generator.

The 750-baud receiver SFR5 is identical to the 1500-baud receiver except for the timing-waveform generator which produces a 750 pulse-per-second waveform as in SFR4. The receiver in fact detects the received 750-baud signal as though it were a 1500-baud signal but only samples every second element, neglecting those which correspond to a comparison of the first and second halves of any element, these being of course all '0's'. The advantage of this receiver over SFR4, when working over telephone circuits, is that there is a larger preponderance of single errors, thus enabling a more useful reduction in the error rate to be achieved by a relatively simple single-error-correcting code. SFR5 also has the same tolerance as SFR2 to frequency-modulation effects on the telephone circuit and each of these receivers will tolerate a frequency translation of the received signal-spectrum, having twice the magnitude of that tolerated by SFR4. The performance of SFR4 in this respect is however more than adequate for practical purposes. Again SFR5 has a more economical design in that its delay network only produces half the delay of that in SFR4. The main disadvantage of SFR5 is that its tolerance to attenuation and delay distortions in the transmission path is considerably lower than that of SFR4. Its tolerance to white noise in the absence of signal distortion is also some 2 dB lower. The tolerance of SFR5 to noise and distortion is however still quite adequate for reliable operation over the good majority of switched telephone-circuits. A further disadvantage of SFR5 when compared with SFR4 and SFR2, is that for correct operation over the full ambient-temperature range of 0° to +55° C, the maximum number of consecutive '0's' in any signal pattern is limited to between 10 and 20, as compared with 20 and 40 for both SFR4 and SFR2. On the other hand, correct operation over the full temperature-range is obtained with SFR5 for any number of consecutive '1's' in the received signal.

The SWIFT receiver SFR3 provides a choice of 1500 or 750-baud operation and is in fact a combination of SFR2 and SFR5. Since the only difference between SFR2 and SFR5 lies in the timing-waveform generator, the only additional components used in SFR3, when compared with either SFR2 or SFR5, are a high-Q resonant circuit together with a few other minor components and of course the switch. One of the valuable features of SFR3 is therefore that it has a most economical design for a two-speed receiver.

The SWIFT receivers SFR4 and SFR5 are designed for use over switched telephone-circuits and the receivers SFR2 and SFR3 are designed for use over private circuits. In SFR4 and SFR5 no delay equalizers are used either for the transmitter output and receiver input filters or for the telephone circuit. In SFR2 and SFR3 a delay equalizer is added to the receiver input-filter to equalize the group delay of the transmitter output and receiver input filters, and a further delay-equalizer, which may have one of four different group-delay characteristics, may also be added to give partial delay-equalization for the telephone circuit. The additional delay-equalization has been made available, not only to ensure correct operation at 1500 bauds over very inferior private circuits, but also to enable reliable 1500 baud operation to be obtained over switched circuits in those countries where the permissible transmitted-signal-level in the frequency band 420 to 900 c/s is the same as that in the rest of the available bandwidth.

A special circuit in each SWIFT receiver indicates the loss of the incoming signal if the level drops below a predetermined value (normally between −40 and −50 dBm) for more than a few milliseconds. The value of this 'off time-delay', as measured for the filtered input-signal, is around 20 milliseconds for SFR4 and SFR5 whereas for SFR2 and SFR3 it is about 5 milliseconds. When correctly used this circuit prevents undetected errors being caused by all but the most brief disconnections in the telephone line. The circuit also enables no-signal to be used if required as a stand-by condition between the transmission of different messages.

Each SWIFT transmitter and receiver is protected against damage from large current surges on the line. The equipment may therefore safely be used in those countries where lightning and other sources of high-level interference, are a serious hazard over telephone circuits.

2.4. *Acknowledgment Signal*

When data are transmitted between two terminals in only one direction, one of the facilities which is nearly always required is an automatic means of informing the transmitting terminal from the receiving terminal as to whether or not the data are being correctly received.

Provision has been made at the SWIFT receiver for the transmission to line of a continuous tone or 'acknowledgment signal' during the correct reception of data. The acknowledgment signal has a frequency of 410 c/s with SFR4 and SFR5 and 320 c/s with SFR2 and SFR3. A higher frequency is used over switched telephone-circuits in order to avoid the large attenuations sometimes experienced here at frequencies near 300 c/s.

The acknowledgment signal is fed via a band-pass filter to a terminal on the SWIFT receiver. It is fed from here to line via a buffer stage and the line transformer. The problem of interference of the acknowledgment signal with the correct operation of the SWIFT receiver has been effectively overcome by designing the receiver input-filter to attenuate by at least 60 dB at frequencies around that of the acknowledgment signal and by ensuring that the total harmonic-distortion of the acknowledgment signal when it reaches the input filter never exceeds 0·1% of the signal voltage, or −60 dB, neglecting any distortion components of this signal which may be reflected back from the line.

When the acknowledgment signal reaches the SWIFT transmitter at the other end of the line, it is fed via the line transformer and a buffer stage inside the SWIFT transmitter to a terminal on this equipment. From this terminal it is fed via a band-pass filter to the acknowledgment-signal detector, whose output signal indicates whether or not the received signal level exceeds a predetermined value (normally around −50 dBm). The SWIFT transmitter output-filter has been designed to attenuate severely at frequencies around that of the acknowledgment signal and so it prevents the transmitted SWIFT-signals from interfering with the correct operation of the acknowledgment-signal detector.

The generator and detector of the acknowledgment signal together with their associated band-pass filters do not form part of the SWIFT equipment itself and these are mounted as separate units.

The acknowledgment signal is at present used in all commercial applications of 750-baud SWIFT equipment and it has been used in many trials of the equipment over telephone circuits. It has proved to be a simple and most effective means for controlling the operation of the transmission equipment.

3. Tolerance to Noise and Distortion

3.1. *Method of Testing*

A series of tests has been carried out to determine the tolerance of SWIFT to white noise for different attenuation and delay characteristics in the transmission path. In each of these tests the SWIFT equipment was adjusted in the normal manner, for correct operation with no attenuation and delay distortions in the transmission path, and care was taken not to readjust the equipment in any way when it was then tested in the presence of such distortion. The carrier phase relative to that of the modulating waveform in the transmitter was always set to 90 deg.

In every case band-limited white noise was used. This means that the noise energy was confined to the signal frequency-band so that the large majority of the noise energy was accepted by the receiver input-filter. The frequency spectrum of the white noise was also flat over the frequency band 750 to 2250 c/s.

White noise has been used in these tests, not only because it is relatively easy to produce in the laboratory and to analyse theoretically but also because it appears that the relative tolerance to white noise of different data-transmission systems is an approximate measure of their relative tolerance to the additive noise normally obtained over telephone circuits.[2, 9] This is true even though white noise itself is rarely a significant source of errors when working over telephone circuits. The tolerance to white noise of a data-transmission system is therefore one of the important measures of its suitability for use over telephone circuits.

When a network is included in the transmission path, its effect on the performance of SWIFT is quoted as the reduction in tolerance to white noise. In the tests this was measured at the receiver input as the increase in the signal/noise ratio required to maintain an error rate of 1 bit in 10^5, when the network was added to the transmission path which was previously distortionless. The arrangement used was also such that the characteristics and bandwidth of the white noise reaching the receiver input terminals were not affected by the addition of the network into the transmission path.

The reduction in tolerance to white noise can not only be measured much more accurately than the absolute value of the signal/noise ratio, since the more important sources of error tend to cancel out, but also this quantity is largely independent of the bandwidth of the white noise, so long as the latter is constant and always occupies the main portion of the signal frequency-band. It is therefore a most useful measure whereby the effects of given attenuation and delay distortions on different data-transmission systems can be compared.

The signal/noise ratios quoted in this paper or used here to determine the reductions in tolerance to white noise, are in every case expressed as direct power-ratios and not as normalized signal/noise ratios.

In each of the tests carried out with SFR4 and SFR5, the transmitter output and receiver input filters used were delay-equalized filters designed for the 1500-baud SWIFT equipment SFT2 and SFR2. With the arrangement described, these filters introduce negligible attenuation and delay distortions into the signal frequency-band of a 750 or 1500 baud SWIFT signal. Therefore when the filters are used with one of the networks A, B or C (Figs. 4–6) in the transmission path, they do not produce resultant attenuation and delay characteristics which have a noticeably different effect on the performance of the SWIFT equipment than do the characteristics of the network alone. In this way a more accurate comparison can be obtained under similar conditions, not only between the performance of SFR4 or SFR5 and that of SFR2 but also between the performance of any of these SWIFT equipments and that of any other data-transmission system whose line filters again introduce negligible attenuation and delay distortions. The performance of SFR4 or SFR5 is in fact little affected by using the different line filters, and the results of the tests can therefore be applied to the practical models of SFR4 and SFR5 with no serious errors. The point has however been made because sometimes when the equipment line-filters introduce appreciable attenuation and delay distortions, very misleading and unduly optimistic results may be obtained by the chance combination of the attenuation and delay characteristics of the equipment line-filters with those of a particular line-simulator such as network A. The normal result of this is that a quite excessive degradation in performance is then obtained for other equally probable attenuation and delay characteristics in the transmission path.

In no test was any equalization of any kind used for the attenuation and delay characteristics of the transmission path.

Network A (Fig. 4) simulates the typical worst attenuation and delay distortions which are likely to be found over switched telephone-circuits in Great Britain. Its characteristics have been designed to be the same as those shown in Fig. 2 of the diagram LR2009, Issue 4, produced by the Post Office Engineering Department. The tests over this network clearly provide the most important results obtained in this series. Networks B and C (Figs. 5 and 6) are not intended to simulate the actual characteristics obtained over telephone circuits, but rather to provide a contrast to network A. These networks introduce considerably more delay-distortion but much less attenuation-distortion into the signal frequency-band than does network A. However the general form of these distortions is similar to those sometimes obtained over telephone circuits.

In each series of measurements to determine the value of the signal/noise ratio at which the error rate was equal to 1 bit in 10^5, measurements of the error rate were carried out for different values of the signal/noise ratio with each of the following repetitive 16-bit signal-patterns:

1010110011110000, 1100101011110000,

1111010100010001 and 0101111100010001.

Wherever a noticeable difference in the signal/noise ratio for an error rate of 1 bit in 10^5 was observed between any of these signal patterns or wherever a noticeable variation in the level of the signal distortion at the output of the post-detection low-pass filter was observed over a range of different signal patterns, a further series of measurements was carried out to determine first the repetitive 16-bit signal-patterns requiring the largest signal/noise ratio for an error rate of 1 bit in 10^5 and then the value of this signal/noise ratio.

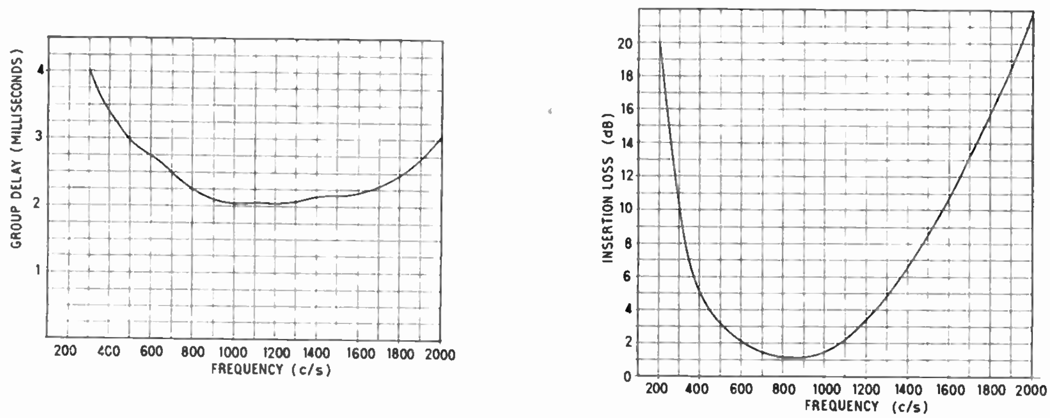


Fig. 4. Attenuation and delay characteristics of network A.

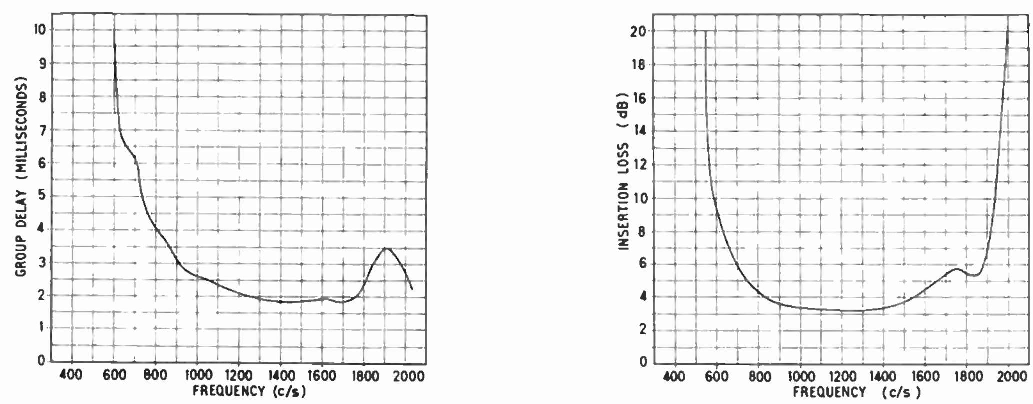


Fig. 5. Attenuation and delay characteristics of network B.

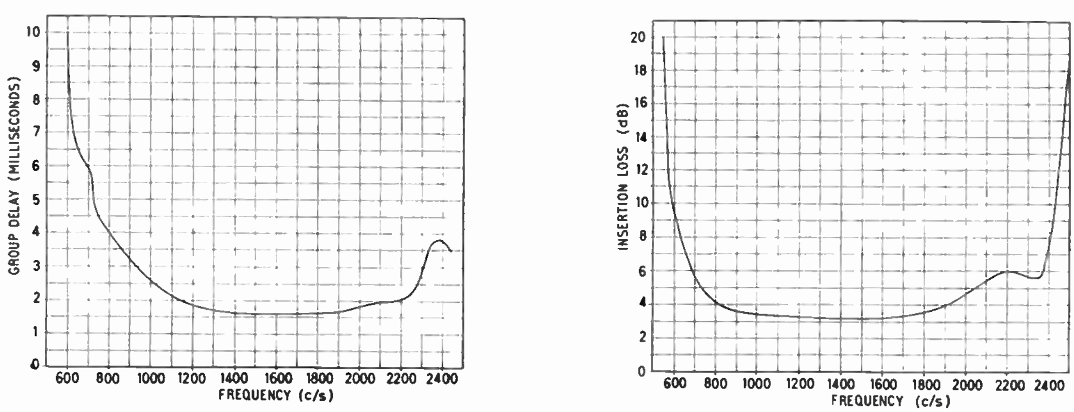


Fig. 6. Attenuation and delay characteristics of network C.

With either no distortion or the network A in the transmission path, the signal/noise ratio giving an error rate of 1 bit in 10^5 was the same for each of the four standard 16-bit signal-patterns mentioned above and only slightly greater for the less favourable patterns, so long of course as these had at least one '1' and one '0'.

With the network B or C in the transmission path, however, the signal/noise ratio required was much more seriously affected by the signal pattern. Thus in the results obtained with the latter networks, two values of the reduction in tolerance to white noise are quoted: one for the average or typical 16-bit signal-patterns, such as the four standard patterns, and the other for the 16-bit signal-patterns requiring the largest value of the signal/noise ratio for an error rate of 1 bit in 10^5.

To avoid repetition in the following sections, it will be assumed that wherever correct operation is claimed for all 16-bit signal-patterns, these patterns do not include the two special cases of all '0's' and all '1's', for which the correct output signal is not obtained from the timing-waveform generator in either SFR2 or SFR4.

3.2. *Results of Tests*

Under conditions of no attenuation and delay distortions in the transmission path, the signal/noise ratio which gives an error rate of 1 bit in 10^5 with band-limited white noise, is 8 dB for SFR4 and 10 dB for both SFR2 and SFR5.

In each of these results the signal/noise ratio is given to the nearest whole decibel above the value actually measured. This has been done in an attempt to avoid an over-optimistic assessment of the equipment, bearing in mind that the overall accuracy of the measurements is unlikely to have been better than $\pm\frac{1}{2}$ dB. With wide-band white noise and no receiver input-filter, the advantage of SFR4 over SFR2 and SFR5 is in fact 3 dB.

The reductions in tolerance to white noise experienced by the different receivers, when working with one of the networks A, B or C in the transmission path, are shown in Table 1.

Table 1

SWIFT receiver	Signalling speed (bauds)	Reduction in tolerance to white noise (dB) with the connection of each of the following networks into a previously distortionless transmission-path: Network A (Fig. 4)	Network B (Fig. 5) Average 16-bit signal patterns	Network B (Fig. 5) Worst 16-bit signal patterns	Network C (Fig. 6) Average 16-bit signal patterns	Network C (Fig. 6) Worst 16-bit signal patterns
SFR4	750	1	$1\frac{1}{2}$	$2\frac{1}{2}$		
SFR5	750	15	8	12		
SFR2	1500				10	12

The receiver SFR4 clearly has a much greater tolerance than SFR5 to the attenuation and delay distortions of the type experienced over telephone circuits. SFR5 will nevertheless operate correctly with a transmission path having attenuation and delay characteristics equal to the worst which are likely to be found over the switched telephone-network in Great Britain (Network A).

These and other tests with SFR4, over different transmission paths having attenuation and delay characteristics appreciably worse than the poorest characteristics to be expected over the switched telephone-network, have shown only a small reduction in tolerance to white noise for the signal patterns most seriously affected and also a correspondingly small distortion in the receiver detected-signal. Correct operation together with a high tolerance to noise should therefore be obtained with the 750-baud receiver SFR4 over the large majority of switched telephone-circuits in Great Britain, with no need for any equalization of the attenuation or delay characteristics of these circuits.

Correct operation for all 16-bit signal-patterns cannot be obtained with SFR2 over network B (Fig. 5). This network removes part of the upper sideband of the 1500-baud p.m. signal and also introduces considerable delay-distortion over the remaining part of the signal frequency-band.

The network A (Fig. 4) represents fairly accurately one combination of the worst attenuation and delay distortions in the transmission path, which still just permits the correct operation of SFR2 in the absence of noise for all 16-bit signal-patterns. The signal patterns most seriously affected by distortion are here on the verge of being incorrectly detected. The fact that the attenuation and delay distortions introduced by network A are considerably more severe than the worst which are likely to be experienced over private telephone-circuits, implies that the latter are unlikely to cause any serious reduction in the tolerance of SFR2 to white noise.

These and other tests over networks having various attenuation and delay characteristics, indicate that correct operation with the 1500-baud receiver SFR2 should be obtained over the large majority of private telephone-circuits in Great Britain, without any additional delay-equalization or attenuation-equalization being used for the telephone circuit. Where partial delay-equalization of the telephone circuit can be achieved by the use of one of two or three different delay-equalizers, correct operation with SFR2 should be obtained not only over the private circuits in Great Britain but also over the majority of switched circuits in those countries where the permissible transmitted-signal-level in the frequency band 420 to 900 c/s is the same as that in the rest of the available bandwidth.

Further tests have also been carried out on both 750 and 1500 baud SWIFT equipment to determine the effects of different transmitter output and receiver input filters. These tests have shown that correct operation can be maintained with an acceptable distortion level in the receiver detected-signal, for all 16-bit signal-patterns, even when a large part of the upper sideband of the transmitted signal is removed. This applies so long as the lower sideband is transmitted without excessive attenuation and delay distortions. Tests on a SWIFT receiver arranged for vestigial-sideband operation have also shown that, as with double-sideband operation, the receiver has a more than adequate tolerance to frequency-translation effects in the transmission path. It therefore appears that SWIFT equipment using a vestigial-sideband signal should operate satisfactorily over telephone circuits. However because of the inevitable reduction in tolerance to white noise of such a vestigial-sideband system, its use is not recommended and double-sideband operation should always be used wherever permitted by the available frequency-band.

4. Performance over Telephone Circuits

A series of tests was carried out with SWIFT in transmitting data at 750 bauds over a switched telephone-circuit of about 150 miles in length (circuit D). The distribution of the errors obtained was analysed and the results are given in Table 2. In each of these tests a character having a fixed pattern of ten or sixteen bits was transmitted repetitively.

Two further series of measurements of the error rate were carried out on 750 baud SWIFT equipment, where this formed part of a complete paper-tape to paper-tape data link operating over the switched telephone-network. Two different switched telephone-circuits were used (circuits D and E, where circuit E was about 200 miles in length). The results of these tests are given in Table 3. The information transmitted here was normal data or its equivalent.

The tests carried out over circuit D show a very typical average-error-rate but a rather high proportion of the larger groups of multiple errors. The results obtained in working over circuit E show a higher average-error-rate but a much lower proportion of the larger groups of multiple errors.

Since the various tests were limited to two different routes on the switched network, they serve only to illustrate very approximately the type of performance that may be expected from the 750-baud SWIFT equipment. However, at least two other independent series of trials have been carried out with SWIFT equipment over equivalent telephone-networks and such information as has been available from these tests indicates a general standard of performance similar to that obtained in the tests described here.

Table 2

Analysis of the errors obtained in transmission at 750 bauds over the switched telephone-circuit D

SWIFT receiver	SFR5	SFR4	SFR4
Transmitted character	1010110100	1010110100 or 1111100000	1011001110001100
Number of bits per character	10	10	16
Total number of bits transmitted (millions)	37	62	27
Bit error-rate (bits in 10^5)	1·1	0·95	1·1
Total number of characters transmitted (millions)	3·7	6·2	1·7
Character error-rate (characters in 10^5)	6·6	4·8	8·1
% of error characters containing: 1 bit error	58·0	41·0	36·4
2 ,, errors	25·5	34·8	32·8
3 ,, ,,	10·9	13·7	15·0
4 ,, ,,	2·8	5·9	9·3
5 ,, ,,	2·0	4·0	3·6
6 or more bit errors	0·8	0·6	2·9
% of isolated error-groups containing: 1 bit	77·0	62·9	59·5
2 adjacent bits	18·5	27·5	29·0
3 ,, ,,	3·6	6·3	9·0
4 ,, ,,	0·6	2·3	2·5
5 or more adjacent bits	0·3	1·0	0·0

In the results of the tests over the telephone-circuit D (Table 2) SFR4 shows only a very small advantage in the bit error-rate over SFR5. This is due to the fact that the signal distortion over the telephone circuit was quite low and therefore the advantage in tolerance to white noise of SFR4 over SFR5 was of the order of only 2 dB. Had there been appreciably higher levels of attenuation and delay distortions over this circuit, the bit error-rate obtained with SFR4 would almost certainly have been significantly lower than that of SFR5.

5. Comparison of SWIFT with a 600/1200-baud F.M. System

5.1. *Method of Testing*

A series of comparative tests has been carried out with SWIFT and a good 600/1200-baud synchronous f.m. data-transmission system designed to the C.C.I.T.T. recommendations. The object of these tests was to obtain an accurate experimental comparison between a p.m. system and an f.m. system, in which the results were in no way influenced by the particular circuit design of either system. The achievement of this has been made possible by the fact that the f.m. system which was used for these tests, has a design closely analogous to that of SWIFT.

In the f.m. transmitter a square-wave modulating waveform is fed to a linear frequency-modulator. This provides the optimum value of the cross-correlation coefficient for the two binary-elements transmitted[2, 5] and eliminates phase-discontinuities at the boundaries between these elements. The f.m. receiver slices the received signal and produces a detected-signal voltage which is linearly proportional to the carrier frequency. This frequency is measured by the time interval between adjacent zero-level crossings in the received signal-carrier. The whole system is so designed that the inherent jitter and signal distortion in the receiver detected-signal, are reduced to a very low level. This basic design not only uses the best available transmitted signal which also satisfies the C.C.I.T.T. recommendations, but it also makes the most effective possible use of the information in the received f.m. signal. It should therefore have the best overall tolerance to noise and distortion which can reasonably be obtained in a practical f.m. system designed to the C.C.I.T.T. recommendations.

The individual circuits in the f.m. equipment have been designed throughout to the same standard as those in SWIFT. Furthermore the timing-waveform generator in the f.m. receiver has exactly the same basic design as that in the SWIFT receiver, so that neither system gains any advantage over the other due to the performance of its timing-waveform generator.

The transmitter output and receiver input filters in the f.m. equipment are delay equalized, as in the case

Table 3

Errors obtained in transmission at 750 bauds over the switched telephone-network, using a complete paper-tape to paper-tape data-link

Telephone circuit	D	E
SWIFT receiver	SFR5	SFR5
Number of bits per character	10	12
Total number of bits transmitted (millions)	24	204
Bit error-rate (bits in 10^5)	1·1	3·1
Total number of characters transmitted (millions)	2·4	17
Character error-rate (characters in 10^5)	8	32
% of error characters containing 1 bit error	67·5	85
2 „ errors		14
3 „ „		1

of the SWIFT equipment which was used for the comparative tests. In these tests there was therefore no serious tendency for the line filters used with either system to bias the results in favour of one or the other. However, since the pass-band of the line filters in the f.m. system was not quite the same as that of the line filters in the p.m. system, there was a possibility of a small bias due to this difference. Thus in every test over one of the networks A, B or C, the f.m. system and SWIFT were each used without their line filters and associated delay-equalizers In the tests over a distortionless transmission-path, where the line filters could not be omitted without introducing other sources of error, the bias due to the difference between the line filters was carefully checked and found to be negligible.

In each test the SWIFT equipment was set up as described at the beginning of section 3.1 and the various adjustments in the f.m. equipment were also carefully optimized. Every reasonable precaution was therefore taken to avoid all the known sources of error in these tests.

When comparing the tolerance to white noise of the 750-baud p.m. receiver SFR4 with that of the 600-baud f.m. receiver, the source of band-limited white noise mentioned in section 3.1 was used with each system. Band-limited as opposed to wide-band white noise was of course essential for the comparative tests with the line filters omitted.

In describing the results of the tests with white noise, the quoted advantage in tolerance to white noise of the p.m. system over the f.m. system, is in every case ½ dB less than the measured value. This ensures that the quoted results are biased slightly in favour of the f.m. system, so that these results correspond to a comparison between SWIFT and an optimum f.m. system designed to the C.C.I.T.T. recommendations.

No measurements were made of the tolerance to band-limited white noise of the 1200-baud f.m. receiver. This was largely because it was not considered that the bandwidth of the white noise adequately covered the frequency band of the 1200-baud f.m. signal and the results of such measurements could not therefore be relied upon to an acceptable degree of accuracy.

In all tests with white noise, the repetitive 16-bit signal-patterns mentioned in Section 3.1 were used, whereas in all tests to measure distortion in the receiver detected-signal, two different pseudo-random patterns were used, the first pattern having a cycle of 1736 bits and the second a cycle of 60 bits.

5.2. *Results of Tests*

In the first series of tests, the tolerance to white noise of the 750-baud p.m. receiver SFR4 was compared with that of the 600-baud f.m. receiver for different attenuation and delay characteristics in the transmission path. The results obtained are shown in Table 4.

Table 4

Frequency characteristics of transmission path:	No attenuation or delay distortions	Network A (Fig. 4)	Network B (Fig. 5)
Advantage in tolerance to white noise of the 750-baud p.m. receiver SFR4 over the 600-baud f.m. receiver (dB):	$3\frac{1}{2}$	8	3

The advantage in tolerance to white noise of the 750-baud p.m. receiver over the 600-baud f.m. receiver, is expressed as the direct power-ratio of the transmitted level of the f.m. signal to that of the p.m. signal, for an error rate of 1 bit in 10^5 in each receiver and with band-limited white noise of a fixed level fed to the receiver input terminals. This method of comparison makes full allowance for any difference between the attenuations experienced over the transmission path by the f.m. and p.m. signals. For typical signal-patterns, the p.m. signal experiences $\frac{1}{2}$ dB more attenuation over network A than the f.m. signal and nearly 1 dB more attenuation over network B. Thus a comparison of the signal/noise ratios which are needed at the receiver input terminals to give an error rate of 1 bit in 10^5 in each receiver, will show correspondingly more favourable results for the p.m. receiver with network A or B in the transmission path, than those shown in Table 4.

In the absence of attenuation and delay distortions in the transmission path, a binary p.m. data-transmission system such as SWIFT (using SFR2 or SFR4) will theoretically operate in the presence of white noise to give an error rate of 1 bit in 10^5, when the normalized signal/noise ratio is 3 dB less than that required for an optimum binary f.m. data-transmission system.[2,5,6,7,8] If the frequency shift in the f.m. signal is reduced from its optimum value, where the shift in c/s is equal to the modulation rate in bauds, to 2/3 of this value as in the f.m. system recommended by the C.C.I.T.T., the normalized signal/noise ratio which gives an error rate of 1 bit in 10^5, is increased by a further $1\frac{1}{2}$ dB.[5] This assumes a rectangular modulating-waveform for the f.m. signal, which is the optimum arrangement for the applications considered here. The p.m. system therefore gains an overall advantage in tolerance to white noise of $4\frac{1}{2}$ dB. Although the 750-baud p.m. system requires about the same frequency-band in the transmission path as does the 600-baud f.m. system recommended by the C.C.I.T.T., it has a 25% greater modulation rate. Thus the theoretical advantage in tolerance to white noise of the 750-baud p.m. receiver SFR4 over the 600-baud f.m. receiver, is reduced from $4\frac{1}{2}$ to $3\frac{1}{2}$ dB, when the signal/noise ratios are converted from normalized to direct power-ratios. The experimental results quoted in Table 4 are expressed as direct power-ratios. The theoretical result is therefore in very good agreement with the experimental value obtained for no distortion in the transmission path.

It has been shown experimentally and confirmed by theoretical analysis that, in transmission over network A (Fig. 4), the more severely attenuated of the 1300 and 1700-c/s binary-elements in the 600-baud f.m. signal are attenuated by a little over $4\frac{1}{2}$ dB more than the more severely attenuated of the two binary-elements in the 750-baud p.m. signal, even though the p.m. carrier is at the centre frequency of the f.m. signal. Since the delay characteristic of network A has an almost negligible effect on either system in comparison with the attenuation characteristic of this network and since the network A introduces in either system only slight signal-distortion at the output of the post-detection low-pass filter, it can be shown by simple theoretical analysis that for the average signal patterns, in which there are as many '1's' as '0's', the p.m. system gains a further advantage in tolerance to white noise of $4\frac{1}{2}$ dB over the f.m. system, when a distortionless transmission path is replaced by the network A. This is essentially because of the greater attenuation introduced in the f.m. signal. Thus with a transmission path having the frequency characteristics of network A, the 750-baud p.m. receiver SFR4 has a theoretical advantage in tolerance to white noise of 8 dB over the 600-baud f.m. receiver, where this advantage is expressed as a direct power-ratio. This result is in very good agreement with the experimental value obtained under the same conditions (Table 4).

The significant feature of the type of attenuation distortion normally experienced over telephone circuits is that it is likely to produce an appreciably greater attenuation in one of the two binary-elements of the f.m. signal than in either of the two binary-elements of the equivalent p.m. signal. Because a definite limit is set on the power level fed into the telephone circuit for either element, no amount of attenuation equalization either at the transmitter or at the receiver can lessen the advantage in tolerance to additive noise gained by the p.m. system because of this effect. This places the f.m. system at an undoubted disadvantage.

From the tests over the networks A and B, it appears that, whereas a given degree of attenuation distortion in the transmission path has a much more serious effect on the 600-baud f.m. system than it does on the 750-baud p.m. system, a given degree of delay distortion has a rather similar effect on either system.

With network A as the transmission path, the 1500-baud p.m. receiver SFR2 will just operate correctly for all 16-bit signal-patterns, whereas the 1200-baud f.m. receiver fails completely. Neither the 1500-baud p.m. system nor the 1200-baud f.m. system will operate correctly over network B (Fig. 5), but both systems will operate correctly for all 16-bit signal-patterns over network C (Fig. 6).

In the second series of tests, a comparison was made of the signal distortion at the output of the post-detection low-pass filter in the p.m. receiver (SFR4 or SFR2) with that in the corresponding f.m. receiver, with each of the networks A, B and C connected in turn into the transmission path. These tests showed that the 750 or 1500-baud p.m. system normally has a distortion level in the receiver detected-signal similar to that in the corresponding 600 or 1200-baud f.m. system when the attenuation and delay characteristics of the transmission path are of the type obtained over telephone circuits and when the same degrees of attenuation and delay distortions are introduced into the signal frequency-band of either system. The one significant exception to this is the complete failure of the 1200-baud f.m. system over network A. For the purpose of these tests it is assumed that the bandwidth of the 750 or 1500-baud p.m. signal is the same as that of the corresponding 600 or 1200-baud f.m. signal. The 1200-baud f.m. signal uses frequencies of 1300 and 2100 c/s for the two binary-elements.

The interesting point here is that when the received signal is sliced at the receiver input, the distortion level in the receiver detected-signal is not in itself a measure of the reduction in tolerance to white noise, since it takes no account of the relative attenuations of the two binary-elements over the transmission path. This is a major source of reduction in tolerance to white noise in the case of the f.m. system, when working over a transmission path which introduces appreciable attenuation-distortion.

Taking all the available experimental and theoretical evidence, it appears clear that when using a transmission path having attenuation and delay characteristics of the type which are likely to be experienced over telephone circuits, the p.m. receivers SFR2 and SFR4 have a very useful advantage over the 1200 and 600-baud f.m. receivers respectively, not only because of the 25% increase in signalling speed but also because at this greater speed the p.m. equipment has in each case a significant advantage in tolerance to white noise. This advantage is likely to be maintained in operation over telephone circuits, since the relative tolerance to white noise of different data-transmission systems is an approximate measure of their relative tolerance to the additive noise normally obtained over these circuits.[2, 9] Also the tolerance of the p.m. equipment to the multiplicative noise over telephone circuits is similar to that of an equivalent f.m. system of very good design.

The average error rate to be expected when using SWIFT over telephone circuits is typically less than half that which would be obtained if the equivalent f.m. system were used instead. This applies when the SWIFT receiver is either SFR2 or SFR4.

The 750-baud p.m. receiver SFR5 has a slight advantage in tolerance to white noise over a good 600-baud f.m. receiver, when these are working under conditions of no signal distortion, but it has a very much lower tolerance to attenuation and delay distortions. Thus over telephone circuits having significant attenuation and delay distortions, the error rate obtained with SFR5 may be higher than that of the f.m. system. A much better performance would therefore be obtained over such circuits when using SFR4 instead of SFR5.

One conclusion which can clearly be drawn from the many contributions to the meeting in the autumn of 1963 of the Special Study Group A of the C.C.I.T.T., is that a significantly higher transmission rate is obtainable with a binary p.m. system than with a binary f.m. system, when these are working over telephone circuits. Several experiments over telephone circuits to compare the performance of SWIFT with that of equivalent f.m. data-transmission systems, have further confirmed these conclusions. For instance, comparative tests with 750-baud SWIFT equipment and a good 600-baud f.m. system, showed that correct operation was obtained with SWIFT not only over all circuits where the f.m. system would operate correctly but also over some others where the latter would not.

5.3. *Comparison of Equipment Complexities*

Not only has the 750/1500-baud SWIFT equipment (using SFR4 or SFR2 respectively) a useful advantage in performance over a very good 600/1200-baud f.m. system designed to the C.C.I.T.T. recommendations, but also the SWIFT equipment gains a further advantage in that its basic circuit design as well as its test and setting-up procedure are undoubtedly simpler. A careful comparison of the designs of f.m. and p.m. systems has fully confirmed the economic advantages of p.m. systems, where these are compared with f.m. systems on the basis of the same standard of design. Obviously if full account is not taken in each case of the standard of design and of the degree of sophistication in the individual circuits, the comparison between f.m. and p.m. systems can be most misleading.

A more economical design can be obtained for a p.m. system, for the following reasons. Firstly, a single oscillator or frequency source can be used in the p.m. transmitter both to generate the carrier and to determine the modulation rate, instead of the two separate frequency-sources normally required in an f.m. system. Secondly, as a result of the phase coherence between the carrier and modulating waveform obtained in the p.m. transmitter through the use of a single frequency-source, useful economies can be achieved in the filters at the p.m. transmitter. If the line filters of the p.m. system are delay equalized, similar economies can also be achieved in the delay equalizer at the receiver. Thirdly, by using for the phase modulator in the p.m. transmitter, a phase switch fed by a square-wave modulating waveform, the design of the phase modulator need certainly be no more complex and may well be simpler than that of the corresponding frequency-modulator in the f.m. transmitter. This applies even when the frequency modulator is the means for controlling the frequency of a multivibrator which itself generates the signal carrier. Fourthly, since a buffer stage is always required between the transmitter input-terminal and the modulator in order to ensure reliable operation, the code converter which is a simple divide-by-two stage, involves no additional equipment complexity for the p.m. transmitter. Finally, in the large majority of applications the number of transmitters will exceed the number of receivers. Thus if f.m. systems were replaced by the equivalent p.m. systems in any complete installation, the reduction in equipment complexity which would be achieved at the transmitters would be of significant value and would in general appreciably more than compensate for the slightly more complex receiver design in the p.m. systems.

One possible explanation for the persistent belief that f.m. systems are basically more economical in design than p.m. systems, is that the majority of p.m. systems contain not only the modulator and demodulator but also two timing-waveform generators, one of which is used at the transmitter and the other at the receiver (see Figs. 2 and 3). On the other hand, the majority of f.m. systems contain only the modulator and demodulator, leaving the timing-waveform generators to be provided in the associated equipment. Each of the two timing-waveform generators is however essential to the correct operation of any data-transmission system.[2] Obviously if this is not taken into account, the majority of p.m. systems will appear considerably more complex than the corresponding f.m. systems. Again, although p.m. systems are normally designed for synchronous operation whereas f.m. systems tend to be used in a start-stop mode, there is no reason at all why a p.m. system should not be designed for start-stop operation and still maintain all its various advantages over the equivalent f.m. system.

The only advantage that can reasonably be claimed for the f.m. system is that the modulator and demodulator in this system can be used over a range of signalling speeds below each of the standard or maximum rates, with no further modifications to the design. The modulator and demodulator in a p.m. system such as SWIFT could also be designed for use over a wide range of signalling speeds and when used under these conditions the p.m. equipment would fully maintain its advantage in performance over the corresponding f.m. equipment. However the circuit design of the resultant p.m. equipment would become at least as complex as that of the f.m. equipment and so great an increase in circuit complexity does not seem to be justified by the advantage of being able to use the equipment over a range of signalling speeds. This is particularly so as the equipment is not likely to be used at non-standard speeds (or modulation rates) except in a small minority of its applications. Furthermore, since in these applications the data-transmission system makes a relatively inefficient use of the transmission channel, the correct technical approach is to modify the peripheral equipment wherever possible so that the latter can make full use of the standard signalling speeds. This would avoid complicating the design of the data-transmission equipment at the expense of the large majority of users who require to transmit data only at the standard rates.

6. Conclusions

The good performance obtained with SWIFT in tests over telephone circuits and in the laboratory under various conditions of noise and distortion has fully justified the choice of the basic design. A careful comparison has also shown that in applications over telephone circuits a p.m. system such as SWIFT is superior to the equivalent f.m. systems in every important respect, both technically and economically.

The good performance obtained with a p.m. data-transmission system such as SWIFT could in addition be maintained at signalling speeds which are multiples of 600 bauds. The speed of 600 bauds itself has however little technical merit for such a system, since it is unnecessarily slow, but the higher speeds of 1200 or 1800 bauds offer useful possibilities.

In every case the delay used in the detector of the p.m. receiver must be equal to the duration of one signal element, as in SFR2 and SFR4, in order to obtain the maximum tolerance to noise and distortion.

At a signalling speed of 1200 bauds and using a binary double-sideband signal with a carrier frequency of 1800 c/s, correct operation should be obtained with the p.m. system over the large majority of switched telephone-circuits in Great Britain. When using this system, reliable operation would be obtained over many of the poorer circuits over which the equivalent 1200-baud f.m. system designed to the C.C.I.T.T. recommendations, would give a poor performance or fail completely.

At a signalling speed of 1800 bauds and using a binary double-sideband signal with a carrier frequency of 1800 c/s, correct operation with the p.m. system should be obtained over the large majority of private circuits in Great Britain.

From purely technical and economic considerations, however, the signalling speeds of 750 and 1500 bauds, using in each case a carrier frequency of 1500 c/s, undoubtedly give the most effective and efficient design for the basic p.m. system. This is of course because these signalling speeds were chosen as the optimum for the particular modulation-method used and not as a compromise with the rather different requirements of other modulation methods.

Where operation at 1800 or 2400 bits per second is required over switched circuits and where operation at 2400, 3000 or 3600 bits per second is required over private circuits, the data-transmission system giving the best overall performance for a given degree of complexity in the design, is very probably a synchronous system using a single phase-modulated carrier with a quaternary double-sideband signal and employing differentially-coherent detection at the receiver.

7. Acknowledgments

The author wishes to acknowledge the considerable contribution of his colleagues towards the design and testing of the SWIFT equipment, and is indebted to the Director of Research, British Telecommunications Research Ltd., for permission to publish this paper.

8. References

1. A. P. Clark, "A high speed signalling system for use over telephone circuits", *A.T.E. J.*, **15**, No. 2, pp. 157–72, April 1959.
2. A. P. Clark, "Considerations in the choice of the optimum data transmission systems for use over telephone circuits", *J.Brit.I.R.E.*, **23**, No. 5, pp. 331–55, May 1962.
3. D. J. Dace, "Data collection and distribution", *J.Brit.I.R.E.*, **23**, No. 5, pp. 357–64, May 1962.
4. D. J. Dace, "Experience in the practical use of data transmission", *Computer J.*, **6**, No. 1, pp. 17–23, April 1963.
5. G. L. Turin, "Error probabilities for binary symmetric ideal reception through nonselective slow fading and noise", *Proc. Inst. Radio Engrs*, **46**, pp. 1603–19, September 1958.
6. A. B. Glenn, "Comparison of psk vs fsk and psk-am vs fsk-am binary coded transmission systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-8**, No. 2, pp. 87–100, June 1960.
7. E. Arthurs and H. Dym, "On the optimum detection of digital signals in the presence of white Gaussian noise—a geometric interpretation and a study of three basic data transmission systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-10**, No. 4, pp. 336–72, December 1962.
8. P. Besslich, "Error probabilities of binary transmission systems due to noise and fading", *Arch. Elekt. Ubertragung*, **17**, No. 4, pp. 185–97, April 1963.
9. J. Wegner, "Data transmission tests on the telephone network of the Federal German Post Office", *Nachrichtentech. Z.*, **15**, No. 12, pp. 629–34, December 1962.

Manuscript first received by the Institution on 17th July 1963 and in final form on 13th December 1963. (Paper No. 887.)

Letter to the Editor

Comment on technical and other matters appropriate to *The Radio and Electronic Engineer* is welcomed for consideration for publication under this heading. Unless otherwise stated, the opinions expressed are those of the writers and do not necessarily reflect the views of the Institution.

DEAR SIR,

On re-reading your September 1963 issue† I noted two errors in the editorial on "International Co-operation in Space Research".

You state that the first joint space research venture took place in April 1962 with the launching of *Ariel I*. Ten months prior to that, on 22nd June, 1961, Canada's Defence Research Telecommunications Establishment installed equipment in *Transit IIA*. The purpose was to measure cosmic noise levels at the intended altitude of Canada's first satellite, *Alouette I*.

Later on in your editorial you mention that the communications satellites have been completely American in design. Engineers at RCA Victor Company Limited in Montreal designed and built the wide-band and narrow-band receivers, and the c.w. beacons for *Relay I* and *Relay II*. While the company is owned by Radio Corporation of America I think it only fair to give credit to the Canadian personnel who worked on this project. Incidentally, some of the scientists and engineers at RCA Victor were born and educated in England or Scotland.

† *The Radio and Electronic Engineer*, **26**, No. 3, p. 189, Sep. 1963.

International co-operation in space research should be quite prominent in the new programme announced recently by the United States National Aeronautics and Space Administration, and Canada's Defence Research Board. Ionospheric satellites *Alouette II*, ISIS A, B, and C may contain experiments suggested by scientists from many countries. In this programme Canada will have primary responsibility for the design and construction of the satellites. The United States will provide launching and tracking facilities. Invitations will be extended by the two countries to scientists in other lands to submit proposals for experiments.

This type of international co-operation is very desirable.

Yours sincerely,
IAN R. DUTTON,
Associate Editor,
Canadian Electronics Engineering

481 University Avenue,
Toronto 2, Ontario, Canada.
5th February, 1964.

[We welcome this opportunity provided by Mr. Dutton's letter to give proper credit to the work of Canadian scientists and engineers in space research and development, and on page 208 of this issue appears an article reprinted from his journal, dealing with future space programmes.

The Canadian satellite *Alouette I* was the subject of a paper read at the Institution's 1961 Convention and subsequently published in *J.Brit.I.R.E.*, **23**, No. 1, pp. 61–68, January 1962.—Editor, *The Radio and Electronic Engineer*].

STANDARD FREQUENCY TRANSMISSIONS

(Communication from the National Physical Laboratory)

Deviations, in parts in 10^{10}, from nominal frequency for **February 1964**

February 1964	GBR 16kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.	February 1964	GBR 16 kc/s 24-hour mean centred on 0300 U.T.	MSF 60 kc/s 1430–1530 U.T.	Droitwich 200 kc/s 1000–1100 U.T.
1	− 151·9	− 151·3	+ 6	16	− 150·5	− 150·5	− 5
2	—	− 151·5	− 10	17	− 150·5	− 150·6	− 5
3	− 151·5	− 150·5	− 11	18	− 150·6	− 150·2	− 4
4	− 150·5	− 150·2	− 12	19	− 150·7	− 150·8	− 4
5	− 149·1	− 149·6	− 11	20	− 150·6	− 151·2	− 4
6	− 149·2	− 149·5	− 9	21	− 151·2	− 150·4	− 4
7	− 149·0	− 150·1	− 7	22	− 150·3	− 150·4	− 3
8	− 151·0	− 150·2	− 7	23	− 150·1	− 150.5	− 3
9	− 150·6	− 150·1	− 7	24	− 150·8	− 150·8	− 5
10	− 150·6	− 150·4	− 7	25	− 151·1	—	− 5
11	− 150·4	− 149·9	− 7	26	− 151·1	—	− 4
12	− 150·9	− 150·2	− 7	27	—	− 151·3	− 2
13	− 150·7	− 150·4	− 9	28	—	− 149·9	− 4
14	− 150·3	− 150·2	− 8	29	− 149·5	− 149·4	—
15	− 150·7	− 150·9	− 6				

Nominal frequency corresponds to a value of 9 192 631 770 c/s for the caesium F,m (4,0)–F,m (3,0) transition at zero field.

U.D.C. 621.317.76: 621.374.32

Electronic Digital Ratio Equipment

By

A. RUSSELL†

Summary: The equipment provides an accurate method of multiplying a frequency by M/N, where M and N are integers and $M \leqslant N$. Two batching counters are used to determine the output frequency in the long term and the equipment described restores the correct phase relationship. The equipment is transistorized and makes use of modern techniques for counting, gating and parallel addition and subtraction. A brief description is also given of a relay-operated decimal-to-binary converter which has been designed for the ratio equipment.

1. Introduction

In the Machine Tools and Metrology Division of the National Engineering Laboratory measurements are made to determine the accuracy of gear wheels. One device for this work is known as a single-flank gear tester[1,2] on which the gear under test is meshed with a master gear of known accuracy, and rotated. Radial gratings are fixed to both gears and these provide pulse trains which indicate errors in tooth form and tooth position of the gear under test; these errors are apparent as phase changes in the pulse trains.

The number of teeth in gears under test may vary widely; to deal with this and to permit identical gratings to be used for both gears, two types of electronic equipment to convert the frequency of the pulse train from the master gear to that of the gear under test, so providing a perfect electronic 'gear train', have been developed.

One of these systems[2] uses analogue methods; the other proposed by Nairn and described here, uses digital methods. Either equipment can be used where, the nominal gear ratio being M/N, $M \leqslant N$ and both are integers. Applications of this nature occur in speed control, machine-tool control, liquid blending and other mechanical fields, and in electrical frequency control.

2. Principles of Operation

The measuring system is shown in Fig. 1. The master gear radial grating generates a frequency, f; a multiplier and divider are used to convert this to a frequency $(M/N)f$ accurately phase-locked to f; $(M/N)f$ is the frequency that would be generated by the test gear grating if the test gear was perfect. The actual frequency generated by the test gear grating, $(M'/N')f$, includes errors caused by tooth discrepancies in the test gear. A comparison of $(M/N)f$ with $(M'/N')f$ in a phase comparator gives a measure of these errors.

† National Engineering Laboratory, Department of Scientific and Industrial Research, East Kilbride, Glasgow.

Figure 2 shows the operation of the digital multiplier and divider for the simple ratio of $M/N = 4/5$. The master gear pulse train, of frequency f, is shown in row (*a*) where a pulse occurs at each instant g. The g pulse is used to allow four pulses of a higher order, known as 'clock' pulses, to occur as shown in row (*b*). The frequency now averages $4f$. By selecting every fifth pulse of row (*b*), the pulse train shown in row (*c*), with a frequency of $(4/5)f$, is developed. This pulse train, although accurate in the long term, is very irregular in the short term. Thus the time between P_{01} and P_{02} is long and is followed by short intervals between P_{02} and P_{03} and between P_{03} and P_{04}. It is shown in Appendix 1 that the P_0 pulse spacing can be made uniform by introducing a time delay t_d after each P_0 pulse such that

$$t_d = \frac{S}{M} R \qquad \ldots\ldots(1)$$

where S = number of 'clock' pulses between g and P_0

R = number of unused pulses (in row (*d*)).

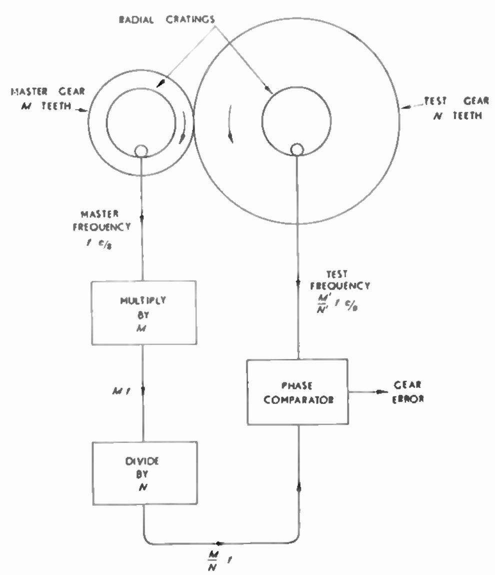


Fig. 1. Measuring system.

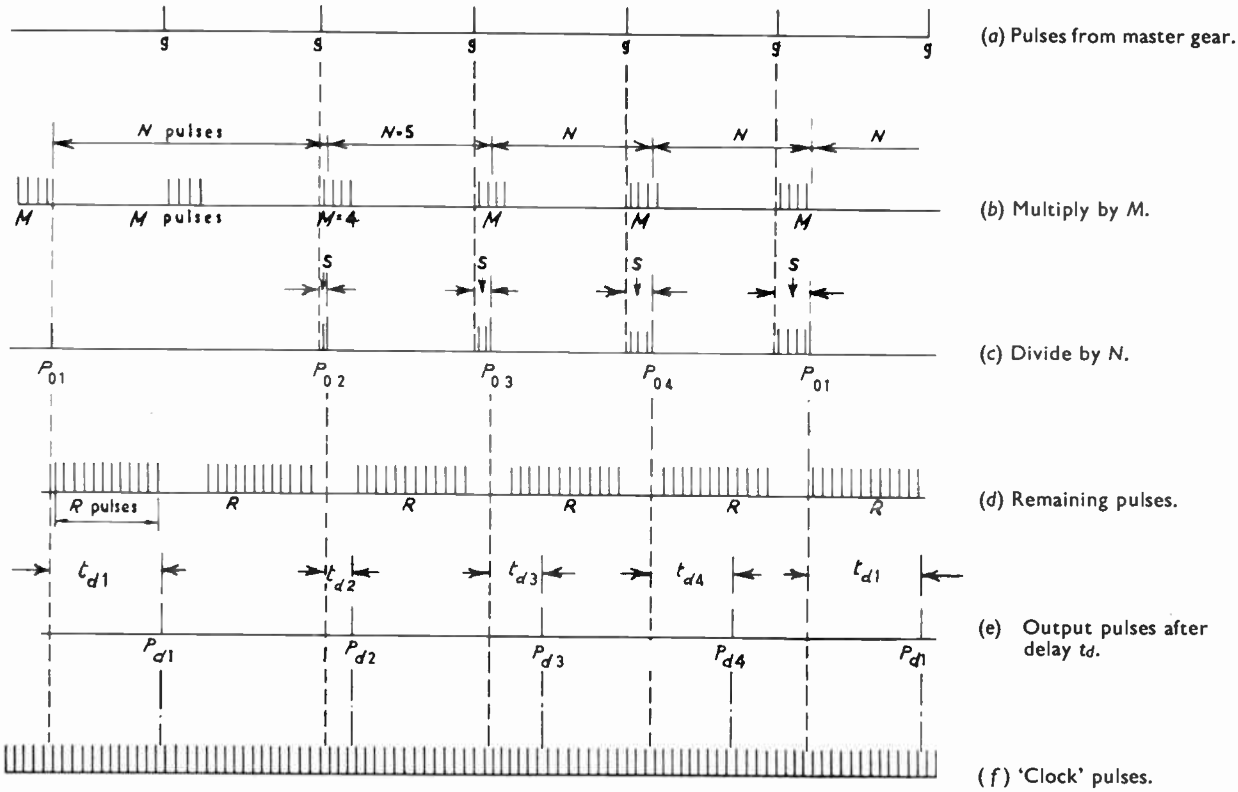


Fig. 2. Spacing P_0 pulses evenly by adding $t_d = \frac{S}{M} R$

With a fixed ratio M/N and a given 'clock' frequency f_c, ($f_c \gg f$), both M and R will remain constant. In the example $M = 4$ and $R = 12$, this means that the delay necessary in each case will be given by

$$t_d = \frac{S}{M} R = \frac{12}{4} S = 3S \qquad \text{......(2)}$$

Thus for

P_{01}, $S = 4$, so $t_d = 12$ 'clock' pulse periods
P_{02}, $S = 1$, so $t_d = 3$,, ,, ,,
P_{03}, $S = 2$, so $t_d = 6$,, ,, ,,
P_{04}, $S = 3$, so $t_d = 9$,, ,, ,,

Row (*e*) shows the output pulses P_{d1}, P_{d2} etc. obtained by delaying P_{01}, P_{02} etc. by the appropriate t_{d1}, t_{d2} etc. and having a uniform spacing of twenty 'clock' pulses. In practice some variation in the relationship between the mechanically derived g pulses and the 'clock' pulses will be experienced, thus R will vary and the value of $t_d = 3S$ (eqn. (2)) will not be valid. Provision must therefore be made in the equipment forming t_d for this variation of R.

3. Logical Design of Multiplier and Divider

The method used to accomplish the time delay t_d, $S \times R/M$ of eqn. (1), is shown in Fig. 3. Each train of R pulses is diverted to a binary counter and stored. The number stored in this counter is then gated into the main accumulator S times, so the count in the main accumulator will then be $S \times R$ pulses. The value of M is set on a decimal-to-binary converter and

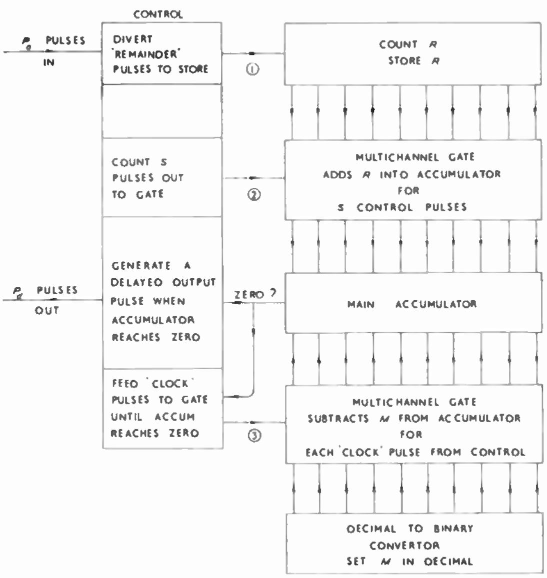


Fig. 3. Delaying P_0 pulses by $t_d = \frac{S}{M} R$ to give P_d pulses.

fed to a multi-channel gate. This gate is opened at each 'clock' pulse after the arrival of P_0 causing M to be subtracted each time until the main accumulator returns to zero. The number of 'clock' pulses required to return the main accumulator to zero is the desired delay for the output pulse P_d and completes the term $S \times R/M$.

4. Logical Design of Control Stage

In Fig. 4 CFF1, 2 and 3 represent three flip-flops which perform the necessary logical steps to form the M, S, R and $P_0 - P_d$ pulses. In the convention adopted throughout for flip-flops an input to the '1' end will drive the flip-flop to the '1' state and an output signal will be present at the '1' end; an input at the '0' end will drive the flip-flop to the '0' state and an output signal will be present at the '0' end. A pulse applied to the base of the flip-flop will reverse the state of the flip-flop.

In the initial condition CFF1 will be in the '0' state and gate G2 will be passing 'clock' pulses to the R counter. A g pulse will drive CFF1 into the '1' condition, closing G2 and opening G1, and 'clock' pulses will then be fed to the M batching counter via G1. At the fourth 'clock' pulse (to continue the example in Fig. 2) the M batching counter will deliver an output pulse which will drive CFF1 to '0' state, open G2 and close G1. Thus a train of four M pulses will be generated by each g pulse; these appear at G1 and are fed to the N batching counter; they are illustrated in row (b), Fig. 2.

The g pulse will also have driven CFF2 to the '1' state allowing G3 to feed S pulses to the main accumulator. The N batching counter will deliver an output pulse (P_0) at each fifth M pulse (Fig. 2, row (c)) which will drive CFF2 to the '0' state, thus the number of S pulses between each g pulse and the subsequent P_0 pulse from the N batching counter will thus be gated. Each S pulse will cause the R count to be added into the main accumulator which will then hold the $S \times R$ portion of eqn. (1), $\left(\frac{S}{M} R = t_d\right)$

It is to pulse P_0 that delay must be applied to give the in-phase pulse P_d. The number M, four in the example, is stored in a decimal-to-binary converter and the number of times that this number will divide into the $S \times R$ count represents the desired delay in 'clock' pulses. If M is subtracted from $S \times R$ at each 'clock' pulse then the number of 'clock' pulses required to return the $S \times R$ count to zero will represent the desired delay and a P_d pulse is generated when this happens.

The simplest way to subtract in binary is to negate and add, a negated binary number being the 'ones' complement of the positive number with unity added;

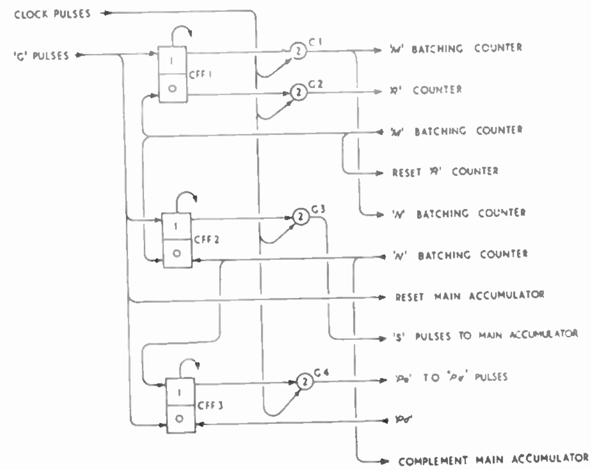


Fig. 4. Control logics.

for example, the simple subtraction $5-3$ can best be performed in binary terms by adding -3 to $+5$

$$5 = 000\,101 \quad \text{and} \quad 3 = 000\,011$$

$$\text{add} \quad -3 = 111\,101$$

$$\text{spill} \quad 1 \quad \overline{\underline{000\,010}} = 2 \qquad \ldots\ldots(3)$$

To divide M into $S \times R$ until $S \times R$ becomes zero, it is convenient to negate $S \times R$ then add M to this negative number until it becomes zero. In practice it is difficult to add unity, so the 'ones' complement of $S \times R$, and not the negation, is used. The maximum error which can accrue from this is one 'clock' pulse only, and normally there will be a 'remainder' into which this pulse is assimilated. This will 'spill' into an extra binary stage (AFF16) and be detected there; this stage is not complemented at P_0. Pulse P_0 from the N batching counter is used to complement the total $S \times R$ in the main accumulator, also to drive CFF3 to the '1' state which gates, at G4, 'clock' pulses to cause M to be added to the complement of $S \times R$ until this reaches zero and delivers pulse P_d and drives CFF3 to the '0' state.

This completes one cycle of control logics from the g pulse to the final output P_d; to meet cases as in rows (b) and (c) in Fig. 2 where no P_0 is delivered during the second group of M pulses, it is arranged that the output pulse from the M batching counter will drive CFF2 to the '0' state to prevent a false count being added into the main accumulator.

5. Logical Design of Main Accumulator

The main accumulator has 16 binary flip-flop stages, with associated half-adders and two-gates. Fifteen of the stages cater for a binary number of 32 767 which is adequate for a 'clock' g pulse ratio of 350:1; this should not be exceeded. The highest count to be

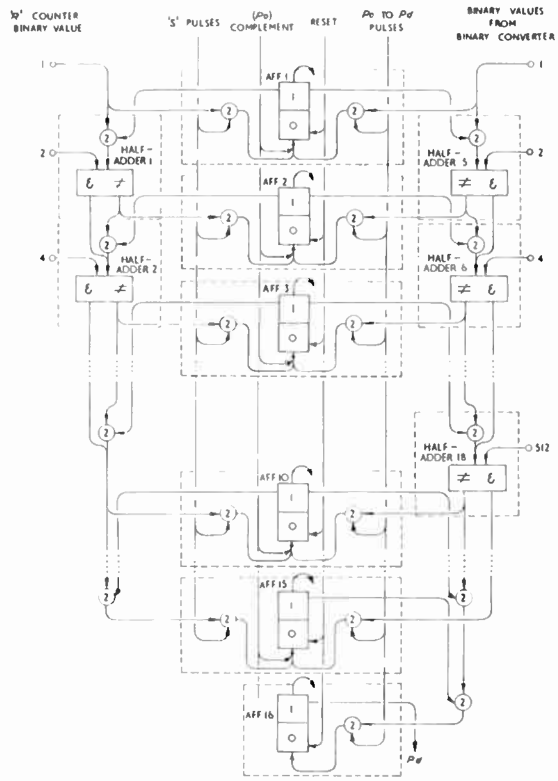


Fig. 5. Main accumulator logics.

expected from this ratio would be $S = 175$ and $R = 175$ which gives $S \times R = 30\ 625$. The accumulator can be extended as necessary.

The logical arrangement is shown in Fig. 5. Stages AFF1 to AFF9 provide for parallel inputs from the R counter and the decimal-to-binary converter; AFF10 provides for a parallel input from the converter only as there are only 9 inputs present from the R counter. AFF15 and 16 have no provision for parallel inputs but accommodate the 'spills' of the binary count in the same parallel mode. The logical sequence is as follows.

(1) The arrival of a g pulse (Fig. 4) will drive all flip-flops to the '0' state.

(2) The first S pulse (Fig. 4) will gate the current number stored in the R counter into the main accumulator (AFF1–AFF9). The action of the half-adders and 2-gates will be described later.

(3) Each succeeding S pulse will add the R number to the one already stored in the main accumulator, the total number 'spilling' up the binary registers as necessary.

(4) At the end of the S pulse train a complement pulse from control will complement the total. AFF16 is not complemented and will remain in the '0' state.

(5) The $P_0 - P_d$ pulses will add the number which has been set on the converter to the complemented number until the main accumulator goes positive and 'spills' into AFF16 driving it into the '1' condition, a P_d pulse being delivered.

(6) $P_0 - P_d$ pulses will cease after delivery of the P_d pulse; the next g pulse will drive all flip-flops to the '0' state, and cycle will recommence.

Figure 5 shows that the binary 1 channel from the R counter can be fed direct to AFF1, via a 2-gate where it is 'clocked' by S pulses, there being no possibility of 'carry' levels from a previous flip-flop or binary input. At the binary 2 channel from the R counter, however, before the input is 'clocked' to AFF2 we have to consider

(*a*) the state of AFF1 (1 or 0);

(*b*) the state of binary 1 from R counter; and

(*c*) the state of binary 2 from R counter.

Thus any one of eight different configurations may be set up at the following S pulse. These are shown in Table 1.

Table 1

	(*a*)	(*b*)	(*c*)	(*d*)	(*e*)	(*f*)	(*g*)	(*h*)
AFF1	0	1	0	0	1	1	0	1
Binary 1	0	0	1	0	1	0	1	1
Binary 2	00	00	00	10	00	10	10	10
	00	01	01	10	10	11	11	100

To 'sense' the state of each of the three inputs a half-adder[3] is used in conjunction with a 2-gate. The requirements of a half-adder are shown in Appendix 2.

Operation of the half-adders (Fig. 5) in conjunction with their 2-gates can best be verified if we apply the configurations above to HA1 column by column. The summations are decimal.

(*a*) AFF1 in '0' state, no input from binary 1 so no input from 2-gate; also no input from binary 2 so no output from HA1. AFF1 remains in '0' state ($0+0+0 = 0$) AFF2 remains in '0' state.

(*b*) AFF1 in '1' state, no input from binary 1 so no input from 2-gate; also no input from binary 2 so no output from HA1. AFF1 remains in '1' state ($1+0+0 = 1$) AFF2 remains in '0' state.

(*c*) AFF1 in '0' state, input from binary 1 so no input from 2-gate; also no input from binary 2 so no output from HA1. AFF1 driven to '1' state ($0+1+0 = 1$). AFF2 remains in '0' state.

(*d*) AFF1 in '0' state, no input from binary 1 so no input from 2-gate; there is input from binary 2 so HA1 gives a sum output of 2. AFF1 remains in '0' state ($0+0+2 = 2$). AFF2 driven to '1' state.

(*e*) AFF1 in '1' state, input from binary 1 so 2-gate gives an input to HA1; also no input from binary 2 so HA1 gives a 'sum' output. AFF1 driven to '0' state; AFF2 driven to '1' state ($1+1+0 = 2$).

(*f*) AFF1 in '1' state, no input from binary 1 so no input from 2-gate; there is an input from binary 2 so HA1 gives a 'sum' output. AFF1 remains in '1' state ($1+0+2 = 3$). AFF2 driven to '1' state.

(*g*) AFF1 in '0' state, input from binary 1 so no input from 2-gate; there is an input from binary 2 so HA1 gives a 'sum' output. AFF1 driven to '1' state ($0+1+2 = 3$). AFF2 driven to '1' state.

(*h*) AFF1 in '1' state, input from binary 1 so 2-gate gives an input to HA1; there is input from binary 2 so HA1 gives a 'carry' output. AFF1 driven to '0' state. AFF2 remains in '0' state. AFF3 driven to '1' state ($1+1+2 = 4$).

The nine binary channels from the *R* counter are 'clocked' into the accumulator in this parallel fashion as are also the ten binary channels from the decimal-to-binary converter giving firstly the $S \times R$ product and then the $\frac{S \times R}{M}$ quotient in terms of 'clock' pulses.

6. Flip-flop Stage Circuits

The flip-flop section contained within the dotted lines of Fig. 5 is shown in Fig. 6. Transistors VT4 and VT5 form, along with the associated diodes, two conventional 'and' or '2' gates. When the input diodes are at '1' (-6V) level the transistors will conduct and raise the emitter level to almost -6V. When either input diode has a '0' (earth) level applied the transistor will be held in a non-conducting state and a '0' will be present at the emitter. The negative-going output pulse is differentiated and inverted at VT3 into a suitable positive-going pulse for triggering the conventional binary flip-flop VT1/VT2. The negative reset pulses are applied via terminal 2 to the base of VT2; this drives VT2 into conduction to achieve a '0' level at the collector. Negative-going

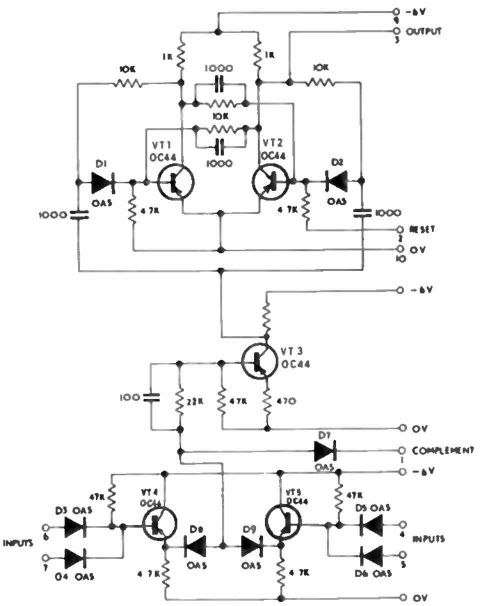


Fig. 6. Main accumulator flip-flop and two-gate.

complement pulses are applied via isolating diode D7 to the inverter stage. Diodes D8 and D9 isolate the '2' gate outputs.

The *R* counter flip-flops are identical with VT1/VT2 in Fig. 6, but without 2-gates or inverter, and the positive output is taken from the emitter of VT1 and fed direct to the next stage.

7. Half-adder Stage Circuits

The half-adder and 2-gate section contained within the dotted lines of Fig. 5 is shown in Fig. 8. Transistor VT6 and diodes D1, D2 and D3 form the 2-gate while transistors VT1, VT2, VT3, VT4 and VT5 form the half-adder. Diode D4 isolates the 'carry' output for merging with the output of the 2-gate of the succeeding half-adder stage.

The operation of the 2-gate has already been described in Section 6. The inputs at terminals 3 and 4 are represented in Fig. 7 as the A and B inputs, while the functions of NOR 1, 2, 3, 4 and 5 are performed (Fig. 8) by transistors VT1, 2, 3, 4 and 5.

With no inputs present, the transistor NOR gates will be shut off due to the positive bias at the base via the 33 kΩ resistor from the +6V supply; no current will flow and a '1' (-6V) will be present at collector. With an input (-6V) present, base current will flow and cause the transistor to bottom, bringing the collector down to almost zero volts, which represents a '0' output. Thus the transistor acts as a natural NOR gate and the 'sum' and 'carry' outputs are propagated by connecting them in half-adder fashion as shown logically in Fig. 7.

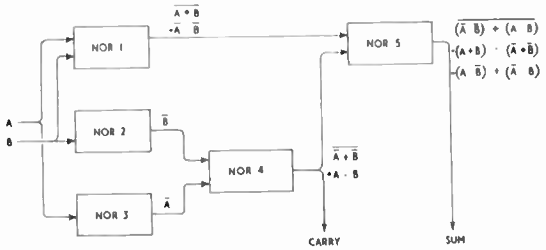


Fig. 7. NOR element half-adder.

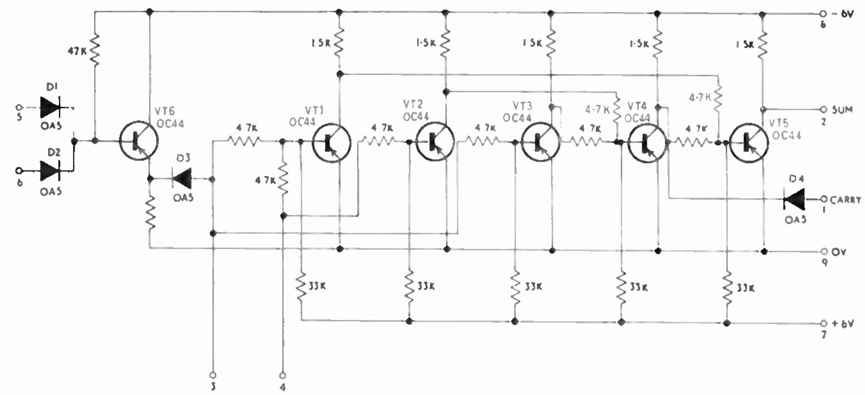


Fig. 8. Half-adder and two-gate.

8. Decimal-to-Binary Converter

The circuits have been simplified by using pure binary numbers throughout the computations, but this raises the problem of providing decimal number insertion facilities for the operators of the equipment. The M and N batching counters have decimal switches fitted[5] and it was decided to develop a simple decimal-to-binary converter using decimal switches at the input and providing a pure binary output from 000 to 999.

Full details of this converter are given elsewhere.[6] Basically, the powers of 10 from the decimal switches are routed to their equivalent binary channels (9×10^0 would be routed to binary 8 channel and binary 1 channel, 9×10^1 would be routed to binary 64, 16, 8 and 2 channels) and the resultant binary channel inputs are summed by means of relay half-adders. An example is given in Fig. 9 of two binary channels, one providing for two inputs using one half-adder and the other providing for three inputs and a 'carry' from the previous channel using three half-adders. The requirements for half-adders are discussed in Appendix 2 and the logics are given in Fig. 7; as the switching speed was unimportant a relay was used to perform the function of half-adder very simply.

Figure 10 shows this novel half-adder using a double-coil relay with coils wired in opposition. When both coils are energized, owing to the presence of signals A and B, the effect is cancelled out and contact AB2 remains closed and a 'carry' signal is generated ($1+1$ = '0' carry '1'). When either coil

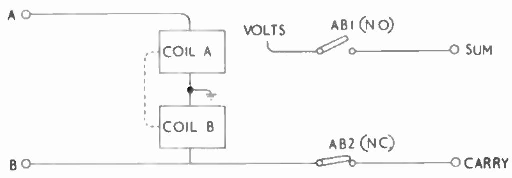


Fig. 10. Relay half-adder.

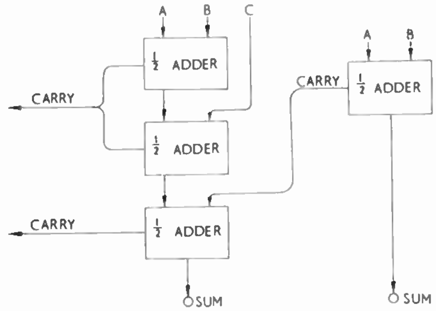


Fig. 9. Summing of binary channel.

A or B is energized, but not both, then the relay operates and contact AB1 closes and generates a 'sum' output (($1+0 = 1$ carry 0) or ($0+1 = 1$ carry 0)). Contact AB2 will be opened and no 'carry' will be generated. If no inputs are present the contacts will remain inoperative but no 'carry' will be generated through AB2 as input B is absent.

Twenty such half-adders are necessary to handle decimal numbers up to 999 and the binary output is available on ten wires representing binary channels 512 to 1.

9. Construction

Figure 11 shows the novel method[8] of construction used to avoid the use of plugs and sockets and yet give easy access to component parts. The components are inserted into perforated boards and terminals are arranged at the ends only. Ordinary wander plugs are inserted into slots cut on the sides of the boards and these permit a stacking arrangement which can be opened up in book fashion for inspection purposes.

Main accumulator flip-flop stages and half-adder stages are accommodated four per board and the control stage is on one board. The computer is fitted in a standard 19-in panel and is housed, along with the batching counters and decimal-to-binary

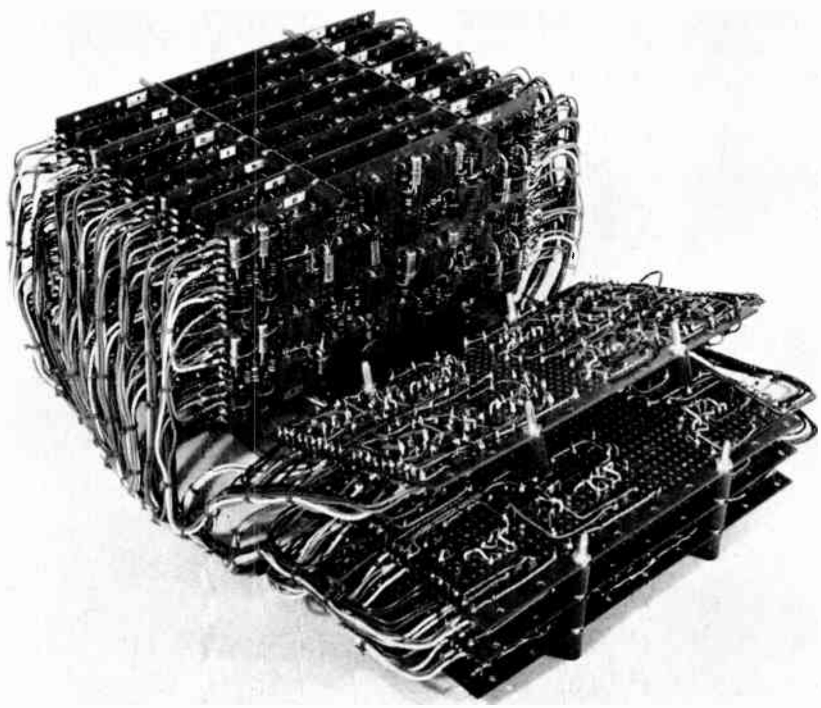


Fig. 11. Computer open for inspection showing construction.

converter, in a standard 19-in cabinet. Interconnections are made internally and the only external connections are supplies, 24V d.c., input frequency (f) and output frequency [$(M/N)f$] (Fig. 12).

10. Operation and Conclusions

The equipment is simple to operate. The desired ratio M/N is set up on the M and N batching counters; value M is also set up on the decimal-to-binary converter. The frequency f is inserted at the appropriate socket and the output frequency is available at socket $(M/N)f$. Input frequency f can be pulse or sine wave and an internal pulse shaping unit only requires that the waveform passes through zero twice per cycle. The peak-to-peak amplitude can range from 1 to 240V; the shaper is similar to one used and described in ref. 7.

For present purposes the output $(M/N)f$ is actually taken from CFF3 (Fig. 4, control logic) and the waveform is differentiated, then has positive spike removed to give a negative pulse suitable for comparing with the input pulse.

11. Acknowledgments

The author acknowledges the advice and encouragement of Mr. W. H. P. Leslie, the early logical proving work carried out by Mr. D. Nairn and the

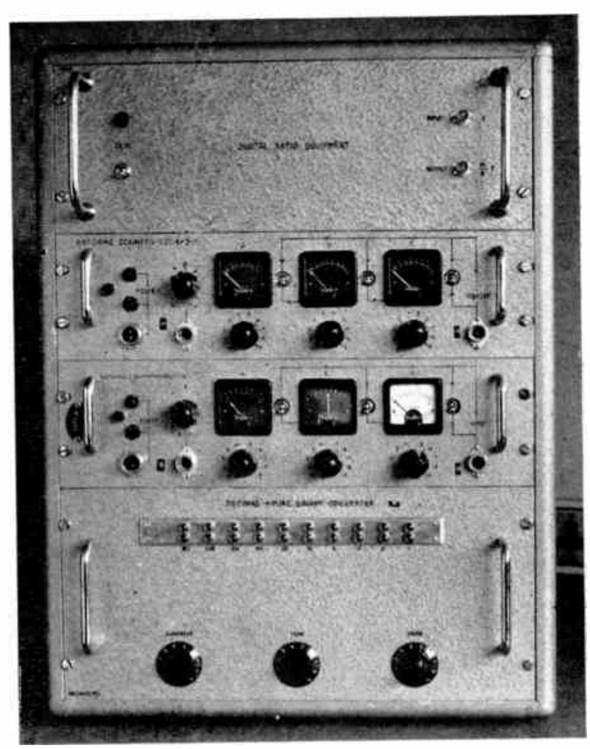


Fig. 12. Complete digital ratio equipment.

work of Mrs. C. Underhill in the preparation of Appendix 1. The author also thanks the Director of the National Engineering Laboratory for permission to publish this paper.

12. References

1. C. Timms, "Recent developments in spur and helical gears", *Prodn Engr*, **39**, No. 6, pp. 321–40, 1959.
2. W. H. P. Leslie, "Widening the application of diffraction gratings", *Internat. J. Mach. Tool Des. and Res.*, **2**, pp. 393–411, 1962.
3. C. V. L. Smith, "Electronic Digital Computers", (McGraw-Hill, New York, 1959).
4. J. Boswell, "NOR logic", *Instrum. and Control Syst.*, **33**, No. 9, pp. 1523–25, 1960.
5. D. Nairn, "An improved batching counter", *Electronic Engng*, **34**, No. 407, pp. 38–39, 1962.
6. D. Nairn, "A simple decimal-to-binary decoder using relays", NEL Report, No. 44, National Engineering Laboratory, East Kilbride, Glasgow, 1962.
7. H. I. Messer and W. H. P. Leslie, "An automatic self-checking transistor counter with digital display", *Electronic Engng*, **33**, No. 402, pp. 484–89, 1961.
8. R. Howatson, "An easy-access unit assembly", *Electronic Engng*, **35**, No. 421, p. 165, 1963.
9. W. H. P. Leslie and D. Nairn, "A fast counter for adding or subtracting randomly related pulse trains", *Electronic Engng*, **34**, No. 410, pp. 227–33, 1962.

13. Appendix 1

Mathematical Treatment of Delay Equation

To transform a frequency f into a frequency $(M/N)f$ we can consider the problem for $f = 1$. In a period of N units of time, the beginning of this period beginning with a pulse, N input pulses will occur. Introduce set of pulses with frequency q, where q is so large that the original pulses occurring at times $1, 2 \ldots N$ coincide nearly enough with a q pulse.

Take the first M pulses in each group of q pulses and eliminate the others. Each M pulse can be specified by a pair of numbers (α, β) where the time of the pulse from time 0 ($1, 2 \ldots N$ pulses) is

$$\alpha + \frac{\beta}{q} (0 \leqslant \beta \leqslant M)$$

We can number the M pulses from 0 to NM, where pulse (α, β) has the number $\alpha M + \beta$.

Choose every Nth pulse starting at $t = 0$. There are M of these pulses; they are pulses $N, 2N, 3N \ldots MN$.

Consider pulse λN, $0 \leqslant \lambda \leqslant M$

$$\lambda N = \frac{\lambda MN}{M} = \frac{\lambda N}{M} M$$

$$= M\left[\frac{\lambda N}{M}\right] + M\left\{\frac{\lambda N}{M}\right\}$$

where $[x]$ denotes integral part of x, and

$\{x\}$ denotes remainder of x.

This pulse is number $\alpha M = \beta$

where $$\alpha = \left[\frac{\lambda N}{M}\right] \qquad \beta = M\left\{\frac{\lambda N}{M}\right\};$$

its distance from time 0 is

$$\alpha + \frac{\beta}{q} = \left[\frac{\lambda N}{M}\right] + \frac{M}{q}\left\{\frac{\lambda N}{M}\right\};$$

its distance, S, from the nearest N pulse to the left is

$$\frac{M}{q}\left\{\frac{\lambda N}{M}\right\} = S$$

If the M pulses are now equally distributed along time N the distance of the λth pulse from 0 is $\frac{\lambda N}{M}$,

therefore $$t_d = \frac{\lambda N}{M} - \left(\left[\frac{\lambda N}{M}\right] + \left\{\frac{\lambda N}{M}\right\}\frac{M}{q}\right)$$

$$= \left\{\frac{\lambda N}{M}\right\}\left(1 - \frac{M}{q}\right)$$

But $$R = q - M,$$

i.e. $$1 - \frac{M}{q} = \frac{R}{q},$$

therefore $$t_d = \left\{\frac{\lambda N}{M}\right\}\frac{R}{q} = \frac{S \times R}{M}$$

14. Appendix 2

Half-adder Requirements

The 'half-adder' must satisfy the terms $0+0 = 0$, $1+0 = 1$, $0+1 = 1$ and $1+1 = 0$ carry 1. It provides for two inputs and possible 'sum' and 'carry' outputs. Several combinations of logical elements will perform this task, but for simplicity it was decided to use NOR elements[4] throughout and this arrangement is shown in Fig. 7 along with the Boolean algebraic terms at each stage; handling of these terms is done by use of De Morgan's theorem which states that a logic expression can be inverted simply by

(1) inverting each term of the expression;

(2) changing each AND (.) element into an OR (+) element; and

(3) changing each OR element into an AND element.

The NOR element is a combination of a NOT and OR gate. Two inputs, A and B, fed to a NOR element will yield an output when the input conditions are $\overline{A+B}$ (not, A or B) which simplifies (according to De Morgan) to $\overline{A} \cdot \overline{B}$ (not A and not B).

15. Appendix 3

Performance

Tests carried out on the equipment have shown the performance to be as predicted. Figure 13 shows the jitter present in the output pulse. The sweep speed is 50 μs/cm and the jitter is about 30 μs, which represents one 'clock' pulse interval at 30 kc/s. The jitter shows up as a continuously variable effect due to the random relationship between input frequency and 'clock' frequency encountered in practice. The camera exposure for this oscillogram was 1 second.

Figure 14 shows in the top trace the output waveform P_d, and in the lower trace S pulses above and $P_0 - P_d$ pulses below the base line. All are for an M/N ratio of 4/5. Thus the pattern of S pulses achieved is that shown in Fig. 2. The phase relationship of 'clock' pulses to 'gear' pulses for this oscillogram was controlled by deriving 'gear' pulses from 'clock' pulses via a batching counter.

When used in practice with the single-flank gear tester, the M/N ratio is of the order of 30/80 but, to produce readable oscillograms, a lower ratio is required. Some interesting waveforms at a ratio of 12/13 are shown in Fig. 15. The top trace shows the waveform at CFF3 (Fig. 4), the output pulse P_d coinciding with the positive-going edges of this wave-

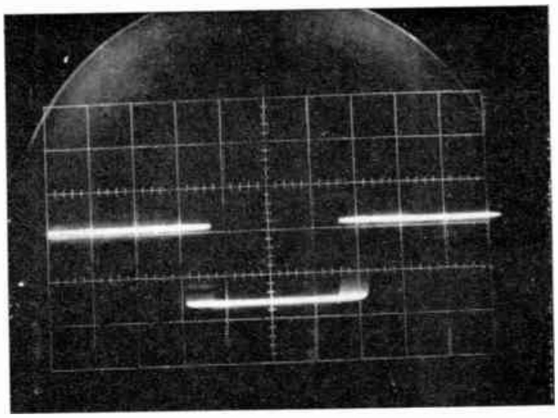


Fig. 13. Output waveform oscillogram showing 'jitter'.

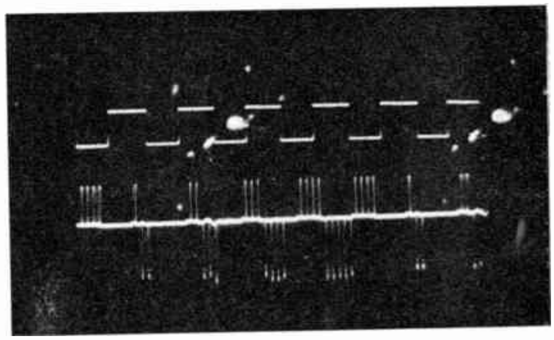


Fig. 14. Oscillogram of $M/N = 4/5$.

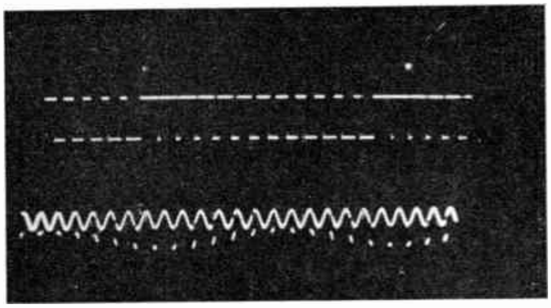


Fig. 15. Oscillogram of $M/N = 12/13$.

form. The bottom trace shows the pulses P_d in relation to the sine-wave input frequency for the same ratio (12/13). This method of display indicates clearly almost two cycles of 12 pulses which are at various levels on 13 input cycles.

Before using the electronic digital ratio equipment with the single-flank gear tester it is essential to check its performance. By suitable switching at a control panel the arrangement shown in Fig. 16 is obtained. The two outputs f/N should be locked in phase for any desired ratio of M/N; the only stipulation is that

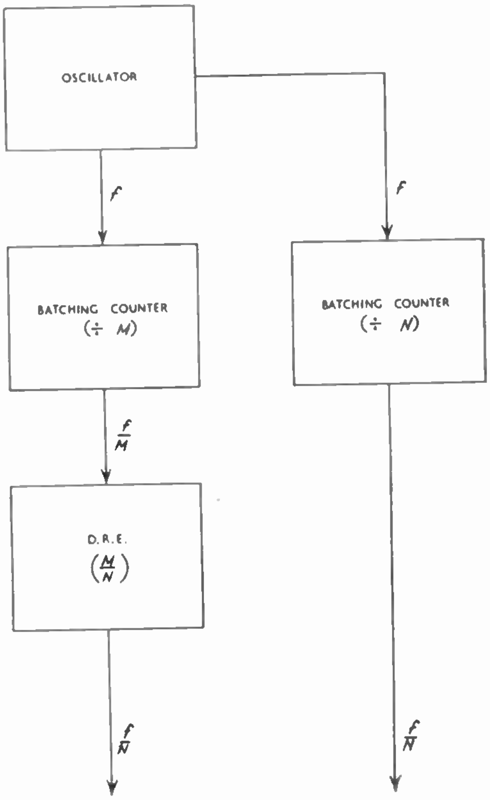


Fig. 16. Block schematic of test apparatus.

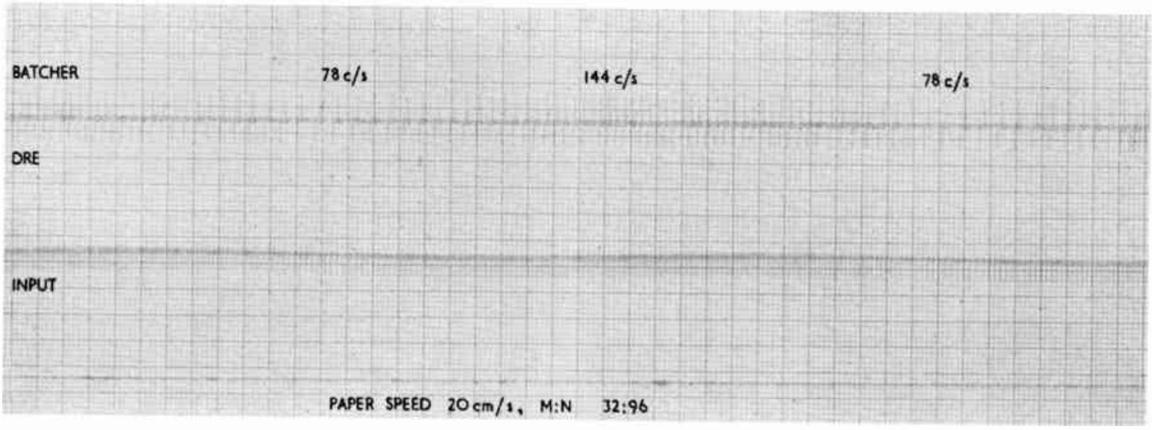


Fig. 17. Record of f.m. tests, ratio 32/96.

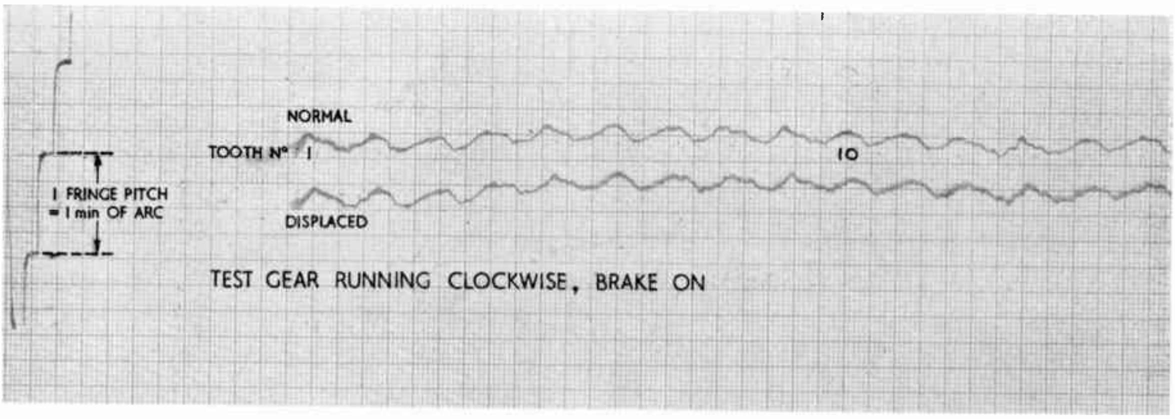


Fig. 18. Part of a record of gear-wheel irregularities.

the oscillator frequency should be kept above $(200 \times M)$ c/s so that the capacity of the main accumulator may not be exceeded.

Figure 17 is a record of frequency modulation tests carried out to determine the frequency response of the entire equipment. Any error caused by a change in input frequency is due to the remainder pulses R being accumulated during any cycle, say n, and being applied to the delay equation $S \times R/M$ during cycle $n+1$. If the error in the remainder in cycle n is called R_e then the error in cycle $n+1$ will be

$$P_{dc(n+1)} = \frac{S \times R_{e(n)}}{M} \qquad \ldots.(4)$$

S can never be larger than M and the worst possible case will be when $S = M$. Equation (4) will then be

$$P_{dc(n+1)} = R_{e(n)} \qquad \ldots.(5)$$

Thus, in the worst possible case, the difference of a single 'clock' pulse period between two adjacent cycles will result in an error of one 'clock' pulse in P_d in the second cycle. Alternatively, it can be stated that, at worst, the percentage error will be equal to the percentage change in wavelength per cycle.

The practical application of the equipment to a single-flank gear tester gives a trace as shown in Fig. 18. The phase comparator used to obtain this record is a differential counter described elsewhere.[9] The 'stepped' waveform shows the levels corresponding to one count in the counter which, in turn, corresponds to one fringe pitch on the radial grating and this gives a calibration of 1 minute of arc to the recording. Two recordings of the same 96-tooth gear-wheel are given to show the reproducibility of the entire system. There is a slight phase error between the graphs due to manual operation of the recorder 'on-off' during second recording.

16. Appendix 4

When the digital ratio equipment was used in practice with an input frequency train which was completely random with respect to the 'clock' frequency train, unexpected errors were found to be present and were traced to the random coincidence of pulses at several gates in the control unit (Fig. 4). For example, in the formation of M and R pulses at CFF1, G1 and G2, there was, at certain instances, coincidence between the arrival of a 'clock' pulse and the arrival of a 'gear' pulse. This can result in one of several possibilities (see Fig. 19). Assuming an M count of 2, the formation of M and R pulses at A and B is correct, but at C an R pulse has been generated plus an M pulse which has 'squeezed' through G1 when CFF1 was switched 'on'. At D a 'clock' pulse has occurred simultaneously with the switching of CFF1 and this has resulted in neither an R nor M pulse being generated. To provide for such possibilities, the logics for the insertion of g pulses has been modified as shown in Fig. 20. In this diagram the additional elements are CFFA and GA, flip-flop and 2-gate respectively. In operation the action is as follows: the arrival of a g pulse switches CFFA into the '1' state, thus storing the event; the next clock pulse passes via GA to the one-shot multivibrator, simultaneously generating, via G2, a further R pulse. To avoid switching CFF1 too quickly the output from the one-shot multivibrator is slightly delayed and becomes a time-corrected g pulse. This is used as such throughout the remainder of the control circuit. In cases of coincidence at GA the one-shot multivibrator decides whether or not the output from GA is treated as a pulse by either responding or not. If the one-shot multivibrator generates an output pulse then CFFA and CFF1 are reset and set respectively; otherwise switching is delayed until the next 'clock' pulse time. There can be no ambiguity at G1 and G2 where the M and R pulses are generated. The remainder of the circuit is as Fig. 4 with corrected g pulses used.

In the main accumulator some trouble was experienced due to flip-flops switching too rapidly and opening other clock gates. To overcome this with a safe margin the circuit shown in Fig. 21, in which the output is taken from across the capacitor, has been added to the output of each flip-flop.

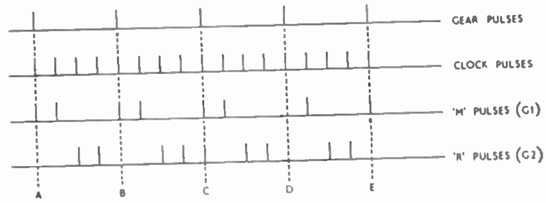


Fig. 19. Random coincidence of pulses.

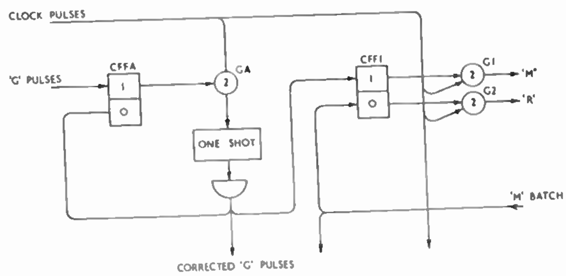


Fig. 20. Circuit modification giving corrected g pulses.

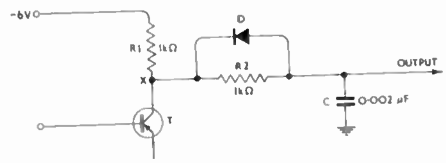


Fig. 21. Circuit modification to give equal time constants.

When T conducts, point X falls to approximately zero volts, the negative charge on C leaks off via R2, and D is biased off. When T is non-conducting point X rises to approximately -6 V, C is charged to this level via R1, and R2 is short-circuited by D which is now conducting. This system ensures an equal time constant during the on/off transitions.

Manuscript received by the Institution on 6th September 1963. (Paper No. 888/C68.)

CANADIAN PROGRAMME FOR IONOSPHERIC RESEARCH BY SATELLITE

Agreement has just been concluded between Defence Research Board (D.R.B.) and the U.S. National Aeronautics and Space Administration (N.A.S.A.) for the multi-satellite upper atmosphere research programme reported last year.

The schedule has been revised from earlier reports, probably due, in part, to the fact that *Alouette I* is continuing to work at full capacity. The new programme calls for comprehensive studies of the ionosphere from the approaching period of minimum solar activity through to the next period of maximum activity. Four satellites will be launched between 1965 and 1970 and bear the names *Alouette II*, ISIS A, B and C. (ISIS is derived from International Satellites for Ionospheric Studies.)

Under the terms of the new agreement D.R.B. will design, construct and test the spacecraft, provide the basic experiment (the topside sounder), and operate at least one ground station capable of supplying data on spacecraft operation.

N.A.S.A.'s primary obligations include providing four launch vehicles, placing the spacecraft in orbit, and providing and launching up to five sounding rockets for testing components and subsystems and determining the compatibility of experiments.

A joint working group, composed of representatives of the two government agencies, will be responsible for implementing the programme. All data obtained by the satellites will be available to the scientific community.

Alouette II, resembling its predecessor in size, shape and weight, will repeat some of the experiments of *Alouette I* for confirmation, but will extend them substantially as it passes over the earth's poles in an elliptical orbit varying in altitude from 460 to about 1,600 statute miles. (*Alouette I* is following a circular orbit over the poles at a 630-mile altitude.)

Along with *Alouette II*, N.A.S.A. will launch a Direct Measurement Explorer satellite similar to *Explorer VIII*.

Components similar to those carried in *Alouette I* are being designed to conform with the increased altitudes of the new satellites. A magnetometer and solar aspect sensor will be included to provide information on the satellite's orientation at all times.

Experience gained from *Alouette I*, combined with the planned increase in altitude of *Alouette II*'s orbit, have convinced scientists they will need greater frequency range for the new sounder. Power will also be increased because of the 1,600-mile apogee.

Alouette II will conduct five experiments. They are (*a*) sound the ionosphere from inside and above the main layers; (*b*) measure cosmic noise; (*c*) monitor v.l.f. emissions, or whistlers; (*d*) detect energetic particles; and (*e*) determine the temperature of electrons in the vicinity of the satellite. The first four experiments are Canadian, but the fifth is designed by N.A.S.A. Information will be telemetered to 13 world-wide ground stations on a real-time basis.

Launching is planned for the first half of 1965 from the Pacific Missile Range, California.

Plans for the ISIS satellites are still in the preliminary stages. ISIS A will be a completely new spacecraft, but will carry experiments similar to those in *Alouette II*. Additional experiments, such as electron and ion probes, which measure the temperatures and densities of the hydrogen, helium and oxygen present at satellite altitudes, and an ion mass spectrometer to confirm the identification of the individual ions, will probably be included.

Firm specifications for ISIS A have not been determined. Design began last month and launching is planned for 1967.

Because ISIS B will conduct more advanced experiments than its predecessors, its design will depend on data returned from *Alouette II* and the Direct Measurement Explorer. Design and launch time have not been discussed in specific terms.

The design of ISIS C, which will closely resemble ISIS A, is expected to begin in 1966 for launching during the next maximum sunspot activity, expected early in 1969.

Associated with Defence Research Telecommunications Establishment (D.R.T.E.) thoughout the programme will be research laboratories in the U.S., Britain and other countries. N.A.S.A.'s Office of Space Science and Applications will have overall responsibility for U.S. participation in the programme. The Goddard Space Flight Centre will participate in data analysis and carry out the companion sounding rocket experiments and the Direct Measurement Explorer project.

Canadian industry will play a major role in the new satellite programme. Companies will work closely with D.R.T.E. scientists and engineers on design and construction of *Alouette II* and ISIS A. It is expected that industry will assume most of the responsibility for ISIS B and C, leaving D.R.T.E. personnel to plan experiments and interpret the scientific information derived.

RCA Victor Co. Ltd. has been named prime contractor for the four new satellites, and The De Havilland Aircraft of Canada Ltd. has been named associate contractor.

The ISIS series is intended to fill the need for obtaining detailed information about the ionosphere for the remainder of this decade. Invitations will be extended to other scientific investigators to suggest and design experiments for the programme. In Canada, D.R.B. will invite scientific agencies interested in the upper atmosphere, primarily Canadian universities, to propose experiments. N.A.S.A. will solicit proposals from the U.S. and countries other than Canada.

Reprinted from *Canadian Electronics Engineering*, February 1964.

U.D.C. 621.397.743

Pay Television by Wire

By

P. BASS†

Presented at a meeting of the Television Group in London on 29th May 1963.

Summary: The distribution of pay television programmes by wire eliminates coding and decoding equipment required for an over-the-air system in order to prevent unauthorized reception. An outline is given of the advantages of h.f. multi-pair type networks in this and other respects. Two wired systems of pay television are described, and some factors which have governed their development are discussed.

1. Introduction

It is not always fully appreciated that the success of pay television will depend to a large extent on the technical system employed being comparatively simple in design and reliable in operation, two factors which by the very nature of things generally go hand in hand. A lesson soon learned by the system designer is that the cost of providing and maintaining the necessary equipment in each home must be recovered from the subscriber's expenditure on programmes, and any preconceived notions of developing elaborate control schemes requiring complex and expensive subscriber equipment must be dismissed, or commercial failure of the operation will inevitably follow. At the same time, comprehensive facilities are demanded for pay-television operations: it should be possible to offer a subscriber the alternatives of paying for programmes on a credit basis, or by direct contributions to a coin unit; the method of pricing should be sufficiently flexible to accommodate a wide range of programme material; and means for evaluating the total receipts for each programme at the time of transmission should be provided.

All these requirements are met by two particular wired systems of pay television which have been developed by the author's company. The systems differ chiefly in the method of charging for programmes. With one system, a subscriber pays in full the particular price specified for each programme he decides to accept, while with the other he pays for his viewing time at a rate which is adjusted in relation to the value of each programme concerned. Both systems will serve any type of television receiver in general use, and in no instance is it necessary to make inside connections to a receiver or in any way to interfere with the reception of existing programmes. The systems have also been designed to be adaptable for colour, 525 and 625-line operations. Before describing these two systems in more detail, it is proposed to mention briefly a number of fundamental factors which have governed their development.

† R. & R. Research Ltd., 1–3A Westville Road, London, W.12.

2. Choice of Network System

The operation of a pay television system over wire, as opposed to over the air, has the very great advantage that given a choice of the basic wired network to be employed it is possible to avoid the very costly business of scrambling at the transmitter and unscrambling at each receiver in order to prevent unauthorized reception. V.h.f. coaxial networks, or communal aerial networks as they are sometimes called, are not altogether satisfactory in this respect. Such networks invariably radiate to some extent, so that anybody in the vicinity having a sensitive receiver with a high-gain aerial might find that he could tune in a programme, and it would be almost impossible to trace such acts of piracy. It would also be no easy matter to prevent a pay-television subscriber from receiving a programme until he had 'paid', or to isolate permanently an existing wired subscriber who was not prepared to accept the pay-television service at all; the design of the necessary filters capable of rejecting the pay channels while leaving the other channels unaffected presents a formidable problem, particularly at Band III frequencies. Some form of scrambling or out-of-band transmission of the pay programme signals would thus seem inescapable if v.h.f. networks were used.

H.f. multi-pair networks, on the other hand, present no undue difficulties over maintaining security. On networks of this type each programme is normally carried by a separate pair of wires and all the vision signals occupy similar or overlapping frequency bands. Any radiation from the networks would therefore comprise a mixture of several programme signals which could not be separated by a pirate. At all events, the programme sound signals are normally distributed at audio frequencies and could not be obtained. The problem of designing filters to reject the pay programme channels does not arise, since reception can be prevented by simply disconnecting the pairs concerned from the input circuit to the subscriber's receiver. These networks hold another important advantage for pay television in that the individual pairs may also be used for the direct control of the subscriber coin and credit units, thus avoiding the

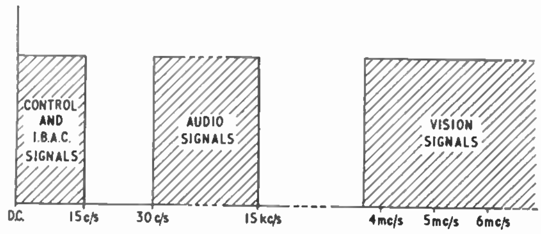


Fig. 1. Frequency bands occupied by pay television signals.

need for amplifying and demodulating stages at each installation which would otherwise be necessary with a carrier-operated control scheme. In addition, the same pairs may be used to measure the total receipts for each programme by means of a simple and accurate method of audience counting known as i.b.a.c. (instantaneous broadcast audience counting). In view of these very considerable advantages, the designs of the two pay television systems under consideration have been based on the h.f. multi-pair type of network.

3. Control and Audience Counting Circuits

The frequency bands occupied by the various pay-television signals are shown in Fig. 1. In accordance with standard practice, the programme sound signals are distributed at audio frequencies and the vision signals in the h.f. band; each vision carrier is actually on a frequency of approximately 5 Mc/s with full upper sidebands. The price control and i.b.a.c. signals are confined to the sub-audio frequencies and d.c., and thus interfere with neither sound nor vision reception. With the system designed to sell programmes rather than time, a subscriber in accepting a programme causes a latching switch to operate which connects the network pair concerned to the input circuit of his receiver and to the pay television control and i.b.a.c. circuits; the latching switch remains locked in the 'accept' position until the end of the programme and cannot be released by the subscriber, although he is free to go over to the I.T.A. and B.B.C. channels at any time. At the end of the pay programme, control and i.b.a.c. signals of the form shown in Fig. 2 are transmitted. First of all a series of short pulses with relatively fast rise-times and each representing two-

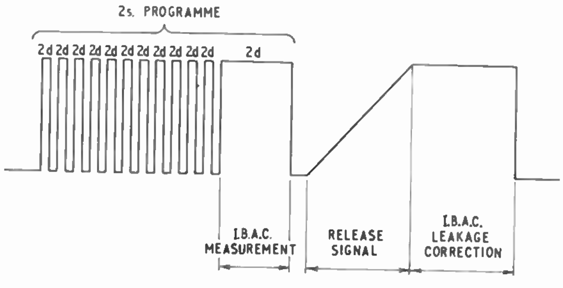


Fig. 2. Pay television control and i.b.a.c. signals.

pence in value are transmitted to register the price of the programme. The last of these pulses is of longer duration than the rest to enable the i.b.a.c. measurement to be made. This is followed by a 'ramp' signal which is recognized by an integrating circuit in the subscriber's unit and is applied to release the latching switch. The cycle is concluded with a measurement of the network leakage current necessary to correct the i.b.a.c. result.

Figure 3 shows the general arrangement of the subscriber's control and i.b.a.c. circuits. S1 is the latching switch which, for the convenience of the subscriber, is operated by a key; it is thought that a subscriber might wish at times to prevent his children or other members of the household from running up a large bill while he is away from the house. The control and i.b.a.c. signals are filtered from the sound and vision signals by the h.f. and l.f. inductors, and are applied to the coin mechanism or credit meter, as the case may be, through a differentiating circuit and a constant current regulator. Pulses with fast rise-times are passed by the differentiating circuit thus actuating the price registering solenoid and, at the

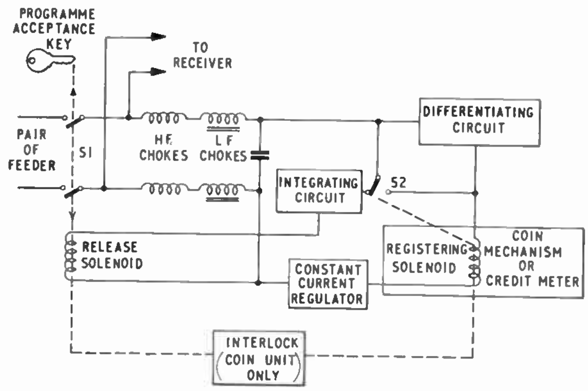


Fig. 3. General arrangement of subscriber control and i.b.a.c. circuits.

same time, causing switch S2 to bypass the differentiating circuit. In this condition, the control current drawn from the network is held to a predetermined constant value by the regulator, and during the last pricing pulse of the sequence, which it will be recalled is longer than the rest, the total current flowing into the network is measured at the central station. This in fact is the main i.b.a.c. measurement, and it will be appreciated that the reading obtained represents the total control current drawn by all the subscribers plus the network leakage current which will vary with the atmospheric conditions. The i.b.a.c. measurement is followed by the transmission of a ramp function rising at a constant rate which is disregarded by the differentiating circuit and accepted by the integrating circuit thus causing the release solenoid to return the latching switch to the 'off' position.

Finally, when all the subscribers have been disconnected from the programme circuit in this manner, the network leakage current is measured and subtracted from the previous i.b.a.c. reading to give the true value of total control current. Since each subscriber equipment is designed to draw the same control current irrespective of the network loading conditions, a further simple calculation provides a figure for the number of subscribers who have accepted the programme and hence for the total revenue from it, from which the programme supplier is paid his due proportion. With some systems of pay television which do not incorporate this feature, individual viewing recorders are provided and the system operator must go to the additional expense of periodically collecting and analysing the records produced. I.b.a.c. therefore represents another important step in the reduction of costs of pay-television.

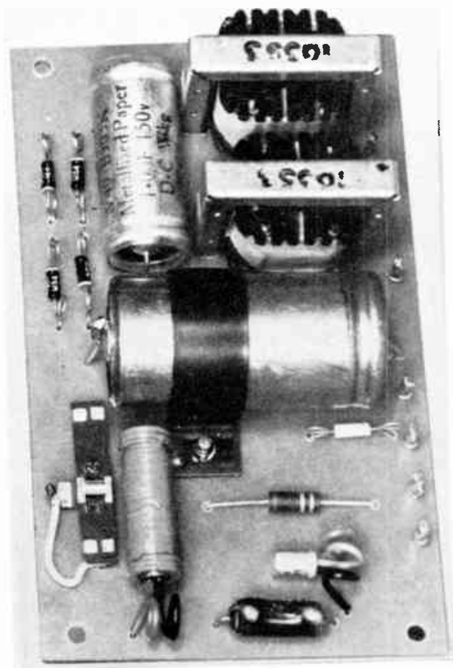


Fig. 4. Printed circuit assembly of subscriber control and i.b.a.c. components.

4. Programme Payment System

Returning to the basic control scheme for the programme payment system, it will be evident that the charge for each programme is normally not registered until the end of the programme and this has the important operational advantage that if a programme is interrupted for any reason or cancelled altogether, the charge may be adjusted accordingly or not made at all, so avoiding the necessity to refund the subscribers' money or alter the credit accounts. An idea of the degree of design simplicity that has been achieved in the system may be obtained from the printed circuit assembly shown in Fig. 4, which embodies all the components of the subscriber control and i.b.a.c. circuits with the exception of the latching switch; the circuits are completely passive and require no local power supply. Figure 5 shows the subscriber coin mechanism which has been designed to accept all silver coins of current denominations. The charging pulses are applied to the incorporated registering solenoid which causes the indication on the displayed credit dial to decrease from zero into the red region by a series of small steps, the total decrease corresponding to the price of the programme previously accepted. (The illustration shows a registered debit for a six-shilling programme.) An interlock system between the coin mechanism and the latching switch prevents the subscriber from accepting a further programme until he has inserted sufficient coins to cancel the debit.

Fig. 5. Subscriber multi-coin mechanism.

He may, should he so wish, insert a greater amount to store credit for future programmes, in which case the indication on the dial would enter the credit region to the extent of the excess amount. On the subscriber credit meter, used as an alternative to the coin mechanism, the registering solenoid simply increases the displayed total reading by an amount equal to the price of the programme previously accepted. The meter itself has been designed to be installed remotely from the subscriber's receiver, either outside the house or in the garage, so that it can be read without needing access to the house and disturbing the subscriber.

5. Time Payment System

The second system under consideration, with which a subscriber pays on the basis of his viewing time, is similar in many respects to the programme payment system just described. The main points of difference in the design of the subscriber unit are that the integrating circuit is omitted altogether, and the latching switch is replaced by a simple off/on switch operated directly by the acceptance key. As regards the general control scheme, the charging pulses in this instance are transmitted individually at intervals throughout each programme and only those pulses which occur while the subscriber's unit is actually switched on and adjusted to receive the programme are registered. The rate of charge is varied with the type of programme material concerned by altering the repetition rate of the pulses. An i.b.a.c. measurement is made as each pulse is transmitted, thus indicating the number of subscribers receiving the programme at the time, and the results are automatically recorded and totalled to provide a figure for the revenue which the programme has attracted. The design of the subscriber coin mechanism is also slightly different for the time payment system. No interlock with the acceptance key is needed, but a cam-operated switch is incorporated to disconnect the subscriber from the pay channels when his credit has been exhausted. The design of the remote credit meter, on the other hand, is identical with that for the programme payment system.

Fig. 6. Pay television coin unit for aerial receiver.

To all outward appearances the pay television unit installed with a subscriber's receiver is similar for both systems. Figure 6 shows a coin-operated unit for a receiver using an aerial for the I.T.A. and B.B.C. programmes. The central station equipment needed for two pay channels is shown in Fig. 7. The two bays on the left of the picture contain programme distribution equipment, the two shorter bays price control equipment, and the bay on the far right contains the i.b.a.c. audience measurement equipment.

Manuscript received by the Institution on 29th July 1963. (Paper No. 889/T24.)

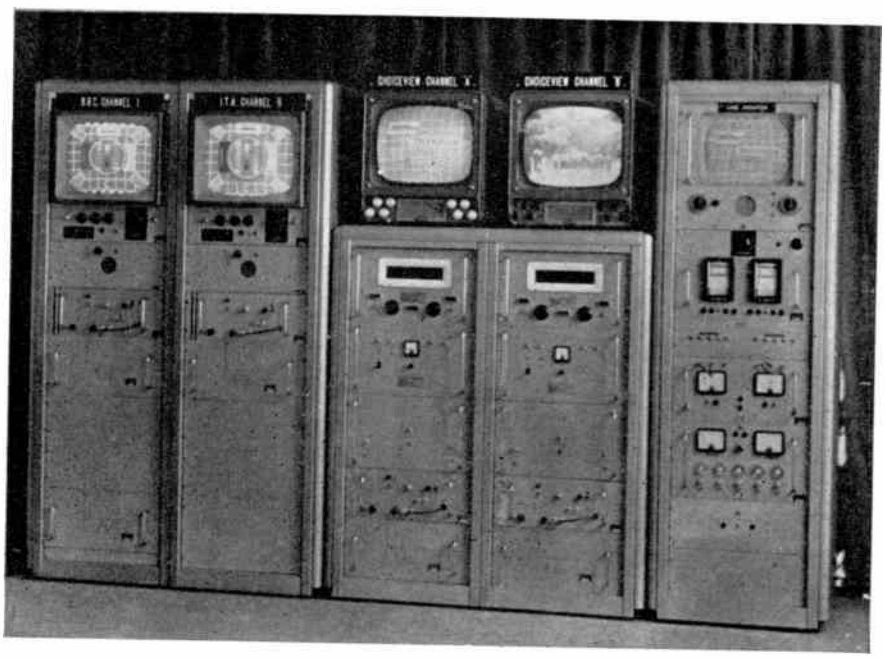


Fig. 7. Central station equipment for two pay channels.

U.D.C. 621.397.743

Standardization of Signal Specifications for Pay Television Systems

By
D. W. HEIGHTMAN
(*Member*) †

Presented at a meeting of the Television Group in London on 29th May 1963.

Summary: Desirable system standards are discussed in the light of operational requirements and particular reference is made to a proposed universal control signal which provides for a variety of pricing methods.

1. Introduction

Before pay television becomes established on a national scale it will be generally advantageous to agree on system standards, particularly in respect of control signal specifications. If systems are developed independently the ultimate change to standardization will take a great deal longer because of the capital investment in apparatus already installed.

In the context of this paper the description 'pay-television system' refers to (*a*) the transmission method, (*b*) any additions or modifications (coding) to the national vision and sound standards for pay-television purposes, (*c*) terminal units for installation at subscribers' homes.

2. Operational Requirements

The following is a summary of the main operational requirements from the viewpoint of the author's company:

(1) *Security.* There must be a basic security within the system to give protection against unauthorized reception.

(2) *Payment.* There should be provision as alternatives for both (*a*) prepayment by coin and (*b*) acceptance by the subscriber of charge on credit terms for each programme viewed. Full payment will be required whether the subscriber switches on at the commencement or during the run of a programme. There should be a minimum of delay in the operation of the payment mechanism—preferably not more than 15 seconds.

(3) *Free Reception.* The system should allow 'free' reception for pre-view, announcements, etc. In this connection a separate sound announcement channel as used on some North American systems, is not considered essential. 'Charged programme' signal/s are required to actuate the subscribers' control boxes (*a*) so that reception of charged programmes is only possible when the requisite coins have been inserted or acceptance of the charge has been recorded and (*b*) so that the payment procedure has to be repeated for any subsequent programme—unless it is a repeat performance for which no charge is made.

(4) *Price Indication.* Whilst not essential for credit operation (2*b*) it is desirable for coin prepayment that the system should provide a repetitive signal indicating price of current programme.

(5) *Local Record of Transactions.* For confirmation of charges in cases of dispute, or for credit operation, it should be possible to obtain a simple and reliable record of the programmes viewed by each subscriber, over a period of at least one month.

(6) *Audience Measurement.* It should be possible at convenient centres to obtain at least a sample measurement of the viewing audience at any particular time.

(7) *Future Development including Automatic Central Account Recording.* The system should, from the start, be capable of incorporating without major modification such refinements as 'central billing', where response signals from subscribers are generated, to provide central records for accounting purposes (i.e. without the need to visit individual homes).

(8) *No Television Receiver Modifications.* The system should not call for internal modifications to the normal television receiver—assuming the receiver is fitted with coils to give reception on the required v.h.f. channels, which may well be to the 625-line standard.

(9) *Multiple Programme Capability.* Whilst it is recognized that the cheapest and simplest system would cater for only one pay-television programme, it is regarded as essential that any system adopted eventually should allow for the simultaneous transmission of at least three programmes.

(10) *Economy.* The system should permit the use of relatively low cost and reliable subscriber control boxes, i.e. of the order of £10.

(11) *Control Signals.* Any signals used for control purposes (see (3) and (4) above) should, not only in the interest of channel space conservation, but also to

† British Home Entertainment Ltd., 72 Dean Street, London, W.1.

simplify subscriber control apparatus, be preferably contained within the normal television channel width or should be carried over the same cable at low frequency in order not to reduce the television channel capacity.

(12) *System Standardization.* It is highly desirable that there should be, on a national basis, general technical agreement on the pay-television control signal system to be adopted. In an experimental period it is recognized that several systems may be tested.

3. Meeting the Requirements

3.1. *Security*

Since the Government have laid down that 'over the air' transmission will not be permitted in the United Kingdom, the complex coding/decoding systems and equipment needed for radiated transmission will not be considered and this paper will deal with cable transmission only.

There are two main versions of cable transmission:

(1) The multi-wire h.f. type (Rediffusion and B.R.W.) using receivers with a special i.f. amplifier.

(2) Coaxial v.h.f. type (used by the remainder of the relay companies) using normal aerial-fed receivers.

Multi-wire. In multi-wire working, all television channels are transmitted in the 8 Mc/s region, one channel over each pair of wires in the cable, the present limit being normally four channels, i.e. eight wires or four pairs. Whilst attractive in the first case from an economy viewpoint, this system is less advantageous for future development for, say, three pay-television channels plus integration with other new services. In this paper it is taken for granted that cable systems will essentially carry a minimum of two B.B.C. and one I.T.A. as well as pay television.

Security, in the case of the multi-wire system, is simply obtained by restricting the pay-television channels to one (or more) pair of wires and switching to the receivers through suitable payment mechanisms. Viewers not subscribing to the pay-television service would not be connected to the multi-wire pair carrying this service.

Special systems, the technical details of which have not yet been released have been developed by B.R.W. and Rediffusion for multi-wire operation and will not, therefore, be further considered in this paper.

Coaxial. In coaxial working, several television channels are sent over the same cable, the number of channels being mainly dependent on the band-width and cross-modulation performance of the repeater amplifiers along the cable and the receiver adjacent-channel-rejection characteristics. The proposal for security here is that the pay television channels should be transmitted on non-standard frequencies within the h.f. to v.h.f. band, i.e. so that reception cannot be obtained on an ordinary receiver. The pay television control box within the subscriber's home will then, by means of a simple frequency changer tuner unit, convert the pay-television channel to an available channel on the receiver, i.e. one not used in that particular locality by B.B.C. or I.T.A. As a further measure of security the sound and vision carriers of the pay-television channel can be reversed in the spectrum so that they only appear in normal sequence when passed through the frequency changer unit.

It is anticipated that pay television will only be transmitted on the 625-line system. Receivers should, therefore, have provision for 625-line v.h.f. reception. (It is understood that many so-called dual standard receivers only provide for 625-line reception on u.h.f.)

3.2. *Payment and Control Signals*

Provided the alternative methods of payment likely to be required are borne in mind in the conception and design of the control signal specification, there would appear to be no reason why, on any given town installation, operators should not be given the optional alternatives of using coin prepayment subscriber control boxes or credit recording boxes on the one cable network. On new networks or old ones suitably adapted, when response signals from individual subscribers can be fed back over the networks (by suitable low-pass filters, etc.) the control system can also allow for central recording of subscriber accounts. Such a system will additionally give an almost minute by minute record of the viewing audience.

In considering the specification of any final control signal to meet the several requirements, it is logical first to set down the simplest and very minimum requirement for a control signal and then to build the more complex signal on this base.

3.2.1. Credit payment

It will be found that a simple credit system calls for the minimum of complexity in a control signal, whereas coin prepayment requires a price-coded control signal. For a simple credit system with a local record, film, paper tape, etc., in the subscriber's home, all that is required is a d.c. signal which is off during periods of 'free' reception and on at the commencement and through the run of a programme. (See Fig. 1.)

The operation of this simple system would be approximately as follows: a subscriber switches his control box selector to the required pay television channel. This physical movement aligns a cam movement to indicate the channel chosen. If he has switched on before programme commencement time he sees some minutes of preview and is reminded of the charge for the programme. He is also reminded

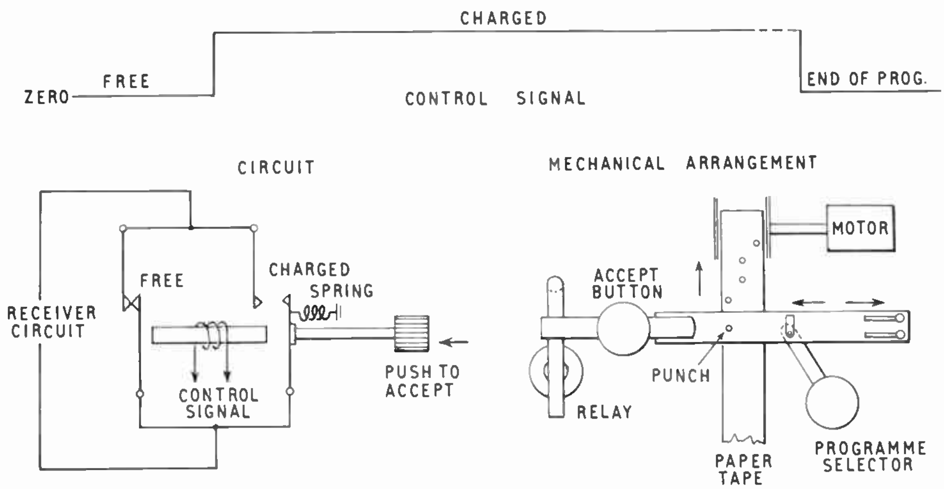


Fig. 1. Diagrams of main elements for simple local credit system for pay television.

to press his acceptance button before the programme can be seen. Some 15 to 30 seconds before the programme starts the control signal is applied; this energizes the 'charge' relay which opens the 'free' contacts and stops reception. The subscriber pushes the acceptance button. This is now held down by the energized 'charge' relay or solenoid. If the subscriber switched on during the run of a programme he would have to press the acceptance button before any reception could be obtained. The movement of the button is also used to mark or punch the film or paper tape which is passed slowly and continuously by a simple clock or similar movement in convenient relationship to the channel selector.

The subscriber can now view the programme until its end when the control signal will release the 'charge' relay, the 'free' contacts making again—until the next programme is due, when the sequence will be repeated. Periodically the tape or film record will be collected and this will show a series of marks indicating the programme viewed and time of selection. Processing this record to produce periodical accounts would be a simple matter. In developing more sophisticated systems from this very simple and reliable one, it will be important to remember that any modulation or coding should not break the control signal for any appreciable length of time, i.e. so that any pulses, etc., can be integrated to form a continuous control current.

3.2.2. Coin payment

There appear to be three practical possibilities for coin payment systems as follows:

Method 1. Variable length programme time for a given unit coin.

Method 2. Full payment at end of programme.

Method 3. Prepayment in full when viewing commences.

In method 1, which has the advantage of simplicity of apparatus, the subscriber control box will be fitted with a small motor and movement similar to that used in an electricity supply, or television receiver rental meter. The time for which the motor runs is, however, controlled by a periodically intermittent control signal. Thus the charge for a programme can be varied by adjusting the time period expiring for a unit coin.

Owing to the fact that the charge for the programme will not generally exactly equal the value of coins inserted, there can be a carry-over of coins in credit from one programme to the next. Alternatively, by viewing only part of a programme the subscriber need not pay in full. Thus, overall, there would not be a very satisfactory relationship between individual programme cost and coin value inserted.

Method 2 has the advantage of requiring only one set of price coded information on the control signal at, or after, the end of a programme. This arrangement simplifies the subscriber control box because, unlike the repetitive price coded signal required for method 3, only one price-code sequence has to be dealt with. The control signal, only necessary at the end of programme, can consist of a set of relatively slow pulses to operate a ratchet relay or similar incremental device and also stop reception by some means. The number of pulses would equal the charge for the programme in units of, say, 6 pence. Alternatively, the end of programme price signal would be a variable length d.c. one, the on-time related to programme charge.

This method, requiring payment for the previous programme before the next one can be viewed, is

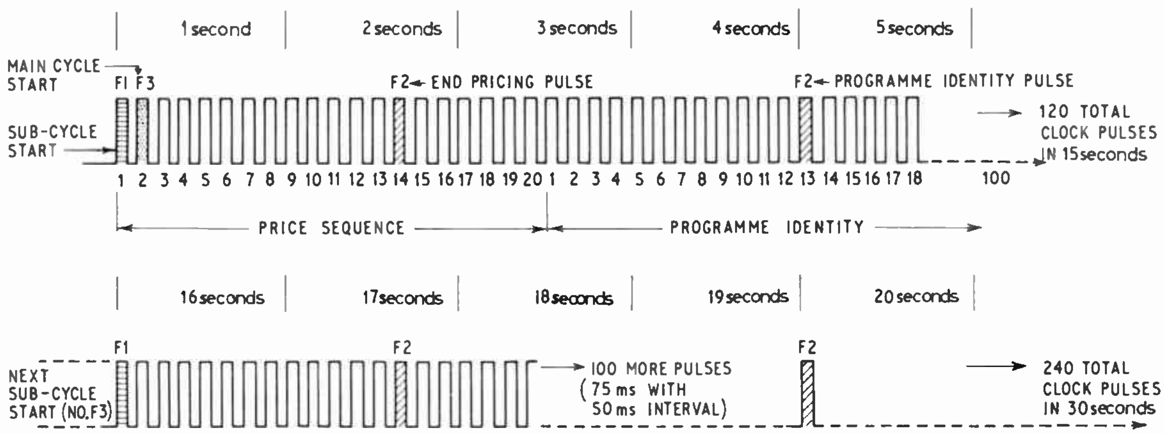


Fig. 2. Proposed universal control signal waveform.

F1, F2 and F3 represent modulation of clock pulses at suitably different audio (or r.f.) frequencies. F2 programme identity pulse can have same frequency modulation as F2 end-pricing pulse, because coin mechanism will not respond after receipt of first F2 pulse. The clock pulse rate is shown at 8 per second with a 60/40 duty cycle but these parameters could be varied if necessary.

considered to have operational disadvantages, particularly if a high-priced programme had to be paid for in order to view a much cheaper one. To ensure payment, arrangements in the control box will have to be made to 'hold' the subscriber until the programme end (even though he might switch his receiver off), and this also can be considered an operational disadvantage.

Thus, method 3 becomes the preferred one from an operational viewpoint. Systems to cover this type of operation have been previously described† and are, in fact, in operation in North America, particularly by International Telemeter. It is felt, however, that modifications to the Telemeter control signal specification could make for a greater flexibility of application and simplification of apparatus. The requirement that a subscriber should be able to switch on at any time means the provision of a repetitive price-code sequence at time intervals which are negligible. In the Telemeter system this cycle is 6 seconds. For reasons to be given below, it would be desirable for this period (later referred to as a 'sub-cycle') to be increased to the order of 15 seconds.

3.3. *Universal Control Signal*

Proposals for the control signal wave form are shown in Fig. 2. The 8 per second 'clock pulses' form the basis for all information carried. It will be noted that the first price-pulse sequence is similar to that of the Telemeter system, and could be used with a coin mechanism of that type, i.e. a 'pricing solenoid' stepping up a price indicator drum in unit coin increments, according to the number of price pulses. The proposed start pricing pulse, however, is identified by audio modulation F1 of the first clock pulse and similarly the end pricing pulse is a modulation F2 of the price pulse, which sets the price for a particular programme. Twenty pulses (increments) are allowed for pricing purposes but this allowance could be extended if necessary. The further proposals, in regard to the development of a universal control signal, represent, however, a considerable departure from the Telemeter scheme.

3.3.1. Coin mechanisms

Important requirements of such mechanisms and associated circuitry with a system of this type are:

(1) Price indicator mechanism must only start at start of price-code series, i.e. must not indicate if switched on during price-pulse sequence.

(2) Having recorded price level once the mechanism must not respond to subsequent price-code cycles during a given programme.

(3) The mechanism must return to the start position when released at the end of programme.

For these reasons the start-pricing and end-pricing pulses must be readily discriminated from the actual price pulses by simple and reliable circuits. Figure 3 shows the main elements of an electro-mechanical coin device to operate on the proposed signal.

Various other electro-mechanical devices and electronic circuits for dealing with this prepayment type of operation will, undoubtedly, occur to design engineers, and it is not proposed to go into details on this aspect. The aim will be rather to consider the development of the 'universal' control signal waveform.

† P. R. J. Court, *et al.*, "How a closed-circuit pay t.v. system works", *Electronics*, 33, 19th August 1960, pp. 49–55.

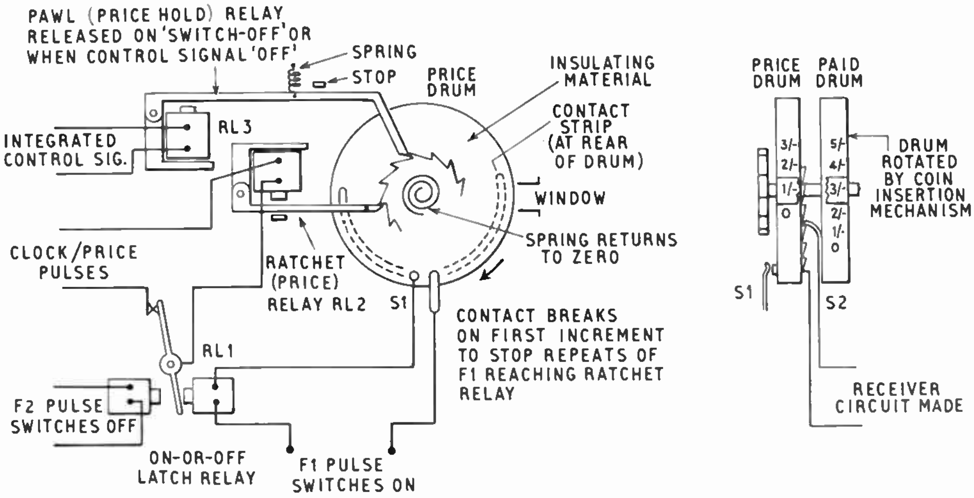


Fig. 3. Diagram of main elements of a coin payment system to operate on universal control signal for pay television.

Operation Sequence

On switch on: Contacts S1 *made* (from spring return to zero at end of previous programme): Contacts on relay RL1 *open.* First F1 pulse to be received makes RL1 contact. This puts clock pulses on to relay RL2 which steps up price drum. Pawl relay held by control signal. F2 pulse opens relay RL1 which cuts off clock pulses from RL2 at end of price sequence. Further F1 pulses will not repeat sequence because S1 contact open.

End of programme or switch off: pawl relay RL3 releases price drum to return to zero.

Paid drum rotates according to coin value inserted. Ratchet contact strip holds paid drum until end of programme when RL3 releases both drums. Excess payment catered for by over-travel on S2.

The coin mechanism can be considerably simplified and, therefore, reduced in price if designed for use with one value of coin (say, one florin) only. At most it is considered that two coin values are adequate.

3.3.2. Programme identification

The pulses from No. 21 to No. 120 can be used to identify the programme being received by a simple a.f. modulation of any one of the pulses. The counting would be from start pulse F1 to programme identity pulse No. X.

3.4. *Central Recording*

In order to render subscribers' accounts based on information automatically returned over the cable networks the minimum information required is:

A. Subscriber identity.

B. Programme identity (where more than one channel is available).

The problem is to pass this information from individual subscribers, several hundred, possibly, being connected on one cable. It is not practical to have individual lines from each subscriber to the exchange—as with the telephone. The two obvious methods by which the information can be carried over the single network are:

(i) Each subscriber's installation responds simultaneously on a different frequency (which might be anything from a.f. to h.f.).

(ii) Each subscriber's installation responds sequentially, over a relatively short time period, on one frequency.

Because (i) would entail a miscellany of different frequencies on the network and attendant interference problems, (ii) is to be preferred, even though the sequential arrangement would appear to call for more complex apparatus. Basic requirements for sequential response signalling are (*a*) the provision of a periodic 'start' sync. pulse signal from the central recording point, (*b*) variable delay devices or mechanisms in the subscriber control boxes which commence their operating cycle on receipt of the start sync. pulses, (*c*) simple response signal generator (oscillator) circuits in the subscriber control boxes which emit single pulses, the timings of which are different for each subscriber and are decided by the delay introduced by the devices from receipt of the common start pulses. Short cycle times are not essential, i.e. the A and B information can be adequately provided on a 2 to 3 minute cycle for all subscribers, thus quite low pulse repetition rates can be used.

The response signals emitted from subscriber boxes can be simple audio frequency bursts of short duration, i.e. 20–40 milliseconds. The actual audio frequency would identify the particular pay-television channel to which a subscriber was tuned. The subscriber identity would be established purely by the arrival time of the individual a.f. pulses in relation to the main start pulse. At the same time these subscriber responses would provide audience measurement information as between the channels. It would be necessary for coaxial (or other) cable networks using the system to be provided with low-pass filters, across the line repeater amplifiers, etc., to allow the subscriber response signals to reach the central recording points.

Referring again to Fig. 2, the clock pulses are used as dual purpose pulses which also act as the synchronizing pulses for response signalling. In the proposed sub-cycle time of 15 seconds, the first 2·5 seconds (max.) would be occupied by the price pulses (coin operation) and the remaining 12·5 seconds for programme identification. At the same time the 'clock' pulses at a repetition rate of 8/second are available for response signalling to ensure further the accuracy in timing of the individual sequential subscriber pulses by triggering simple gate circuits.

In the sub-cycle time of 15 seconds there are approximately 120 clock pulses available for subscriber responses. Since there will often be more than 120 up to, say, 500 subscribers on a line, it is proposed that the sub-cycle is repeated four or five times to form the 'main' cycle. To distinguish the main cycle start pulse the proposal is that it should be modulated at audio frequency F_3, whereas the intermediate start pricing pulse would be modulated at F_1 (see Fig. 4). Thus, using the one composite control signal waveform, it would be possible to use coin prepayment subscriber boxes, simple local credit or central credit recording on one cable network.

The outgoing, narrow band, control signal could either be sent on d.c. or v.l.f. where line attenuation would be negligible, or within the television signal channel as an additional carrier adjacent to the sound carrier.

4. Acknowledgment

This paper is published by permission of the directors of British Home Entertainment Limited.

Certain subject matter has been covered by Patent Applications No. 22,298/62 and 15,988/63.

Manuscript received by the Institution on 26th June 1963. (Paper No. 890/T25.)

U.D.C. 681.32

Low Cost Electronic Arithmetic

By
D. W. THOMASSON
(Associate Member)†

Presented at the Convention on "Electronics and Productivity" in Southampton on 18th April 1963.

Summary: The paper deals with electronic computing devices intended to fill the gap between the more complex mechanical calculators and the simpler forms of general purpose electronic computer. After a discussion of design considerations, the overall concept of a particular solution to this problem is described, with some details of the resulting equipment. Some configurations suitable for known applications are outlined, and the useful limits of the concept are discussed.

1. Introduction

The apparatus needed to carry out an arithmetic calculation may be as simple as pencil and paper or as complex as a large general-purpose computer. The purpose of the present study is to consider the design and use of devices lying roughly midway between these extremes, devices that are more complex and versatile than purely mechanical calculators, but simple enough to avoid problems which arise when a full size computer is used.

Such devices seem to have received comparatively little attention, perhaps because electronic computing and processing methods show maximum economic advantage when they are applied on a large scale using standardized general-purpose equipment adapted for a particular purpose by suitable programming. The simpler, more specialized, equipment is still needed.

The design of devices of this type can be approached in two ways, by refinement and improvement of mechanical devices or by simplification of ordinary electronic calculator techniques. In either case entirely new techniques and concepts arise as development progresses, but the two approaches never quite meet. The ultimate design must either be a mechanical device with electrical or electronic appendages, or an electronic device with mechanical appendages. Attempts to achieve a full symbiosis between electronics and mechanical devices in this field seem fated to failure.

The basically mechanical approach can be illustrated by a device in which ordinary mechanical calculators are adapted by replacing the keyboard by a bank of solenoids and putting read-out switches in place of the printing mechanism. The input is produced by a similarly adapted typewriter which also prints out the results. The input, calculation, and output processes are controlled by a further switching system operated by movement of the typewriter carriage.

A comparable illustration of the electronic approach would be to use electronic logic to perform the calculation and control the action of the system. The typewriter remains as a reminder that electronic devices must usually rely on mechanical input and output equipment, but it is outside the calculating system.

Nevertheless, the typewriter will limit speed and make enough noise to render the silence of the electronic calculator of little value, and it might seem that there is no point in using electronics in such a case as this. If the calculations are extremely simple, the mechanical device will probably be quite adequate, but for more complex requirements the electronic approach can show considerable advantages.

Taking a topical example, there is much concern over the impending plan to decimalize currency. An electronic device can convert pounds, shillings, and pence to decimal form as the figures are typed, and re-convert the result of a calculation to sterling without delaying print-out perceptibly. At the touch of a switch, moreover, it can handle decimal currency with equal ease, this facility having little or no impact on manufacturing cost.

This example shows that fast calculation may be useful, even with a typewriter-controlled system, and it also illustrates the basic flexibility of the electronic method. These assets show to even greater advantage where part of the input data is read automatically from punched tape or card, especially if this method is taken to the point where the typewriter is not connected directly to the calculating system, and merely used for tape preparation and print-out from tape.

A potential user may well question the cost of the advantages conferred by electronics. A completely valid comparison with mechanical calculators is rarely possible, since performance is rarely comparable, but it is only to be expected that electronic devices will often show a higher cost. The increase can be

† Ultra Electronics Ltd., Western Avenue, London, W.3.

Fig. 1. The Contotronic computer.

kept within reasonable bounds by careful consideration of the essential requirements, and in the more complex configurations the electronic equipment may well be cheaper than the mechanical equivalent.

Initial cost, however, is not the only economic factor. Electronic devices do not suffer the effects of constant mechanical wear, and therefore require less maintenance. A greater degree of automatic operation can be achieved by electronic systems, thus reducing the amount of skill needed to work them. Since the operator's wages often represent a high proportion of total running costs, this may be of appreciable significance.

In general, it can be expected that electronic methods will offer advantages where:

(*a*) The calculations are fairly complex (for example, sterling invoices).

(*b*) The volume of work is sufficient to ensure that increased operating speed is put to good use.

(*c*) The nature of the work is such that a fair degree of automatic operation is possible.

The rest of this paper describes a particular series of devices aimed at making the most of the potential advantages of the electronic approach.

2. Logical Design

2.1. *Limits of Study*

The work to be described began as a general study of electronic calculating methods. The aim was to determine whether such methods could be used with advantage in place of electro-mechanical calculation. An arbitrary upper limit of selling price was set at £5,000, this being inclusive of input and output facilities.

Initially, there was no special bias towards commercial, technical, or scientific applications, and problems from each of these fields were considered in assessing various configurations.

Apart from such particular studies, the investigation was conducted on a broad basis, an important objective being to determine which methods offered the lowest cost compatible with sufficient reliability, flexibility and speed for serious use in a wide range of applications.

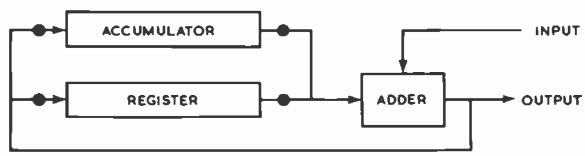
2.2. *Basic Approach*

It became apparent from the outset that an early decision must be taken regarding the basic type of equipment to be developed. The choice lay, broadly speaking, between:

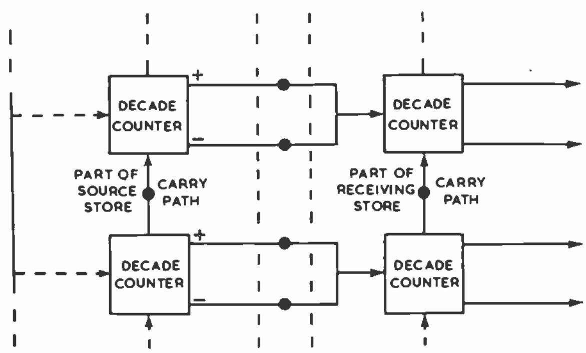
(*a*) The orthodox computer approach, with simple storage and centralized arithmetic logic.

(*b*) Combined storage and arithmetic logic.

Examples of these alternatives are outlined in Fig. 2. The orthodox system in Fig. 2(*a*) needs little comment. Numbers may be stored in any location, but must be brought into the accumulator or register for addition or subtraction, these being used alternatively in multiplication and division carried out on an iterative basis.



(*a*) Orthodox computer system.



(*b*) Combined arithmetic and storage logic.

Fig. 2. Alternative types of calculating system.

The system shown in Fig. 2(*b*) operates on a basis of addition by counting. Each storage decade can signal its contents as N by producing a train of N pulses. These are added to the count in the corresponding decade of the store to which the number is to be added. Subtraction can be obtained by using complementary pulse trains, and 'carries' are handled as for normal decade counters. This is sometimes known as the 'abacus' or 'cross counting' method.

This type of system has been used in small calculating devices, particularly in special-purpose devices, where the number of transfer routes is small. Three types of decade store have been used:

(*a*) 10-state stores in which the state is indicated by position (e.g. dekatrons).

(*b*) 10-state stores in which the state is indicated by voltage level (e.g. step counters).

(*c*) 10-state counters, constructed from bistable circuits.

Of these, only the last can be considered as satisfactory in the present context. The 'position' devices (*a*) are either too slow or too expensive, and the 'level' devices (*b*) are not considered reliable enough in extended service.

The same bistable circuits, however, can be arranged to form shift registers storing binary code and working with centralized arithmetic, and in devices on the scale envisaged here the cost of this approach is less because the transfer gating is much simpler. On balance, it appeared that the centralized arithmetic technique offered the best choice.

2.3. *Number Representation*

In a large computer, the cost of storage is of major importance, and pure binary number representation offers maximum economy. The resulting need for conversion between binary and decimal representation at input and output is an insignificant addition to the overall program.

In the relatively small device under consideration, on the other hand, the cost of storage is less significant, whereas binary/decimal conversion may represent a considerable extension of the calculating facilities. Pure binary notation is not such an obvious choice in this case.

By arranging the storage bistable circuits to form 4-channel shift registers, a binary-coded decimal notation could be used. This requires 20% more bit-storage locations than a corresponding binary system, but the shift-pulse rate can be appreciably lower for a given word transfer time, allowing the use of cheaper bistable circuits.

This arrangement was adopted; it allowed a simple input/output process, requiring only the reversal of decimal-digit order, and permitted a simple means of reading out store contents directly when checking operation of the equipment.

2.4. *Addition*

Adoption of 4-channel binary-coded-decimal working entailed the use of a decimal adder. Two types were considered, one working on the basis of cross-counting and the other on the tabular basis, in which a logic analogue of the addition table is used.

The latter type had the advantages of being more amenable to circuit standardization (Fig. 3). It consists basically of four full binary adders, to which are fed the two binary quartets to be added and any 'carry' from the previous pair of decimal digits.

The output of this arrangement is correct if there is no output 'carry', but where the sum exceeds 9 the result will be in pure binary, not in binary-coded decimal. For example, $7+8 = 15$ will produce 1111, whereas what is required is 0101 with a decimal 'carry'.

The full adders are therefore followed by further adder elements which are arranged to add 6 to the original result if it exceeds 9. Thus, in the case quoted, $15+6 = 21 = 16+5$. The most-significant-bit, representing 16, spills over to provide the output carry, leaving the result as 5.

The complete analogue, including all adder and detector elements, is built up from 20 identical gate elements, each of the (ab+cd) type. (The output is reversed if a and b or c and d are in the '1' state.) Identical elements built in sets of four, are also used to control signal routes and for such purposes as generating complements for subtraction. The complete adder complex, including the adder proper, complementer, sign logic and routing gates, uses 68 of these gates (Fig. 4).

Subtraction is performed by complementing each decimal digit with respect to 9 and adding the result. To save the need for 'carry round' from the most significant digit, a false carry is generated to add unity to the least significant digits. The result is that the system works effectively in complements with respect to 10^N, where the word length is N decimal digits. The 'carry' from the most significant digit is used only to control condition programme branching and sign logic.

The basic arithmetic consists of two storage locations, the accumulator and register, which can be used separately or coupled in series for double-length working and other purposes. Either store can be fed into the adder, with data from any other storage location, the sum being returned to the original source store.

2.5. *Storage*

The accumulator and register are identical four-channel ten-stage bistable shift registers. The same configuration, with one extra stage, is used for the input and output stores, which are required for transfer speed adjustment and number reversal. Units identical to the accumulator can be used to provide up to two backing stores in small configurations.

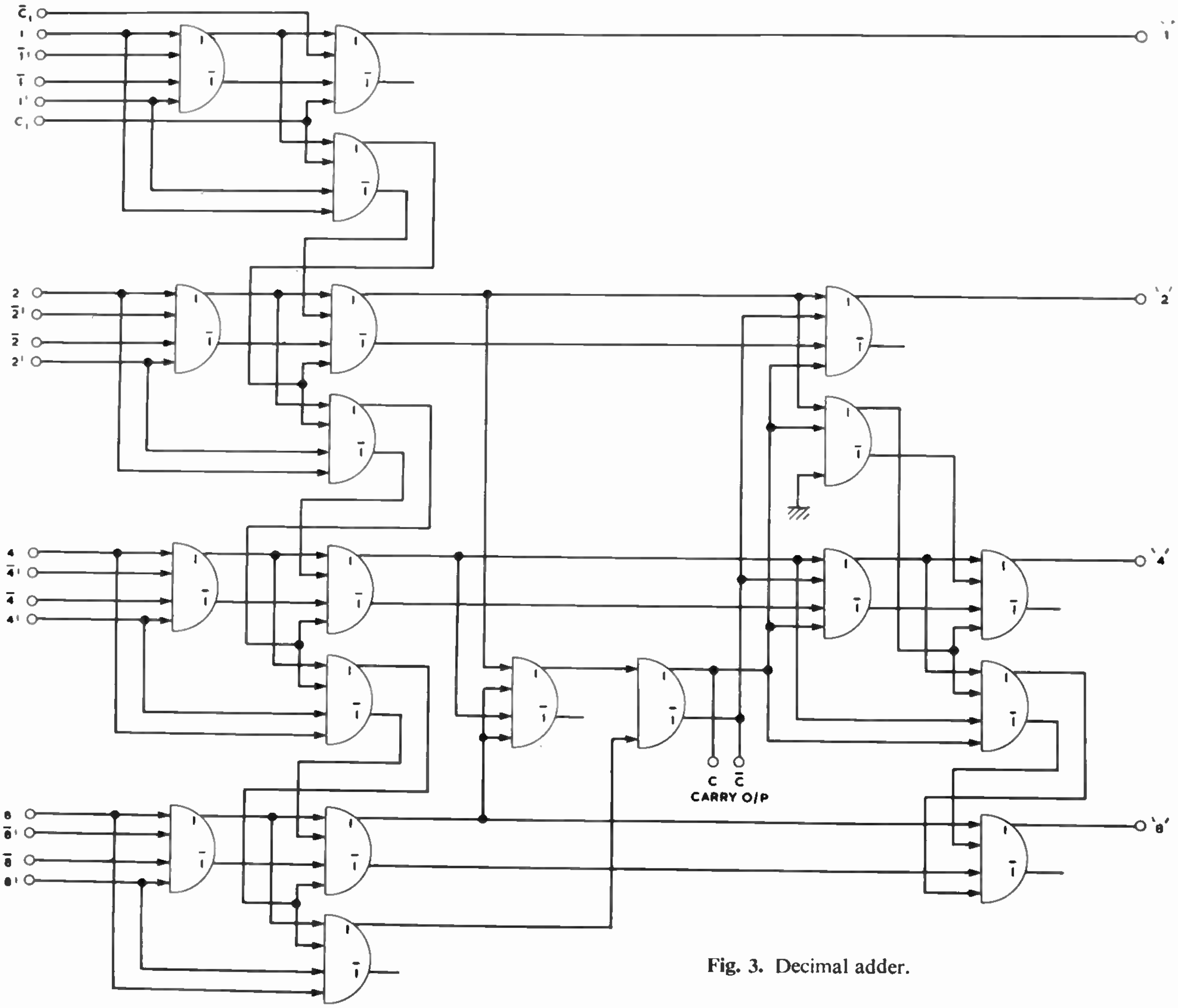


Fig. 3. Decimal adder.

Where more than two backing stores are needed, the shift register, based on bistable circuits, becomes uneconomic and, to provide further storage, a 16-address core store was created. The number of addresses was chosen arbitrarily to meet experimental needs, but has since proved suitable for production requirements.

The core-store unit includes a standard shift-register store into which the contents of one core-store address can be transferred by a single pulse, after which transfer occurs to another location normally. The original contents can be read back into the core store to obtain non-destructive readout.

The shift-register stores, complete with access and circulation gates, show a manufacturing cost of about 35*s*. per bit, while the core store reduces this figure to about 7*s*. 6*d*. (These are initial manufacturing costs, based on limited production.)

2.6. *Program Control*

The basic needs in respect of programming were:

(1) The program must be simple to set up.

(2) It must be economic.

(3) It must be possible to modify programs by external signals.

The normal technique of stored programming, using the same stores as those holding numeric words, was considered and rejected. Instead, a programming system was built up on the basis of a matrix-type plugboard. This offered a form of fixed storage at an economic price, and matched conveniently with sub-routine groups constructed on the same basis but with only the essential links wired in.

Each store location was numbered, as follows:

00(0000) Unity generator/Drain.
01(0001) Input 1/Output 1.
02(0010) Input 2/Output 2.
03(0011) Accumulator.
04(0100) Register.
05(0101)–15(1111) } Core stores for normal access.
16(——)–20(——) } Core stores only accessible via sub-routines.

The instruction word begins with a specification of the sending address, followed by the receiving address, each being represented by a four-bit binary group. A third similar group follows to specify the operation required.

0000	Add	0100	Shift × 10
0001	Add, clear SA	0101	Shift ÷ 10
0010	Subtract	0110	Shift × 10^N
0011	Subtract, clear SA	0111	Shift ÷ 10^N

1000–1111 Sub-routines, including multiply and divide.

The instruction word is completed by a branching signal identification (if required) and a branching jump instruction stating the next row to be selected if the identified branching signal is on.

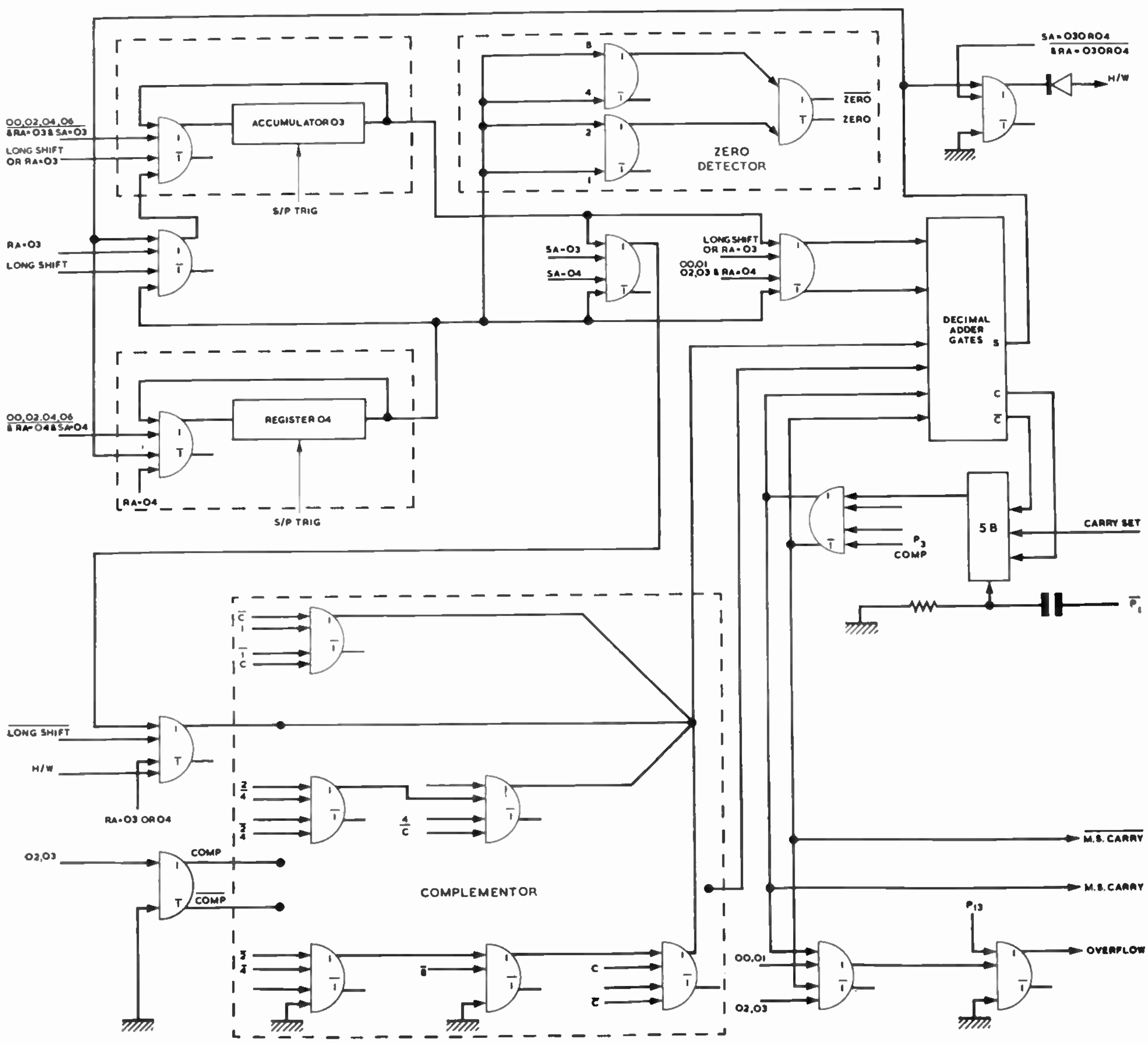


Fig. 4. Complete adder complex.

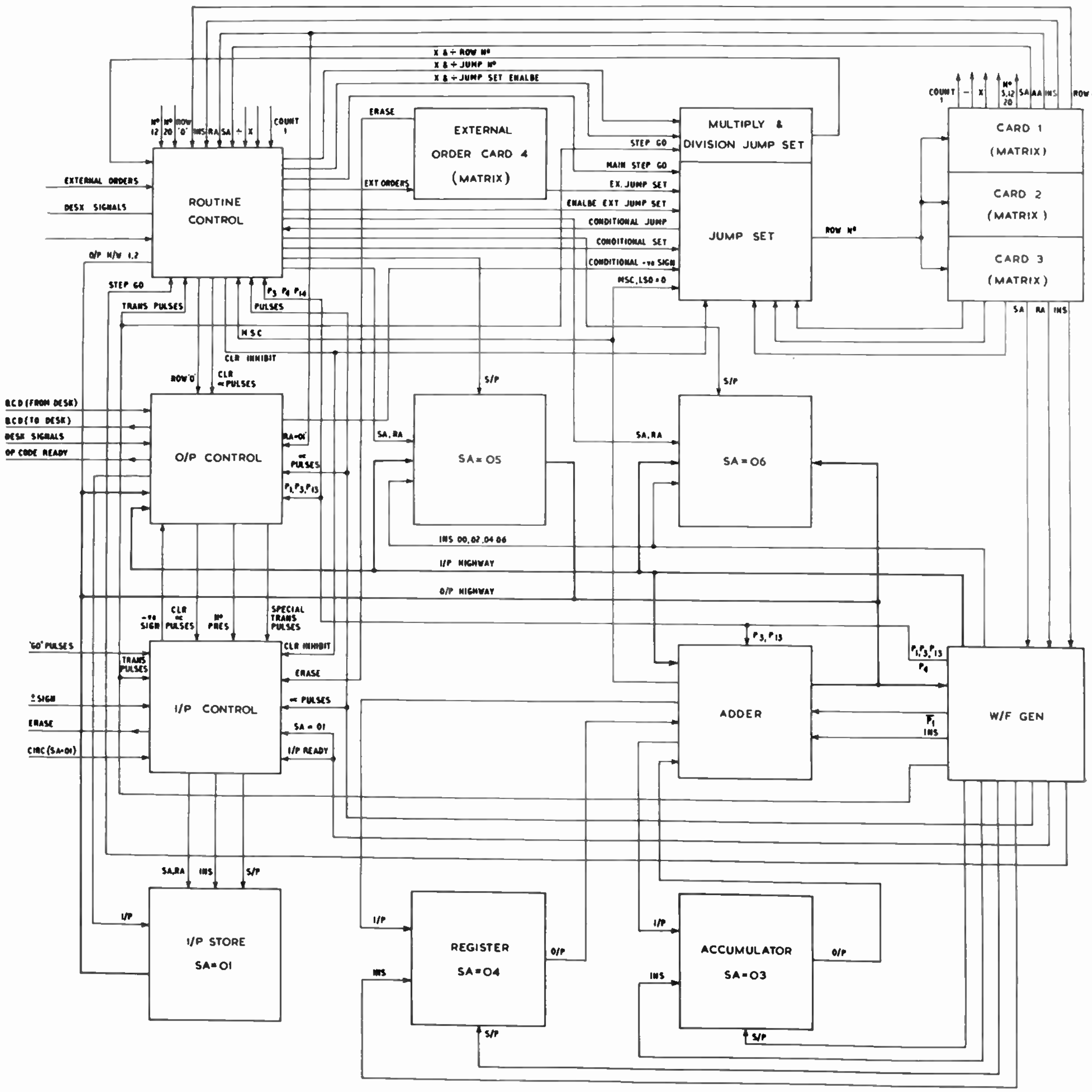


Fig. 5. Block diagram of complete system.

The basic system used in the experimental unit provided 32 program rows, with up to 8 backing sub-routines, which could be varied to suit the application. Later models either eliminated the plugboard entirely for a fixed purpose machine, or introduced a third programming 'level', to facilitate use of 'third level' sub-routines by 'second level' sub-routines. The basic principle remained the same, however. It has proved sufficiently flexible, and although a few 'special' instructions have been found useful these are not essential.

2.7. *Overall Concept*

The complete scheme may be illustrated by one of the possible simpler configurations (Fig. 5). This has two backing stores, and is intended for invoicing and similar calculations.

Incoming data are fed to the input store, where the order of the digits is reversed. The store may be cleared immediately prior to the arrival of a number for use in calculation; the input may, therefore, be connected permanently to an information highway.

The first digit fed on after clearance is shifted to the least-significant (output) end of the store. The second digit is shifted to the same position, the first moving one place towards the most-significant end of the store. After the whole number (which may be of any length up to 10 digits) has been fed in, the end of the number and its sign are signalled. The number may now be read into the calculating system proper, the least-significant digit leading, as is necessary for the calculation processes.

Meanwhile, a signal from the information source has stated the operation to be performed, setting the program control matrix to an appropriate row. As soon as completion of the input process allows the program to proceed, the instructions set up for the selected row are carried out. When the arithmetic operations are complete, the program will move to a new row called up by the last instruction. When the sequence of instructions is complete, the program returns to row 0, which is a 'waiting row', giving no instructions.

When a result is available, it is transferred to the input store one digit at a time, reversal of digit order occurring as before, then read out at a speed appropriate to the associated equipment.

The scheme is clearly open to many variations. In this case a common input/output reversing store is adequate, because the unit is normally tied to a typewriter, and input and output never occur simultaneously. When working with tape, separate input and output channels would allow faster working. With another 14 backing stores no basic difference in the system is needed although the program needs to be more complex.

Most of the circuit elements required were either bistable or gates of the (ab + cd) form already discussed. The design study had shown the need for system clock speeds in the 10–40 kc/s region. This called for a maximum response time in the individual circuit elements of about 5 μs, since several elements must act sequentially in the intervals between clock pulses.

To meet this need simple transistor/diode logic was chosen. No attempt was made to carry circuit standardization to the ultimate limit, and the bistable circuits and gates were designed independently for minimum cost compatible with required performance. The success attained may be judged from production cost: each bistable, complete with steering gates, costs less than £1, and each gate complex, with two inverters, costs approximately 25*s*. (These figures, again, are based on limited production.)

These two circuit elements make up some 70% of the equipment. The remaining circuit groups include matrices for encoding and decoding, special amplifiers, and pulse generators.

In addition to the calculating devices, the circuit elements also serve for a variety of other types of equipment. The typewriter/tape desks used in 'on-line' configurations require some specialized elements for solenoid drive, but are otherwise built from standard circuits, which are also used in the construction of test gear.

The combination of printed-circuit boards into convenient sub-units was based, as far as possible, on a logical grouping of the various parts of the system. The most widely used group is the shift-register store, complete with access and circulating gates and pulse generators. The core store is built in two similar sub-units.

Wherever possible, the sub-units were made suitable for more than one system configuration. Thus it was possible to establish quite rapidly a number of configurations of varying price, complexity and capability, so allowing a wide coverage of the intended field of application.

The configurations established have either two or sixteen backing stores, are arranged to work 'on-line' with a typewriter, or with readers and punches for 'off-line' working, and have either preset or changeable programming. A word length of 10 decimal digits is normal, but this could be changed without difficulty.

3. Equipment

It was clear from the start that standardization of circuitry must be sought in some degree. A standard rack system was already in existence, based on printed circuit boards measuring approximately 10 in × 5 in (Fig. 6). These fitted into a connector across the narrow dimension, and were retained in nylon guides held in metal framing. The framing itself was designed to fit 19 in racks or be built into one of two sizes of free-standing case.

Examination of the first logic diagrams showed that this system alone would not suffice, so sub-units were devised to hold smaller printed-circuit boards measuring 4 in × 4·7 in (Fig. 7). The sub-units used the original nylon guides, and were connected up by multi-pin connectors. The circuit boards were set parallel to the front plane of the rack and were connected up entirely by soldered cables.

The use of edge connectors for each board was avoided, since experience had shown that a high proportion of the faults which would require servicing attention could be traced to the use of edge connectors on the larger boards.

Eliminating these connectors makes access for servicing a little more difficult (although it reduces the need for such access) while giving improved reliability. It also cuts the cost by some 10%.

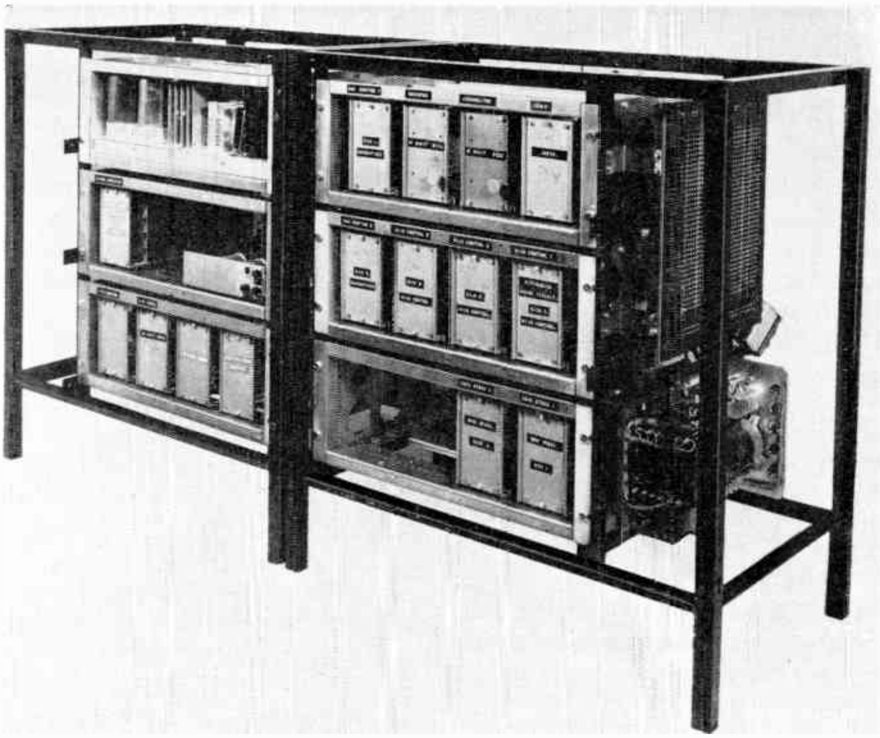


Fig. 6. Experimental unit.

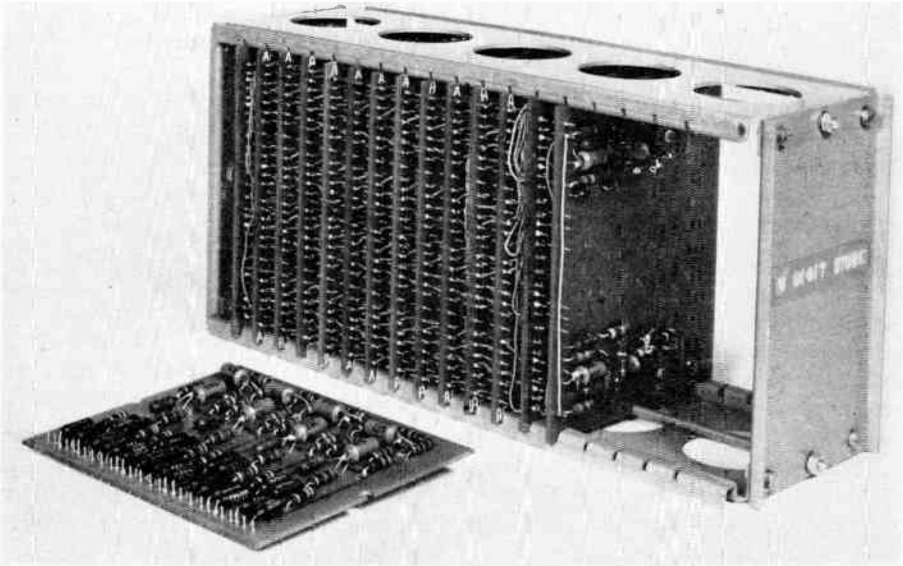


Fig. 7. Sub-unit rack with printed circuit board.

4. Applications

There are so many possible applications of this kind of device that an exhaustive review would be too extensive for inclusion here. Some particular cases will be examined, however, to indicate, in very approximate terms, the circumstances in which such devices can be used with advantage.

4.1. *Invoicing*

The configuration already taken as an example of a complete system was designed for invoice calculations. For very simple invoices in decimal currency electro-mechanical methods are quite adequate, although simplified forms of electronic multiplier are being used to reduce the calculation time. Where the invoice is in sterling, or where complex quantities are used, electronics begin to show an advantage, especially in simplicity of operation.

Two invoicing configurations have been set up. The simpler form, with two backing stores, covers sterling and decimal invoices having only one running total, with facilities for calculating percentages up and down on either the gross or the net total. The calculating unit is associated with a typewriter desk having facilities for tape or card input and output, allowing much of the necessary data to be fed in automatically, while a tape record can be created for subsequent checking or other processing. In the extreme, the operator only needs to hand-set quantities, the rest of the operation being controlled automatically.

The other invoicing configuration has 16 backing stores, and deals with three running totals, purchase tax and other 'line percentage' calculations, group totals and discounts, complex quantities, such as areas or volumes in feet and inches, and also provides day totals.

A most interesting point is that the more complex calculator only costs approximately 25% more than the simpler system, the considerable extension of the facilities requiring the use of only a 16-address core store (in place of the original two shift-register stores) and an extended programme. It does require more operator training, however, due to its greater versatility, and the simpler machine may therefore be preferable in some cases.

Either of these machines can be readily re-programmed for other types of work, and a simple independent keyboard and visual readout can be fitted to allow them to be used as separate calculating devices if the occasion should arise.

4.2. *Other Commercial Work*

The configurations described above are equally suitable for other commercial work such as dividend calculations, preparation of statements, and in many instances no alteration to the internal program needs to be made.

Payroll and personal income tax problems have been considered in relation to the same devices, but it was found that these could be handled more effectively by a version with more flexible programming facilities and tape input and output. This equipment, using medium speed tape reader and punch, can deal with 1000 feet of tape in 20–30 minutes. In income tax problems, the tax code number is replaced by 'annual free pay', the rest of the calculation presenting no difficulty.

4.3. *Experimental Test Results*

In such work as aero-engine testing, particularly in the development stage, large masses of data have to be processed before the results of the test can be determined. If, as in one case, test results have to be passed to a computer centre for analysis, the resulting delays may be serious.

Much work of this type can be handled by the 'on-line' typewriter configurations, especially where the bulk of the intermediate results are required in printed form. It may, however, be better to use the 'off-line' tape version, with an independent print-out unit for recording final results.

4.4. *Sorting and Merging*

There are many potential advantages in the use of paper tape as a means of data storage, especially where block lengths vary considerably, providing that the following operations can be carried out:

(*a*) *Sorting* the items into a pre-determined order.

(*b*) *Checking* that the sequence is correct.

(*c*) *Selecting* a given item or class of items.

(*d*) *Merging* two streams of data into sorted order.

The essential common factor in all these operations is the comparison of pairs of block references, which may be alpha-numeric. The 'off-line' configuration can handle numeric references without difficulty, and can be readily adapted to compare 8-bit code groups by programming the stores to work in pairs, the input and output channels being duplicated.

4.5. *Unsuitable Applications*

The equipment described has insufficient storage to invert large matrices or store tables of constants, nor would any appreciable increase in storage be justifiable with the present programming arrangements. At the other end of the scale, it cannot compete with a simple add-list machine on its own ground.

It is difficult to define the limits of utility more precisely in general terms. A 'second-generation' device, particularly appropriate to the 'on-line' configurations, is already under review, and this and similar future developments will extend the range of applications appreciably.

An ultimate objective might well be a fully electronic desk calculator which can compete in terms of price-for-performance with the more sophisticated mechanical equivalents, while retaining the flexibility, reliability, and checked accuracy possible with orthodox computer methods. This objective is not yet within reach, but neither is it entirely out of sight.

5. Conclusions

The fundamental decision to use centralized arithmetic places a limit on the degree of simplification which can be justified economically. A simple one-store adder system still requires an adder unit and input/output facilities, and the optimum approach to this type of requirement would be quite different.

A system based on distributed arithmetic would compare more closely with mechanical-calculator methods, and if the initial object had been a frontal attack on the mechanical-calculator market this system might have been a better choice.

Such a frontal attack, however, was not envisaged. An essential feature of the project was that it should be possible to provide a fair number of holding stores, with corresponding programming complexity. This caters for a field in which there is little to attack, except for a few specialized electro-mechanical hybrid systems.

In relation to the small full-scale stored program computers, the cost of the devices described herein is comparatively small. The saving has been achieved by drastic reduction in storage, by simplified program methods, and by accepting a lower working speed. These can be off-set to a considerable extent by proper work organization; the equipment has capabilities which are well beyond the limits which might be expected on casual consideration.

It is doubtful whether these capabilities have been fully charted, since there are many possible fields of application still to be studied. The boundaries already known enclose a large enough area to indicate that the low-price electronic calculator has a definite job to do, and can offer economic advantages in many applications.

6. Acknowledgments

The author thanks the Directors of Ultra Electronics for permission to publish this paper.

Manuscript first received by the Institution on 28th January 1963, in revised form on 7th March 1963 and in final form on 27th November 1963. (Paper No. 891/C69.)

U.D.C. 621.395.4: 621.315

The Telephone Channel Capacity of Transmission Lines

By

D. A. GRAY, B.E.E.†

Presented at the I.R.E. Australia Convention in Sydney in March 1961, and reprinted from The Proceedings of the Institution of Radio Engineers Australia.

Summary: The attenuation of transmission lines used for telephone transmission varies very nearly as the square root of frequency when dielectric loss is small and conductor resistance is controlled by skin effect. In this paper the maximum number of telephone channels which can be obtained on such a line using a.m.-f.d.m., f.m.-f.d.m., and p.c.m.-t.d.m. is calculated as a function of line attenuation at a reference frequency, transmitted power level, and signal/noise ratio. At moderate and high transmitting levels s.s.b.-a.m. usually permits the greatest number of channels to be derived.

1. Introduction

Long-distance telephone transmission systems operating over land lines and sea cables have, for about 40 years, made use of single-sideband suppressed carrier (s.s.b.) modulation, since this requires a channel bandwidth very little greater than the bandwidth of the telephone signal itself. The bandwidth utilized by each channel of such a carrier system, when operating over long lines is kept to a minimum because of the rapid increase of line attenuation with increasing frequency. (We exclude from consideration here those waveguide systems in which the reverse is true.)

In considering the application of some modern modulation techniques to line transmission one recollects the ability of some forms of modulation, such as f.m. for example, to operate under conditions of low carrier/noise ratios. This ability is always obtained at the expense of an increased bandwidth requirement for the modulated signal. Here are two conflicting factors; with signals such as f.m. where bandwidth may be exchanged with signal/noise ratio, it is obvious that there will be an optimum arrangement which will give a maximum number of channels in a line transmission system. Since the attenuation-frequency characteristics of transmission lines can often be represented in a mathematically simple form, the problem can be dealt with mathematically. This has been done for f.m. in this paper; and comparisons are drawn with the channel capacity obtainable with s.s.b., d.s.b. and pulse-code modulation (p.c.m.).

† Research Laboratories, Australian Post Office, Melbourne, Victoria.

The calculations are made on the assumption that the line noise spectrum intensity and the available transmitter power are the same at all frequencies.

In a long system, repeaters must be inserted at intervals to overcome the line loss; if these repeaters are identical then their maximum amplification must be equal to the maximum line loss of one repeater section. Therefore a system using wide bandwidths must either use high-gain repeaters, or repeaters must be inserted at frequent intervals. (Cross-talk considerations tend to limit the maximum gain which can be used in a repeater.) In long systems where repeater cost is an important part of total system cost, this consideration leads to the use of the minimum-bandwidth system, i.e. s.s.b. But in short systems repeater cost may be sufficiently unimportant so that wide-band systems can be seriously considered.

2. Propagation Constant of Transmission Lines

The propagation constant of a transmission line can be expressed in terms of its 'primary constants' per unit length, namely the series inductance (L), the shunt capacitance (C), the shunt conductance (G) and the series resistance (R). At frequencies for which $2\pi fL \gg R$ and $2\pi fC \gg G$ (this applies over a very large part of the frequency spectrum of the systems considered in this paper), the propagation constant is given sufficiently well by the well-known equation

$$P = \tfrac{1}{2}(R/Z_0 + GZ_0) + j2\pi f\sqrt{(LC)} \qquad \ldots\ldots(1)$$

where $Z_0 = \sqrt{(L/C)}$, the characteristic impedance of the transmission line. If, further, the dielectric losses of the line, expressed by means of the primary constant G, are negligible as is frequently the case, then the real part of the propagation constant is directly

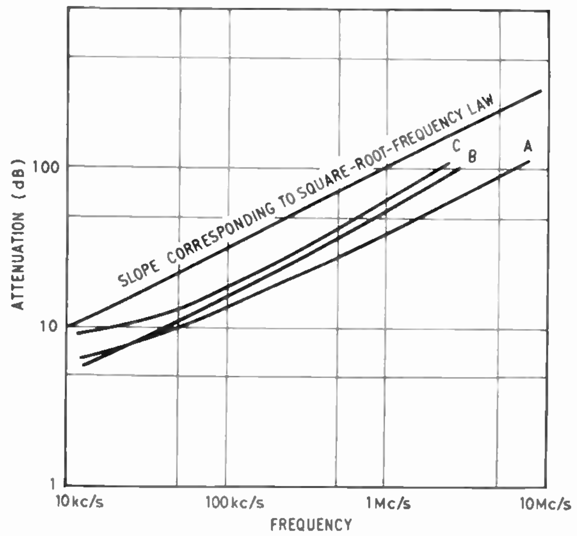


Fig. 1. Attenuation of some typical transmission lines.
Curve (A) 10 miles of trunk coaxial cable (⅜ in polythene disk insulated).
Curve (B) 50 miles of trunk aerial line (200 lb hard-drawn copper).
Curve (C) 2 miles of local trunk cable (10 lb paper insulated quad local).

proportional to R, the conductor resistance. The variation of R with frequency is controlled primarily by skin effect. If the skin depth is small, as at high frequencies, then R is proportional to $\sqrt{F}$ and hence the attenuation constant $R/2Z_0$ is proportional to the square root of frequency. This fact enables us to calculate the width of the usable spectrum and hence the telephone channel capacity of such a transmission line.

Figure 1 shows the attenuation-frequency relationships for some typical transmission lines. In this paper the attenuation-frequency relationship is expressed by

$$\alpha = \alpha_1\sqrt{F} \qquad \ldots\ldots(2)$$

where F is the frequency in Mc/s and α_1 is the attenuation in decibels at 1 Mc/s.

3. Noise Level at the End of a Repeater Section

In this paper, it is assumed that the noise arising in the transmission line has the characteristics of thermal agitation noise. At the end of a repeater section the transmission line will be terminated by a receiving amplifier and equalizing networks. The repeater section, though having a finite transmission delay, will produce no attenuation and no phase distortion from a low frequency (assumed here to be indistinguishable from zero frequency) up to the maximum frequency F_t required by the transmission system. At the output of the repeater section the noise power per unit bandwidth will be proportional to $\exp(0{\cdot}2303\alpha_1\sqrt{F})$. (See Fig. 2.)

Denoting the frequency spectrum bandwidth of a single telephone channel (usually 4 kc/s) as F_v Mc/s and L_i as the line noise level in decibels in this bandwidth, the total noise power at the end of the repeater section can be obtained by integration of the above-mentioned exponential relationship.

Thus the noise power in an elemental bandwidth $\mathrm{d}F$ at the receiving amplifier input is $\dfrac{\mathrm{d}F}{F_v}\cdot 10^{L_i/10}$ units of reference power (e.g. milliwatts). At the repeater section output the total noise power P_t is

$$P_t = \int_{F_1}^{F_2} 10^{L_i/10}\,.\,10^{\alpha_1\sqrt{F}/10}.\,\frac{\mathrm{d}F}{F_v} \qquad \ldots\ldots(3)$$

$$= 10^{L_i/10}/F_v \int_{F_1}^{F_2} 10^{\alpha_1\sqrt{F}/10}\mathrm{d}F \qquad \ldots\ldots(4)$$

This may readily be evaluated by a change of variable. Putting $10^{L_i/10} = \exp(0{\cdot}2303L_i)$ we have

$$P_t = \frac{2\exp(0{\cdot}2303L_i)}{(0{\cdot}2303)^2\alpha_1^2F_v}\times$$
$$\times\left[(0{\cdot}2303\alpha_1\sqrt{F}-1)\exp(0{\cdot}2303\alpha_1\sqrt{F})\right]_{F_1}^{F_2} \qquad \ldots\ldots(5)$$

For the frequency range 0 to F_t Mc/s, the noise level

$$L_t = 10\lg P_t$$
$$= L_i + 15{\cdot}76 +$$
$$+10\lg\frac{(0{\cdot}2303\alpha_1\sqrt{F_t}-1)(\exp 0{\cdot}2303\alpha_1\sqrt{F_t})+1}{F_v\alpha_1^2} \qquad \ldots\ldots(6)$$

where '$\lg x$' means '$\log_{10} x$'.

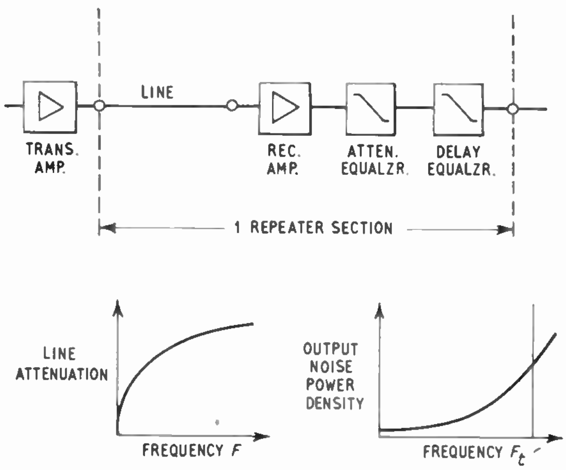


Fig. 2. Conditions in a line repeater section.

For values of line loss ($\alpha_1\sqrt{F_t}$) greater than 10 dB it is sufficiently accurate to omit the +1 term, giving

$$L_t = L_i + \alpha_1\sqrt{F_t} + 15{\cdot}76 + 10\lg\frac{0{\cdot}2303\alpha_1\sqrt{F_t}-1}{F_v\alpha_1^2} \quad \ldots(7)$$

Where, in the following, the noise level in a narrow channel of width $(F_2 - F_1)$ centred on the carrier frequency F_c is required, the following approximation is used

$$L_t = L_i + \alpha_1\sqrt{F_c} + 10\lg[(F_2-F_1)/F_v] \quad \ldots\ldots(8)$$

4. Circuits Provided by Single-sideband Suppressed Carrier

For circuits provided by single-sideband suppressed carrier (s.s.b.) systems a channel of bandwidth equal to F_v is required in the line frequency spectrum for each telephone circuit provided. Let n_a be the total number of s.s.b. circuits provided and L_s be the transmitted level in decibels; then, at the centre frequency of the highest frequency channel, the level at the input to the receiving amplifier will be

$$L_r = L_s - \alpha_1\sqrt{[F_v(n_a - \tfrac{1}{2})]}$$

If L_i is the line noise level in the band occupied by this channel then the signal/noise level separation

$$S_r = L_r - L_i = (L_s - L_i) - \alpha_1\sqrt{[F_v(n_a-\tfrac{1}{2})]}$$

$(L_s - L_i)$ is the signal/noise level difference at the sending end and we write this as S_s. Thus

$$S_r = S_s - \alpha_1\sqrt{[F_v(n_a-\tfrac{1}{2})]}$$

and hence

$$n_a = \frac{(S_s - S_r)^2}{\alpha_1^2 F_v} + \tfrac{1}{2} \quad \ldots\ldots(9)$$

In order that useful comparisons may be drawn between s.s.b. and frequency modulation (f.m.), L_s is to be understood as referring to the sending level for circuit overload and not to the usual 'line-up level'. A corresponding interpretation is to be placed on S_r and S_s.

5. Circuits Provided by Frequency Modulation

5.1. *Bandwidth Required per Channel*

We concern ourselves here only with f.m.-f.d.m., i.e. channels multiplexed on a frequency-division basis after frequency modulation.

In radio-telephone practice, the bandwidth requirement in Mc/s is accepted[1] as being

$$B = 2F_v(1+mK) \quad \ldots\ldots(10)$$

where m is the f.m. index and K is a constant, usually 1·0 for commercial telephony.

The practical effect of band-limiting in f.m. channels is to cause non-linear distortion; this matter is well treated by Küpfmüller.[2]

For small values of modulation index used in narrow-band frequency modulation (n.b.f.m.) we may put $K = 0$ and therefore transmit only the first pair of sidebands at the highest modulation frequency. The resulting distortion of a single frequency modulating signal is predominantly third order and is given[2] by

$$k_3 = [J_1^2(m)]/[J_0^2(m) + 2J_1^2(m)] \quad \ldots\ldots(11)$$

In eqn. (11), k_3 is the distortion factor for the third harmonic and J_0 represents the Bessel function of the first kind and order zero. The permissible value of k_3 at full deviation may be inferred from specifications[3] for non-linear distortion in the circuits provided by multi-channel s.s.b. carrier telephone systems. A distortion factor of 0·05 at a signal level 2 dB below the onset of channel limiting is specified. Applying eqn. (11), we find that for n.b.f.m. transmitting only one pair of sidebands, $k_3 = 0{\cdot}05$ when $m = 0{\cdot}45$ and hence the maximum modulation index must be restricted to $0{\cdot}45 \times 10^{2/20}$ or 0·57. For 'white' noise, f.m. at a modulation index of 0·57 gives a signal/noise ratio very nearly equal to that provided at the same carrier level by double-sideband amplitude modulation, which of course requires the same bandwidth as n.b.f.m. The use of n.b.f.m. at modulation index less than 0·57 does not permit any further reduction in bandwidth.

At low received carrier levels, the use of w.b.f.m. (wideband f.m.) is necessary in order to obtain the desired noise improvement available with f.m. For w.b.f.m., using the recommended value of 1·0 for K in eqn. (10), the channel bandwidth is such as to permit at the highest modulation frequency, the transmission of two pairs of sidebands for $m = 1$, and so on. The non-linear distortion which results can be calculated[2] from

$$k_3 = [6/m][J_{n-2}(m).J_{n+1}(m) + J_{n-1}(m).J_{n+2}(m) + J_n(m).J_{n+3}(m)] \quad \ldots(12)$$

where n is the number of pairs of sidebands transmitted. The following results are obtained for some integral values of m.

m	k_3
1·0	0·097
2·0	0·067
3·0	0·050
4·0	0·039

When it is remembered that, for channel bandwidths determined by eqn. (10), the above distortion factors apply to the maximum modulation frequency transmitted at full deviation and that the distortion factor for lower modulation frequencies is less (because of the greater number of sidebands accepted by the channel), the recommended value of K in eqn. (10) represents a reasonable compromise.

5.2. *Channels Available with F.M. at Constant m*

If n_f channels are provided using the whole of the available spectrum from zero frequency and if all channels use the same deviation and bandwidth, then the carrier frequency of the highest channel will be

$$F_{nc} = n_f[2F_v(1+mK)] - F_v(1+mK)$$
$$= (2n_f - 1)(1+mK)F_v \quad \ldots\ldots(13)$$

At this frequency the received carrier level at the input to the receiver amplifier will be

$$L_r = L_s - \alpha_1\sqrt{F_{nc}}$$
$$= L_s - \alpha_1\sqrt{[(2n_f-1)(1+mK)F_v]} \quad \ldots(14)$$

If the line noise level is L_i dB for a bandwidth of F_v then it will be $L_i + 3{\cdot}01$ dB for the bandwidth required for d.s.b.-a.m. The signal/noise level difference for d.s.b.-a.m. is

$$S_{am} = L_r - (L_i + 3{\cdot}01)$$
$$= L_s - \alpha_1\sqrt{[(2n_f-1)(1+mK)F_v]} - (L_i + 3{\cdot}01)$$

and for f.m. of modulation index m, with 'white' noise, the derived circuit noise, expressed in decibels below a tone producing full deviation is

$$S_r = S_{am} + 20\lg m\sqrt{3}$$
$$= L_s - \alpha_1\sqrt{[(2n_f-1)(1+mK)F_v]} - $$
$$- L_i + 20\lg m\sqrt{3/2} \quad \ldots(15)$$

Thus, with $L_s - L_i = S_s$

$$n_f = [S_s - S_r + 20\lg m\sqrt{3/2}]^2/[\alpha_1^2 2F_v(1+mK)] + \tfrac{1}{2} \quad \ldots\ldots(16)$$

where the levels are expressed in decibels, α_1 is the line attenuation at 1 Mc/s and F_v is the derived circuit bandwidth.

If, in eqn. (16), we put $m = 0{\cdot}57$ and $K = 0$ then the resulting expression gives the number of n.b.f.m. channels available

$$n_n = (S_s - S_r - 3{\cdot}14)^2/(\alpha_1^2 . 2F_v) + \tfrac{1}{2} \quad \ldots\ldots(17)$$

There is a value of m which, in eqn. (16), gives a maximum value of n_f. This optimum value, m_0, is given by

$$m_0 K[S_s - S_r + 20\lg m_0\sqrt{3/2} - 17{\cdot}38] - 17{\cdot}38 = 0 \quad \ldots\ldots(18)$$

Note that the optimum value of modulation index is independent of the attenuation α_1 of the transmission line. A plot of $(S_s - S_r)$ against m_0, for $K = 1$ and $K = 2$ is given in Fig. 3.

In Appendix 1 the effect of the f.m. threshold phenomenon on the validity of eqn. (18) is considered.

For practical values of $(S_s - S_r)$, systems using constant deviation f.m. (i.e. the same modulation

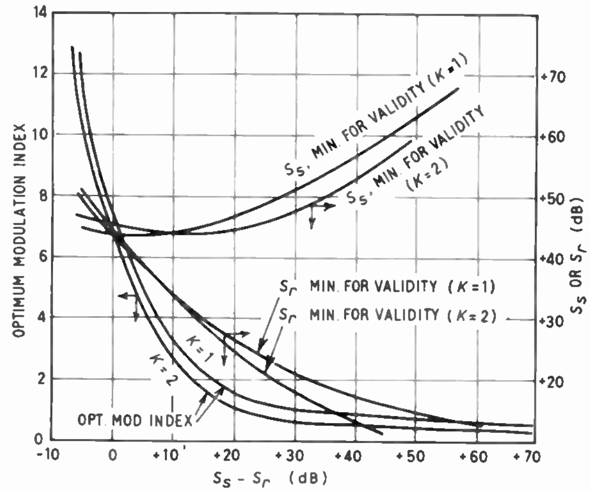


Fig. 3. Optimum modulation index for constant-deviation f.m. channels.

index in all channels irrespective of the position in the frequency spectrum) do not yield as many circuits as s.s.b.-a.m., n.b.f.m. or f.m. with frequency deviation adjusted in each channel to suit conditions applying in the relevant part of the spectrum. This latter case will now be considered.

5.3. *Channels Available with F.M. (Adjusted m)*

In this case we suppose that the modulation index in each channel is so adjusted that the desired signal/noise ratio is just reached in each channel. This requires that the modulation index and channel bandwidth are increased as the channel carrier frequency is increased. There is, however, nothing to be gained, as explained in the discussion of n.b.f.m. from using modulation indices less than 0·57. A transmission line fully loaded with 'adjusted m' channels would therefore, in the lower part of the frequency spectrum, utilize n.b.f.m. channels, the number available being given by eqn. (17). In the upper part of the spectrum, w.b.f.m. would be required.

The modulation index required in any channel to produce a desired value of S_r can be obtained by rearranging eqn. (15), writing for the expression $(2n_f - 1)(1+mK)F_v$ the channel carrier frequency F.

$$S_r = L_s - \alpha_1\sqrt{F} - L_i + 20\lg m\sqrt{3/2}$$
$$= S_s - \alpha_1\sqrt{F} + 20\lg m + 1{\cdot}76$$

Therefore

$$\lg m = (S_r - S_s + \alpha_1\sqrt{F} - 1{\cdot}76)/20 \quad \ldots\ldots(19)$$

Hence

$$m = \exp[(S_r - S_s + \alpha_1\sqrt{F} - 1{\cdot}76)/8{\cdot}686] \quad (20)$$

The channel bandwidth required is obtained by inserting the value of m given by eqn. (20) into eqn. (10). The number of f.m. channels per megacycle is then

$$\frac{1}{B}=\frac{1}{2F_v(1+mK)}$$
$$=\frac{1}{2F_v\{1+K\exp[(S_r-S_s+\alpha_1\sqrt{F}-1\cdot76)/8\cdot686]\}}$$
$$=\frac{1}{2F_v\left\{1+K\exp\left[\frac{S_r-S_s-1\cdot76}{8\cdot686}\right]\exp\left[\frac{\alpha_1\sqrt{F}}{8\cdot686}\right]\right\}} \qquad (21)$$

The total number of channels which can be accommodated between two frequency limits F_1 and F_2 is therefore

$$n_w=\int_{F_1}^{F_2}\frac{\mathrm{d}F}{2F_v\left\{1+K\exp\left[\frac{S_r-S_s-1\cdot76}{8\cdot686}\right]\exp\left[\frac{\alpha_1\sqrt{F}}{8\cdot686}\right]\right\}} \qquad \ldots\ldots(22)$$

This integral is of the form

$$n_w=\frac{1}{2F_v}\int_{F_1}^{F_2}\frac{\mathrm{d}F}{1+M\exp(N\sqrt{F})} \qquad \ldots\ldots(23)$$

where M and N are independent of F. The method of solution is outlined in Appendix 2. The result is

$$n_w=[(F_2-F_1)/2F_v]-$$
$$-[8\cdot686/F_v][\sqrt{F_2}\ln(1+m_2K)-$$
$$-\sqrt{F_1}\ln(1+m_1K)]+$$
$$+[75\cdot43/\alpha_1^2F_v][\mathrm{f}(m_2K)-\mathrm{f}(m_1K)] \quad (24)$$

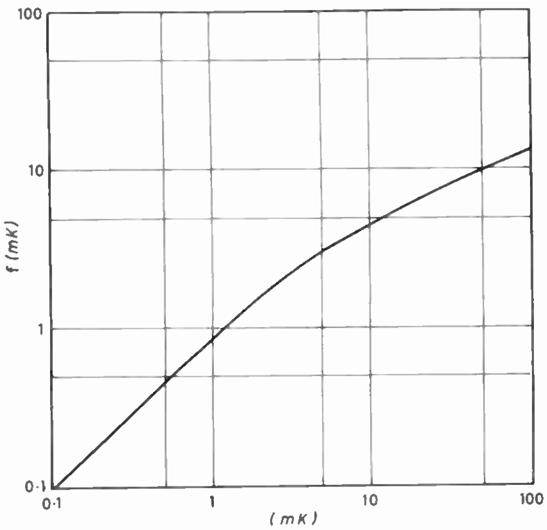


Fig. 4. The function f(mK).

where m_1, m_2 are the modulation indices required at frequencies F_1 and F_2 respectively as given by eqn. (20) and

$$\ln x = \text{natural logarithm of } x$$

The function f(mK) cannot be obtained in explicit form and has been evaluated by series expansions. It is plotted in Fig. 4.

It now remains to determine F_1 and F_2. F_1 is determined by the number of n.b.f.m. channels which can be applied to the transmission line and hence

$$F_1 = 2n_nF_v \qquad \ldots\ldots(25)$$

since only first-order sidebands are transmitted. F_2 is not so specifically determined since it depends upon the evaluation of the signal level necessary to reach the threshold of f.m. noise improvement. This threshold signal level is a function of frequency deviation needed in the channel at full modulation. It can be reduced by feedback arrangements within the receiver, but the effects of such arrangements will not be considered in this paper.

It is known from fundamental considerations of f.m. that when the signal/noise ratio falls to approximately unity, noise predominates in the receiver output irrespective of the amount of frequency modulation of the signal. The noise signal implied by the term 'white noise' has occasional peaks which greatly exceed the mean power. Unless a satisfactory signal/noise ratio is maintained these occasional peaks will suppress the signal in the receiver output and in the signal gaps there will be very strong noise bursts. Following conventional practice, we take a peak factor of 12 dB for 'white noise'. Such peaks will not be exceeded, in either the positive or negative direction, for more than 0·006% of the time.

The frequency F_2 is therefore the upper limit of the channel for which the received carrier level exceeds the noise level in the band B by 12 dB

$$B = 2F_v(1+mK)$$

where the m value to be used is determined by conditions at the carrier frequency of that channel. Moderate inaccuracies in the value of F_2 will produce only small errors in the total number of channels obtainable because the channels will be spread very thinly at the upper end of the frequency spectrum.

If the noise level in bandwidth F_v is L_i dB, then the noise level in the wide-band channel will be

$$L_w = L_i + 10\lg 2(1+mK) \;\; \text{dB}$$

and since in the last channel $mK \gg 1$ (in all practical cases) then

$$L_w = L_n + 10\lg 2mK \qquad \ldots\ldots(26)$$

Substituting for lg m from eqn. (19) we find that the carrier to noise level difference is

$$L_r - L_w = L_s - \alpha_1\sqrt{F} - L_i - 10\lg 2K - \\ -\tfrac{1}{2}(\alpha_1\sqrt{F} + S_r - S_s - 1{\cdot}76)$$

The frequency for which this level difference is just 12 dB is found to be given by

$$\sqrt{F} = (3S_s - S_r - 28{\cdot}23 - 20\lg K)/3\alpha_1 \quad \ldots(27)$$

This determines the carrier frequency of the highest frequency channel, and the modulation index and channel bandwidth can be obtained by the application of eqns. (20) and (10) and thus F_2 may be fixed and m_2 determined. The necessary data are then available for the calculation of the number of wide-band f.m. channels by use of eqn. (24).

6. Pulse-Code Modulation

For 7-digit p.c.m., probably the most suitable p.c.m. arrangement for commercial telephone transmission, a signal/noise level difference of 14 dB is sufficient[4] to ensure that quantization noise will predominate over the line noise in its effect on the derived circuit noise. The signal level in this case is that which will just reach the detector threshold.

Now from eqn. (7) we have the total noise power from zero to F_t at the end of a repeater section at a point where the signal level is L_s dB. For the above requirement to hold then

$$L_s - L_t = 14$$

Therefore

$$L_s - L_i - 15{\cdot}76 - \\ -10\lg(0{\cdot}2303\alpha_1\sqrt{F_t} - 1)/\alpha_1^2 F_v - \alpha_1\sqrt{F_t} = 14$$

or since $L_s - L_i = S_s$

$$S_s = 29{\cdot}76 + 10\lg[(0{\cdot}2303\alpha_1\sqrt{F_t} - 1)/\alpha_1^2 F_v] + \alpha_1\sqrt{F_t} \quad \ldots\ldots(28)$$

This equation gives (implicitly) the frequency band F_t which is available in terms of S_s, α_1 and desired derived circuit bandwidth F_v.

The number of p.c.m. channels which can be derived from bandwidth F_t is theoretically[4]

$$n_p = F_t/d.F_v$$

where d is the number of binary digits used to represent each signal sample; $d = 7$ for 7-digit p.c.m. In present day practice, however, the bandwidth requirements[5] for 7-digit p.c.m. are such that the number of channels available is

$$n_p = F_t/15F_v \quad \ldots\ldots(29)$$

Equations (28) and (29) can be used to calculate the value of n_p for any specific transmission line using a specific sending level. Figure 5 shows the frequency bands and numbers of p.c.m. channels available, as a function of S_s, for a few transmission lines used in telephone plant.

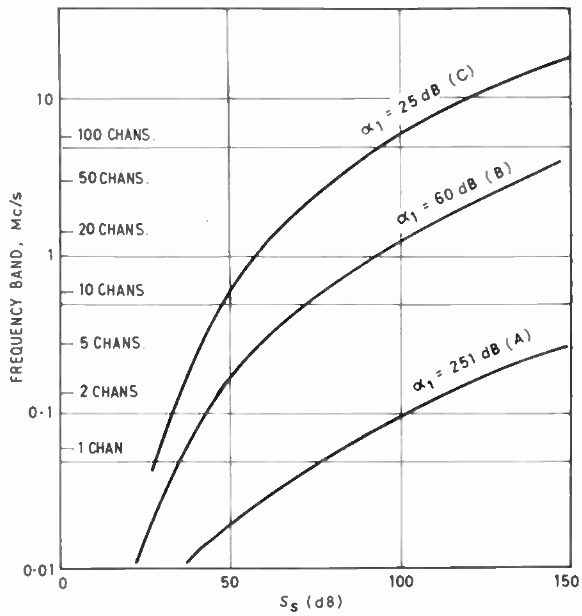


Fig. 5. Available frequency band and number of p.c.m. channels for three representative transmission lines.
Curve (A) 10 miles of 20 lb p.i.q.l.
Curve (B) 15 miles of ⅜ in coaxial cable.
Curve (C) 6¼ miles of ⅜ in coaxial cable.

7. Conclusion

Figure 6 shows the results of the above calculations applied to the case of a transmission line having an attenuation of 60 dB at 1 Mc/s (i.e. $\alpha_1 = 60$). This is approximately the attenuation of 1 mile of 6½-lb

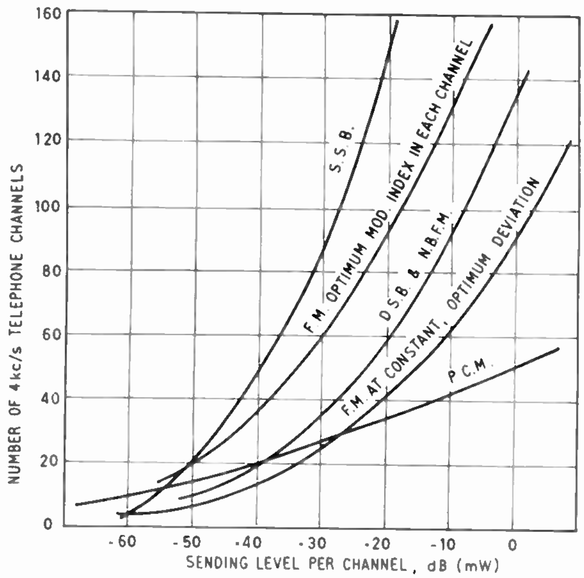


Fig. 6. Telephone channel capacity of line having an attenuation of 60 dB at 1 Mc/s.

paper-insulated quadded local type cable. The 'sending level per channel' in Fig. 6 means the transmitted level corresponding to channel overload in the case of s.s.b.-a.m., and in the case of p.c.m. it means the level corresponding to peak pulse power. In the other cases the carrier power level is referred to. Line noise is assumed to be thermal agitation noise only. The curves are calculated for a derived channel bandwidth of 4 kc/s and a 70-dB signal/noise level difference S_r.

It will be observed that s.s.b.-a.m. provides the greatest number of channels over the predominant part of the power level range plotted in Fig. 6. These power levels represent practical values for present day techniques using semiconductor active devices in the repeaters.

8. Acknowledgments

The permission of the Engineer-in-Chief, Australian Post Office, to publish this paper is gratefully acknowledged. The work was carried out as a result of a suggestion made by J. C. Wilson of the A.P.O. Research Laboratories.

9. References

1. "Reference Data for Radio Engineers", 4th ed., p. 23. (International Telephone and Telegraph Corp., New York, 1956.)
2. K. Küpfmüller, "Die Systemtheorie der Elektrischen Nachrichtenübertragung". (S. Hirzel Verlag, Zürich, 1949.)
3. "Twelve-channel Open-wire Carrier System". Commonwealth of Australia, Postmaster-General's Department, Specification 820.
4. E. Bedrosian, "Weighted p.c.m.", *Trans. Inst. Radio Engrs, (Information Theory)*, IT-4, p. 45, March 1958.
5. M. J. Kelly, "The trends of telecommunications as affected by solid-state electronic instrumentation", *Alta Frequenza*, 5, p. 333, October 1957.

10. Appendix 1: Effect of F.M. Threshold Phenomenon on the Validity of Eqn. (18)

As discussed in Section 5.3, the effect of noise bursts due to breakthrough of the f.m. threshold will be satisfactorily small if the signal/r.m.s. noise level difference at the demodulator input is equal to or greater than 12 dB. Thus,

$$L_s - \alpha_1\sqrt{F_c} - [L_i + 10\lg 2(1+mK)] \geqslant 12 \quad \text{...(30)}$$

where F_c is the unmodulated carrier frequency. In the highest frequency channel, the carrier frequency is given by eqn. (13). Substituting from eqns. (13) and (15)

$$S_r = L_s - \alpha_1\sqrt{F_{nc}} - L_i + 20\lg[m\sqrt{3/2}]$$

and since $L_s - L_i = S_s$

$$S_s - S_r = \alpha_1\sqrt{F_{nc}} - 20\lg[m\sqrt{3/2}] \quad \text{......(31)}$$

Combining (30) and (31) we obtain, since $F_{nc} = F_c$

$$S_r \geqslant 16{\cdot}77 + 10\lg[m^2(1+mK)] \quad \text{......(32)}$$

Equation (18) gives a valid value for m_0, the optimum modulation index, provided eqn. (32) is satisfied. When the modulation index is optimum, then the threshold requirement can be expressed as

$$S_s \geqslant 17{\cdot}38(1{\cdot}965 + 1/m_0K) + 10\lg[(2/3)(1+m_0K)] \quad \text{......(33)}$$

Expressions (32) and (33) are plotted on Fig. 3.

11. Appendix 2: Evaluation of Eqn. (23)

The calculation of the number of w.b.f.m. channels available in the frequency band F_1 to F_2 involves the evaluation of eqn. (23) of the text:

$$n_w = L\int_{F_1}^{F_2} \frac{dF}{1+M\exp(N\sqrt{F})} \quad \text{......(23)}$$

where

$$L = 1/(2F_v), \qquad M = K\exp[(S_r - S_s - 1{\cdot}76)/8{\cdot}686]$$

and

$$N = \alpha_1/8{\cdot}686$$

We substitute $Y = \exp N\sqrt{F}$, noting that $MY = mK$. Then

$$n_w = \frac{2L}{N^2}\int_{Y_1}^{Y_2} \frac{\ln Y}{Y(1+MY)}\,dY \quad \text{......(23a)}$$

where $Y_1 = \exp N\sqrt{F_1}$, etc. Separating into partial fractions

$$n_w = \frac{2L}{N^2}\int_{Y_1}^{Y_2} \frac{\ln Y}{Y}\,dY - \frac{2LM}{N^2}\int_{Y_1}^{Y_2} \frac{\ln Y}{1+MY}\,dY \quad \text{(23b)}$$

The first term in this equation reduces to

$$\frac{2L}{N^2}\int_{Y_1}^{Y_2} \frac{\ln Y}{Y}\,dY = L(F_2 - F_1) \quad \text{......(23c)}$$

The integral in the second term, on integrating by parts, becomes

$$\int_{Y_1}^{Y_2} \frac{\ln Y}{1+MY}\,dY = \frac{N}{M}[\sqrt{F_2}\ln(1+m_2K) - \sqrt{F_1}\ln(1+m_1K)] - \frac{1}{M}\int_{Y_1}^{Y_2} \frac{\ln(1+MY)}{Y}\,dY \quad \text{......(23d)}$$

where m_2 is the modulation index required for a carrier frequency F_2 and is given by eqn. (20). The integral remaining in eqn. (23d) must be evaluated in series form using the series

$$\ln(1+x) = x - x^2/2 + x^3/3 \ldots$$

where $x^2 < 1$. Then, for $(MY)^2 < 1$

$$\int \frac{\ln(1+MY)}{Y} \mathrm{d}Y = MY - \frac{(MY)^2}{2^2} + \frac{(MY)^3}{3^2} - \ldots \quad \ldots\ldots(23e)$$

and for $(MY)^2 > 1$

$$\int \frac{\ln(1+MY)}{Y} \mathrm{d}Y =$$

$$\tfrac{1}{2}(\ln MY)^2 - 1/(MY) + 1/2^2(MY)^2 - 1/3^2(MY)^3 + \ldots \quad \ldots\ldots(23f)$$

Because of the limited range of validity of each of these series, the definite integral

$$\int_{Y_1}^{Y_2} \frac{\ln(1+MY)}{Y} \mathrm{d}Y$$

must be evaluated in two sub-ranges, Y_1 to $1/M$ and $1/M$ to Y_2. The function and its integral are, however, continuous for all positive values of M and Y. The definite integral is a function of MY_1 and MY_2 and hence of m_1K and m_2K. Thus we write

$$\int_{Y_1}^{Y_2} \frac{\ln(1+MY)}{Y} \mathrm{d}Y = [\mathrm{f}(m_2K) - \mathrm{f}(m_1K)] \quad (23g)$$

Re-writing eqn. (23b) using the results given by eqns. (23c), (23d) and (23g), we get

$$n_w = L(F_2 - F_1) - (2L/N)[\sqrt{F_2}\ln(1+m_2K) - $$
$$- \sqrt{F_1}\ln(1+m_1K)] +$$
$$+ (2L/N^2)[\mathrm{f}(m_2K) - \mathrm{f}(m_1K)] \quad \ldots(23h)$$

which, on substitution for L, M, and N, gives eqn. (24) of the text.

*Reprinted from The Proceedings of the Institution of Radio Engineers Australia, **24**, No. 7, July 1963. (Paper No. 892.)*

U.D.C. 621.396.662: 621.382.3

Analysis of Series Regulators as Active Two-port Networks

By

R. A. ZAKAREVICIUS,

B.Sc., M.Eng.Sc.†

Summary: Regulators for regulated power supplies are here treated as active two-port networks. To characterize the performance, two parameters: the regulation factor $\mathcal{R}$ and the internal impedance Z_{out} are defined for the ideal case of infinite load impedance and zero source impedance of the unregulated side. A general theory is developed and specific circuit arrangements are shown to be particular cases of the general theory. For the simplest series transistor regulator it is found that $\mathcal{R} = h_{rb}$ and $Z_{out} = h_{ib}$, h_{rb} and h_{ib} being standard hybrid parameters of the transistor. A more advanced series transistor regulator is then dealt with and extension of the theory to more complicated cases is discussed. The treatment is from a transistor point of view, but the method is also applicable to vacuum tubes.

List of Symbols

A	Feedback amplifier gain.
$\lvert A \rvert$	Magnitude of A.
$A_{a.c.}$	*Magnitude* of A for 100 c/s.
$A_{d.c.}$	*Magnitude* of A for d.c.
a	As a last subscript refers to transistor VTa.
β	Common-emitter short-circuit current transfer ratio of a single transistor or of two transistors compound-connected as in Fig. 5.
E_{in}, E_{out}	Large-signal input and output voltage of a regulator.
e_{in}, e_{out}	Small-signal input and output voltage of a regulator.
e'_{in}, e'_{out}	Small-signal input and output voltage of a modified regulator.
e_f	Small-signal feedback voltage.
h	Standard hybrid parameter of a single transistor or of two transistors compound-connected as in Fig. 5.
h'	Standard hybrid parameter of three transistors compound-connected as in Fig. 5.
I_{in}, I_{out}	Large-signal input and output current of a regulator.
i_{in}, i_{out}	Small-signal input and output current of a regulator.
R_i	Current sampling resistance.
$\mathcal{R}$	Regulation factor of a regulator.
$\mathcal{R}'$, $\mathcal{R}''$	Regulation factor of a modified regulator.
$\mathcal{R}_{a.c.}$	Regulation factor for 100 c/s.
$\mathcal{R}_{d.c.}$	Regulation factor for d.c.
s	As a last subscript refers to VTs, either as a single or a compound-connected transistor.
VTa	Transistor providing feedback amplification.
VTs	Either a single series transistor or a compound-connected series transistor.
x	Input sampling fraction.
$x_{a.c.}$	Value of x for 100 c/s.
$x_{d.c.}$	Value of x for d.c.
y	Output sampling fraction.
z	Dynamic impedance of either a Zener diode or an ordinary diode.
Z_f	Output impedance of feedback amplifier.
$Z_{f(a.c.)}$	Z_f for 100 c/s.
$Z_{f(d.c.)}$	Z_f for d.c.
$Z_{f(reg)}$	Z_f in Fig. 4.
Z_g	Output impedance of the unregulated input to a regulator.
Z_L	Impedance of a load.
Z_{out}	Output impedance of a regulator.
Z'_{out}, Z''_{out}	Output impedance of a modified regulator.
$Z_{out(d.c.)}$	Value of Z_{out} for d.c.
1, 2, 3	As subscripts for transistor parameters, the numbers refer to transistors numbered as in Fig. 5.

1. Introduction

Regulators for stabilized power supplies are normally analysed by writing down the relevant circuit equations and solving them in the usual way to obtain expressions for the performance parameters in terms of circuit parameters.[1, 2] This procedure is not ideally suited for design purposes as it has essentially to be repeated for every major circuit alteration. The following analysis is a development of the ideas of Hill.[3] It views the regulator as an active linear network and treats individual circuit arrangements as particular cases of a general network. This viewpoint allows regulator analysis to be done by a simple step-

† School of Electrical Engineering, University of Sydney, New South Wales.

by-step procedure, applicable to many series regulators, which has a number of conceptual advantages. The analysis is presented from a transistor point of view, but many of the ideas are also applicable to the vacuum-tube case. In Section 2 the general theory is developed and is applied to particular cases in Section 3.

2. General Theory

2.1. *General Regulator*

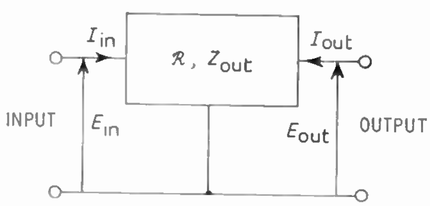
A general regulator is shown in Fig. 1(*a*). If E_{in} and I_{out} are taken as the independent variables, $E_{out} = f(E_{in}, I_{out})$. Let the regulation factor $\mathscr{R}$ and the output impedance Z_{out} be defined, following Hill,[3] by

$$\mathscr{R} = \left.\frac{\partial E_{out}}{\partial E_{in}}\right|_{\Delta I_{out}=0} \quad \text{and} \quad Z_{out} = \left.\frac{\partial E_{out}}{\partial I_{out}}\right|_{\Delta E_{in}=0}$$

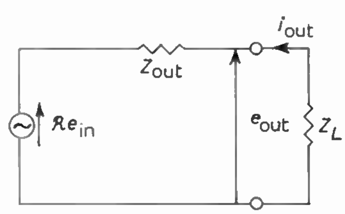
$\mathscr{R}$ and Z_{out} are thus parameters which give the variation in output voltage due to variations in input voltage and output current. Using incremental quantities of current and voltage we have

$$e_{out} = \mathscr{R}e_{in} + Z_{out} \cdot i_{out}$$

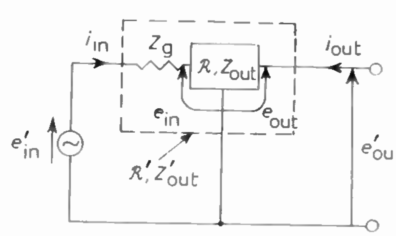
This equation is represented by Thevenin's theorem equivalent circuit of Fig. 1(*b*).



(*a*) General regulator.



(*b*) Equivalent circuit of the general regulator.



(*c*) Inclusion of input source impedance.

Fig 1.

$\mathscr{R}$ is the actual regulation factor in the case of infinite load impedance and Z_{out} is the output impedance for the ideal case of zero source impedance of the unregulated side. Yet $\mathscr{R}$ and Z_{out} are sufficient to characterize completely the regulating properties, as simple relations exist[3] between the regulation factor and output impedance in an actual case and $\mathscr{R}$, Z_{out}. For example, from Fig. 1(*b*) it is seen that the regulation factor for finite load is defined by

$$\mathscr{R}' = \left.\frac{\partial E_{out}}{\partial E_{in}}\right|_{Z_L=\text{constant}}$$

and is simplified to

$$\mathscr{R}' = \frac{\mathscr{R}Z_L}{Z_{out}+Z_L}$$

The effect of the unregulated supply source impedance, Z_g, can be found by considering the modified regulator, with intrinsic parameters $\mathscr{R}'$, Z'_{out}, of Fig. 1(*c*). If $i_{out} \simeq -i_{in}$, as is true in a practical case, it is simply shown[3] that $\mathscr{R}' \simeq \mathscr{R}$ and $Z'_{out} \simeq \mathscr{R}Z_g + Z_{out}$. In what follows $\mathscr{R}'$ and Z'_{out} in general refer to the regulation factor and output impedance of a modified regulator, the modification being applied to a regulator with intrinsic parameters $\mathscr{R}$, Z_{out}.

In order completely to characterize the regulating properties, a third parameter, relating the variation in output voltage to variations in reference standards, should be defined. The effect of these variations, however, can in most cases be estimated simply by inspection and a third parameter would complicate matters unnecessarily. In what follows ideal references are assumed.

2.2. *Feedback Regulator*

In Fig. 2(*a*) the simple regulator of Fig. 1(*a*) is incorporated in a feedback regulator. At this stage assume that $Z_f = 0$. Simple manipulation of circuit equations gives

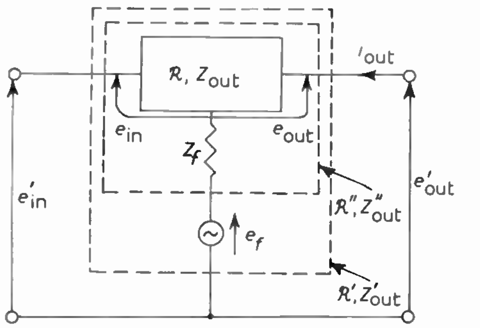
$$e'_{out} = \mathscr{R}e'_{in} + (1-\mathscr{R})e_f + Z_{out} \cdot i_{out}$$

This equation is represented by the equivalent circuit of Fig. 2(*b*). The most general case is represented by $e_f = Ae'_{out} + xe'_{in} + R_i i_{out}$. The Ae'_{out} term represents output voltage feedback, xe_{in} represents input sampling and $R_i i_{out}$ represents current feedback. A is the gain of an amplifier and x and R_i include the effect of amplifying stages and thus may be positive or negative quantities.

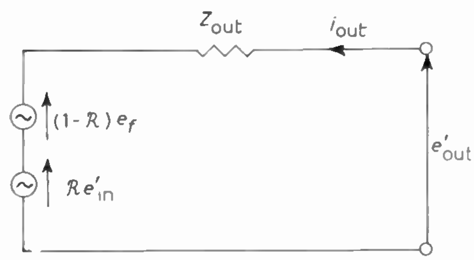
Further manipulation yields

$$\mathscr{R}' = \frac{\mathscr{R}+x(1-\mathscr{R})}{1-A(1-\mathscr{R})} \simeq \frac{\mathscr{R}+x}{1-A} \quad (\text{for } \mathscr{R} \ll 1)$$

$$Z'_{out} = \frac{Z_{out}+R_i(1-\mathscr{R})}{1-A(1-\mathscr{R})} \simeq \frac{Z_{out}+R_i}{1-A} \quad (\text{for } \mathscr{R} \ll 1)$$



(*a*) Feedback regulator.



(*b*) Equivalent circuit of the feedback regulator with $Z_f = 0$.

Fig 2.

If $Z_f \neq 0$, its effect can be included by considering first the $\mathscr{R}''$, Z''_{out} block of Fig. 2(*a*). The above results can be used directly by replacing $\mathscr{R}$ by $\mathscr{R}''$ and Z_{out} by Z''_{out}. It only remains to evaluate $\mathscr{R}''$, Z''_{out} in terms of the parameters of the $\mathscr{R}$, Z_{out} block and Z_f. This is shown in Section 3.

3. The Transistor Series Regulator

3.1. *The Simplest Transistor Regulator*

The simplest series transistor regulator is shown in Fig. 3. In this case Z_f is the parallel combination of R_1 and z. Both A and R_i are zero, but

$$x = \frac{z}{z+R_1} \simeq \frac{z}{R_1}.$$

If e_f and Z_f are zero, we would have $\mathscr{R} = h_{rb}$ and $Z_{out} = h_{ib}$, the h's being standard common-base

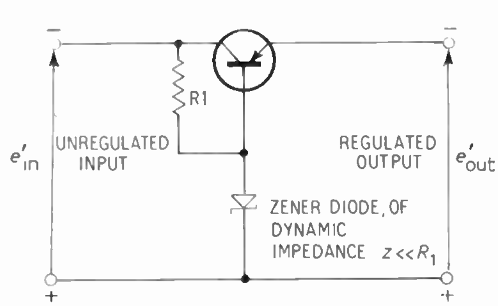


Fig. 3. The simplest transistor regulator.

hybrid parameters. This follows from their definition, although the terminals here designated as 'input', would normally be the output terminals if the transistor were used as a common base amplifier, and vice versa. With $Z_f = z$, manipulation shows that

$$\mathscr{R}'' = \frac{h_{rb}+zh_{ob}}{1+zh_{ob}}$$

$$Z''_{out} = h_{ib} + \frac{z(1+h_{fb})(1-h_{rb})}{1+zh_{ob}}$$

Generally

$$zh_{ob} \ll 1, \; h_{rb} \ll 1$$

and
$$1+h_{fb} = 1/(1+\beta) \simeq 1/\beta.$$

Thus $\mathscr{R}'' \simeq h_{rb}+zh_{ob}$ and $Z''_{out} \simeq h_{ib}+z/\beta$. The inclusion of input sampling with $x = z/R_1$ makes $\mathscr{R}' \simeq h_{rb}+zh_{ob}+z/R_1$ and $Z'_{out} \simeq h_{ib}+z/\beta$.

It is seen that for an ideal regulator with $e_f = 0$ and $Z_f = 0$ only two of the four h-parameters are needed to specify the performance. With non-zero Z_f the expressions for $\mathscr{R}'$ and Z'_{out} involve the other two h-parameters.

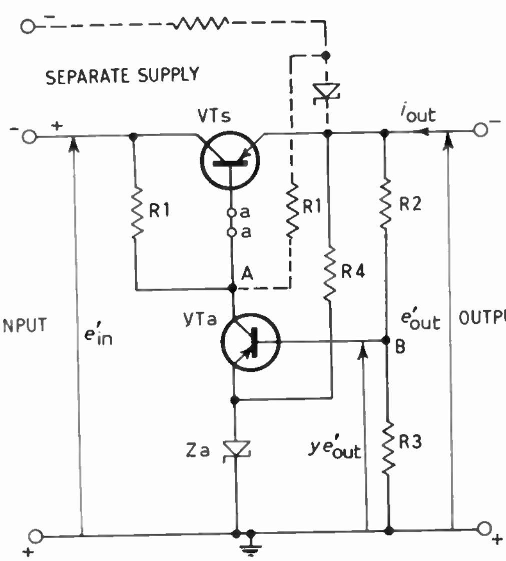


Fig. 4. Simple feedback regulator.

3.2. *Simple Feedback Regulator: An Actual Circuit*

Figure 4 is a simple voltage regulator with an amplifier in the feedback loop provided by the transistor VTa. A modification for better performance is drawn in dashed lines. Referring now to the full line circuit, $e_f = xe'_{in} + Aye'_{out}$. Here x is close to unity and is approximately equal to $1/(1-R_1h_{oea})$, where h_{oea} is the h_{oe} of VTa. A is the open-circuit gain of the feedback amplifier (i.e. it is the gain from B to A with a—a open-circuited); y is the output sampling fraction defined in Fig. 4. Z_f here is $Z_{f(reg)}$, the output impedance of the amplifier at the point A, which is approximately equal to the parallel combination R_1 and $1/h_{oea}$. If it is assumed that R_2, R_3 and R_4 provide negligible loading of the output and if $R_1\ h_{oea} \ll 1$, then it is clear that

$$\mathscr{R}' = \frac{1+h_{rbs}+Z_{f(reg)}h_{obs}}{1-Ay} \simeq \frac{1}{|A|y+1}$$

$$Z'_{out} = \frac{h_{ibs}+Z_{f(reg)}/\beta_s}{1-Ay} = \frac{h_{ibs}+Z_{f(reg)}/\beta_s}{|A|y+1}$$

The subscript s refers to the series transistor VTs.

The analysis clearly shows that the feedback amplifier prevents the utilization of the intrinsic regulating action of the series transistor by introducing an x which is unity and positive and which completely supresses the effect of h_{rbs}. The advantage of introducing the feedback amplifier, then, lies in the reduction of Z'_{out}, yet maintaining $\mathscr{R}' = 1/(|A|y+1)$ at a satisfactory level. The dashed modification is seen as a way of minimizing the unwanted input sampling. Likewise, the practice of splitting R1 into two resistors and connecting a capacitor from the junction to ground is seen as a considerable reduction in x for a.c. ripple, with only a relatively small reduction in A.

The effect of a non-zero dynamic impedance of the Zener diode in Fig. 4 is here included in A. If the loading effect of R2, R3 and R4 cannot be neglected, the effect is equivalent to the application of an external load of magnitude equal to the parallel combination of R4 and R2+R3. It is interesting to note that the loading by VTs of the feedback amplifier affects the expressions for $\mathscr{R}'$ and Z'_{out} differently. This is because in the determination of $\mathscr{R}'$ i_{out} is in effect set to zero and any current in the base of VTs must be due purely to the input voltage fluctuations being impressed across $1/h_{obs}$. In effect the total load of amplifier becomes $Z_{f(reg)}$ in parallel with $1/h_{obs}$. On the other hand, in the determination of Z'_{out}, e'_{in} is zero, the current in the base of VTs is $i_{out}/(\beta_s+1)$ and the total load of the amplifier becomes, in effect, $Z_{f(reg)}$ in parallel with $h_{ibs}\ (\beta_s+1)$.

In practical circuits a number of transistors in the compound-emitter connection are often used for VTs. To use the analysis which has been developed, it is necessary to obtain an expression for the h-parameters of the connection in terms of the h-parameters of the individual transistors. These can be readily calculated and the results are given in Appendix 1. Although the exact expressions are complicated, the approximate expressions are simple and still fairly accurate. However, it can be readily seen that rarely would it be necessary to use fully even the approximate expressions. Often the compound-emitter connection is modified by the introduction of various resistors. The effect of these, if it were desired, could be taken into account by first deriving the effect of these resistors on the h-parameters of the individual transistors. Such derivations have appeared in the literature.[4] Some of the exact expressions are complicated, but again approximate expressions are simple and accurate and rarely would it be necessary to use them in full.

4. Conclusions

The usual way of analysing regulators involves essentially the solution of the system equations. This essentially has to be repeated for every major circuit alteration, and it is not always easy to predict in advance the effect of circuit changes. Furthermore, the errors introduced by simplifying assumptions cannot be directly ascertained. This paper treats series transistor regulators as linear, active, two-port networks. A general theory is presented and specific circuit arrangements are seen as particular cases of the general theory. The two-port treatment allows the analysis to be confined to the ideal case of infinite load impedance and zero source impedance of the unregulated side. This already provides significant simplification, with no loss of generality, as simple relations exist between the performance parameters with other values of these impedances and the parameters in the ideal case. The further application of the general theory to series transistor regulators, using h-parameter representation, results in expressions which either are simple themselves or become simple on the application of approximations, most of which would introduce errors of less than 1%. These approximations correspond in general to the simplifying assumptions normally made in traditional analyses, but they are in effect applied to the end results so that their effect can be directly ascertained. The treatment of the individual circuit configurations as special cases of a general theory allows performance parameters to be evaluated almost by inspection and permits the ready visualization of the effect of various circuit connections.

The approach adopted here can also be used with vacuum-tube regulators. If in Fig. 3 the transistor is replaced by a vacuum-tube triode, with the grid, cathode and anode replacing base, emitter and collector respectively, the polarity of the supply voltage and of the Zener diode are reversed, the simplest vacuum-tube regulator is obtained. Then

$$\mathscr{R} = \frac{1}{\mu+1}$$

and

$$Z_{out} = \left(g_m + \frac{1}{r_a}\right)^{-1} = \frac{r_a}{\mu+1}$$

where g_m is the mutual conductance of a vacuum-tube triode, μ its amplification factor and r_a is the anode resistance. If grid current and capacitive effects are neglected, $h_{fb} = 1$ and $h_{ob} = 0$, and there are no correction terms due to Z_f. More complicated circuits can be treated in analogous fashion.

5. Acknowledgments

The author is indebted to Professor R. E. Aitchison and Mr. C. T. Murray for constructive criticism. This paper is based on part of a M.Eng.Sc. thesis in electrical engineering at the University of Sydney. The research was financed by the Electrical Research Board.

6. References

1. R. P. Middlebrook, "Design of transistor regulated power supplies", *Proc. Inst. Radio Engrs*, **45**, pp. 1502–9, November 1957.
2. F. A. Benson, "Voltage Stabilized Supplies". (Macdonald, London, 1957.)
3. W. R. Hill, jun., "Analysis of voltage-regulator operation", *Proc. Inst. Radio Engrs*, **33**, pp. 38–45, January 1945.
4. R. E. Aitchison, "Transistor circuit design using modified hybrid parameters", *Proc. Inst. Radio Engrs Aust.*, **20**, pp. 673–9, November 1959.

7. Appendix 1

Calculations of *h*-parameters

The expressions for the common-base parameters of a compound-emitter connection in terms of the common-base parameters of the individual transistors are presented below. The subscripts 1, 2, 3 refer to correspondingly numbered transistors in Fig. 5. For

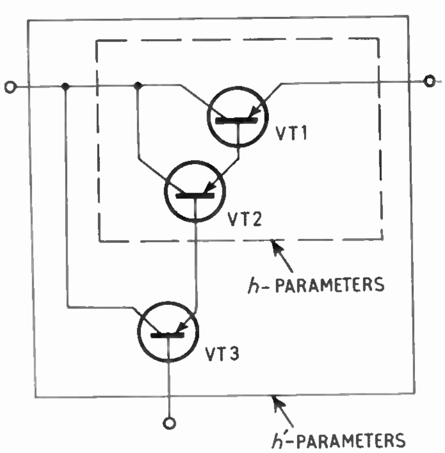


Fig. 5. The compound-emitter connection.

a combination of two transistors the expressions are:

$$h_i = h_{i1} + \frac{h_{i2}(1+h_{f1})(1-h_{r1})}{1+h_{i2}h_{o1}}$$
$$\simeq h_{i1} + h_{i2}/\beta_1$$

$$h_r = h_{r2} + \frac{(h_{r1}+h_{i2}h_{o1})(1-h_{r2})}{1+h_{i2}h_{o1}}$$
$$\simeq h_{r1} + h_{r2} + h_{i2}h_{o1}$$

$$h_f = h_{f1} + \frac{(h_{f2}-h_{i2}h_{o1})(1+h_{f1})}{1+h_{i2}h_{o1}}$$
$$\simeq h_{f1} + h_{f2}(1+h_{f1})$$

$$h_o = h_{o2} + \frac{h_{o1}(1-h_{r2})(1+h_{f2})}{1+h_{i2}h_{o1}}$$
$$\simeq h_{o2} + h_{o1}/\beta_2$$

The relation between h_f's may be expressed in another approximation as $\beta \simeq \beta_1 \beta_2$. For three transistors the expressions are:

$$h'_i \simeq h_{i1} + \frac{h_{i2}}{\beta_1} + \frac{h_{i3}}{\beta_1\beta_2}$$

$$h'_r \simeq h_{r1} + h_{r2} + h_{r3} + h_{i2}h_{o1} + h_{i3}h_{o2} + \frac{h_{i3}h_{o1}}{\beta_2}$$

$$\beta' \simeq \beta_1\beta_2\beta_3$$

$$h'_o \simeq h_{o3} + \frac{h_{o2}}{\beta_3} + \frac{h_{o1}}{\beta_2\beta_3}$$

The approximations above hold for normal transistors with

$$\left.\begin{matrix} h_{r1} \\ h_{r2} \\ h_{r3} \end{matrix}\right\} \ll 1, \quad \left.\begin{matrix} 1/h_{o1} \\ 1/h_{o2} \\ 1/h_{o3} \end{matrix}\right\} \gg \left\{\begin{matrix} h_{i1} \\ h_{i2} \\ h_{i3} \end{matrix}\right. \quad \text{and} \quad \left.\begin{matrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{matrix}\right\} \gg 1$$

The approximate expressions for three transistors were derived using the approximate expressions for two transistors.

8. Appendix 2

Numerical Example

It is required to design a regulated power supply with the following properties:

(1) Output voltage	12 V nominal.
(2) Output current	300 mA.
(3) D.c. output impedance $Z_{out(d.c.)}$	$< 0{\cdot}1\ \Omega$.
(4) Output voltage stability for line voltage fluctuations of $\pm 10\%$	$< 0{\cdot}3\%$.
(5) Line frequency ripple to be as small as practicable and in any case	$< 0{\cdot}01\%$.

The regulator is to be fed by an unregulated supply giving 20 V at 300 mA with 0·9 V peak-to-peak double line-frequency ripple and having $Z_g = 5{\cdot}0\ \Omega$.

Thus the regulator must have

$$\mathscr{R}_{d.c.} < \frac{12 \times 0{\cdot}003}{20 \times 0{\cdot}2} \simeq 1 \times 10^{-2}$$

and

$$\mathcal{R}_{\text{a.c.}} < \frac{12\times10^{-4}}{0{\cdot}9} \simeq 1\times10^{-3}$$

The subscript a.c. refers to 100 c/s.

(*a*) Consider the circuit in Fig. 3 with ASZ16 for the transistor, OAZ213 for the Zener diode and $R_1 = 1$ kΩ. With $h_{ib} = 0{\cdot}32\,\Omega$, $\beta = 140$ and $z = 10\,\Omega$,

$$Z_{\text{out(d.c.)}} = 0{\cdot}32 + \frac{10}{140} = 0{\cdot}39\Omega$$

This is too large. A feedback amplifier is called for and to provide enough gain with a low power transistor the compound-emitter connection must be used for the series transistor.

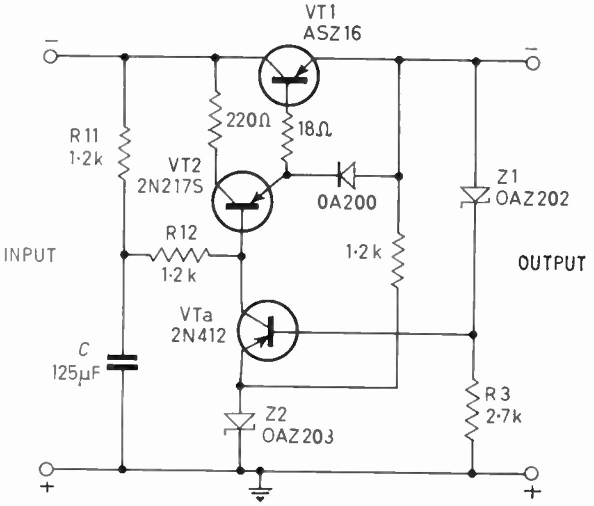


Fig. 6. Circuit of the regulator for the supply described in the numerical example.

(*b*) Consider now the circuit in Fig. 6. This employs a feedback amplifier to drive two transistors in compound-emitter connection. R_1 is split into R_{11} and R_{12} for better a.c. performance. As $A_{\text{a.c.}}$ is approximately proportional to R_{12} and $x_{\text{a.c.}}$ is approximately inversely proportional to R_{11}, optimum performance is obtained with $R_{11} = R_{12}$, as is done here. The value of 125 μF for *C* is about the largest practicable. A h.f. transistor is used in the feedback amplifier to avoid instability problems. The OA200 and 18 Ω resistor are inserted to limit the overload current. The 220 Ω resistor limits the dissipation in VT2.

8.1. *Parameters of the Compound-emitter Connection*

The following are the relevant parameters of the transistors and diodes of Fig. 6 at 300 mA output current obtained by actual measurement at the appropriate points.

	VT1: ASZ16		VT2: 2N217S
h_{ib1}	0·32 Ω	h_{ib2}	15 Ω
h_{rb1}	3×10^{-4}	h_{rb2}	4×10^{-4}
β_1	140	β_2	210
h_{ob1}	33×10^{-6} mho	h_{ob2}	$0{\cdot}21\times10^{-6}$ mho

	VTa: 2N412
h_{iea}	640 Ω
β_a	68
h_{oea}	$0{\cdot}12\times10^{-3}$ mho

Dynamic impedance of Z1, z_1 82 Ω
Dynamic impedance of Z2, z_2 3·6 Ω
Dynamic impedance of OA200, $z_3 \gtrsim 0{\cdot}3$ MΩ.

The effect of z_3 upon the parameters of VT1 can be neglected as it introduces errors at the most of the order $\beta_1 h_{ib1}/z_3$ (= 0·01%) or $h_{rb1}/h_{ob1} z_3$ (= 0·003%).[4] Likewise the effect of the 220 Ω resistor upon the parameters of VT2 can be neglected as it introduces errors at the most only of the order 220 h_{rb2}/h_{ib2} (= 0·6%) or 220 $\beta_2 h_{ob2}$ (= 1%). However, the 18 Ω resistor must be included. Clearly it has exactly the same effect as increasing h_{ib2} by 18 Ω. The overall parameters of the series transistors are thus:

$$h_{ibs} = 0{\cdot}32 + \frac{18+15}{140} = 0{\cdot}56\,\Omega$$

$$h_{rbs} = 3\times10^{-4}+4\times10^{-4}+(18+15)\times33\times10^{-6}$$
$$= 1{\cdot}8\times10^{-3}$$

$$\beta_s = 140\times210 = 2{\cdot}9\times10^4$$

$$h_{obs} = 0{\cdot}21\times10^{-6} + \frac{33\times10^{-6}}{210} = 0{\cdot}37\times10^{-6}\text{ mho}$$

8.2. *Feedback Parameters*

Input impedance of the feedback amplifier in parallel with R3

$$= \frac{R_3(h_{iea}+\beta_a z_2)}{R_3+h_{iea}+\beta_a z_2} = 660\,\Omega$$

Thus

$$y = \frac{660}{660+z_1} = 0{\cdot}89$$

$$A_{\text{d.c.}} = \frac{\beta_a(R_{11}+R_{12})}{(h_{iea}+z_2\beta_a)(1+(R_{11}+R_{12})h_{oea})} = 140$$

$$A_{\text{a.c.}} = \frac{\beta_a R_{12}}{(h_{iea}+z_2\beta_a)(1+R_{12}h_{oea})} = 80$$

$$x_{\text{d.c.}} = (1+(R_{11}+R_{12})h_{oea})^{-1} = 0{\cdot}78$$

$$x_{\text{a.c.}} = (2\pi 100 C R_{11})^{-1}(1+R_{12}h_{oea})^{-1} = 9{\cdot}3\times10^{-3}$$

$$Z_{f(d.c.)} = \frac{(R_{11}+R_{12})}{1+(R_{11}+R_{12})h_{oea}} = 1{\cdot}9\ \text{k}\Omega$$

$$Z_{f(a.c.)} = \frac{R_{12}}{1+R_{12}h_{oea}} = 1{\cdot}1\ \text{k}\Omega$$

8.3. *Parameters of the Regulator*

$$\mathscr{R}_{d.c.} = \frac{x_{d.c.}+h_{rbs}+Z_{f(d.c.)}h_{obs}}{yA_{d.c.}}$$

$$= \frac{0{\cdot}78+1{\cdot}8\times10^{-3}+0{\cdot}7\times10^{-3}}{0{\cdot}89\times140}$$

$$= 6{\cdot}3\times10^{-3}$$

$$\mathscr{R}_{a.c.} = \frac{x_{a.c.}+h_{rbs}+Z_{f(a.c.)}h_{obs}}{yA_{a.c.}}$$

$$= \frac{(9{\cdot}3+1{\cdot}8+0{\cdot}4)\times10^{-3}}{0{\cdot}89\times80}$$

$$= 1{\cdot}6\times10^{-4}$$

$$Z_{out(d.c.)} = \frac{h_{ibs}+\dfrac{Z_{f(d.c.)}}{\beta_s}}{A_{d.c.}y} = \frac{0{\cdot}56+0{\cdot}07}{140\times0{\cdot}89} = 5{\cdot}1\,\text{m}\Omega$$

8.4. *Parameters of the Complete Power Supply*

$$\mathscr{R}_{d.c.} = \mathscr{R}_{d.c.} \text{ of regulator only} = 6{\cdot}3\times10^{-3}$$

$$Z_{out(d.c.)} = \mathscr{R}_{d.c.}Z_g + Z_{out(d.c.)} \text{ (regulator only)}$$

$$= 37\ \text{m}\Omega$$

These values, together with the $\mathscr{R}_{a.c.}$, meet the requirements set out at the beginning.

8.5. *Experimental Confirmation*

Measurements on the circuit (shown in Fig. 6) at 300 mA output current gave the values shown in the table below, which also includes the calculated values for comparison.

		Calculated	Measured
Regulator only	$\mathscr{R}_{d.c.}$	$6{\cdot}3\times10^{-3}$	$6{\cdot}8\times10^{-3}$
	$\mathscr{R}_{a.c.}$	$1{\cdot}6\times10^{-4}$	$1{\cdot}3\times10^{-4}$
	$Z_{out(d.c.)}$	5·1 mΩ	4·7 mΩ
Complete power Supply	$\mathscr{R}_{d.c.}$	$6{\cdot}3\times10^{-3}$	$6{\cdot}4\times10^{-3}$
	$Z_{out(d.c.)}$	37 mΩ	38 mΩ

Manuscript first received by the Institution on 26th August 1963 and in final form on 30th November 1963. (Paper No. 893).

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".

BURIED AERIALS

In 1951 Dr. H. Paul Williams (now with the Air Defence Technical Centre of S.H.A.P.E.) was awarded the Institution's Clerk Maxwell Premium for a paper on "Subterranean communication by electric waves". In a recent lecture he explored rather wider aspects, namely buried antennæ and showed that, except for transmissions between deeply buried locations, the best performance is always obtained when a buried antenna is used horizontally. This results in a figure-of-eight radiation diagram in the horizontal plane, the maximum being along the axis of the antenna. The overall radiation diagram in the air has the shape of half a torus. A useful form of buried antenna is the resonant half-wave dipole. Due to the presence of the earth, its actual length is usually between $\lambda/6$ and $\lambda/8$, while the input resistance is of the order of 20 ohms. The resistance is a loss resistance associated with the induction field; the radiation resistance is very small since the efficiency of the antenna is only about $\frac{1}{2}n^2$, where n is the refractive index. For example, for a frequency of 500 kc/s and a conductivity of 0·01 mhos/m the efficiency would be about 0·1%. By making an array of 20 such dipoles, the efficiency could be increased to 2·5%.

"Buried antennas", H. P. Williams. *Tijdschrift van het Nederlands Elektronica-en Radiogenootschap*, **28,** No. 4, pp. 271–89, 1963.

PRE-DETECTION RECORDING

The techniques of recording before detection offer many advantages over the customary methods. The process of demodulation involves obtaining a reproduction of the information originally transmitted; and a receiving device with a conventional demodulator cannot give the highest accuracy of all kinds of information. All types of modulation can be equally well recorded by this technique without previous knowledge of the method of modulation used for the information. The system described by an engineer in France allows signals of very-low frequency, even down to d.c., to be stored for future use. It is claimed that the technique makes possible accuracy in the recording of information equal or superior to that obtained by the conventional system and moreover many methods of modulation, difficult to manage by conventional methods, are in this way handled with ease.

"Recording before detection", W. Fletcher. *L'Onde Électrique*, **43,** No. 440, pp. 1176–82, November 1963.

AUTOMATIC RADAR DATA PROCESSING

The video output of a surveillance radar receiver supplies several echo pulses from each aircraft during each rotation of the antenna, and noise appears superimposed as interference on this output signal. For automatic air traffic control systems aircraft have to be detected from this signal and expressed in co-ordinate form. For equal error probability and equal echo-signal/noise ratio the probability of detection by this detector is not substantially reduced when the radar video signal in front of the detector is digitized in unit bits. Using the theory of statistical decisions, a German paper quotes criteria for this form of data processing which ensure an optimum probability of detection and a minimum scatter of indication of the angles. Two types of digital detectors for which the criteria of probability of detection are slightly simplified, have been investigated mathematically and experimentally.

"Automatic evaluation of digitized radar signals", W. Storzand and W. D. Wirth. *Nachrichtentechnische Zeitschrift*, **16,** No. 12, pp. 643–8, December 1963.

ELECTRON TUNNELLING

The simplest type of electron tunnel device, the tunnel cathode, consists of two metal layers separated by a thin insulating film. A potential difference between the two metals causes electrons to tunnel from the one at lower potential to the one at higher potential. Electrons which have tunnelled into the second metal film can traverse this film if it is thin enough, and then be emitted into a vacuum. The energy distribution of the emitted electrons has been measured in an Australian valve laboratory and, from these results, it is possible to obtain values for the mean free path between collisions for hot electrons in a metal as a function of electron energy. These values can be correlated with those obtained by measurements of photo-excited electrons. Two other tunnel devices having possible future application are described. These are a three-terminal transistor-like device, consisting of successive layers of metal-insulator-metal-semiconductor and a device consisting of layers of semiconductor-metal-semiconductor. Although the work on these devices is still in the preliminary stages, they both show promise of having application in electronic circuitry.

"Theory and application of electron tunnelling", R. E. Collins. *Proceedings of the Institution of Radio Engineers Australia*, **24,** No. 10, p. 727, October 1963.