The Bell System Technical Journal

Vol. XXIII *July, 1944* *No. 3*

Effect of Telegraph Distortion on the Margins of Operation of Start-Stop Receivers

By W. T. REA

Recent practical and theoretical investigations of the effect of signal distortion on the margins of operation of start-stop telegraph receivers have led to the development of improved methods of testing and adjusting receivers, have enabled criteria of distortion tolerance to be set up for subscribers' and monitoring receivers and regenerative repeaters, and have made possible the application of more convenient and accurate standards of telegraph transmission. This paper describes the causes of distortion occurring both externally and internally to the receiver and the effects of such distortion on the operating margins. Methods of determining the internal distortion of a receiver are described and some of the more important considerations involved in establishing distortion tolerance criteria are discussed.

DURING the past decade the proportion of Bell System telegraph service operated on a start-stop teletypewriter basis has shown a continuous increase. Whereas in 1930 about 65% of telegraph long-distance circuit mileage was manual Morse, the present proportion of teletypewriter and teletypesetter service stands at 92%. The rapid growth of teletypewriter switching facilities has been an important factor in this development.

Naturally, this situation has made increasingly important a thorough understanding of the factors which affect the performance of start-stop receivers. In the present paper, an effort will be made to show some relationships between signal distortion and the operating margins of start-stop receivers.

A properly designed start-stop telegraph receiver requires only a small portion of the time of each signal element to permit a selection to be made; i.e. to determine whether the signal element in question is marking or spacing. The remainder of the signal element gives an operating margin, and serves as a reserve to take care of imperfections in the receiver or distortions which the telegraph signals may suffer in their passage over lines and through repeaters. The greater the signal distortion is, the smaller will be the margin which remains in the receiver to overcome the effect of such factors as wear of parts, variation of adjustments, or differences in speed between transmitter and receiver.

A consideration of the effects of telegraph distortion on the margins of

operation of start-stop receivers may well begin with a brief review of the nature and causes of the various types of distortion commonly experienced by telegraph signals. Telegraph distortion is generally considered to be divided into three types or components: bias, characteristic distortion, and fortuitous distortion.[1] The magnitude of the distortion is expressed in per cent of a unit pulse.

The Components of Telegraph Distortion

Bias, which is the simplest and most common component of distortion, may be positive (marking) or negative (spacing). Positive bias appears as a uniform lengthening of all marking pulses and an equal uniform shortening of all spacing pulses. Conversely, negative bias appears as a uniform lengthening of all spacing pulses and an equal uniform shortening of all marking pulses.

Bias is caused by an improper relation between the levels at which the relay or other receiving device responds and the steady-state marking and spacing levels of the signal. For example, Fig. 1(B) shows the signals of Fig. 1(A) as they might appear as a symmetrical wave on a line. With such a wave zero bias will be received when the currents at which the receiving relay operates from spacing to marking and from marking to spacing are symmetrically located with respect to the average of the steady-state marking and spacing currents. That is, zero bias will be received if the relay operates from spacing to marking and from marking to spacing at *B-B*, or if the relay operates from spacing to marking at *A-A* and from marking to spacing at *C-C*, or if the relay operates from spacing to marking at *C-C* and from marking to spacing at *A-A*. Negative bias will be received if the relay operates in both directions at *A-A*, and positive bias will be received if it operates in both directions at *C-C*.

In Fig. 1(C) is shown an unsymmetrical wave, in which the transient from space to mark is more rapid than that from mark to space. In this case, positive bias will result when the relay operates in both directions at *B-B* or at *C-C*, but no bias will result if the relay operates in both directions at *A-A*.

In the remaining diagrams of Fig. 1 it is assumed that the relay operates in both directions at a level midway between the steady marking and spacing levels. Fig. 1(D) shows a wave in which the transients are of such duration that the steady-state value is not attained in the shortest pulse length. It will be seen that the operation of the relay is delayed less after a short pulse than after a long one, and that this is true whether the pulse be marking or spacing. This effect is known as *negative characteristic distortion*, and it tends to shorten short pulses and lengthen long pulses. When a series of unbiased dots (called telegraph reversals) is transmitted, a steady-state condition is reached, in which the delays become equal on all transitions.

Hence, the signals are received as sent. When biased reversals are transmitted, the longer pulses are further lengthened and the shorter pulses are further shortened, causing the bias of the received signals to be of greater magnitude than that of the transmitted signals.

Fig. 1(E) shows a wave in which the current overswings the steady-state value, and fails to complete the return to steady state within the duration of the shortest pulse. It will be seen that the operation of a relay will be

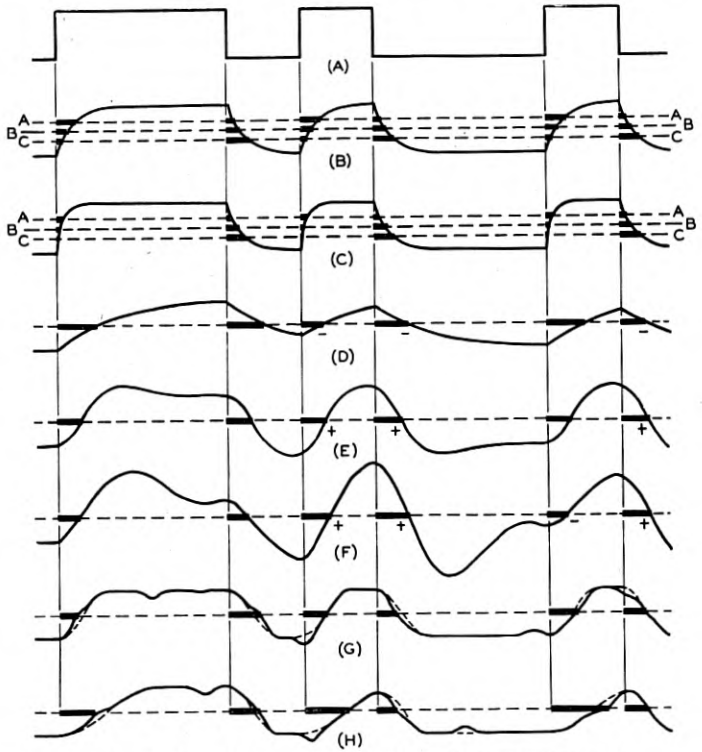


Fig. 1—Signal diagrams illustrating causes of distortion.

delayed more after a short pulse than after a long one, and that this is true whether the pulse in question be marking or spacing. This effect is known as *positive characteristic distortion*, and it tends to shorten long pulses and lengthen short ones. When unbiased reversals are transmitted, a steady-state condition is reached, in which the delays become equal on all transitions. Hence, the signals are received as sent. When biased reversals are transmitted, the shortening of the long pulses and lengthening of the short pulses causes the bias of the received signals to be less than that of the transmitted signals.

Fig. 1(F) shows a wave which performs a damped oscillation before settling to a steady state. This type of wave tends to produce a negative characteristic effect on certain transitions and a positive characteristic effect on others.

In general, if, on a given transition, the sum of all previous transients is such as to delay the operation of the receiving device, positive characteristic distortion is said to occur. If, on the other hand, the sum of all previous transients is such as to advance the operation of the receiving device, negative characteristic distortion is said to occur.

Bias and characteristic distortion, considered together, are called "systematic" distortion, because they occur with some regularity, and obey certain constant laws. There is another type of distortion that is not systematic. This is known as *fortuitous distortion*. It may be caused by the effect of various interfering currents on the receiving device. Fig. 1(G) shows a wave upon which interfering currents have been superposed. It will be noted that, for a given magnitude of interfering current, the more sloping the wave is in the region of the operating level of the receiving device, the greater will be the resulting fortuitous distortion.

Fortuitous distortion may also occur, in cases of extremely sloping wave-shape, due to the "indecision" of the receiving device, or, in other words, due to small variations of its effective operating level from signal to signal.

Fig. 1(H) shows a wave that is affected by interfering currents and in which the mark-to-space and space-to-mark transients have different slopes in the region of the operating level of the receiving device. The interfering current therefore causes fortuitous distortion of different magnitudes on mark-to-space and space-to-mark transitions. It will be shown later that distortion of this type affects a start-stop receiver in a particular manner which differs from the effect of distortion of the type illustrated in Fig. 1(G).

These, then are the generally-recognized components of telgraph distortion. More complicated effects ensue when characteristic distortion occurs on waves having dissimilar transients in the mark-to-space and space-to-mark directions, but a consideration of such phenomena is outside the scope of an elementary explanation of telegraph distortion, and is not necessary to an understanding of the effects of distortion on the margins of operation of start-stop receivers.

Start-Stop Displacements

The basic principles of operation of start-stop receivers have been described in previous articles[2,3]. A brief review of these principles will, therefore, suffice here.

The start-stop signal train consists of a start pulse, which is generally spacing, several selective pulses, each of which may be either marking or

spacing, and a stop pulse which is generally marking. The receiving mechanism is started by the transition at the beginning of the start pulse, and its speed is such that it arrives at the stop position before the end of the stop pulse occurs, and remains stopped until the succeeding start transition takes place. Thus any speed difference between the transmitter and receiver is prevented from cumulating for more than the duration of one signal train.

Since the receiving device starts anew at each start transition, and the instants of selection of the selective pulses are spaced in time relative to the instant of starting, as shown in Fig. 2(A), the start transition acts as a basic reference point to which all other instants of time during the selective cycle may be referred.

The advances and delays of the transitions of the start-stop signal train from their normal times of occurrence, relative to the start transition, are known as "start-stop displacements." Fig. 2(B) shows the four types of displacement that may occur: *MB* or "marking beginning displacement," which is the advance of a space-to-mark transition (beginning of a marking pulse) relative to the start transition; *SB* or "spacing beginning displacement," which is the delay of a space-to-mark transition relative to the start transition; *SE* or "spacing end displacement," which is the advance of a mark-to-space transition (end of a marking pulse) relative to the start transition; and *ME* or "marking end displacement," which is the delay of a mark-to-space transition relative to the start transition.

Effect of Bias on Displacement

Since bias affects all pulses alike, and since in the usual start-stop receiver the start transition is mark-to-space, the succeeding mark-to-space transitions of the signal train are not shifted relative to the start transition. Hence the total effect of the bias appears on the space-to-mark transitions. Positive bias causes MB displacement alone, as shown in Fig. 2(C). Negative bias causes SB displacement alone, as illustrated in Fig. 2(D).

The total range through which the selective periods may be shifted, relative to the start transition, without producing an incorrect selection is known as the orientation range of the receiver. Its limits are read on a scale calibrated from 0 to 100 in per cent of a unit pulse-length. Figure 3 is a graph of teletypewriter orientation range versus input signal bias, for a receiver whose range is from 10 to 90 on unbiased signals. Diagrams of this type are called "bias parallelograms."

Effect of Characteristic Distortion on Displacement

Characteristic distortion does not affect all pulses of miscellaneous signals alike, because, as explained above, the effect on each transition depends

upon the signal combinations that have previously been sent over the circuit. Hence the start transition and the transitions occurring between selective pulses are, in general, delayed by varying amounts. All four types of displacement shown in Fig. 2(B) occur, depending upon whether the transition in question is mark-to-space or space-to-mark and whether it has been

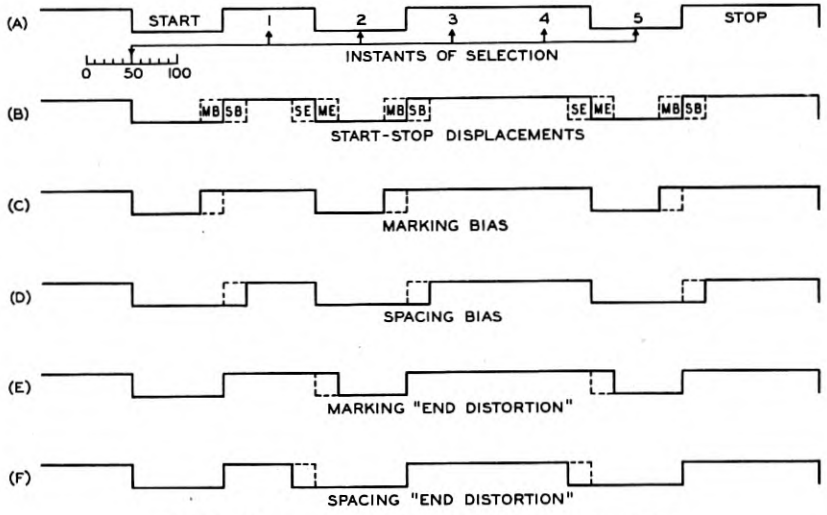


Fig. 2—Diagrams illustrating start-stop displacements.

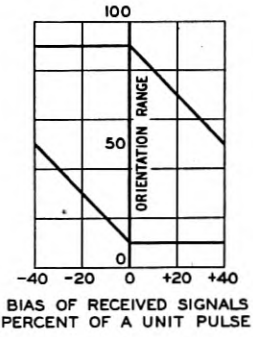


Fig. 3—The bias parallelogram.

delayed more or less than the start transition. For example, if a space-to-mark transition is delayed less (on an absolute time basis) than the start transition, *MB* displacement occurs; if more, *SB* displacement. If a mark-to-space transition is delayed less than the start transition, *SE* displacement occurs, if more, *ME* displacement.

Maximum Displacements Caused by Characteristic Distortion

The *maximum MB displacement* will occur when the start transition is delayed as much as possible and some space-to-mark selective transition is delayed as little as possible. This will take place, in the case of negative characteristic distortion, when as long a marking signal as is possible precedes the start transition and a combination of pulses as predominantly marking as possible precedes the space-to-mark transition in question. A marking signal sufficiently long to permit a steady state to be attained, followed by any signal train having the first selective pulse marking satisfies this condition, as shown at "*X*" in Fig. 4(B), but it will be noted that the *MB* displacement extends into the start pulse, where, in the case of a start-stop receiver, no selection is made. Hence it will not affect the margin of operation of the receiver, provided it is not so large as to prevent the receiver from starting. This particular distortion will, however, affect a start-stop distortion measuring set[4] or regenerative repeater which is so designed that measurements or selections are made during both the selective pulses and the start pulse. As far as a start-stop teletypewriter, in which no selection occurs during the start pulse, is concerned, the *maximum MB* displacement occurs on the fourth transition of the letter *K* following as long a marking signal as possible, as shown at "*Y*" in Fig. 4(B). This space-to-mark transition, being preceded by a spacing pulse of unit length which, in turn, was preceded by signals which are predominantly marking, is delayed for a short time, whereas the mark-to-space start transition, which was preceded by a long marking signal, is delayed for a longer time. Except in the case of unusual wave forms, there will be very little difference between the magnitudes of the displacements shown at "*X*" and "*Y*" unless they are both very large, since the wave will usually attain steady state during the steady marking interval constituted by the first, second, third and fourth selective signal intervals.

In the usual case of positive characteristic distortion, the maximum *MB* displacement will occur when the start transition is preceded by a combination of pulses as predominantly spacing as possible, and some space-to-mark transition is preceded by the longest spacing signal possible in the start-stop code. These conditions are met by repeated, "BLANK" signal trains, shown in Fig. 4(E).

The *maximum SB displacement* will occur when the start transition is delayed as little as possible, and some space-to-mark selective transition is delayed as long as possible. This takes place, in the case of negative characteristic distortion, when a combination of pulses as predominantly spacing as possible precedes the start transition and the longest possible spacing signal precedes the space-to-mark transition in question. As noted in the

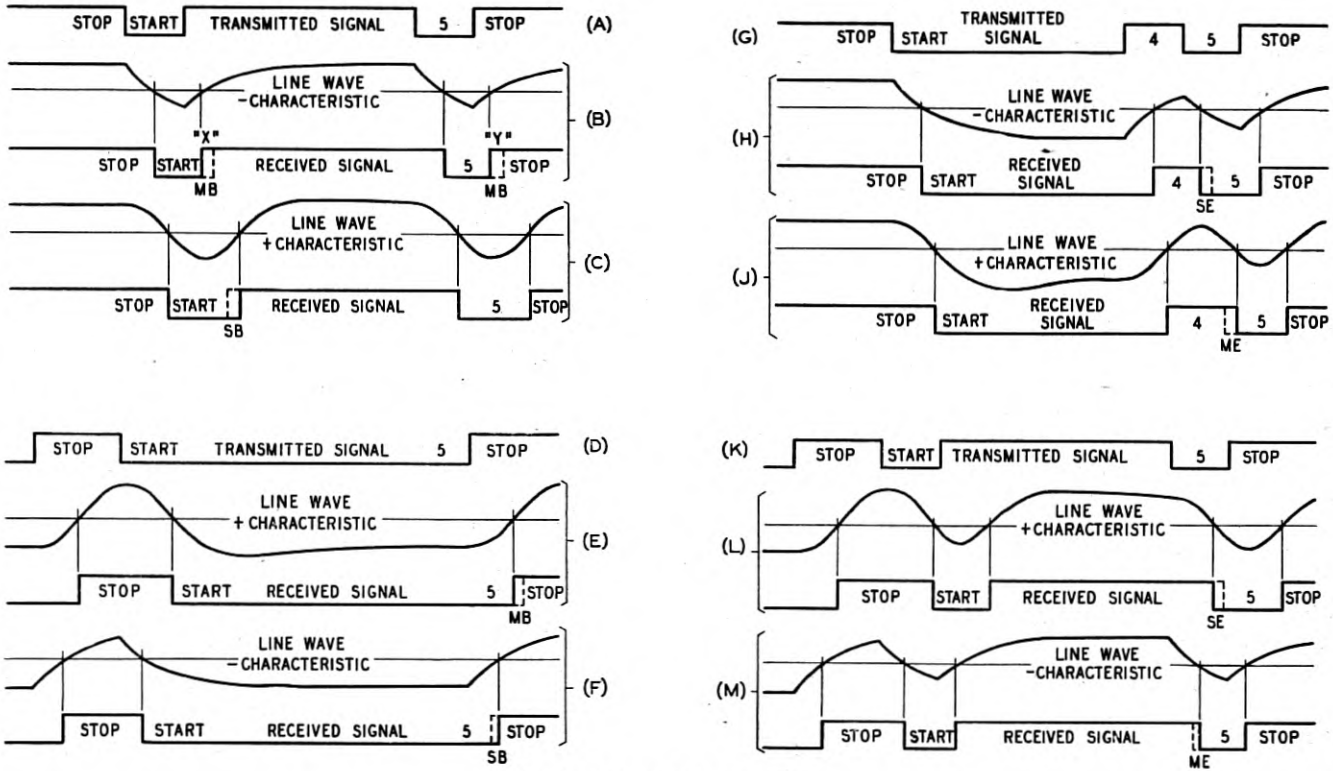


Fig. 4—Characteristic distortion.

preceding paragraph, repeated "BLANK" signals satisfy this description. Fig. 4(F) illustrates the effect of negative characteristic distortion on this signal combination. It will be noted that the resulting *SB* displacement extends into the stop pulse, where usually no selection takes place, and hence it would not affect the margin of operation of the start-stop receiver except, of course, in the case of a type of receiver, such as a regenerative repeater, in which selection of the stop pulse does occur. For the case of a receiver which does not select the stop pulse, the maximum *SB* displacement affecting the margin of operation occurs at the second transition of the letter "*T*" following repeated "BLANK" signals. Except in the case of very large distortions, this displacement will be of nearly the same magnitude as that illustrated in Fig. 4(F).

In the usual case of positive characteristic distortion, the maximum *SB* displacement will occur when the start transition is preceded by a long marking signal and some space-to-mark selective transition is preceded by a combination of pulses as predominantly marking as possible. As noted previously, this description is satisfied by a sufficiently long marking signal to permit the attainment of steady state, followed by any signal train having the first selective pulse marking. Figure 4(C) illustrates the effect of positive characteristic distortion on this type of signal.

The *maximum SE displacement* will occur when the start transition is delayed as much as possible and some mark-to-space selective transition is delayed as little as possible. This will take place, in the case of negative characteristic distortion, when a long marking signal precedes the start transition and a combination of pulses as predominantly spacing as possible precedes the mark-to-space transition in question. This description is answered by a long marking signal followed by a "CARRIAGE RETURN" signal train, as shown in Fig. 4(H). The *SE* displacement occurs at the end of the fourth selective pulse.

In the usual case of positive characteristic distortion, the maximum *SE* displacement will occur when the start transition is preceded by a combination of pulses which is as predominantly spacing as possible, and some mark-to-space selective transition is preceded by the longest possible marking signal. This description is satisfied by repeated "BLANK" signal trains followed by the letter "*K*," and, as shown in Fig. 4(L), the *SE* displacement occurs at the end of the fourth selective pulse.

The *maximum ME displacement* will occur when the start transition is delayed as little as possible and some mark-to-space selective transition is delayed as much as possible. This will take place, in the case of negative characteristic distortion, when the start transition is preceded by a combination of pulses which is as predominantly spacing as possible, and some mark-to-space selective transition is preceded by the longest possible marking

signal. As noted in the previous paragraph, the letter "*K*" preceded by repeated "BLANK" signal trains satisfies this description, and the *ME* displacement is as illustrated in Fig. 4(M).

In the usual case of positive characteristic distortion, the maximum *ME* displacement will occur when the start transition is preceded by a long marking signal, and some mark-to-space selective transition is preceded by a combination of pulses as predominantly spacing as possible. As seen previously, this description is answered by a long marking signal followed by a "CARRIAGE RETURN" signal train. Fig. 4(J) illustrates the *ME* displacement.

Effect of Characteristic Distortion on Orientation Limits

In the usual start-stop system which employs a stop pulse longer than the unit selecting pulse, characteristic distortion affects the upper and lower limits of orientation differently. This effect is due mainly to the longer stop pulse, although the fact that the start transition is always mark-to-space contributes to the effect.

In the case of negative characteristic distortion, the displacements (*MB* and *SE*) which affect the upper end of the orientation range are those in which the start transition suffers a long delay and a selective transition suffers a short delay. The delay of the start transition can be quite large, since it may be preceded by a long marking pulse. Moreover, the delay of the selective transition may be very short, since the pulse which precedes the transition can be of unit length, and this, in turn, may be preceded by a signal of the opposite type which may be of as much as four units duration. Hence these displacements, being the difference between a large and a small delay, are large.

On the other hand, the displacements, *SB* and *ME*, which affect the lower end of the range are those in which the start transition suffers only a fairly short delay and a selective transition suffers a long delay. The delay of the start transition can not be very short for two reasons: first, the start pulse cannot be preceded by a steady spacing pulse; and second, what is of more importance, the stop pulse is of greater than unit length. The delay of a selective transition can be long, as when the transition is preceded by a pulse of four or five units in length. (This delay may not be so long as that suffered by a start transition which follows a steady-state marking condition, but it is not much shorter.) Hence the *SB* and *ME* displacements, being the difference between a long selective transition delay and only a fairly short start transition delay, are smaller than the *MB* and *SE* displacements.

For this reason negative characteristic distortion affects the upper end of the range more than it does the lower.

In the case of what we have termed "the usual type of positive characteristic distortion," the displacements (*SB* and *ME*) which affect the lower end of the orientation range are those in which the start transition suffers a short delay and a selective transition suffers a long delay. The delay of the start transition can be quite short, since it may be preceded by a long marking signal. Moreover, the delay of the selective transition may be very long, since the pulse which precedes the transition can be of unit length and this, in turn, may be preceded by a signal of the opposite type which may be four or more units in length. Hence these displacements, being the difference between a short and a long delay, are large.

On the other hand, the displacements (*MB* and *SE*) which affect the upper end of the range are, in this type of distortion, those in which the start transition suffers only a fairly long delay and a selective transition suffers a short delay. The delay of the start transition cannot be very short for the two reasons mentioned previously. The delay of the selective transition can be short, as when the transition is preceded by a pulse four to six units in length. Hence the *MB* and *SE* displacements, being the difference between a short selective transition delay and only a fairly long start transition delay, are smaller than the *SB* and *ME* displacements.

For this reason positive characteristic distortion of this type affects the lower end of the range more than it does the upper.

In the case of a wave which oscillates, causing positive characteristic distortion on some transitions and negative on others, no such general statements as are made above are applicable. In practice, cases have been observed in which one end of the orientation range was cut and the other was actually extended.

Due to the fact that characteristic distortion delays the start transition by different amounts from character to character, it causes the character length to vary during continuous automatic transmission. The maximum variation in character length is roughly of the same magnitude as the maximum displacement affecting the selective pulses.

Effect of Fortuitous Distortion on Displacement

Fortuitous distortion causes the start transition to be delayed more or less than normal, and has the same effect on the selective transitions. Since it is usually equally probable that the maximum fortuitous effects will occur on mark-to-space or space-to-mark transitions and will increase or decrease their delay, this type of distortion generally produces the four types of displacement in equal magnitude, and this magnitude is equal to the maximum increase or decrease in the length of pulse.

An exception to the above statement occurs when the mark-to-space and space-to-mark transients give the wave different slopes at the point where the

receiving device operates. Then the magnitude of the fortuitous effect is different on mark-to-space and space-to-mark transitions. If the effect is greater on the space-to-mark transitions, *MB* and *SB* displacements are greater than *SE* and *ME*. If the opposite, *SE* and *ME* are greater than *MB* and *SB*. In all cases, however, the orientation range is reduced equally at both ends.

Fortuitous distortion also lengthens and shortens the character since it does not affect all transitions alike.

Internal Distortion

Telegraph signal distortion may occur within the start-stop receiver, and it should be expected that the components of distortion will have the same effect on the margins of operation as the same components external to the receiver. Consequently, it should be possible to determine the magnitudes of the various components of internal distortion by their effects on the margins of operation.

As mentioned previously, the upper end of the orientation range is determined by whichever of the displacements *MB* and *SE* is the greater; and the lower end by whichever of the displacements *SB* and *ME* is the greater. To discover the magnitude of the smaller type of displacement it is necessary to reduce the larger displacement by distorting the transmitted signals. For example, if a receiver has a large internal marking bias, the upper limit of orientation is determined by *MB* displacement, and hence the amount of *SE* displacement caused by internal distortion is concealed. However, by transmitting signals affected by *SB* displacement (in other words, signals biased to spacing), the total *MB* displacement is decreased until it is less than the internal *SE* displacement, whose effect on margin can then be found. Thus the internal distortion may be determined by observing the effect of external distortion on the margins of operation.

It is convenient to regard any start-stop receiver as a theoretically perfect receiver affected by certain types of internal distortion. The internal distortion is usually considered to be composed of bias, "skew" (defined later) and fortuitous distortion. (The internal characteristic distortion is generally included in "internal fortuitous distortion," since it is usually very small, and a fairly elaborate testing procedure is required to separate its effects from those of internal fortuitous distortion.) Internal bias and internal fortuitous distortion are of the same nature as the external effects previously described. Skew is said to occur when there exists the type of distortion, previously mentioned, in which the fortuitous effect on space-to-mark transitions differs in magnitude from that on mark-to-space transitions. When the former is greater the skew is said to be positive; when the latter, negative. Hence in positive skew, *MB* and *SB* displacements tend to

be larger; in negative skew, ME and SE displacements tend to be larger. The magnitude of the skew is defined as the difference between the magnitudes of the fortuitous effects on space-to-mark and mark-to-space transitions.

Figure 3 showed the bias parallelogram of a receiver which had a local margin of 10 to 90. Figure 5 shows the bias parallelogram of a perfect

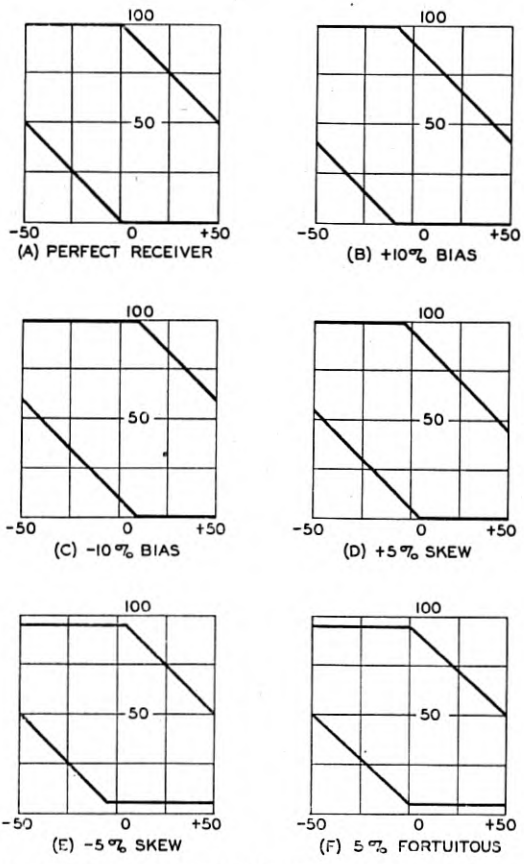


Fig. 5—Effect of internal distortion on bias parallelogram.

receiver and illustrates how the components of internal distortion affect the shape of the bias parallelogram. The skewing of the corners of the parallelograms shown in Fig 5(D) and (C) led to the use of the term "skew" for this effect.

In telegraph transmission systems skew may be caused by the effect of interference on a wave which has different slopes during mark-to-space and

space-to-mark transitions. It may also result from an equivalent electromechanical effect in a start-stop receiver, as will be described later.

Measurements of Receiver Distortion Tolerance

In measurements of the distortion tolerance of start-stop receivers there is used a distributor which is arranged to transmit signals having any of the four types of displacement *MB*, *SB*, *SE* and *ME*. Positively biased signals are transmitted for *MB* displacement and negatively biased signals for *SB* displacement. The test signals having *SE* or *ME* displacement are said to be affected by "end distortion." These differ from any experienced on transmission circuits in that only the mark-to-space transitions of the selective pulses are shifted relative to the start transition, being delayed for *ME* displacement and advanced for *SE* displacement, as shown in Fig. 2(E) and (F). "End distortion" simulates the mark-to-space displacements produced by characteristic and fortuitous distortion, and it has been found in practice that it yields results which enable a receiver's tolerance to these components of distortion to be predicted with a high degree of accuracy.

When fixed values of displacement are transmitted, the limits of orientation are measured by means of the range scale of the receiver. Alternately, a distributor may be used in which the magnitude of displacement may be continuously varied, and this enables measurements of internal distortion to be conducted with the orientation fixed, or, indeed, on receivers having no means or a limited means of varying the orientation.

Orientation Settings for Best Tolerance to Test Distortions

Obviously, the best orientation setting is that which permits the receiver to tolerate the greatest amount of any distortion which is expected. If all four types of displacement are considered equally likely, the orientation should be set at that point at which the minimum tolerance to any type of displacement is as large as possible. For example, consider a receiver which, with an orientation setting of 49, has the following tolerances to test displacements:

MB	44
SB	38
SE	42
ME	44

Let the orientation setting be raised 2 per cent, to 51. Then the tolerances are as follows:

MB	42
SB	40
SE	40
ME	46

The shift of orientation has increased the minimum tolerance (spacing bias) from 38 to 40. Any further shift would make the tolerance to spacing "end distortion" less than the tolerance to spacing bias. This setting is called the "center of fortuitous distortion tolerance," since at this point the receiver will tolerate the maximum amount of fortuitous distortion.

If, on the other hand, bias is considered more probable than distortions which produce "end distortion" effects, the orientation might be adjusted to the point at which the tolerances to marking and spacing bias are equal. For example, suppose the orientation setting of the receiver under consideration were raised 1 per cent to 52. The tolerances would then be

MB	41
SB	41
SE	39
ME	47

This setting is called the "center of bias tolerance," since at this point the receiver will tolerate the maximum amount of bias regardless of the sign of the bias.

There is one more setting that is of interest. It is that at which the tolerances to marking and spacing "end distortion" are equal. Suppose the orientation of the receiver were lowered 4 per cent to 48. The tolerances would then be

MB	45
SB	37
SE	43
ME	43

This setting is called the "center of end distortion tolerance," since at this point the receiver will tolerate the maximum amount of "end distortion" regardless of its sign.

Calculation of Components of Internal Distortion

Figure 6 illustrates how the components of internal distortion are determined from measurements of distorted signals. Each diagram shows a portion of a teletypewriter character consisting of a start pulse, a marking selective pulse and a spacing selective pulse. The solid lines show an undistorted signal. The dashed lines show the displacement of a transition due to internal bias. The shaded area defines the fortuitous effect which is skew; that is, the transition in question may fall anywhere within the shaded area during repeated transmission of the signal. The arrows below the figure show the extent of the displacement occurring on each transition due to the presence of a given displacement of the transmitted signals. The four types of displacement are of equal magnitude D. The arrows above the diagram designated L_B and L_E show the lower limits of orientation with, respectively,

spacing bias and marking "end distortion" (SB and ME displacements). The arrow U_B and U_E show the upper limits or orientation with, respectively, marking bias and spacing "end distortion" (MB and SE displacements).

Figure 6(A) shows the case of positive internal bias and positive skew; Fig. 6(B), positive bias and negative skew; Fig. 6(C), negative bias and positive skew; and Fig. 6(D), negative bias and negative skew. The following relationships hold, bearing in mind that $MB = SB = SE = ME = D$:

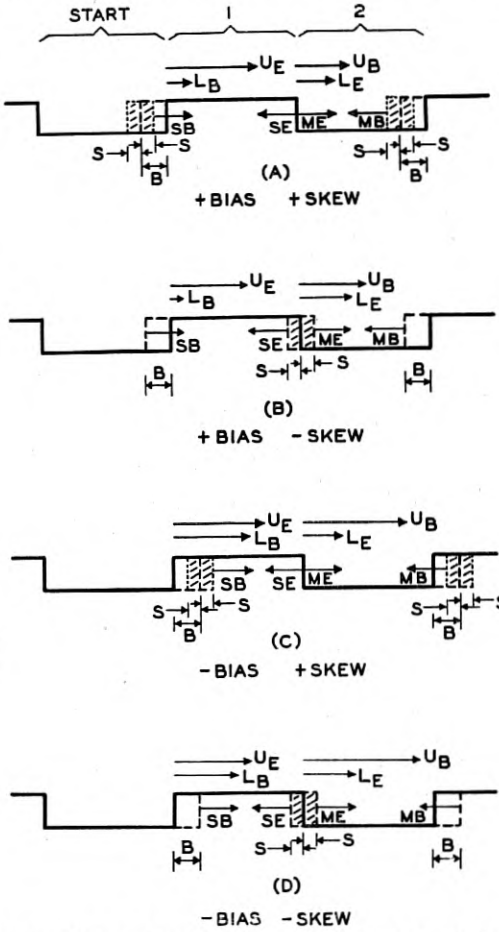


Fig. 6—Use of distorted test signals in measuring internal distortion.

Fig.	*Bias*	*Skew*	L_B	U_B	L_E	U_E
(A)	+	+	$D+s-b$	$1-b-s-D$	D	$1-D$
(B)	+	−	$D-b$	$1-b-D$	$D+(-s)$	$1-(-s)-D$
(C)	−	+	$D+s+(-b)$	$1+(-b)-s-D$	D	$1-D$
(D)	−	−	$D+(-b)$	$1+(-b)-D$	$D+(-s)$	$1-(-s)-D$

In any figure $L_B - L_E = \quad s - b$
and $\qquad U_B - U_E = -s - b$
Adding and subtracting, we find that:

$$\text{Internal bias} = \frac{U_E + L_E}{2} - \frac{U_B + L_B}{2}$$

$$\text{Skew} = \frac{U_E - L_E}{2} - \frac{U_B - L_B}{2}$$

Any $\frac{U+L}{2}$ is the center of an orientation range. Hence it may be stated that the internal bias is equal to the difference between the *centers* of tolerance to "end distortion" and bias. It will also be noted that any $\frac{U-L}{2}$ is half of an orientation range. When the test signal displacements determining the range limits are equal, the amount of tolerance equals $\frac{U-L}{2} + D$ (assuming no curvature in the distortion parallelogram). Hence the skew is equal to the difference between the *amounts* of tolerance to "end distortion" and bias.

For example, the receiver cited previously has the following characteristics

$$\text{Internal bias} = 48 - 52 = -4\%$$
$$\text{Skew} = 43 - 41 = +2\%$$

Incidentally, this means that internal bias does not reduce the total bias tolerance of a receiver, but merely shifts the center of bias tolerance with relation to the center of "end distortion" tolerance. Hence the effects of internal bias may be compensated for, as far as the bias tolerance of the receiver is concerned, by setting the orientation at the center of bias tolerance. However, internal bias does reduce the minimum "end distortion" tolerance of a receiver whose orientation is adjusted to the center of bias tolerance.

"Switched" Bias

When biased signals are produced by the action of a biasing current on a relay driven by a symmetrical wave, and the sign of bias is suddenly reversed during the transmission of a teletypewriter character, all the succeeding transitions of that character are affected, not by bias, but by "end distortion." This is shown in Fig. 7, of which (A) shows the original unbiased signals, (B) shows the signals affected by bias which changes from positive to negative at time T, and (C) shows the effect on the same signals when the bias is changed from negative to positive.

Signals such as these, in which the sign of bias is changed at intervals, are said to be affected by "switched bias." Since all four types of displace-

ment are present in equal magnitude in switched bias signals, the effect on a start-stop receiver resembles that of fortuitous distortion. Thus the center of switched bias tolerance is the center of fortuitous distortion tolerance and the amount of switched bias tolerance is the amount of fortuitous distortion tolerance. This center is also the center of orientation in a receiver having no curvature or symmetrical curvature of the displacement-vs.-orientation-limit characteristic. The switched bias tolerance is, of course, one-half the orientation range in a receiver having no curvature of the characteristic.

In actual field practice, switched bias signals, applied at a central office, are used as a test of tolerance of the teletypewriter at a subscriber station in combination with the subscriber loop. They provide a more accurate measure of transmission capabilities than an orientation range measurement with undistorted signals from the central office, since not only is the curvature of the distortion parallelogram taken into account, but the character

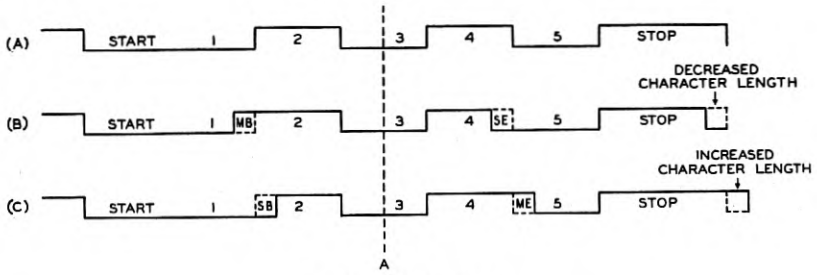


Fig. 7—Switched bias.

length changes in much the same manner as in signals affected with characteristic or fortuitous distortion.

The components of internal distortion of a receiver may be estimated from bias and switched bias measurements, but they cannot be accurately specified thereby. Figure 8 illustrates the difficulty in separating bias and skew by means of measurements of the difference between the amounts and centers of tolerance to steady bias and switched bias. Figure 8(A) shows the bias and end displacement parallelograms of a receiver having +24 per cent bias and +16 per cent skew. The center of tolerance to switched bias is 4 per cent above the center of steady bias tolerance and the steady bias tolerance is 4 per cent greater than the switched bias tolerance. Figure 8(B) shows the parallelograms of a receiver having +4 per cent bias and −4 per cent skew. Again, the center of tolerance to switched bias is 4 per cent above the center of steady bias tolerance and the steady bias tolerance is 4 per cent greater than the switched bias tolerance.

Of course, the components of internal distortion can be measured by

observing both ends of the orientation range with positive and negative bias rather than observing the upper end with positive bias and the lower end with negative bias. This type of measurement is merely equivalent to using a fairly large percentage of bias and zero per cent of end distortion. The disadvantage of this measurement is that no account is taken of the curvature

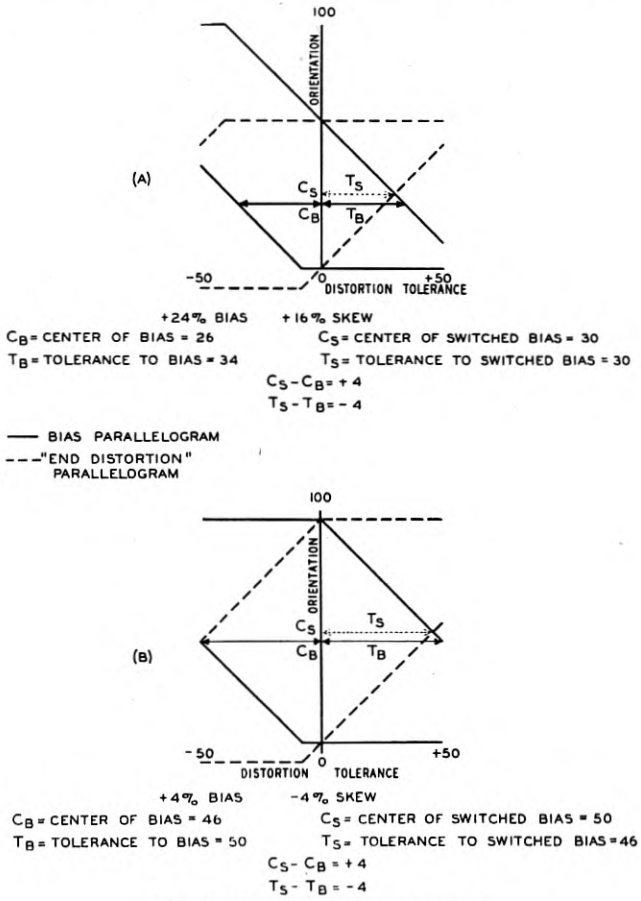


Fig. 8—Switched bias measurements

of the end displacement parallelogram, and hence the indicated values of tolerance may not be an accurate measure of the receiver's ability to receive distorted signals.

Internal Fortuitous Distortion

It is usually considered, in measurements of miscellaneous signals, that the difference between the maximum distortion tolerance and 50 per cent (the

latter being the tolerance of a perfect receiver) is due to internal fortuitous effects, even though part of it may be due to the effects of internal characteristic distortion. Hence the internal fortuitous distortion is usually defined as the difference between 50 and the tolerance to bias or end distortion, whichever of the latter may be the larger.

For example, in the sample receiver considered on page 220, the internal fortuitous distortion is:

$$50 - 43 = 7 \text{ per cent}$$

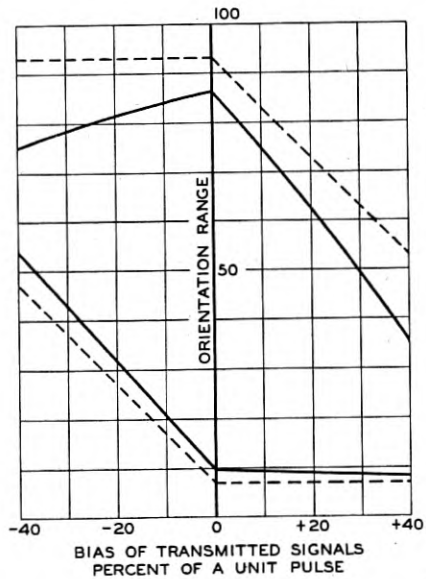


Fig. 9—Effect of negative characteristic distortion on bias parallelogram.

Internal Characteristic Distortion

In practice it is found that the relation between displacement and reduction of margin is sometimes not strictly linear. Especially at large values of displacement, the reduction in margin is often greater than the displacement causing it. This effect is due to internal characteristic distortion, which causes an increase in the distortion of shortened pulses. Internal characteristic distortion, like any other form of characteristic distortion, is caused by the failure of some circuit or mechanical element to attain steady state before the occurrence of a succeeding transition. Figure 9 shows an example

of the bias parallelogram of a receiver suffering from internal negative characteristic distortion.

Some Considerations Involved in the Measurement and Adjustment of Start-Stop Receivers

Because of the effects of characteristic distortion, it cannot be assumed that the ultimate tolerance of a receiver is equal to the sum of the displacement of the received test signals and one-half the remaining orientation range, especially if the latter is large. To attain accurate results, the ultimate tolerance must be measured with the orientation adjusted to the center of tolerance.

For the same reason (the curvature of the "parallelogram" caused by internal characteristic distortion) measurements of internal distortion on a receiver which is, itself, to be used to measure distortion should be made with displacements of approximately the same magnitude as the distortions which the receiver is to measure. In a receiver which is to be used to measure small distortions, we are interested in the properties of the linear portion of the parallelograms. Hence we measure the receiver's internal bias and skew using small amounts of displacement in the measuring signals. The internal fortuitous distortion may generally be neglected, since it does not affect the shape, but only the size, of the distortion-vs-margin characteristic.

On the other hand, in a receiver which is to be used for receiving signals we are interested not so much in the shape of the characteristic as in the ultimate tolerance to telegraph distortion at an optimum setting of the orientation mechanism. For this reason, a receiver destined for service use is best tested with signals containing fairly large displacements. Internal fortuitous distortion is deleterious in such a receiver, since it decreases the tolerance to displacement of all kinds. Skew, depending upon its sign, affects the tolerance to either space-to-mark or mark-to-space displacements.

It should be realized that the removal of skew does not necessarily improve a service receiver. In the case of bias or characteristic distortion the introduction of distortion of a given sign will remove internal distortion of the opposite sign, and thus improve the performace of the receiver. But since skew is the difference between two fortuitous distortion effects, it may be removed either by reducing the larger or increasing the smaller effect. The former procedure will increase the receiver's total tolerance to distortion, whereas the latter will reduce it.

In practice bias tolerance is generally considered to be more desirable than "end distortion" tolerance. The reason for this is that most transmission circuits suffer from some bias (of unpredictable sign and amount) which uses up some of the receiver's bias tolerance but none of its "end distortion"

tolerance. This is why the orientation of a service receiver is generally adjusted to the center of bias tolerance, and small amounts of internal bias or negative skew are not considered objectionable, since they do not affect the tolerance to bias at the center of bias tolerance. By the same token, the presence of positive skew, which indicates a lowered bias tolerance, usually calls for a readjustment of the receiver to reduce the fortuitous effect on the space-to-mark transitions. As explained above, removing the skew by introducing a fortuitous effect on the mark-to-space transitions will not, of course, improve the bias tolerance.

It is the present practice in the field to specify a minimum bias tolerance about 5 per cent greater than the minimum permissible "end distortion" tolerance, the orientation being adjusted to the center of bias tolerance for both measurements.

Some Causes of Internal Distortion

Up to this point internal distortion has been considered without regard to its probable causes. The more obvious causes will be found to be analogous to those which produce equivalent distortions in telegraph transmission circuits.

Bias will result when an element (whether electrical, mechanical, or electronic) of a receiver possesses dissymmetry toward marking or spacing. For example, a mechanical element may travel more slowly from spacing to marking than from marking to spacing and thus cause spacing bias, or its range of travel may be divided unequally into marking and spacing portions, thus producing an equivalent effect.

Characteristic distortion will result when an element (whether electrical or mechanical) of a receiver fails to attain a steady state before being acted upon by a succeeding transition, or otherwise depends, in its action, upon the previous history of the signal train. An example of characteristic distortion is found in the 20-milliampere holding magnet selector when it is equipped with a resistive shunt. In this type of selector the armature is actuated by a cam, which presents it to the pole-face at about the middle of each pulse, and then disengages it. The armature is then free to release or remain operated, according as the received pulse is spacing or marking. The shunt that is normally used presents so low an impedance to the magnet winding that the motional impedance effect which is produced by the sudden mechanical presentation of the armature to the pole-faces causes a sizeable reduction in the magnet current. In the case of a short marking pulse, the current fails to attain steady state before the next mark-to-space transition occurs. The magnet therefore releases sooner than it does at the end of a long marking pulse, during which the current has had time to attain steady state. It will be seen that this is really a characteristic distortion effect, since it is due to a failure to reach steady state and depends upon the previous history

of the signal train. However, when miscellaneous signals are being received the effect appears similar to a fortuitous distortion occurring on mark-to-space selective transitions, and hence it is usually thought of as negative skew.

Fortuitous distortion will result when an element is irregular in its action, and if such action is more irregular on one type of transition than on the other, the result will appear as skew. For example, irregular action of the receiving clutch affects the selector alike in regard to all selective transitions, and appears as internal fortuitous distortion. Another source of internal fortuitous distortion is the period of indecision that occurs during the passage of a selective element past a locking member, at which time the choice between marking and spacing is largely fortuitous.

A common cause of skew in teletypewriters may occur in the following manner: If the armature stops are so adjusted that, for example, the armature travel is greater on the marking side than on the spacing side of the armature lock, positive internal bias results. If, now, this bias is compensated for by so adjusting the armature air-gap and retractive spring tension as to cause the receiving magnet to operate in a negatively biased manner (rather than by correcting the improper armature travel), the armature will be forced to operate in a region of the operating wave that is more sloping than the region in which it releases. Hence, it will operate more irregularly than it releases, and thus will be affected by positive skew.

Selector Action

Over and above the sources of internal distortion which are analogous in effect to sources of distortion encountered in telegraph transmission circuits, there is another whose action in causing internal distortion is not so obvious as those just described. This source of internal distortion may be termed "*selector action*," and it depends upon the relation between the operating time of a selector element and the period of time allowed for said element to act. For the purpose of explaining the effect of time relations within the selector on internal distortion, selector mechanisms may be classified as of three basic types: M, S, and P.

In a mechanism of type M each selector is initially in the spacing condition and either remains spacing or operates to marking when subjected to the action of the corresponding received signal element. When it attains the marking condition it becomes locked for the duration of the character. Early types of start-stop printers having an individual selector magnet for each pulse of the code and employing a separate receiving distributor,[2] are illustrative of type M.

In a device of type S each selector is initially in the marking condition and either remains marking or operates to spacing when subjected to the action of the corresponding received signal element. When it attains the

spacing condition it becomes locked and cannot again operate to marking during that character. The Siemens-Halske five-selector teleprinter[5] is an example of this type.

In a mechanism of type *P*, the selector may be in either the marking or spacing condition initially, according to the type of the previous signal element to which it has responded. When subjected to the action of a received pulse the selector may go in either direction, and it remains responsive to the action of the signal during the entire selecting interval. The No. 14 and No. 15 teletypewriters[2] (not equipped with holding magnet selector) of the Teletype Corporation are examples of type *P*.

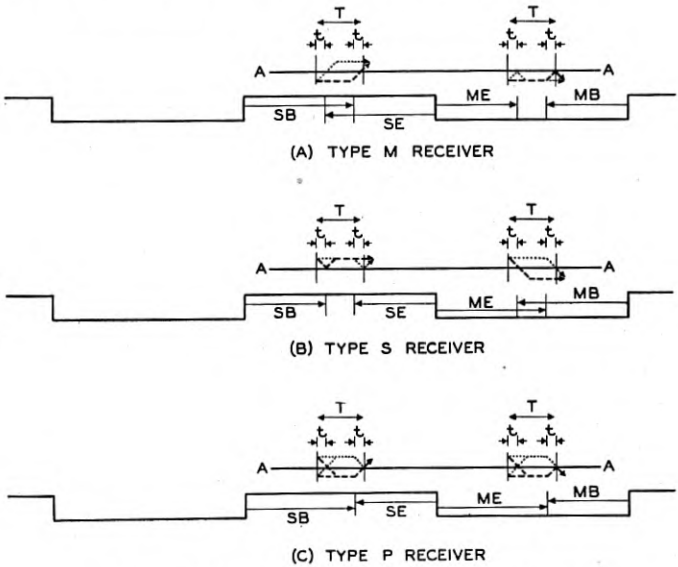


Fig. 10—Effect of selector action on internal distortion.

Figure 10 (A) illustrates the action of a type *M* selector. A portion of a teletypewriter character is shown, consisting of the spacing start pulse, a marking first selective pulse and a spacing second selective pulse. The undistorted signal is shown in solid lines. The maximum amounts of marking and spacing bias that the receiver will tolerate are shown by dashed lines and are designated *MB* and *SB*. The limiting amounts of marking and spacing end displacement are shown by dotted lines and are designated *ME* and *SE*. Above the signal train is shown a schematic representation of the action of the selective system. The periods of time *T* are those during which the selector is subject to the action of the received signal, and *t* is the time that the selector must be subjected to the operative force in order that it

operate. The line *A-A* indicates the boundary between the marking and spacing positions of the selectors. In this type of receiver, as mentioned previously, when the selector crosses to the marking or upper side of line *A-A* it becomes locked and cannot again go to spacing even though the signal should subsequently become spacing during the selective period T.

It will be noted that the limits of end displacement tolerance occur at time t after the beginning of the selective period. This instant is sometimes called the "instant of decision for end displacement." On the other hand, the limiting tolerances to bias are determined at a time t before the end of the selective period, sometimes known as the "instant of decision for bias." If the selective periods were advanced relative to the start transition by lowering the orientation until the bias tolerances were equal, the instants of decision for bias would correspond with the center of bias tolerance. If, then, the selective periods were delayed, by raising the orientation, by an amount $T - 2t$, the instants of decision for end displacement would correspond with the center of end displacement tolerance. Since the difference between the center of end displacement tolerance and the center of bias tolerance is equal to the internal bias of the receiver, it will be obvious that the internal bias is also equal to the difference between the instant of decision for bias and the instant of decision for end displacement. In this type of receiver the internal bias is $T - 2t$, and will be positive, zero, or negative according as $2t$ is less than, equal to, or greater than T.

Figure 10 (B) shows the action of a type S selector. Here the instant of decision for bias occurs at time t before the end of the selective period and that for end displacement at time t after the beginning of the selective period. Hence, the internal bias is equal to $2t - T$.

The action of a type P receiver is illustrated in Fig. 10 (C). It is assumed in this figure that the selector operates toward marking at the same rate as toward spacing, since the effect of unequal rates of operation has been described previously. In a selector of this type, both instants of decision occur at time t before the end of the selective period and hence the internal bias is not dependent upon the relation between T and t. If, however, t is so long that the selector cannot pass from one extreme of travel to the other, attain a steady state, and return to the center position within time T, a sort of characteristic distortion occurs, in which the instant of decision depends upon whether the selector began the selective period in the same or the opposite condition from that finally selected. In measurements of miscellaneous signals this appears similar to a fortuitous effect, since it decreases all tolerances equally. Hence it is usually considered as internal fortuitous distortion.

Receivers equipped with holding magnet selectors are of Type S, since the armature may be released, but not operated, by the magnet. In this

type of mechanism, the armature generally drives a subsidiary selective member, and the time T extends from the instant at which the armature is disengaged by its operating cam until the instant when the subsidiary selector becomes locked. As this period is often long in relation to the magnet releasing time t_1 and the subsidiary selector operating time t_2, holding magnet selectors are often subject to negative internal bias. In those mechanisms in which the subsidiary selector is flexibly coupled to the magnet armature, the former's operation is of type P. It, therefore, may be subject to the characteristic distortion effect noted in the description of type P operation, except that the effect, when it occurs in this type of mechanism, affects only the instant of decision for end displacement and hence resembles negative skew rather than internal fortuitous distortion.

An interesting, but somewhat unusual, effect occurs in any receiver, of whatever type, in which the lengths of selective period or selector operate time, or both, differ for the various selective pulses, or in which the spacing of the selective periods is improper. In a case of this sort, the receiver exhibits an internal bias equal to the difference between the average instant of decision for bias and the average instant of decision for end displacement, an internal fortuitous distortion equal to the variation of the instant of decision having the smaller variation, and a skew equal to the difference between the variations of the instant of decision for bias and the instant of decision for end displacement.

Conclusions

A working knowledge of the effect of telegraph distortion on the margins of operation of start-stop receivers is essential in dealing with a plant in which the use of teletypewriters, regenerative repeaters and start-stop distortion measuring sets is as widespread as it is in the Bell System. When a major portion of the communication system operates on a start-stop basis, it is desirable that transmission measurements be made on the same basis.

The knowledge of this subject that has been gained in recent years has made possible many improvements in technique both in the field and in the laboratory, and these have led to corresponding improvements in the mechanisms used in telegraph service. The analysis of new start-stop devices may now be carried out efficiently and accurately, and this often permits the formulation of suggestions leading to improved operation of the devices.

The general level of service excellence has been raised by the setting up of criteria for the distortion tolerances of station teletypewriters, regenerative repeaters and other start-stop devices used in service, including those provided for switching. The sources of distorted test signals that are now available are useful not only in measuring the tolerances of service receivers,

but also in determining the characteristics, and hence the accuracy, of start-stop distortion measuring sets and monitoring teletypewriters.

Finally, there has resulted an improved ability to analyze and predict the performance of transmission links from the results of distortion measurements made on a start-stop basis.

References

1. "Measurement of Telegraph Transmission," H. Nyquist, R. B. Shanck, S. I. Cory, *Jour., A. I. E. E.*, March 1927, p. 231.
2. "Fundamentals of Teletypewriters Used in the Bell System," E. F. Watson, *Bell Sys. Tech. Jour.*, October 1938, p. 620.
3. "A Transmission System for Teletypewriter Exchange Service," R. E. Pierce and E. W. Bemis, *Bell Sys. Tech. Jour.*, October 1936, p. 529.
4. "Recent Developments in the Measurement of Telegraph Transmission," R. B. Shanck, F. A. Cowan, S. I. Cory, *Bell Sys. Tech. Jour.*, January 1939, p. 143.
5. "Der Spielraum des Siemens—Springschreibers," M. J. de Vries, *Telegraphen-und Fernsprech-Technik*, January 1934, p. 7.

CHAPTER XIII*

The Mounting and Fabrication of Plated Quartz Crystal Units

By R. M. C. GREENIDGE

13.1 Introduction

THIS paper is one of a series on piezoelectric quartz plates and deals primarily with the methods employed in mounting crystal plates operating up to approximately one megacycle for practical utilization in communication equipment. The theoretical aspects of mounting crystals have been covered in Chapter VII. The discussion is confined to plates[1] having definite nodal lines or points, such as $+5°$ and $-18°\ 25'$ X cuts, GT, CT, DT, MT and NT cuts. The mounting of high-frequency crystal plates such as AT and BT cuts, which vibrate in thickness shear modes, is not included. It should also be noted that the subject matter is treated descriptively and that no attempt is made to go into the more intricate details of design or to give performance characteristics. These matters will be dealt with fully in a later paper. The designs and methods outlined are up to date for each type of unit, the results of many years of development on the part of Bell System engineers to evolve practical designs for commercial manufacture and use. Expanding on the contributions of the early investigators mentioned by W. P. Mason in Chapter I,[1] these engineers had, in the ten years prior to 1939, worked out practical designs and developed suitable tools and processes for wide commercialization in telephone applications. In the last five years, under the impetus of war, further improvements have been made in the design and manufacture of crystal units, particularly those for use by the Armed Forces.

The term "Crystal Unit", originally adopted by the Bell System to designate the complete assembly of a crystal plate in its mounting and case, has now been standardized quite generally in the art, replacing a variety of names by which these devices were formerly called. The basic design features of a crystal unit involve the use of:

1. Electrodes, on or near to the crystal surfaces for impressing voltage across the plate,

* Chapters IX, X, XI and XII, which will be included in a forthcoming volume are omitted from the *Technical Journal* because they deal largely with details of manufacturing operations.

[1] "Quartz Crystal Applications", W. P. Mason, *B.S.T.J.*, Vol. XXII, Page 191, July 1943.

2. Supports for holding the crystal plate in its mount, and
3. A sealed outer case having the necessary terminals, and provisions for incorporating the unit electrically and mechanically into the apparatus.

Two distinct types of crystal units have been evolved, one embodying the use of pressure pins or anvils for supporting and holding the crystal plate and the other involving the suspension of the crystal plate by means of fine wires.[2] These designs are known, respectively, as the Pressure Type and Wire Supported Type, and will be discussed later under these headings. However, there are several details of fabrication common to both types which can best be discussed at this point.

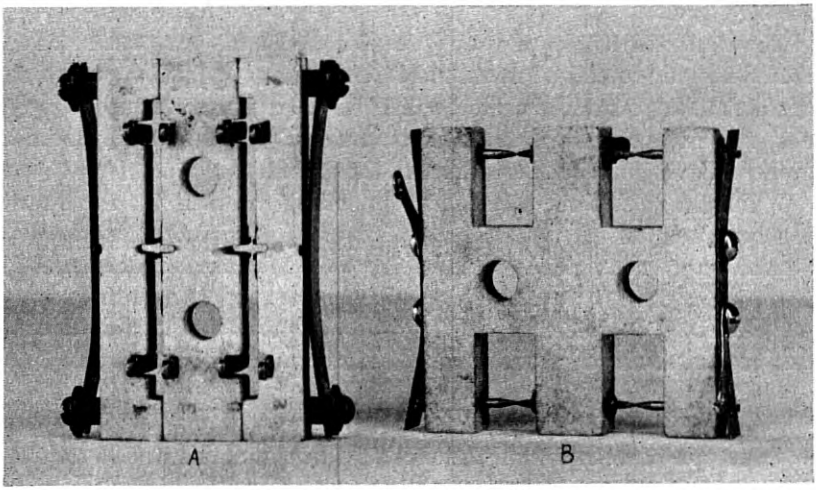


Fig. 13.1—Pressure-type holders.

Irrespective of the type of mounting, realization of the desired performance in a crystal unit depends to a considerable extent on the processing of the quartz plate itself. Previous articles[1, 3] have brought out the significance of such factors as the precision of angular orientation and linear dimensions on the fundamental characteristics of the plate. The plate must also be virtually free of impurities or imperfections.[4] In the preparation of a quartz plate it must be lapped using increasingly finer abrasive materials until the final dimensions are reached. Depending upon the type of crystal unit, No. 400, No. 600 carborundum or finer abrasives are now

[2] A. W. Ziegler, Patent 2,275,122, March 3, 1942.

[3] "The Use of X-Rays for Determining the Orientation of Quartz Crystals", W. L. Bond and E. J. Armstrong, *B.S.T.J.*, Vol. XXII, Oct. 1943.

[4] "Raw Quartz, Its Imperfections and Inspection," G. W. Willard, *B.S.T.J.*, Vol. XXII, Oct. 1943.

employed for the final stage grinding. Following this, the plate is thoroughly cleaned by acid treatments or by the use of solvents and detergents followed by copious washing. It is then etched in commercial hydrofluoric acid to remove all loose particles of quartz that might have remained on the surfaces or in the crevices after cleaning. Etching also smooths off the roughness of the ground surfaces. The effect of this treatment reduces energy dissipation in the plate itself and increases by many times the efficiency of the crystal units. Etching also improves the stability of performance of the crystal unit. Standard designs of crystal units require etching of the plates, uniformly on all surfaces, for a period of thirty to forty minutes. For units of highest precision and efficiency longer etching periods are employed.

The electrodes employed with types of crystal units being described consist of metallic coatings, generally aluminum, silver, or gold deposited over the major surfaces of the crystal plate. These coatings are applied by the evaporation process which results in an extremely thin and uniform coating of metal having excellent adherence to the quartz.

With reference to the mechanical supporting members for the plate, it has been brought out that such supports should be confined as closely as possible to the nodal points or nodal lines where the motion for all practical purposes is zero. It is common practice for the supporting members to be made of metal so that they will serve also as a means of making electrical connections to the electrode coatings on the surfaces of the plates.

13.2 Pressure Type Crystal Units

This type of crystal unit was initially developed for use in telephone filters. Up to about five years ago it was employed in virtually all commercial designs of filters. Depending upon the mode of vibration and size of the crystal plate the design of the mounting varies. However the principles employed for clamping are essentially the same in all cases. Where small longitudinal or face shear plates are involved one pair of pressure pins is used unless two are required for electrical reasons as explained later. For medium size plates of the same type or for face flexure plates, two pairs of pins are employed. In the case of large low-frequency longitudinal plates double anvils are used instead of pins in order to obtain firmer clamping of the plate to prevent translation or rotational movement which might cause wear in the electrode surface at the pressure point with resultant variations in frequency and resistance. The blocks are usually composed of molded steatite and the springs for exerting the necessary pressure are of phosphor-bronze.

Figure 13.1 (B) shows a pressure mounting for holding four crystals which have single coatings on each of their major surfaces. The main require-

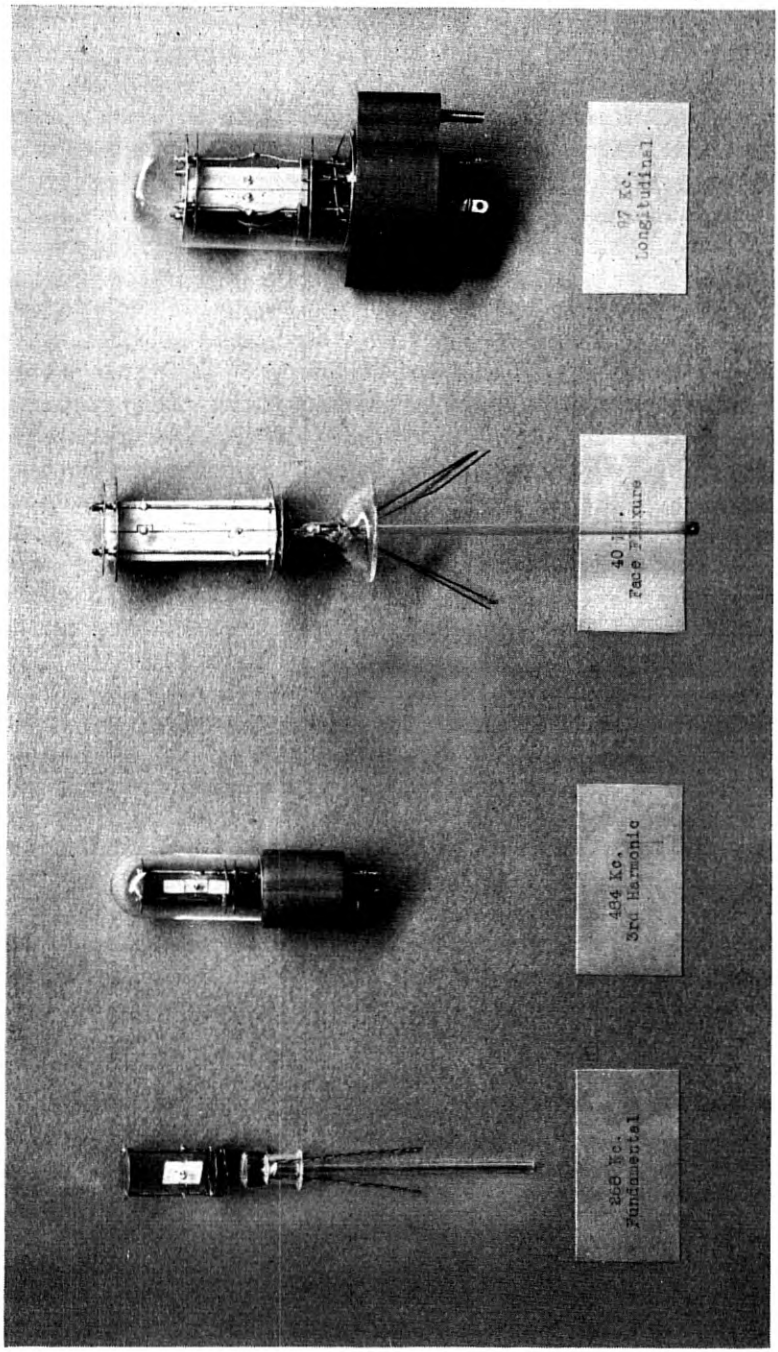


Fig. 13.2—Wire-supported crystal units.

ments which must be met for such a mounting are small areas, accurate alignment, and adequate pressure to hold the plate in place. Some designers make slight indentations in the quartz at the point of contact to improve the mechanical stability of the plate. For crystal plates of the order of one-half inch square or smaller, points having an area of about 10 mils in diameter are employed and the pressures used for holding the plate range from one to two pounds. For larger plates correspondingly larger areas of points and increased pressures are employed. The accuracy of alignment required for ten mil points is of the order of two or three mils. This is obtained in the mountings shown in Figure 13.1 by using concentric sleeves for holding the points which are brought into alignment by means of a straight rod and then cemented in place. One of the points is fixed while the other point slides in its sleeve and the pressure required is obtained by the spring which presses on the outer end of the sliding point. In balanced filter structures it is desirable to use crystal plates with the coating on each side divided into two equal areas. This reduces to one-half the number of plates that would otherwise be required. In mounting plates with divided coating, it is necessary to provide a mounting which makes double contact on each side of the plate. Figure 13.1 (A) shows a pressure type mounting which accomplishes this. This is the mounting which has been used for several years in holding the plates used for the 75-type crystal channel filters[5] for the standard terminal common to all broad-band telephone systems. The crystal is mounted in the holder in such a way that the two pairs of points clamp the crystal along the nodal line. The rectangular dimensions of the points used for this type mounting for crystals operating in the frequency range from 60 kc up to 120 kc are about 35 mils long in the direction of the nodal line and from 10 to 15 mils wide. A very important requirement for such a mounting is that the flat area of the points on each side of the plate fall in the same plane. This is accomplished by a precise milling operation after the points are assembled in the mounting. The pressure applied to the pair of points is furnished by the flat spring shown and is equalized by the action of the roller centrally located under the springs. The pressure employed is of the order of four to five pounds for each pair of points.

The most commonly used coating for crystal plates held in pressure-type mountings is aluminum.[6] Aluminum has been found to be most satisfactory for this type of unit because its hard surface is more resistant to wear at the points of clamping than other metals such as silver, or gold.

Except for a few designs, which are mounted in sealed metal or glass

[5] "Crystal Channel Filters for Carrier Cable Systems," C. E. Lane, *B.S.T.J.*, Vol. XVII, Page 125.

[6] The details of processing aluminum-coated crystals are similar to those described for silver-coated crystals in paragraphs 13.42 and 13.43.

containers, pressure-type crystal units are not sealed in individual containers. However, the entire filter in which they are employed is dried and sealed off after filling with dry air.

In pressure-type units of this type, variations in frequency of the order of .01% may be expected if crystals are transferred from one mounting to another or relocated in the same mounting. Consequently, if high-frequency precision is desired, it is necessary to make the final frequency adjustment with the crystal located in its final position. Due to the inherent difficulties of adjustment, coupled with the close manufacturing tolerances on parts, and precision adjustments necessary for the holders during assembly, these designs have been virtually discarded in favor of wire supported designs. However, more recent developments by J. F. Barry on pressure-mounted-type units have brought forth some new ideas which might prove in for wider future application.

13.3 Wire Supported Crystal Units

This type of mounting is being used extensively on crystal applications and has superseded the earlier pressure-type units. Various designs of this type of crystal are shown on Fig. 13.2. The wire-mounted crystal possesses the definite advantage in that after the supporting wires are attached to the plate, they remain fixed in position throughout the subsequent manufacturing process thus facilitating adjustment of the frequency or the frequency-temperature characteristic. Moreover, the supporting wires can be formed so as to provide a spring mounting for the crystal plate which protects it from any shocks or vibration it may encounter in shipment or use. In the wire-supported design the suspension wire is also employed as a means of connecting the electrical circuit to the electrode plating on the crystal. By virtue of the solder bond between the wire and the electrode, this type of unit is free from the possibility of instability in frequency performance due to slight changes in position and variations in contact resistance prevalent in pressure type designs, and for this reason the stability and frequency precision of wire supported crystals is superior.

From a manufacturing standpoint the wire-supported unit involves a greater number of processing operations than the pressure-type unit, but the adjustment operations are considerably easier. Moreover, it is possible to realize greater precision of frequency adjustment by a factor of at least two or even three. The crystals are also more uniform in their effective resistance. Since the mount or cage in which the crystal is suspended is comparatively inexpensive, there should in general be little difference in the manufacturing costs of the two types. Consequently, in the wire-supported crystal units a very appreciable improvement should be gained in performance without increasing the cost of the unit.

13.4 Fabrication of Wire Supported Unit

13.41 Silver Spotting

13.411 Application of Silver Paste

Starting with the crystal plate, the first step in manufacture is to apply silver spots to the surfaces of the plate. These spots serve as footings to which the supporting wires are ultimately soldered or sweated. They are placed on the nodal points or along the nodal lines of the plate in order to detract as little as possible from the intrinsic characteristics of the plate itself. The areas of the silver spots cover the range from about 40 to 90 mils in diameter depending upon the amount of solder to be used in attaching the wire to the plate. Before spotting the plates it is essential that they be free from any contamination such as grease or organic material that might affect the fusion of the silver spots into the surface of the quartz. One of the best methods to ensure cleanliness is to boil the plates in aqua regia, followed by copious rinsing in water. Detergents such as sodium meta-silicate are also employed followed by a rinsing. The plates may finally be boiled in distilled water and carefully dried. Throughout the subsequent processes the plates should be handled with clean tweezers or gloved fingers and kept away from any source of contamination. Prefiring of the plates at 950°F prior to spotting has also been used as a positive way of ensuring freedom from any contamination that would affect the fusion of the spots, but this process is not necessary if the first mentioned process is properly controlled.

In spotting, small quantities of a prepared silver paste are placed on the areas of the plate to which the wires will ultimately be attached. The paste consists of a compound of finely divided silver and low melting point glass (lead borate) thoroughly mixed with a suitable vehicle to facilitate application. For spotting purposes it has been found that a paste having a specific gravity of between 2.3 and 2.6 gives best results. In use, the materials must be constantly agitated or stirred in order to prevent the solid ingredients from settling out. This is important, for, unless the concentration of silver is maintained around 90 to 95 per cent of the solid matter, it will not be possible to obtain good wetting of the solder in making the wire attachment.

The placement of the semi-liquid material on the plate is accomplished by means of a small stylus, the crystal plate being held in a clamp or vise and the stylus guided so as to place the material at the exact location on the plate as desired. A typical tool for doing this work is shown in Figure 13.3. The point of the stylus should have a slightly rounded end. With the rounded point the tendency of the paste to spread out is minimized and consequently the diameter of the spot is substantially the same as that of

the stylus. The rounded stylus also results in a more uniform distribution of the material. The material is applied to the end of the stylus by spreading a small amount of the paste on a glass plate from which it is transferred to the stylus and then deposited on the crystal. The material on the glass plate should be wiped off and replaced quite often due to settling and drying out of the mixture, in order to insure uniformly good spots. Generally speaking, anywhere from two to six spots at the most should be possible from one loading of the transfer plate depending upon the speed of the operator.

Regarding the character of the crystal surface, aside from cleanliness, and its effect on the ultimate strength of adhesion of the silver spot, experience

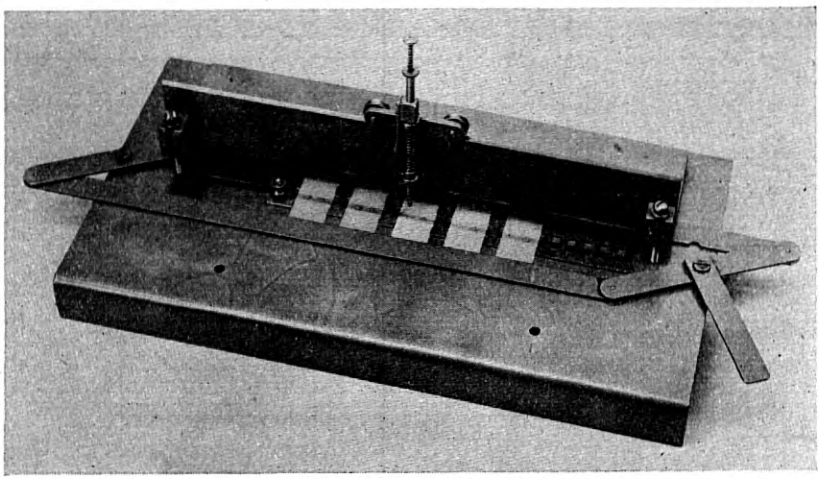


Fig. 13.3—Tool for applying silver spots.

so far with all the various cuts of plates does not indicate that this is a factor. Spots have been found to adhere to polished surfaces as well as ground and etched surfaces with about the same degree of strength.

13.412 Firing of Silver Spots

Following the application of the paste to the plates it is desirable to predry the spots in order to remove the low volatile constituents of the vehicle prior to firing at high temperature. This may be done in a ventilated oven or over a hot plate at approximately 300°F for about 15 minutes. The plates are then placed in a furnace and heated up to between 975 to 1010°F and held at that temperature for a sufficient length of time to obtain good fusion of the spot to the plate. Ordinarily this reaction takes only a few minutes after the plate has reached the proper temperature. After firing, the plates are allowed to cool in the furnace to the point where they can be

removed without danger of the crystal cracking due to cold shock. It is essential to control the temperature of the furnace so that the temperature of the crystal plates does not reach 1063°F, otherwise the crystals may become electrically twinned and consequently useless. In order to avoid shattering of quartz plates due to thermal shocks while heating up and cooling, fairly long cycles have heretofore been specified. However, more recent experience has shown that much shorter cycles can be employed especially where small crystals are involved. Moreover, there are indications that the faster heating, particularly during the last two or three hundred degrees temperature rise, results in better spots. During the firing

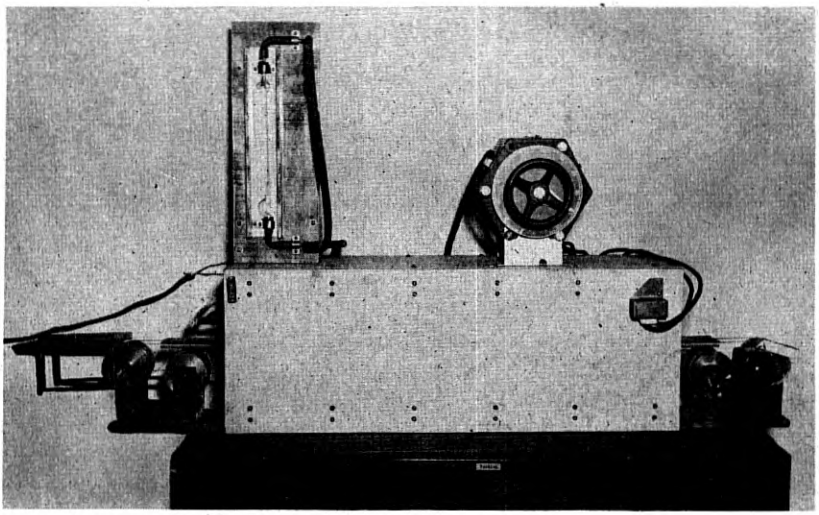


Fig. 13.4—Continuous-belt furnace for firing silver spots.

operation the crystal plates may be placed on nichrome wire mesh trays or Pyrex dishes provided ample provision is made for air to circulate around the spots. This precaution is essential with the types of pastes employed as the baking reaction must take place in the presence of oxygen so that the lead borate will not be reduced to lead, leaving only a partially bonded mixture of silver and lead on the plate. This type of spot when encountered usually has a dull appearance after burnishing as compared with the bright surface obtained with a good silver spot, and is quite difficult to wet with solder. For firing silver spots, ventilated continuous-belt-type furnaces with open ports are being used with very satisfactory results. Figure 13.4 shows a furnace of this type developed by C. J. Christensen for this purpose.

After the firing operation, the spots should be examined to ensure that they are satisfactory. Besides a visual inspection, it is desirable that a

small percentage of the plates be used as a control sample to which mounting wires are attached and pull tested. Satisfactory spots should withstand for a few seconds a force of at least two pounds. The average pull-off strength of commercial attachments using 6-mil hooked or headed wires is between three and four pounds. Unsatisfactory plates can be reclaimed at this stage by stripping off the spots by means of aqua regia and ammonium hydroxide, and reprocessing in the manner described.

13.42 Silver Plating

At this point of the process the surfaces of the plates are coated with silver electrodes by the evaporation process previously mentioned. Four milligrams per square inch of silver is the weight of coating generally employed, which amounts to a thickness of .024 mil. Except for harmonic and other special types of crystal units, these coatings are required only on the major surfaces of the plates. However, during the evaporation process, the silver is deposited to some degree on the minor surfaces or edges as well, and it is necessary to remove it. This process called "edge cleaning", is done by lapping the edges of the crystal on a flat plate covered with a mixture of pumice or a finely divided abrasive such as No. 600 carborundum and water or kerosene in the form of a paste. Rubbing the edge of the plate lightly over very fine abrasive cloth is also satisfactory. The pumice is preferable, however, since while it readily removes the silver, it is much softer than quartz and consequently does not remove any material from the plate. The use of harder abrasives has a tendency to chip the edge of the quartz unless the operation is performed very carefully. After the edge cleaning is completed the plates are washed, dried and inspected by testing for insulation resistance at 500 volts d.c. to make certain that no conducting material remains between the silver coatings on the major surfaces.

13.43 Division of Coating

For circuit reasons all but a few types of crystal units require a balanced pair of electrodes on each side of the plate. Division of coatings along the longitudinal axis is essential on flexure mode crystals in order to make the plate vibrate in flexure. Typical divisions of coating can be noted on the crystals shown in Figure 13.2. In the case of the flexural crystal the dividing line is carried around the wire attachments in such a manner that each of the divided surfaces is connected to one of the wires. One method for dividing coating involves the use of a low voltage (two to three volts) impressed between the coating to be divided and a stylus.[7] When the fine point of the stylus is brought in contact with the silver plating and moved along the desired line of division, the silver is burned away, leaving a small

[7] W. L. Bond, Pat. #2,248,057.

gap in the plating between 8 and 18 mils wide depending upon the point of the stylus. Following the burning operation, the plate is immersed in a photographic hypo solution to remove all traces of the burned residue after

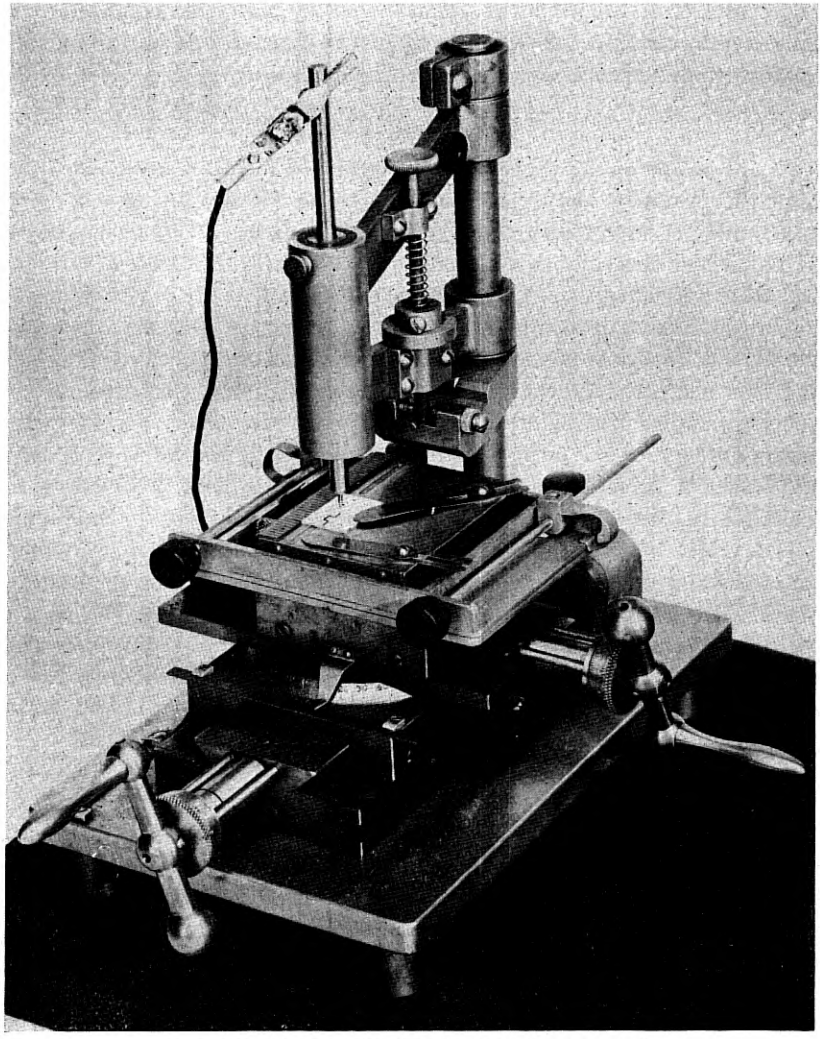


Fig. 13.5—Tool for dividing plating (electric stylus).

which it is carefully washed and rinsed in water and dried. In this process it is important that the plates are not kept in the hypo solution for longer than two or three minutes, otherwise discoloration of the silver coating will

result. The gap in then tested for presence of metallic particles by gradually impressing voltages up to 1000 volts a-c across it. If no flashover occurs the division is satisfactory. If flashover occurs the voltage is maintained until the slivers are burned out. The hypo and burning treatments are repeated until a good division is obtained. Figure 13.5 shows an electric dividing tool developed for this purpose. In using this method it has been found that the arc at the point of the stylus may cause twinning of the quartz to a minor extent along the dividing line. This effect is usually insignificant although it may be objectionable especially where precise values of crystal inductance or frequency-temperature performance are required. Where more complicated divisions are necessary, as in the case of face flexure and harmonic plates, the electric stylus method is employed, although methods and tools for performing this operation by other means to avoid twinning are being developed.

13.44 Attachment of Wire Supporting Leads

Phosphor-bronze wire is employed in wire supported crystal units primarily because of its high tensile strength, and excellent fatigue resistance characteristics. Five- and six-mil diameter wires are the most widely used sizes, depending upon the mass of the crystal plate, the desired electrical performance, and the severity of treatment it is likely to encounter in use. To facilitate soldering the wires to the spot on the crystal plate and to the crystal support system the phosphor-bronze wire is given a heavy electrotinned finish. 59.5–34.5 per cent tin-lead eutectic solder saturated with approximately 6 per cent silver at 570°F is employed for attaching these fine wires to the silver spots of the crystals. This solder solidifies at approximately 360°F with practically no mushy stage. The reason for saturating the solder with silver is to discourage migration of the silver in the spot to the solder during the soldering operation. Even with this solder it is advisable to limit the time for heating of the joint to a minimum.

One method of attaching the wires to the crystal plate is by means of a special machine developed for the purpose. Such a machine is illustrated in the photograph on Fig. 13.6. The wire is fed from a spool through the head in the movable arm. The head contains a wire guide having a hole only slightly larger than the diameter of the wire and a small vise for firmly clamping the wire. The crystal plate is clamped in the vise on the hot plate which is thermostatically controlled at approximately 240°F. The position of the arm carrying the wire is lined up with respect to the crystal plate by means of guides so that the wire will be placed exactly on the nodal point or line of the crystal. In making the attachment, with everything lined up, the wire is fed through the guide until it touches the spot on the plate and the vise closed. Since the curvature of the wire can never be entirely

eliminated the distance between the tip of the guide and the crystal plate is kept as small as possible. A small disc of solder is then punched in the press at the left, the little disc remaining in a round slot whose position is also lined up with respect to the arm carrying the wire. The movable arm

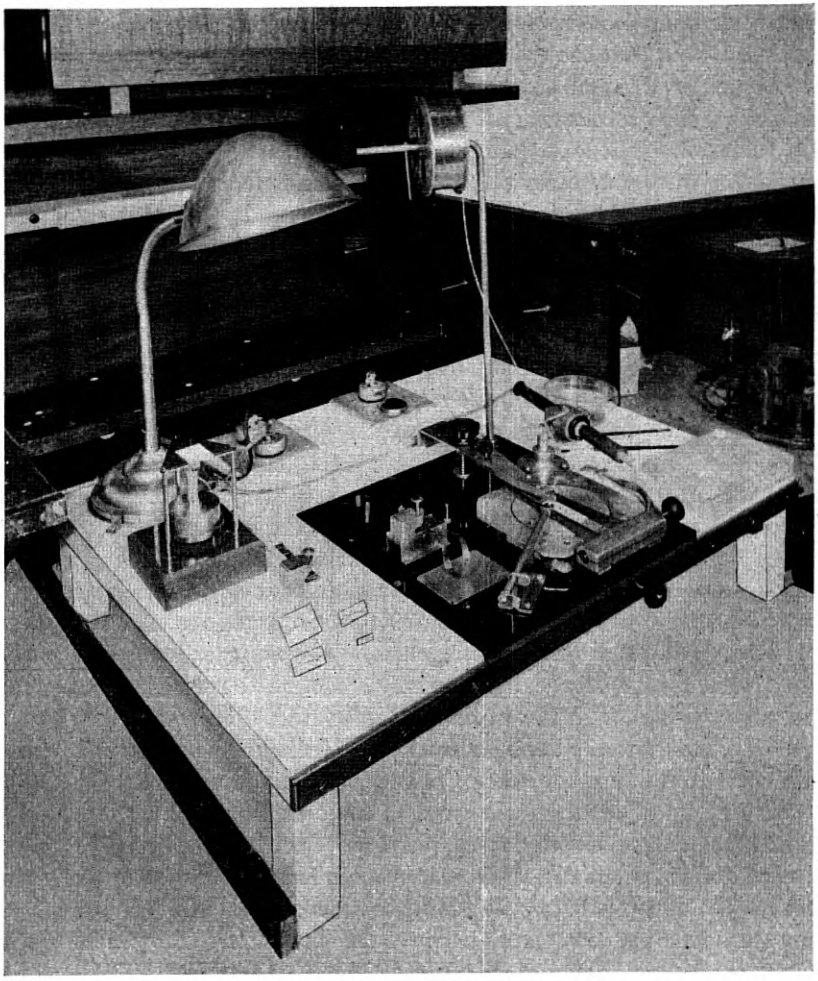


Fig. 13.6—Wire-soldering machine for straight or hooked wires.

is then rotated until the wire is directly over the solder disc at which point it falls into a guide and comes down and spears the disc. The arm is then lifted, picking up the solder at the end of the wire. The solder and the end of the wire is then wetted with rosin-alcohol flux and the arm rotated

until it falls into its original position over the crystal plate. The solder is then fused to the wire and plate by means of a special aluminum tipped soldering iron as shown in the illustration or by a controlled hot air blast focused on the joint to melt the solder. In this operation a fillet or conical button is formed around the wire attaching it to the silver spot. To promote good wetting of the solder, the spot should be clean and well burnished. Rubbing the spot on a hard polished metal surface or burnishing with a blunt pointed tool of agate, are the best methods found so far.

The hot air blast has now replaced the iron entirely in commercial use. It consists of a tube through which air at about one inch water pressure is passed over a hot filament and through a nozzle directed at the solder. The head of the filament is adjusted so that the temperature of the blast is just hot enough to melt the solder and complete the attachment in 10 to 15 seconds time. In using this method with large plates care must be taken to insure that the temperature of the crystal has reached that of the hot plate and that the blast is brought up to the plate slowly, for otherwise the heat shock of the localized blast may cause the crystal to crack. For very small plates the use of a hot plate may be dispensed with if the hot blast is brought up slowly enough to preheat the crystal plate. Other means of melting the solder such as a hot radiant wire or ribbon or the use of a minute flame have been considered, but so far no extensive trials of these methods have been made. The advantage of the hot blast over the other methods mentioned is that it can be better controlled since little is left to the judgment of the operator. If an iron is used it must actually be touched to the solder with the possibility of displacing the position of the wire. Moreover, as already mentioned, the iron must be equipped with a special aluminum tip to prevent removal of solder from the joint on withdrawal of the iron. Considerable maintenance is required to keep such irons in satisfactory operating condition.

13.45 Type of Wire Attachments

The type of attachment described above wherein the part of the wire embedded in the solder cone is straight was used in the first designs of wire-supported crystals. However, it was found that with such attachments vibration of the crystal plate caused breakage of the bond between the solder cone and the wire with resultant failure of the attachments, especially in large plates. Because of this the use of straight wires is recommended only for small size plates. In order to eliminate the above difficulty a little hook has been placed at the end of the wire embedded in the solder in order to obtain a better anchorage. The hook is formed in the wire by means of a special tool affixed to the soldering machine. The basic methods described for straight wires are otherwise used for this type of attachment. Instead

of spearing the little solder discs as with the straight wire, the solder is punched in the shape of a horseshoe and squeezed in place on the hook or positioned by tool with the hooked wire in place on the spot. Hooked wire attachments will withstand pulls of the order of three to four pounds before pulling off. Under severe vibration hooked-wire attachments have the same tendencies as straight wires towards breaking away of the wires from the top of the solder cone forming a small crater in the latter. However, the crater does not progress deeply enough into the solder cone to impair their strength or cause failure under ordinary conditions.

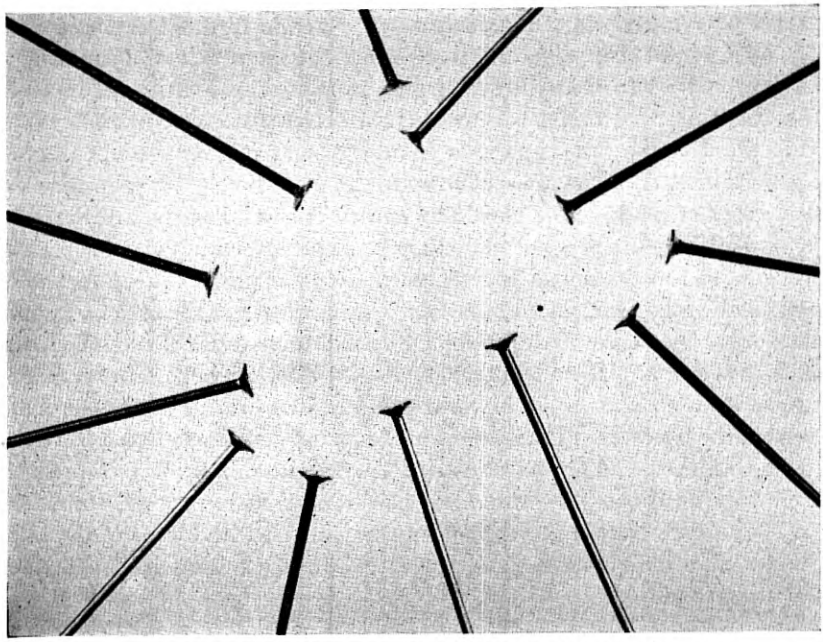


Fig. 13.7—Examples of headed phosphor-bronze wire.

The most recent development for wire supports involves the use of headed phosphor-bronze wires as worked out by A. W. Ziegler. In this procedure individual wire lengths are cut and one end upset in a cold heading tool which provides a little cone-shaped head with a base of about 22 mils as shown in Figure 13.7 for 6-mil diameter wires. The head is carefully pre-tinned, leaving a small globule of solder at the end. Depending upon the size of the crystal plate, globules of 1000, 3000 or 7000 cubic mils of solder are used. The attachments are made in a wire soldering tool which attaches the wires to both sides of the plate simultaneously. This tool is illustrated in Figure 13.8. The prepared wires are fed into positioning guides, which

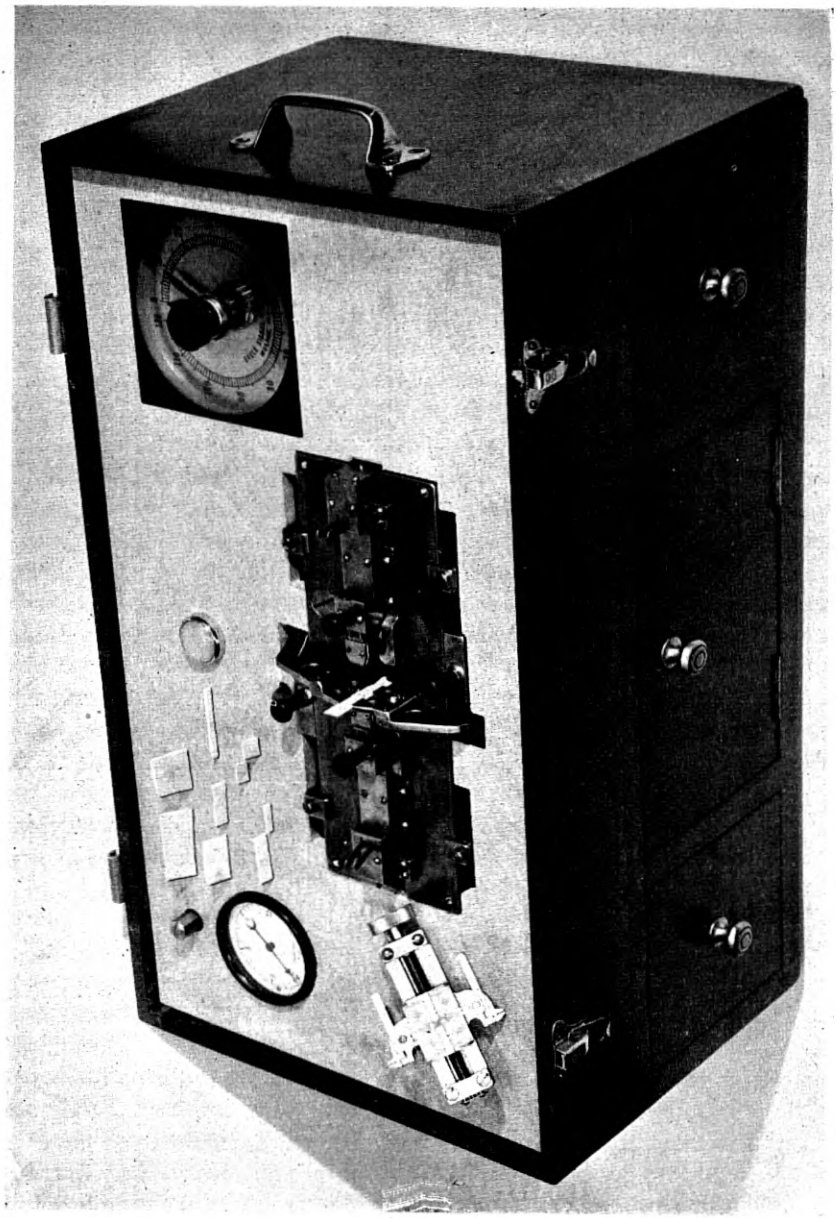


Fig. 13.8—Machine for attaching headed wires.

are aligned with respect to the crystal plate, which is properly located and held between sliding jaws as shown. Prior to the operation the silver spots are burnished and fluxed. The wires in their guides are then slid into contact with the plate. The hot blasts are raised from below so as to aim directly at the work. The solder is melted and the attachment completed in about 10 seconds when the blasts are withdrawn. Slight pressure is maintained on the wires during this operation by springs in the guides to force the head of the wire to seat on the spot. A small fillet is obtained around the head making the solder cover an area of about 35 to 40 mils diameter.

Crystal units using headed wire attachments have many advantages over those made with straight or hooked wires. The pull-off strength is more uniform and averages slightly better than that of hooked wires, despite the fact that only a fraction of the amount of solder used with hooked wires is employed. With this type of attachment the cratering effects encountered with previous methods have also been eliminated. The reduced quantity of solder on the face of the crystal plate effects a decided improvement in the temperature coefficient of the crystal unit as well as in its efficiency and stability. Although the heading of the wires and the subsequent cleaning and tinning operations involve more work, the process of making the attachments is simpler and quicker, since the use of individual wires is better adapted to making all the attachments in one operation. Headed wire attachments are more uniform in size and shape and give a more workmanlike finish to the job. This type of attachment has now replaced those using straight and hooked wires in virtually all designs of telephone type crystal units using wire supports. While the potential advantages of a headed wire type of attachment for crystal support wires had been known for many years, the practical exploitation of the idea depended on finding commercial means for producing the headed wires. The development of a suitable machine for this purpose was carried out by the Western Electric Company in close collaboration with the Laboratories. Figure 13.9 shows such a machine. The fine wire is fed through the lower mechanism to a die in which it is firmly clamped with a predetermined amount extending above the plate. This part of the wire is then cold-worked by multiple punches in the head of the machine until a conical shaped head is formed in the die cavity. As the individually headed wires are formed they are cut off to a definite length and expelled as the vise is released and the next wire brought into position. Cold heading of the wires is necessary in order to retain the elastic properties of the phosphor-bronze springs employed in the suspension. The operation of the tool is simple after the precise alignments of the die and punches have been made.

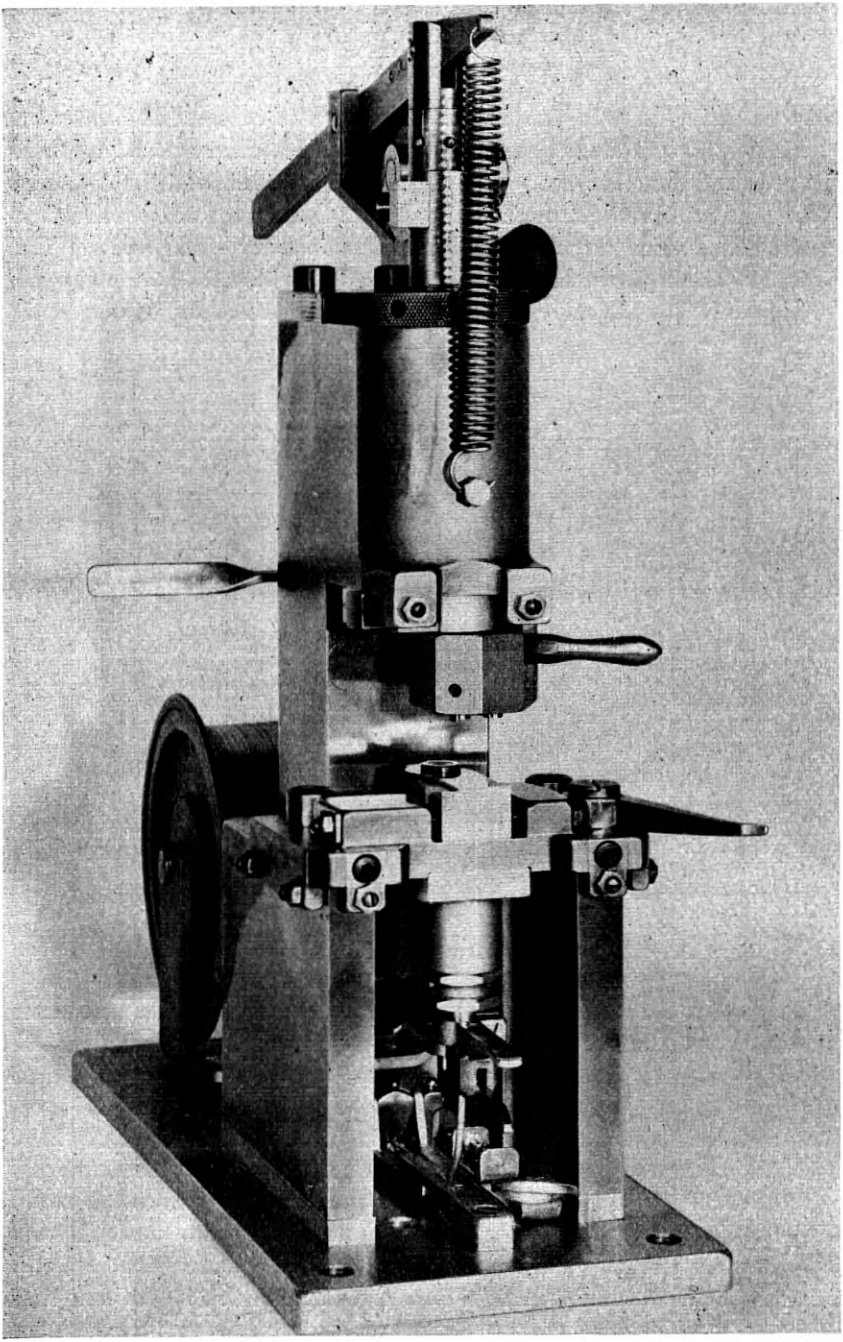


Fig. 13.9—Tool for cold-heading phosphor-bronze wires.

13.46 Mounting the Crystal Plate

After the suspension wires are affixed to the plate, they are then bent to serve as springs and to permit soldering into the cages as illustrated in Fig. 13.2. Two different types of springs are used, one of them involving one bend and the other two. The direction of the bends and the distances between them have been worked out so that the crystal will be displaced to about the same extent in all three directions for equal forces. The cages are of simple construction being made up of mica stampings and metal rods. The assembly of these parts is performed by welding little

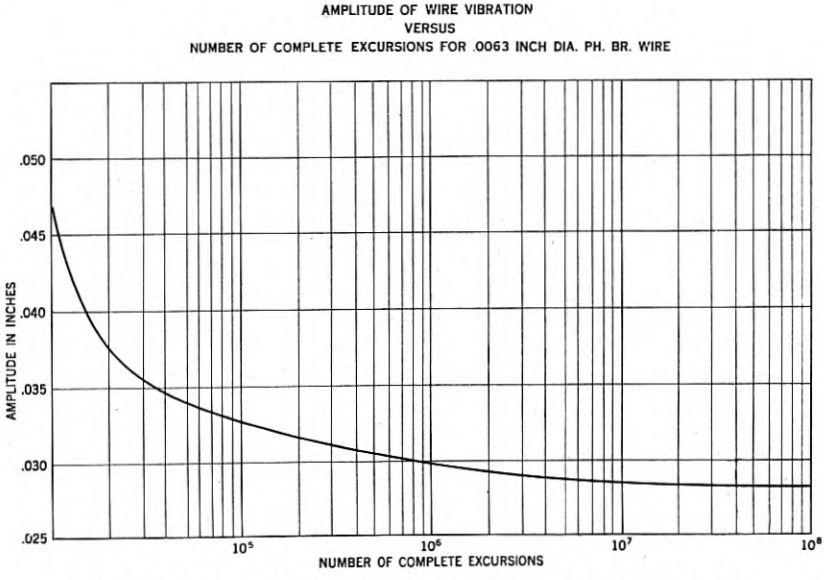


Fig. 13.10—Characteristic performance of phosphor-bronze spring wires.

eyelets, which are staked into the micas, to the rods. In the structures shown, the inside micas are provided with rectangular slots which limit the sidewise movement of the crystal plate from 25 to 30 mils. The end micas are spaced so as to limit the movement of the plate in the lengthwise direction by the same amount. Aside from being used as parts of the cage, the micas therefore serve as "bumpers" to prevent excessive displacement, and possible breakage of the wires or plate if the crystal units are subject to extreme vibration or shock. Figure 13.10 is an experimental curve showing the minimum number of excursions made by wire-mounted crystal plates vibrated at different amplitudes before wire failure occurs for 6.3 mil phosphor bronze spring wires with single bends. On the basis of these data, the

chosen spacing of 25–30 mils between the crystal plate and the bumper should ensure against any service failure of the unit in this regard. In order to center the crystal laterally and longitudinally in the bumper system, the plate is assembled first in the cage by means of spacers. The fine wires are then soldered to the vertical rods or "straights" as they are called, and the spacers removed leaving the plate suspended in position. In order not to set up any strains in the junctions of the wire to the straight which might tend to displace the plate after this operation, the spring wires are usually pre-formed to come within about 5 to 20 mils of the straight. The junction is made by immersing the intersection of the wire and straight in a ball of molten solder. As the wires are withdrawn the ball of molten solder comes with them, solidifying in the air and thus joining the fine wire to the straight without strain.

It will be noted from Fig. 13.2 that the wires to the longitudinal crystal are equipped with little weights close to the plate. This practice has been found desirable on virtually all types of crystal units to alleviate problems of wire resonance[8] which arise in occasional units thereby causing high resistance as well as a shift in the frequency of the plate. Initially, while these effects were noted to some extent in the course of laboratory developments, it was not thought that they would be prevalent enough to warrant taking precaution to eliminate them by loading the wires, since they can usually be corrected by refloating and resoldering the crystal plate thereby changing the effective length of the wire. However, it has turned out that in manufacture a large enough percentage of crystals contain resonant wires to warrant the use of weights. For low-frequency crystals (up to 200 kc) solder balls are placed on the wire at the desired location using a method worked out in conjunction with the Western Electric Company. The process is performed in somewhat the same manner as that described above for connecting the crystal support wires to the straights, except that the weight of the solder deposited and the distance from the plate is more critically controlled. For higher-frequency crystals above 200 kc in which more precise positioning of the weight is essential, small metal discs are employed. They are threaded onto the mounting wire and held in the correct position by a definite amount of solder on the back to obtain the desired loading. Since the free length of wire must be accurately controlled, the manufacturing aspects of this job have been greatly simplified by the use of headed wires in which the variation in height of the solder cones is very small. The chart shown in Fig. 13.11 shows the weights of solder balls or discs and the position they should take on crystals having frequencies up to about one megacycle. It should be noted that the chart covers .0063″ phosphor

[8] "Principles of Mounting Quartz Plates," R. A. Sykes, *B.S.T.J.*, April 1944.

bronze wire. For any other diameter, d, of phosphor bronze wire, the new distance

$$X' = X\sqrt{\frac{d}{.0063}}.$$

Earlier, it was mentioned that the supporting wires for the plates were formed with one or two bends. In addition to the function of suspension these bends also introduce changes in impedance along the wire thus mini-

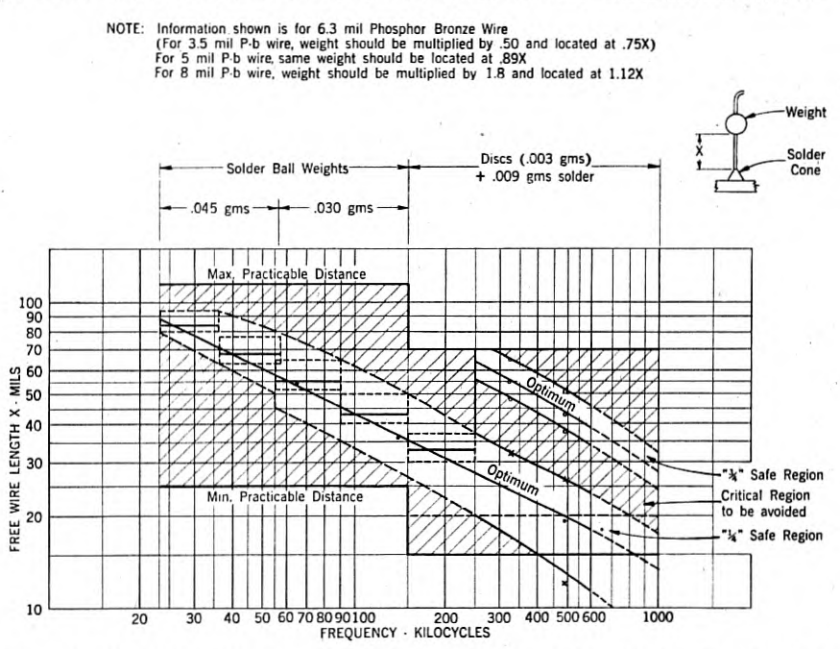


Fig. 13.11—Graph for determining placement of weights on wires for damping vibration.

mizing the possibility of trouble due to wire resonance. The use of a greater number of bends in the wire would tend to accomplish the same result as that of weights. However, the use of weights is considered more practical and has been adopted. As a result of this change, it is possible to employ wire supports having only one bend in virtually all crystals. In low-frequency crystals (below 2 kc) where the wave-length of the flexural wave in the wire is relatively long it is unnecessary to use weights since the wire length can be controlled adequately by the termination of the support wire at the straight. Depending upon the frequency, the desired length of wire is obtained by using either two or three direction bends.

13.47 *Housing of Crystal Units*

For the pressure-type units first discussed no provision was made for protecting or sealing them other than the hermetically sealed containers in which all the other associated components of the filter were enclosed. However, the wire-supported designs have been worked out so that each unit is sealed in its own individual container. Fortunately, the sizes of virtually all crystal units are in the range which permits the use of relatively inexpensive radio tube parts for these housings. There are many obvious advantages to the individually sealed unit. After adjustment and sealing it can be handled more readily in subsequent assembly operations. It is not subject to variations due to changes in ambient humidity and consequently does not restrict the assembly of apparatus to conditioned space. It can be made up and stored or shipped as an individual unit. It has a higher degree of stability. There is one small effect, however, in the case of units which are sealed in vacuum. Due to the absence of any gaseous medium around the crystal, a slight change in frequency is encountered when the tube is evacuated. However, this change is always the same for each particular type and size of crystal and can be allowed for in the final adjustment before sealing.

Most designs of crystals can be sealed in an atmosphere of dry air although better performance results from the use of vacuum. Some crystals must be sealed in vacuum for this reason. A decided advantage in favor of vacuum-sealed crystals is the elimination of acoustic effects from air resonance.

Both metal and glass tubes are used for housing crystal units. Initially it appeared that metal tube radio parts were ideally adapted to crystal use, and it was felt that, instead of welding the stem to tube, this sealing operation could be done by soldering. However, it was found that while sound solder joints could be obtained, extreme precautions were necessary to protect the button-type glass seals, through which the leads emerge, during the pre-tinning and soldering operations. Even with such precautions, it would have been essential to include in every vacuum type tube a means of detecting whether or not a leak had developed. For air-filled tubes at atmospheric pressure this would not have been necessary since minute leaks can be tolerated with little likelihood of the crystal being affected over a long period of time. The possibility of welding as is done in the case of radio tubes was considered but did not appear justified on the basis of equipment cost. Moreover, even with welding there still appeared to be problems from leakage and outgassing of the metal since, after the crystal is enclosed, the assembly cannot be exposed to high temperature to drive off adsorbed gases during the evacuation process. In view of these draw-

backs the use of metal tubes has been discarded in favor of glass tubes except in the case of a few special designs.

The procedure of mounting a crystal unit on a stem and sealing it in glass is much the same as for a radio tube. Figure 13.2 shows crystal units mounted on stems ready for sealing and also shows units sealed in glass and based. The extensions of the straights through the bottom micas are welded to the formed wires emerging from the glass seals. In the glass-sealing operation care must be taken not to heat up the assembly to the point where the solder attachments will be melted or even softened enough to permit the crystal to change position. To accomplish this it is necessary to use hot, sharp-pointed fires localized to the region where the seal will be made. The use of oxygen-gas flames is virtually essential to accomplish the seal quickly. Having the fires strike the bulb at tangency is also desirable. The ordinary type of glass-sealing head for use with gas-air fires is not well adapted to this work since the rotating pillars require the fires to be held too far away from the work thereby necessitating larger flames and consequently more heating up of the crystal unit assembly. The screening effect of the pillars as they revolve also slows up the work of the fires thus increasing the over-all heating of the assembly. A special glass-sealing machine developed for sealing crystal units is shown in Fig. 13.12. Immediately following the sealing operation the glass units should be placed in a suitable annealing box or leer where they can cool off very slowly. A large wooden block equipped with holes to admit the individual bulbs is convenient. The holes may be covered with a cloth to prevent air circulation.

After the units have cooled they are placed on a vacuum pumping station and evacuated. During the first half hour of pumping they are enclosed in a heated oven in which the ambient temperature is maintained at about 240°F. This drives off any traces of moisture that might have entered the tube prior to sealing. Following the heating interval, the tubes are pumped for another half-hour during which time they will have cooled down to room temperature. At this point the pressure in the tubes should be at the minimum of which the pump is capable of attaining. This value should be at most 20 microns and preferably less. However, with a six- or eight-tube station better than 15–20 microns is not likely to be attained unless a liquid nitrogen trap is employed in the system for eliminating moisture.

After the pumping period, vacuum-type units are sealed off, with pump running, by melting the glass tubulation with a fine-pointed oxygen-gas flame as close to the stem as possible. If air is to be admitted, the pump is closed off from the system and dry air admitted to the tubes after which they are sealed off. After testing the crystal unit to see that it meets its requirements, the unit is equipped with a base in the same manner as followed for radio vacuum tubes.

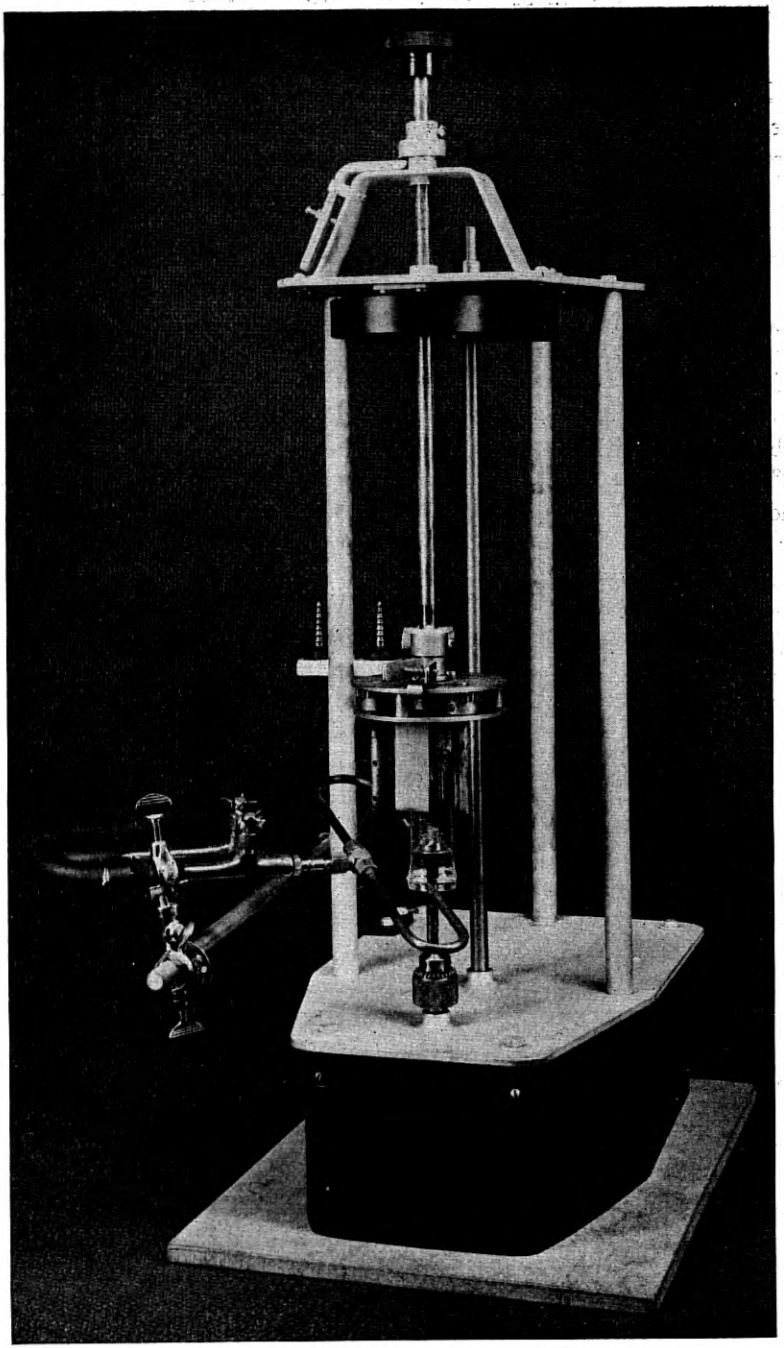


Fig. 13.12—Glass-sealing machine.

13.48 Stabilization of Crystal Units

Despite the close dimensional tolerances applying to the manufacture of the individual crystal plates, the exact frequencies are rarely realized in the mounted crystals. To bring the crystal to frequency it is therefore necessary to grind off minute layers from the lengthwise or widthwise edges of the plate depending on the mode of vibration. This adjustment, which causes superficial disruption of the quartz areas affected, results in unstable operation of the unit with respect to frequency and resistance. Unless the crystal plate is properly treated after these operations, considerable drift in these characteristics will take place, particularly during the initial service life of the unit To alleviate this condition, the crystal units are first rough-adjusted to the approximate frequency and then heat-aged in an oven which subjects the units to several heating and cooling cycles between 240°F and 75°F. The units are then mounted in their cages as previously described and fine-adjusted after which they are again aged. This operation also tends to drive off any moisture which might be troublesome. With this type of accelerated aging the crystals are stabilized to the point where changes in performance can be detected only by the use of the most precise measuring equipment over long periods of time. Crystals so stabilized may generally be depended upon, at any one temperature, maintaining their frequency indefinitely within two or three parts per million provided they are not subjected to excess voltage.

13.49 Cleaning of Crystal Units

Throughout the manufacturing process it is essential that every precaution be taken to keep the crystal plate and associated parts absolutely free from contamination and dirt. The rigorous cleaning necessary before the spotting operation has already been discussed. In all the subsequent operations care must be taken to prevent the plates coming in contact with substances that might tend to cause corrosion. Any particles of foreign matter that may have accumulated on the plate or wires before rough and fine adjustments should be carefully washed off. Otherwise the performance and life of the crystal may be adversely affected. A suitable method for cleaning crystal units before sealing consists of washing and rinsing in chemically pure carbon-tetrachloride or other suitable solvent to remove grease, followed by washing and rinsing in hot distilled water at about 150°F. To facilitate removal of unwanted substances, the parts should be scrubbed gently with a soft brush or agitated in the solution during this process. The use of pure alcohol (95%) in addition to carbon-tetrachloride is also good for this process, but is not essential. The cleansed crystal units should be carefully dried out and protected from further contamination prior to the sealing operation.

13.5 Conclusion

In ending I should like to acknowledge the aid and useful suggestions given me by Mr. C. E. Lane and my other associates in preparing this article. I should also like to reiterate the fact that the status of the art as described was reached after many years of pioneering development by many engineers. In some cases the names of individuals associated with specific contributions of a major nature have been mentioned.

CHAPTER XIV

Effects of Manufacturing Deviations on Crystal Units for Filters

By A. R. D'HEEDENE

14.1 The Effect of Deviations in the Characteristics of Crystal Units on Filter Performance

THIS chapter emphasizes primarily the need for close control in the manufacture of crystal units for use in filters. The first telephone use of crystal units in the commercial manufacture of filters was made by the Western Electric Company in about 1936. To make such commercial manufacture practical, it was necessary to establish accurate design information and allowable manufacturing tolerances. The quantitative data collected for this purpose provided the chief source of material for this chapter. While the data is quite extensive, it will be observed that there are still some factors which must be treated qualitatively.

While filter crystal units are like oscillator crystal units in that they must have low internal dissipation and a close control of resonant frequency, they are different in that many additional characteristics of the filter crystal units must also be controlled accurately. Two typical illustrations will demonstrate how characteristics other than resonant frequency and Q may react on filter performance.

The first characteristic considered is the slope of the reactance with frequency curve in the vicinity of the series resonant frequency. This slope is sometimes referred to as the impedance level of the crystal unit. A convenient measure is the inductance of the equivalent electrical circuit. When this inductance departs from its nominal value, the performance of the filter using the crystal unit may undergo appreciable change. This is particularly true of filters in which the schematic contains a lattice or some other type of bridge circuit with crystal units contained in all the bridge arms. For example, in Fig. 14.1 the solid curve illustrates the transmission characteristic obtained from a lattice-type crystal filter, in which both the series branches and the diagonal branches contain two balanced crystal units. High loss results from a close impedance balance between the branches of the lattice. When the inductance of any of the crystal units departs from its nominal value, the bridge balance is disturbed and the transmission characteristic of the filter is changed. The two dotted curves of Fig. 14.1 illustrates the characteristics that result when the inductance

values of the crystal unit in either branch depart from their nominal values by about one per cent. A negative departure in one branch results in about the same effect on performance as a positive departure in the other branch. The difference between the two curves shown on Fig. 14.1 is that one assumes a positive departure and the other a negative departure for the inductance of a branch.

Due to the close impedance balance which is required for these filters, the effect of small departures in resonant frequency will produce rather large variations in the transmission characteristic. For example, departures of about 10 cycles per second in the crystal units of either branch will produce variations in discrimination of about the same type and magnitude as those

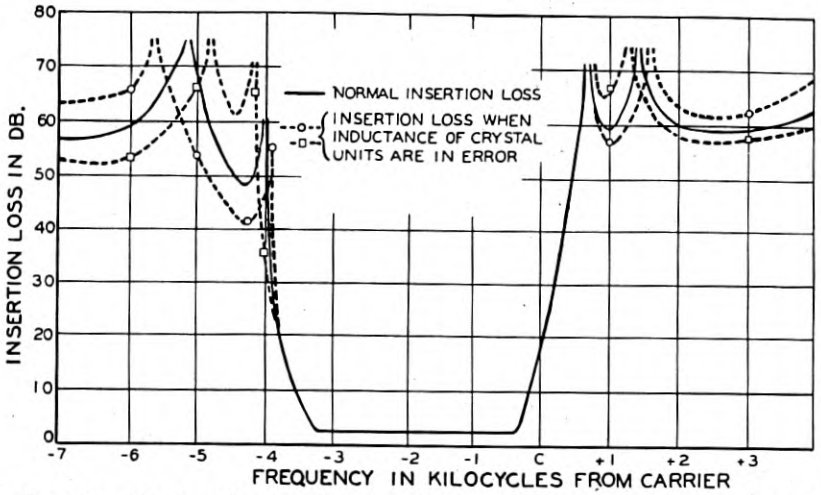


Fig. 14.1.—The insertion loss characteristic of a crystal band-pass filter as affected by deviations in the inductance of the crystal units.

illustrated in Fig. 14.1 for departures in inductance. On the other hand, if the crystal units of both branches exhibit equal departures the entire transmission characteristic will be shifted by the frequency departure of the crystal units, and there will be no loss in discrimination.

Another way in which deviations in the properties of crystal units may react on filter performance is illustrated by the schematic and curves shown in Fig. 14.2. The schematic is the equivalent electrical circuit of a narrow band filter, using two balanced quartz crystal units. The filter is designed to provide a passed band of about 10 cycles per second with distortion of less than 0.2 db. The insertion loss characteristics show that the desired transmission can be obtained for various magnitudes of effective resistance as long as the resistances in the series and diagonal branches are equal.

However, if the effective resistance in one branch is twice as large as that in the other branch a highly distorted characteristic results as shown by the middle curve of Fig. 14.2.

Both of these illustrations show that filter performance is degraded rapidly if the crystal units of the lattice have characteristics which depart from their nominal values by different extents for the two branches. A similar effect is produced when the temperature coefficient of resonant frequency for the crystal units in one branch differs from the temperature coefficient of the units in the other branch. Deviations occurring in a single unit may also affect filter performance. Such deviations include the presence of unwanted resonances of even weak amplitude, inadequate insulation resistance between the metallized coatings or unbalance between the halves of plates on which the coating has been divided.

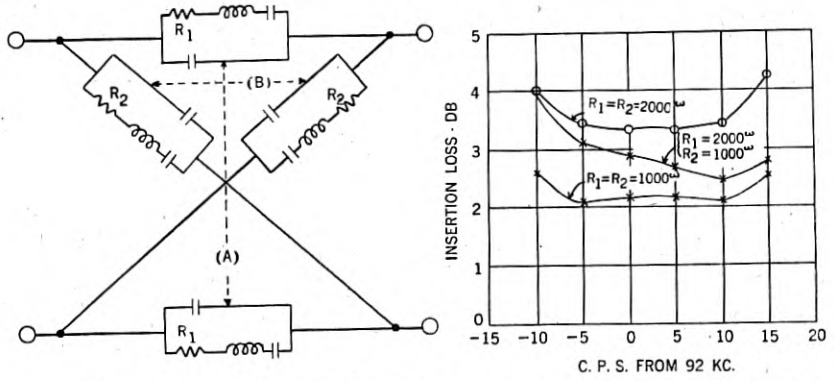


Fig. 14.2.—Effect of deviation in the effective resistance of crystal units on the distortion characteristic of a crystal filter.

The importance of controlling the electrical characteristics of the crystal units is indicated from the above considerations. It is pertinent to correlate deviations in the mechanical properties of the crystal unit with the deviations in electrical characteristics. This is the subject of the succeeding sections. Consideration is restricted to the plates commonly used in filters, that is, X-cut plates, vibrating in extensional or flexural modes, and GT-cut plates.

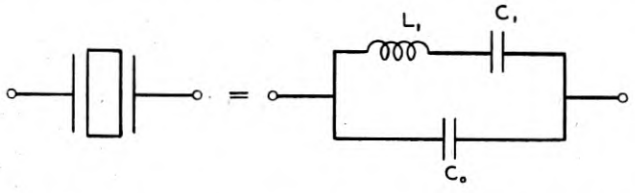
14.2 The Effect of Deviations That Occur in the Manufacture of Quartz Plates

Quartz is an anisotropic material. Accordingly, plates cut from a quartz crystal exhibit elastic and piezo-electric properties which depend on the orientation of the plates with respect to the principal axes of the crystal.

For that reason, any deviation in the orientation of the plates from nominal will affect the electrical characteristics of the crystal units. In addition, these characteristics are affected by imperfections in the plates due to deviations in linear dimensions, to the presence of flaws, or to the condition of the surface of the plates. The effects of these deviations differ for various cuts of crystal plates, for plates of various shapes and for the various modes of vibration. In the following paragraphs, each type of deviation will be considered in turn and data will be presented to show its effect on the characteristics of crystal units using the various types of plates.

14.21 Deviations in the Angle of Orientation

Accurate information is available on the effect of deviation in angle of orientation on the characteristics of X-cut plates vibrating in the exten-



$$C_0 = \frac{K\, w\, l}{4\pi\, t} \times \frac{1}{9\times10^{11}} \text{ FARADS} \qquad f_R = \frac{1}{2\, l\sqrt{\rho\, s'_{22}}} \text{ CYCLES PER SECOND}$$

$$L_1 = \frac{\rho\, s'^{2}_{22}\, t\, l}{8\, d'^{2}_{12}\, w} \times 9\times10^{11} \text{ HENRIES} \qquad C_1 = \frac{8\, d'^{2}_{12}\, w\, l}{\pi^2\, s'_{22}\, t} \times \frac{1}{9\times10^{11}} \text{ FARADS}$$

Fig. 14.3.—Equivalent electrical circuit of piezoelectric crystal.

sional mode. The relation between the electrical characteristics of this type of vibration and the properties of the quartz are shown in Fig. 14.3. This information, with minor changes, is reproduced from a preceding publication.[1] In Fig. 14.3: l,w and t are the length, width and thickness respectively of the plate; K is the dielectric constant; ρ is the density; d'_{12} is the piezo-electric constant; and s'_{22} is the modulus of compliance (inverse of Young's modulus). All these individual quantities are expressed in electrostatic units. The quantities which depend on the orientation of the plates are the piezo-electric constant and the modulus of compliance. The symbols for these quantities usually are primed when they are used for a generalized orientation. When unprimed, the symbols designate quantities measured along the principal axes. For X-cut plates, deviations of the plane of the major surface from the YZ plane have relatively

[1] "Electrical Wave Filters Employing Crystals with Normal and Divided Electrodes", W. P. Mason and R. A. Sykes, *B. S. T. J.*, April 1940, page 222.

small effect, while variations in the angle of rotation about the X-axis have a relatively large effect on these quantities.

Mason has shown[2] how the magnitudes of the piezo-electric constants and the moduli of compliance for any angle of rotation may be derived from their magnitudes along the principal axes of quartz. Using these equations and the magnitudes for the principal axes tabulated in a recent paper[3] by Mason, d'_{12} and s'_{22} have been calculated as a function of the angle of rotation of the plates about the X-axis. In turn, the frequency and inductance constants have been calculated as a function of the angle of rotation, using the relations shown in Fig. 14.3. Figure 14.4 is a plot of the frequency and inductance constants as a function of the angle of rotation for angles between about $-70°$ and $+70°$. It shows how the inductance and resonant

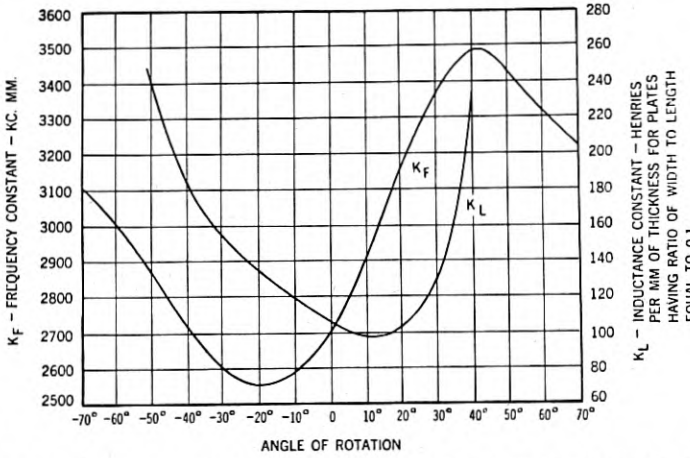


Fig. 14.4.—Frequency and inductance constants of X-cut quartz plates as a function of their angle of rotation around the X-axis.

frequency of these plates will change with deviations in the angle of rotation. The frequency constant used is the product of the resonant frequency in kilocycles and the length of the plate in millimeters. The inductance constant used is the inductance per millimeter of thickness of a plate having a width-to-length ratio equal to 0.1.

The change of inductance and resonant frequency with deviations in angle of rotation about the X-axis is of particular interest for the angles of rotation most commonly used, that is for $-18.5°$ and for $+5°$. The calculations indicate that for an X-cut plate rotated $-18.5°$, a deviation of $\pm 1°$ in the angle of rotation will change the inductance by $\mp 1.2\%$

[2] "Electrical Wave Filters Employing Quartz Crystals as Elements", W. P. Mason, *B. S. T. J.*, July 1934, equations on page 451; also in Chapter VI.
[3] "Quartz Crystal Applications," W. P. Mason, *B. S. T. J.*, July 1943.

respectively, and the resonant frequency by +.05% and −.02%, respectively. For an X-cut plate rotated +5°, a deviation of ±1° will change the inductance −0.9%, and +0.6%, respectively, and the resonant frequency by ±0.7%.

Deviations in angle of rotation will, in general, affect temperature coefficient. The effect is illustrated by Fig. 1.19 of Chapter I,[3] which, for X-cut plates, shows the relation between temperature coefficient and angle of rotation. Ths curve shows that the temperature coefficient is practically zero for an angle of rotation of +5°. For that reason this particular cut is used whenever a low-temperature coefficient is desired. The curve also shows that at this point the slope of temperature coefficient as a function of angle of rotation is zero. Hence, for the +5° X-cut plate, which is most important from the standpoint of temperature coefficient, there will be little change due to a deviation in the angle of rotation.

In GT-cut plates the effect of deviation in the angle of orientation must be considered in combination with deviations in linear dimensions. Mason shows[4] that for an angle of rotation of +51 degrees 7.5 minutes and a width-to-length ratio of .859, a temperature coefficient close to zero may be obtained from −25° C to +75° C. He also has shown that this temperature coefficient varies with both the angle of rotation and the width-to-length ratio. Because of this, it has been found possible to compensate for small deviations in the angle of rotation by adjusting the linear dimensions. The net effect of a deviation in angle of rotation, after it has been so compensated, is to raise (or lower) the temperature region of zero temperature coefficient by 11° C for each 10-minute increase (or decrease) in angle of rotation. In GT-cut plates, the width dimension directly controls the primary resonance. For this reason, it is preferable to adjust temperature coefficient by varying the length dimension rather than the width. The crystal plates are cut larger than desired. The frequency of resonance and the temperature coefficient are then adjusted simultaneously by grinding either the width or length dimension as required. Experimental work carried on by L. F. Willey of the Laboratories shows that an increase of 1.0 per cent in the width-to-length ratio results in an increase in temperature coefficient of approximately +1.35 parts per million per degree C. In his experimental work Willey used the ratio of the secondary to the primary frequency as a convenient measure of the width-to-length ratio. The inductance constant for GT plates, which is about 17 henries per millimeter of thickness, will increase by less than 1% for deviations in any of the angles of rotation of as much as 30 minutes. However, the inductance may depart appreciably from nominal due to the adjustment of width and length dimensions.

[4] "New Quartz-Crystal Plate, GT, Produces Constant Frequency Over Wide Temperature Range", W. P. Mason, *Proc. I. R. E.*, May, 1940, page 220.

14.22 Deviations in Linear Dimensions

In the case of X-cut plates, the length dimension is used to control the location of their resonant frequency. The length is lapped to its final dimension after all other processes have been completed, so that this dimension will be such as to compensate for the effect of any other deviations that may have occurred. The sensitivity of this adjustment depends on the mode of vibration. For plates vibrating in their extensional mode, the resonant frequency is inversely proportional to the length, as shown in Fig. 14.3, while for plates vibrating in their flexural mode, the resonant frequency is inversely proportional to the square of the length. The amount of the adjustment required depends on the magnitude of the frequency errors that may have been introduced due to deviations in the width or in the angular orientation of the plates or due to still other causes. The magnitude of such frequency errors, in turn, depends to a considerable degree on the angle of rotation of the plates. For example, it was shown in Section 14.21 that a deviation in the angle of rotation of a $+5°$ plate changes its resonant frequency more than ten times as much as a similar deviation in a $-18.5°$ plate. It must be noted that the adjustment of length compensates for frequency errors only and that errors in inductance or temperature coefficient may be increased by such adjustment.

Deviations in the thickness dimension principally affect the impedance level of the plates. As shown by Fig. 14.3, the inductance is directly and the capacity inversely proportional to the thickness. In the case of GT-cut plates the thickness dimension is important also because it controls the location of the most prominent unwanted resonances, which arise from vibrations in thickness flexure. However, plates are designed to avoid critical thicknesses and small deviations from the nominal thickness will not usually result in plates having unwanted resonances.

Deviations in the width dimension affect the equivalent electrical characteristics appreciably. The effect of deviations in width on the frequency of X-cut plates vibrating in their extensional mode can be deduced from the curves of Fig. 14.5. The curves show that this effect is more pronounced for larger values of the width-to-length ratio where coupling with the width extensional mode becomes appreciable. For a $-18.5°$ plate with a width-to-length ratio of 0.8 an increase of 1% in the width dimension will decrease the frequency by about .04%. For a $+5°$ plate with a width-to-length ratio of 0.4 an increase of 1% in the width dimension will also decrease the frequency by about .04%, but for a ratio of 0.6 the decrease in frequency will amount to 0.13%.

Similar information is available for crystals vibrating in width flexure from the measurements published by Harrison[5] and the calculations pub-

[5] "Piezo-Electric Resonance and Oscillatory Phenomena with Flexural Vibration in Quartz Plates", J. R. Harrison, *Proc. I. R. E.*, Dec. 1927.

lished by Mason.[6] This information shows that for small ratios of axes the resonant frequency will be directly proportional to the width dimension. However, as the ratio is increased to 0.5, a change of 1% in the width dimension will change the frequency by only 0.5%.

The effect of the width dimension on the inductance of the plates frequently is important. Fig. 14.6 illustrates the relation between inductance and the ratio of axes. From these curves the effects of deviations in width can be deduced. For the two longitudinal plates, the inductance is almost inversely proportional to width. For the flexure plate, the decrease of

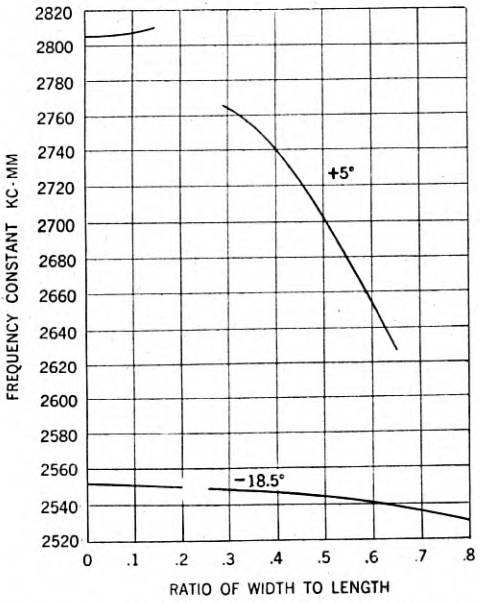


Fig. 14.5.—Frequency constant of the longitudinal mode of X-cut quartz plates as a function of their ratio of width to length.

inductance with increase in width is much more rapid. With a ratio of axes of 0.6 the inductance decreases about as a square power, while with a ratio of 0.1 the decrease is about as the third power.

The width dimension of the +5° plate has an appreciable effect on the temperature coefficient of the plate. Mason has shown[3] that while the temperature coefficient is zero for a long narrow bar, it increases quite rapidly as the width dimension increases, due to coupling between the face shear and the longitudinal modes. In the case of an −18.5° plate, coupling with other modes is relatively weak. Hence its temperature co-

[6] "Motion of a Bar Vibrating in Flexure Including the Effects of Rotary and Lateral Inertia", W. P. Mason, *Jour. Acous. Soc. America*, April, 1935, pages 246–249.

efficient, which is about 25 parts per million per degree C, does not change appreciably with changes in width. For a +5° plate vibrating in its flexure mode, Fig. 14.7 illustrates measurements made on the variation of temperature coefficient with ratio of axes. For all these X-cut crystals, it may be observed that deviations of 1% in the width dimension will not

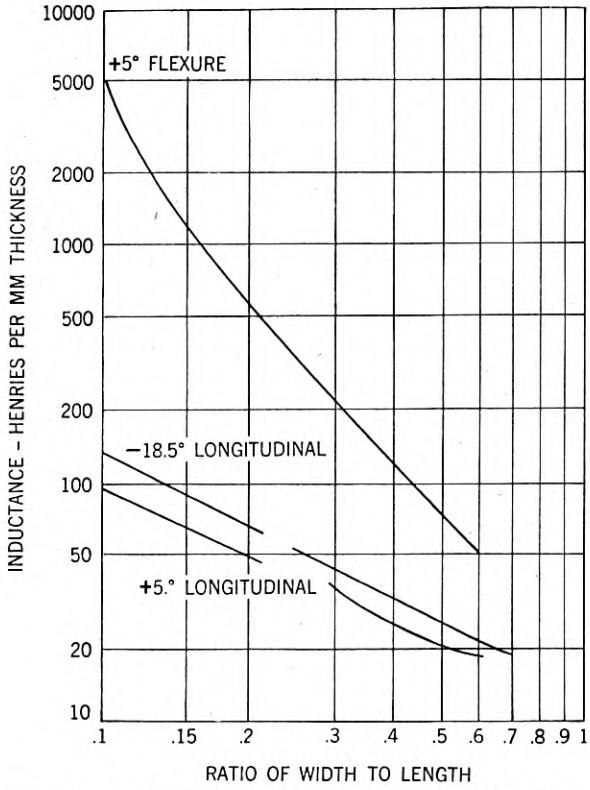


Fig. 14.6.—Inductance of the crystal units used in filters as a function of the cuts of the plates and their ratio of width to length.

change the temperature coefficient by more than 5%. Such changes are usually negligible.

14.23 Internal Defects

Internal defects in the quartz plates may have a large effect on their electrical characteristics. These defects vary so widely in type, size and concentration that it is impossible to predict the effects quantitatively. General comments regarding the results that may be expected for various

defects are described in Chapter IV[7]. The conclusions drawn there for oscillator plates are also applicable to filter plates. These are: (1) Evidence that a particular defect is permissible in a given type of plate does not prove that a similar defect is permissible in some other type of plate, and (2) proof that a particular defect is permissible in a given type of plate can be obtained only by a statistical study.

Some qualitative statements can be made regarding the effect of mechanical flaws. Cracks result in instability of resonant frequency and effective resistance and must be avoided. The effect of inclusions or chips depends

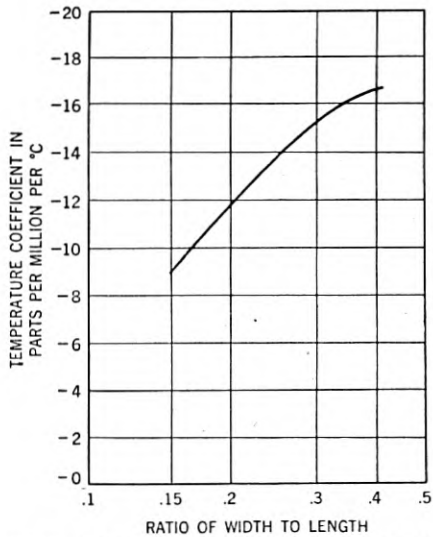


Fig. 14.7.—Temperature coefficient for +5° flexural crystal units as a function of the ratio of width to length of the plates.

on the size of these defects relative to the size of the finished plates and also on their location in the plate.

Twinning in quartz may be either of the optical type (Brazil twin) or of the electrical type (Dauphiné twin)[8]. The effect of these two types of twinning on the performance of oscillator crystal units has been described thoroughly in Chapter V[9].

When optical twinning is present, the plate will exhibit the same elastic properties throughout, but the two portions of the plate will tend to expand

[7] "Raw Quartz, its Defects and Inspection", G. W. Willard, *B. S. T. J.*, October 1943, pp. 338–361.

[8] "The Properties of Silica", R. B. Sosman, Chapter XII.

[9] "Use of the Etch Technique for Determining Orientation and Twinning in Quartz Crystals", G. W. Willard, *B. S. T. J.*, January 1944, pp. 11–51.

and contract in opposite phase. Hence there is little change in frequency constant or temperature coefficient, but there will be a large change in inductance. Th change in inductance can be estimated roughly by comparing the twinned plate with an untwinned plate in which activity has been reduced by removing electrical charge from part of the surface. The area of the surface from which this charge is removed would be twice the twinned area and located at about the same position in the plate.

It is believed that a small amount of electrical twinning is more serious than a similar amount of optical twinning, because the twinned areas are of opposite angular sense. Each of the two areas has a different modulus of compliance and the effective modulus of the plate has a value intermediate between the two different values of modulus. Therefore, the frequency constant of the plate will be intermediate between that of the desired cut and its electrical twin. For a small amount of twinning, the direction and rate of change of frequency can be estimated from the comparison shown on Table I between the standard filter cuts and their electrical twins.

TABLE I

Filter Plate	Frequency Constant—kc. m.m.	Electrical Twin	Frequency Constant—kc. m.m
−18.5°	2560	+18.5°	3120
+5°	2815	−5°	2650
+51.1° (GT)	3280	−51.1°	2610

This verifies the experimentally observed fact that for −18.5° X-cut plates, twinning increases the frequency, while for +5° X-cut and GT plates, twinning decreases the frequency. Even for small amounts of twinning the inductance will increase rapidly for plates of any orientation. When the amount of twinning becomes large, the equivalent inductance approaches infinity. That is, the crystal will not be set in motion by an applied voltage.

The quantitative effect of twinning (probably electrical) has been measured on one set of plates by R. M. Jensen. Figure 14.8 includes a photograph of the plates used, illustrating the extent of the twinning in each. All of the plates are −18.5° X-cut plates, having the dimensions 30.88 x 10.56 x .86 mm. The tabulation below the photograph compares the inductance and resonant frequency measured for each of the plates with the one, designated AN-3, which shows the least effect of twinning. While there is a good correlation between the amount of twinning in the plates and their change in electrical performances, it is not practical to estimate accurately the effect of a given amount of twinning. For this reason, crystal plates having any twinning should not be used for crystal units for filters.

14.24 Etching

The surface condition of the quartz plates also has some effect on crystal characteristics. This surface condition is determined in large part by the lapping operation used to obtain final dimensions. As described in Chapter

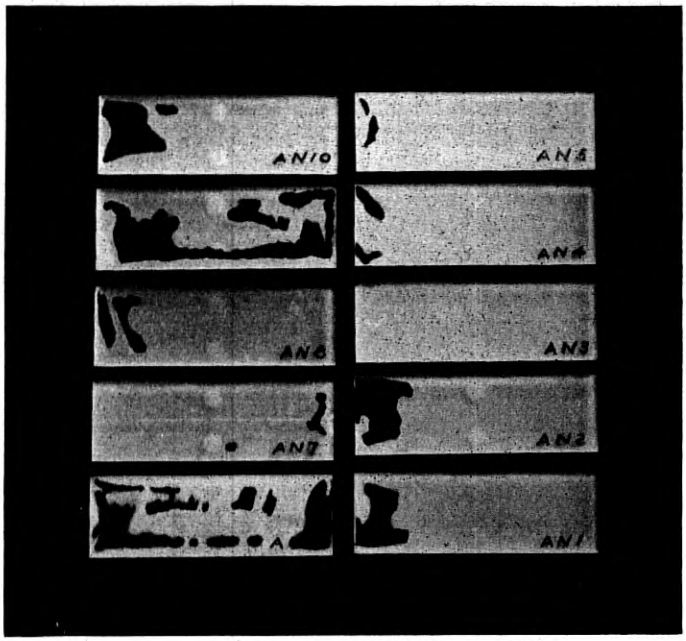


Plate Designation	Percentage Increase over Values Measured for AN-3	
	Inductance	Resonant Frequency
AN-1	+22.12	+.50
AN-2	+27.20	+.59
AN-3	0	0
AN-4	+3.03	+.04
AN-5	+3.12	+.01
AN-6	+276.54	+5.01
AN-7	+1.58	−.03
AN-8	+14.21	+.37
AN-9	+738.00	+6.63
AN-10	+32.44	+.83

Fig. 14.8.—Effect of various degrees of twinning on the performance of −18.5° X-cut quartz crystal plates.

XIII[10], the plates are given a final lap with 400 or 600-mesh carborundum. This, in turn, is followed by an etching bath which removes foreign particles. A short etch, about eight minutes in 47% hydrofluoric acid, has been found adequate to ensure firm adherence of the metal coating to the quartz. On

[10] "The Mounting and Fabrication of Plated Quartz Crystal Units," R. M. C. Greenidge, this issue of the *B. S. T. J.*

the other hand, the use of a relatively long etch, 30 minutes or more, is desirable when a high Q is desired. The long etch also results in an improved stability of the resonant frequency as a function of current. This will be discussed in a subsequent paragraph. A disadvantage of a long etch is the difficulty of controlling the etching process within close tolerances. The variations in rate of removing material may be sufficient to affect the uniformity of the linear dimensions of the plates.

These factors indicate that etching is an important process in preparing crystal plates. A close control must be maintained on the strength of the acid, the uniformity with which the surfaces of the plates are exposed and the duration of the exposure.

14.3 The Effects of Deviations during Fabrication of Wire-Supported Unit

As described in Chapter XIII[10], two types of mountings have been developed for supporting crystal plates, the Pressure Type and the Wire-Supported Type. The wire-supported type of mounting is the more recent development and has resulted in crystal units which have a much higher degree of stability and can be reproduced within much closer tolerances than the units using the pressure type of mounting. Since this chapter is concerned chiefly with the problem of obtaining a high degree of precision in crystal units, the discussion is restricted to the wire-supported type of mounting.

14.31 Silver Spotting

For the wire-supported type of mounting the first operation is to bake small silver spots on the surface of the crystal plates. In the application of these silver spots to the crystal plates three factors are of importance in their effect on the characteristics of the plate, namely, the size of the spot, its location, and the firing temperature. Since in all crystal designs to date the silver spots are applied at or near the nodal line of the crystal plate the principal effect of the spots is to increase the stiffness of the plate, so slightly increasing the frequency of resonance. Variations of an appreciable magnitude in either the amount of silver paste used (that is, the size of the spot) or in the location of the spot with respect to the nodal line will change the resonant frequency of the plate. Such changes could be corrected later, when the plates are adjusted for resonant frequency, as long as the length is increased sufficiently to allow such adjustment. However, if the length be increased sufficiently to allow for extreme cases, average adjusting time will be increased materially, while if the allowance is insufficient some of the plates may be unusable. For this reason, close control of the size and location of the silver spots is well justified.

In baking the silver spots, care must be taken to prevent "heat" twinning. If the temperature of a quartz plate is raised above the inversion point (573°C) and then is reduced again, the plate will be electrically twinned.[8] The firing temperature of the silver paste currently used for the spots is not many degrees below this inversion point. Hence, the firing temperature may easily become so high as to result in twinned plates. In addition, it has been observed that the twinning may occur at a considerably lower temperature if the plate is subjected to large thermal stress. For this reason, care must be taken to heat the plates uniformly during the baking operation.

14.32 Division of Coating

The next operation is to evaporate a coating of silver on the surface of the quartz plates. The plates must be thoroughly cleaned before this coating is applied in order to ensure firm adherence of the coating. Poor adherence may cause the coating to peel off the plate, changing all of the electrical characteristics of the plate. In many cases the coating must also be divided.[1,11] Two methods are in general use for dividing the silver coating on crystal plates, namely, an abrasive method and an electrical stylus method. In general, the abrasive method of dividing the coating is superior to the electric stylus for all cases requiring a simple straight line division, but it has not been found practical for complicated divisions such as are desirable for harmonic longitudinal plates and flexure plates.

In using the abrasive method for dividing the coating only two factors are likely to change the characteristics of the crystal plate, these being the location and the width of the dividing line. Deviations in the location of the dividing line from the lengthwise center line for a longitudinal plate will affect the capacity and inductance balance between the two halves of the plate. Deviations in the width of a properly centered dividing line will cause changes in the inductance of the plate since for a given plate the inductance is a function of the ratio of the plated area to the total area of the plate. So, for a wide crystal plate deviations in the width of the dividing line will be negligible while for narrow plates these deviations can cause an appreciable change in the inductance of the plates.

When the electric stylus is used for dividing the coating, the location and the width of the dividing line again will affect the performance of the plates. In addition, varying amounts of twinning will occur along the division line apparently due to instantaneous high temperature gradients introduced by burning of the silver at the point of contact of the stylus. In measure-

[1] Loc. cit.
[8] Loc. cit.
[11] "Crystal Channel Filters for the Cable Carrier System", C. E. Lane, *B. S. T. J.*, January 1938, pp. 125–136.

ments made on a group of −18.5° X-cut crystals on which the coating has been divided carefully with an electric stylus, the increase in the inductance of the plates ranged from 1.4% to 2.6%. Any twinning resulting from the dividing operation will also change the resonant frequency of the plates.

14.33 Soldering of Wires to Plates

The next process, that of soldering the supporting wires to the crystal plate may have considerable effect on the performance of the unit. The deviations which may be introduced depend on the amount of solder used, the location of the solder button with respect to the nodal line of the plate,

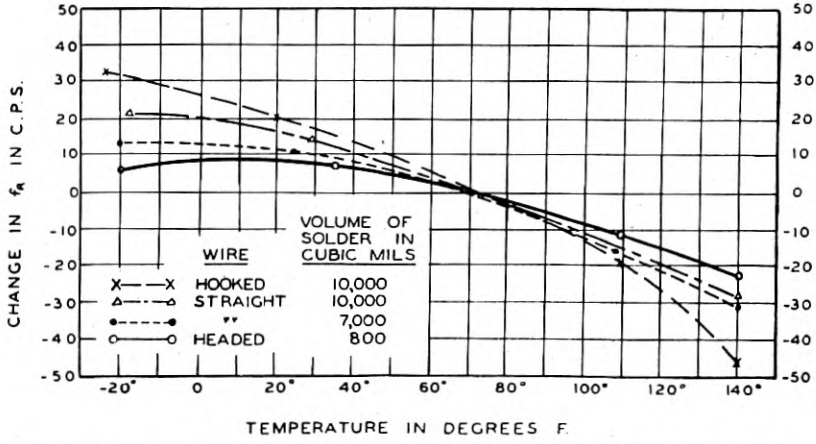


Fig. 14.9.—Change of resonant frequency with temperature of +5° X-cut quartz crystal plates. The curves show that when the volume of solder used for joining the plates to the supporting wires is appreciable compared to the volume of the plates, the frequency-temperature coefficient characteristic is affected by the volume of solder.

the shape of the solder button, and the possible twinning of the plate during the soldering operation.

The amount of solder used in forming the joint of the wire to the plate becomes extremely important when the plate is small. For example, Fig. 14.9 illustrates the changes in the frequency-temperature characteristic resulting from the use of varying amounts of solder for a particular size of plate. The units on which these measurements were made used X-cut +5° plates of 16 mm x 6 mm x 0.5 mm. The types of wire referred to in the figure, that is, hooked, straight and headed, were described in Chapter XIII[10]. The frequency-temperature characteristic expected on the basis of measurements made on crystal units using larger plates is approximated

[10] Loc. cit.

closely by the solid curve. This solid curve actually was obtained from measurements made on crystal units using plates supported with headed wires and using a very small amount of solder. The other curves indicate that the temperature coefficient may be increased appreciably due to the presence of a larger amount of solder. Further, when the larger amounts of solder are used, the characteristics depend on the exact amount of the solder, so that the characteristics represented by the dashed curves are hard to reproduce.

The amount of solder used in this operation also affects the Q of the crystal unit and its resonant frequency. Measurements using several crystal plates of relatively small sizes have shown improvements in Q of as much as 25 per cent when headed wires are used over that obtained with other wires using larger amounts of solder.

Variations in the consistency of the solder joint will, of course, affect the adherence of the supporting wire to the plate. A poor joint will result in a high effective resistance for the crystal unit and will generally cause instability both in resistance and in resonance frequency.

In soldering the supporting wire to the crystal plate two methods have been used for melting the solder; namely, the soldering iron, and the hot-air blast. With either method, lack of sufficient control can seriously change the electrical characteristics of the plate due to twinning. It has been observed that this twinning occurs when there is a large temperature gradient in the quartz, even at temperatures well below the inversion point. Experimental work by G. W. Willard has indicated that it may occur even when the temperature of the soldering iron is as low as 300°C. To avoid such twinning during the soldering operation, it has been found desirable to raise the temperature of the entire plate to just below the melting point of the solder.

Twinning, when it occurs, will affect the crystal plate by causing an increase in inductance, a change in the resonant frequency, increased effective resistance, and a change in the temperature coefficient, as stated previously. Also, in crystals with divided plating there will be an inductance unbalance between the two halves of the crystal plate set up due to unequal amounts of twinning. Several measurements made, using GT plates at 160 kc, showed that twinning during the soldering operation decreased the resonant frequency in a range from 200 to 100 cps and the temperature coefficient of the units ranged from 2 to 6 times that of units using untwinned crystal plates.

14.34 Effects due to Wire Resonance

As described in Chapter VIII[12], the characteristics of crystal units may be changed due to vibrations set up in the supporting wires. When any

[12] "Methods of Mounting and Holding Crystals", R. A. Sykes, *B. S. T. J.*, April 1944.

one of the wires is not located exactly on a node of the plate, the plate will set the wire into vibration. For certain critical lengths of the wire, it will offer considerable resistance to this motion and there will be a rapid increase in effective resistance and some change in resonant frequency of the crystal plate.

The effect of wire vibration can be described in terms of its electrical analogy. The vibrating wire, clamped at its far end, may be considered a rather special electrical transmission line open-circuited at its far end. When viewed from the crystal plate the impedance changes rapidly with

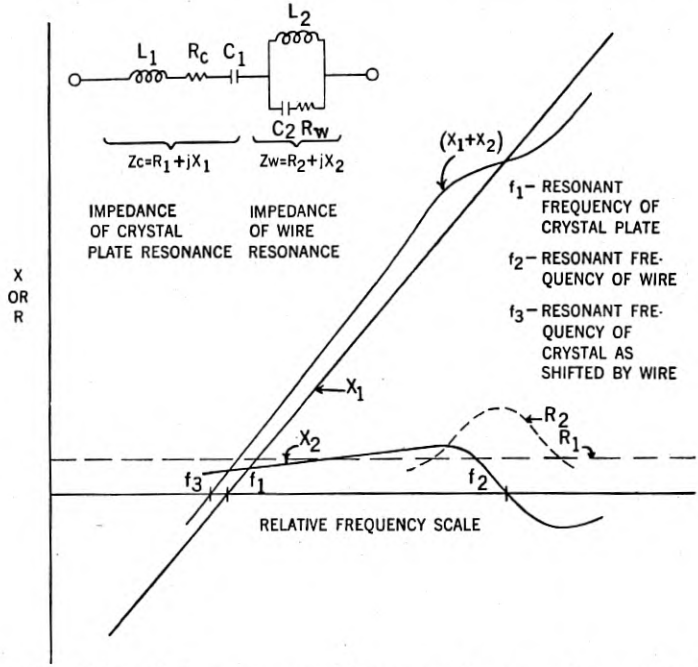


Fig. 14.10.—Effect of wire resonance on the resonant frequency of a crystal unit.

frequency in a succession of pronounced resonances and anti-resonances. In the vicinity of an anti-resonance the electrical equivalent of the vibrating wire may be approximated by a coil and condenser in parallel as shown by L_2 and C_2 of Fig. 14.10. This acts in series with the mechanical resonance of the quartz plate, represented by L_1 and C_1. The impedance curves illustrate the effect of the wire resonance on the crystal impedance. R_1, the equivalent resistance of the crystal plate, is constant for frequencies in the vicinity of resonance. X_1, the equivalent reactance of the crystal plate, increases rapidly as the frequency departs from resonance. R_2 and X_2, the equivalent resistance and reactance of the wire resonance, are typical of an

anti-resonant electrical network. The curve labeled $(X_1 + X_2)$ shows the effect of the wire resonance on the response of the crystal plate. It may be observed that the apparent resonance has been reduced by a small frequency decrement. The amount of frequency shift and the increase in effective resistance depend on the Q of the wire resonance, its frequency location compared with the resonance of the crystal plate, the mass of the wire relative to that of the quartz plate, and the distance from the node to the point at which the wire is actually fastened to the plate.

The slope of the frequency-reactance characteristic corresponding to the mechanical resonance of the quartz plate is very steep and the effect of the wire resonance will be noticed only when an anti-resonant frequency of the wire is close to the resonant frequency of the plate. The changes in resonant frequency and effective resistance due to wire resonance have been measured for some filter crystal units and the measurements are tabulated in Table II.

TABLE II

EFFECT OF WIRE VIBRATIONS ON THE RESISTANCE OF A QUARTZ CRYSTAL PLATE

Crystal Type	Mode of Vibration for Crystal	Resonant Frequency	Crystal Mass in Grams	Distance of Wire from Nodal Line	Maximum Frequency Shift CPS	Maximum Increase in Resistance
+5° X-Cut	Flexural	12 kc	.51	(N) .060″	±2.0	250%
+5° X-Cut	Longitudinal	164 kc	.142	(N) .012″	±30	640%
−18° X-Cut	Longitudinal	335 kc	.075	(M) .002″	±90	360%
−18° X-Cut	Longitudinal	552 kc	.068	(N) 0.0″	±75	1100%
5th Harmonic GT	Longitudinal	164 kc	.98	(N) .011″	±12	370%

(N) Specified Dimension.
(M) Measured Dimension.

The relation between the length of a wire and the frequencies at which it will resonate in flexural modes is expressed by the following equation:

$$l = m\sqrt{\frac{vd}{8\pi f}}$$

where v is the velocity of sound in the wire
d is the diameter of the wire
l is the length of the wire
f is the frequency of wire resonance in cycles per second
m is a number that depends on the manner in which the ends of the wire can move.

At a particular frequency and for wire of a particular material and diameter there is a series of critical wire lengths which must be avoided. The critical lengths are those which cause the wire to present a high impedance to the motion of the plate. This high impedance may be considered, from the electrical point of view, as corresponding to an anti-resonance of the wire.

The critical lengths are defined by the series of numbers $m \doteq (n + \frac{1}{2})\pi$ where n takes the values 1, 2, 3, etc. and apply to successive modes of a bar clamped at both ends. Beyond the first mode, the critical wire lengths are spaced at equal intervals, corresponding to increments of m each equal to π.

There is also a series of wire lengths which will present minimum impedance to the motion of the plate. These may be considered as corresponding to a resonance of the wire. These minima of impedance are obtained for lengths of wire defined by the series of numbers $m \doteq (n - \frac{1}{4})\pi$. They apply to a bar which is clamped at one end and, while free to vibrate at the other end, is constrained to a slope perpendicular to the plate.

In selecting a desirable length for the supporting wire, it is not essential that this length be such as to cause the wire to present minimum impedance to motion of the plate. As a matter of fact, since the wire is of relatively low characteristic impedance a small departure from the critical length is sufficient to avoid trouble from wire resonance. In order to allow for as wide a manufacturing tolerance as possible the supporting wire is usually designed to have a length half-way between two successive critical lengths. For a 6.3-mil phosphor-bronze wire, the spacing between successive critical lengths ranges from about 58 mils at 100 kc to about 15 mils at 1000 kc. Hence, even at 100 kc the length of the supporting wire must be controlled within a tolerance of about 20 mils.

These supporting wires are formed to have definite bends along their length and the location of these bends varies slightly from one wire to another. In addition, the wires are terminated by solder at both ends. Because of these complications it is impractical to meet such close tolerances on the effective length of the wires. Furthermore, a wire that does have a suitable effective length at room temperatures may exhibit sufficient change of properties with variations in temperature so that it becomes of critical length at some other operating temperature.

Much of the difficulty due to wire resonance is avoided by use of a solder ball on the supporting wire, as described in Chapters VIII and XIII. The solder ball is located near the quartz plate. Since it serves as a clamp at that point, it makes the supporting wire short. By locating and forming the solder ball accurately, the length of the supporting wire is controlled within a close tolerance. Further, since the wire is shortened by use of the ball it is less affected by changes in temperature. Experience at about 500 kc indicates that a tolerance of about 10 mils in locating the solder ball is practical and has provided satisfactory operation between −40 C and +85 C.

14.4 Need For Cleanliness and Low Relative Humidity

One of the most serious difficulties encountered in manufacturing quartz crystal plates is that of assuring sufficient cleanliness. Even minute par-

ticles of foreign matter will introduce appreciable changes in crystal performance.

Usually, the presence of foreign matter will act to load the crystal and will reduce the resonant frequency but there are also instances where the added matter tends to stiffen the plate and increase its frequency. The latter has been observed to occur as the result of the deposit of a thin film of rosin on the surface of the plate. In the presence of foreign matter on the surface of the plates, the performance will be unstable with time and temperature even after the plate is sealed into a container. Also, erratic variations are observed as the plate is shifted from a normal atmosphere to a container which is evacuated or filled with dry air. Experience has shown

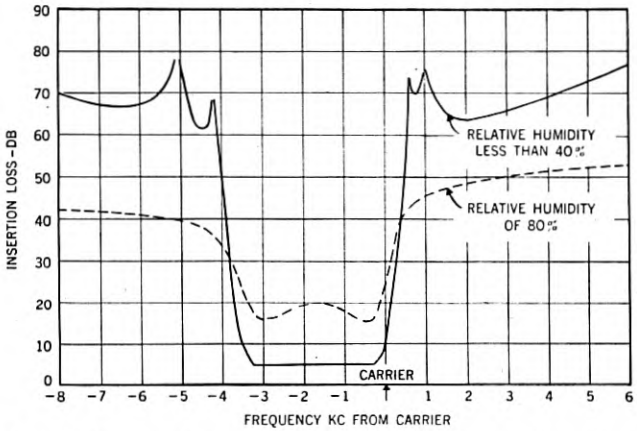


Fig. 14.11.—Effect of humidity on the discrimination of channel crystal filters. To prevent decrease of discrimination with increase of relative humidity, the crystal units must be hermetically sealed.

that elaborate precautions for insuring cleanliness are justified by the time saved in the adjusting processes.

The need for cleanliness is closely related to the effect of humidity on the insulation resistance of crystal units. As used in filters, crystal units must provide extremely high impedances at their anti-resonant frequencies. These impedances may be as high as 100 megohms. With clean crystal plates in relatively dry atmospheres, such insulation resistance can be maintained up to 1000 kc. However, even a trace of salts or other types of contamination will make the insulation resistance highly sensitive to moisture in the adjacent air. While it is relatively difficult to measure insulation resistance at high carrier frequencies, the effect of the reduced insulation due to moisture is evident on inspection of the discrimination characteristics of the filters. For example, Fig. 14.11 illustrates the result of high relative

humidity on the transmission characteristic of a typical crystal filter. It may be observed that the discrimination almost disappears for a relative humidity of 80%. These measurements were made on a filter containing well cleaned crystal plates. It will be found frequently that an unsatisfactory discrimination characteristic is produced by considerably lower values of relative humidity when the plates are not so clean. Experience has shown that it is impractical to let the relative humidity surrounding the crystal plate exceed 40% for satisfactory filter performance. When a high degree of accuracy is required, the plates are assembled

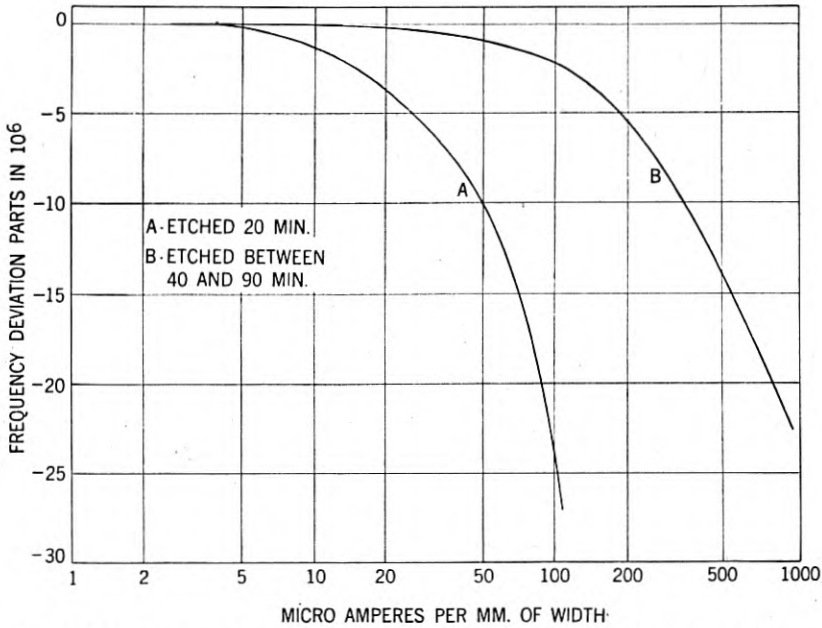


Fig. 14.12.—Change of resonant frequency of GT-cut plates due to increase of transmitted current.

in a unit which is either evacuated or filled with air at a relative humidity of less than 5%.

14.5 Effect of Current Level

Crystal units will undergo change in effective resistance and in frequency of resonance as the current transmitted is increased. Some change might be expected due to the heating of the plate by the dissipative loss associated with the transmission of current. However, the effects are not identical with those obtained with a change in ambient temperature. Appreciable

changes have been observed even when using GT-cut plates adjusted to zero temperature coefficient as shown, for example, in the curves of Fig. 14.12. Also, it has been observed that after a plate has been driven hard and the transmitted current then reduced, the original resonant frequency is restored only after a considerable time interval. The data of Fig. 14.12 provides a rough correlation between stability and current levels. For example, if the stability desired for a crystal unit using a GT type plate be in the order of one part per million, the circuit design should be such as to keep the current level in the plate below about 10 microamperes per millimeter of width.

The parameter used in these paragraphs for measuring current levels is the current per unit of width. This appears to be useful as a common basis for comparing various plates of any one cut and mode of vibration. Theoretically, in the case of a plate vibrating longitudinally, the current, I, per unit of width, w, is directly proportional to the elongation per unit of length, y_y, as shown by following equation:

$$I/w = K\, y_y$$

where K is a constant which depends on the cut of plate and mode.

Figure 14.12 also illustrates the importance of the surface condition of the plates. Curve A is the average frequency-current characteristic for a group of crystal units using plates etched for twenty minutes in 47% hydrofluoric acid and curve B the average characteristic for a group of crystal units using plates etched for over forty minutes but less than ninety minutes. Evidently, crystal units using plates which have been etched for a long period exhibit a frequency-current characteristic which is appreciably more constant than those using plates etched for a shorter period.

Mathematical Analysis of Random Noise

By S. O. RICE

Introduction

THIS paper deals with the mathematical analysis of noise obtained by passing random noise through physical devices. The random noise considered is that which arises from shot effect in vacuum tubes or from thermal agitation of electrons in resistors. Our main interest is in the statistical properties of such noise and we leave to one side many physical results of which Nyquist's law may be given as an example.[1]

About half of the work given here is believed to be new, the bulk of the new results appearing in Parts III and IV. In order to provide a suitable introduction to these results and also to bring out their relation to the work of others, this paper is written as an exposition of the subject indicated in the title.

When a broad band of random noise is applied to some physical device, such as an electrical network, the statistical properties of the output are often of interest. For example, when the noise is due to shot effect, its mean and standard deviations are given by Campbell's theorem (Part I) when the physical device is linear. Additional information of this sort is given by the (auto) correlation function which is a rough measure of the dependence of values of the output separated by a fixed time interval.

The paper consists of four main parts. The first part is concerned with shot effect. The shot effect is important not only in its own right but also because it is a typical source of noise. The Fourier series representation of a noise current, which is used extensively in the following parts, may be obtained from the relatively simple concepts inherent in the shot effect.

The second part is devoted principally to the fundamental result that the power spectrum of a noise current is the Fourier transform of its correlation function. This result is used again and again in Parts III and IV.

A rather thorough discussion of the statistics of random noise currents is given in Part III. Probability distributions associated with the maxima of the current and the maxima of its envelope are developed. Formulas for the expected number of zeros and maxima per second are given, and a start is made towards obtaining the probability distribution of the zeros.

When a noise voltage or a noise voltage plus a signal is applied to a non-

[1] An account of this field is given by E. B. Moullin, "Spontaneous Fluctuations of Voltage," Oxford (1938).

linear device, such as a square-law or linear rectifier, the output will also contain noise. The methods which are available for computing the amount of noise and its spectral distribution are discussed in Part IV.

ACKNOWLEDGEMENT

I wish to thank my friends for many helpful suggestions and discussions regarding the subject of this paper. Although it has been convenient to acknowledge some of this assistance in the text, I appreciate no less sincerely the considerable amount which is not mentioned. In particular, I am indebted to Miss Darville for computing the curves in Parts III and IV.

SUMMARY OF RESULTS

Before proceeding to the main body of the paper, we shall state some of the principal results. It is hoped that this summary will give the casual reader an over-all view of the material covered and at the same time guide the reader who is interested in obtaining some particular item of information to those portions of the paper which may possibly contain it.

Part I—Shot Effect

Shot effect noise results from the superposition of a great number of disturbances which occur at random. A large class of noise generators produce noise in this way.

Suppose that the arrival of an electron at the anode of the vacuum tube at time $t = 0$ produces an effect $F(t)$ at some point in the output circuit. If the output circuit is such that the effects of the various electrons add linearly, the total effect at time t due to all the electrons is

$$I(t) = \sum_{k=-\infty}^{+\infty} F(t - t_k) \tag{1.2-1}$$

where the k^{th} electron arrives at t_k and the series is assumed to converge. Although the terminology suggests that $I(t)$ is a current, and it will be spoken of as a noise current, it may be any quantity expressible in the form (1.2–1).

1. Campbell's theorem: The average value of $I(t)$ is

$$\overline{I(t)} = \nu \int_{-\infty}^{+\infty} F(t)\, dt \tag{1.2-2}$$

and the mean square value of the fluctuation about this average is

$$\text{ave. } [I(t) - \overline{I(t)}]^2 = \nu \int_{-\infty}^{+\infty} F^2(t)\, dt \tag{1.2-3}$$

where ν is the average number of electrons arriving per second at the anode. In this expression the electrons are supposed to arrive independently and at random. $\nu e^{-\nu t}\,dt$ is the probability that the length of the interval between two successive arrivals lies between t and $t + dt$.

2. Generalization of Campbell's theorem. Campbell's theorem gives information about the average value and the standard deviation of the probability distribution of $I(t)$. A generalization of the theorem gives information about the third and higher order moments. Let

$$I(t) = \sum_{-\infty}^{+\infty} a_k F(t - t_k) \tag{1.5-1}$$

where $F(t)$ and t_k are of the same nature as those in (1.2–1) and $\cdots a_1$, a_2, $\cdots a_k$, $\cdots$ are independent random variables all having the same distribution. Then the n^{th} semi-invariant of the probability density $P(I)$ of $I = I(t)$ is

$$\lambda_n = \nu\,\overline{a^n} \int_{-\infty}^{+\infty} [F(t)]^n\,dt \tag{1.5-2}$$

The semi-invariants are defined as the coefficients in the expansion of the characteristic function $f(u)$:

$$\log_e f(u) = \sum_{n=1} \frac{\lambda_n}{n!} (iu)^n \tag{1.5-3}$$

where

$$f(u) = \text{ave. } e^{iIu} = \int_{-\infty}^{+\infty} P(I) e^{iIu}\,dI$$

The moments may be computed from the λ's.

3. As $\nu \to \infty$ the probability density $P(I)$ of the shot effect current approaches a normal law. The way it is approached is given by

$$P(I) \sim \sigma^{-1}\varphi^{(0)}(x) - \frac{\lambda_3\sigma^{-4}}{3!}\varphi^{(3)}(x) + \left[\frac{\lambda_4\sigma^{-5}}{4!}\varphi^{(4)}(x) + \frac{\lambda_3^2\sigma^{-7}}{72}\varphi^{(6)}(x)\right] + \cdots \tag{1.6-3}$$

where the λ's are given by (1.5–2) and

$$\sigma^2 = \lambda_2 \qquad x = \frac{I - \bar{I}}{\sigma} \qquad \varphi^{(n)}(x) = \frac{1}{\sqrt{2\pi}} \frac{d^n}{dx^n} e^{-x^2/2}$$

Since the λ's are of the order of ν, σ is of the order of $\nu^{1/2}$ and the orders of σ^{-1}, $\lambda_3\sigma^{-4}$, $\lambda_4\sigma^{-5}$ and $\lambda_3^2\sigma^{-7}$ are $\nu^{-1/2}$, ν^{-1}, $\nu^{-3/2}$ and $\nu^{-3/2}$ respectively. A

possible use of this result is to determine whether a noise due to random independent events occuring at the rate of ν per second may be regarded as "random noise" in the sense of this work.

4. When $I(t)$, as given by (1.5–1), is analyzed as a Fourier series over an interval of length T a set of Fourier coefficients is obtained. By taking many different intervals, all of length T, many sets of coefficients are obtained. If ν is sufficiently large these coefficients tend to be distributed normally and independently. A discussion of this is given in section 1.7.

Part II—Power Spectra and Correlation Functions

1. Suppose we have a curve, such as an oscillogram of a noise current, which extends from $t = 0$ to $t = \infty$. Let this curve be denoted by $I(t)$. The correlation function of $I(t)$ is $\psi(\tau)$ which is defined as

$$\psi(\tau) = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I(t)I(t+\tau)\,dt \qquad (2.1\text{–}4)$$

where the limit is assumed to exist. This function is closely connected with another function, the power spectrum, $w(f)$, of $I(t)$. $I(t)$ may be regarded as composed of many sinusoidal components. If $I(t)$ were a noise current and if it were to flow through a resistance of one ohm the average power dissipated by those components whose frequencies lie between f and $f + df$ would be $w(f)\,df$.

The relation between $w(f)$ and $\psi(\tau)$ is

$$w(f) = 4\int_0^\infty \psi(\tau)\cos 2\pi f\tau\,d\tau \qquad (2.1\text{–}5)$$

$$\psi(\tau) = \int_0^\infty w(f)\cos 2\pi f\tau\,df \qquad (2.1\text{–}6)$$

When $I(t)$ has no d.c. or periodic components,

$$w(f) = \operatorname*{Limit}_{T\to\infty} \frac{2\,|\,S(f)\,|^2}{T} \qquad (2.1\text{–}3)$$

where

$$S(f) = \int_0^T I(t)e^{-2\pi i f t}\,dt.$$

The correlation function for

$$I(t) = A + C\cos(2\pi f_0 t - \varphi)$$

is

$$\psi(\tau) = A^2 + \frac{C^2}{2}\cos 2\pi f_0\tau \qquad (2.2\text{–}3)$$

These results are discussed in sections 2.1 to 2.4 inclusive.

2. So far we have supposed $I(t)$ to be some definite function for which a curve may be drawn. Now consider $I(t)$ to be given by a mathematical expression into which, besides t, a number of parameters enter. $w(f)$ and $\psi(\tau)$ are now obtained by averaging the integrals over the possible values of the parameters. This is discussed in section 2.5.

3. The correlation function for the shot effect current of (1.2–1) is

$$\psi(\tau) = \nu \int_{-\infty}^{+\infty} F(t)F(t+\tau)\,dt + \left[\nu \int_{-\infty}^{+\infty} F(t)\,dt\right]^2 \qquad (2.6\text{–}2)$$

The distributed portion of the power spectrum is

$$w_1(f) = 2\nu \mid s(f) \mid^2$$

where

$$s(f) = \int_{-\infty}^{+\infty} F(t)e^{-2\pi i f t}\,dt \qquad (2.6\text{–}5)$$

The complete power spectrum has in addition to $w_1(f)$ an impulse function representing the d.c. component $\overline{I}(t)$.

In the formulas above for the shot effect it was assumed that the expected number, ν, of electrons per second did not vary with time. A case in which ν does vary with time is briefly discussed near the end of Section 2.6.

4. Random telegraph signal. Let $I(t)$ be equal to either a or $-a$ so that it is of the form of a flat top wave, and let the lengths of the tops and bottoms be distributed independently and exponentially. The correlation function and power spectrum of I are

$$\psi(\tau) = a^2 e^{-2\mu|\tau|} \qquad (2.7\text{–}4)$$

$$w(f) = \frac{2a^2\mu}{\pi^2 f^2 + \mu^2} \qquad (2.7\text{–}5)$$

where μ is the expected number of changes of sign per second.

Another type of random telegraph signal may be formed as follows: Divide the time scale into intervals of equal length h. In an interval selected at random the value of $I(t)$ is independent of the value in the other intervals and is equally likely to be $+a$ or $-a$. The correlation function of $I(t)$ is zero for $\mid \tau \mid > h$ and is

$$a^2\left(1 - \frac{\mid \tau \mid}{h}\right)$$

for $0 \leq \mid \tau \mid < h$ and the power spectrum is

$$w(f) = 2h\left(\frac{a \sin \pi f h}{\pi f h}\right)^2 \qquad (2.7\text{–}9)$$

5. There are two representations of a random noise current which are especially useful. The first one is

$$I(t) = \sum_{n=1}^{N} (a_n \cos \omega_n t + b_n \sin \omega_n t) \tag{2.8-1}$$

where a_n and b_n are independent random variables which are distributed normally about zero with the standard deviation $\sqrt{w(f_n)\Delta f}$ and where

$$\omega_n = 2\pi f_n, \qquad f_n = n\Delta f$$

The second one is

$$I(t) = \sum_{n=1}^{N} c_n \cos (\omega_n t - \varphi_n) \tag{2.8-6}$$

where φ_n is a random phase angle distributed uniformly over the range $(0, 2\pi)$ and

$$c_n = [2w(f_n)\Delta f]^{1/2}$$

At an appropriate point in the analysis N and Δf are made to approach infinity and zero, respectively, in such a manner that the entire frequency band is covered by the summations (which then become integrations).

6. The normal distribution in several variables and the central limit theorem are discussed in sections 2.9 and 2.10.

Part III—Statistical Properties of Noise Current

1. The noise current is distributed normally. This has already been discussed in section 1.6 for the shot-effect. It is discussed again in section 3.1 using the concepts introduced in Part II, and the assumption, used throughout Part III, that the average value of the noise current $I(t)$ is zero. The probability that $I(t)$ lies between I and $I + dI$ is

$$\frac{dI}{\sqrt{2\pi\psi_0}} e^{-I^2/2\psi_0} \tag{3.1-3}$$

where ψ_0 is the value of the correlation function, $\psi(\tau)$, of $I(t)$ at $\tau = 0$

$$\psi_0 = \psi(0) = \int_0^\infty w(f)\, df, \tag{3.1-2}$$

$w(f)$ being the power spectrum of $I(t)$. ψ_0 is the mean square value of $I(t)$, i.e., the r.m.s. value of $I(t)$ is $\psi_0^{1/2}$.

The characteristic function (ch. f.) of this distribution is

$$\text{ave. } e^{iuI(t)} = \exp -\frac{\psi_0}{2} u^2 \tag{3.1-6}$$

2. The probability that $I(t)$ lies between I_1 and $I_1 + dI$, and $I(t + \tau)$ lies between I_2 and $I_2 + dI_2$ when t is chosen at random is

$$[\psi_0^2 - \psi_\tau^2]^{-1/2} \frac{dI_1\, dI_2}{2\pi} \exp\left[\frac{-\psi_0 I_1^2 - \psi_0 I_2^2 + 2\psi_\tau I_1 I_2}{2(\psi_0^2 - \psi_\tau^2)}\right] \quad (3.2\text{–}4)$$

where ψ_τ is the correlation function $\psi(\tau)$ of $I(t)$:

$$\psi(\tau) = \int_0^\infty w(f) \cos 2\pi f\tau\, df \quad (3.2\text{–}3)$$

The ch. f. for this distribution is

$$\text{ave. } e^{iuI(t)+ivI(t+\tau)} = \exp\left[-\frac{\psi_0}{2}(u^2 + v^2) - \psi_\tau uv\right] \quad (3.2\text{–}7)$$

3. The expected number of zeros per second of $I(t)$ is

$$\frac{1}{\pi}\left[-\frac{\psi''(0)}{\psi(0)}\right]^{1/2} = 2\left[\frac{\int_0^\infty f^2 w(f)\, df}{\int_0^\infty w(f)\, df}\right]^{1/2} \quad (3.3\text{–}11)$$

assuming convergence of the integrals. The primes denote differentiation with respect to τ:

$$\psi''(\tau) = \frac{d^2}{d\tau^2}\psi(\tau).$$

For an ideal band-pass filter whose pass band extends from f_a to f_b the expected number of zeros per second is

$$2\left[\frac{1}{3}\frac{f_b^3 - f_a^3}{f_b - f_a}\right]^{1/2} \quad (3.3\text{–}12)$$

When f_a is zero this becomes $1.155\, f_b$ and when f_a is very nearly equal to f_b it approaches $f_b + f_a$.

4. The problem of determining the distribution function for the length of the interval between two successive zeros of $I(t)$ seems to be quite difficult. In section 3.4 some related results are given which lead, in some circumstances, to approximations to the distribution. For example, for an ideal narrow band-pass filter the probability that the distance between two successive zeros lies between τ and $\tau + d\tau$ is approximately

$$\frac{d\tau}{2} \frac{a}{[1 + a^2(\tau - \tau_1)^2]^{3/2}}$$

where

$$a = \sqrt{3}\,\frac{(f_b + f_a)^2}{f_b - f_a}, \qquad \tau_1 = \frac{1}{f_b + f_a}$$

f_b and f_a being the upper and lower cut-off frequencies.

5. In section 3.5 several multiple integrals which occur in the work of Part III are discussed.

6. The distribution of the maxima of $I(t)$ is discussed in section 3.6. The expected number of maxima per second is

$$\frac{1}{2\pi}\left[-\frac{\psi_0^{(4)}}{\psi_0''}\right]^{1/2} = \left[\frac{\int_0^\infty f^4 w(f)\,df}{\int_0^\infty f^2 w(f)\,df}\right]^{1/2} \tag{3.6-6}$$

For a band-pass filter the expected number of maxima per second is

$$\left[\frac{3}{5}\,\frac{f_b^5 - f_a^5}{f_b^3 - f_a^3}\right]^{1/2} \tag{3.6-7}$$

For a low-pass filter where $f_a = 0$ this number is $0.775\,f_b$.

The expected number of maxima per second lying above the line $I(t) = I_1$ is approximately, when I_1 is large,

$$e^{-I_1^2/2\psi_0} \times \tfrac{1}{2}[\text{the expected number of zeros of } I \text{ per second}] \tag{3.6-11}$$

where ψ_0 is the mean square value of $I(t)$.

For a low-pass filter the probability that a maximum chosen at random from the universe of maxima lies between I and $I + dI$ is approximately, when I is large,

$$\frac{\sqrt{5}}{3}\,y e^{-y^2/2}\,\frac{dI}{\psi_0^{1/2}} \tag{3.6-9}$$

where

$$y = \frac{I}{\psi_0^{1/2}}$$

7. When we pass noise through a relatively narrow band-pass filter one of the most noticeable features of an oscillogram of the output current is its fluctuating envelope. In sections 3.7 and 3.8 some statistical properties of this envelope, denoted by R or $R(t)$, are derived.

The probability that the envelope lies between R and $R + dR$ is

$$\frac{R}{\psi_0}\,e^{-R^2/2\psi_0}\,dR \tag{3.7-10}$$

where ψ_0 is the mean square value of $I(t)$. The probability that $R(t)$ lies between R_1 and $R_1 + dR_1$ and at the same time $R(t + \tau)$ lies between R_2 and $R_2 + dR_2$ when t is chosen at random is obtained by multiplying (3.7–13) by $dR_1\, dR_2$. For an ideal band-pass filter, the expected number of maxima of the envelope in one second is

$$.64110(f_b - f_a) \tag{3.8–15}$$

When R is large, say $y > 2.5$ where

$$y = \frac{R}{\psi_0^{1/2}}, \qquad \psi_0^{1/2} = \text{r.m.s. value of } I(t),$$

the probability that a maximum of the envelope, selected at random from the universe of such maxima, lies between R and $R + dR$ is approximately

$$1.13(y^2 - 1)e^{-y^2/2}\frac{dR}{\psi_0^{1/2}}$$

A curve for the corresponding probability density is shown for the range $0 \leq y \leq 4$. Curves which compare the distribution function of the maxima of R with other distribution functions of the same type are also given.

8. In section 3.9 some information is given regarding the statistical behavior of the random variable:

$$E = \int_{t_1}^{t_1+T} I^2(t)\, dt \tag{3.9–1}$$

where t_1 is chosen at random and $I(t)$ is a noise current with the power spectrum $w(f)$ and the correlation function $\psi(\tau)$. The average value m_T of E is $T\psi_0$ and its standard deviation σ_T is given by (3.9–9). For a relatively narrow band-pass filter

$$\frac{\sigma_T}{m_T} \sim \frac{1}{\sqrt{T(f_b - f_a)}}$$

when $T(f_b - f_a) \gg 1$. This follows from equation (3.9–10). An expression which is believed to approximate the distribution of E is given by (3.9–20).

9. In section 3.10 the distribution of a noise current plus one or more sinusoidal currents is discussed. For example, if I consists of two sine waves plus noise:

$$I = P \cos pt + Q \cos qt + I_N, \tag{3.10–20}$$

where p and q are incommensurable and the r.m.s. value of the noise current I_N is $\psi_0^{1/2}$, the probability density of the envelope R is

$$R\int_0^\infty rJ_0(Rr)J_0(Pr)J_0(Qr)e^{-\psi_0 r^2/2}\, dr \tag{3.10–21}$$

where $J_0(\ )$ is a Bessel function.

Curves showing the probability density and distribution function of R, when $Q = 0$, for various ratios of P/r.m.s. I_N are given.

10. In section 3.11 it is pointed out that the representations (2.8–1) and (2.8–6) of the noise current as the sum of a great number of sinusoidal components are not the only ones which may be used in deriving the results given in the preceding sections of Part III. The shot effect representation

$$I(t) = \sum_{-\infty}^{+\infty} F(t - t_k)$$

studied in Part I may also be used.

Part IV—Noise Through Non-Linear Devices

1. Suppose that the power spectrum of the voltage V applied to the square-law device

$$I = \alpha V^2 \tag{4.1–1}$$

is confined to a relatively narrow band. The total low-frequency output current $I_{t\ell}$ may be expressed as the sum

$$I_{t\ell} = I_{dc} + I_{\ell f} \tag{4.1–2}$$

where I_{dc} is the d.c. component and $I_{\ell f}$ is the variable component. When none of the low-frequency band is eliminated (by audio frequency filters)

$$I_{t\ell} = \frac{\alpha R^2}{2} \tag{4.1–6}$$

where R is the envelope of V. If V is of the form

$$V = V_N + P \cos pt + Q \cos qt, \tag{4.1–4}$$

where V_N is a noise voltage whose mean square value is ψ_0, then

$$I_{dc} = \alpha\left(\psi_0 + \frac{P^2}{2} + \frac{Q^2}{2}\right)$$

$$\overline{I_{\ell f}^2} = \alpha^2\left[\psi_0^2 + P^2\psi_0 + Q^2\psi_0 + \frac{P^2Q^2}{2}\right] \tag{4.1–16}$$

2. If instead of a square-law device we have a linear rectifier,

$$I = \begin{cases} 0 & V < 0 \\ \alpha V, & V > 0 \end{cases} \tag{4.2–1}$$

the total low-frequency output is

$$I_{t\ell} = \frac{\alpha R}{\pi} \tag{4.2–2}$$

When V is a sine wave plus noise, $V_N + P \cos pt$,

$$I_{dc} = \alpha\left(\frac{\psi_0}{2\pi}\right)^{1/2} {}_1F_1(-\tfrac{1}{2}; 1; -x) \tag{4.2-3}$$

$$\overline{I_{l\ell}^2} = \frac{\alpha^2}{\pi^2}(P^2 + 2\psi_0) \tag{4.2-6}$$

where ${}_1F_1$ is a hypergeometric function and

$$x = \frac{P^2}{2\psi_0} = \frac{\text{Ave. sine wave power}}{\text{Ave. noise power}} \tag{4.2-4}$$

When x is large

$$\overline{I_{l\ell}^2} \sim \frac{\alpha^2\psi_0}{\pi^2}\left[1 - \frac{1}{4x} \cdots\right] \tag{4.2-7}$$

If V consists of two sine waves plus noise, I_{dc} consists of a hypergeometric function of two variables. The equations running from (4.2–9) to (4.2–15) are concerned with this case. About the only simple equation is

$$\overline{I_{l\ell}^2} = \frac{\alpha^2}{\pi^2}[2\psi_0 + P^2 + Q^2] \tag{4.2-14}$$

3. The expressions (4.1–6) and (4.2–2) for $I_{l\ell}$ in terms of the envelope R of V, namely

$$\frac{\alpha R^2}{2} \quad \text{and} \quad \frac{\alpha R}{\pi},$$

are special cases of a more general result

$$I_{l\ell} = A_0(R) = \frac{1}{2\pi}\int_C F(iu)J_0(uR)\,du. \tag{4.3-11}$$

In this expression $J_0(uR)$ is a Bessel function. The path of integration C and the function $F(iu)$ are chosen so that the relation between I and V may be expressed as

$$I = \frac{1}{2\pi}\int_C F(iu)e^{iVu}\,du. \tag{4A-1}$$

A table giving $F(iu)$ and C for a number of common non-linear devices is shown in Appendix 4A.

If this relation is used to study the biased linear rectifier.

$$I = \begin{cases} 0, & V < B \\ V - B, & V > B \end{cases}$$

for the case in which V is $V_N + P \cos pt$, we find

$$I_{dc} \sim -\frac{B}{2} + \frac{P}{\pi} + \frac{B^2 + \psi_0}{2\pi P}$$
$$\overline{I_{\ell f}^2} \sim \frac{P^2 - B^2}{\pi^2 P^2}\psi_0 \qquad (4.3\text{–}17)$$

when $P \gg |B|$, $P^2 \gg \psi_0$ where ψ_0 is the mean square value of V_N.

4. When V is confined to a relatively narrow band and there are no audio-frequency filters, the probability density and all the associated statistical properties of $I_{t\ell}$ may be obtained by expressing $I_{t\ell}$ as a function of the envelope R of V and then using the probability density of R. When V is $V_N + P \cos pt + Q \cos qt$ this probability density is given by the integral, (3.10–21) (which is the integral containing three Bessel functions stated in the above summary of Part III). When V consists of three sine waves plus noise there are four J_0's in the integrand, and so on. Expressions for $\overline{R^n}$ when R has the above distribution are given by equations (3.10–25) and (3.10–27).

When audio-frequency filters remove part of the low-frequency band the statistical properties, except the mean square value, of the resulting current are hard to compute. In section 4.3 it is shown that as the output band is chosen narrower and narrower, the statistical properties of the output current approach those of a random noise current.

5. The sections in Part IV from 4.4 onward are concerned with the problem: Given a non-linear device and an input voltage consisting of noise alone or of a signal plus noise. What is the power spectrum of the output? A survey of the methods available for the solution of this problem is given in section 4.4.

6. When a noise voltage V_N with the power spectrum $w(f)$ is applied to the square-law device

$$I = \alpha V^2 \qquad (4.1\text{–}1)$$

the power spectrum of the output current I is, when $f \neq 0$,

$$W(f) = \alpha^2 \int_{-\infty}^{+\infty} w(x)w(f - x)\, dx \qquad (4.5\text{–}5)$$

where $w(-x)$ is defined to equal $w(x)$. The power spectrum of I when V is either $P \cos pt + V_N$ or

$$Q(1 + k \cos pt) \cos qt + V_N$$

is considered in the portion of section 4.5 containing equations (4.5–10) to (4.5–17).

7. A method discovered independently by Van Vleck and North shows that the correlation function $\Psi(\tau)$ of the output current for an unbiased linear rectifier is

$$\Psi(\tau) = \frac{\psi_\tau}{4} + \frac{\psi_0}{2} {}_2F_1\left[-\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{1}{2}; \frac{\psi_\tau^2}{\psi_0^2}\right] \tag{4.7-6}$$

where the input voltage is V_N. The correlation function $\psi(\tau)$ of V_N is denoted by ψ_τ and the mean square value of V_N is ψ_0. The power spectrum $W(f)$ of I may be obtained from

$$W(f) = 4 \int_0^\infty \Psi(\tau) \cos 2\pi f\tau \, d\tau \tag{4.6-1}$$

by expanding the hypergeometric function and integrating termwise using

$$G_n(f) = \int_0^\infty \psi_\tau^n \cos 2\pi f\tau \, d\tau. \tag{4C-1}$$

Appendix 4C is devoted to the problem of evaluating the integral for $G_n(f)$.

8. Another method of obtaining the correlation function $\psi(\tau)$ of I, termed the "characteristic function method," is explained in section 4.8. It is illustrated in section 4.9 where formulas for $\Psi(\tau)$ and $W(f)$ are developed when the voltage $P \cos pt + V_N$ is applied to a general non-linear device.

9. Several miscellaneous results are given in section 4.10. The characteristic function method is used to obtain the correlation function for a square-law device. The general formulas of section 4.9 are applied to the case of a ν^{th} law rectifier when the input noise spectrum has a normal law distribution. Some remarks are also made concerning the audio-frequency output of a linear rectifier when the input voltage V is

$$Q(1 + r \cos pt) \cos qt + V_N.$$

10. A discussion of the hypergeometric function ${}_1F_1(a; c; x)$, which often occurs in problems concerning a sine wave plus noise, is given in Appendix 4B.

PART I

THE SHOT EFFECT

The shot effect in vacuum tubes is a typical example of noise. It is due to fluctuations in the intensity of the stream of electrons flowing from the cathode to the anode. Here we analyze a simplified form of the shot effect.

1.1 The Probability of Exactly K Electrons Arriving at the Anode in Time T

The fluctuations in the electron stream are supposed to be random. We shall treat this randomness as follows. We count the number of electrons flowing in a long interval of time T measured in seconds. Suppose there are K_1. Repeating this counting process for many intervals all of length T gives a set of numbers $K_2, K_3 \cdots K_M$ where M is the total number of intervals. The average number ν, of electrons per second is defined as

$$\nu = \operatorname*{Lim}_{M\to\infty} \frac{K_1 + K_2 + \cdots + K_M}{MT} \tag{1.1-1}$$

where we assume that this limit exists. As M is increased with T being held fixed some of the K's will have the same value. In fact, as M increases the number of K's having any particular value will tend to increase. This of course is based on the assumption that the electron stream is a steady flow upon which random fluctuations are superposed. The probability of getting K electrons in a given trial is defined as

$$p(K) = \operatorname*{Lim}_{M\to\infty} \frac{\text{Number of trials giving exactly } K \text{ electrons}}{M} \tag{1.1-2}$$

Of course $p(K)$ also depends upon T. We assume that the randomness of the electron stream is such than the probability that an electron will arrive at the anode in the interval $(t, t + \Delta t)$ is $\nu\Delta t$ where Δt is such that $\nu\Delta t \ll 1$, and that this probability is independent of what has happened before time t or will happen after time $t + \Delta t$.

This assumption is sufficient to determine the expression for $p(K)$ which is

$$p(K) = \frac{(\nu T)^K}{K!} e^{-\nu T} \tag{1.1-3}$$

This is the "law of small probabilities" given by Poisson. One method of derivation sometimes used can be readily illustrated for the case $K = 0$. Thus, divide the interval, $(0, T)$ into M intervals each of length $\Delta t = \frac{T}{M}$. Δt is taken so small that $\nu\Delta t$ is much less than unity. (This is the "small probability" that an electron will arrive in the interval Δt). The probability that an electron will not arrive in the first sub-interval is $(1 - \nu\Delta t)$. The probability that one will not arrive in either the first or the second sub-interval is $(1 - \nu\Delta t)^2$. The probability that an electron will not arrive in any of the M intervals is $(1 - \nu\Delta t)^M$. Replacing M by $T/\Delta t$ and letting $\Delta t \to 0$ gives

$$p(0) = e^{-\nu T}$$

The expressions for $p(1), p(2), \cdots p(K)$ may be derived in a somewhat similar fashion.

1.2 Statement of Campbell's Theorem

Suppose that the arrival of an electron at the anode at time $t = 0$ produces an effect $F(t)$ at some point in the output circuit. If the output circuit is such that the effects of the various electrons add linearly, the total effect at time t due to all the electrons is

$$I(t) = \sum_{k=-\infty}^{+\infty} F(t - t_k) \tag{1.2-1}$$

where the k^{th} electron arrives at t_k and the series is assumed to converge.

Campbell's theorem[2] states that the average value of $I(t)$ is

$$\overline{I(t)} = \nu \int_{-\infty}^{+\infty} F(t)\, dt \tag{1.2-2}$$

and the mean square value of the fluctuation about this average is

$$\overline{(I(t) - \overline{I(t)})^2} = \nu \int_{-\infty}^{+\infty} F^2(t)\, dt \tag{1.2-3}$$

where ν is the average number of electrons arriving per second.

The statement of the theorem is not precise until we define what we mean by "average". From the form of the equations the reader might be tempted to think of a time average; e.g. the value

$$\operatorname*{Lim}_{T\to\infty} \frac{1}{T} \int_0^T I(t)\, dt \tag{1.2-4}$$

However, in the proof of the theorem the average is generally taken over a great many intervals of length T with t held constant. The process is somewhat similar to that employed in (1.1) and in order to make it clear we take the case of $\overline{I(t)}$ for illustration. We observe $I(t)$ for many, say M, intervals each of length T where T is large in comparison with the interval over which the effect $F(t)$ of the arrival of a single electron is appreciable. Let ${}_nI(t')$ be the value of $I(t)$, t' seconds after the beginning of the n^{th} interval. t' is equal to t plus a constant depending upon the beginning time of the interval. We put the subscript in front because we wish to reserve the usual place for another subscript later on. The value of $\overline{I(t')}$ is then defined as

$$\overline{I(t')} = \operatorname*{Limit}_{M\to\infty} \frac{1}{M} [{}_1I(t') + {}_2I(t') + \cdots + {}_MI(t')] \tag{1.2-5}$$

and this limit is assumed to exist. The mean square value of the fluctuation of $I(t')$ is defined in much the same way.

[2] *Proc. Camb. Phil. Soc.* 15 (1909), 117–136, 310–328. Our proof is similar to one given by J. M. Whittaker, *Proc. Camb. Phil. Soc.* 33 (1937), 451–458.

Actually, as the equations (1.2–2) and (1.2–3) of Campbell's theorem show, these averages and all the similar averages encountered later turn out to be independent of the time. When this is true and when the M intervals in (1.2–5) are taken consecutively the time average (1.2–4) and the average (1.2–5) become the same. To show this we multiply both sides of (1.2–5) by dt' and integrate from 0 to T:

$$\begin{aligned}\overline{I(t')} &= \operatorname*{Limit}_{M\to\infty} \frac{1}{MT} \sum_{m=1}^{M} \int_0^T {}_mI(t')\, dt' \\ &= \operatorname*{Limit}_{M\to\infty} \frac{1}{MT} \int_0^{MT} I(t)\, dt\end{aligned} \tag{1.2–6}$$

and this is the same as the time average (1.2–4) if the latter limit exists.

1.3 Proof of Campbell's Theorem

Consider the case in which exactly K electrons arrive at the anode in an interval of length T. Before the interval starts, we think of these K electrons as fated to arrive in the interval $(0, T)$ but any particular electron is just as likely to arrive at one time as any other time. We shall number these fated electrons from one to K for purposes of identification but it is to be emphasized that the numbering has nothing to do with the order of arrival. Thus, if t_k be the time of arrival of electron number k, the probability that t_k lies in the interval $(t, t + dt)$ is dt/T.

We take T to be very large compared with the range of values of t for which $F(t)$ is appreciably different from zero. In physical applications such a range usually exists and we shall call it Δ even though it is not very definite. Then, when exactly K electrons arrive in the interval $(0, T)$ the effect is approximately

$$I_K(t) = \sum_{k=1}^{K} F(t - t_k) \tag{1.3–1}$$

the degree of approximation being very good over all of the interval except within Δ of the end points.

Suppose we examine a large number M of intervals of length T. The number having exactly K arrivals will be, to a first approximation $M\, p(K)$ where $p(K)$ is given by (1.1–3). For a fixed value of t and for each interval having K arrivals, $I_K(t)$ will have a definite value. As $M \to \infty$, the average value of the $I_K(t)$'s, obtained by averaging over the intervals, is

$$\begin{aligned}\overline{I_K(t)} &= \int_0^T \frac{dt_1}{T} \cdots \int_0^T \frac{dt_K}{T} \sum_{k=1}^{K} F(t - t_k) \\ &= \sum_{k=1}^{K} \int_0^T \frac{dt_k}{T} F(t - t_k)\end{aligned} \tag{1.3–2}$$

and if $\Delta < t < T - \Delta$, we have effectively

$$\overline{I_K(t)} = \frac{K}{T} \int_{-\infty}^{+\infty} F(t)\, dt \tag{1.3–3}$$

If we now average $I(t)$ over all of the M intervals instead of only over those having K arrivals, we get, as $M \to \infty$,

$$\begin{aligned} \overline{I(t)} &= \sum_{K=0}^{\infty} p(K)\overline{I_K(t)} \\ &= \sum_{K=0}^{\infty} \frac{K}{T} \frac{(\nu T)^K}{K!} e^{-\nu T} \int_{-\infty}^{+\infty} F(t)\, dt \\ &= \nu \int_{-\infty}^{+\infty} F(t)\, dt \end{aligned} \tag{1.3–4}$$

and this proves the first part of the theorem. We have used this rather elaborate proof to prove the relatively simple (1.3–4) in order to illustrate a method which may be used to prove more complicated results. Of course, (1.3–4) could be established by noting that the integral is the average value of the effect produced by one arrival, the average being taken over one second, and that ν is the average number of arrivals per second.

In order to prove the second part, (1.2–3) of Campbell's theorem we first compute $\overline{I^2(t)}$ and use

$$\begin{aligned} \overline{(I(t) - \overline{I(t)})^2} &= \overline{I^2(t)} - 2\,\overline{I(t)\overline{I(t)}} + \overline{I(t)}^2 \\ &= \overline{I^2(t)} - \overline{I(t)}^2 \end{aligned} \tag{1.3–5}$$

From the definition (1.3–1) of $I_K(t)$,

$$I_K^2(t) = \sum_{k=1}^{K} \sum_{m=1}^{K} F(t - t_k)F(t - t_m)$$

Averaging this over all values of $t_1, t_2, \cdots t_K$ with t held fixed as in (1.3–2),

$$\overline{I_K^2(t)} = \sum_{k=1}^{K} \sum_{m=1}^{K} \int_0^T \frac{dt_1}{T} \cdots \int_0^T \frac{dt_K}{T} F(t - t_k)F(t - t_m)$$

The multiple integral has two different values. If $k = m$ its value is

$$\int_0^T F^2(t - t_k) \frac{dt_k}{T}$$

and if $k \neq m$ its value is

$$\int_0^T F(t - t_k) \frac{dt_k}{T} \int_0^T F(t - t_m) \frac{dt_m}{T}$$

Counting up the number of terms in the double sum shows that there are K of them having the first value and $K^2 - K$ having the second value. Hence, if $\Delta < t < T - \Delta$ we have

$$\overline{I_K^2(t)} = \frac{K}{T}\int_{-\infty}^{+\infty} F^2(t)\,dt + \frac{K(K-1)}{T^2}\left[\int_{-\infty}^{+\infty} F(t)\,dt\right]^2$$

Averaging over all the intervals instead of only those having K arrivals gives

$$\overline{I^2(t)} = \sum_{K=0}^{\infty} p(K)\,\overline{I_K^2(t)}$$

$$= \nu\int_{-\infty}^{+\infty} F^2(t)\,dt + \overline{I(t)}^2$$

where the summation with respect to K is performed as in (1.3–4), and after summation the value (1.3–4) for $\overline{I(t)}$ is used. Comparison with (1.3–5) establishes the second part of Campbell's theorem.

1.4 The Distribution of $I(t)$

When certain conditions are satisfied the proportion of time which $I(t)$ spends in the range I, $I + dI$ is $P(I)dI$ where, as $\nu \to \infty$, the probability density $P(I)$ approaches

$$\frac{1}{\sigma_I\sqrt{2\pi}}\,e^{-(I-\bar{I})^2/2\sigma_I^2} \qquad (1.4\text{–}1)$$

where $\bar{I}$ is the average of $I(t)$ given by (1.2–2) and the square of the standard deviation σ_I, i.e. the variance of $I(t)$, is given by (1.2–3). This normal distribution is the one which would be expected by virtue of the "central limit theorem" in probability. This states that, under suitable conditions, the distribution of the sum of a large number of random variables tends toward a normal distribution whose variance is the sum of the variances of the individual variables. Similarly the average of the normal distribution is the sum of the averages of the individual variables.

So far, we have been speaking of the limiting form of the probability density $P(I)$. It is possible to write down an explicit expression for $P(I)$, which, however, is quite involved. From this expression the limiting form may be obtained. We now obtain this expression. In line with the discussion given of Campbell's theorem, we seek the probability density $P(I)$ of the values of $I(t)$ observed at t seconds from the beginning of each of a large number, M, of intervals, each of length T.

Probability that $I(t)$ lies in range $(I, I + dI)$

$$= \sum_{K=0}^{\infty} \text{(Probability of exactly } K \text{ arrivals)} \times$$

(Probability that if there are exactly K arrivals, $I_K(t)$ lies in $(I, I + dI)$).

Denoting the last probability in the summation by $P_K(I)dI$, using notation introduced earlier, and cancelling out the factor dI gives

$$P(I) = \sum_{K=0}^{\infty} p(K)P_K(I) \tag{1.4-2}$$

We shall compute $P_K(I)$ by the method of "characteristic functions"[3] from the definition

$$I_K(t) = \sum_{k=1}^{K} F(t - t_k) \tag{1.3-1}$$

of $I_K(t)$. The method will be used in its simplest form: the probability that the sum

$$x_1 + x_2 + \cdots + x_K$$

of K independent random variables lies between X and $X + dX$ is

$$dX \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-ixu} \prod_{k=1}^{K} (\text{average value of } e^{ix_k u})\, du \tag{1.4-3}$$

The average value of $e^{ix_k u}$, i.e., the characteristic function of the distribution of x_k, is obtained by averaging over the values of x_k. Although this is the simplest form of the method it is also the least general in that the integral does not converge for some important cases. The distribution which gives a probability of $\frac{1}{2}$ that $x_k = -1$ and $\frac{1}{2}$ that $x_k = +1$ is an example of such a case. However, we may still use (1.4–3) formally in such cases by employing the relation

$$\int_{-\infty}^{+\infty} e^{-iau}\, du = 2\pi\delta(a) \tag{1.4-4}$$

where $\delta(a)$ is zero except at $a = 0$ where it is infinite and its integral from $a = -\epsilon$ to $a = +\epsilon$ is unity where $\epsilon > 0$.

When we identify x_k with $F(t - t_k)$ we see that the average value of $e^{ix_k u}$ is

$$\frac{1}{T} \int_0^T \exp\,[iuF(t - t_k)]\, dt_k$$

[3] The essentials of this method are due to Laplace. A few remarks on its history are given by E. C. Molina, *Bull. Amer. Math. Soc.*, 36 (1930), pp. 369–392. An account of the method may be found in any one of several texts on probability theory. We mention "Random Variables and Probability Distributions," by H. Cramér, Camb. Tract in Math. and Math. Phys. No. 36 (1937), Chap. IV. Also "Introduction to Mathematical Probability," by J. V. Uspensky, McGraw-Hill (1937), pages 240, 264, and 271–278.

All of the K characteristic functions are the same and hence, from (1.4–3), $P_K(I)dI$ is

$$dI \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iIu} \left(\frac{1}{T} \int_0^T \exp [iuF(t - \tau)] \, d\tau \right)^K du$$

Although in deriving this relation we have taken $K > 0$, it also holds for $K = 0$ (provided we use (1.4–4)). In this case $P_0(I) = \delta(I)$, because $I = 0$ when no electrons arrive.

Inserting our expression for $P_K(I)$ and the expression (1.1–3) for $p(K)$ in (1.4–2) and performing the summation gives

$$P(I) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \left(-iIu - \nu T + \nu \int_0^T \exp [iuF(t - \tau)] \, d\tau \right) du \quad (1.4\text{–}5)$$

The first exponential may be simplified somewhat. Using

$$\nu T = \nu \int_0^T d\tau$$

permits us to write

$$-\nu T + \nu \int_0^T \exp [iuF(t - \tau)] \, d\tau = \nu \int_0^T (\exp [iuF(t - \tau)] - 1) \, d\tau$$

Suppose that $\Delta < t < T - \Delta$ where Δ is the range discussed in connection with equation (1.3–1). Taking $| F(t - \tau) | = 0$ for $| t - \tau | > \Delta$ then enables us to write the last expression as

$$\nu \int_{-\infty}^{+\infty} [e^{iuF(t)} - 1] \, dt \quad (1.4\text{–}6)$$

Placing this in (1.4–5) yields the required expression for $P(I)$:

$$P(I) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \left(-iIu + \nu \int_{-\infty}^{+\infty} [e^{iuF(t)} - 1] \, dt \right) du \quad (1.4\text{–}7)$$

An idea of the conditions under which the normal law (1.4–1) is approached may be obtained from (1.4–7) by expanding (1.4–6) in powers of u and determining when the terms involving u^3 and higher powers of u may be neglected. This is taken up for a slightly more general form of current in section 1.6.

1.5 Extension of Campbell's Theorem

In section 1.2 we have stated Campbell's theorem. Here we shall give an extension of it. In place of the expression (1.2–1) for the $I(t)$ of the shot effect we shall deal with the current

$$I(t) = \sum_{k=-\infty}^{+\infty} a_k F(t - t_k) \tag{1.5–1}$$

where $F(t)$ is the same sort of function as before and where $\cdots a_1, a_2, \cdots a_k, \cdots$ are independent random variables all having the same distribution. It is assumed that all of the moments $\overline{a^n}$ exist, and that the events occur at random

The extension states that the nth semi-invariant of the probability density $P(I)$ of I, where I is given by (1.5–1), is

$$\lambda_n = \nu \overline{a^n} \int_{-\infty}^{+\infty} [F(t)]^n \, dt \tag{1.5–2}$$

where ν is the expected number of events per second. The semi-invariants of a distribution are defined as the coefficients in the expansion

$$\log_e (\text{ave. } e^{iIu}) = \sum_{n=1}^{N} \frac{\lambda_n}{n!} (iu)^n + o(u^N) \tag{1.5–3}$$

i.e. as the coefficients in the expansion of the logarithm of the characteristic function. The λ's are related to the moments of the distribution. Thus if $m_1, m_2, \cdots$ denote the first, second $\cdots$ moments about zero we have

$$\text{ave. } e^{iIu} = 1 + \sum_{n=1}^{N} \frac{m_n}{n!} (iu)^n + o(u^N)$$

By combining this relation with the one defining the λ's it may be shown that

$$\begin{aligned} \overline{I} &= m_1 = \lambda_1 \\ \overline{I^2} &= m_2 = \lambda_2 + \lambda_1 m_1 \\ \overline{I^3} &= m_3 = \lambda_3 + 2\lambda_2 m_1 + \lambda_1 m_2 \\ &\cdots\cdots\cdots \end{aligned}$$

It follows that $\lambda_1 = \overline{I}$ and $\lambda_2 = \text{ave. } (I - \overline{I})^2$. Hence (1.5–2) yields the original statement of Campbell's theorem when we set n equal to one and two and also take all the a's to be unity.

The extension follows almost at once from the generalization of expression (1.4–7) for the probability density $P(I)$. By proceeding as in section 1.4 and identifying x_k with $a_k F(t - t_k)$ we see that

$$\text{ave. } e^{ix_k u} = \frac{1}{T} \int_{-\infty}^{+\infty} q(a) \, da \int_0^T \exp [iuaF(t - t_k)] \, dt_k$$

where $q(a)$ is the probability density function for the a's. It turns out that the probability density $P(I)$ of I as defined by (1.5–1) is

$$P(I) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-iIu + \nu \int_{-\infty}^{+\infty} q(a)\, da \int_{-\infty}^{+\infty} [e^{iuaF(t)} - 1]\, dt\right) du \quad (1.5\text{–}4)$$

The logarithm of the characteristic function of $P(I)$ is, from (1.5–4),

$$\nu \int_{-\infty}^{+\infty} q(a)\, da \int_{-\infty}^{+\infty} [e^{iuaF(t)} - 1]\, dt = \sum_{n=1}^{\infty} \frac{(iu)^n}{n!} \nu \int_{-\infty}^{+\infty} q(a)\, da\, a^n \int_{-\infty}^{+\infty} F^n(t)\, dt$$

Comparison with the series (1.5–3) defining the semi-invariants gives the extension of Campbell's theorem stated by (1.5–2).

Other extensions of Campbell's theorem may be made. For example, suppose in the expression (1.5-1) for $I(t)$ that $t_1, t_2, \cdots t_k, \cdots$ while still random variables, are no longer necessarily distributed according to the laws assumed above. Suppose now that the probability density $p(x)$ is given where x is the interval between two successive events:

$$t_2 = t_1 + x_1 \quad (1.5\text{–}5)$$

$$t_3 = t_2 + x_2 = t_1 + x_1 + x_2$$

and so on. For the case treated above

$$p(x) = \nu e^{-\nu x}. \quad (1.5\text{–}6)$$

We assume that the expected number of events per second is still ν.

Also we take the special, but important, case for which

$$F(t) = 0, \qquad t < 0 \quad (1.5\text{–}7)$$

$$F(t) = e^{-\alpha t}, \qquad t > 0.$$

For a very long interval extending from $t = t_1$ to $t = T + t_1$ inside of which there are exactly K events we have, if t is not near the ends of the interval,

$$I(t) = a_1 F(t - t_1) + a_2 F(t - t_1 - x_1) + \cdots + a_{K+1} F(t - t_1 - x_1 \cdots - x_K)$$

$$= a_1 F(t') + a_2 F(t' - x_1) + \cdots + a_{K+1} F(t' - x_1 - \cdots - x_K)$$

where $q(a)$ is the probability density function for the a's. It turns out that the probability density $P(I)$ of I as defined by (1.5–1) is

$$P(I) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left(-iIu + \nu \int_{-\infty}^{+\infty} q(a)\, da \int_{-\infty}^{+\infty} [e^{iuaF(t)} - 1]\, dt\right) du \quad (1.5\text{–}4)$$

The logarithm of the characteristic function of $P(I)$ is, from (1.5–4),

$$\nu \int_{-\infty}^{+\infty} q(a)\, da \int_{-\infty}^{+\infty} [e^{iuaF(t)} - 1]\, dt = \sum_{n=1}^{\infty} \frac{(iu)^n}{n!} \nu \int_{-\infty}^{+\infty} q(a)\, da\, a^n \int_{-\infty}^{+\infty} F^n(t)\, dt$$

Comparison with the series (1.5–3) defining the semi-invariants gives the extension of Campbell's theorem stated by (1.5–2).

Other extensions of Campbell's theorem may be made. For example, suppose in the expression (1.5-1) for $I(t)$ that $t_1, t_2, \cdots t_k, \cdots$ while still random variables, are no longer necessarily distributed according to the laws assumed above. Suppose now that the probability density $p(x)$ is given where x is the interval between two successive events:

$$t_2 = t_1 + x_1 \quad (1.5\text{–}5)$$

$$t_3 = t_2 + x_2 = t_1 + x_1 + x_2$$

and so on. For the case treated above

$$p(x) = \nu e^{-\nu x}. \quad (1.5\text{–}6)$$

We assume that the expected number of events per second is still ν.

Also we take the special, but important, case for which

$$F(t) = 0, \qquad t < 0 \quad (1.5\text{–}7)$$

$$F(t) = e^{-\alpha t}, \qquad t > 0.$$

For a very long interval extending from $t = t_1$ to $t = T + t_1$ inside of which there are exactly K events we have, if t is not near the ends of the interval,

$$I(t) = a_1 F(t - t_1) + a_2 F(t - t_1 - x_1) + \cdots + a_{K+1} F(t - t_1 - x_1 \cdots - x_K)$$

$$= a_1 F(t') + a_2 F(t' - x_1) + \cdots + a_{K+1} F(t' - x_1 - \cdots - x_K)$$

$$I^2(t) = a_1^2F^2(t') + a_2^2F^2(t' - x_1) + \cdots + a_{K+1}^2F^2(t' - x_1 \cdots - x_K)$$
$$+ 2a_1a_2F(t')F(t' - x_1) + \cdots + 2a_1a_{K+1}F(t')F(t' - x_1 \cdots - x_K)$$
$$+ 2a_2a_3F(t' - x_1)F(t' - x_1 - x_2) + \cdots + \cdots$$

where $t' = t - t_1$. If we integrate $I^2(t)$ over the entire interval $0 < t' < T$ and drop the primes we get approximately

$$\int_0^T I^2(t)dt = (a_1^2 + \cdots + a_{K+1}^2)\varphi(0)$$
$$+ 2a_1a_2\varphi(x_1) + 2a_1a_3\varphi(x_1 + x_2) + \cdots + 2a_1a_{K+1}\varphi(x_1 + \cdots + x_K)$$
$$+ 2a_2a_3\varphi(x_2) + \cdots + \cdots + 2a_Ka_{K+1}\varphi(x_K)$$

where

$$\varphi(x) = \int_{-\infty}^{+\infty} F(t)F(t - x)\,dx$$

When we divide both sides by T and consider K and T to be very large,

$$\frac{K}{T}\frac{a_1^2 + \cdots + a_{K+1}^2}{K}\varphi(0) \approx \nu\bar{a^2}\varphi(0)$$

$$\frac{1}{T}[a_1a_2\varphi(x_1) + a_2a_3\varphi(x_2) + \cdots + a_Ka_{K+1}\varphi(x_K)] = \frac{K}{T}\text{ average } a_ka_{k+1}\varphi(x_k)$$

$$\approx \nu\bar{a}^2\int_0^\infty \varphi(x)p(x)\,dx$$

$$\frac{1}{T}[a_1a_3\varphi(x_1 + x_2) + \cdots] = \frac{K-1}{T}\text{ ave. } a_ka_{k+3}\varphi(x_k + x_{k+1})$$

$$\approx \nu\bar{a}^2\int_0^\infty dx_1\int_0^\infty dx_2\,p(x_1)p(x_2)\varphi(x_1 + x_2)$$

Consequently

$$\overline{I^2(t)} = \operatorname*{Lim}_{T\to\infty}\frac{1}{T}\int_0^T I^2(t)\,dt$$
$$= \nu\bar{a^2}\varphi(0) + 2\nu\bar{a}^2\left[\int_0^\infty p(x)\varphi(x)\,dx\right.$$
$$\left. + \int_0^\infty dx_1\int_0^\infty dx_2\,p(x_1)p(x_2)\varphi(x_1 + x_2) + \cdots\right]$$

For our special exponential form (1.5–7) for $F(t)$,

$$\varphi(x) = \frac{e^{-\alpha x}}{2\alpha}$$

and the multiple integrals occurring in the expression for $\overline{I^2(t)}$ may be written in terms of powers of

$$q = \int_0^\infty p(x)e^{-\alpha x}\,dx \tag{1.5-8}$$

Thus

$$2\alpha\overline{I^2(t)} = \nu\bar{a}^2 + 2\bar{a}^2\nu\,\frac{q}{1-q}$$

and since

$$\bar{I}(t) = \nu\bar{a}\int_{-\infty}^{+\infty} F(t)\,dt = \nu\bar{a}/\alpha$$

we have

$$\overline{I^2(t)} - \overline{I(t)}^2 = \frac{\nu\bar{a}^2}{2\alpha} + \left(\frac{\nu\bar{a}}{\alpha}\right)^2\left[\frac{\alpha q}{\nu(1-q)} - 1\right] \tag{1.5-9}$$

Equations (1.5–8) and (1.5–9) give us an extension of Campbell's theorem subject to the restrictions discussed in connection with equations (1.5–5) and (1.5–7). Other generalizations have been made[4] but we shall leave the subject here. The reader may find it interesting to verify that (1.5–9) gives the correct answer when $p(x)$ is given by (1.5–6), and also to investigate the case when the events are spaced equally.

1.6 Approach of Distribution of I to a Normal Law

In section 1.5 we saw that the probability density $P(I)$ of the noise current I may be expressed formally as

$$P(I) + \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\left[-iIu + \sum_{n=1}^{\infty}(iu)^n\lambda_n/n!\right]du \tag{1.6-1}$$

where λ_n is the nth semi-invariant given by (1.5–2). By setting

$$\lambda_2 = \sigma^2$$

$$x = \frac{I-\lambda_1}{\sigma} = \frac{I-\bar{I}}{\sigma} \tag{1.6-2}$$

[4] See E. N. Rowland, *Proc. Camb. Phil. Soc.* 32 (1936), 580–597. He extends the theorem to the case where there are two functions instead of a single one, which we here denote by $I(t)$. According to a review in the Zentralblatt für Math., 19, p. 224, Khintchine in the Bull. Acad. Sci. URSS, sér. Math. Nr. 3 (1938), 313–322, has continued and made precise the earlier work of Rowland.

expanding

$$\exp \sum_{n=3}^{\infty} (iu)^n \lambda_n / n!$$

as a power series in u, integrating termwise using

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} (iu\sigma)^n \exp\left[-iu\sigma x - \frac{u^2\sigma^2}{2}\right] du = (-)^n \sigma^{-1} \varphi^{(n)}(x),$$

$$\varphi^{(n)}(x) = \frac{1}{\sqrt{2\pi}} \frac{d^n}{dx^n} e^{-x^2/2}$$

and finally collecting terms according to their order in powers of $\nu^{-1/2}$, gives

$$P(I) \sim \sigma^{-1}\varphi^{(0)}(x) - \frac{\lambda_3 \sigma^{-4}}{3!} \varphi^{(3)}(x) + \left[\frac{\lambda_4 \sigma^{-5}}{4!} \varphi^{(4)}(x) + \frac{\lambda_3^2 \sigma^{-7}}{72} \varphi^{(6)}(x)\right] + \cdots \tag{1.6-3}$$

The first term is $0(\nu^{-1/2})$, the second term is $0(\nu^{-1})$, and the term within brackets is $0(\nu^{-3/2})$. This is Edgeworth's series.[5] The first term gives the normal distribution and the remaining terms show how this distribution is approached as $\nu \to \infty$.

1.7 The Fourier Components of $I(t)$

In some analytical work noise current is represented as

$$I(t) = \frac{a_0}{2} + \sum_{n=1}^{N} \left(a_n \cos \frac{2\pi nt}{T} + b_n \sin \frac{2\pi nt}{T}\right) \tag{1.7-1}$$

where at a suitable place in the work T and N are allowed to become infinite. The coefficients a_n and b_n, $1 \leq n \leq N$, are regarded as independent random variables distributed about zero according to a normal law.

It appears that the association of (1.7–1) with a sequence of disturbances occurring at random goes back many years. Rayleigh[6] and Gouy suggested that black-body radiation and white light might both be regarded as sequences of irregularly distributed impulses.

Einstein[7] and von Laue have discussed the normal distribution of the coefficients in (1.7–1) when it is used to represent black-body radiation, this radiation being the resultant produced by a great many independent os-

[5] See, for example, pp. 86–87, in "Random Variables and Probability Distributions" by H. Cramér, *Cambridge Tract* No. 36 (1937).

[6] *Phil. Mag.* Ser. 5, Vol. 27 (1889) pp. 460–469.

[7] A. Einstein and L. Hopf, *Ann. d. Physik* 33 (1910) pp. 1095–1115.
M. V. Laue, *Ann. d. Physik* 47 (1915) pp. 853–878.
A. Einstein, *Ann. d. Physik* 47 (1915) pp. 879–885.
M. V. Laue, *Ann. d. Physik* 48 (1915) pp. 668–680.
I am indebted to Prof. Goudsmit for these references.

cillators. Some argument arose as to whether the coefficients in (1.7–1) were statistically independent or not. It was finally decided that they are independent.

The shot effect current has been represented in this way by Schottky.[8] The Fourier series representation has been discussed by H. Nyquist[9] and also by Goudsmit and Weiss. Remarks made by A. Schuster[10] are equivalent to the statement that a_n and b_n are distributed normally.

In view of this wealth of information on the subject it may appear superfluous to say anything about it. However, for the sake of completeness, we shall outline the thoughts which lead to (1.7–1).

In line with our usual approach to the shot effect, we suppose that exactly K electrons arrive during the interval $(0, T)$, so that the noise current for the interval is

$$I_K(t) = \sum_{k=1}^{K} F(t - t_k) \tag{1.7–2}$$

The coefficients in the Fourier series expansion of $I_K(t)$ over the interval $(0, T)$ are a_{nK} and b_{nK} where

$$\begin{aligned} a_{nK} - ib_{nK} &= \frac{2}{T}\sum_{k=1}^{K}\int_0^T F(t - t_k)\exp\left[-i\frac{2\pi nt}{T}\right]dt \\ &\approx \frac{2}{T}\sum_{k=1}^{K}\int_{-\infty}^{+\infty} F(t)\exp\left[-i\frac{2\pi n}{T}(t + t_k)\right]dt \\ &= R_n e^{-i\varphi_n}\sum_{k=1}^{K} e^{-in\theta_k} \end{aligned} \tag{1.7–3}$$

In this expression

$$\begin{aligned} \theta_k &= \frac{2\pi t_k}{T} \\ R_n e^{-i\varphi_n} &= C_n - iS_n = \frac{2}{T}\int_{-\infty}^{+\infty} F(t)e^{-i2\pi nt/T}\,dt \end{aligned} \tag{1.7–4}$$

In the earlier sections the arrival times $t_1, t_2, \cdots t_K$ were regarded as K independent random variable each distributed uniformly over the interval (O, T). Hence the θ_k's may be regarded as random variables distributed uniformly over the interval 0 to 2π.

Incidentally, it will be noted that in (1.7–3) there occurs the sum of K randomly oriented unit vectors. When K becomes very large, as it does

[8] *Ann. d. Physik*, 57 (1918) pp. 541–567.

[9] Unpublished Memorandum, "Fluctuations in Vacuum Tube Noise and the Like," March 17, 1932.

[10] Investigation of Hidden Periodicities, Terrestrial Magnetism, 3 (1898), pp. 13–41. See especially propositions 1 and 2 on page 26 of Schuster's paper.

when $\nu \to \infty$, it is known that the real and imaginary parts of this sum are random variables, which tend to become independent and normally distributed about zero. This suggests the manner in which the normal distribution of the coefficients arises. Averaging over the θ_k's in (1.7–3) gives when $n > 0$

$$\bar{a}_{nK} = \bar{b}_{nK} = 0 \tag{1.7-5}$$

Some further algebra gives

$$\begin{aligned} \overline{a_{nK}^2} &= \overline{b_{nK}^2} = \frac{K}{2} R_n^2 \\ \overline{a_{nK} b_{nK}} &= \overline{a_{nK} a_{mK}} = \overline{b_{nK} b_{mK}} = 0 \end{aligned} \tag{1.7-6}$$

where $n \neq m$ and $n, m > 0$.

So far, we have been considering the case of exactly K arrivals in our interval of length T. Now we pass to the general case of any number of arrivals by making use of formulas analogous to

$$\overline{a_n^2} = \sum_{K=0}^{\infty} p(K) \overline{a_{nK}^2} \tag{1.7-7}$$

as has been done in section 1.3. Thus, for $n > 0$,

$$\begin{aligned} \bar{a}_n &= \bar{b}_n = 0 \\ \overline{a_n^2} &= \overline{b_n^2} = \frac{\nu T}{2} R_n^2 = \sigma_n^2 \\ \overline{a_n b_n} &= \overline{a_n a_m} = \overline{b_n b_m} = 0, \qquad n \neq m \end{aligned} \tag{1.7-8}$$

In the second line we have used σ_n to denote the standard deviation of a_n and b_n. We may put the expression for σ_n^2 in a somewhat different form by writing

$$f_n = \frac{n}{T} = n\Delta f, \qquad \Delta f = \frac{1}{T} \tag{1.7-9}$$

where f_n is the frequency of the nth component. Using (1.7-4),

$$\sigma_n^2 = 2\nu\Delta f \left| \int_{-\infty}^{+\infty} F(t) e^{-i2\pi f_n t}\, dt \right|^2 \tag{1.7-10}$$

Thus, σ_n^2 is proportional to ν/T.

The probability density function $P(a_1, \cdots a_N, b_1, \cdots b_N)$ for the $2N$ coefficients, $a_1, \cdots a_N, b_1, \cdots b_N$ may be derived in much the same fashion as was the probability density of the noise current in section 1.4. Here N

is arbitrary but fixed. The expression analogous to (1.4–5) is the $2N$ fold integral

$$P(a_1, \cdots, b_N) = (2\pi)^{-2N} \int_{-\infty}^{+\infty} du_1 \cdots \int_{-\infty}^{+\infty} dv_N \qquad (1.7\text{–}11)$$
$$\exp\left[-i(a_1 u_1 + \cdots + b_N v_N) - \nu T + \nu T E\right]$$

where

$$E = \frac{1}{2\pi} \int_0^{2\pi} d\theta \exp\left[i \sum_{n=1}^{N} (u_n C_n + v_n S_n) \cos n\theta + (v_n C_n - u_n S_n) \sin n\theta \right] \qquad (1.7\text{–}12)$$

in which $C_n - iS_n$ is defined as the Fourier transform (1.7–4) of $F(t)$.

The next step is to show that (1.7–11) approaches a normal law in $2N$ dimensions as $\nu \to \infty$. This appears to be quite involved. It will be noted that the integrand in the integral defining E is composed of N factors of the form

$$\exp\left[i\rho_n \cos(n\theta - \psi_n)\right]$$
$$= J_0(\rho_n) + 2i \cos(n\theta - \psi_n) J_1(\rho_n) - 2 \cos(2n\theta - 2\psi_n) J_2(\rho_n) + \cdots$$

where

$$\rho_n^2 = (u_n^2 + v_n^2)(C_n^2 + S_n^2) = \frac{2}{\nu T} \sigma_n^2 (u_n^2 + v_n^2).$$

As ν becomes large, it turns out that the integral (1.7–11) for the probability density obtains most of its contributions from small values of u and v. By substituting the product of the Bessel function series in the integral for E and integrating we find

$$E = \prod_{n=1}^{N} J_0(\rho_n) + A + B + C$$

where A is the sum of products such as

$$-2i \cos(\psi_{k+\ell} - \psi_k - \psi_\ell) J_1(\rho_k) J_1(\rho_\ell) J_1(\rho_{k+\ell}) \text{ times } N - 3 \; J_0\text{'s}$$

in which $0 < k \leq l$ and $2 \leq k + l \leq N$. Similarly B is the sum of products of the form

$$-2i \cos(\psi_{2k} - 2\psi_k) J_1(\rho_{2k}) J_2(\rho_k) \text{ times } N - 2 \; J_0\text{'s}$$

C consists of terms which give fourth and higher powers in u and v. There are roughly $N^2/4$ terms of form A and $N/2$ terms of form B.

Expanding the Bessel functions, neglecting all powers above the third and

proceeding as in section 1.4, will give us the normal distribution plus the first correction term. It is rather a messy affair. An idea of how it looks may be obtained by taking the special case in which $F(t)$ is an even function of t and neglecting terms of type B. Then

$$P(a_1, \cdots a_N, b_1, \cdots b_N) = (1 + \eta) \prod_{n=1}^{N} \frac{e^{-(x_n^2+y_n^2)/2}}{2\pi\sigma_n^2} \quad (1.7\text{–}12)$$

where

$$x_n = \frac{a_n}{\sigma_n}, \qquad y_n = \frac{b_n}{\sigma_n}$$

$$\eta = (2\nu T)^{-1/2} \sum_{k,\ell} [x_{k+\ell}(x_k x_\ell - y_k y_\ell) + 2\, y_{k+\ell} y_k y_\ell] \quad (1.7\text{–}13)$$

and the summation extends over $2 \leq k + l \leq N$ with $k \leq l$.

It is seen that if T and N are held constant, the correction term η approaches zero as ν becomes very large. A very rough idea of the magnitude of η may be obtained by assuming that unity is a representative value of the x's and y's. Further assuming that there are N^2 terms in the summation each one of which may be positive or negative suggests that magnitude of the sum is of the order of N. Hence we might expect to find that η is of the order of $N(2\nu T)^{-1/2}$.

PART II

POWER SPECTRA AND CORRELATION FUNCTIONS

2.0 Introduction

A theory for analyzing functions of time, t, which do not die down and which remain finite as t approaches infinity has gradually been developed over the last sixty years. A few words of its history together with an extensive bibliography are given by N. Wiener in his paper on "Generalized Harmonic Analysis".[11] G. Gouy, Lord Rayleigh and A. Schuster were led to study this problem in their investigations of such things as white light and noise. Schuster[12] invented the "periodogram" method of analysis which has as its object the discovery of any periodicities hidden in a continuous curve representing meteorological or economic data.

[11] *Acta Math.*, Vol. 55, pp. 117–258 (1930). See also "Harmonic Analysis of Irregular Motion," *Jour. Math. and Phys.* 5 (1926) pp. 99–189.

[12] The periodogram was first introduced by Schuster in reference 10 cited in Section 1.7. He later modified its definition in the Trans. Camb. Phil. Soc. 18 (1900), pp. 107–135, and still later redefined it in "The Periodogram and its Optical Analogy," Proc. Roy. Soc., London, Ser. A, 77 (1906) pp. 136–140. In its final form the periodogram is equivalent to $\frac{1}{2}w(f)$, where $w(f)$ is the power spectrum defined in Section 2.1, plotted as a function of the period $T = (2\pi f)^{-1}$.

The correlation function, which turns cut to be a very useful tool, apparently was introduced by G. I. Taylor.[13] Recently it has been used by quite a few writers[14] in the mathematical theory of turbulence.

In section 2.1 the power spectrum and correlation function of a specific function, such as one given by a curve extending to $t = \infty$, are defined by equations (2.1–3) and (2.1–4) respectively. That they are related by the Fourier inversion formulae (2.1–5) and (2.1–6) is merely stated; the discussion of the method of proof being delayed until sections 2.3 and 2.4. In section 2.3 a discussion based on Fourier series is given and in section 2.4 a parallel treatment starting with Parseval's integral theorem is set forth. The results as given in section 2.1 have to be supplemented when the function being analyzed contains a d.c. or periodic components. This is taken up in section 2.2.

The first four sections deal with the analysis of a specific function of t. However, most of the applications are made to functions which behave as though they are more or less random in character. In the mathematical analysis this randomness is introduced by assuming the function of t to be also a function of suitable parameters, and then letting these parameters be random variables. This question is taken up in section 2.5. In section 2.6 the results of 2.5 are applied to determine the average power spectrum and the average correlation function of the shot effect current. The same thing is done in 2.7 for a flat top wave, the tops (and bottoms) being of random length. The case in which the intervals are of equal length but the sign of the wave is random is also discussed in 2.7. The representation of the noise current as a trigonometrical series with random variable coefficients is taken up in 2.8. The last two sections 2.9 and 2.10 are devoted to probability theory. The normal law and the central limit theorem, respectively, are discussed.

2.1 Some Results of Generalized Harmonic Analysis

We shall first state the results which we need, and then show that they are plausible by methods which are heuristic rather than rigorous. Suppose that $I(t)$ is one of the functions mentioned above. We may think of it as being specified by a curve extending from $t = -\infty$ to $t = \infty$. $I(t)$ may be regarded as composed of a great number of sinusoidal components whose frequencies range from 0 to $+\infty$. It does not necessarily have to be a noise current, but if we think of it as such, then, in flowing through a resistance of one ohm it will dissipate a certain average amount of power, say ρ watts.

[13] Diffusion by Continuous Movements, *Proc. Lond. Math. Soc.*, Ser. 2, 20, pp. 196–212 (1920).

[14] See the text "Modern Developments in Fluid Dynamics" edited by S. Goldstein, Oxford (1938).

That portion of ρ arising from the components having frequencies between f and $f + df$ will be denoted by $w(f)df$, and consequently

$$\rho = \int_0^\infty w(f)df \tag{2.1-1}$$

Since $w(f)$ is the spectrum of the average power we shall call it the "power spectrum" of $I(t)$. It has the dimensions of energy and on this account is frequently called the "energy-frequency spectrum" of $I(t)$. A mathematical formulation of this discussion leads to a clear cut definition of $w(f)$.

Let $\Phi(t)$ be a function of t, which is zero outside the interval $0 \leq t \leq T$ and is equal to $I(t)$ inside the interval. Its spectrum $S(f)$ is given by

$$S(f) = \int_0^T I(t)e^{-2\pi ift}\,dt \tag{2.1-2}$$

The spectrum of the power, $w(f)$, is defined as

$$w(f) = \operatorname*{Limit}_{T\to\infty} \frac{2|S(f)|^2}{T} \tag{2.1-3}$$

where we consider only values of $f > 0$ and assume that this limit exists. This is substantially the definition of $w(f)$ given by J. R. Carson[15] and is useful when $I(t)$ has no periodic terms and no d.c. component. In the latter case (2.1–3) must either be supplemented by additional definitions or else a somewhat different method of approach used. These questions will be discussed in section 2.2.

The correlation function $\psi(\tau)$ of $I(t)$ is defined by the limit

$$\psi(\tau) = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I(t)I(t+\tau)\,dt \tag{2.1-4}$$

which is assumed to exist. $\psi(\tau)$ is closely related to the correlation coefficients used in statistical theory to measure the correlation of two random variables. In the present case the value of $I(t)$ at time t is one variable and its value at a different time $t + \tau$ is the other variable.

The spectrum of the power $w(f)$ and the correlation function $\psi(\tau)$ are related by the equations

$$w(f) = 4\int_0^\infty \psi(\tau)\cos 2\pi f\tau\,d\tau \tag{2.1-5}$$

$$\psi(\tau) = \int_0^\infty w(f)\cos 2\pi f\tau\,df \tag{2.1-6}$$

[15] "The Statistical Energy-Frequency Spectrum of Random Disturbances," *B.S.T.J.*, Vol. 10, pp. 374–381 (1931).

It is seen that $\psi(\tau)$ is an even function of τ and that

$$\psi(0) = \rho \tag{2.1-7}$$

When either $\psi(\tau)$ or $w(f)$ is known the other may be obtained provided the corresponding integral converges.

2.2 Power Spectrum for D.C. and Periodic Components

As mentioned in section 2.1, when $I(t)$ has a d.c. or a periodic component the limit in the definition (2.1–3) for $w(f)$ does not exist for f equal to zero or to the frequency of the periodic component. Perhaps the most satisfactory method of overcoming this difficulty, from the mathematical point of view, is to deal with the integral of the power spectrum.[16]

$$\int_0^f w(g)\, dg \tag{2.2-1}$$

instead of with $w(f)$ itself.

The definition (2.1–4) for $\psi(\tau)$ still holds. If, for example,

$$I(t) = A + C \cos (2\pi f_0 t - \varphi) \tag{2.2-2}$$

$\psi(\tau)$ as given by (2.1–4) is

$$\psi(\tau) = A^2 + \frac{C^2}{2} \cos 2\pi f_0 \tau \tag{2.2-3}$$

The inversion formulas (2.1–5) and (2.1–6) give

$$\begin{aligned} \int_0^f w(g)\, dg &= \frac{2}{\pi} \int_0^\infty \psi(\tau) \frac{\sin 2\pi f\tau}{\tau}\, d\tau \\ \psi(\tau) &= \int_0^\infty \cos 2\pi f\tau\, d\left[\int_0^f w(g)\, dg\right] \end{aligned} \tag{2.2-4}$$

[16] This is done by Wiener,[11] loc. cit., and by G. W. Kenrick, "The Analysis of Irregular Motions with Applications to the Energy Frequency Spectrum of Static and of Telegraph Signals," *Phil. Mag.*, Ser. 7, Vol. 7, pp. 176–196 (Jan. 1929). Kenrick appears to be one of the first to apply, to noise problems, the correlation function method of computing the power spectrum (one of his problems is discussed in Sec. 2.7). He bases his work on results due to Wiener. Khintchine, in "Korrelationstheorie der stationären stochastischen Prozesse," *Math. Annalen*, 109 (1934), pp. 604–615, proves the following theorem: A necessary and sufficient condition that a function $R(t)$ may be the correlation function of a continuous, stationary, stochastic process is that $R(t)$ may be expressed as

$$R(t) = \int_{-\infty}^{+\infty} \cos tx\, dF(x)$$

where $F(x)$ is a certain distribution function. This expression for $R(t)$ is essentially the second of equations (2.2–4). Khintchine's work has been extended by H. Cramér, "On the theory of stationary random processes," *Ann. of Math.*, Ser. 2, Vol. 41 (1940), pp. 215–230. However, Khintchine and Cramér appear to be interested primarily in questions of existence, representation, etc., and do not stress the concept of the power spectrum.

where the last integral is to be regarded as a Stieltjes' integral. When the expression (2.2–3) for $\psi(\tau)$ is placed in the first formula of (2.2–4) we get

$$\int_0^f w(g)\,dg = \begin{cases} A^2 & \text{when } 0 < f < f_0 \\ A^2 + \dfrac{C^2}{2}, & \text{" } f > f_0 \end{cases} \tag{2.2-5}$$

When this expression is used in the second formula of (2.2–4), the increments of the differential are seen to be A^2 at $f = 0$ and $C^2/2$ at $f = f_0$. The resulting expression for $\psi(\tau)$ agrees with the original one.

Here we desire to use a less rigorous, but more convenient, method of dealing with periodic components. By examining the integral of $w(f)$ as given by (2.2–5) we are led to write

$$w(f) = 2A^2\delta(f) + \frac{C^2}{2}\delta(f - f_0) \tag{2.2-6}$$

where $\delta(x)$ is an even unit impulse function so that if $\epsilon > 0$

$$\int_0^\epsilon \delta(x)\,dx = \frac{1}{2}\int_{-\epsilon}^{\epsilon} \delta(x)\,dx = \tfrac{1}{2} \tag{2.2-7}$$

and $\delta(x) = 0$ except at $x = 0$, where it is infinite. This enables us to use the simpler inversion formulas of section 2.1. The second of these, (2.1-6), is immediately seen to give the correct expression for $\psi(\tau)$. The first one, (2.1–5), gives the correct expression for $w(f)$ provided we interpret the integrals as follows:

$$\begin{aligned} \int_0^\infty \cos 2\pi f\tau\,d\tau &= \tfrac{1}{2}\delta(f) \\ \int_0^\infty \cos 2\pi f_0\tau \cos 2\pi f\tau\,d\tau &= \tfrac{1}{4}\delta(f - f_0) \end{aligned} \tag{2.2-8}$$

It is not hard to show that these are in agreement with the fundamental interpretation

$$\int_{-\infty}^{+\infty} e^{-i2\pi ft}\,dt = \int_{-\infty}^{+\infty} e^{i2\pi ft}\,dt = \delta(f) \tag{2.2-9}$$

which in its turn follows from a formal application of the Fourier integral formula and

$$\int_{-\infty}^{+\infty} \delta(f)e^{i2\pi ft}\,df = \int_{-\infty}^{+\infty} \delta(f)e^{-i2\pi ft}\,df = 1 \tag{2.2-10}$$

We must remember that $f_0 > 0$ and $f \geq 0$ in (2.2–8) so that $\delta(f + f_0) = 0$ for $f \geq 0$.

The definition (2.1–3) for $w(f)$ gives the continuous part of the power spectrum. In order to get the part due to the d.c. and periodic components, which is exemplified by the expression (2.2–6) for $w(f)$ involving the δ functions, we must supplement (2.1–3) by adding terms of the type

$$2A^2\delta(f) + \frac{C^2}{2}\delta(f - f_0) = \left[\operatorname*{Limit}_{T\to\infty} \frac{2|S(0)|^2}{T^2}\right]\delta(f) + \left[\operatorname*{Limit}_{T\to\infty} \frac{2|S(f_0)|^2}{T^2}\right]\delta(f - f_0) \tag{2.2–11}$$

The correctness of this expression may be verified by calculating $S(f)$ for the $I(t)$ of this section given by (2.2–2), and actually carrying out the limiting process.

2.3 Discussion of Results of Section One—Fourier Series

The fact that the spectrum of power $w(f)$ and the correlation function $\psi(\tau)$ are related by Fourier inversion formulas is closely connected with Parseval's theorems for Fourier series and integrals. In this section we shall not use Parseval's theorems explicitly. We start with Fourier's series and use the concept of each component dissipating its share of energy independently of the behavior of the other components.

Let that portion of $I(t)$ which lies in the interval $0 \leq t < T$ be expanded in the Fourier series

$$I(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{2\pi nt}{T} + b_n \sin\frac{2\pi nt}{T}\right) \tag{2.3–1}$$

where

$$a_n = \frac{2}{T}\int_0^T I(t)\cos\frac{2\pi nt}{T}\,dt$$
$$b_n = \frac{2}{T}\int_0^T I(t)\sin\frac{2\pi nt}{T}\,dt \tag{2.3–2}$$

Then for the interval $-\tau \leq t < T - \tau$,

$$I(t+\tau) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos\frac{2\pi n(t+\tau)}{T} + b_n \sin\frac{2\pi n\,(t+\tau)}{T}\right) \tag{2.3–3}$$

Multiplying the series for $I(t)$ and $I(t+\tau)$ together and integrating with respect to t gives, after some reduction,

$$\frac{1}{T}\int_0^T I(t)I(t+\tau)\,dt = \frac{a_0^2}{4} + \sum_{n=1}^{\infty}\frac{1}{2}(a_n^2 + b_n^2)\cos\frac{2\pi n}{T}\tau + O\left(\frac{\tau I^2}{T}\right) \tag{2.3–4}$$

where the last term represents correction terms which must be added because the series (2.3–3) does not represent $I(t+\tau)$ in the interval $(T-\tau, T)$ when $\tau > 0$, or in the interval $(0, -\tau)$ if $\tau < 0$.

If $I(t)$ were a current and if it were to flow through one ohm for the interval $(0, T)$, each component would dissipate a certain average amount of power. The average power dissipated by the component of frequency $f_n = n/T$ cycles per second would be, from the Fourier series and elementary principles,

$$\begin{aligned} &\frac{1}{2}(a_n^2 + b_n^2) \text{ watts}, && n \neq 0 \\ &\frac{a_0^2}{4} \text{ watts}, && n = 0 \end{aligned} \tag{2.3–5}$$

The band width associated with the nth component is the difference in frequency between the $n+1$ th and nth components:

$$f_{n+1} - f_n = \frac{n+1}{T} - \frac{n}{T} = \frac{1}{T} \text{ cps}$$

Hence if the average power in the band $f, f + df$ is defined as $w(f)df$, the average power in the band $f_{n+1} - f_n$ is

$$w(f_n)(f_{n+1} - f_n) = w\left(\frac{n}{T}\right)\frac{1}{T}$$

and, from (2.3–5), this is given by

$$\begin{aligned} &w\left(\frac{n}{T}\right)\frac{1}{T} = \frac{1}{2}(a_n^2 + b_n^2), && n \neq 0 \\ &w(0)\,\frac{1}{T} = \frac{a_0^2}{4}, && n = 0 \end{aligned} \tag{2.3–6}$$

When the coefficients in (2.3–4) are replaced by their expressions in terms of $w(f)$ we get

$$\begin{aligned} &\frac{1}{T}\int_0^T I(t)I(t+\tau)\,dt + O\left(\frac{\tau I^2}{T}\right) \\ &\quad = \frac{1}{T}\sum_{n=0}^{\infty} w\left(\frac{n}{T}\right)\cos\frac{2\pi n\tau}{T} \\ &\quad = \int_0^{\infty} w\left(\frac{n}{T}\right)\cos\frac{2\pi n\tau}{T}\,\frac{dn}{T} \\ &\quad = \int_0^{\infty} w(f)\cos 2\pi f\tau\,df \end{aligned} \tag{2.3–7}$$

where we have assumed T so large and $w(f)$ of such a nature that the summation may be replaced by integration.

If I remains finite, then as $T \to \infty$ with τ held fixed, the correction term on the left becomes negligibly small and we have, upon using the definitions (2.1–4) for the correlation function $\psi(\tau)$, the second of the fundamental inversion formulas (2.1–6). The first inversion formula may be obtained from this at once by using Fourier's double integral for $w(f)$.

Incidentally, the relation (2.3–6) between $w(f)$ and the coefficients a_n and b_n is in agreement with the definition (2.1–3) for $w(f)$ as a limit involving $|S(f)|^2$. From the expressions (2.3–2) for a_n and b_n, the spectrum $S(f_n)$ given by (2.1–2) is

$$S(f_n) = \frac{T}{2}(a_n - ib_n)$$

Then, from (2.1–3) $w(f_n)$ is given by the limit, as $T \to \infty$, of

$$\frac{2}{T}|S(f_n)|^2 = \frac{2}{T}\cdot\frac{T^2}{4}(a_n^2 + b_n^2)$$

$$= \frac{T}{2}(a_n^2 + b_n^2)$$

and this is the expression for $w\left(\frac{n}{T}\right)$ given by (2.3–6).

2.4 Discussion of Results of Section One—Parseval's Theorem

The use of Parseval's theorem[17] enables us to derive the results of section 2.1 more directly than the method of the preceding section. This theorem states that

$$\int_{-\infty}^{+\infty} F_1(f)F_2(f)\,df = \int_{-\infty}^{+\infty} G_1(t)G_2(-t)\,dt \qquad (2.4\text{–}1)$$

where F_1, G_1 and F_2, G_2 are Fourier mates related by

$$F(f) = \int_{-\infty}^{+\infty} G(t)e^{-i2\pi ft}\,dt$$

$$G(t) = \int_{-\infty}^{+\infty} F(f)e^{i2\pi ft}\,df \qquad (2.4\text{–}2)$$

It may be proved in a formal manner by replacing the F_1 on the left of (2.4–1) by its expression as an integral involving $G_1(t)$. Interchanging the

[17] E. C. Titchmarsh, Introduction to the Theory of Fourier Integrals, Oxford (1937).

order of integration and using the second of (2.4–2) to replace F_2 by G_2 gives the right hand side.

We now set $G_1(t)$ and $G_2(t)$ equal to zero except for intervals of length T. These intervals and the corresponding values of G_1 and G_2 are

$$G_1(t) = I(t), \qquad 0 < t < T \tag{2.4–3}$$

$$G_2(t) = I(-t + \tau), \quad \tau - T < t < \tau$$

From (2.4–3) it follows that $F_1(f)$ is the spectrum $S(f)$ of $I(t)$ given by equation (2.1–2). Since $I(t)$ is real it follows from the first of equations (2.4–2) that

$$S(-f) = S^*(f), \tag{2.4–4}$$

where the star denotes conjugate complex, and hence that $|S(f)|^2$ is an even function of f.

The first of equations (2.4–2) also gives

$$\begin{aligned} F_2(f) &= \int_{\tau-T}^{\tau} I(-t + \tau)e^{-i2\pi ft}\, dt \\ &= \int_0^T I(t)e^{i2\pi f(t-\tau)}\, dt \\ &= S^*(f)e^{-i2\pi f\tau} \end{aligned} \tag{2.4–5}$$

When these G's and F's are placed in (2.4–1) we obtain

$$\int_{-\infty}^{+\infty} |S(f)|^2 e^{-2\pi f\tau}\, df = \int_0^{T-\tau} I(t)I(t + \tau)\, dt \tag{2.4–6}$$

where we have made use of the fact that $G_2(-t)$ is zero except in the interval $-\tau < t < T - \tau$ and have assumed $\tau > 0$. If $\tau < 0$ the limits of integration on the right would be $-\tau$ and T.

Since $|S(f)|^2$ is an even function of f we may write (2.4–6) as

$$\frac{1}{T}\int_0^T I(t)I(t + \tau)\, dt + O\left(\frac{\tau I^2}{T}\right) = \int_0^\infty \frac{2|S(f)|^2}{T} \cos 2\pi f\tau\, df \tag{2.4–7}$$

If we now define the correlation function $\psi(\tau)$ as the limit, as $T \to \infty$, of the left hand side and define $w(f)$ as the function

$$w(f) = \operatorname*{Limit}_{T\to\infty} \frac{2|S(f)|^2}{T}, \qquad f > 0 \tag{2.1–3}$$

we obtain the second, (2.1–6), of the fundamental inversion formulas. As before, the first may be obtained from Fourier's integral theorem.

In order to obtain the interpretation of $w(f)df$ as the average power dissipated in one ohm by those components of $I(t)$ which lie in the band f, $f + df$, we set $\tau = 0$ in (2.4–7):

$$\underset{T\to\infty}{\text{Limit}} \frac{1}{T}\int_0^T I^2(t)\,dt = \int_0^\infty w(f)\,df \qquad (2.4\text{–}8)$$

The expression on the left is certainly the total average power which would be dissipated in one ohm and the right hand side represents a summation over all frequencies extending from 0 to ∞. It is natural therefore to interpret $w(f)df$ as the power due to the components in $f, f + df$.

The preceding sections have dealt with the power spectrum $w(f)$ and correlation function $\psi(\tau)$ of a very general type of function. It will be noted that a knowledge of $w(f)$ does not enable us to determine the original function. In obtaining $w(f)$, as may be seen from the definition (2.1–3) or from (2.3–6), the information carried by the phase angles of the various components of $I(t)$ has been dropped out. In fact, as we may see from the Fourier series representation (2.3–1) of $I(t)$ and from (2.3–6), it is possible to obtain an infinite number of different functions all of which have the same $w(f)$, and hence the same $\psi(\tau)$. All we have to do is to assign different sets of values to the phase angles of the various components, thereby keeping $a_n^2 + b_n^2$ constant.

2.5 Harmonic Analysis for Random Functions

In many applications of the theory discussed in the foregoing sections $I(t)$ is a function of t which has a certain amount of randomness associated with it. For example $I(t)$ may be a curve representing the price of wheat over a long period of years, a component of air velocity behind a grid placed in a wind tunnel, or, of primary interest here, a noise current.

In some mathematical work this randomness is introduced by considering $I(t)$ to involve a number of parameters, and then taking the parameters to be random variables. Thus, in the shot effect the arrival times $t_1, t_2, \cdots t_K$ of the electrons were taken to be the parameters and each was assumed to be uniformly distributed over an interval $(0, T)$.

For any particular set of values of the parameters, $I(t)$ has a definite power spectrum $w(f)$ and correlation function $\psi(\tau)$. However, now the principal interest is not in these particular functions, but in functions which give the average values of $w(f)$ and $\psi(\tau)$ for fixed f and τ. These functions are obtained by averaging $w(f)$ and $\psi(\tau)$ over the ranges of the parameters, using, of course, the distribution functions of the parameters.

By averaging both sides of the appropriate equations in sections 2.1 and

2.2 it is seen that our fundamental inversion formulae (2.1–5) and (2.1–6) are unchanged. Thus,

$$\bar{w}(f) = 4\int_0^\infty \bar{\psi}(\tau)\cos 2\pi f\tau\, d\tau \tag{2.5-1}$$

$$\bar{\psi}(\tau) = \int_0^\infty \bar{w}(f)\cos 2\pi f\tau\, df \tag{2.5-2}$$

where the bars indicate averages taken over the parameters with f or τ held constant.

The definitions of $\bar{w}$ and $\bar{\psi}$ appearing in these equations are likewise obtained from (2.1–3) and (2.1-4)

$$\bar{w}(f) = \underset{T\to\infty}{\text{Limit}} \frac{2\overline{|S(f)|^2}}{T} \tag{2.5-3}$$

and

$$\bar{\psi}(\tau) = \underset{T\to\infty}{\text{Limit}} \frac{1}{T}\int_0^T \overline{I(t)I(t+\tau)}\, dt \tag{2.5-4}$$

The values of t and τ are held fixed while averaging over the parameters. In (2.5–3) $S(f)$ is regarded as a function of the parameters obtained from $I(t)$ by

$$S(f) = \int_0^T I(t)e^{-2\pi i f t}\, dt \tag{2.1-2}$$

Similar expressions may be obtained for the average power spectrum for d.c. and periodic components. All we need to do is to average the expression (2.2–11)

Sometimes the average value of the product $I(t)I(t+\tau)$ in the definition (2.5–4) of $\bar{\psi}(\tau)$ is independent of the time T. This enables us to perform the integration at once and obtain

$$\bar{\psi}(\tau) = \overline{I(t)I(t+\tau)} \tag{2.5-5}$$

This introduces a considerable simplification and it appears that the simplest method of computing $\bar{w}'f)$ for an $I(t)$ of this sort is first to compute $\bar{\psi}(\tau)$, and then use the inversion formula (2.5–1).

2.6 First Example—The Shot Effect

We first compute the average on the right of (2.5–5). By using the method of averaging employed many times in part I, we have

$$\overline{I(t)I(t+\tau)} = \sum_{K=0}^{\infty} p(K)\,\overline{I_K(t)I_K(t+\tau)} \tag{2.6-1}$$

where $p(K)$ is the probability of exactly K electrons arriving in the interval $(0, T)$,

$$p(K) = \frac{(\nu T)^K}{K!} e^{-\nu T} \tag{1.1-3}$$

and

$$I_K(t) = \sum_{k=1}^{K} F(t - t_k) \tag{1.3-1}$$

Multiplying $I_K(t)$ and $I_K(t + \tau)$ together and averaging $t_1, t_2, \cdots t_K$ over their ranges gives

$$\overline{I_K(t)I_K(t+\tau)} = \sum_{k=1}^{K}\sum_{m=1}^{K} \int_0^T \frac{dt_1}{T} \cdots \int_0^T \frac{dt_K}{T} F(t - t_k)F(t + \tau - t_m)$$

This is similar to the expression for $\overline{I_K^2(t)}$ which was used in section 1.3 to prove Campbell's theorem and may be treated in much the same way. Thus, if t and $t + \tau$ lie between Δ and $T - \Delta$, the expression above becomes

$$\frac{K}{T}\int_{-\infty}^{+\infty} F(t)F(t+\tau)\,dt + \frac{K(K-1)}{T^2}\left[\int_{-\infty}^{+\infty} F(t)\,dt\right]^2$$

When this is placed in (2.6–1) and the summation performed we obtain an expression independent of T. Consequently we may use (2.5–5) and get

$$\bar{\psi}(\tau) = \nu \int_{-\infty}^{+\infty} F(t)F(t+\tau)\,dt + \overline{I(t)}^2 \tag{2.6–2}$$

where we have used the expression for the average current

$$\overline{I(t)} = \nu \int_{-\infty}^{+\infty} F(t)\,dt \tag{1.3–4}$$

In order to compute $\bar{w}(f)$ from $\bar{\psi}(\tau)$ it is convenient to make use of the fact that $\psi(\tau)$ is always an even function of τ and hence (2.5–1) may also be written as

$$\bar{w}(f) = 2\int_{-\infty}^{+\infty} \bar{\psi}(\tau)\cos 2\pi f\tau\,d\tau \tag{2.6–3}$$

Then

$$\bar{w}(f) = 2\nu \int_{-\infty}^{+\infty} dt\,F(t) \int_{-\infty}^{+\infty} d\tau\,F(t+\tau)\cos 2\pi f\tau$$

$$+ 2\int_{-\infty}^{+\infty} \overline{I(t)}^2 \cos 2\pi f\tau\,d\tau$$

$$= 2\nu \text{ Real Part of } \int_{-\infty}^{+\infty} dt\, F(t)e^{-2\pi i f t} \int_{-\infty}^{+\infty} dt'\, F(t')e^{2\pi i f t'}$$

$$+ 2\overline{I(t)}^2 \int_{-\infty}^{+\infty} e^{i2\pi f\tau}\, d\tau$$

$$= 2\nu \mid s(f) \mid^2 + 2\overline{I(t)}^2 \delta(f) \tag{2.6-4}$$

In going from the first equation to the second we have written $t' = t + \tau$ and have considered cos $2\pi f\tau$ to be the real part of the corresponding exponential. In going from the second equation to the third we have set

$$s(f) = \int_{-\infty}^{+\infty} F(t)e^{-2\pi i f t}\, dt \tag{2.6-5}$$

and have used

$$\int_{-\infty}^{+\infty} e^{i2\pi f t}\, dt = \delta(f) \tag{2.2-9}$$

The term in $\bar{w}(f)$ involving $\delta(f)$ represents the average power which would be dissipated by the d.c. component of $I(t)$ in flowing through one ohm. It is in agreement with the concept that the average power in the band $0 \leq f < \epsilon$, $\epsilon > 0$ but very small, is

$$\int_0^{\epsilon} \bar{w}(f)\, df = 2\overline{I(t)}^2 \int_0^{\epsilon} \delta(f)\, df$$
$$= \overline{I(t)}^2 \tag{2.6-6}$$

The expression (2.6–4) for $\bar{w}(f)$ may also be obtained from the definition (2.5–3) for $\bar{w}(f)$plus the additional term due to the d.c. component obtained by averaging the expressions (2.2–11). We leave this as an exercise for the reader. He will find it interesting to study the steps in Carson's[15] paper leading up to equation (8). Carson's $R(\omega)$ is related to our $\bar{w}(f)$ by

$$\bar{w}(f) = 2\pi R(\omega)$$

and his $f(i\omega)$ is equal to our $s(f)$.

Integrating both sides of (2.6–4) with respect to f from 0 to ∞ and using

$$\overline{I^2} = \int_0^{\infty} \bar{w}(f)\, df$$

gives the result

$$\overline{I^2} - \overline{I}^2 = 2\nu \int_0^{+\infty} \mid s(f) \mid^2 df \tag{2.6-7}$$

[15] Loc. cit.

This may be obtained immediately from Campbell's theorem by applying Parseval's theorem.

As an example of the use of these formulas we derive the power spectrum of the voltage across a resistance R when a current consisting of a great number of very short pulses per second flows through R. Let $F(t - t_k)$ be the voltage produced by the pulse occurring at time t_k. Then

$$F(t) = R\varphi(t)$$

where $\varphi(t)$ is the current in the pulse. We confine our interest to relatively low frequencies such that we may make the approximation

$$s(f) = \int_{-\infty}^{+\infty} R\varphi(t)e^{-2\pi ift}\,dt$$

$$\approx R\int_{-\infty}^{+\infty} \varphi(t)\,dt = Rq$$

where q is the charge carried through the resistance by one pulse. From (2.6-4) it follows that for these low frequencies the continuous portion of the power spectrum for the voltage is constant and equal to

$$\bar{w}(f) = 2\nu R^2q^2 = 2\bar{I}R^2q \qquad (2.6\text{-}8)$$

where $\bar{I} = \nu q$ is the average current flowing through R. This result is often used in connection with the shot effect in diodes.

In the study of the shot effect it was assumed that the probability of an event (electron arriving at the anode) happening in dt was νdt where ν is the expected number of events per second. This probability is independent of the time t. Sometimes we wish to introduce dependency on time.[18] As an example, consider a long interval extending from 0 to T. Let the probability of an event happening in $t, t + dt$ be $\bar{K}p(t)dt$ where $\bar{K}$ is the average number of events during T and $p(t)$ is a given function of t such that

$$\int_0^T p(t)\,dt = 1$$

For the shot effect $p(t) = 1/T$.

What is the probability that exactly K events happen in T? As in the case of the shot effect, section 1.1, we may divide $(0, T)$ into N intervals each of length Δt so that $N\Delta t = T$. The probability of no event happening in the first Δt is

$$1 - \bar{K}p\left(\frac{\Delta t}{2}\right)\Delta t$$

[18] A careful discussion of this subject is given by Hurwitz and Kac in "Statistical Analysis of Certain Types of Random Functions." I understand that this paper will soon appear in the Annals of Math. Statistics.

The product of N such probabilities is, as $N \to \infty$, $\Delta t \to 0$;

$$\exp\left[-\bar{K}\int_0^T p(t)\,dt\right] = e^{-\bar{K}}$$

This is the probability that exactly 0 events happen in T. In the same way we are led to the expression

$$\frac{\bar{K}^K}{K!}\,e^{-\bar{K}} \tag{2.6-9}$$

for the probability that exactly K events happen in T.

When we consider many intervals $(0, T)$ we obtain many values of K and also many values of I measured t seconds from the beginning of each interval. These values of I define the distribution of I at time t. By proceeding as in section 1.4 we find that the probability density of I is

$$P(I, t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} du \exp\left[-iuI + \bar{K}\int_0^T p(x)(e^{iuF(t-x)} - 1)\,dx\right]$$

The corresponding average and variance is

$$\bar{I} = \bar{K}\int_0^T p(x)F(t-x)\,dx$$

$$\overline{(I-\bar{I})^2} = \bar{K}\int_0^T p(x)F^2(t-x)\,dx \tag{2.6-10}$$

If $S(f)$ is given by (2.1–2) and $s(f)$ by (2.6–5) (assuming the duration of $F(t)$ short in comparison with T) the average value of $|S(f)|^2$ may be obtained by putting (1.3-1) in (2.1-2) to get

$$S_K(f) = s(f)\sum_1^K e^{-2\pi i f t_k}$$

Expressing $S_K(f)\,S_K^*(f)$, where the star denotes conjugate complex, as a double sum and averaging over the t_k's, using $p(t)$, and then averaging over the K's gives

$$\overline{|S(f)|^2} = \bar{K}|s(f)|^2\left[1 + \bar{K}\left|\int_0^T p(x)e^{-2\pi i f x}\,dx\right|^2\right] \tag{2.6-11}$$

This may be used to compute the power spectrum from (2.5–3) provided $p(x)$ is not periodic. If $p(x)$ is periodic then the method of section 2.2 should be used at the harmonic frequencies. If the fluctuations of $p(t)$ are slow in comparison with the fluctuations of $F(t)$ the second term within the brackets of (2.6–11) may generally be neglected since there are no values of

f which make both it and $s(f)$ large at the same time. On the other hand, if both $p(t)$ and $F(t)$ fluctuate at about the same rate this term must be considered.

2.7 Second Example—Random Telegraph Signal[16]

Let $I(t)$ be equal to either a or $-a$ so that it is of the form of a flat top wave. Let the intervals between changes of sign, i.e. the lengths of the tops and bottoms, be distributed exponentially. We are led to this distribution by assuming that, if on the average there are μ changes of sign per second, the probability of a change of sign in $t, t + dt$ is μdt and is independent of what happens outside the interval $t, t + dt$. From the same sort of reasoning as employed in section 1.1 for the shot effect we see that the probability of obtaining exactly K changes of sign in the interval $(0, T)$ is

$$p(K) = \frac{(\mu T)^K}{K!} e^{-\mu T} \tag{2.7-1}$$

We consider the average value of the product $I(t)I(t + \tau)$. This product is a^2 if the two I's are of the same sign and is $-a^2$ if they are of opposite sign. In the first case there are an even number, including zero, of changes of sign in the interval $(t, t + \tau)$, and in the second case there are an odd number of changes of sign. Thus

$$\begin{aligned} &\text{Average value of } I(t)I(t + \tau) \qquad (2.7\text{–}2)\\ &= a^2 \times \text{probability of an even number of changes of sign in } t, t + \tau\\ &- a^2 \times \text{probability of an odd number of changes of sign in } t, t + \tau \end{aligned}$$

The length of the interval under consideration is $| t + \tau - t | = | \tau |$ seconds. Since, by assumption, the probability of a change of sign in an elementary interval of length Δt is independent of what happens outside that interval, it follows that the same is true of any interval irrespective of when it starts. Hence the probabilities in (2.7–2) are independent of t and may be obtained from (2.7–1) by setting $T = | \tau |$. Then (2.7–2) becomes, assuming $\tau > 0$ for the moment,

$$\begin{aligned} \overline{I(t)I(t + \tau)} &= a^2[p(0) + p(2) + p(4) + \cdots]\\ &\qquad - a^2[p(1) + p(3) + p(5) + \cdots]\\ &= a^2 e^{-\mu\tau}\left[1 - \frac{\mu\tau}{1!} + \frac{(\mu\tau)^2}{2!} - \cdots\right]\\ &= a^2 e^{-2\mu\tau} \end{aligned} \tag{2.7-3}$$

[16] Kenrick, cited in Section 2.2.

From (2.5–5), this gives the correlation function for $I(t)$

$$\bar{\psi}(\tau) = a^2 e^{-2\mu|\tau|} \tag{2.7-4}$$

The corresponding power spectrum is, from (2.5–1),

$$\begin{aligned} \bar{w}(f) &= 4a^2 \int_0^\infty e^{-2\mu\tau} \cos 2\pi f\tau \, d\tau \\ &= \frac{2a^2\mu}{\pi^2 f^2 + \mu^2} \end{aligned} \tag{2.7-5}$$

Correlation functions and power spectra of this type occur quite frequently. In particular, they are of use in the study of turbulence in hydrodynamics. We may also obtain them from our shot effect expressions if we disregard the d.c. component. All we have to do is tc assume that the effect $F(t)$ of an electron arriving at the anode at time $t = 0$ is zero for $t < 0$, and that $F(t)$ decays exponentially with time after jumping to its maximum value at $t = 0$. This may be verified by substituting the value

$$F(t) = 2a\sqrt{\frac{\mu}{\nu}} e^{-2\mu t}, \qquad t > 0 \tag{2.7-6}$$

for $F(t)$ in the expressions (2.6–2) and (2.6–4) (after using 2.6–5) for the correlation function and energy spectrum of the shot effect.

The power spectrum of the current flowing through an inductance and a resistance in series in response to a very wide band thermal noise voltage is also of the form (2.7–5).

Incidentally, this gives us an example of two quite different $I(t)$'s, one a flat top wave and the other a shot effect current, which have the same correlation functions and power spectra, aside from the d.c. component.

There is another type of random telegraph signal which is interesting to analyze. The time scale is divided into intervals of equal length h. In an interval selected at random the value of $I(t)$ is independent of the values in the other intervals, and is equally likely to be $+a$ or $-a$. We could construct such a wave by flipping a penny. If heads turned up we would set $I(t) = a$ in $0 < t < h$. If tails were obtained we would set $I(t) = -a$ in this interval. Flipping again would give either $+a$ or $-a$ for the second interval $h < t < 2h$, and so on. This gives us one wave. A great many waves may be constructed in this way and we denote averages over these waves, with t held constant, by bars.

We ask for the average value of $I(t)I(t + \tau)$, assuming $\tau > 0$. First we note that if $\tau > h$ the currents correspond to different intervals for all

values of t. Since the values in these intervals are independent we have

$$\overline{I(t)I(t+\tau)} = \overline{I(t)}\;\overline{I(t+\tau)} = 0$$

for all values of t when $\tau > h$.

To obtain the average when $\tau < h$ we consider t to lie in the first interval $0 < t < h$. Since all the intervals are the same from a statistical point of view we lose no generality in doing this. If $t + \tau < h$, i.e., $t < h - \tau$, both currents lie in the first interval and

$$\overline{I(t)I(t+\tau)} = a^2$$

If $t > h - \tau$ the current $I(t + \tau)$ corresponds to the second interval and hence the average value is zero.

We now return to (2.5–4). The integral there extends from 0 to T. When $\tau > h$, the integrand is zero and hence

$$\bar{\psi}(\tau) = 0, \qquad \tau > h \tag{2.7–7}$$

When $\tau < h$, our investigation of the interval $0 < t < h$ enables us to write down the portion of the integral extending from 0 to h:

$$\int_0^h I(t)I(t+\tau)\,dt = \int_0^{h-\tau} a^2\,dt + \int_{h-\tau}^h 0\,dt$$
$$= a^2(h-\tau)$$

Over the interval of integration $(0, T)$ we have T/h such intervals each contributing the same amount. Hence, from (2.5–4),

$$\begin{aligned}\bar{\psi}(\tau) &= \operatorname*{Limit}_{T\to\infty} \frac{a^2}{T}\cdot\frac{T}{h}(h-\tau) \\ &= a^2\left(1-\frac{\tau}{h}\right), \qquad 0 \le \tau < h\end{aligned} \tag{2.7–8}$$

The power spectrum of this type of telegraph wave is thus

$$\begin{aligned}\bar{w}(f) &= 4a^2\int_0^h \left(1-\frac{\tau}{h}\right)\cos 2\pi f\tau\,d\tau \\ &= 2h\left(\frac{a\sin \pi fh}{\pi fh}\right)^2\end{aligned} \tag{2.7–9}$$

This is seen to have the same general behavior as $\bar{w}(f)$ for the first type of telegraph signal given by (2.7–5), when we relate the average number, μ, of changes of sign per second to the interval length h by $\mu h = 1$.

2.8 Representation of Noise Current

In section 1.8 the Fourier series representation of the shot effect current was discussed. This suggests the representation*

$$I(t) = \sum_{n=1}^{N} (a_n \cos \omega_n t + b_n \sin \omega_n t) \tag{2.8-1}$$

where

$$\omega_n = 2\pi f_n, \qquad f_n = n\Delta f \tag{2.8-2}$$

a_n and b_n are taken to be independent random variables which are distributed normally about zero with the standard deviation $\sqrt{w(f_n)\Delta f}$. $w(f)$ is the power spectrum of the noise current, i.e., $w(f)\,df$ is the average power which would be dissipated by those components of $I(t)$ which lie in the frequency range $f, f + df$ if they were to flow through a resistance of one ohm.

The expression for the standard deviation of a_n and b_n is obtained when we notice that Δf is the width of the frequency band associated with the nth component. Hence $w(f_n)\Delta f$ is the average energy which would be dissipated if the current

$$a_n \cos \omega_n t + b_n \sin \omega_n t$$

were to flow through a resistance of one ohm, this average being taken over all possible values of a_n and b_n. Thus

$$w(f_n)\Delta f = \overline{a_n^2} \cos^2 \omega_n t + \overline{2a_n b_n} \cos \omega_n t \sin \omega_n t + \overline{b_n^2} \sin^2 \omega_n t = \overline{a_n^2} = \overline{b_n^2} \tag{2.8-3}$$

The last two steps follow from the independence of a_n and b_n and the identity of their distributions. It will be observed that $w(f)$, as used with the representation (2.8–1), is the same sort of average as was denoted in section 2.5 by $\bar{w}(f)$. However, $w(f)$ is often given to us in order to specify the spectrum of a given noise.

For example, suppose we are interested in the output of a certain filter when a source of thermal noise is applied to the input. Let $A(f)$ be the absolute value of the ratio of the output current to the input current when a steady sinusoidal voltage of frequency f is applied to the input. Then

$$w(f) = cA^2(f) \tag{2.8-4}$$

* As mentioned in section 1.7 this sort of representation was used by Einstein and Hopf for radiation. Shottky (1918) used (2.8–1), apparently without explicitly taking the coefficients to be normally distributed. Nyquist (1932) derived the normal distribution from the shot effect.

If W is the average power dissipated in one ohm by $I(t)$,

$$W = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I^2(t)\,dt = \int_0^\infty w(f)\,df$$
$$= c\int_0^\infty A^2(f)\,df \tag{2.8–5}$$

which is an equation to determine c when W and $A(f)$ are known.

In using the representation (2.8–1) to investigate the statistical properties of $I(t)$ we first find the corresponding statistical properties of the summation on the right when the a's and b's are regarded as random variables distributed as mentioned above and t is regarded as fixed. In general, the time t disappears in this procedure just as it did in (2.8–3). We then let $N \to \infty$ and $\Delta f \to 0$ so that the summations may be replaced by integrations. Finally, the frequency range is extended to cover all frequencies from 0 to ∞.

The usual way of looking at the representation (2.8–1) is to suppose that we have an oscillogram of $I(t)$ extending from $t = 0$ to $t = \infty$. This oscillogram may be cut up into strips of length T. A Fourier analysis of $I(t)$ for each strip will give a set of coefficients. These coefficients will vary from strip to strip. Our representation ($T\Delta f = 1$) assumes that this variation is governed by a normal distribution. Our process for finding statistical properties by regarding the a's and b's as random variables while t is kept fixed corresponds to examining the noise current at a great many instants. Corresponding to each strip there is an instant, and this instant occurs at t (this is the t in (2.8–1)) seconds from the beginning of the strip. This is somewhat like examining the noise current at a great number of instants selected at random.

Although (2.8–1) is the representation which is suggested by the shot effect and similar phenomena, it is not the only representation, nor is it always the most convenient. Another representation which leads to the same results when the limits are taken is[19]

$$I(t) = \sum_{n=1}^{N} c_n \cos(\omega_n t - \varphi_n) \tag{2.8–6}$$

where $\varphi_1, \varphi_2, \cdots \varphi_N$ are angles distributed at random over the range $(0, 2\pi)$ and

$$c_n = [2w(f_n)\Delta f]^{1/2}, \qquad \omega_n = 2\pi f_n, \qquad f_n = n\Delta f \tag{2.8-7}$$

[19] This representation has often been used by W. R. Bennett in unpublished memoranda written in the 1930's.

In this representation $I(t)$ is regarded as the sum of a number of sinusoidal components with fixed amplitudes but random phase angles.

That the two different representations (2.8–1) and (2.8–6) of $I(t)$ lead to the same statistical properties is a consequence of the fact that they are always used in such a way that the "central limit theorem*" may be used in both cases.

This theorem states that under certain general conditions, the distribution of the sum of N random vectors approaches a normal law (it may be normal in several dimensions**) as $N \to \infty$. In fact from this theorem it appears that a representation such as

$$I(t) = \sum_{n=1}^{N} (a_n \cos \omega_n t + b_n \sin \omega_n t) \tag{2.8–6}$$

where a_n and b_n are independent random variables which take only the values $\pm [w(f_n)\Delta f]^{1/2}$, the probability of each value being $\frac{1}{2}$, will lead in the limit to the same statistical properties of $I(t)$ as do (2.8–1) and (2.8–6).

2.9 The Normal Distribution in Several Variables[20]

Consider a random vector r in K dimensions. The distribution of this vector may be specified by stating the distribution of the K components, $x_1, x_2, \cdots x_K$, of r. r is said to be normally distributed when the probability density function of the x's is of the form

$$(2\pi)^{-K/2} \mid M \mid^{-1/2} \exp [-\tfrac{1}{2} x' M^{-1} x] \tag{2.9–1}$$

where the exponent is a quadratic form in the x's. The square matrix M is composed of the second moments of the x's.

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1K} \\ \cdot & & & \cdot \\ \mu_{1K} & & \cdots & \mu_{KK} \end{bmatrix} \tag{2.9–2}$$

where the second moments are defined by

$$\mu_{11} = \overline{x_1^2}, \qquad \mu_{12} = \overline{x_1 x_2}, \qquad \text{etc.} \tag{2.9–3}$$

$\mid M \mid$ represents the determinant of M and x' is the row matrix

$$x' = [x_1, x_2, \cdots x_K] \tag{2.9–4}$$

x is the column matrix obtained by transposing x'.

* See section 2.10.
** See section 2.9.
[20] H. Cramér, "Random Variables and Probability Distributions," Chap. X., Cambridge Tract No. 36 (1937).

The exponent in the expression (2.9–1) for the probability density may be written out by using

$$x'M^{-1}x = \sum_{r=1}^{K}\sum_{s=1}^{K}\frac{M_{rs}}{|M|}x_r x_s \tag{2.9–5}$$

where M_{rs} is the cofactor of μ_{rs} in M.

Sometimes there are linear relations between the x's so that the random vector r is restricted to a space of less than K dimensions. In this case the appropriate form for the density function may be obtained by considering a sequence of K-dimensional distributions which approach the one being investigated.

If r_1 and r_2 are two normally distributed random vectors their sum $r_1 + r_2$ is also normally distributed. It follows that the sum of any number of normally distributed random vectors is normally distributed.

The characteristic function of the normal distribution is

$$\text{ave. } e^{iz_1x_1+iz_2x_2+\cdots+iz_Kx_K} = \exp\left[-\frac{1}{2}\sum_{r=1}^{K}\sum_{s=1}^{K}\mu_{rs}z_r z_s\right] \tag{2.9–6}$$

2.10 Central Limit Theorem

The central limit theorem in probability states that the distribution of the sum of N independent random vectors $r_1 + r_2 + \cdots + r_N$ approaches a normal law as $N \to \infty$ when the distributions of $r_1, r_2, \cdots r_N$ satisfy certain general conditions.[7]

As an example we take the case in which $r_1, r_2, \cdots$ are two-dimensional vectors[21], the components of r_n being x_n and y_n. Without loss of generality we assume that

$$\bar{x}_n = 0, \qquad \bar{y}_n = 0.$$

The components of the resultant vector are

$$\begin{aligned} X &= x_1 + x_2 + \cdots + x_N \\ Y &= y_1 + y_2 + \cdots + y_N \end{aligned} \tag{2.10–1}$$

and, since $r_1, r_2, \cdots$ are independent vectors, the second moments of the resultant are

$$\begin{aligned} \mu_{11} &= \overline{X^2} = \overline{x_1^2} + \overline{x_2^2} + \cdots + \overline{x_N^2} \\ \mu_{22} &= \overline{Y^2} = \overline{y_1^2} + \overline{y_2^2} + \cdots + \overline{y_N^2} \\ \mu_{12} &= \overline{XY} = \overline{x_1y_1} + \overline{x_2y_2} + \cdots + \overline{x_Ny_N} \end{aligned} \tag{2.10-2}$$

[7] Incidentally, von Laue (see references in section 1.7) used this theorem in discussing the normal distribution of the coefficients in a Fourier series used to represent black-body radiation. He ascribed it to Markoff.

[21] This case is discussed by J. V. Uspensky, "Introduction to Mathematical Probability", McGraw-Hill (1937) Chap. XV.

Apparently there are several types of conditions which are sufficient to ensure that the distribution of the resultant approaches a normal law. One sufficient condition is that[21]

$$\mu_{11}^{-3/2} \sum_{n=1}^{N} \overline{|x_n|^3} \to 0$$
$$\mu_{22}^{-3/2} \sum_{n=1}^{N} \overline{|y_n|^3} \to 0 \tag{2.10-3}$$

The central limit theorem tells us that the distribution of the random vector (X, Y) approaches a normal law as $N \to \infty$. The second moments of this distribution are given by (2.10–2). When we know the second moments of a normal distribution we may write down the probability density function at once. Thus from section 2.9

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{bmatrix}, \qquad M^{-1} = |M|^{-1} \begin{bmatrix} \mu_{22} & -\mu_{12} \\ -\mu_{12} & \mu_{11} \end{bmatrix}$$
$$|M| = \mu_{11}\mu_{22} - \mu_{12}^2$$
$$x' = [X, Y]$$
$$x'M^{-1}x = |M|^{-1}(\mu_{22}X^2 - 2\mu_{12}XY + \mu_{11}Y^2)$$

The probability density is therefore

$$\frac{(\mu_{11}\mu_{22} - \mu_{12}^2)^{-1/2}}{2\pi} \exp\left[\frac{-\mu_{22}X^2 - \mu_{11}Y^2 + 2\mu_{12}XY}{2(\mu_{11}\mu_{22} - \mu_{12}^2)}\right] \tag{2.10-3}$$

Incidentally, the second moments are related to the standard deviations σ_1, σ_2 of X, Y and to the correlation coefficient τ of X and Y by

$$\mu_{11} = \sigma_1^2, \qquad \mu_{22} = \sigma_2^2, \qquad \mu_{12} = \tau\sigma_1\sigma_2 \tag{2.10-4}$$

and the probability density takes the standard form

$$\frac{(1 - \tau^2)^{-1/2}}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2(1 - \tau^2)}\left(\frac{X^2}{\sigma_1^2} - 2\tau\frac{XY}{\sigma_1\sigma_2} + \frac{Y^2}{\sigma_2^2}\right)\right] \tag{2.10-5}$$

[21] This is used by Uspensky, loc. cit. Another condition analogous to the Lindeberg condition is given by Cramer,[20] loc. cit.

(To be concluded)

Abstracts of Technical Articles by Bell System Authors

Crossbar Toll Switching System.[1] L. G. ABRAHAM, A. J. BUSCH, AND F. F. SHIPLEY. A new crossbar toll switching system was placed in service in Philadelphia in August 1943. Important improvements offered by this system include:

1. Transmission objectives are met more readily and substantial economies are obtained in outside plant and in repeater equipment.

2. Extended use of toll dialing results in operating economies and improved service to subscribers. Calls over circuits still on a ringdown basis are also handled more expeditiously and with operating economies.

3. Flexibility due to the use of sender and markers to control establishing connections will be useful in meeting future toll switching requirements.

Low-Frequency Quartz-Crystal Cuts having Low Temperature Coefficients.[2] W. P. MASON AND R. A. SYKES. This paper discusses low-frequency, low-temperature-coefficient crystals which are suitable for use in filters and oscillators in the frequency range from 4 to 100 kilocycles. Two new cuts, the MT and NT, are described. These are related to the +5-degree *X*-cut crystal, which is the quartz crystal having the lowest temperature coefficient for any orientation of a bar cut from the natural crystal. When the width of the +5-degree *X*-cut crystal approaches in value the length, the motion has a shear component, and this introduces a negative temperature coefficient which causes the temperature coefficient of the crystal to become increasingly negative as the ratio of width to length increases.

The MT crystal has its length along nearly the same axis as the +5-degree *X*-cut crystal, but its major surface is rotated by 35 to 50 degrees around the length axis. This results in giving the shear component a zero or positive temperature coefficient and results in a crystal with a uniformly low temperature coefficient independent of the width-length ratio. A slightly higher rotation about the length axis results in a crystal which has a low-temperature coefficient when vibrating in flexure and this crystal has been called the NT crystal. The NT crystal can be used in a frequency range from 4 to 50 kilocycles, while the MT is useful from 50 kilocycles to 500 kilocycles.

A special oscillator circuit is described which can drive a high-impedance NT flexure crystal. This circuit, together with the NT crystal, has been used to control the mean frequency of the Western Electric frequency-modulated radio transmitter.

[1] *Elec. Engg., Transactions Section,* June, 1944.
[2] *Proc. I. R. E.,* April 1944.

Electronics; Today and Tomorrow.[3] JOHN MILLS. John Mills, scientist, teacher, author, telephone and radio engineer, was first introduced to the electron while still a fledgling under the tutelage of that eminent scientist R. A. Millikan at Chicago University. That was before Millikan had, to quote the author, "first put salt on its tail." Electronics, as a special branch of physics and engineering, has come to adulthood under the eyes of this observant author. From intimate association with this and allied fields John Mills writes knowingly. It is a book intended for the intelligent layman rather than for the expert.

"Electronics; Today and Tomorrow" likewise contains much of interest concerning the electronics of yesterday but, as one would expect, very little about the electronics of tomorrow because, as the author points out, much of this "should await the victorious conclusion of the present conflict." The book generally presents an interesting introduction to many things electronic and throughout is interspersed with examples of present day techniques which employ electronic devices. All line drawings, diagrams and photographs have been omitted.

The author begins his latest book with a brief capitulation of familiar engineering applications such as long-distance telephony, broadcasting, sound motion pictures and television which have been achieved through the use of electronic devices. He goes on to an historical account of some underlying discoveries and a discussion of atomic structure. An introduction to static electricity with the classical concepts of positive and negative charge is followed by establishment of the ideas of charges in motion, electrical current, discharges in gases, X-rays and their generation.

The remainder of the book is divided into Part I—Electron Tubes, and Part II—Electronic Devices. In Part I the author follows the development of the art in nearly chronological order from diodes through the modern multi-electrode tubes giving uses and applications of each and explaining the purpose of the grids introduced. Part II discusses more complicated structures such as cathode-ray tubes, kinescopes, iconoscopes, electron microscopes, kodatrons, magnetrons, rumbatrons, klystrons, etc. and elaborates upon their practical applications. This is, in fact, a wondrous field of developments. Finally there is a chapter on cyclotrons which the author says "comes into this book because it requires for its operation a powerful oscillator—or oscillator plus powerful electron amplifier—which will supply a high voltage at a frequency of megacycles." While many industrial applications, such perhaps as tin plating, might be included on the same basis, nevertheless the discussion of the cyclotron does give an opportunity to explain the concepts of modern physics more completely than was undertaken earlier in the book and is very interesting reading.

[3] Published by D. Van Nostrand Company, Inc., New York, N. Y., 1944.

Impedance Concept in Wave Guides.[4] S. A. SCHELKUNOFF. The impedance concept is the foundation of engineering transmission theory. If wave guides are to be fully utilized as transmission systems or parts thereof, their properties must be expressed in terms of appropriately chosen impedances or else a new transmission theory must be developed. The gradual extension of the concept has necessitated a broader point of view without which an exploitation of its full potentialities would be impossible.

In the course of various private discussions it has been noted that there exists some uneasiness with regard to the applicability of the concept at very high frequencies. In part this may be attributed to relative unfamiliarity with the wave guide phenomena and in part to the evolution of the concept itself. Some particular aspects of the concept have to be sacrificed in the process of generalization and although these aspects may be logically unimportant, they frequently become psychological obstacles to understanding in the early stages of the development. For this reason several sections of this paper are devoted to a general discussion of the impedance concept before more specific applications are given; then by way of illustration it is proved that an infinitely thin perfectly conducting iris between two *different* wave guides behaves as if between the admittances of its faces there existed an ideal transformer. This theorem is a generalization of another theorem to the effect that when the two wave guides are alike, the iris behaves as a shunt reactor. Actual calculation of the admittances and the transformer ratio depends on the solution of an appropriate boundary value problem.

More generally, wave guide discontinuities are representable by T-networks. In some special cases these networks lack series branches, and in other cases the shunt branch.

Theory of Cathode Sputtering in Low Voltage Gaseous Discharges.[5] CHARLES HARD TOWNES. To determine the amount of sputtering in a glow discharge three functions must be known: the number of ions of a given energy striking the cathode, the amount of cathode material released from the cathode by each ion, and the fraction of material released from the cathode which diffuses away. Expressions derived for these allow determination of the dependence of total rate of sputtering on the geometry of the discharge, pressure, cathode fall, current, and constants of the gas and cathode surface. The result is most accurate for very low voltage, high pressure discharges. Comparison with experimental data shows quantitative agreement under these conditions.

[4] *Quarterly of Applied Mathematics*, April 1944.
[5] *Phys. Rev.*, June 1 and 15, 1944.

Contributors to this Issue

A. R. D'HEEDENE, B.S. in Mechanical Engineering, New York University, 1924; Bell Telephone Laboratories, 1924–. Mr. D'heedene has been engaged in the development of wave filters, particularly filters using quartz crystal units or operating at very high frequencies.

R. M. C. GREENIDGE, B.S. in Mechanical Engineering, Harvard University, 1924. Engineering Department, Western Electric Company, Inc., 1924–25; Bell Telephone Laboratories, Inc., 1925–. Engaged in the development of crystal units for filters and oscillators.

WILTON T. REA, B.S. in Physics, Princeton University, 1926. American Telephone and Telegraph Company, Department of Development and Research, 1926–1934; Bell Telephone Laboratories, Inc. 1934–. Prior to 1941 Mr. Rea was concerned with telegraph development problems. Since then the group of which he is in charge has been engaged full time on war projects.

S. O. RICE, B.S. in Electrical Engineering, Oregon State College, 1929; California Institute of Technology, 1929–30, 1934–35. Bell Telephone Laboratories, 1930–. Mr. Rice has been concerned with various theoretical investigations relating to telephone transmission theory.