Diode Phase-Sensitive Detectors with Load

R. Chidambaram*, M.A., M.Sc., A.M.I.R.E., and S. Krishnan*, M.Sc.

A theoretical investigation of the operation of the simple diode push-pull phase-sensitive detector with load is carried out. The transfer ratios for the two diodes are found to vary considerably with the signal. The non-linearity in the output due to these variations is evaluated and a table is given from which the suitability of a given detector may be judged immediately. Experiments confirm quantitatively the theoretical results.

(Voir page 637 pour le résumé en français: Zusammenfassung in deutscher Sprache Seite 644)

THE simple push-pull phase-sensitive detector† finds application in various fields such as the generation of very slow sine waves, nuclear magnetic resonance studies, balance indication on a.c. bridges, etc.[1,2]. Though for some purposes, linearity of the detector is not essential, there are other cases where the linearity is of prime importance. The connexion of a load (which may be the measuring instrument itself) between the output terminals of the detector introduces a non-linearity, an effect not mentioned in the literature. This article deals with a theoretical investigation of this effect and the experimental verification of the conclusions reached.

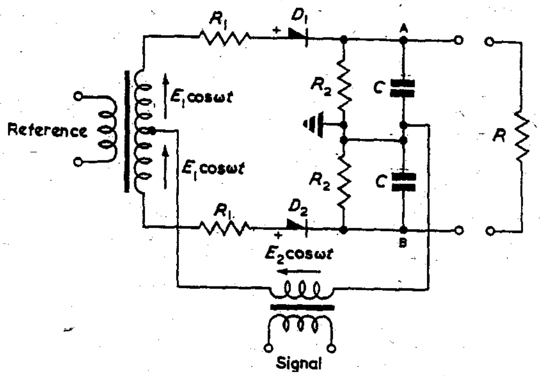

Fig. 1. Circuit of the simple push-pull phase-sensitive detector

The Simple Push-pull Phase-sensitive Detector

Fig. 1 shows the circuit of the detector. E_1 and E_2 are the amplitudes of the reference and signal voltages respectively, assumed to be in phase, as is customary. The sum of the two voltages is applied to one of the diodes, D_1, and the difference to the other diode, D_2. R_1 are the effective series resistances in the two diode circuits. If the value CR_2 is very large compared to the period of the applied sine waves, the voltages E_A and E_B developed at the output nodes, A and B, may be assumed to be steady. The output voltage is denoted by

$$E_o = E_A - E_B \quad \ldots\ldots\ldots\ldots\ldots\ (1)$$

In the absence of the external load R

$$E_A = k_o\ (E_1+E_2),\ E_B = k_o\ (E_1-E_2),\ E_2<E_1$$

$$E_o = 2k_oE_2 \quad \ldots\ldots\ldots\ldots\ldots\ (2)$$

The transfer ratio

$$k_o = \cos\ \theta_o \quad \ldots\ldots\ldots\ldots\ldots\ (3)$$

where $2\theta_o$ is the angle of conduction of either diode, θ_o being given by the following equation derived by Farren[3].

$$R_2/R_1 = \frac{\pi\ \cos\ \theta_o}{\sin\ \theta_o - \theta_o\ \cos\ \theta_o} \quad \ldots\ldots\ldots\ldots\ (4)$$

* *Indian Institute of Science, Bangalore.*

† *This detector has been commonly called a push-pull detector even though the output in general is not balanced.*

If now the external load R is connected across the output terminals, a current will flow through it from node A to node B. The effective resistance between node A and ground will consequently decrease while that between node B and ground will increase. This will make the transfer ratio for D_1 decrease and that for D_2 increase.

Suppose that the diodes, D_1 and D_2, conduct respectively over angles $2\theta_1$ and $2\theta_2$ during a cycle, their transfer ratios, k_1 and k_2, can be written down as

$$k_1 = \cos\theta_1,\ k_2 = \cos\theta_2.$$

$$E_A = k_1(E_1+E_2),\ E_B = k_2(E_1-E_2) \quad \ldots\ldots\ (5)$$

and

$$E_o = k_1(E_1+E_2) - k_2(E_1-E_2) \quad \ldots\ldots\ (6)$$

Since, in general, k_1 and k_2 are unequal, the output E_o is not proportional to the signal amplitude E_2.

Writing $x \equiv E_2/E_1$, equation (6) becomes

$$E_o = E_1\left\{k_1(1+x) - k_2(1-x)\right\} \quad \ldots\ldots\ (7)$$

The average current flowing through D_1 into node A is

$$\frac{1}{2\pi R_1}\int_{-\theta_1}^{\theta_1} E_1(1+x)\ (\cos\omega t - k_1)\ d(\omega t)$$

$$= \frac{E_1(1+x)}{\pi R_1}[\sin\theta_1 - k_1\theta_1]$$

This current may be equated to the current flowing out from node A through R_2 and R.

$$\frac{E_1(1+x)}{\pi R_1}(\sin\theta_1 - k_1\theta_1) = \frac{k_1E_1(1+x)}{R_2} + \frac{k_1E_1(1+x) - k_2E_1(1-x)}{R} \quad \ldots\ldots\ (8)$$

Equation (8) and the corresponding equation for node B can be written as

$$n_1/\pi\ (\sin\theta_1 - \theta_1\cos\theta_1) = \cos\theta_1(1+n_2) - wn_2\cos\theta_2 \quad \ldots\ (9)$$

and

$$n_1/\pi\ (\sin\theta_2 - \theta_2\cos\theta_2) = \cos\theta_2(1+n_2) - (n_2/w)\cos\theta_1,\ w\neq 0$$

where $\ldots\ldots\ldots\ (10)$

$$n_1 \equiv R_2/R_1,\ n_2 \equiv R_2/R \ \text{ and } w \equiv \frac{1-x}{1+x} \quad \ldots\ (11)$$

n_1 and n_2 may for obvious reasons be called the *source impedance factor* and the *load factor* respectively.

For given values of n_1, n_2 and w (or x), a pair of simultaneous transcendental equations have to be solved in θ_1 and θ_2. The method of solution is given in the Appendix.

Behaviour of the Transfer Ratios k_1 and k_2

The calculations have been made for the source impedance factor $n_1 = 1{\cdot}46\pi$, this value being recommended[4] for the elimination of third harmonic distortion because when used in equation (4) it leads to an angle of conduction $2\theta_o = 120°$. (As will be clear below, however, this condition is not really satisfied for either diode for any finite value of the signal).

Fig. 2 gives the variation of k_1 and k_2 against the signal-to-reference voltage ratio x for different values of the load factor n_2. It will be noted that in all cases, in the absence of the signal (i.e. $x = 0$), $k_1 = k_2 = k_o$ (equation (3)). This is also otherwise obvious since $E_A = E_B$ here and no current flows through R. As the signal voltage (or x) is increased, k_1 decreases and k_2 increases continuously. The deviations of k_1 and k_2 from k_o are more pronounced for larger values of the load factor n_2 (i.e. lower load resistances) and, incidentally, also for smaller values of the source impedance factor n_1.

For any value of n_2, there can be seen to exist a value of x at which k_2 reaches unity, corresponding to zero angle of conduction of D_2. This occurs when the voltage transferred

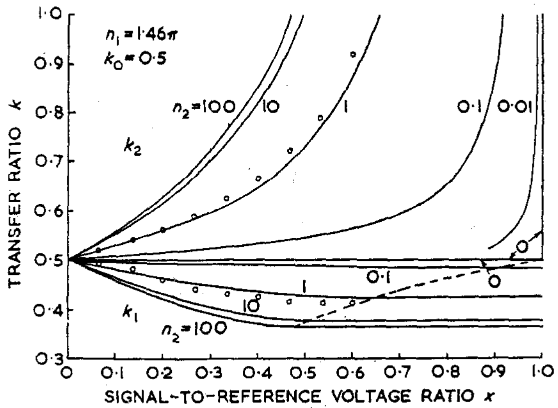

Fig. 2. Variation of the transfer ratios of the two diodes with the signal
The broken line is the locus of $k_{1(crit)}$ the value of k_1 where D_2 stops conducting. Experimentally observed points for $n_2 = 1$ are indicated by circles

from the node A to the node B through R equals the peak value of the voltage applied to D_2. For still higher values of x, D_2 continues to be non-conducting and E_B is obtained from E_A through potential division between R and R_2. E_o is now just a fraction, $1/(n_2 + 1)$, of $k_{1(crit)} E_1 + x)$, where $k_{1(crit)}$ is the value of k_1 at the point where D_2 stops conducting.

From equation (7), it can be seen that the ratio of the output d.c. voltage E_o to the signal voltage amplitude E_2, which may be called 'sensitivity' of the detector, is given by

$$\alpha \equiv E_o/E_2 = k_2 + k_1 - \frac{k_2 - k_1}{x} \quad \ldots\ldots\ldots (12)$$

It is interesting to note that even though k_1, $k_2 \to k_o$ as $x \to 0$, α does not tend to $2k_o$ because $(k_2 - k_1)x$ tends to a finite positive limit. The limit $\alpha_{x=0} \equiv \alpha_o$ has been theoretically evaluated in the Appendix.

Non-Linearity of the Detector

It might appear from an examination of Fig. 2 that the wide variations in k_1 and k_2 might cause correspondingly large variations in the sensitivity α (making the detector highly non-linear). However, it turns out that the non-linearity is not pronounced so long as the diode D_2 does not cease conducting. This is due to the fact that the variations in k_1 and k_2 tend to compensate each other partially.

It has been found that the graphical method described in the Appendix, although accurate enough for the determination of k_1 and k_2, is not satisfactory for estimating this small non-linearity. It is, however, possible to evaluate accurately the sensitivity, α, at $x = 0(\alpha_o)$ and at $x = x_{(crit)}(\alpha_{(crit)})$, where D_2 stops conducting. This has been done in the Appendix. Since it is positively undesirable to use the detector for $x > x_{(crit)}$, the total non-linearity N_o may be defined as

$$N_o \equiv 100 \frac{\alpha_o - \alpha_{(crit)}}{\alpha_o} \text{per cent} \quad \ldots\ldots\ldots\ldots (13)$$

Values of N_o calculated in this manner are given in Table 1 for various values of n_1 and n_2, along with the corresponding values of α_o and $x_{(crit)}$. The suitability of a detector can be judged immediately from this Table.

Experimental Verification

All the significant results of the above analysis have been verified experimentally for the value of source impedance factor $n_1 = 1{\cdot}46\pi$ and load factor $n_2 = 1$.

DETERMINATION OF TRANSFER RATIOS k_1 and k_2

Diode-connected 6AC7's were used for D_1 and D_2 because of their low forward resistance. The signal and reference voltages were obtained from the 50c/s mains.

TABLE 1

n_2 / n_1	100	30	10	3	1	0·3	0·1	0·03	0·01
1	0·00309 0·774 5·5%	0·0101 0·776 5·0%	0·0290 0·783 4·7%	0·0835 0·807 4·2%	0·181 0·851 3·2%	0·306 0·917 1·8%	0·381 0·964 0·8%	0·417 0·988 0·3%	0·429 0·996 0·15%
1·46 π . .	0·0125 0·470 9·8%	0·0405 0·477 9·6%	0·113 0·495 9·3%	0·297 0·564 7·5%	0·558 0·654 5·1%	0·808 0·809 2·4%	0·926 0·916 0·9%	0·977 0·973 0·4%	0·993 0·990 0·15%
10	0·0239 0·322 11.8%	0·0763 0·330 11·4%	0·205 0·352 9·8%	0·498 0·421 8·0%	0·843 0·555 4·9%	1·11 0·753 2·1%	1·23 0·892 0·8%	1·27 0·964 0·3%	1·28 0·988 —
30	0·0550 0·174 13·5%	0·170 0·183 12·4%	0·421 0·213 10·7%	0·870 0·292 7·2%	1·25 0·454 3·8%	1·48 0·696 1·6%	1·56 0·846 0·7%	1·59 0·954 —	1·60 0·985 —
100	0·127 0·084 13·9%	0·363 0·095 12·5%	0·777 0·125 9·4%	1·29 0·213 5·0%	1·60 0·391 2·2%	1·74 0·659 0·9%	1·78 0·849 0·3%	1·80 0·948 —	1·81 0·982 —

The set of three numbers in each square gives in that order the limiting sensitivity α_0, rounded off to three significant figures, the critical value of x, x_{crit}, where the diode D_2 stops conducting and the total non-linearity N_0 in the range $O < x < x_{crit}$.

The component values were:

$R_2 = 240\text{k}\Omega$
$R_1 = 52{\cdot}4\text{k}\Omega$
$R = 240\text{k}\Omega$
$C = 8\mu\text{F}$

The experimental points are marked in Fig. 2 and are seen to fit the corresponding theoretical curves closely.

Determination of Non-linearity

In this case, the signal voltage was maintained constant and the reference voltage varied since the change in the output voltage then indicates directly the non-linearity. Also the meter does not need to be switched to different ranges with consequent variations in loading and sensitivity. The sensitivity α has been plotted in Fig. 3. From

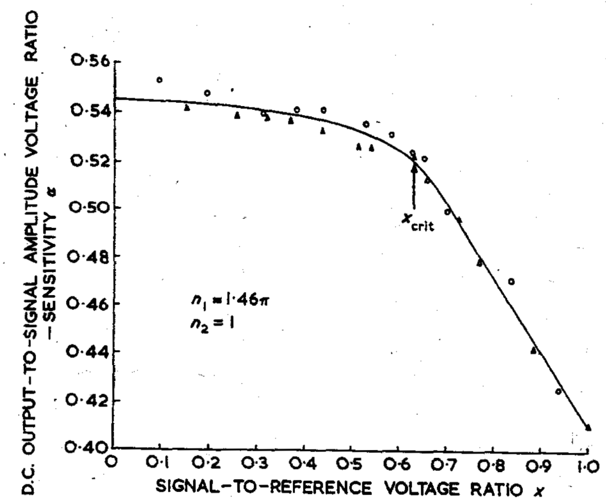

Fig. 3. Variation of the sensitivity α with the signal. *The value of $x_{(crit)}$ indicated has been obtained from Fig. 2*

the curve, using the value of $x_{(crit)}$ obtained in the previous experiment

$\alpha_o = 0{\cdot}545 \quad \alpha_{(crit)} = 0{\cdot}522,$

total non-linearity $N_o = 4{\cdot}2$ per cent.

The corresponding theoretically predicted values (taken from Table 1) are

$\alpha_o = 0{\cdot}558 \quad \alpha_{(crit)} = 0{\cdot}530, \quad N_o = 5{\cdot}1$ per cent.

Appendix

(a) Solution of the Equations (9) and (10) for $\cos\theta_1, \cos\theta_2$

From equations (9) and (10), one immediately obtains

$$w = -(F(\theta_1)/F(\theta_2)) \quad \ldots\ldots\ldots\ldots (14)$$

and

$$n_2 = 1\Big/\left\{\frac{\cos\theta_1}{F(\theta_1)} + \frac{\cos\theta_2}{F(\theta_2)}\right\} \quad \ldots\ldots\ldots\ldots (15)$$

where

$$F(\theta) \equiv (n_1/\pi)(\sin\theta - \theta\cos\theta) - \cos\theta \quad \ldots\ldots (16)$$

It is not possible to solve for θ_1 and θ_2 for given n_2 and w in a straightforward manner. Hence a graphical method of solution is adopted. For a given value of n_1, the functions $F(\theta)$ and $F(\theta)/\cos\theta$ are plotted against $\cos\theta$ for $0 \leqslant \theta \leqslant \pi/2$. It is known that for $x = 0$, the angles of conduction of the two diodes are equal and, as seen from equations (9) and (10), are the single root of the equation

$$F(\theta) = 0, \quad 0 \leqslant \theta \leqslant \pi/2 \quad \ldots\ldots\ldots\ldots (17)$$

From the physical arguments provided in the text, it is clear that for any n_2 ($\neq 0$) and $w(\neq 1)$

$$\theta_1 > \theta_o > \theta_2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (18)$$

Hence to obtain a plot such as is given in Fig. 2, a value of $\cos\theta_1$ is chosen such that $\theta_1 > \theta_o$, the corresponding $F(\theta_1)/\cos\theta_1$ is read from the curve and using equation (15), $F(\theta_2)/\cos\theta_2$ is calculated for the chosen value of n_2 and the corresponding value of θ_2 is read from the curve. The value of w consistent with this set $(\theta_1, n_2, \theta_2)$ is determined by reading the values of $F(\theta_1)$ and $F(\theta_2)$ from the $F(\theta)$ plot and using equation (14).

Note: $F(\theta)/\cos\theta$ and not $\cos\theta/F(\theta)$ is plotted because the latter has a discontinuity at θ_o, approaching the limits $+\infty$ and $-\infty$ from the θ_o+ and θ_o- sides respectively.

(b) Evaluation of $\alpha_o \equiv \operatorname*{Limit}_{x=0} \dfrac{k_1(1+x) - k_2(1-x)}{x}$

For a given small value of x, let the changes, $\Delta\theta_1$ and $\Delta\theta_2$, in the semi-angles of conduction of D_1 and D_2 be δ and $-p$ respectively where δ and p are small positive quantities. The signs have been so chosen because of condition (18).

Applying Taylor's theorem, it is easily shown that

$$F(\theta_o + \Delta\theta) \simeq a\Delta\theta + b(\Delta\theta)^2 \quad \ldots\ldots\ldots\ldots (19)$$

and

$$\frac{F(\theta_o + \Delta\theta)}{\cos(\theta_o + 6\Delta\theta)} \simeq c\Delta\theta + d(\Delta\theta)^2 \quad \ldots\ldots\ldots\ldots (20)$$

neglecting higher powers of $(\Delta\theta)$, where

$$\left.\begin{aligned} a &\equiv ((n_1\theta_o/\pi) + 1)\sin\theta_o \\ b &\equiv \tfrac{1}{2}\left\{(n_1/\pi)\sin\theta_o + ((n_1\theta_o/\pi) + 1)\cos\theta_o\right\} \\ c &\equiv ((n_1\theta_o/\pi) + 1)\tan\theta_o \\ &\text{and} \\ d &\equiv \tfrac{1}{2}\left\{(n_1/\pi)\tan\theta_o + ((n_1\theta_o/\pi) + 1)(\tan^2\theta_o + \sec^2\theta_o)\right\} \end{aligned}\right\} \quad (21)$$

From equations (14) and (11)

$$x \simeq \frac{a(p - \delta) - b(p^2 + \delta^2)}{a(p + \delta) - b(p^2 - \delta^2)} \quad \ldots\ldots\ldots\ldots (22)$$

From equation (5), using Taylor's theorem

$$\left.\begin{aligned} k_1 &\simeq k_o - (\sin\theta_o)\,\delta - k_o\delta^2 \\ k_2 &\simeq k_o + (\sin\theta_o)\,p - k_o p^2 \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots (23)$$

From equations (22) and (23) and the definition of α_o,

$$\alpha_o = 2k_o - \operatorname*{Limit}_{\delta=0} \frac{a(p + \delta)^2 \sin\theta_o}{a(p - \delta) - b(p^2 + \delta^2)} \quad \ldots\ldots\ldots (24)$$

From equations (15) and (20) one can solve for p in terms of δ to obtain

$$p = \delta + ((c/n_2) + (2d/c))\,\delta^2 + 0(\delta^3) \ldots\ldots (25)$$

From equations (24), (25) and (21)

$$\alpha_o = \frac{2k_o}{1 + \dfrac{2n_2}{((n_1\theta_o/\pi) + 1)}} \quad \ldots\ldots\ldots\ldots (26)$$

It may be mentioned that the procedure for the accurate determination of θ_o is similar to the one described below for calculating $\theta_{1(crit)}$.

(c) Evaluation of $\alpha_{(crit)}$

It will be noted that at $x = x_{(crit)}$ when the diode D_2 stops conducting, the equivalent load resistance, R_{eq}, for the diode D_1 consists of the parallel combination of R_2 and $(R + R_2)$. The semi-angle of conduction of D_1, $\theta_{1(crit)}$, at this point is a solution of equation (4) with R_{eq}/R_1 for R_2/R_1. This equation has also to be solved graphically[3]. A single curve giving $\theta_{1(crit)}$ against R_{eq}/R_1 can serve for all phase-sensitive detectors. The value of $\theta_{1(crit)}$ obtained graphically may be refined using Taylor's theorem.

One can now make use of the fact, already mentioned, that at the critical point, the voltage at node B obtained through potential division between R and R_2 is equal to the peak value of the voltage applied to the diode D_2. That is

$$k_{1(\mathrm{crit})}\,(1 + x_{(\mathrm{crit})})\,\frac{R_2}{R + R_2} = (1 - x_{(\mathrm{crit})})$$

remembering that $k_{2(\mathrm{crit})} = 1$. Using equation (12),

$$\alpha_{(\mathrm{crit})} = \frac{2k_{1(\mathrm{crit})}}{n_2 + 1 - n_2 k_{1(\mathrm{crit})}} \quad \ldots\ldots\ldots\ldots (27)$$

Acknowledgments

The authors wish to thank Prof. R. S. Krishnan for his kind interest and encouragement and Dr. G. Suryan for many helpful discussions. One of the authors (S.K.) is grateful to the Council of Scientific and Industrial Research for a research assistantship and the other (R.C.) to the Department of Atomic Energy for a research fellowship.

REFERENCES

1. KRISHNAN, S. Diode Phase Detectors, *Electronic Radio Engr.* *36*, 45 (1959).
2. CHANCE, B. *et. al.* Waveforms, M.I.T. Radiation Lab. Series, Vol. 19, Ch. 7 (McGraw-Hill Book Co., 1949).
3. FARREN, L. I. Phase Detectors. *Wireless Engr.* *23*, 330 (1946).
4. FRANK, R. L. Harmonic Insensitive Rectifiers for A.C. Measurements. *Proc. Nat. Elec. Conference*, *8*, 495 (1952).

Television News Films by Transatlantic Telephone Cable

The BBC Engineering Division has developed a system for transmitting brief television news picture sequences and other short television film sequences of up to one minute's duration over a circuit of the transatlantic telephone cable normally used for sound. These picture transmissions can be sent over the cable in both directions.

The process employs a slow-speed flying-spot film scanner, the video signal from which is used to modulate a carrier for transmission over the cable. At the receiving end the signals are demodulated and used to operate a slow-speed film telerecording equipment.

Consideration of the characteristics of the Atlantic cable indicated that a maximum video frequency of 4·5kc/s could be used. It has therefore been necessary to effect as many economies in the bandwidth of the video signal as are compatible with acceptable picture quality. These economies are:

(1) Restriction of the horizontal definition to that corresponding with a bandwidth of 1·75Mc/s in the 405-line television system.
(2) A reduction to 200 lines using sequential scanning.
(3) The scanning at the transmitting end of only alternate film frames with each frame-scan reproduced on two adjacent film frames at the receiving terminal.

These measures result in reducing the 3Mc/s bandwidth of the British television system to approximately 450kc/s, the remainder of the bandwidth reduction being obtained by a decrease of the scanning speed until the maximum video frequency corresponds with the available 4·5kc/s upper limit. The time required to scan the film is approximately 100 times normal.

16mm film has been chosen because it is almost universally used for television news. The average length of television newsreel sequences is less than half a minute and the new system will make it possible to transmit facsimiles of these sequencies in much less time than it would take to fly them across the Atlantic. The speed of transmission of each film frame is about 75 times faster than previous methods of sending still pictures by facsimile transmission.

The effective picture repetition frequency of $12\frac{1}{2}$ per second results in satisfactory reproduction of most material excepting that in which rapid movement occurs. In this case special provision is made for the transmission of every frame of the original film. This of course doubles the transmitting time.

The new system uses a channel of the type normally used for transmitting music over the cable; such a channel has a nominal bandwidth of 6·4kc/s. In order to limit the variation in the group delay/frequency characteristic to a value which can be corrected, it is necessary to restrict the usable video bandwidth to 4·5kc/s.

Vestigial sideband transmission is used with a special form of negative-going amplitude modulation. The carrier frequency is 5kc/s and the whole of the lower sideband is transmitted, the vestige of the upper sideband extending from 5kc/s to 5·5kc/s.

An additional problem results from the need to remove from the circuit the volume-range compressors and expanders which are normally used. This makes it more difficult to achieve a satisfactory signal-to-noise ratio and therefore a special form of amplitude modulation has been used in which the maximum depth of modulation considerably exceeds 100 per cent. That part of the modulation envelope which extends beyond the normal 100 per cent, or zero carrier, condition is "folded back" in a positive-going direction. This method of modulation results in an increase in the effective depth of modulation and thus also in the signal-to-noise ratio of the system. In order to achieve the synchronous detection needed with this type of signal it is necessary to use a re-generated carrier at the receiving terminal and this must be locked in phase to the original transmitted carrier. The necessary bursts of carrier which occur during the synchronizing signals are used for this purpose.

Another problem which has necessitated careful design in the equipment results from the fact that the line scanning frequency of 25c/s makes it impossible to reduce the effects of mains hum by means of clamping circuits. It has therefore been necessary to keep hum through the system at an extremely low level. Should trouble from hum on the transatlantic circuit be experienced some benefit can be achieved by a further reduction of line scanning frequency at the cost of a slight increase in transmission time. The precise choice of frequency will depend upon whether the hum is predominantly 50c/s or 60c/s.

As in other television systems a synchronizing signal is transmitted at the beginning of each line-scanning period, in this case the full amplitude of the video signal being utilized for the triggering edge of the synchronizing signal. The field synchronizing signal consists of four similar pulses and protection is provided against these pulses interfering with the bursts of reference carrier which are used for oscillator locking.

Negative film will normally be used at the sending terminal, but the equipment can deal with either negative or positive film.

Identical film equipments are used at both terminals of the system. At the sending end this apparatus operates as a flying-spot scanner while at the receiving end it functions as a telerecording channel. The same cathode-ray tube is used for both purposes and is enclosed in a double mu-metal shield in order to minimize mains frequency interference.

The time required for each field scan is approximately 8sec; a separate monitor tube with a long persistence phosphor reproduces a recognizable picture.

The special film traction mechanism, which is operated by the synchronizing signal, pulls down two film frames at a time. Twin optical systems are needed to telerecord simultaneously on two adjacent film frames and very small lenses with a focal length of one inch and an aperture of $f/8$ were developed specially for this purpose.