THE BELL SYSTEM
TECHNICAL JOURNAL

A JOURNAL DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

TABLE OF CONTENTS
AND
INDEX

VOLUME XXIV

1945

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

PRINTED IN U. S. A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXIV, 1945

Table of Contents

Index to Volume XXIV

VOLUME XXIV JANUARY, 1945 NUMBER 1

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are 50 cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

Vol. XXIV January, 1945 No. 1

Intermittent Behavior in Oscillators

By W. A. EDSON

Oscillators of all sorts may, for certain values of the parameters, show low-frequency disturbances. Usually the disturbance takes the form of a low-frequency interruption of the desired oscillation. By the method here presented it is possible to determine whether or not such intermittent behavior will occur in a given oscillator and what circuit modifications are required to promote stability. The intentional generation of a modulated wave by control of the low frequency behavior of an oscillator is also considered. Oscillators of the negative resistance type are not considered.

I. Introduction

IT HAS been known for a long time that all kinds of oscillators are subject to the trouble variously referred to as intermittent oscillation, motor boating, or squegging. In conventional circuits such as the Hartley the phenomenon is most likely to be observed if the grid leak and grid condenser are abnormally large. It is found that the time constant of this combination must be reduced as the frequency is raised and as the Q of the resonant circuit is decreased. At frequencies above a few hundred megacycles the problem of producing a practical circuit with suitable margin of stability is quite difficult.

With the advent of the oscillator having automatic output control the problem assumed a new aspect.[1, 2] By application of an amplified control circuit a high degree of constancy of output together with low harmonic output is obtained. Satisfactory operation is secured, however, only when suitable attention is given to the characteristics of the control circuit.

The intentional generation of pulses by means of intermittent oscillations of relatively high frequency has been studied to some extent, and circuits of this kind are employed in some television systems. Usually the high-frequency oscillation is limited to a small portion of the low-frequency cycle, the charge stored during this period being allowed to dissipate itself relatively slowly during the remainder of the cycle.

In all of these circuits satisfactory performance depends upon a proper proportioning of elements not directly associated with the operating fre-

[1] L. B. Argimbau, "An Oscillator Having a Linear Operating Characteristic," *Proc. I.R.E.*, Vol. 21, p. 14, Jan. 1933.

[2] J. Groszkowski, "Oscillators with Automatic Control of the Threshold of Regeneration," *Proc. I.R.E.*, Vol. 22, p. 145, Feb. 1934.

quency. When continuous oscillation is necessary it is desirable to provide adequate margin against intermittent operation. When intermittent operation is desired the opposite is true. In either case an understanding of the same general problem is necessary.

The present analysis applies only to oscillators of the feedback type. No method of extending it to cover negative resistance oscillators such as the Dynatron and the Transitron has been found. Relaxation oscillators as such are not considered here inasmuch as they are seldom affected by intermittent operation. No specific frequency limits apply but it is sometimes difficult at very high frequencies to achieve desirable values of the constants. At very low frequencies oscillators employing automatic output control are relatively unsuitable because their performance tends to be unduly sluggish.

The term linear oscillator is used to indicate an oscillator in which the range of operation is controlled within such limits that the harmonic content of the output is inappreciable.

The general equation describing a simple amplitude-modulated wave is

$$V = V_0(1 + m \sin 2\pi ft) \sin 2\pi Ft$$

This may be taken as defining the modulation factor m, a complex number which is limited to magnitudes between zero and one.

II. General Theory of Oscillation

It is found that three separate functions are necessary and sufficient for the operation of an oscillator of the feedback type.[3] These are indicated in the block diagram of Fig. 1.

The amplifier must be provided to overcome the losses of the rest of the system. The power output, if any, depends upon the fact that the output of an amplifier is greater than the input.

Selectivity must be provided to insure that the output has a definite frequency. Ordinarily a tuned circuit of relatively high Q is used although some excellent oscillators employ resistance-capacitance networks. The term filter is employed as being sufficiently general to include these extremes.

A limiter of some form is necessary to establish the level at which sustained oscillations occur. In many circuits the functions of amplifier and limiter are performed simultaneously in the vacuum tube. In an important class of oscillators the limiter is a thermal device such as a tungsten lamp. In the Meacham circuit the functions of limiter and filter are combined in a bridge employing a tuned circuit and a tungsten lamp.

To simplify the analysis it is convenient to assume that the amplifier of Fig. 1 is completely linear and operates with equal gain at all frequencies

[3] This topic is discussed more fully in "Hyper and Ultra-High Frequency Engineering," R. I. Sarbacher, and W. A. Edson, John Wiley & Sons, Inc., 1943.

from zero to infinity. Similarly the filter is assumed to consist of linear circuit elements and to have a definite curve of loss versus frequency. Associated with this loss characteristic is some specific phase characteristic.[4] The limiter is assumed to have a loss which is independent of frequency but which is explicitly related to the input (or output) voltage.

Although amplifiers having the ideal performance indicated are not physically realizable there are no new or unfamiliar concepts involved. Similarly the performance of passive networks, such as constitute the filter, has been extensively studied and is well understood. It is therefore appropriate to devote the following section to the third function.

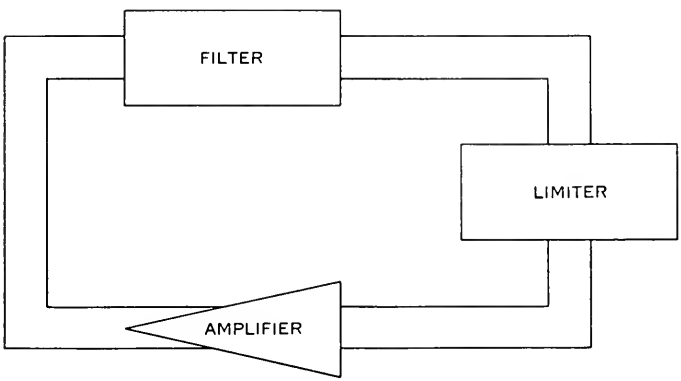


Fig. 1—Functional block diagram of an oscillator.

III. Types of Limiters

The limiters which are now in common use may be separated into four relatively distinct groups.

1. Vacuum tubes in which the gain is decreased by simple overload as the level of oscillation rises. This is the most common form of limiter.

2. Varistors in which the impedance depends upon the instantaneous value of current. Copper oxide, thyrite, and electronic diodes are examples.

3. Thermistors in which the resistance depends upon the rms value of current but does not vary appreciably during any one cycle. Carbon and tungsten filament lamps are the most common examples.

4. Vacuum tubes in which the gain is reduced by application of a bias which depends upon the level of oscillation. Usually the bias is developed by rectifying a portion of the output.

The limiters of the first two groups depend for their operation upon the generation of harmonic voltages and currents. The limiters of the second

[4] H. W. Bode, "Relations Between Attenuation and Phase in Feedback Amplifier Design," *Bell Sys. Tech. Jour.*, Vol. 19, pp. 421–457, July 1940.

two groups operate with very little harmonic distortion. The output of oscillators employing such limiters may, therefore, be made quite free from harmonic voltages. Oscillators of this sort are referred to as linear because the tube or tubes serve as simple Class A linear amplifiers.

IV. Criterion of Self Modulation

The block diagram of Fig. 1 is characterized by the fact that the separate elements are connected to each other in the form of an endless ring. The output may be assumed to come from any of the three junctions. It is this fact of closure which complicates the problem of oscillator study. For purposes of analysis it is convenient to open the loop as shown in Fig. 2. For this example it makes no difference where we choose to make the cut, but in actual circuits some caution must be exercised. This matter is dis-

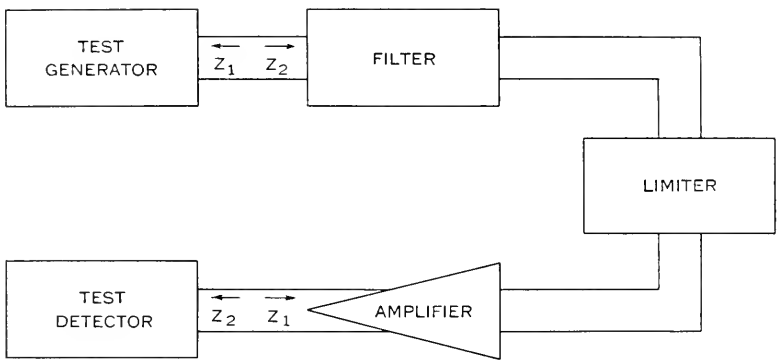


Fig. 2—Test for self-modulation in an oscillator.

cussed more fully later. It is also necessary to choose the impedances of the test generator and test detector so that the operation of the components of the original system is not disturbed.

If a continuous wave of suitable voltage and frequency is supplied by the test generator it will be found that the terminal voltage of the test detector is identical in magnitude and phase with that of the generator. In this condition the requirements which are fundamental to oscillators are satisfied. That is, the frequency and level at which oscillation should occur if the circuit were closed as in Fig. 1 have been established. The net phase shift of the system is zero and the net gain is zero.

Whether the oscillations so produced would be stable or interrupted is now determined by adding amplitude modulation of relatively low frequency and very small magnitude to the test generator. It is clear that this modulation will be transmitted through the amplifier, filter, and limiter to the test detector and that the phase and percentage of the modulation may both be

modified. By examining the transmission of a lightly modulated wave for various frequencies of modulation it is possible to determine whether or not the normal oscillation will be self modulated when the loop is closed as in Fig. 1.

The carrier is held constant at the frequency F and amplitude V for which the input and output are identical, and the frequency f of the modulation is varied from zero to infinity. In the following treatment it is assumed that the significant portion of the characteristic is observed for modulation frequencies small compared to F. The theory is simplified in this way without being seriously restricted in usefulness. The percentage of modulation must be held very low so as not to exceed the normal operating range of the limiter. The criterion is most conveniently stated in terms of the transmission of the modulation envelope which may be considered as a vector quantity.

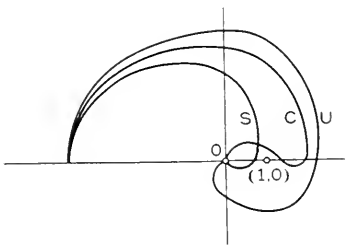


Fig. 3—Nyquist diagram showing magnitude and phase of loop transmission.
Legend: U is unstable
C is conditionally stable
S is absolutely stable

A plot of the vector ratio of output to input modulation for various frequencies is prepared as in Fig. 3. The system characterized by curve U is unstable and will generate a self modulated rather than a continuous wave. The system characterized by curve S is unconditionally stable and will be free from self modulation. The system characterized by curve C is conditionally stable and may generate either a continuous or an interrupted wave depending upon the manner in which the oscillation is started and other factors.

V. Analogy of the Oscillator to the Feedback Amplifier

The behavior of oscillators of the type here considered is entirely dependent upon feedback. It is therefore appropriate to review the fundamental principles which apply to feedback in general.

In the feedback amplifier, negative feedback is applied to improve the linearity, stability, impedance, or frequency characteristics. Considerable improvements in some or all of the properties may be secured if a consider-

able amount of negative feedback is applied and properly controlled. Positive feedback is sometimes used to increase gain or selectivity, but stability under such circumstances is poor. Any considerable amount of positive feedback results in oscillation.

The criterion by which stable feedback systems are distinguished from unstable ones has been presented by Nyquist and verified by others.[5, 6] A closed feedback system having input and output terminals is illustrated in Fig. 4. In Fig. 5 the loop is opened at some arbitrary point and a test

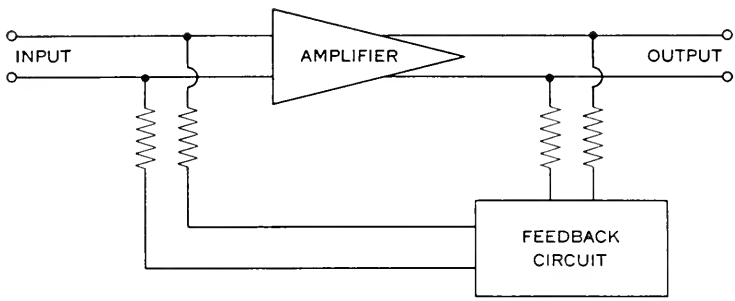


Fig. 4—Typical feedback amplifier.

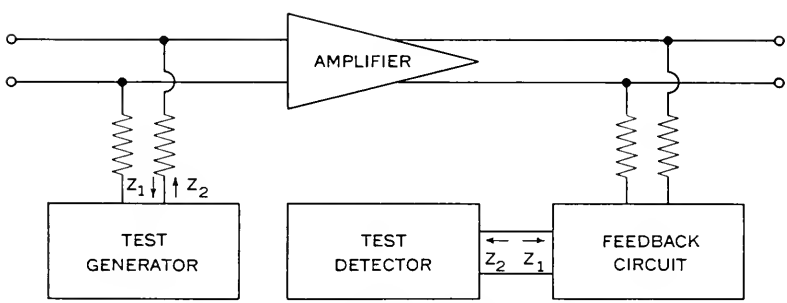


Fig. 5—Test for stability of feedback amplifier.

oscillator and detector are connected. Here as in Fig. 2 certain precautions as to impedance are observed. The test generator must produce a pure sinusoidal wave of such small magnitude that no part of the tested system overloads and the vector ratio of the detector voltage to the generator voltage is observed for a large number of frequencies. The polar plot of Fig. 3 applies directly to the feedback amplifier except that the radius vector represents the transmission of a simple wave rather than of an envelope.

[5] H. Nyquist, "Regeneration Theory," *Bell Sys. Tech. Jour.*, Vol. 11, pp. 126–147, Jan., 1932.

[6] E. Peterson, J. G. Kreer, & L. A. Ware, "Regeneration Theory and Experiment," *Proc. I.R.E.*, Vol. 22, pp. 1191–1210, Oct., 1934.

The conditions of absolute and conditional stability and instability are exactly the same as those already given.

It must be appreciated that Nyquist's criterion supplies no information as to the type or frequency of oscillations which will be generated by an unstable system. This is true because the analysis is limited to linear systems. The only information imparted is that a very small oscillation of some frequency will increase exponentially with time until the amplitude is limited by the action of some non-linear device. A small or relatively large shift of frequency may occur and the oscillation may be regular or intermittent. The present work extends the usefulness of Nyquist's criterion by using it in modified form to determine whether or not a particular unstable system (oscillator) has or lacks stability as to self-modulation. There is no apparent reason why a system lacking in both fundamental and envelope stability might not be analyzed a third time for the stability of the self-modulation.

VI. Analysis of an Oscillator having Automatic Output Control

Figure 6 presents a simple form of feedback oscillator having a separate rectifier as limiter. For small amplitudes of oscillation the tube operates in a linear fashion with cathode self-bias. No bias is produced by the diode rectifier until the peak voltage in the coil L_3 exceeds that of the bias battery B. All voltage in excess of this value is rectified, smoothed by the condenser C, and applied to the resistor r as bias. It is seen that a small percentage change in the output level may result in a large change in the bias. Accordingly an output which is quite stable with respect to the tube condition and applied voltages, except that of B, is to be expected.

The stability of this circuit with respect to self modulation is most conveniently tested by opening the oscillatory loop at the plate of the tube. In so far as the plate resistance of the tube is high with respect to that of the associated circuit it is not necessary to control the impedances of the test generator and detector extremely accurately. A block diagram equivalent to Fig. 6 is presented in Fig. 7. The conditions which must exist for the test of stability are shown in Fig. 8. In both those figures it should be noted that the gain control is actuated by the input, not the output, of the amplifier. It is therefore possible for a marked decrease of output voltage to result from a small increase of input voltage. This behavior is very different from that of the conventional, back-acting, automatic-volume-control amplifier in which the output change is in the same direction as the input change but of reduced magnitude. It is this difference which is the basis of most difficulty with amplitude controlled oscillators.

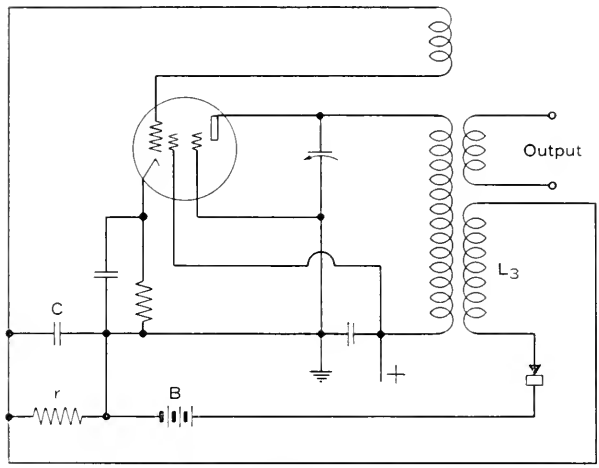


Fig. 6—Oscillator having automatic output control.

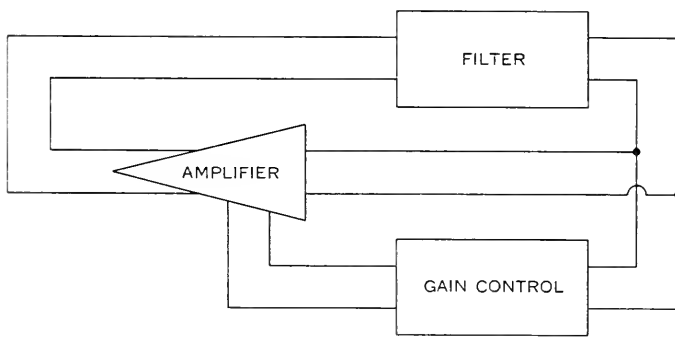


Fig. 7—Block diagram of automatic output control oscillator.

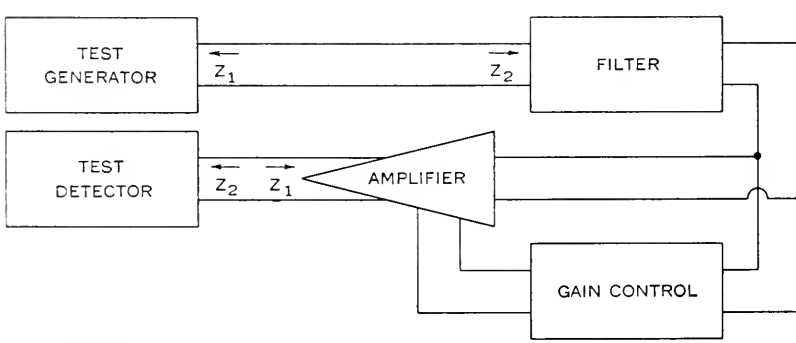


Fig. 8—Test for modulation stability of automatic output control oscillator.

Filter

The filter of Fig. 8 consists of only a single tuned circuit. Its transmission is readily represented in terms of the circuit Q by the familiar universal resonance curve. The transmission of a modulated wave through such a passive network is conveniently determined by separating the wave into its carrier and two sidebands. The carrier will be the frequency F corresponding to zero phase shift which, in this case, is also the frequency of maximum transmission. The sidebands will be shifted in phase by equal

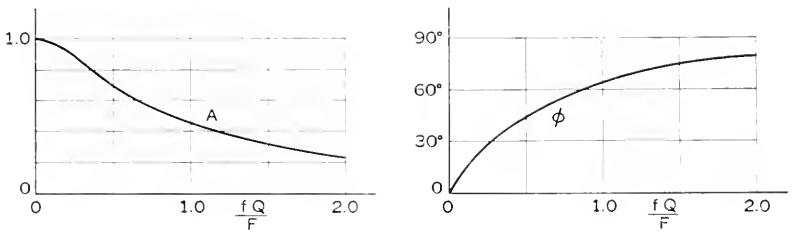


Fig. 9—Envelope transmission of a modulated wave through a single tuned circuit of selectivity Q.

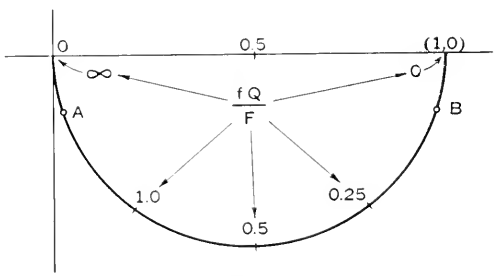


Fig. 10—Data of Fig. 9 plotted in polar form.

and opposite amounts and attenuated according to the frequency f by which they differ from the carrier. This behavior is interpreted in Fig. 9 as transmission and phase shift of the envelope. It is seen that the transmission approaches zero and the phase shift approaches 90° as the modulation frequency is indefinitely increased. The same data is presented in polar form in Fig. 10. Specifically Fig. 10 shows the vector ratio of the modulation factor m of the output wave to that of the input wave for all frequencies. In Fig. 9 the magnitude and phase angle of the ratio are shown separately.

Limiter

The limiting action of the tube and diode combination is determined by direct circuit analysis. For very low modulating frequencies the condenser

C of Fig. 6 serves only as a high-frequency by-pass; the direct voltage across r being the instantaneous difference between the peak voltage induced in L_3 and that of the stabilizing battery B. For very high modulating frequencies the modulation as well as the carrier is by-passed by C and no modulation voltage appears across r. Thus the bias is constant and the output wave is identical with the input wave. This corresponds to an envelope transmission of (1, 0). For intermediate values of the modulating frequency the voltage developed across r varies in magnitude and phase approximately as if a constant current of the modulating frequency f were applied to r and C in parallel.

The output of the amplifier depends not only upon the bias developed across r but also upon the input. For systems having a large amount of control the action of the bias is predominant. Thus for a low modulating frequency the variation of the bias overpowers the initial modulation, the phase of the modulation is reversed, and the percentage magnified by the

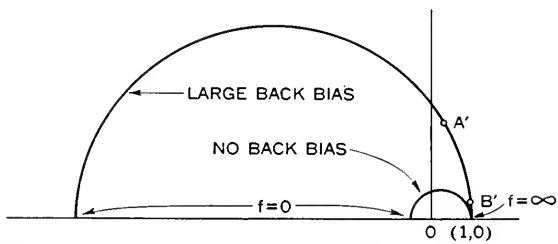


Fig. 11—Envelope transmission of a modulated wave through controlled amplifier.

action of the limiter. In Fig. 11 the envelope transmission is plotted in polar form for conditions of relatively large and relatively small amounts of control.

Loop Transmission

The separate diagrams of Figs. 10 and 11 are combined in Fig. 12 to determine the stability of the system. For any chosen frequency f the vector of Fig. 10 is multiplied by the vector of Fig. 11 corresponding to the same frequency to locate one point of Fig. 12. The resultant vector has an angle which is the sum of the two component angles and a magnitude which is the product of the two component magnitudes.

It is seen that the loop may be made to cross the axis considerably to the left of the point (1, 0) if the points A and A' of the previous figures correspond to the same frequency. Similarly the loop may be made to come very close to the point (1, 0) by increasing the size of C or lowering the Q of the tuned circuit so that the points B and B' correspond to the same frequency. With the circuit elements drawn in Fig. 6 the stability margin

may be reduced to zero, but actual looping of the point (1, 0) is not indicated. Parasitic elements, not here considered, can readily affect the performance enough to produce instability.

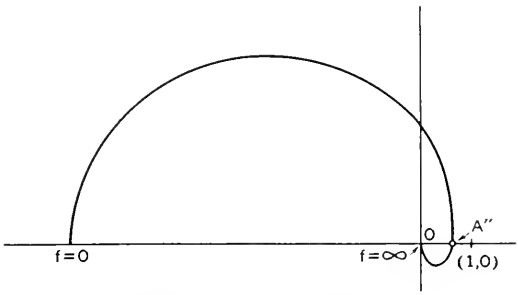


Fig. 12—Nyquist diagram applying to Fig. 6.

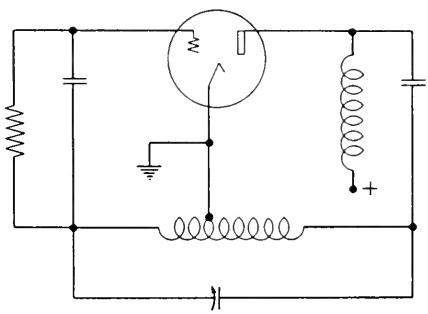


Fig. 13—Hartley circuit.

VII. Analysis of the Hartley Oscillator

The familiar Hartley Oscillator circuit is shown in Fig. 13. In this arrangement the tube serves as amplifier and limiter by the action of overloading. Harmonic voltages and currents are produced but if the selectivity of the tuned circuit is high the voltage returned to the grid of the tube is nearly sinusoidal.

The stability of this circuit is tested in exactly the same way as was that of the previous circuit. The loop is opened at the plate of the tube to determine the transmission of a modulated signal. If, as is usually the case, the coupling of the coil is close, the filter reduces to a single tuned circuit. The limiting action results from bias produced by rectification at the grid. Accordingly the block diagram of Fig. 7 is directly applicable, and the behavior of the filter is correctly given by Fig. 9.

Generally the circuit operates in class "C" with high bias and large grid

voltage swings. If the time constant of the grid-leak-condenser combination is long in comparison to the period of a modulation cycle the bias will not be able to follow the applied voltage and the modulation of the output will be larger than that of the input. Moreover it is in phase with that of the input. When the modulating frequency is low the bias is able to follow the level of modulation and the output modulation is very small. Thus the transmission of a modulated signal is greatest at high modulating frequencies, and the modulation output is in phase with the input. Because of the

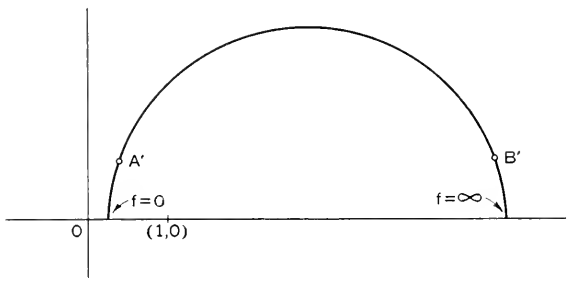


Fig. 14—Envelope transmission of a modulated wave through a grid-leak-biased Class C amplifier.

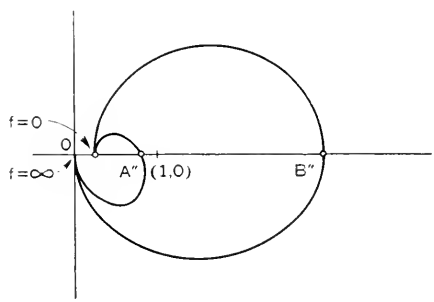


Fig. 15—Nyquist diagram applying to Fig. 13.

action of the grid-leak-condenser network a phase shift at intermediate modulating frequencies occurs. This behavior is represented in polar form in Fig. 14.

The stability of the system is determined by combining in Fig. 15 the separate diagrams of Figs. 14 and 10. As in the previous system a thoroughly stable system results if the element values are such that the points A and A' of Figs. 10 and 14 correspond to the same frequency. If on the other hand the elements are such that B and B' correspond to the same frequency the curve loops (1, 0) indicating instability. In general stability is

promoted by increase of the Q of the tuned circuit and by decrease of the time constant of the grid-leak-condenser combination.

VIII. The Lamp Stabilized Oscillator

The circuit of Fig. 16 is of particular interest because the functions of amplifier, limiter, and filter are performed separately by units which are readily identified with their functions. The present method of analysis was developed in connection with this particular circuit. The output frequency and amplitude are both quite stable and the harmonic content of the output is low.

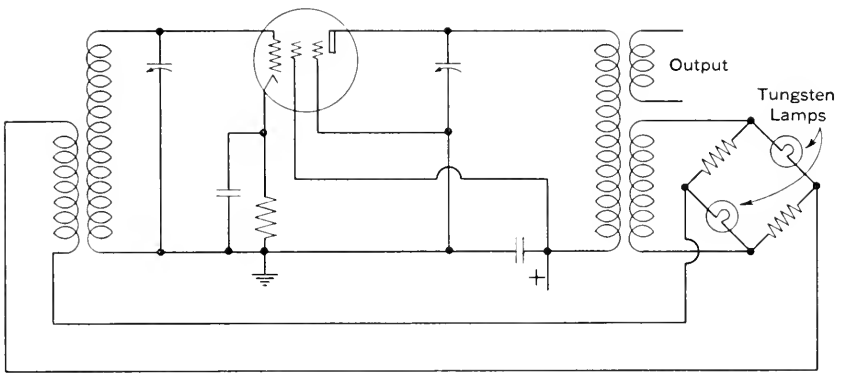


Fig. 16—Schematic diagram of lamp stabilized oscillator.

Under operating conditions the gain of the tuned amplifier, which is ordinarily in the order of 40 db, is equalled by the loss of the lamp bridge. The lamps operate at such a temperature that their resistance is slightly less than that of the associated linear resistors. If the gain of the amplifiers is for any reason somewhat reduced, the current through the lamps decreases, the temperature and resistance of the lamps is reduced, and the loss through the bridge is reduced to the new value of amplifier gain.

The d-c characteristic of a lamp bridge is shown in Fig. 17. A curve identical with Fig. 17 is observed if the measurement is made with an alternating current whose period is very short in comparison to the thermal time-constant of the filaments. Up to L the operation is nearly linear. In the region of M the output is essentially independent of the input. At N the bridge is nearly balanced and a small percentage change in the input voltage results in a large and opposite percentage change in the output. It is thus seen that an alternating current having a small superimposed modulation of low frequency will result in an output having a considerably

larger percentage modulation in the opposite phase. When the modulation frequency exceeds a few hundred cycles the lamps are unable to follow the individual cycles and the output wave is identical in form to the input. At intermediate modulating frequencies the transmission of a modulated wave

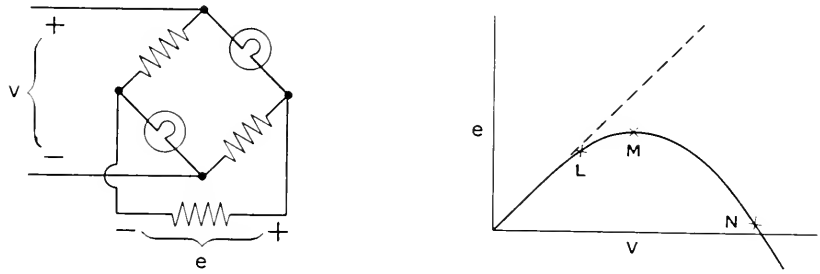


Fig. 17—D-C characteristics of a lamp bridge.

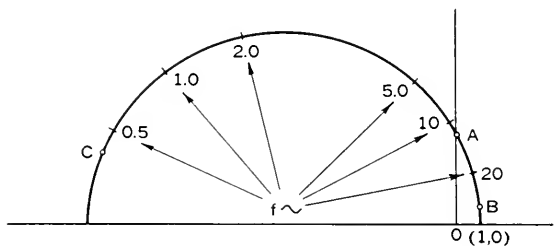


Fig. 18—Envelope transmission of a modulated wave through a lamp bridge.

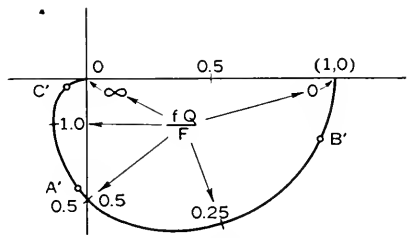


Fig. 19—Envelope transmission of a modulated wave through two similar tuned circuits of selectivity Q.

involves a phase shift. The behavior of a typical lamp bridge is presented in Fig. 18.

If the Q of the grid and plate circuits are both relatively high the filter circuit may be taken as equivalent to two separate tuned circuits. The transmission of each is given by Fig. 9. The combined transmission of the pair is given in polar form in Fig. 19. Because two tuned circuits are

employed, the diagram of Fig. 19 differs markedly from that of Fig. 10. Specifically the phase shift corresponding to a given value of attenuation is greatly increased. As in previous cases the curve of over-all loop transmission may or may not loop the point (1, 0) depending upon the relative frequency scales. Thus if the points A and A' of Figs. 18 and 19 correspond to the same frequency the Nyquist diagram passes near the point (2, 0) indicating instability. If the points B and B' correspond to the same frequency the loop passes very near to the point (1, 0) and instability is likely.

By making the tuned circuits very selective or by reducing the thermal time constant of the lamp circuit the points C and C' may be made to correspond to the same frequency. In this case the loop passes to the left of the point (1, 0) and the system is absolutely stable. The same result may be secured more easily by making one of the tuned circuits much more selective than the other. This is ordinarily accomplished by increasing the Q and impedance level of the grid circuit while keeping the Q and impedance level of the plate circuit much lower so as to provide a suitable power output to operate the lamp bridge.

IX. The Varistor Stabilized Oscillator

A circuit which differs from that of Fig. 16 only in that the lamps are replaced by varistors is shown in Fig. 20. At low levels of oscillation the impedance of the varistors is relatively high, the loss of the limiter is low and the amplitude of oscillation rises. At some higher level the varistor impedance is reduced, the bridge approaches balance to the fundamental frequency, and a stable condition is reached. Because the initial unbalance of the bridge is opposite to that of Fig. 16 a reversal of phase is necessary to establish oscillation.

The stable condition reached differs from that of the lamp stabilized oscillator in that the varistor goes through its entire range during each high-frequency cycle. The lamp resistance changes by only a small amount during any one cycle, its resistance depending on an integration of many previous cycles. Two important facts arise from this difference. Harmonics are produced in the bridge and, in so far as the varistors face reactances of these harmonic frequencies, intermodulation may produce currents of fundamental frequency but shifted in phase with respect to the original. Thus the bridge may produce a phase shift which is a function of level of the oscillation frequency. A degradation of frequency stability results from such a condition. More important to the present problem is the fact that all modulation frequencies are transmitted alike. A small modulation is reversed in phase and magnified by an amount depending upon the bridge balance but not upon the modulation frequency.

Because the limiter introduces no phase shift it follows that the envelope loop transmission is merely an enlarged and reversed copy of that for the filter. This can loop the (1,0) point only if there are at least three shunt elements in the filter section. That is, instability can result only if the phase shift of the filter system exceeds 180° for frequencies relatively near the operating frequency. This circuit is therefore much less likely to produce intermittent operation than any other circuit here considered.

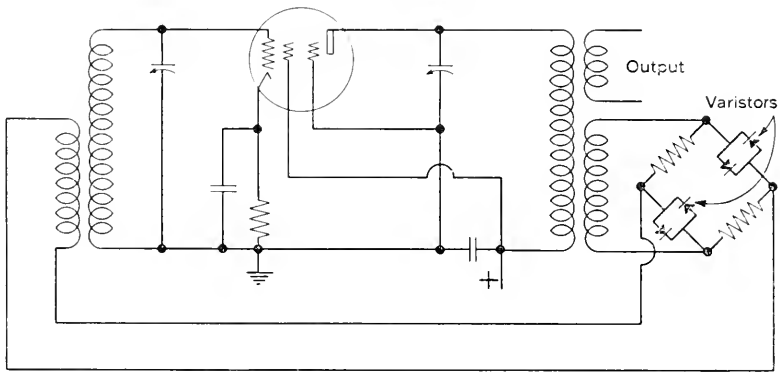


Fig. 20—Schematic diagram of varistor stabilized oscillator.

X. Negative Feedback in Oscillators

Because positive feedback is the necessary condition for the operation of an oscillator it is not obvious that the application of negative feedback is ever desirable. Actually it is frequently possible to introduce negative feedback into an oscillator with no loss of performance and under certain circumstances advantages are gained.

The circuit of Fig. 16 serves as a convenient example. Removal of the cathode by-pass condenser is likely to reduce the amplifier gain by about 6 db and to increase the stability of the gain with respect to applied voltages by a corresponding amount. Coincident with removal of the by-pass condenser the operating level drops a small amount, the bridge loss decreases 6 db to reestablish equilibrium, and the stabilizing effect of the bridge is cut in half. Accordingly the over-all stability of the output with respect to applied voltages is unchanged. The advantages gained are that the loss which must be held in the bridge is reduced so that stray reactances are less likely to disturb the operation, and that the harmonic content of the output is reduced.

Stated in a different way, the output stability of an oscillator using a non-feedback amplifier is limited in practice by the bridge balance which may be maintained. After this value of gain has been reached additional stability

may be secured by supplying increased inherent gain which is offset by direct negative feedback.

XI. Design of a Controlled Oscillator

To clarify the material already presented and to convey some additional concepts an oscillator having a large amount of control will be designed. The block diagram is to be that of Fig. 7 and the circuit is to be similar to that of Fig. 6.

It may readily be seen that the gain control must satisfy two fundamental requirements. It must deliver a d-c bias which increases rapidly with increase of the level of oscillation and it must not return any appreciable voltage of oscillation frequency. Otherwise the frequency will be affected by the elements in the control circuit as well as those in the filter, and the performance will be generally poor. Because of its balance a push-pull rectifier is helpful in meeting the latter requirement. The principal requirement is achieved by amplification and by the use of a constant counter emf or back bias. No bias is produced until the level of oscillation exceeds some threshold value. Above this threshold the bias increases approximately volt for volt with the peak value of the signal. The same amplifier which is used to increase the control may be used advantageously as a buffer so that appreciable power outputs may be produced without degrading the frequency or amplitude stability.

It will be assumed that a Q of 100 is available in the coil and that a frequency of one megacycle is to be generated. The transmission of a modulated wave in terms of the sideband displacement through such a one-circuit filter is shown in Fig. 21. Because the cutoff occurs very slowly it will be convenient to incorporate a rapid cutoff in the auxiliary filter of the gain control, thus avoiding an excessive phase shift at any one frequency.

The circuit features already discussed are shown in Fig. 22. A basic oscillator with a single tuned coil, a buffer amplifier having little selectivity and therefore contributing very little to the equivalent filter section, a source of biasing voltage, a balanced rectifier, and an auxiliary low-pass filter are shown. The condenser C is only large enough to allow the rectifier to be driven without serious loss at one megacycle. It has relatively little effect upon the modulation performance.

It is assumed that the buffer-amplifier, rectifier, etc. are so chosen that a modulation of very low frequency of one part per million applied at the plate terminal of the oscillator will result in a modulation of one part in a thousand returned to that point. This is equivalent to saying that the envelope gain is 60 db at low frequencies, and corresponds to 60 db of negative feedback in a conventional amplifier.

The auxiliary filter will be designed to approximate the attenuation and

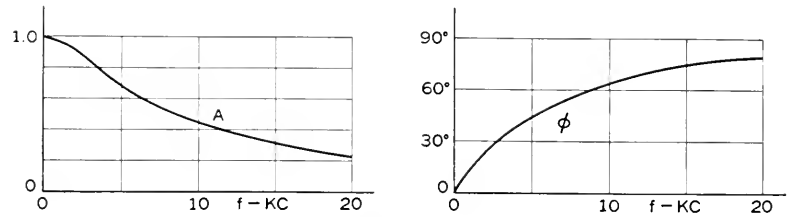


Fig. 21—Envelope transmission through tuned circuit.

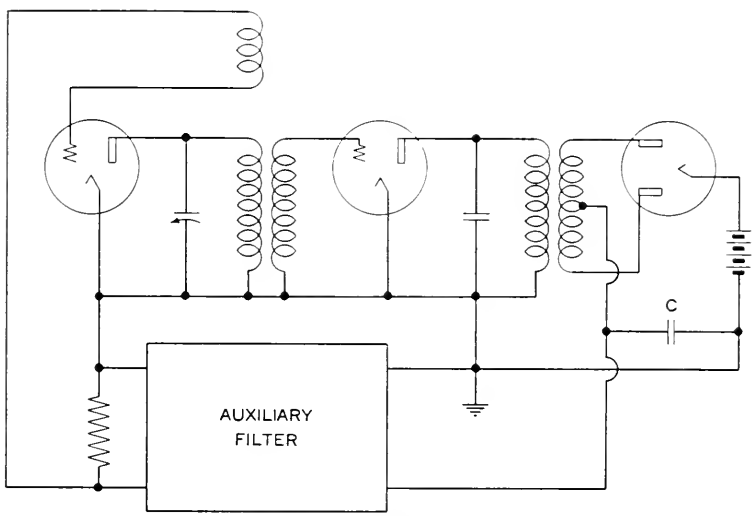


Fig. 22—Special A-V-C oscillator.

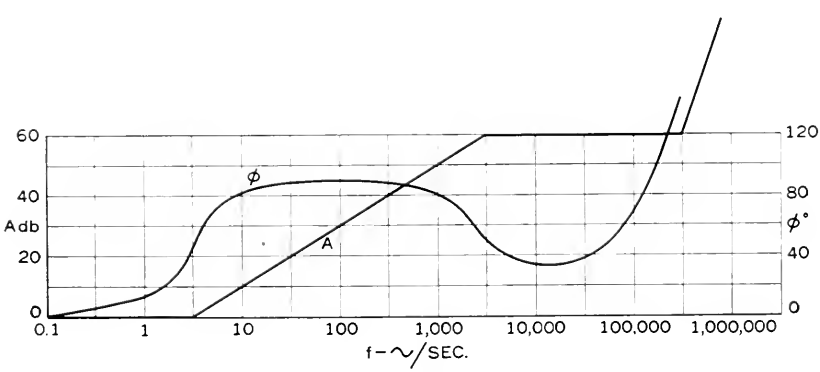


Fig. 23—Characteristics of auxiliary filter.

phase characteristics shown in Fig. 23. The choice of this particular shape is best explained by reference to Fig. 24 which presents the over-all envelope loop transmission of the system. It is seen that the phase shift is relatively constant at 90° over a wide band of frequencies and that the gain falls off approximately linearly over the same band. In particular the gain becomes zero around 5000 cycles whereas the phase does not reach zero below 500,000

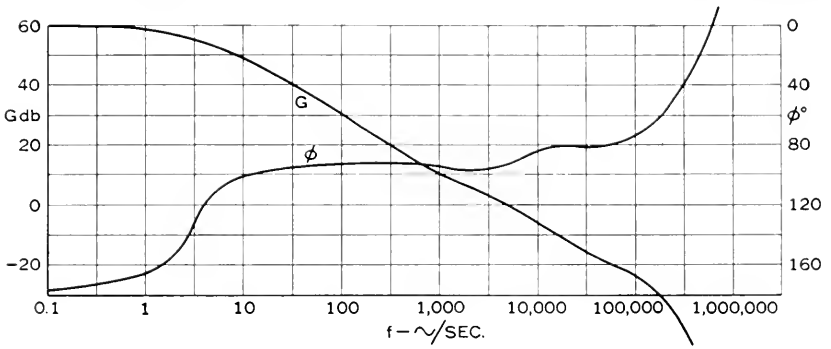


Fig. 24—Overall envelope transmission of Fig. 22.

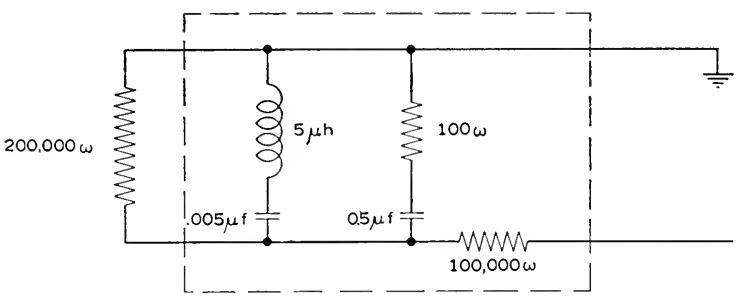


Fig. 25—Configuration of auxiliary filter.

cycles. In terms of Nyquist's criterion this represents a very stable system which is little disturbed by transient effects. A system having even greater stability could be achieved by beginning the cut-off at lower frequencies. It would then be found that the output was somewhat sluggish in reaching a new equilibrium after being disturbed. Such a behavior is not uncommon but is generally undesirable.

Elements which give approximately the characteristics called for in Fig. 23 are shown in Fig. 25. The peak of loss at one megacycle is contributed by the series resonant trap. The rest of the behavior is due to the 0.5 μf condenser in combination with the associated resistors.

XII. Auxiliary Control of Thermally Limited Oscillators

In the Meacham and certain other oscillator circuits a thermistor is associated with reactive elements in a bridge circuit which functions as both limiter and filter. In these circuits a large increase in the frequency stability is observed. This may sometimes be conveniently expressed as a magnification of the effective Q of the filter.

The advantages of great frequency stability and good amplitude stability of these systems are accompanied by an undesirable tendency toward intermittent operation. The thermal constants of the thermistor are not readily adjustable. Moreover adjustment of the reactances to secure

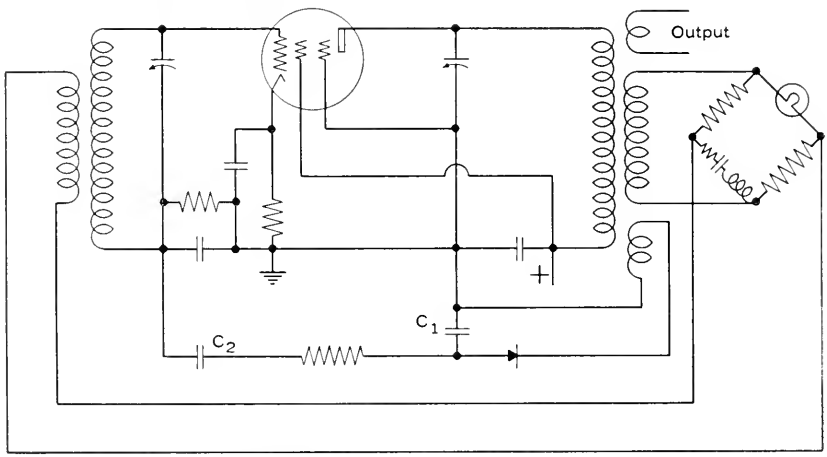


Fig. 26—Meacham circuit with auxiliary control.

suitable envelope stability is likely to impair the frequency or amplitude stability for which the circuit is chosen.

This dilemma may be resolved by the addition of an auxiliary network which does not affect the envelope transmission to very low frequencies but does modify the behavior at higher frequencies in such a way as to promote the stability of the system.

A simple circuit illustrating the principle appears in Fig. 26. It will be noticed that the circuit is so arranged that the average bias applied to the tube is only that due to the cathode resistor. The steady voltage developed across C_1 by the rectifier is unable to affect the bias because of the blocking condenser C_2. Accordingly the rectifier circuit does not affect the normal operating condition, which is characterized by a bridge loss equal to the amplifier gain. The added elements come into play only if there is a tendency toward self-modulation. Then displacement currents of modulation frequency flow through C_2 in such a magnitude and phase as to modify the tube gain and compensate the modulation returned from the bridge.

The exact nature of the control which must be added is best ascertained by opening the circuit at the plate of the tube. The loop transmission of a modulation envelope may then be determined, either experimentally or analytically. If instability is found an auxiliary circuit must be designed to produce an over-all system which is stable. In general the elements of the auxiliary circuit are to be chosen so that the loop transmission is considerably less than unity in the region of zero phase. This is ordinarily accomplished by increasing the final cutoff frequency at which the over-all loop envelope transmission is negligible.

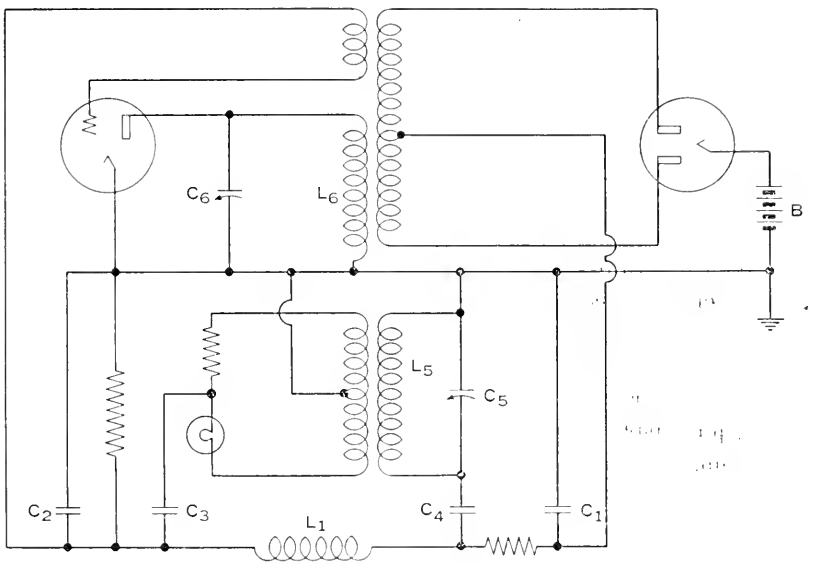


Fig. 27—Self-modulating oscillator.

XIII. A Self Modulated Oscillator

The previous sections have been devoted primarily to the problem of preventing self-modulation in oscillators. Let us now consider an oscillator having envelope instability. The Nyquist diagram indicates that self-modulation will occur and tells the approximate frequency of the envelope wave. More detailed analysis of the circuit is necessary to determine the wave form of the envelope and the manner in which its amplitude is limited.

If a circuit is to function well as an oscillator the Nyquist diagram for the operating frequency must loop the $(1, 0)$ point with considerable margin. This is necessary so that a small loss of gain will not stop oscillation. At the operating level the limiter reduces the loop transmission to unity. In the region of $(1, 0)$ amplitude stability is favored if the rate of change of gain

with respect to level is high. Similarly the frequency stability is favored if the rate of change of phase with respect to frequency is high.

If a circuit is to function well as a self-modulated oscillator, the above conditions must be met and in addition the Nyquist diagram for the envelope must meet similar requirements. That is, there must be a limiter and filter in addition to the effective amplifier in the envelope system.

A circuit which meets these requirements is shown in Fig. 27. It is seen to be similar to that of Fig. 6 but to have a more complicated low-frequency path. The operation is best explained in terms of the relative size of the various elements. The by-pass condensers C_1 and C_2 are comparatively small. The blocking condensers C_3 and C_4 are quite large. The choke L_1 is large. Thus these elements serve as open or short circuits but do not enter into the setting of either of the frequencies.

The stability tests are carried out by opening the mesh at the plate of the tube. At the operating frequency, as defined by the plate coil and condenser the loop gain is high at low levels. Thus the fundamental conditions for oscillation exist.

The next step in the analysis is to supply a signal of suitable magnitude and frequency to reduce the loop transmission to $(1, 0)$. A small modulation of very low frequency is returned magnified and reversed in phase, as with previous systems. The phase of the envelope transmission changes with increase of modulating frequency until it is zero at the resonant frequency of L_5 and C_5. At this frequency a considerable gain exists so that the Nyquist diagram for the envelope also loops the point $(1, 0)$.

The tungsten lamp in conjunction with the other impedances of the bridge serves to limit the degree of self-modulation. The operating frequency may be set by means of C_6 in conjunction with a suitable value of L_6. The operating amplitude may be controlled by adjustment of the bias battery B. The frequency of the self-modulation is set by means of C_5 in conjunction with L_5.

XIV. Conclusions

A method of applying known feedback theory to the problem of self-modulation in oscillators has been presented. Although the discussion has been limited to electrical circuits it is clear that the analysis is applicable to other systems, such as electromechanical or mechanical oscillators.

The analysis has been applied to several familiar oscillators to illustrate the method and to clarify some details of their operation. A sample design of a bias controlled oscillator is presented to show application to new designs.

The application of bias control to thermistor stabilized oscillators is described. The design of a self-modulated oscillator is undertaken to show how intentional modulation may be introduced and controlled.

Evaluating the Relative Bending Strength of Crossarms

By RICHARD C. EGGLESTON

OVER a million crossarms are produced annually in the United States. In the open wire lines of the Bell System alone there are now about 20 million arms in use. It is natural, therefore, that public utility engineers should have an interest in the strength of such an important item of outside plant material; and, consequently, an interest in any tool or means of evaluating the strength of such material. It is believed that the moment diagram is a convenient and reasonably reliable tool for estimating the loads an arm will support, for measuring the effect of knots of various sizes and of pinhole locations on arm strength, and for answering similar questions relating to the bending strength of crossarms under vertical loads.

Two moment diagrams are shown in Fig. 1 for Bell System Type A crossarms; and in the pages that follow are presented the method used in constructing the diagrams and a discussion of their use. While the calculation results apply particularly to the type and quality of arm referred to, they would also be of value as a time saving reference in future studies that may be proposed relating to the strength of the same or other types of arms involving different knot allowances.

The resisting moment of a beam is the product of its section modulus by the unit stress on the remotest fiber of the beam. The section modulus of a beam of uniform cross-section is constant and readily determinable. The section modulus, however, of a beam of nonuniform cross-section, such as a crossarm, varies because of the different cross-sectional shapes and dimensions involved.

In this study the following five different shapes were recognized:

(1) Roofed section between pinholes
(2) Roofed pinhole section
(3) Roofed brace bolt hole section
(4) Rectangular pole bolt hole section
(5) Rectangular section without bolt holes

The dimensions of the sections investigated were as follows:

Section of Arm	Dimensions	
	Minimum	Nominal
	(Inches)	(Inches)
Roofed section, except at end of arm	$3\frac{3}{16}$ x $4\frac{3}{32}$	$3\frac{1}{4}$ x $4\frac{3}{16}$
Roofed section at end of arm	$3\frac{3}{16}$ x 4	$3\frac{1}{4}$ x $4\frac{3}{16}$
Unroofed sections	$3\frac{3}{16}$ x $4\frac{3}{16}$	$3\frac{1}{4}$ x $4\frac{1}{4}$

Since there is little, if any, engineering interest in the strength of structural members of maximum size, no investigations were made of sections of maximum dimensions.

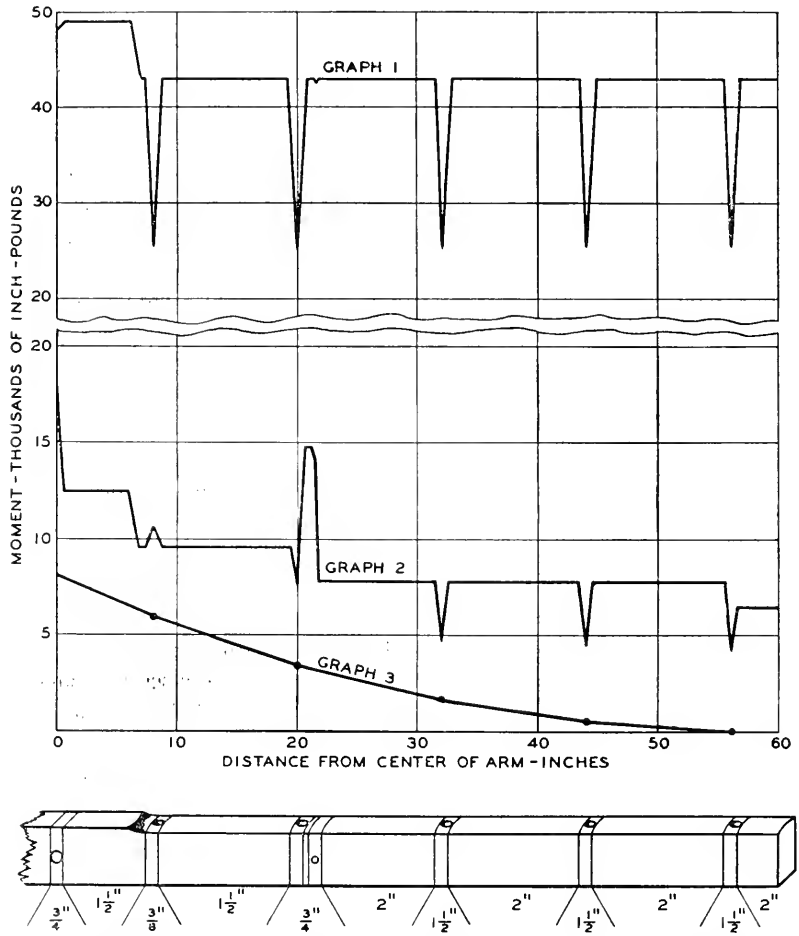


Fig. 1—Moment diagram for Type A southern pine and Douglas fir crossarms per Specification AT-7075:

Graph 1—Resisting moments of arms of nominal dimensions, straight grained and free from knots. (Fiber stress 5000 psi)

Graph 2—Resisting moments of arms of minimum dimensions, having maximum slant grain (1″ in 8″), and containing knots of the maximum sizes permitted (viz., sizes shown at bottom of arm sketch). (Fiber stress 3250 psi)

Graph 3—Bending moments from a load of 50 pounds at each pin position.

Section modulus calculations were made of each shape of minimum and nominal size, both with and without knots. Tests have shown that, be-

cause of the distortion of the grain that occurs around them, knots are fully as injurious to the strength of structural timbers as knot holes.[1] Therefore, in dealing with sections containing knots, it was assumed for the purposes of this study that the knot extended across the section in the same manner as a hole having a diameter equal to the diameter of the knot. It was also assumed that the knot was located in, or reasonably close to, the most damaging position in the arm section.

In the calculations of the section modulus of all roofed arm sections, it was necessary first to compute the moments of inertia of the whole or parts of the top segments of such sections (viz. nominal and minimum sections

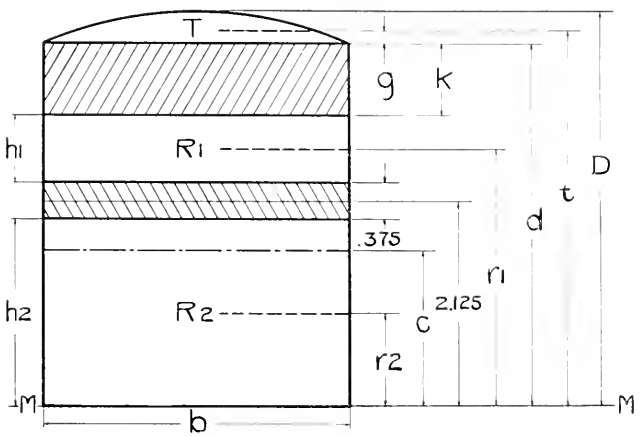


Fig. 2—Brace bolt hole section containing a $\frac{3}{4}$ inch knot located immediately below the top segment (knot and bolt hole shaded).

between pinholes, and nominal and minimum pinhole sections). Accordingly, four such computations were made and the results used in calculating the section moduli of all the roofed sections investigated. The details of the four computations are shown in the Appendix. To insure uniformity in the results, the degree of precision used in these computations was considerably greater than is ordinarily employed in dealing with timber products. All of the work, however, was done on a computing machine, and it was just about as easy to carry the operations to eight decimal places (which was the capacity of the machine used) as to a lesser number. As a matter of interest in this connection, it was found by actual trial in Computation I that absurd results would occur if fewer than five decimal places were used.

For convenience, all of the section modulus calculations were made in tabular form. In such form the procedure employed would not be readily

[1] Pg. 6 Dept. Circular 295, U. S. Dept. of Agriculture, "Basic Grading Rules and Working Stresses for Structural Timbers," by J. A. Newlin and R. P. A. Johnson.

apparent. Therefore, a sample calculation follows showing the method of finding the section modulus of the brace bolt hole section containing a $\frac{3}{4}$ inch knot.

Sample Calculation

Referring to Fig. 2, it will be noted that the knot and bolt hole divide the section into three parts: the top segment (T) and two rectangular portions ($R1$ and $R2$). The moment of inertia (I) of such a compound section about its neutral axis (at a distance c from M-M) is equal to the sum of the moments of inertia (IT, $IR1$ and $IR2$) of the component parts T, $R1$ and $R2$ about axes through their own centers of gravity, plus the areas of the component parts multiplied by the squares of the distances of their own centers of gravity from the neutral axis of the compound section. The section modulus (S) of this section is found, of course, by dividing its moment of inertia by the distance (y) from the neutral axis of the section to the most remote fiber.

Dimensions:

$b = 3.1875''$ (Width of Section)
$k = 0.7500''$ (Diameter of Knot)
$d = 3.7625''$ (See Computation I in Appendix)
$h1 = 0.7000''$ $(d - 2.125'' - 0.1875'' - k)$
$h2 = 1.9375''$ $(2.125'' - 0.1875'')$
$g = 0.1330''$ (See Computation I)
$t = 3.8955''$ $(d + g)$
$r1 = 2.6625''$ $(\frac{1}{2}\ h1 + 2.3125'')$
$r2 = 0.96875''$ $(\frac{1}{2}\ h2)$
$D = 4.09375''$ (Depth of Section)

Areas:

$T = 0.7099$ sq. ins. (See Computation I)
$R1 = 2.2313$ " $(bh1)$
$R2 = 6.1758$ " $(bh2)$

9.1170 sq. ins.

Moments about $M - M$:

$Tt = 2.7654$
$R1r1 = 5.9408$
$R2r2 = 5.9828$

$14.6890 = 9.1170\ c$; and hence
$c = 1.6112$

Moments of Inertia:

$IT = 0.0053$ (See Computation I)
$IR1 = 0.0911$ $(bh1^3 \div 12)$
$IR2 = 1.9319$ $(bh2^3 \div 12)$
$T(t - c)^2 = 3.7043$
$R1(r1 - c)^2 = 2.4661$
$R2(c - r2)^2 = 2.5490$

$I = 10.7477$
$y = 2.48255$ (D − c)

Section Modulus:

$$S = \frac{I}{y} = 4.3293$$

The same general procedure shown in this sample calculation was followed in dealing with the other cross-sectional shapes. For this reason, only the final results of the several calculations are presented; although, for

TABLE 1.—*Section Modulus of Roofed Sections between Pinholes*

	Knot Diameter—Inches								
	No Knot	½	¾	1	1¼	1½	2	2½	3
Calculation 1: (Knots located at top of section) Section Size*:									
Minimum		6.86		5.08		3.57	2.33	1.35	0.64
End. Min.				4.78			2.13		0.53
Nominal		7.37		5.50		3.91	2.59	1.54	0.76
Calculation 2: (Knots located at bottom of section) Section Size*:									
Minimum		6.11	5.24	4.42	3.71	3.05	1.92	1.05	0.47
Calculation 3: (Knots located immediately below top segment) Section Size*:									
Minimum	8.03	5.45	4.56	3.86	3.34	2.95	2.50	2.34	2.37
End. Min.	7.65			3.65					
Nominal	8.60			4.16					

TABLE 2.—*Section Modulus of Roofed Pinhole Sections*

	Knot Diameter—Inches						
	No Knot	¼	½	¾	1	1½	2
Calculation 4: (Knots vertical) Section Size*:							
Minimum	4.50	3.84	3.21	2.63	2.25		
End. Min.	4.29			2.50			
Nominal	5.11			2.88			
Calculation 5: (Knots horizontal) Section Size*:							
Minimum		3.63	2.96	2.40	1.97	1.41	1.11
End. Min.				2.26		1.33	
Nominal				2.76		1.64	

* Section Sizes:
Minimum = $3\frac{3}{16}'' \times 4\frac{3}{32}''$
End. Min. = $3\frac{3}{16}'' \times 4''$ (viz. minimum at end of arm)
Nominal = $3\frac{1}{4}'' \times 4\frac{3}{16}''$

convenience, reference is made to the calculations by number in the pages that follow. These results are shown in Tables 1, 2, 3 and 4 and a brief discussion of the scope and use made of them follows.

TABLE 3.—*Section Modulus of Bolt Hole Sections*

	Knot Diameter—Inches								
	No Knot	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{3}{4}$	1	$1\frac{1}{4}$	$1\frac{1}{2}$	
Calculation 6:									
Brace bolt hole section									
Section Size*:									
Minimum	7.97	6.47		5.28	4.33	3.58		2.62	
Nominal	8.55				4.71			2.78	
Calculation 7:									
Pole bolt hole section								$2\frac{1}{2}$" *Knot*	3" *Knot*
Section Size*:									
Minimum	9.25		7.42		5.63		3.24	1.51	.75
Nominal	9.74				6.05		3.61	1.66	.85

* Section Sizes:
Minimum = $3\frac{3}{16}$" x $4\frac{3}{32}$"
End. Min. = $3\frac{3}{16}$" x 4" (viz. minimum at end of arm)
Nominal = $3\frac{1}{4}$" x $4\frac{3}{16}$"

TABLE 4.—*Section Modulus of Rectangular Section without Bolt Holes (Calculation 8)*

Section Size	Knot Diameter	Section Modulus
Minimum ($3\frac{3}{16}$" x $4\frac{3}{16}$")	(No Knot)	9.32
	$\frac{1}{4}$	8.24
	$\frac{1}{2}$	7.22
	$\frac{3}{4}$	6.28
	1	5.40
	$1\frac{1}{2}$	3.84
	2	2.54
	$2\frac{1}{2}$	1.51
	3	.75
Nominal ($3\frac{1}{4}$" x $4\frac{1}{4}$")	(No Knot)	9.78
	$\frac{1}{4}$	8.67
	$\frac{1}{2}$	7.62
	$\frac{3}{4}$	6.64
	1	5.72
	$1\frac{1}{2}$	4.10
	2	2.74
	$2\frac{1}{2}$	1.66
	3	.85

ROOFED SECTIONS BETWEEN PINHOLES

As indicated in Table 1, three tabular calculations were made for roofed sections between pinholes. In Calculations 1, 2 and 3 it was assumed that the knots present were located (1) at the top, (2) at the bottom, and (3) immediately below the top segment of the section, respectively. The results relating to the $3\frac{3}{16}$" x $4\frac{3}{32}$" section are plotted as Curves 1, 2 and 3, respectively, in Fig. 3.

With respect to the knot positions considered, it is apparent from an examination of the three curves (Fig. 3) that knots up to approximately $1\frac{1}{2}$″ in diameter are most damaging when located immediately below the roofed portion of the arm; and that the worst position for knots over $1\frac{1}{2}$″ in diameter is at the bottom of the arm. However, since under usual loading

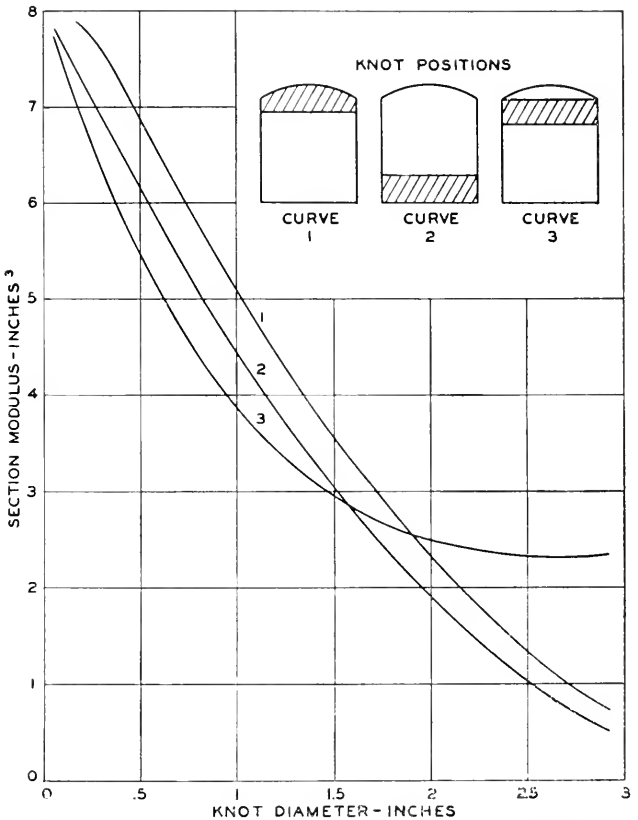


Fig. 3—Sections between pinholes. Section modulus of crossarm sections containing knots of the sizes shown on the base line and located in the positions indicated. The data apply to sections of minimum size ($3\frac{3}{16}$″ x $4\frac{3}{32}$″).

conditions knots at the bottom of an arm section are in compression, and thus would have less influence on strength than they would have on the tension side,[2] it was felt that the strength value shown by Curve 2 may be ignored; and that the values shown by a smooth curve, combining the values

[2] On Page 69 of U. S. Dept. of Agriculture Tech. Bul. 479, "Strength and Related Properties of Woods Grown in the United States" by L. J. Markwardt and T. R. C. Wilson, is the following statement: "Knots have approximately one-half as much effect on compressive as on tensile strength."

of Curve 3 up to the $1\frac{1}{2}''$ knot point with those of Curve 1 for 2″ and larger knots, would be the practical minimum section moduli for roofed sections between pinholes. Accordingly, such a smooth curve was constructed and is shown as Curve 2 in Fig. 4. The results of Calculations 1 and 3 for nom-

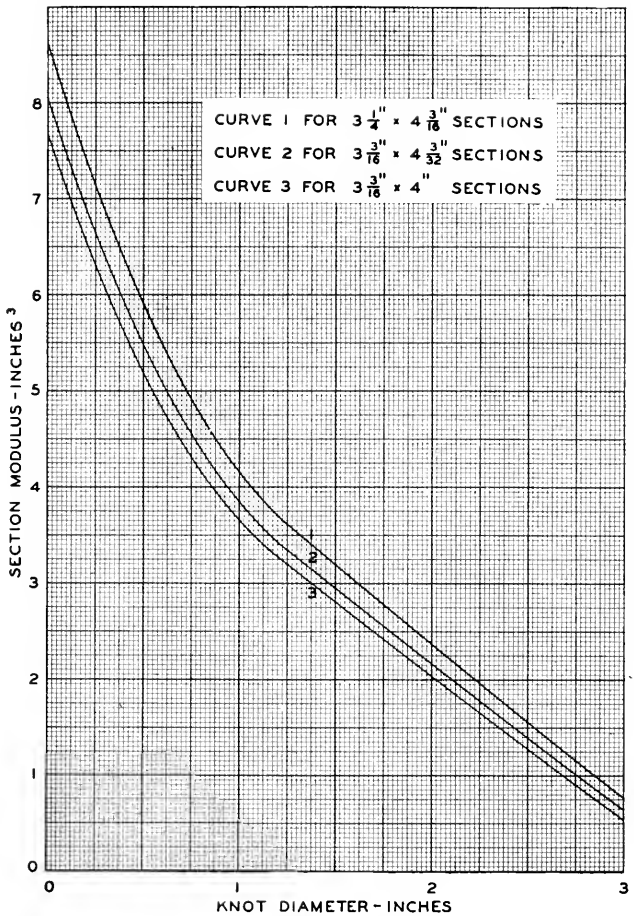


Fig. 4—Sections between pinholes. Section modulus of crossarm sections containing knots of the sizes shown on the base line and located in damaging positions.

inal and arm-end minimum sections were also plotted, and Curves 1 and 3 drawn for those sections.

Roofed Pinhole Sections

Two calculations were made for the pinhole sections: Calculation 4, in which the knots were assumed to be located adjacent to the pinhole in a

vertical position; and Calculation 5, in which the knots were assumed to be immediately below the top segment in a *horizontal* position. The results of these two calculations are shown in Table 2. It has heretofore been generally assumed that in pinhole sections knots less than 1″ in diameter were more damaging in a vertical position than in a horizontal position. The results of Calculations 4 and 5, however, show that the horizontal knots immediately below the top segment are the more damaging. In order to compare the effect of knots so located with the effect of knots at the extreme

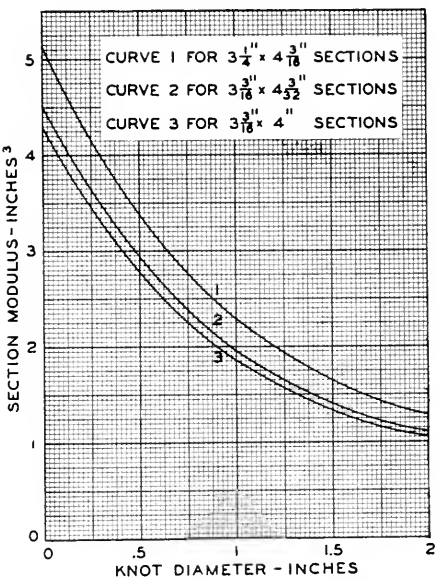


Fig. 5—Pinhole sections. Section modulus of crossarm sections containing knots of the sizes shown on the base line and located in damaging positions.

top of the section, the following two computations assumed 1″ and 2″ horizontal knots at the latter location:

1″ Knot at Section Top:

$$\tfrac{1}{2}S = \frac{.02875\ (3.09375)^2}{6} = 1.48156$$

$$S = 2.9631$$

2″ Knot at Section Top:

$$\tfrac{1}{2}S = \frac{.92875\ (2.09375)^2}{6} = .67857$$

$$S = 1.3571$$

As the section modulus (S) values for sections containing 1″ and 2″ horizontal knots located immediately below the top segment are 1.97 and 1.11, respectively, (Calculation 5, Table 2) it is apparent that in pinhole sections horizontal knots immediately below the top segment are the more damaging. The results of Calculation 5 were accordingly plotted in Fig. 5 and smooth curves drawn to show the section modulus for each of the three sections containing knots of any size.

Roofed Brace Bolt Hole Section

The worst position for knots in the brace bolt hole section was assumed to be substantially the same as in the roofed sections between pinholes; and in Calculation 6, the results of which are shown in Table 3, knots up to $1\frac{1}{2}$″ in diameter were assumed to be so located, viz. immediately below the top segment.

To check this assumption with respect to worst position, the following analysis was made of the minimum sections:

Distance from top of section:

To top of bolt hole	1.78″
To bottom of bolt hole	2.16″

Distance from bottom of top segment:

To top of bolt hole	1.45″
To bottom of bolt hole	1.83″

It is apparent that any knot ranging in diameter from 1.78″ to 2.16″, when located at the top of the section, would enter the bolt hole. The section modulus of any section containing a knot within that size range would be the section modulus of the remaining portion of the section, or $\frac{bd^2}{6}$, where b is the width of the section and d the depth below the bolt hole. Thus

$$S \text{ (minimum arm)} = \frac{3.1875\ (1.9375)^2}{6} = 1.9943$$

It is also evident that any knot from 1.45″ to 1.83″ in diameter, when located immediately below the top segment, would likewise enter the bolt hole; and that the section modulus, on this basis of knot location, would be the same for any section containing a knot within the size range mentioned. Continuing the analysis the following tests were made:

2″ Knot:

The distance between the top segment and the bottom of the bolt hole of a minimum section is 1.83″. Therefore, a 2″ knot located immediately

below the top segment would extend beyond the hole; and its effect would be the same as in Calculation 3 (Table 1), where the section modulus of a section containing a 2″ knot similarly located was found to be 2.50. On the other hand, since a 2″ knot is within the limits 1.78″ and 2.16″, the section modulus of a section containing such a knot located at its top would be 1.99.

1.78″ Knot:

A knot of this size immediately below the top segment would enter the bolt hole since it is within the 1.45″ and 1.83″ limits, and the section modulus value associated with it would be the same as shown in the Calculation 6 results (Table 3) for a section containing a $1\frac{1}{2}$″ knot, or $S = 2.62$. But, as evident from previous discussion, the section modulus associated with this knot, if located at the top of the section, would be 1.99.

1.5″ Knot:

It can be shown that the section modulus of a section containing a knot of this size located at the top of the section would be 2.55; and that the section modulus associated with a similarly located *1″ knot* would be 4.55.

The foregoing analysis for minimum sections may be summarized as follows:

Knot Size	Section Modulus	
	Knot at Top	Knot below Top Segment
(Inches)	(Inches^3)	(Inches^3)
2.0	1.99	2.50
1.78	1.99	2.62
1.5	2.55	2.62
1.0	4.55	3.58

A study of this summary shows that knots $1\frac{1}{2}$″ and over are more damaging when located at the section top; and that knots under $1\frac{1}{2}$″ are more damaging when located immediately below the top segment. The section modulus values associated with $2\frac{1}{2}$″ and 3″ knots would be the same as shown in the Calculation 1 results (Table 1).

By a similar analysis for arms of nominal size it can be shown:

(1) That the more damaging position for knots $1\frac{1}{2}$″ and under is immediately below the top segment;

(2) That the more damaging position for any knot within the diameter range from 1.875″ to 2.25″ and all the larger knots is at the top of the section;

(3) That the section modulus associated with 1.875″ to 2.25″ knots would be $\frac{3.25(1.9375)^2}{6} = 2.0334$; and

(4) That the section modulus values associated with $2\frac{1}{2}$″ and 3″ knots would be the same as shown in the Calculation 1 results (Table 1).

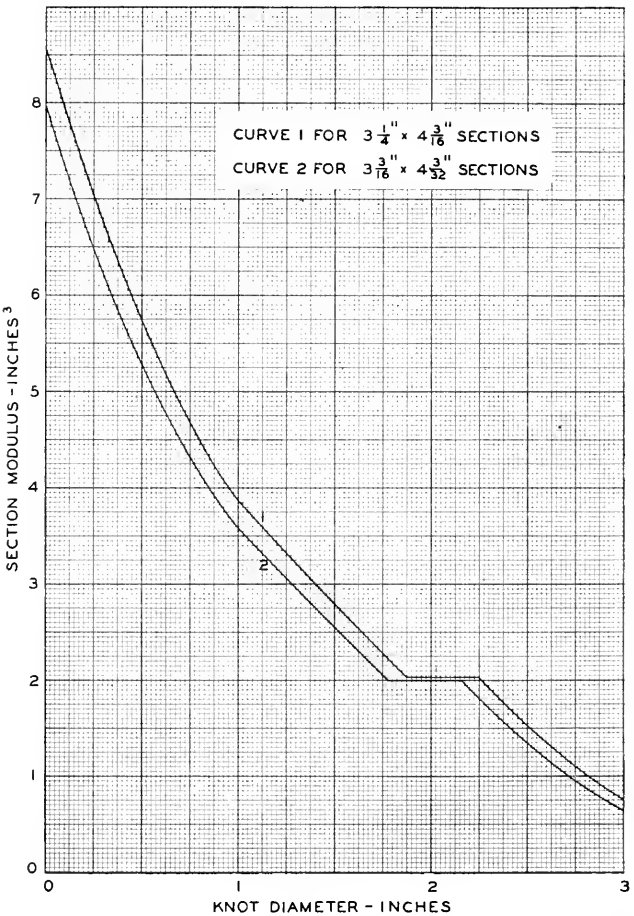


Fig. 6—Brace bolt hole sections. Section modulus of crossarm sections containing knots of the sizes shown on the base line and located in damaging positions.

The results of Calculation 6 (Table 3), and of the foregoing analyses, together with the Calculation 1 results for $2\frac{1}{2}$″ and 3″ knots, were plotted in Fig. 6 for both minimum and nominal sections.

Rectangular Pole Bolt Hole Section

The most damaging position for knots in the pole bolt hole section was assumed to be at the top of the section. They were so figured in Calcula-

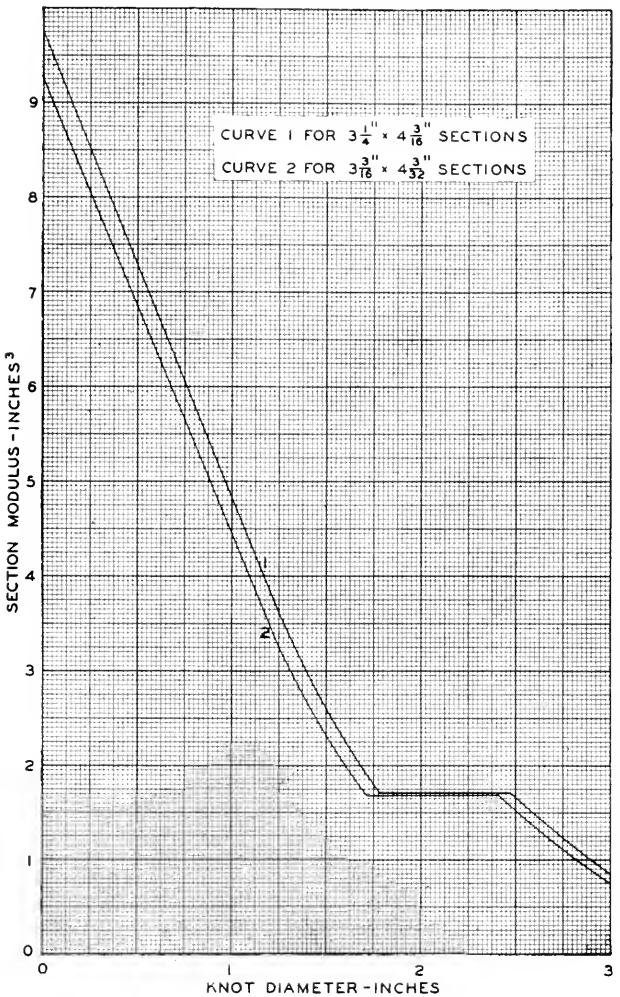


Fig. 7—Pole bolt hole section. Section modulus of crossarm section containing knots of the sizes shown on the base line and located in damaging positions.

tion 7, the results of which are shown in Table 3 and plotted in Fig. 7 for both minimum and nominal arms.

Rectangular Sections without Bolt Holes

Here too the most damaging position for knots was assumed to be at the top of the section. In Calculation 8 the section moduli of sections containing knots from $\frac{1}{4}''$ to $3''$ in diameter were determined for both minimum and nominal sections. The results are shown in Table 4. As section modulus values for sections containing knots of other sizes than those shown may be found so simply by the formula for rectangular sections, $S = \frac{bd^2}{6}$, no curves of the results of this calculation were drawn.

Moment Diagrams

From the results of this study as shown in Table 4 and in Figs. 4, 5, 6 and 7, section modulus values for clear arms and for arms containing knots of various sizes may be read and multiplied by appropriate fiber stresses to determine the resisting moments throughout the length of such arms. For example, the section moduli of clear arms of nominal dimensions, and of arms of minimum dimensions with the maximum knots permitted under the current Bell System crossarm specification (AT-7075) are as follows:

Section of Arm	Arms of nominal size and free from knots	Arms of minimum size with maximum knots	
	Section Modulus	Section Modulus	Diameter of Max. Knots
Pole bolt hole	9.74	5.63	$\frac{3}{4}''$
Brace bolt holes	8.55	4.33	$\frac{3}{4}''$
Pole pinholes	5.11	3.28	$\frac{3}{4}''$
Other pinholes in middle section[3]	5.11	2.38	$\frac{3}{4}''$
End pinholes	5.11	1.33	$1\frac{1}{2}''$
Other pinholes in end sections[3]	5.11	1.41	$1\frac{1}{2}''$
Unroofed part of middle section	9.78	3.84	$1\frac{1}{2}''$
Roofed part of middle section	8.60	2.95	$1\frac{1}{2}''$
Solid part of brace bolt hole zones[4]	8.60	4.56	$\frac{3}{4}''$
Between pinholes in end sections	8.60	2.17	$2''$
Extreme ends	8.60	2.03	$2''$

These section modulus values were used in preparing the moment diagrams shown in Fig. 1. The clear arm of nominal dimensions was also assumed to be straight grained. The fiber stress factor used for it was 5000 psi, which is the ultimate fiber stress value that has been employed in the Bell System for many years for sawn southern pine and Douglas fir. The

[3] For the purposes of specifying knot limitations, crossarms under Specification AT-7075 are divided into a middle section (between brace bolt holes) and end sections (beyond brace bolt holes).

[4] Where a brace bolt hole zone is less than four (4) inches from a pinhole zone, these zones and the portion of the arm between them are considered as a single zone.

fiber stress factor used in computing the resisting moments for the arm of minimum size with maximum slant grain and maximum knots was 3250 psi, which is simply 5000 psi discounted 35% to allow for slant grain of 1″ in 8″, which is the maximum permitted by Specification AT-7075. A discount is, of course, unnecessary for the presence of knots, since allowance for their effect on strength was made in the section modulus values used.

Since the 5000 psi value is an ultimate fiber stress and not a working stress, and since the arms were assumed to be made of clear, straight grained material, Graph 1 (Fig. 1) represents an idealized condition. The resisting moments shown are probably the maximum that may be expected from any commercial lots of southern pine or Douglas fir crossarms,[5] notwithstanding the fact that the dimensions of some of the arms may exceed the nominal specified. With respect to Graph 2 (Fig. 1), the objection may be raised that 35% is not a sufficient discount for a 1″ to 8″ slant of grain and that the 3250 psi value makes no allowance for the effect of long continued loading. On the other hand, the graph assumes the simultaneous occurrence of the maximum knot in a most damaging position in every section of an arm of minimum dimensions and having the maximum slant of grain allowed. Since the probability of such simultaneous occurrence of these defects and conditions is extremely small, it is felt that the resisting moments of Graph 2 represent the minimum strength of any arm of the two species concerned that may be furnished under Specification AT-7075.

Under the assumptions made, Graphs 1 and 2 (Fig. 1) may be regarded as the upper and lower limits of the bending strength of specification crossarms. On the same diagram may be plotted the graph or graphs of the moments resulting from any given load at each pin position, or any single load concentrated at any point on the arm. As an illustration, Graph 3, showing the bending moments from a load of 50 pounds per pin, is shown in the diagram (Fig. 1). A load of 50 pounds per pin is calculated to be the load of size 165 wire coated with ice having a radial thickness of $\frac{1}{4}$ inch in span lengths of 235 feet, or of wire of the same size in 100 foot spans where the radial thickness of the ice coating is $\frac{1}{2}$ inch. Since Graph 3 is wholly below Graph 2, even an arm of lowest specification quality would support the assumed loads with some margin of strength to spare. This margin or factor of safety, would, of course, be increased greatly if the quality of the arm under consideration approached the quality assumed in Graph 1. As previously indicated, the probability is extremely remote that any single arm will ever be furnished of a quality as low as assumed in Graph 2. It

[5] Graphs 1 and 2 (Fig. 1) are for southern pine and Douglas fir crossarms. It is estimated that the resisting moments of comparable graphs for the other woods included in Specification AT-7075 should be about 20% lower.

follows, therefore, that the average strength of any lots of southern pine or Douglas fir arms produced under Specification AT-7075 may be expected to lie well above the Graph 2 limit.

Graph 2 and a bending moment graph for vertical loads at each pin position are of considerable value to the material design engineer, since the degree of parallelism between the two will show whether a consistent strength relationship exists throughout the length of the crossarm. As a matter of interest in this connection, moment diagrams were used as a guide in setting the knot limitations shown in Specification AT-7075.

Resisting and bending moment graphs may also be used to determine the location of the critical section of a crossarm by noting the point of coincidence between a *maximum* bending moment graph and the resisting moment graph for a clear arm. It can be shown by such graphs that this point in all types of Bell System crossarms, designed for vertical loads, is located at the pole pinholes. If the comparison were made between a maximum bending moment graph and the resisting moment graph of an arm containing all of the maximum defects permitted, the location of the point of coincidence between the graphs might or might not fall at the pole pinholes, depending on the magnitude and location of the defects allowed. It should be noted, however, that for such arms the critical section locations so determined apply only when the arms are actually of the assumed minimum quality; and, since the probability of such being the case is so extremely remote, it is concluded that the maximum stress or critical section locations in arms of that quality are of academic interest only, and that for all practical purposes the critical section of any $3\frac{1}{4}''$ x $4\frac{1}{4}''$ x $10'$ crossarm is located at the pole pinhole.

This conclusion does not mean that every arm broken in service or under test will break at the pole pinhole; for, obviously, if some other section is relatively weaker because of some hidden defect which reduces its section modulus or its fiber strength, it will break at such section regardless of any mathematical determination of the break location. But the conclusion does mean that, generally speaking, when a crossarm breaks the break will occur at, or be closely related to, the pole pinholes. To check the accuracy of this conclusion, an examination was made of all available crossarm strength test data in which the break locations were recorded. The examination revealed that, out of 258 arms tested, the breaks in 219, or 85 per cent, were either at, or directly related to, the pole pinholes. Six per cent of the breaks were located between the two pole pinholes, and 9 per cent at points outside the pole pinholes.

As an illustration of another use to which such a moment diagram may be put, the following specific example is cited. Before the present standard Bell System specification for crossarms was drafted, it was decided to

include a new type ("W6") with 16 pin positions. It was felt that, if the additional pin holes in the type W6 did not unduly weaken the arm, it could not only replace the old type "JW" arm with 8 pin positions but also be used in installations where greater flexibility in wire spacings might be required.

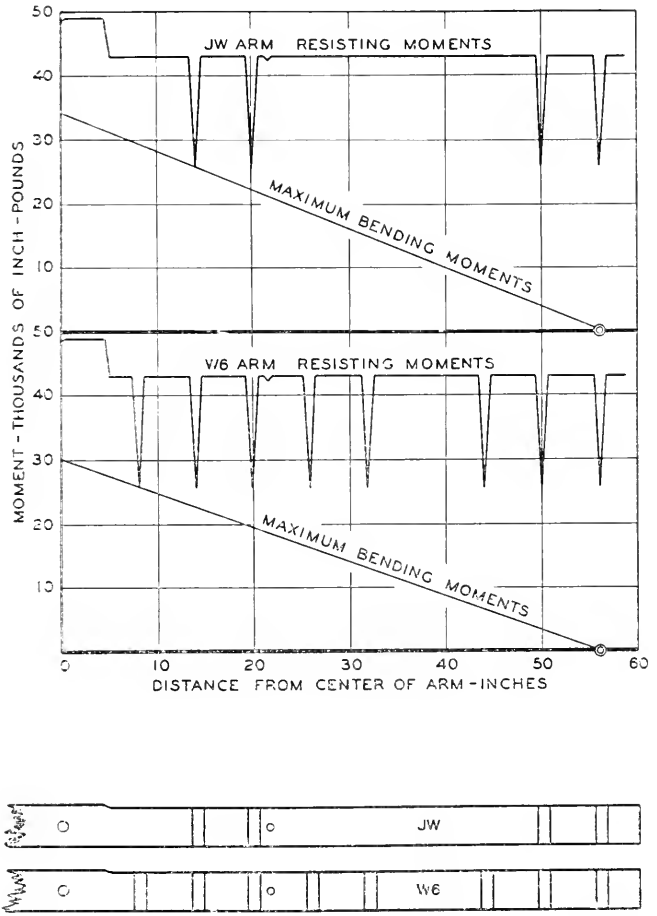


Fig. 8—Resisting moments and maximum bending moments for clear JW and W6 crossarms.

In order to obtain an estimate of the strength relationship between the two types, strength tests were made of 10 matched arms of each type. The test arms were made of air-seasoned, clear Douglas fir. The dimensions of the crossarm blanks were $3\frac{1}{4}''$ x $4\frac{1}{4}''$ x 20′. In selecting the 10 blanks from which the test arms were made, only straight grained pieces free from

evidence of manufacturing and other defects were chosen. Each blank chosen was cut into two 10′ lengths, one of which was made into a JW arm and the other into a W6 arm, making 10 matched arms of each type. The tests were made on an Amsler testing machine. The average breaking load at the end pinholes was 1159 pounds for the JW arms and 1002 pounds for W6 arms.

At the same time an estimate was made of the theoretical *strength relationship* between the two types by means of the moment diagrams shown in Fig. 8. In this figure are shown the graphs of the resisting moments (fiber stress factor—5000 psi) of clear JW and clear W6 arms, together with the graphs of the bending moments due to the maximum loads these arms would withstand when the loads are concentrated at the end pinholes. These maximum loads were determined by dividing the moments at the points of coincidence between the graphs (critical pole pinhole sections) by the distances to the end pinholes. The maximum loads, so determined, are 608 pounds for the JW arm and 532 pounds for the W6 arm. The fact that these loads are low as compared with the actual breaking loads shows, of course, that the average ultimate fiber stress developed by these selected arms was considerably greater than 5000 psi, which is not surprising in view of their exceptionally high quality. However, so far as the information sought is concerned—namely, to determine not the actual strength but the strength relationship between the two types—the result would be the same regardless of the fiber stress factor used in the moment diagram.

The ratio of the strength of the W6 arm to that of the JW arm as shown both by the actual strength tests and by the moment diagrams was as follows:

	Strength Ratio W6 to JW
	(Per cent)
Actual strength tests — $\frac{1002}{1159} \times 100 =$	86.5
Moment diagrams — $\frac{532}{608} \times 100 =$	87.5

These ratios show a remarkably close agreement between theory and actuality and justify the belief that the crossarm moment diagram may be employed to obtain reasonably accurate estimates of relative bending strength.

Summary

The results of this study may be summarized as follows:

1. The moment diagram is a useful guide in setting specification limitations on defects.

2. It is shown that the critical section of a crossarm is located at the pole pinholes. The practical value of this observation is that it emphasizes the need for keeping the pole pinhole sections and the portion of the arm between them reasonably free from strength reducing defects.

3. Only by breaking tests can the *actual* bending strength of crossarms be determined. The *relative* bending strengths, however, of two or more arms of different types or quality may be estimated with sufficient accuracy by means of the moment diagram, regardless of the fiber stress used in its construction.

4. If the fiber stress factor employed is dependable, the moment diagram may be used to estimate the minimum factor of safety that would obtain for an arm of any type or any assumed quality. In this connection, it is believed that the strength of Bell System crossarms is well above the minimum required to support the loads ordinarily carried.

5. The section modulus curves of Figs. 4, 5, 6 and 7 will simplify the construction of moment diagrams for arms of the same sizes shown in the figures but differing with respect to type and quality.

The uses listed lead to the general conclusion that the crossarm moment diagram is a convenient and reasonably reliable engineering tool.

APPENDIX

Computation I. Moment of Inertia of Top Segment of Minimum ($3\frac{3}{16}$" x $4\frac{3}{32}$") *Section between Pinholes:*

The moment of inertia (IT) of a segment (T) with respect to an axis through its center of gravity and parallel to its base may be found by the formula

$$IT = I_{BB} - Ax^2$$

where I_{BB} is the moment of inertia of the segment about the axis BB, A the area of the segment and x the distance between the two axes. The values I_{BB}, A and x are given by:

$$I_{BB} = \tfrac{1}{4}Ar^2\left[1 + \frac{2\sin^3 a\cos a}{a - \sin a\cos a}\right] \tag{1}$$

$$A = \tfrac{1}{2}r^2(2a - \sin 2a) \tag{2}$$

$$x = \tfrac{2}{3}\frac{r^3\sin^3 a}{A} \tag{3}$$

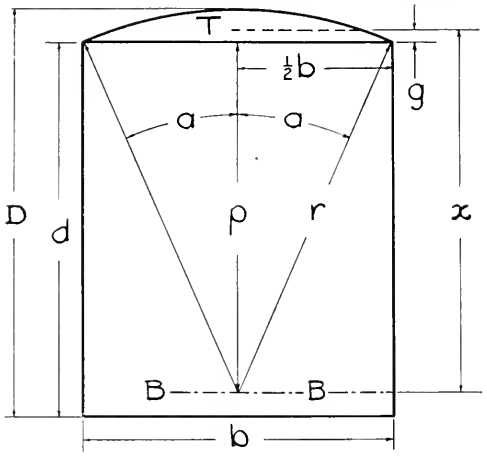


Fig. 9—Crossarm section between pinholes.

The significance of r and a in these formulae, and of the other symbols used in the computations that follow will be clear from a glance at Fig. 9.

$D = 4.09375''$

$b = 3.1875''$

$\frac{1}{2}\,b = 1.59375''$

$r = 4''$

$r^2 = 16.000000$

$(1/2\,b)^2 = 2.540039$

$p^2 = 13.459961$

$p = 3.668782''$

$$\text{Sin } a = \frac{1/2\,b}{r} = 0.39843750$$

$a = 23^\circ\ 28'\ 49.93''$

$a = 0.40981266$ radians

$2\,a = 46^\circ\ 57'\ 39.86''$

$\text{Sin}^3 a = 0.063252925$

$\text{Sin } 2\,a = 0.73089017$

$\text{Cos } a = 0.91719548$

$\text{Sin } a \text{ Cos } a = 0.36544507$

$d = p + (D - r) = 3.7625''$

$A = 0.7099$ sq. ins. [Area of T by Formula (2)]

$x = 3.8018''$ [By Formula (3)]

$g = x - p = 0.1330''$

$I_{BB} = 10.2654$ [By Formula (1)]

$Ax^2 = 10.2601$

$IT = 0.0053$

(Note: While the results of this and the following computations are shown to four decimal places, the actual work was done by machine and carried to eight decimal places as mentioned in the text.)

Since the width of the section in this computation and the radius of its roof is the same as for the minimum $3\frac{3}{16}''$ x $4''$ section at the end of the arm, the top segments of the two are identical, and the only value that will differ

will be the depth (d) of the rectangular portion of the section, which for the smaller will be $p + (D - r)$, or

$$3.6688 + (4 - 4) = 3.6688''$$

Computation II. Moment of Inertia of Top Segment of Nominal ($3\frac{1}{4}''$ *x* $4\frac{3}{16}''$) *Section between Pinholes:*

As this computation was made in exactly the same manner as Computation I, only the results are here shown:

$$d = 3.8593''$$
$$g = 0.1317''$$
$$A = 0.7168 \text{ sq. ins}$$
$$IT = 0.0053$$

Computation III. Moment of Inertia of Top Segment of Minimum ($3\frac{3}{16}''$ *x* $4\frac{3}{32}''$) *Pinhole Section:*

It will be noted in Fig. 10 that the top segment is divided into four parts: the small segment (T_1) at the top of the pinhole, the rectangular portion

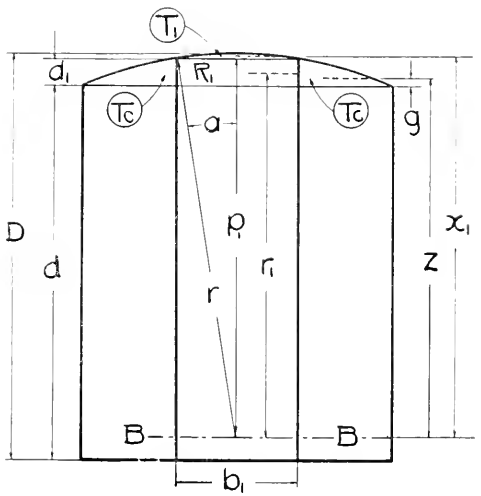


Fig. 10—Crossarm pinhole section.

R_1, with a width of b_1 and a depth of d_1, and two portions designated Tc. The purpose of this computation is to determine the moment of inertia of one of the Tc portions with respect to its gravity axis parallel to its base. The moment of inertia of the two Tc portions about the axis BB may be

found by deducting the moments of inertia of T_1 and R_1 about this axis from the moment of inertia of the entire top segment about the same axis.

$D = 4.09375''$

$b_1 = 1.33''$

$\frac{1}{2} b_1 = 0.665''$

$r = 4.00''$

$r^2 = 16.000000$

$(1/2\ b_1)^2 = 0.442225$

$p_1^2 = 15.557775$

$p_1 = 3.9443345''$

$$\text{Sin } a = \frac{1/2\ b_1}{r} = 0.16625$$

$a = 9° 34' 11.49''$

$a = 0.16702554$ radians

$2a = 19° 8' 22.98''$

$\text{Sin}^3 a = 0.0045949941$

$\text{Sin } 2a = 0.32787285$

$\text{Cos } a = 0.98608364$

$\text{Sin } a \text{ Cos } a = 0.16393640$

d (Computation I) $= 3.7625''$

$d_1 = p_1 + (D - r) - d = 0.2756''$

$r_1 = p_1 - 1/2 d_1 = 3.8065''$

Area $R_1 = b_1 d_1 = 0.3665$ sq. ins.

$A_1 = 0.0494$ sq. ins. [Area of T_1 by Formula (2)]

$x_1 = 3.9666''$ [By Formula (3)]

By Computation I, $IT_{BB} = 10.2654$

IT_{1BB} [Formula (1)] $= 0.7777$

$$IR_{1BB} = \frac{b_1 d_1^3}{12} + R_1 r_1^2 = 5.3126$$

6.0903

$2ITc_{BB} = 4.1751$

The moment of inertia of the 2 Tc areas with respect to the axis through their own centers of gravity is given by

$$2ITc = 2ITc_{BB} - 2Tc\ z^2$$

where

2 Tc is the area of the two Tc portions of the top segment and is given by

$$2Tc = A - (A_1 + R_1)$$

in which A is the area of the entire top segment as shown in Computation I; and

where, by the principle of moments,

$$z = \frac{Tx - T_1 x_1 - R_1 r_1}{2Tc}$$

in which Tx, $T_1 x_1$ and $R_1 r_1$ are the moments of the areas of T, T_1 and R_1, respectively, about the axis BB. ($Tx = Ax$ of Computation I.)

Thus

$2Tc = 0.2940$ sq. ins.

$z = 3.7680''$

As previously shown, $2ITc_{BB} = 4.1751$

$2Tc\,z^2 = 4.1738$

$2ITc = 0.0013$

$ITc = 0.0007$

$D - r = 0.09375''$

$z = 3.7680''$

$3.8618''$

$d = 3.7625''$

g for $Tc = 0.0993''$

The results of this computation apply also to the minimum $3\frac{3}{16}''$ x $4''$ pinhole section at the ends of the arm. The depth (d) of the rectangular portion of the end pinhole sections will be the same as at the extreme ends of the arm, viz. 3.6688".

Computation IV. Moment of Inertia of Top Segment of Nominal ($3\frac{1}{4}''$ x $4\frac{3}{16}''$) *Pinhole Section:*

Since this computation was made in the same manner as Computation III, only the results are here shown:

$d = 3.8593''$

$g = 0.1019''$

$Tc = 0.1630$ sq. ins.

$ITc = 0.0008$

Mathematical Analysis of Random Noise

BY S. O. RICE

(Concluded from July 1944 issue)

PART III

STATISTICAL PROPERTIES OF RANDOM NOISE CURRENTS

3.0 Introduction

In this section we use the representations of the noise currents given in section 2.8 to derive some statistical properties of $I(t)$. The first six sections are concerned with the probability distribution of $I(t)$ and of its zeros and maxima. Sections 3.7 and 3.8 are concerned with the statistical properties of the envelope of $I(t)$. Fluctuations of integrals involving $I^2(t)$ are discussed in section 3.9. The probability distribution of a sine wave plus a noise current is given in 3.10 and in 3.11 an alternative method of deriving the results of Part III is mentioned. Prof. Uhlenbeck has pointed out that much of the material in this Part is closely connected with the theory of Markoff processes. Also S. Chandrasekhar has written a review of a class of physical problems which is related, in a general way, to the present subject.[22]

3.1 The Distribution of the Noise Current[23]

In section 1.4 it has been shown that the distribution of a shot effect current approaches a normal law as the expected number of events per second, ν, increases without limit.

In line with the spirit of this Part, Part III, we shall use the representation

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

to show that $I(t)$ is distributed according to a normal law. This is obtained at once when the procedure outlined in section 2.8 is followed. Since a_n and b_n are distributed normally, so are $a_n \cos \omega_n t$ and $b_n \sin \omega_n t$ when t is regarded as fixed. $I(t)$ is thus the sum of $2N$ independent normal variates and consequently is itself distributed normally.

[22] Stochastic Problems in Physics and Astronomy, *Rev. of Mod. Phys.*, Vol. 15, pp. 1–89 (1943).

[23] An interesting discussion of this subject by V. D. Landon and K. A. Norton is given in the *I.R.E. Proc.*, 30 (Sept. 1942) pp. 425–429.

The average value of $I(t)$ as given by (2.8–1) is zero since $\bar{a}_n = \bar{b}_n = 0$:

$$\bar{I}(t) = 0 \tag{3.1–1}$$

The mean square value of $I(t)$ is

$$\begin{aligned} \overline{I^2}(t) &= \sum_{n=1}^{N} (\overline{a_n^2} \cos^2 \omega_n t + \overline{b_n^2} \sin^2 \omega_n t) \\ &= \sum_{n=1}^{N} w(f_n)\Delta f \\ &\to \int_0^\infty w(f)\, df = \psi(0) \equiv \psi_0 \end{aligned} \tag{3.1–2}$$

In writing down (3.1–2) we have made use of the fact that all the a's and b's are independent and consequently the average of any cross product is zero. We have also made use of

$$\overline{a_n^2} = \overline{b_n^2} = w(f_n)\Delta f, \qquad f_n = n\Delta f, \qquad \omega_n = 2\pi f_n$$

which were given in 2.8. $\psi(\tau)$ is the correlation function of $I(t)$ and is related to $w(f)$ by

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

as is explained in section 2.1. In this part we shall write the argument of $\psi(\tau)$ as a subscript in order to save space.

Since we know that $I(t)$ is normal and since we also know that its average is zero and its mean square value is ψ_0, we may write down its probability density function at once. Thus, the probability of $I(t)$ being in the range $I, I + dI$ is

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

This is the probability of finding the current between I and $I + dI$ at a time selected at random. Another way of saying the same thing is to state that (3.1–3) is the fraction of time the current spends in the range $I, I + dI$.

In many cases it is more convenient to use the representation (2.8–6)

$$I(t) = \sum_{n=1}^{N} c_n \cos(\omega_n t - \varphi_n), \qquad c_n^2 = 2w(f_n)\Delta f \tag{2.8–6}$$

in which $\varphi_1, \cdots \varphi_n$ are independent random phase angles. In order to deduce the normal distribution from this representation we first observe

that (2.8–6) expresses $I(t)$ as the sum of a large number of independent random variables

$$I(t) = x_1 + x_2 + \cdots + x_N$$
$$x_n = c_n \cos (\omega_n t - \varphi_n)$$

and hence that as $N \to \infty$ $I(t)$ becomes distributed according to a normal law. In order to make the limiting process definite we first choose N and Δf such that $N\Delta f = F$ where

$$\int_F^\infty w(f)\, df < \epsilon \int_0^\infty w(f)\, df$$

where ϵ is some arbitrarily chosen small positive quantity. We now let $N \to \infty$ and $\Delta f \to 0$ in such a way that $N\Delta f$ remains equal to F. Then

$$\begin{aligned} A = \overline{x_1^2} + \overline{x_2^2} + \cdots + \overline{x_N^2} &= \sum_1^N 2w(f_n)\Delta f\, \overline{\cos^2 (\omega_n t - \varphi_n)} \\ &= \sum_1^N w(f_n)\Delta f \to \int_0^F w(f)\, df \end{aligned} \tag{3.1–4}$$

$$\begin{aligned} B = \overline{|x_1|^3} + \cdots + \overline{|x_N|^3} &= \sum_1^N (2w(f_n)\Delta f)^{3/2}\, \overline{|\cos (\omega_n t - \varphi_n)|^3} \\ &< 4(\Delta f)^{1/2} \int_0^F [w(f)]^{3/2}\, df \end{aligned}$$

where the bars denote averages with respect to the φ's, t being held constant. If we assume that the integrals are proper, the ratio $BA^{-3/2} \to 0$ as $N \to \infty$, and consequently the central limit theorem* may be used if $w(f) = 0$ for $f > F$. Since we may make F as large as we please by choosing ϵ small enough, we may cover as large a frequency range as we wish. For this reason we write ∞ in place of F.

Now that the central limit theorem has told us that the distribution of $I(t)$, as given by (2.8–6), approaches a normal law, there remains only the problem of finding the average and the standard deviation:

$$\begin{aligned} \overline{I(t)} &= \sum_1^N c_n\, \overline{\cos (\omega_n t - \varphi_n)} = 0 \\ \overline{I^2(t)} &= \sum_1^N c_n^2\, \overline{\cos^2 (\omega_n t - \varphi_n)} \\ &\to \int_0^\infty w(f)\, df = \psi_0 \end{aligned} \tag{3.1–5}$$

* Section 2.10.

This gives the probability density (3.1-3). Hence the two representations lead to the same result in this case. Evidently, they will continue to lead to identical results as long as the central limit theorem may be used. In the future use of the representation (2.8-6) we shall merely assume that the central limit theorem may be applied to show that a normal distribution is approached. We shall omit the work corresponding to equations (3.1-4).

The characteristic function for the distribution of $I(t)$ is

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

3.2 The Distribution of $I(t)$ and $I(t + \tau)$

We require the two dimensional distribution in which the first variable is the noise current $I(t)$ and the second variable is its value $I(t + \tau)$ at some later time τ. It turns out that this distribution is normal[24], as we might expect from the analogy with section 3.1. The second moments of this distribution are

$$\begin{aligned} \mu_{11} &= \overline{I^2(t)} = \psi_0 = \int_0^\infty w(f)\, df \\ \mu_{22} &= \psi_0 \\ \mu_{12} &= \overline{I(t)I(t+\tau)} \\ &= \psi_\tau \end{aligned} \tag{3.2-1}$$

The expression for μ_{12} is in line with our definition (2.1-4) for the correlation function:

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

In order to get the distribution from the representation (2.8-6) we write

$$I_1 = I(t) = \sum_1^N c_n \cos(\omega_n t - \varphi_n)$$

$$I_2 = I(t+\tau) = \sum_1^N c_n \cos(\omega_n t - \varphi_n + \omega_n \tau)$$

[24] It seems that the first person to obtain this distribution in connection with noise was H. Thiede, *Elec. Nachr. Tek.* 13 (1936), 84-95.

From the central limit theorem for two dimensions it follows that I_1 and I_2 are distributed normally. As in (3.1)

$$\mu_{11} = \overline{I_1^2} = \sum_1^N c_n^2 \cdot \tfrac{1}{2} \to \int_0^\infty w(f)\, df = \psi_0$$

$$\mu_{22} = \overline{I_2^2} = \overline{I_1^2} = \psi_0 \tag{3.1-2}$$

$$\mu_{12} = \overline{I_1 I_2} = \sum_1^N c_n^2 \text{ ave. } \{\cos(\omega_n t - \varphi_n) \cos(\omega_n t - \varphi_n + \omega_n \tau)\}$$

Now the quantity within the parenthesis is

$$\cos^2(\omega_n t - \varphi_n) \cos \omega_n \tau - \cos(\omega_n t - \varphi_n) \sin(\omega_n t - \varphi_n) \sin \omega_n \tau$$

and when we take the average with respect to φ_n the second term drops out, giving

$$\mu_{12} = \sum_1^N c_n^2 \cdot \tfrac{1}{2} \cos \omega_n \tau \to \int_0^\infty w(f) \cos 2\pi f \tau\, df = \psi_\tau \tag{3.2-3}$$

where we have used $\omega_n = 2\pi f_n$ and the relation (2.1–6) between $w(f)$ and $\psi(\tau)$.

The probability density function for I_1 and I_2 may be stated. From the discussion of the normal law in 2.9 it is

$$\frac{[\psi_0^2 - \psi_\tau^2]^{-1/2}}{2\pi} \exp\left[\frac{-\psi_0 I_1^2 - \psi_0 I_2^2 + 2\psi_\tau I_1 I_2}{2(\psi_0^2 - \psi_\tau^2)}\right] \tag{3.2-4}$$

For a band pass filter whose range extends from f_a to f_b we have

$$\begin{aligned} \psi_\tau &= \int_{f_a}^{f_b} w_0 \cos 2\pi f \tau\, df \\ &= w_0 \frac{\sin \omega_b \tau - \sin \omega_a \tau}{2\pi\tau} \\ &= \frac{w_0}{\pi\tau} \sin \pi\tau(f_b - f_a) \cos \pi\tau(f_b + f_a) \\ \psi_0 &= w_0(f_b - f_a) \end{aligned} \tag{3.2-5}$$

where w_0 is the constant value of $w(f)$ in the pass band and

$$\begin{aligned} \omega_b &= 2\pi f_b \\ \omega_a &= 2\pi f_a \end{aligned} \tag{3.2-6}$$

According to our formula (3.2–4), I_1 and I_2 are independent when ψ_τ is zero. For the τ's which make ψ_τ zero, a knowledge of I_1 does not add to our knowledge of I_2. For example, suppose we have a narrow filter. Then

$$\psi_\tau = 0 \text{ when } \tau = [2(f_b + f_a)]^{-1}$$

$$\psi_\tau \text{ is nearly } -\psi_0 \text{ when } \tau = [f_b + f_a]^{-1}$$

For the first value of τ, all we know is that I_2 is distributed about zero with $\overline{I_2^2} = \psi_0$. For the second value of τ I_2 is likely to be near $-I_1$. This is in line with the idea that the noise current through a narrow filter behaves like a sine wave of frequency $\frac{1}{2}(f_b + f_a)$ (and, incidentally, whose amplitude fluctuates with an irregular frequency of the order of $\frac{1}{2}(f_b - f_a)$). The first value of τ corresponds to a quarter-period of such a wave and the second value to a half-period. By drawing a sine wave and looking at points separated by quarter and half periods, the reader will see how the ideas agree.

The characteristic function for the distribution of I_1 and I_2 is

$$\text{ave. } e^{iuI_1+ivI_2} = \exp\left[-\frac{\psi_0}{2}(u^2 + v^2) - \psi_\tau uv\right] \tag{3.2-7}$$

The three dimensional distribution in which

$$I_1 = I(t)$$
$$I_2 = I(t + \tau_1)$$
$$I_3 = I(t + \tau_1 + \tau_2)$$

where τ_1 and τ_2 are given and t is chosen at random is, as we might expect, normal in three dimensions. The moments, from which the distribution may be obtained by the method of Section 2.9, are

$$\mu_{11} = \mu_{22} = \mu_{33} = \psi_0$$
$$\mu_{12} = \psi_{\tau_1}$$
$$\mu_{23} = \psi_{\tau_2}$$
$$\mu_{13} = \psi(\tau_1 + \tau_2) = \psi_{\tau_1+\tau_2}$$

The characteristic function for I_1, I_2, I_3 is

$$\text{ave. } e^{iz_1I_1+iz_2I_2+iz_3I_3}$$
$$= \exp\left[-\frac{\psi_0}{2}(z_1^2 + z_2^2 + z_3^2) - \mu_{12}z_1z_2 - \mu_{23}z_2z_3 - \mu_{13}z_1z_3\right] \tag{3.2-8}$$

3.3 Expected Number of Zeros per Second

We shall use the following result. Let y be given by

$$y = F(a_1, a_2, \cdots a_N; x), \tag{3.3-1}$$

and let the a's be random variables. For a given set of a's, this equation gives a curve of y versus x. Since the a's are random variables we shall call this curve a random curve. Let us select a short interval x_1, $x_1 + dx$,

and then draw a batch of a's. The probability that the curve obtained by putting these a's in (3.3–1) will have a zero in x_1, $x_1 + dx$ is

$$dx \int_{-\infty}^{+\infty} |\eta| p(0, \eta; x_1)\, d\eta \tag{3.3–2}$$

and the expected number of zeros in the interval (x_1, x_2) is

$$\int_{x_1}^{x_2} dx \int_{-\infty}^{+\infty} |\eta| p(0, \eta; x)\, d\eta \tag{3.3–3}$$

In these expressions $p(\xi, \eta; x)$ is the probability density function for the variables

$$\begin{aligned} \xi &= F(a_1, \cdots a_N; x) \\ \eta &= \frac{\partial F}{\partial x} \end{aligned} \tag{3.3–4}$$

Since the a's are random variables so are ξ and η, and their distribution will contain x as a parameter. This is indicated by the notation $p(\xi, \eta; x)$.

These results may be proved in much the same manner as are similar results for the distribution of the maxima of a random curve. This method of proof suffers from the restriction that the a's are required to be bounded.[25] Results equivalent to (3.3–2) and (3.3–3) have been obtained independently by M. Kac.[26] His method of proof has the advantage of not requiring the a's to be bounded.

Here we shall sketch the derivation of a closely related result: The probability that y will pass through zero in x_1, $x_1 + dx$ with positive slope is

$$dx \int_0^{\infty} \eta p(0, \eta; x_1)\, d\eta \tag{3.3–5}$$

We choose dx so small that the portions of all but a negligible fraction of the possible random curves lying in the strip $(x_1, x_1 + dx)$ may be regarded as straight lines. If $y = \xi$ at x_1 and passes through zero for $x_1 < x < x_1 + dx$, its intercept on $y = 0$ is $x_1 - \frac{\xi}{\eta}$ where η is the slope. Thus ξ and η must be of opposite sign and

$$x_1 < x_1 - \frac{\xi}{\eta} < x_1 + dx$$

[25] S. O. Rice, *Amer. Jour. Math.* Vol. 61, pp. 409–416 (1939). However, L. A. MacColl has pointed out to me that a set of sufficient conditions for (3.3–5) to hold is: (a) $p(\xi, \eta; x)$ is continuous with respect to (ξ, η) throughout the $\xi\eta$-plane; and (b) that the integral

$$\int_0^{\infty} p(a\eta, \eta; x_1)\, d\eta$$

converges uniformly with respect to a in some interval $-a_1 \leq a \leq a_2$, where a_1 and a_2 are positive. These conditions are satisfied in all the applications we shall make use of (3.3–5).

[26] M. Kac, *Bull. Amer. Math. Soc.* Vol. 49, pp. 314–320 (1943).

According to the statement of our problem, we are interested only in positive values of η, and we therefore write our inequality as

$$-\eta\, dx < \xi < 0$$

For a given random curve i.e. for a given set of a's ξ and η have the values given by

$$\xi = F(a_1, \cdots a_N ; x_1)$$

$$\eta = \left[\frac{\partial F}{\partial x}\right]_{x=x_1}$$

If these values of ξ and η satisfy our inequality, the curve goes through zero in x_1, $x_1 + dx$. The probability of this happening is[27]

$$\int_0^\infty d\eta \int_{-\eta\, dx}^0 d\xi\, p(\xi, \eta; x_1) = \int_0^\infty [0 - (-\eta\, dx)] p(0, \eta; x_1)\, d\eta$$

where we have made use of the fact that dx is so very small that ξ is effectively zero. The last expression is the same as (3.3–5).

In the same way it may be shown that the probability of y passing through zero in x_1, $x_1 + dx$ with a negative slope is

$$-dx \int_{-\infty}^0 \eta p(0, \eta; x_1)\, d\eta \qquad (3.3\text{–}6)$$

Expression (3.3–2) is obtained by adding (3.3–5) and (3.3–6).

We are now ready to apply our formulas. We let t, $I(t)$ and φ_n play the roles of x, y, and a_n, respectively, and use

$$I(t) = \sum_{n=1}^{N} c_n \cos(\omega_n t - \varphi_n), \qquad c_n^2 = 2w(f)\Delta f \qquad (2.8\text{–}6)$$

[27] MacColl has remarked that the step from the double integral on the left hand side of this equation to the final result (3.3–5) may be made as follows:
It is easily seen that the probability density we are seeking is

$$\left[\frac{d}{d(\Delta x)} \int_0^\infty d\eta \int_{-\eta \Delta x}^0 p(\xi, \eta; x)\, d\xi\right]_{\Delta x = 0}$$

Proceeding formally, without regard to conditions validating the analytical operations (for such conditions see the footnote on page 52), we have

$$\frac{d}{d\Delta x} \int_0^\infty d\eta \int_{-\eta \Delta x}^0 p(\xi, \eta; x)\, d\xi = \int_0^\infty \eta p(-\eta \Delta x, \eta; x)\, d\eta$$

and hence the required probability density is

$$\int_0^\infty \eta p(0, \eta; x)\, d\eta$$

The first step is to find the probability density function of the two random variables

$$\begin{aligned} \xi &= \sum_{n=1}^{N} c_n \cos(\omega_n t_1 - \varphi_n) \\ \eta &= I'(t_1) = -\sum_{n=1}^{N} c_n \omega_n \sin(\omega_n t_1 - \varphi_n) \end{aligned} \tag{3.3-7}$$

where the prime denotes differentiation with respect t. From section 2.10

$$\begin{aligned} \mu_{11} &= \overline{\xi^2} = \psi_0 \\ \mu_{22} &= \overline{\eta^2} = \sum_{n=1}^{N} c_n^2 \omega_n^2 \overline{\sin^2(\omega_n t_1 - \varphi_n)} \\ &= \sum_{n=1}^{N} (2\pi f_n)^2 w(f_n) \Delta f \\ &\to 4\pi^2 \int_0^\infty f^2 w(f)\, df = -\psi_0'' \\ \mu_{12} &= \overline{\xi\eta} = -\sum_{n=1}^{N} c_n^2 \omega_n \overline{\cos(\omega_n t_1 - \varphi_n) \sin(\omega_n t_1 - \varphi_n)} \\ &= 0 \end{aligned}$$

The expression for μ_{22} arises from (2.1–6) by differentiation. In this expression ψ_0'' denotes the second derivative of $\psi(\tau)$ with respect to τ at $\tau = 0$:

$$\psi''(\tau) = -4\pi^2 \int_0^\infty f^2 w(f) \cos 2\pi f\tau \, df \tag{3.3-8}$$

Hence the probability density is

$$p(\xi, \eta; t) = \frac{[-\psi_0 \psi_0'']^{-1/2}}{2\pi} \exp\left[-\frac{\xi^2}{2\psi_0} + \frac{\eta^2}{2\psi_0''}\right] \tag{3.3-9}$$

where ψ_0'' is negative. It will be observed that the expression on the right is independent of t. Hence the probability of having a zero in t_1, $t_1 + dt$,

$$dt \int_{-\infty}^{+\infty} |\eta| \frac{[-\psi_0 \psi_0'']^{-1/2}}{2\pi} e^{\eta^2/2\psi_0''} \, d\eta = \frac{dt}{\pi}\left[-\frac{\psi''(0)}{\psi(0)}\right]^{1/2} \tag{3.3-10}$$

which follows from (3.3–3), is independent of t.

The expected number of zeros per second, which may be obtained from (3.3–3) by integrating (3.3–10) over an interval of one second, is

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

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

$$2\left[\frac{\frac{1}{3}f_b^3 - f_a^3}{f_b - f_a}\right]^{1/2} \tag{3.3-12}$$

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

In a recent paper M. Kac[28] has given a result which, after a slight generalization, leads to

$$e^{-I^2/2\psi_0}\,\frac{1}{2\pi}\left[-\frac{\psi_0''}{\psi_0}\right]^{1/2} dt \tag{3.3-13}$$

for the probability that the noise current will pass through the value I with positive slope during the interval $t, t + dt$. The expected number of such passages per second is

$$e^{-I^2/2\psi_0} \times [\tfrac{1}{2} \text{ the expected number of zeros per second}] \tag{3.3-14}$$

The expression (3.3–13) may also be derived from analogue of (3.3–5) obtained by replacing the zero in $p(0, \eta; x_1)$ by y.

In some cases the integral

$$\psi_0'' = -4\pi^2 \int_0^\infty f^2 w(f)\, df$$

does not converge.

An example occurs when we apply a broad band noise voltage to a resistance and condenser in series. The power spectrum of the voltage across the condenser is of the form

$$w(f) = \frac{1}{f^2 + a^2} \tag{3.3-15}$$

Although ψ_0'' is infinite, ψ_0 is finite and equal to $\pi/2a$. A straightforward substitution in our formula (3.3–11) gives infinity as the expected number of zeros per second.

Some light is thrown on this breakdown of our formula when we consider a noise current consisting of two bands of noise. One band is confined to relatively low frequencies, and its power spectrum will be denoted by $w_1(f)$. The other band is very narrow and is centered at the relatively high frequency f_2. The complete power spectrum of our noise is then

$$w(f) = w_1(f) + A^2\delta(f - f_2)$$

[28] On the Distribution of Values of Trigonometric Sums with Linearly Independent Frequencies, *Amer. Jour. Math.*, Vol. LXV, pp 609–615, (1943).

where the unit impulse function δ is used to represent the very narrow band. The power spectrum of the narrow band is approximately the same as that of the wave $A\sqrt{2}\cos 2\pi f_2 t$.

The integrals occurring in our formula are

$$\int_0^\infty w(f)\,df = \int_0^\infty w_1(f)\,df + A^2$$
$$= W + A^2$$
$$\int_0^\infty w(f)f^2\,df = \int_0^\infty f^2 w_1(f)\,df + A^2 f_2^2$$
$$= U + A^2 f_2^2$$

We suppose that A and f_2 are such that

$$W \gg A^2$$
$$U \ll A^2 f_2^2 .$$

Then our formula (3.3-11) gives us the expected number of zeros

$$2\frac{Af_2}{W^{1/2}}$$

We may give a qualitative explanation of this formula if we regard our noise current as composed of a small component

$$I_2 = 2^{1/2} A \cos 2\pi f_2 t$$

due to the narrow band superposed on a large, slowly varying component due to the lower band. Since the r.m.s. value of the second component is $W^{1/2}$ we may assign it a representative frequency f_1 and write it approximately as

$$I_1 = (2W)^{1/2} \cos 2\pi f_1 t$$

The zeros of the noise current are clustered around the zeros of the second wave. Near such a zero

$$I_1 = \pm(2W)^{1/2} 2\pi f_1 \Delta t$$

where Δt is the distance from the zero. The oscillations of I_1 produce zeros when $|I_1|$ is less than the amplitude of I_2 or when

$$A > W^{1/2} 2\pi f_1 |\Delta t|$$

and the interval over which zeros are produced is given by

$$2\Delta t = \frac{AW^{-1/2}}{\pi f_1}$$

The number of zeros is this multiplied by $2f_2$. Since there are $2f_1$ such intervals per second the number of zeros per second is

$$\frac{4}{\pi} AW^{-1/2} f_2$$

This differs from the result given by our formula by a factor of $2/\pi$. This discrepancy is due to our representing the two bands by the sine waves I_1 and I_2.

From this example we obtain the picture that when the integral for ψ_0 converges corresponding to $A \to 0$, while at the same time the integral for ψ_0'' diverges, corresponding to $f_2 \to \infty$ in such a way that $Af_2 \to \infty$, the noise current behaves something like a continuous function which has no derivative. It seems that for physical systems the integrals will always converge since parasitic effects will have the effect of making $w(f)$ tend to zero rapidly enough. The frequency which represents the region where this occurs is of the order of the frequency of the microscopic wiggles.

So far we have been considering the formulas of this section in the most favorable light possible. There are experiments which indicate the possibility of the formulas breaking down in some cases. Prof. Uhlenbeck has pointed out that if a very broad band fluctuation current be forced[29] to flow through a circuit consisting of a condenser, C, in parallel with a series combination of inductance, L, and resistance, R, equation (3.3–11) says that the expected number of zeros per second of the current, I, flowing through R (and L) is independent of R. It is simply $\frac{1}{\pi}(LC)^{-1/2}$. The differential equation for I is the same as that which governs the Brownian motion of a mirror suspended in a gas[30], the gas pressure playing the role of R. Curves are available for this motion and it is seen that their character depends greatly upon the pressure[31]. Unfortunately, it is difficult to tell from the curves whether the expected number of zeros is independent of the pressure. The differences between the curves for various pressures indicates that there may be some dependence*.

3.4 The Distribution of Zeros

The problem of determining the distribution function for the distance between two successive zeros seems to be quite difficult and apparently

[29] For example, by putting the circuit in series with a diode.

[30] This problem in Brownian motion is discussed by G. E. Uhlenbeck and S. Goudsmit, *Phys., Rev.*, 34 (1929), 145–151.

[31] E. Kappler, *Annalen d. Phys.*, 11 (1931) 233–256.

* Since this was written M. Kac and H. Hurwitz have studied the problem of the expected number of zeros using quite a different method of approach which employs the "shot-effect" representation (Sec. 3.11). Their results confirm the correctness of (3.3–11) when the integrals converge. When the integrals diverge the average number of electrons, per sec. producing the shot effect must be considered.

nobody has as yet given a satisfactory solution. Here we shall give some results which are related to the general problem and which give an idea of the form of the distribution for the region of small spacings between the zeros.

We shall show (in the work starting with equation (3.4–12)) that the probability of the noise current, I, passing through zero in the interval $\tau, \tau + d\tau$ with a negative slope, when it is known that I passes through zero at $\tau = 0$ with a positive slope, is

$$\frac{d\tau}{2\pi}\left[\frac{\psi_0}{-\psi_0''}\right]^{1/2}\left[\frac{M_{23}}{H}\right](\psi_0^2 - \psi_\tau^2)^{-3/2}[1 + H\cot^{-1}(-H)] \quad (3.4\text{–}1)$$

where M_{22} and M_{23} are the cofactors of $\mu_{22} = -\psi_0''$ and $\mu_{23} = -\psi_\tau''$ in the matrix

$$M = \begin{bmatrix} \psi_0 & 0 & \psi_\tau' & \psi_\tau \\ 0 & -\psi_0'' & -\psi_\tau'' & -\psi_\tau' \\ \psi_\tau' & -\psi_\tau'' & -\psi_0'' & 0 \\ \psi_\tau & -\psi_\tau' & 0 & \psi_0 \end{bmatrix}, \quad (3.4\text{–}2)$$

$$H = M_{23}[M_{22}^2 - M_{23}^2]^{-1/2}.$$

We choose $0 \leq \cot^{-1}(-H) \leq \pi$, the value π being taken at $\tau = 0$, and the value $\pi/2$ being approached as $\tau \to \infty$. It should be remembered that we are writing the arguments of the correlation functions as subscripts, e.g., $-\psi_\tau''$ is really

$$-\psi''(\tau) = 4\pi^2 \int_0^\infty f^2 w(f) \cos 2\pi f\tau \, df \quad (3.3\text{–}8)$$

As τ becomes larger and larger the behavior of I at τ is influenced less and less by the fact that it goes through zero with a positive slope at $\tau = 0$. Hence (3.4–1) should approach the probability that, for any interval of length $d\tau$ chosen at random, I will go through zero with a negative slope. Because of symmetry, this is half the probability that it will go through zero. Thus (3.4–1) should approach, from (3.3–10),

$$\frac{d\tau}{2\pi}\left[\frac{-\psi_0''}{\psi_0}\right]^{1/2} \quad (3.4\text{–}3)$$

as $\tau \to \infty$. It actually does this since M approaches a diagonal matrix and both M_{23} and H approach zero with $M_{23}/H \to M_{22} \to -\psi_0^2\psi_0''$. For a low pass filter cutting off at f_b (3.4–3) is

$$d\tau f_b 3^{-1/2} \quad (3.4\text{–}4)$$

The behavior of (3.4–1) as $\tau \to 0$ is quite a bit more difficult to work out.

M_{22} and M_{23} go to zero as τ^4, $M_{22}^2 - M_{23}^2$ as τ^{10}, and consequently H goes to infinity as τ^{-1}. The final result is that (3.4–1) approaches

$$d\tau \frac{\tau}{8}\left[\frac{\psi_0\psi_0^{(4)} - \psi_0''^2}{-\psi_0\psi_0''}\right] \tag{3.4–5}$$

as $\tau \to 0$, assuming $\psi^{(4)}$ exists. Here the superscript (4) indicates the fourth derivative at $\tau = 0$,

$$\psi_0^{(4)} = 16\pi^4 \int_0^\infty f^4 w(f)\, df \tag{3.4–6}$$

For a low pass filter cutting off at f_b (3.4–5) is

$$d\tau \frac{\tau}{30}(2\pi f_b)^2 \tag{3.4–7}$$

When (3.4–1) is applied to a low pass filter, it turns out that instead of τ the variable

$$\varphi = 2\pi f_b \tau, \qquad d\varphi = 2\pi f_b\, d\tau \tag{3.4–8}$$

is more convenient to handle. Thus, if we write (3.4–1) as $p(\varphi)\, d\varphi$, it follows from (3.4–4) and (3.4–7) that

$$\begin{aligned} p(\varphi) &\to \frac{1}{2\pi\sqrt{3}} = .0919 \quad \text{as} \quad \varphi \to \infty \\ p(\varphi) &\to \frac{\varphi}{30} \quad \text{as} \quad \varphi \to \infty \end{aligned} \tag{3.4–9}$$

$p(\varphi)$ has been computed and plotted on Fig. 1 as a function of φ for the range 0 to 9. From the curve and the theory it is evident that beyond 9 $p(\varphi)$ oscillates about 0.0919 with ever decreasing amplitude.

We may take $p(\varphi)\, d\varphi$ to be the probability that I goes through zero in $\varphi, \varphi + d\varphi$, when it is known that I goes through zero at $\varphi = 0$ with a slope opposite to that at φ. $p(\varphi)\, d\varphi$ exceeds the probability that I goes through zero at $\varphi = 0$ and in $\varphi, \varphi + d\varphi$ with no zeros in between. This is because $p(\varphi)\, d\varphi$ includes all curves of the latter class and in addition those which may have an even number of zeros between 0 and φ. From this it follows that the curve giving the probability density of the intervals between zeros must be underneath the curve of $p(\varphi)$.

A partial check on the curve for $p(\varphi)$ may be obtained by comparing it with a probability density function obtained experimentally by M. E. Campbell for the intervals between 754 successive zeros. He passed thermal noise through a band pass filter, the lower cutoff being around 200 cps and the upper cutoff being around 3000 cps. The upper cutoff was rather grad-

ual and it is difficult to assign a representative value. The crosses on figure 1 are obtained from his data when we assume that his filter behaves like a low pass filter with a cutoff at $f_b = 2850$, this choice being made in order to make the maximum of his curve coincide with that of $p(\varphi)$.

It is seen that some of the crosses lie above $p(\varphi)$. This is probably due to the fact that the actual filter differs somewhat from the assumed low pass filter.

On Fig. 1 there is also plotted a function closely related to (3.4–1). It is the low pass filter form of the following: The probability of I passing

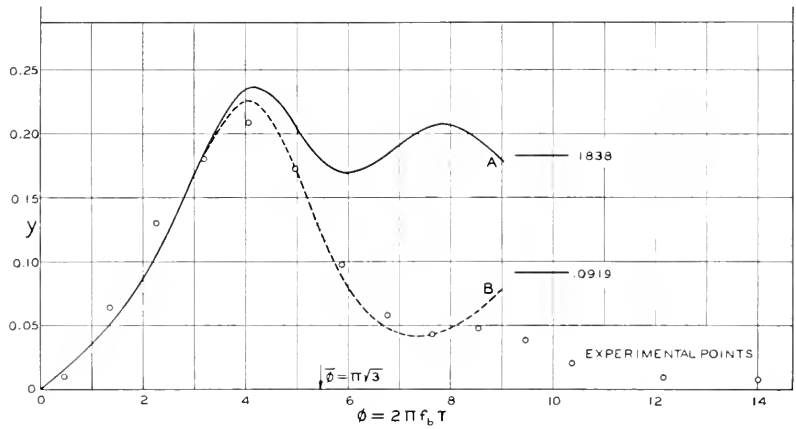


Fig. 1—Distribution of intervals between zeros—low-pass filter
$y_A \Delta\varphi$ is probability of a zero in $\Delta\varphi$ when a zero is at origin.
$y_B \Delta\varphi$ is probability of a zero in $\Delta\varphi$ when a zero is at origin and slopes at zeros are of opposite signs.
$y_B = p(\varphi), f_b$ = filter cutoff, τ = time between zeros.

through zero in τ, $\tau + d\tau$ when it is known that I passes through zero at $\tau = 0$ is

$$\frac{d\tau}{\pi}\left[\frac{\psi_0}{-\psi_0''}\right]^{1/2}\left[\frac{M_{23}}{H}\right](\psi_0^2 - \psi_\tau^2)^{-3/2}[1 + H \tan^{-1} H] \qquad (3.4\text{–}10)$$

where the notation is the same as in (3.4–1) and $-\frac{\pi}{2} \leq \tan^{-1} H \leq \frac{\pi}{2}$. This curve should always lie above $p(\varphi)$ and the small difference between the curves out to $\varphi = 4$ indicates that the true distribution of zeros is given closely by $p(\varphi)$ out to this point.

When (3.4–1) is applied to a relatively narrow band pass filter or some similar device we may make some approximations and obtain an expression somewhat simpler than (3.4–1). As a guide we consider our usual ideal

band pass filter whose range extends from f_a to f_b. The correlation function is given by (3.2–5).

$$\psi_\tau = \frac{w_0}{\pi\tau} \sin \pi\tau(f_b - f_a) \cos \pi\tau(f_b + f_a)$$
$$\psi_0 = w_0(f_b - f_a) \tag{3.2–5}$$

From physical considerations we know that in a narrow filter most of the distances between zeros will be nearly equal to

$$\tau_1 = \frac{1}{f_b + f_a}$$

i.e., nearly equal to the distance between the zeros of a sine wave having the mid-band frequency. We therefore expect (3.4–1) to have a peak very close to τ_1. We also expect peaks at $3\tau_1$, $5\tau_1$ etc. but we shall not consider these. We wish to examine the behavior of (3.4–1) near τ_1.

It turns out that M_{23} is nearly equal to M_{22} so that H is large and (3.4–1) becomes approximately

$$\frac{d\tau}{2}\left[\frac{\psi_0}{-\psi_0''}\right]^{1/2} \frac{M_{23}}{[\psi_0^2 - \psi_\tau^2]^{3/2}}$$

where τ is near τ_1.

In order to see that M_{23} is nearly equal to M_{22} we use the expressions

$$M_{22} = -\psi_0''(\psi_0^2 - \psi_\tau^2) - \psi_0\psi_\tau'^2$$
$$M_{23} = \psi_\tau''(\psi_0^2 - \psi_\tau^2) + \psi_\tau\psi_\tau'^2$$
$$M_{22} + M_{23} = (\psi_0 - \psi_\tau)[(\psi_0 + \psi_\tau)(\psi_\tau'' - \psi_0'') - \psi_\tau'^2]$$
$$= (\psi_0 - \psi_\tau)[B + C]$$
$$M_{22} - M_{23} = (\psi_0 + \psi_\tau)[(\psi_0 - \psi_\tau)(-\psi_\tau'' - \psi_0'') - \psi_\tau'^2]$$
$$= (\psi_0 + \psi_\tau)[-B + C]$$
$$B = \psi_0\psi_\tau'' - \psi_\tau\psi_0''$$
$$C = -\psi_0\psi_0'' + \psi_\tau\psi_\tau'' - \psi_\tau'^2$$

From (3.2–5) it is seen that ψ_τ may be written as

$$\psi_\tau = A \cos \beta\tau, \qquad \beta = \pi(f_b + f_a)$$

where $\beta\tau_1 = \pi$ and A is a function of τ which varies slowly in comparison with $\cos \beta\tau$. We see that near τ_1, ψ_τ is nearly equal to $-\psi_0$. Likewise

ψ_τ' hovers around zero and ψ_τ'' is nearly equal to $-\psi_0''$. Differentiating with respect to τ gives

$$\psi_\tau' = A' \cos \beta\tau - A\beta \sin \beta\tau$$
$$\psi_\tau'' = (A'' - A\beta^2) \cos \beta\tau - 2A'\beta \sin \beta\tau$$
$$\psi_0'' = A_0'' - A_0\beta^2, \qquad \psi_0 = A_0$$

where A_0 and A_0'' are the values of A and its second derivative at τ equal to zero. These lead to

$$B = (A_0A'' - AA_0'') \cos \beta\tau - 2A_0A'\beta \sin \beta\tau$$
$$C = (AA'' - A'^2) \cos^2 \beta\tau - A_0A_0'' + (A_0^2 - A)^2\beta^2$$

We wish to show that $C + B$ and $C - B$ are of the same order of magnitude. If we can do this, it follows that $M_{22} - M_{23}$ is much smaller than $M_{22} + M_{23}$ since $\psi_0 - \psi_{\tau_1}$ is approximately $2\psi_0$ while $\psi_0 + \psi_{\tau 1}$ is quite small. Consequently we will have shown that M_{23} is nearly equal to M_{22}.

So far we have made no approximations. We now express the slowly varying function A as a power series in τ. Since ψ_0' and ψ_0''' must be zero for the type of functions we consider, it follows that

$$A = A_0 + \frac{\tau^2}{2} A_0'' + \cdots$$
$$A' = \tau A_0'' + \cdots$$
$$A'' = A_0'' + \frac{\tau^2}{2} A_0^{(4)} + \cdots$$

where we neglect all powers higher than the second. Multiplication and squaring gives

$$A^2 - A_0^2 = \tau^2 A_0 A_0''$$
$$AA'' - A'^2 = A_0A_0'' + \frac{\tau^2}{2} (A_0A_0^{(4)} - A_0''^2)$$
$$= A_0A_0'' + F$$
$$A_0A'' - AA_0'' = \frac{\tau^2}{2} (A_0A_0^{(4)} - A_0''^2) = F$$

Since, for small τ, A and A'' are nearly equal to A_0 and A_0'', respectively we see that the difference on the left is small relative to A_0A_0'', i.e.,

$$|F| << |A_0A_0''|$$

Our expression for B and C become approximately

$$B = F \cos \beta\tau - 2A_0 A_0'' \beta\tau \sin \beta\tau$$

$$C = F \cos^2 \beta\tau - A_0 A_0'' \sin^2 \beta\tau - A_0 A_0'' \beta^2 \tau^2$$

When τ is near τ_1, $\beta\tau$ is approximately π. Hence both $C + B$ and $C - B$ are approximately $-A_0 A_0'' \pi^2$ and are of the same order of magnitude. Consequently M_{22} and M_{23} are both nearly equal and

$$M_{23} = \psi_0[C + B]$$

$$= -A_0^2 A_0'' \pi^2$$

When this expression for M_{23} is used our approximation to (3.4-1) gives us the result: If the correlation function is of the form

$$\psi_\tau = A \cos \beta\tau$$

where A is a slowly varying function of τ, 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 is positive and

$$a^2 = \frac{A_0 \beta^2}{-A_0'' \tau_1^2}, \qquad \tau_1 = \frac{\pi}{\beta}$$

For our ideal band pass filter with the pass band $f_b - f_a$,

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

and the average value of $|\tau - \tau_1|$ is a^{-1}. Thus

$$\frac{\text{ave.}\,|\tau - \tau_1|}{\tau_1} = \frac{1}{a\tau_1} = \frac{f_b - f_a}{\sqrt{3}\,(f_b + f_a)} = \frac{1}{2\sqrt{3}} \frac{\text{band width}}{\text{mid-frequency}}$$

When the correlation function cannot be put in the form assumed above but still behaves like a sinusoidal wave with slowly varying amplitude we may use our first approximation to (3.4-1). Thus, the probability that the distance between two successive zeros lies between τ and $\tau + d\tau$ is approximately

$$\frac{b\, d\tau}{[\psi_0^2 - \psi_\tau^2]^{3/2}}$$

when τ lies near τ_1 where τ_1 is the smallest value of τ which makes ψ_τ approximately equal to $-\psi_0$. This probability is supposed to approach

zero rapidly as τ departs from τ_1, and b is chosen so that the integral over the effective region around τ_1 is unity.

It seems to be especially difficult to get an expression for the distribution of zeros for large spacing. One method, suggested by Prof. Goudsmit, is to amend the conditions leading to (3.4–1) by adding conditions that I be positive at equally spaced points along the time axis between 0 and τ. This leads to integrals which are hard to evaluate. For one point between 0 and τ the integral is of the form (3.5–7).

Another method of approach is to use the method of "in and exclusion" of zeros between 0 and τ. Consider the class of curves of I having a zero at $\tau = 0$. Then, in theory, our methods will allow us to compute the functions $p_0(\tau)$, $p_1(r, \tau)$, $p_2(r, s, \tau)$, associated with this class where

$p_0(\tau)\, d\tau$ is probability of curve having zero in $d\tau$

$p_1(r, \tau)\, d\tau\, dr$ is probability of curve having zeros in $d\tau$ and dr

$p_2(r, s, \tau)\, d\tau\, dr\, ds$ is probability of curve having zeros in $d\tau$, dr, and ds

In fact $p_0(\tau)\, d\tau$ is expression (3.4–10). The method of in and exclusion then leads to an expression for $P_0(\tau)\, d\tau$, the probability of having a zero at 0 and a zero in τ, $\tau + d\tau$ but none between 0 and τ. It is

$$P_0(\tau) = p_0(\tau) - \frac{1}{1!}\int_0^\tau p_1(r, \tau)\, dr + \frac{1}{2!}\int_0^\tau\int_0^\tau p_2(r, s, \tau)\, dr\, ds - \frac{1}{3!}\int_0^\tau\int_0^\tau\int_0^\tau p_3(r, s, t, \tau)\, dr\, ds\, dt + \cdots \tag{3.4–11}$$

Here again we run into difficult integrals. Incidentally, (3.4–11) may be checked for events occurring independently at random. Thus if $\nu\, d\tau$ is the probability of an event happening in $d\tau$, then, if ν is a constant and the events are independent, we have p_0, p_1, p_2, $\cdots$ given by ν, ν^2, ν^3, $\cdots$. From (3.4–11) we obtain the known result $P_0(\tau) = \nu e^{-\nu\tau}$.

We shall now derive (3.4–1). The work is based upon a generalization of (3.3–5): If y is a random curve described by (3.3–1), the probability that y will pass through zero in x_1, $x_1 + dx_1$ with a positive slope and through zero in x_2, $x_2 + dx_2$ with a negative slope is

$$-dx_1\, dx_2 \int_0^{+\infty} d\eta_1 \int_{-\infty}^{0} d\eta_2\, \eta_1\, \eta_2\, p(0, \eta_1, x_1; 0, \eta_2, x_2) \tag{3.4–12}$$

where $p(\xi_1, \eta_1, x_1; \xi_2, \eta_2, x_2)$ is the probability density function for the four random variables

$$\xi_i = F(a_1, a_2, \cdots, a_N; x_i)$$

$$\eta_i = \left[\frac{\partial F}{\partial x}\right]_{x=x_i}, \qquad i = 1, 2.$$

The x_1 and x_2 play the role of parameters in (3.4–12). This result may be established in much the same way as (3.3–5).

When we identify F with one of our representations, (2.8–1) or (2.8–6), of the noise current $I(t)$ it is seen that p is normal in four dimensions. We may obtain the second moments directly from this representation, as has been done in the equations just below (3.3–7). The same results may be obtained from the definition of $\psi(\tau)$, and for the sake of variety we choose this second method. We set $x_1 = t_1$, $x_2 = t_1 + \tau$. Then

$$\begin{aligned}
\overline{\xi_1^2} &= \overline{\xi_2^2} = \overline{I^2(t)} = \psi_0 \\
\overline{\xi_1 \xi_2} &= \overline{I(t)I(t+\tau)} = \psi_\tau \qquad (3.4\text{–}13) \\
\overline{\eta_1 \eta_2} &= \overline{\left(\frac{\partial I}{\partial t}\right)_t \left(\frac{\partial I}{\partial t}\right)_{t+\tau}} = \operatorname*{Limit}_{T\to\infty} \frac{1}{T} \int_0^T I'(t+\tau)I'(t)\, dt
\end{aligned}$$

where primes denote differentiation with respect to the arguments. Integrating by parts:

$$\int_0^T I'(t+\tau)\, dI(t) = [I'(t+\tau)I(t)]_0^T - \int_0^T I''(t+\tau)I(t)\, dt$$

We assume that I and its derivative remains finite so that the integrated portion vanishes, when divided by T, in the limit. Since

$$I''(t+\tau) = \frac{\partial^2}{\partial \tau^2} I(t+\tau)$$

we have

$$\overline{\eta_1 \eta_2} = -\frac{\partial^2}{\partial \tau^2} \psi(\tau) = -\psi_\tau''$$

Setting $\tau = 0$ gives

$$\overline{\eta_1^2} = \overline{\eta_2^2} = -\psi_0''$$

in agreement with the value of μ_{22} obtained from (3.3–7). In the same way

$$\begin{aligned}
\overline{\xi_1 \eta_2} &= \operatorname*{Limit}_{T\to\infty} \frac{1}{T} \int_0^T I'(t+\tau)I(t)\, dt = \frac{\partial}{\partial \tau} \psi(\tau) \\
&= \psi_\tau' \\
\overline{\xi_2 \eta_1} &= \operatorname*{Limit}_{T\to\infty} \frac{1}{T} \int_0^T I'(t)I(t+\tau)\, dt \\
&= \quad " \quad (-)\frac{1}{T} \int_0^T I'(t+\tau)I(t)\, dt \\
&= -\psi_\tau'
\end{aligned}$$

where we have integrated by parts in getting $\overline{\xi_2 \eta_1}$. Setting $\tau = 0$ and using $\psi_0' = 0$ gives

$$\overline{\xi_1 \eta_1} = \overline{\xi_2 \eta_2} = 0$$

In order to obtain the matrix M of the second moments μ_{rs} in a form fairly symmetrical about its center we choose the 1, 2, 3, 4 order of our variables to be ξ_1, η_1, η_2, ξ_2. From equations (3.4–13) etc. it is seen that this choice leads to the expression (3.4–2) for M.

When we put ξ_1 and ξ_2 equal to zero, we obtain for the probability density function in (3.4–12) the expression

$$\frac{|M|^{-1/2}}{4\pi^2} \exp\left[-\frac{1}{2|M|}(M_{22}\eta_1^2 + 2M_{23}\eta_1\eta_2 + M_{33}\eta_2^2)\right]$$

Because of the symmetry of M, M_{22} is equal to M_{33}. When, in the integral (3.4–12) we make the change of variable

$$x = \left[\frac{M_{22}}{2|M|}\right]^{1/2} \eta_1, \qquad y = -\left[\frac{M_{22}}{2|M|}\right]^{1/2} \eta_2$$

we obtain

$$\frac{dx_1\, dx_2}{\pi^2} \frac{|M|^{3/2}}{M_{22}^2} \int_0^\infty x\, dx \int_0^\infty dy\, y e^{-x^2-y^2+2(M_{23}/M_{22})xy}$$

The double integral may be evaluated by (3.5–4). Let

$$\varphi = \cos^{-1}\left(-\frac{M_{23}}{M_{22}}\right) = \cot^{-1}(-H), \qquad H = M_{23}[M_{22}^2 - M_{23}^2]^{-1/2}$$

where H is the same as that given in (3.4–2). Our expression now becomes

$$\frac{dx_1\, dx_2}{4\pi^2} \frac{|M|^{3/2}}{M_{22}^2 - M_{23}^2}[1 + H \cot^{-1}(-H)]$$

From a property of determinants

$$M_{22}M_{33} - M_{23}^2 = |M|(\psi_0^2 - \psi_\tau^2)$$

Using this to eliminate $|M|$ and dividing by

$$\frac{dx_1}{2\pi}\left[\frac{-\psi_0''}{\psi_0}\right]^{1/2}$$

which, from (3.3–10), is the probability of going through zero in x_1, $x_1 + dx_1$ with positive slope, gives the probability of going through zero in dx_2 with

negative slope when it is known that I goes through zero at x_1 with positive slope:

$$\frac{dx_2}{2\pi}\left[\frac{\psi_0}{-\psi_0''}\right]^{1/2}[M_{22}^2 - M_{23}^2]^{1/2}(\psi_0^2 - \psi_\tau^2)^{-3/2}[1 + H\cot^{-1}(-H)]$$

This is the same as (3.4–1).

The expression (3.4–10) is the same as the probability of I going through zero in $d\tau$ when it is known that I goes through zero at the origin with positive slope. This second probability may be obtained from (3.4–1) by adding the probability that I goes through $d\tau$ with positive slope when it is known to go through zero with positive slope. Thus we must add the expression containing the integral in which the integration in both η_1 and η_2 run from 0 to ∞. In terms of x and y this integral is

$$\int_0^\infty x\,dx \int_0^\infty dy\; y e^{-x^2-y^2-2(M_{23}/M_{22})xy}$$

This is equivalent to a change in the sign of M_{23} and hence of H. After this addition we must consider

$$\begin{aligned} &1 + H\cot^{-1}(-H) + 1 - H\cot^{-1} H \\ &= 2 + H[\cot^{-1}(-H) - \cot^{-1} H] \\ &= 2 + H[\pi - 2\cot^{-1} H] \\ &= 2[1 + H\tan^{-1} H] \end{aligned}$$

and this leads to (3.4–10).

3.5 Multiple Integrals

We wish to evaluate integrals of the form

$$J = \int_0^\infty dx_1 \int_0^\infty dx_2\, e^{-x_1^2 - 2\alpha x_1 x_2 - x_2^2} \tag{3.5-1}$$

Our method of procedure is to first reduce the exponent to the sum of squares by a suitable linear change of variable and then change to polar coordinates. This method appears to work also for triple integrals of the same sort, but when it is applied to a four-fold integral, the last integration apparently cannot be put in closed form.

The reduction of the exponent to the sum of squares is based upon the transformation: If*

$$\begin{aligned} x_1 &= h_1 y_1 + h_2 D_{21} y_2 + h_3 D_{31} y_3 + \cdots + h_n D_{n,1} y_n \\ x_2 &= 0 \quad + h_2 D_{22} y_2 + \quad \cdots \qquad\quad + h_n D_{n,2} y_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n &= 0 \quad + 0 \qquad + \quad \cdots \quad + 0 \quad + h_n D_{n,n} y_n \end{aligned} \tag{3.5-2}$$

* T. Fort, Am. Math. Monthly, 43 (1936), pp. 477–481. See also Scott and Mathews, Theory of Determinants, Cambridge (1904), Prob. 63, p. 276.

where $D_0 = 1$, $D_1 = a_{11}$, $D_{r,r} = D_{r-1}$, and D_{rs} is the cofactor of a_{sr} (or of a_{rs} because they are equal) in D_r :

$$D_r = \begin{vmatrix} a_{11} & a_{12} \cdots a_{1r} \\ a_{12} & a_{22} \\ a_{1r} & \cdots \quad a_{rr} \end{vmatrix}, \qquad h_r = [D_{r-1}D_r]^{-1/2},$$

then, if none of the D_r's is zero,

$$\sum_1^n a_{rs} x_r x_s = y_1^2 + y_2^2 + \cdots + y_n^2$$

From (3.5–2); the Jacobian $\partial(x_1, \cdots x_n)/\partial(y_1, \cdots y_n)$ is equal to $D_n^{-1/2}$.

Applying our transformation to the exponent:

$$x_1 = y_1 - aD_2^{-1/2}y_2$$
$$x_2 = 0 + D_2^{-1/2}y_2$$
$$D_2 = 1 - a^2$$

Since x_2 runs from 0 to ∞ so must y_2. The expression for x_1 shows that y_1 runs from $a\,D_2^{-1/2}y_2$ to ∞. The integral is therefore

$$J = D_2^{-1/2} \int_0^\infty dy_2 \int_{a D_2^{-1/2} y_2}^\infty e^{-y_1^2 - y_2^2}\, dy_1$$

We now change to polar coordinates:

$$y_1 = \rho \cos\theta \qquad dy_1\, dy_2 = \rho\, d\rho\, d\theta$$
$$y_2 = \rho \sin\theta$$
$$y_2 \geq 0 \text{ gives } 0 \leq \theta \leq \pi$$
$$y_1 \geq aD_2^{-1/2}y_2 \text{ gives } \cot\theta \geq aD_2^{-1/2}$$

and obtain

$$J = D_2^{-1/2} \int_0^{\cot^{-1} aD_2^{-1/2}} d\theta \int_0^\infty \rho e^{-\rho^2}\, d\rho$$
$$= \tfrac{1}{2} D_2^{-1/2} \cot^{-1}(aD_2^{-1/2})$$

where the arc-cotangent lies between 0 and π. This may be written in the simpler form

$$J = \tfrac{1}{2}(1 - a^2)^{-1/2} \cos^{-1} a = \tfrac{1}{2}\varphi \csc\varphi$$

where

$$a = \cos\varphi,$$

it being understood that $0 \leq \varphi \leq \pi$.

Other integrals may be obtained by differentiation. Thus from

$$\int_0^\infty dx \int_0^\infty dy\, e^{-x^2-y^2-2xy\cos\varphi} = \tfrac{1}{2}\varphi \csc\varphi \tag{3.5-3}$$

we obtain

$$\int_0^\infty dx \int_0^\infty dy\, xy\, e^{-x^2-y^2-2xy\cos\varphi} = \tfrac{1}{4}\csc^2\varphi(1 - \varphi\cot\varphi) \tag{3.5-4}$$

By using the same transformation we may obtain

$$\int_0^\infty dx \int_0^\infty dy\, ye^{-x^2-y^2-2axy} = \frac{\sqrt{\pi}}{4}\frac{1}{1+a} \tag{3.5-5}$$

Of course, we may expand part of the exponential in a power series and integrate termwise but this leads to a series which has to be summed in each particular case:

$$\int_0^\infty dx \int_0^\infty dy\, x^n y^m e^{-x^2-y^2-2axy}$$

$$= \frac{1}{4}\sum_{r=0}^{\infty}\frac{(-2a)^r}{r!}\Gamma\left(\frac{n+r+1}{2}\right)\Gamma\left(\frac{m+r+1}{2}\right)$$

If we take $-1 < R(m) < -\frac{1}{2}$, $-1 < R(m) < -\frac{1}{2}$, the series may be summed when $a = 1$. The result stated just below equation (3.8-9) is obtained by continuing m and n analytically.

The same methods will work when the limits are $\pm\infty$. We obtain, when m and n are integers,

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\, x^n y^m e^{-x^2-y^2-2xy\cos\varphi}$$

$$= \begin{cases} 0, & n+m \text{ odd} \\ (-)^n\sqrt{\pi}\,\dfrac{\Gamma\left(\dfrac{m+n+1}{2}\right)}{(\sin\varphi)^{n+m+1}} & \\ \qquad F\left(-n, -m; \dfrac{1-n-m}{2}; \dfrac{1-\cos\varphi}{2}\right), & n+m \text{ even} \end{cases} \tag{3.5-6}$$

The hypergeometric function may also be written as

$$F\left(-\frac{n}{2}, -\frac{m}{2}; \frac{1-n-m}{2}; \sin^2\varphi\right)$$

By transformations of this we are led to the following expression for the integral

$$\begin{cases} 0, & n + m \text{ odd}, \\ \dfrac{\Gamma\left(\dfrac{m+1}{2}\right)\Gamma\left(\dfrac{n+1}{2}\right)}{(\sin\varphi)^{n+m+1}} F\left(-\frac{n}{2}, -\frac{m}{2}, \frac{1}{2}; \cos^2\varphi\right), & m, n \text{ both even}, \\ -2\dfrac{\Gamma\left(1+\dfrac{n}{2}\right)\Gamma\left(1+\dfrac{m}{2}\right)}{(\sin\varphi)^{n+m+1}} \cos\varphi F\left(\frac{1-m}{2}, \frac{1-n}{2}; \frac{3}{2}; \cos^2\varphi\right), & m, n \text{ odd} \end{cases}$$

As was mentioned earlier, the method used to evaluate the double integrals may also be applied to similar triple integrals. Here we state two results obtained in this way.

$$\int_0^\infty dx \int_0^\infty dy \int_0^\infty dz \exp\left[-x^2 - y^2 - z^2 - 2cxy - 2bzx - 2ayz\right]$$
$$= \frac{1}{4}\left[\frac{\pi}{D_3}\right]^{1/2}[\alpha + \beta + \gamma - \pi]$$
$$\int_0^\infty dx \int_0^\infty dy \int_0^\infty dz\, yz \exp\left[-x^2 - y^2 - z^2 - 2cxy - 2bzx - 2ayz\right]$$
$$= \frac{\sqrt{\pi}}{8D_3}\left[\frac{1 + a - b - c}{1 + a} - \frac{a - bc}{D_3^{1/2}}(\alpha + \beta + \gamma - \pi)\right] \qquad (3.5\text{–}7)$$

where β and γ are obtained by cyclic permutation of a, b, c from

$$\alpha = \cos^{-1}\frac{a - cb}{(1 - c^2)^{1/2}(1 - b^2)^{1/2}} = \sin^{-1}\left[\frac{D_3}{(1 - c^2)(1 - b^2)}\right]^{1/2}$$
$$= \cot^{-1}\frac{a - bc}{D_3^{1/2}}$$

where α, β, γ all lie in the range $0, \pi$ and where

$$D_3 = \begin{vmatrix} 1 & c & b \\ c & 1 & a \\ b & a & 1 \end{vmatrix} = 1 + 2abc - a^2 - b^2 - c^2$$

For reference we state the integrals which arise from the definition of the normal distribution given in section (2.9)

$$\int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n \exp\left[-\sum_1^n a_{rs}x_r x_s\right] = \left[\frac{\pi^n}{|a|}\right]^{1/2}$$
$$\int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n\, x_t x_u \exp\left[-\sum_1^n a_{rs}x_r x_s\right] = \left[\frac{\pi^n}{|a|^3}\right]^{1/2}\frac{A_{tu}}{2} \qquad (3.5\text{–}8)$$

where the quadratic form is positive definite and $|a|$ is its determinant. A_{tu} is the cofactor of a_{tu}. Incidentally, these may be regarded as special cases of

$$\int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n f\left(\sum_1^n a_{rs} x_r x_s\right) F\left(\sum_1^n b_r x_r\right)$$

$$= \frac{2}{\Gamma\left(\frac{n-1}{2}\right)} \left[\frac{\pi^{n-1}}{|a|}\right]^{1/2} \int_{-\infty}^{+\infty} dx \int_0^{\infty} dy\, y^{n-2} f(x^2 + y^2) \tag{3.5-9}$$

$$F\left\{x\left[\frac{\sum_1^n A_{rs} b_r b_s}{|a|}\right]^{-1/2}\right\},$$

which is a generalization of a result given by Schlömilch.*

3.6 Distribution of Maxima of Noise Current

Here we shall use a result similar to those used in sections 3.3 and 3.4. Let y be a random curve given by (3.3–1),

$$y = F(a_1 \cdots a_N ; x). \tag{3.3-1}$$

If suitable conditions are satisfied, the probability that y has a maximum in the rectangle $(x_1, x_1 + dx_1, y_1, y_1 + dy_1)$, dx_1 and dy_1 being of the same order of magnitude, is[32]

$$-dx_1\, dy_1 \int_{-\infty}^{0} p(y_1, 0, \zeta)\zeta\, d\zeta \tag{3.6-1}$$

and the expected number of maxima of y in $a \leq x \leq b$ is obtained by integrating this expression over the range $-\infty \leq y_1 \leq \infty$, $a \leq x_1 \leq b$. $p(\xi, \eta, \zeta)$ is the probability density function for the random variables

$$\begin{aligned} \xi &= F(a_1, \cdots, a_N ; x_1) \\ \eta &= \left(\frac{\partial F}{\partial x}\right)_{x=x_1} \\ \zeta &= \left(\frac{\partial^2 F}{\partial x^2}\right)_{x=x_1} \end{aligned} \tag{3.6-2}$$

* Höheren Analysis, Braunschweig (1879), Vol. 2, p. 494, equ. (29).

[32] *Am. Jour. Math.*, Vol. 61 (1939) 409–416. A similar problem has been studied by E. L. Dodd, The Length of the Cycles Which Result From the Graduation of Chance Elements, Ann. Math. Stat., Vol. 10 (1939) 254–264. He gives a number of references to the literature dealing with the fluctuations of time series.

In our application of this result we replace x and y by t and I as before. Then

$$\xi = I = \sum_{1}^{N} c_n \cos(\omega_n t - \varphi_n)$$

$$\eta = I'$$

$$\zeta = I''$$

where the primes denote differentiation with respect to t. According to the central limit theorem the distribution of ξ, η, ζ approaches a normal law. The second moments defining this law may be obtained either from the above definitions of ξ, η, ζ, or may be obtained from the correlation function as was done in the work following equation (3.4–13).

$$\overline{\xi^2} = \psi_0, \qquad \overline{\eta^2} = -\psi_0'', \qquad \overline{\xi\eta} = 0$$

$$\overline{\eta\zeta} = \overline{I'(t)I''(t)} = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I'(t)I''(t)\,dt$$

$$= \operatorname*{Limit}_{T\to\infty} \frac{1}{2T}[I'^2(T) - I'^2(0)] = 0$$

$$\overline{\xi\zeta} = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I(t)I''(t)\,dt$$

$$= \operatorname*{Limit}_{\tau\to 0} \frac{\partial^2\psi(\tau)}{\partial\tau^2} = \psi_0''$$

$$\overline{\zeta^2} = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I''(t)I''(t)\,dt$$

$$= \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I^{(4)}(t)I(t)\,dt$$

$$= \psi_0^{(4)}$$

where the superscript (4) represents the fourth derivative. The matrix M of the moments is thus

$$M = \begin{bmatrix} \psi_0 & 0 & \psi_0'' \\ 0 & -\psi_0'' & 0 \\ \psi_0'' & 0 & \psi_0^{(4)} \end{bmatrix}$$

The determinant $|M|$ and the cofactors of interest are

$$|M| = -\psi_0''(\psi_0\psi_0^{(4)} - \psi_0''^2) \tag{3.6-3}$$

$$M_{11} = -\psi_0''\psi_0^{(4)}, \qquad M_{13} = \psi_0''^2, \qquad M_{33} = -\psi_0''\psi_0$$

The probability density function in (3.6–1) is

$$p(I, 0, \zeta) = (2\pi)^{-3/2} \mid M \mid^{-1/2} \exp \left[-\frac{1}{2 \mid M \mid} (M_{11} I^2 + M_{33} \zeta^2 + 2M_{13} I\zeta) \right] \tag{3.6–4}$$

and when this is put in (3.6–1) and the integration with respect to ζ performed we get

$$dI\, dt \frac{(2\pi)^{-3/2}}{M_{33}} \left[\mid M \mid^{1/2} e^{-M_{11} I^2/2\mid M\mid} + M_{13} I \left(\frac{\pi}{2M_{33}} \right)^{1/2} e^{-I^2/2\psi_0} \left(1 + \operatorname{erf} \frac{M_{13} I}{[2 \mid M \mid M_{33}]^{1/2}} \right) \right] \tag{3.6–5}$$

for the probability of a maximum occurring in the rectangle $dI\, dt$. As is mentioned just below expression (3.6–1), the expected number of maxima in the interval t_1, t_2 may be obtained by integrating (3.6–1) from t_1 to t_2 after replacing x by t, and I from $-\infty$ to $+\infty$ after replacing y by I. When we use (3.6–4) it is easier to integrate with respect to I first. The expected number is then

$$-\int_{t_1}^{t_2} dt \frac{M_{11}^{-1/2}}{2\pi} \int_{-\infty}^{0} \zeta \exp \left[-\frac{\zeta^2}{2 \mid M \mid} \left(M_{33} - \frac{M_{13}^2}{M_{11}} \right) \right] d\zeta$$

$$= (t_2 - t_1) \frac{\psi_0^{(4)}}{2\pi} M_{11}^{-1/2} = \frac{t_2 - t_1}{2\pi} \left[\frac{\psi_0^{(4)}}{-\psi_0''} \right]^{1/2}$$

Hence 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

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

where f_b and f_a are the cut-off frequencies. Putting $f_a = 0$ so as to get a low pass filter,

$$f_b \left[\frac{3}{5} \right]^{1/2} = .775 f_b \tag{3.6–8}$$

From (3.6–8) and (3.6–5) we may obtain the probability density function for the maxima in the case of a low pass filter. Thus the probability that a maximum selected at random from the universe of maxima will lie in I, $I + dI$ is

$$\frac{dI}{3\sqrt{2\pi\psi_0}}\left[2e^{-9y^2/8} + \left(\frac{5\pi}{2}\right)^{1/2} ye^{-y^2/2}\left(1 + \operatorname{erf} y\left(\frac{5}{8}\right)^{1/2}\right)\right] \qquad (3.6\text{–}9)$$

where

$$y = \frac{I}{\sqrt{\psi_0}}$$

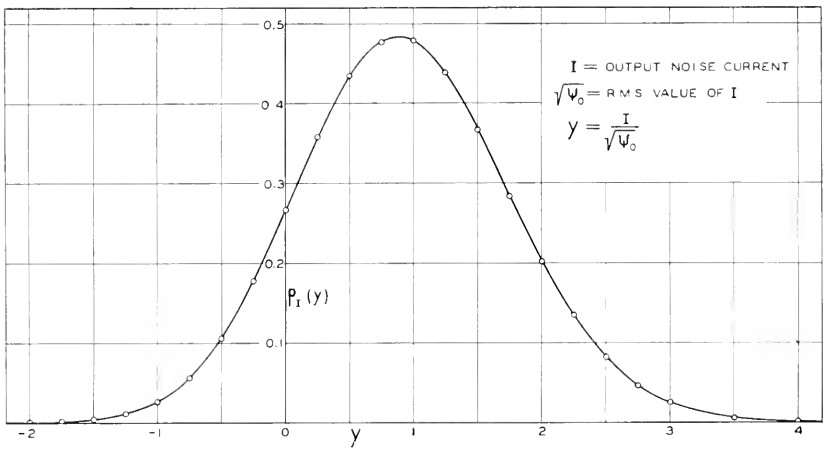


Fig. 2—Distribution of maxima of noise current. Noise through ideal low-pass filter. $\frac{p_I(y)}{\sqrt{\psi_0}}\,dI$ = probability that a maximum of I selected at random lies between I and $I + dI$.

When y is large and positive (3.6–9) is given asymptotically by

$$\frac{dI}{\sqrt{\psi_0}}\,\frac{\sqrt{5}}{3}\,ye^{-y^2/2}$$

If we write (3.6–9) as $p_I(y)\,dy$, the probability density $p_I(y)$ of y may be plotted as a function of y. This plot is shown in Fig. 2. The distribution function $P(I_{\max} < y\sqrt{\psi_0})$ defined by

$$P(I_{\max} < y\sqrt{\psi_0}) = \int_{-\infty}^{y} p_I(y)\cdot dy$$

and which gives the probability that a maximum selected at random is less than a specified $y\sqrt{\psi_0} = I$, is one of the four curves plotted in Fig. 4.

If I is large and positive we may obtain an approximation from (3.6–5). We observe that

$$\frac{M_{11}}{|M|} = \frac{\psi_0^{(4)}}{\psi_0\psi_0^{(4)} - \psi_0''^2} > \frac{1}{\psi_0}$$

so that when I is large and positive

$$e^{-M_{11}I^2/2|M|} \ll e^{-I^2/2\psi_0}$$

Also, in these circumstances the 1 + erf is nearly equal to two. Thus retaining only the important terms and using the definitions of the M's gives the approximation to (3.6–5):

$$\frac{dI}{2\pi\psi_0}\frac{dt}{}\left[\frac{-\psi_0''}{\psi_0}\right]^{1/2} Ie^{-I^2/2\psi_0} \tag{3.6–10}$$

From this it follows that the expected number of maxima per second lying above the line $I = I_1$ is approximately[33] when I_1 is large,

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

It is interesting to note that the approximation (3.6–11) for the expected number of maxima above I_1 is the same as the exact expression (3.3–14) for the expected number of times I will pass through I_1 with positive slope.

3.7 Results on the Envelope of the Noise Current

The noise current flowing in the output of a relatively narrow band pass filter has the character of a sine wave of, roughly, the midband frequency whose amplitude fluctuates irregularly, the rapidity of fluctuation being of the order of the band width. Here we study the fluctuations of the envelope of such a wave.

First we define the envelope. Let f_m be a representative midband frequency. Then if

$$\omega_m = 2\pi f_m \tag{3.7–1}$$

the noise current may be represented, see (2.8–6), by

$$\begin{aligned} I &= \sum_{n=1}^{N} c_n \cos(\omega_n t - \omega_m t - \varphi_n + \omega_m t) \\ &= I_c \cos \omega_m t - I_s \sin \omega_m t \end{aligned} \tag{3.7–2}$$

where the components I_c and I_s are

$$\begin{aligned} I_c &= \sum_{n=1}^{N} c_n \cos(\omega_n t - \omega_m t - \varphi_n) \\ I_s &= \sum_{n=1}^{N} c_n \sin(\omega_n t - \omega_m t - \varphi_n) \end{aligned} \tag{3.7–3}$$

[33] This expression agrees with an estimate made by V. D. Landon, *Proc. I. R. E.*, 29 (1941), 50–55. He discusses the number of crests exceeding four times the r.m.s. value of I. This corresponds to $I_1^2 = 16\psi_0$.

The envelope, R, is a function of t defined by

$$R = [I_c^2 + I_s^2]^{1/2} \tag{3.7-4}$$

It follows from the central limit theorem and the definitions (3.7–3) of I_c and I_s that these are two normally distributed random variables. They are independent since $\overline{I_c I_s} = 0$. They both have the same standard deviation, namely the square root of

$$\overline{I_c^2} = \overline{I_s^2} = \overline{I^2} = \int_0^\infty w(f)\, df = \psi_0 \tag{3.7-5}$$

Consequently, the probability that the point (I_c, I_s) lies within the elementary rectangle $dI_c dI_s$ is

$$\frac{dI_c\, dI_s}{2\pi\psi_0} \exp\left[-\frac{I_c^2 + I_s^2}{2\psi_0}\right] \tag{3.7-6}$$

In much of the following work it is convenient to introduce another random variable θ where

$$\begin{aligned} I_c &= R\cos\theta \\ I_s &= R\sin\theta \end{aligned} \tag{3.7-7}$$

Since I_c and I_s are random variables so are R and θ. The differentials are related by

$$dI_c dI_s = R d\theta dR \tag{3.7-8}$$

and the distribution function for R and θ is obtainable from (3.7–6) when the change of variables is made:

$$\frac{d\theta}{2\pi}\,\frac{R\,dR}{\psi_0}\, e^{-R^2/2\psi_0} \tag{3.7-9}$$

Since this may be expressed as a product of terms involving R only and θ only, R and θ are independent random variables, θ being uniformly distributed over the range 0 to 2π and R having the probability density[34]

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

Expression (3.7–10) gives the probability density for the value of the envelope. Like the normal law for the instantaneous value of I, it depends only upon the average total power

$$\psi_0 = \int_0^\infty w(f)\, df$$

[34] See V. D. Landon and K. A. Norton, *I.R.E. Proc.*, 30 (1942), 425–429.

We now study the correlation between R at time t and its value at some later time $t + \tau$. Let the subscripts 1 and 2 refer to the times t and $t + \tau$, respectively. Then from (3.7–3) and the central limit theorem it follows that the four random variables I_{c1}, I_{s1}, I_{c2}, I_{s2} have a four dimensional normal distribution. This distribution is determined by the second moments

$$\overline{I_{c1}^2} = \overline{I_{s1}^2} = \overline{I_{c2}^2} = \overline{I_{s2}^2} = \psi_0 = \mu_{11}$$

$$\overline{I_{c1} I_{s1}} = \overline{I_{c2} I_{s2}} = 0$$

$$\overline{I_{c1} I_{c2}} = \overline{I_{s1} I_{s2}} = \frac{1}{2} \sum_{n=1}^{N} c_n^2 \cos (\omega_n \tau - \omega_m \tau)$$

$$\rightarrow \int_0^\infty w(f) \cos 2\pi(f - f_m)\tau \, df = \mu_{13} \tag{3.7–11}$$

$$\overline{I_{c1} I_{s2}} = -\overline{I_{c2} I_{s1}} = \frac{1}{2} \sum_{n=1}^{N} c_n^2 \sin (\omega_n \tau - \omega_m \tau)$$

$$\rightarrow \int_0^\infty w(f) \sin 2\pi(f - f_m)\tau \, df = \mu_{14}$$

The moment matrix for the variables in the order I_{c1}, I_{s1}, I_{c2}, I_{s2} is

$$M = \begin{bmatrix} \psi_0 & 0 & \mu_{13} & \mu_{14} \\ 0 & \psi_0 & -\mu_{14} & \mu_{13} \\ \mu_{13} & -\mu_{14} & \psi_0 & 0 \\ \mu_{14} & \mu_{13} & 0 & \psi_0 \end{bmatrix}$$

and from this it follows that the cofactors of the determinant $| M |$ are

$$\begin{aligned} M_{11} = M_{22} = M_{33} = M_{44} &= \psi_0(\psi_0^2 - \mu_{13}^2 - \mu_{14}^2) \\ &= \psi_0 A, \qquad A = \psi_0^2 - \mu_{13}^2 - \mu_{14}^2 \\ M_{12} = \quad M_{34} &= 0 \\ M_{13} = \quad M_{24} &= -\mu_{13} A \\ M_{14} = -M_{23} &= -\mu_{14} A \\ | M | &= A^2 \end{aligned} \tag{3.7–12}$$

The probability density of the four random variables is therefore

$$\frac{1}{4\pi^2 A} \exp - \frac{1}{2A} [\psi_0(I_1^2 + I_2^2 + I_3^2 + I_4^2)$$

$$- 2\mu_{13}(I_1 I_3 + I_2 I_4) - 2\mu_{14}(I_1 I_4 - I_2 I_3)]$$

where we have written I_1, I_2, I_3, I_4 for I_{c1}, I_{s1}, I_{c2}, I_{s2}. We now make the transformation

$$I_1 = R_1 \cos \theta_1 \qquad I_3 = R_2 \cos \theta_2$$

$$I_2 = R_1 \sin \theta_1 \qquad I_4 = R_2 \sin \theta_2$$

and average the resulting probability density over θ_1 and θ_2 in order to get the probability that R_1 and R_2 lie in dR_1 and dR_2. It is

$$\frac{R_1 R_2 \, dR_1 \, dR_2}{4\pi^2 A} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \exp$$

$$- \frac{1}{2A} [\psi_0 R_1^2 + \psi_0 R_2^2 - 2\mu_{13} R_1 R_2 \cos (\theta_2 - \theta_1) - 2\mu_{14} R_1 R_2 \sin (\theta_2 - \theta_1)]$$

Since the integrand is a periodic function of θ_2 we may integrate from $\theta_2 = \theta_1$ to $\theta_2 = \theta_1 + 2\pi$ instead of from 0 to 2π. This integration gives the Bessel function, I_0, of the first kind with imaginary argument. The resulting probability density for R_1 and R_2 is

$$\frac{R_1 R_2}{A} I_0 \left(\frac{R_1 R_2}{A} [\mu_{13}^2 + \mu_{14}^2]^{1/2} \right) \exp - \frac{\psi_0}{2A} (R_1^2 + R_2^2) \qquad (3.7\text{–}13)$$

where, from (3.7–12),

$$A = \psi_0^2 - \mu_{13}^2 - \mu_{14}^2$$

μ_{13} and μ_{14} are given by (3.7–11). Of course, R_1 and R_2 are always positive.

For an ideal band pass filter with cut-offs at f_a and f_b we set

$$f_m = \frac{f_b + f_a}{2}, \qquad w(f) = w_0 \quad \text{for} \quad f_a < f < f_b$$

and obtain

$$\psi_0 = w_0(f_b - f_a)$$

$$\mu_{13} = \int_{f_a}^{f_b} w_0 \cos 2\pi(f - f_m)\tau \, df = \frac{w_0 \sin \pi(f_b - f_a)\tau}{\pi\tau}$$

$$\mu_{14} = \int_{f_a}^{f_b} w_0 \sin 2\pi(f - f_m)\tau \, df = 0$$

The I_0 term in (3.7–13), which furnishes the correlation between R_1 and R_2, becomes

$$I_0 \left(\frac{R_1 R_2}{\psi_0} \frac{\frac{\sin x}{x}}{1 - \frac{\sin^2 x}{x^2}} \right)$$

where x is $\pi(f_b - f_a)\tau$. When x is a multiple of π, R_1 and R_2 are independent random variables. When x is zero R_1 and R_2 are equal. Hence we may say, roughly, that the period of fluctuation of R is the time it takes x to increase from 0 to π or $(f_b - f_a)^{-1}$. This is related to the result given in the next section, namely that the expected number of maxima of the envelope is .641 $(f_b - f_a)$ per second.

3.8 Maxima of R

Here we wish to study the distribution of the maxima of R.* Our work is based upon the expression, cf. (3.6–1),

$$-dR\,dt \int_{-\infty}^{0} p(R, 0, R'')R''\,dR'' \tag{3.8-1}$$

for the probability that a maximum of R falls within the elementary rectangle $dR\,dt$. $p(R, R', R'')$ is the probability density for the three dimensional distribution of R, R', R'' where the primes denote differentiation with respect to t.

We shall determine $p(R, R', R'')$ from the probability density of I_c, I_s', I_c'', I_s, I_c', I_s'', which we shall denote by x_1, x_2, $\cdots$ x_6. The interchange of I_s' and I_c' is suggested by the later work. It is convenient to introduce the notation

$$\begin{aligned} b_n &= (2\pi)^n \int_0^\infty w(f)(f - f_m)^n\,df \\ b_0 &= \psi_0 \end{aligned} \tag{3.8-2}$$

where f_m is the mid-band frequency, i.e., the frequency chosen in the definition of the envelope R. b_n is seen to be analogous to the derivatives of $\psi(\tau)$ at $\tau = 0$.

From the definitions (3.7–3) of I_c and I_s we obtain the second moments

$$\begin{aligned} \overline{x_1^2} &= \overline{I_c^2} = \psi_0 = b_0 \\ \overline{x_4^2} &= \overline{I_s^2} = b_0 \\ \overline{x_2^2} &= \overline{I_s'^2} = \sum_1^N w(f_n)\Delta f 4\pi^2(f_n - f_m)^2 = b_2 \\ \overline{x_5^2} &= \overline{I_c'^2} = b_2 \\ \overline{x_3^2} &= \overline{I_c''^2} = b_4 \\ \overline{x_6^2} &= \overline{I_s''^2} = b_4 \end{aligned}$$

* Incidentally, most of the analysis of this section was originally developed in a study of the stability of repeaters in a loaded telephone transmission line. The envelope, R, was associated with the "returned current" produced by reflections from line irregularities. However, the study fell short of its object and the only results which seemed worth salvaging at the time were given in reference[25] cited in Section 3.3.

$$\overline{x_1 x_2} = \overline{I_c I_s'} = \sum_1^N w(f_n)\Delta f 2\pi(f_n - f_m) = b_1$$

$$\overline{x_4 x_5} = \overline{I_s I_c'} = -b_1$$

$$\overline{x_1 x_3} = \overline{I_c I_c''} = -\sum_1^N w(f)\Delta f 4\pi^2(f_n - f_m)^2 = -b_2$$

$$\overline{x_4 x_6} = \overline{I_s I_s''} = -b_2$$

$$\overline{x_2 x_3} = \overline{I_s' I_c''} = -b_3$$

$$\overline{x_5 x_6} = \overline{I_c' I_s''} = b_3$$

All of the other second moments are zero. The moment matrix M is thus

$$M = \begin{bmatrix} b_0 & b_1 & -b_2 & 0 & 0 & 0 \\ b_1 & b_2 & -b_3 & 0 & 0 & 0 \\ -b_2 & -b_3 & b_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_0 & -b_1 & -b_2 \\ 0 & 0 & 0 & -b_1 & b_2 & b_3 \\ 0 & 0 & 0 & -b_2 & b_3 & b_4 \end{bmatrix}$$

The adjoint matrix is

$$\begin{bmatrix} B_0 & B_1 & -B_2 & 0 & 0 & 0 \\ B_1 & B_{22} & -B_3 & 0 & 0 & 0 \\ -B_2 & -B_3 & B_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & B_0 & -B_1 & -B_2 \\ 0 & 0 & 0 & -B_1 & B_{22} & B_3 \\ 0 & 0 & 0 & -B_2 & B_3 & B_4 \end{bmatrix}$$

$$\begin{aligned} B_0 &= (b_2b_4 - b_3^2)B & B_{22} &= (b_0b_4 - b_2^2)B \\ B_1 &= -(b_1b_4 - b_2b_3)B & B_3 &= -(b_0b_3 - b_1b_2)B \\ B_2 &= (b_1b_3 - b_2^2)B & B_4 &= (b_0b_2 - b_1^2)B \end{aligned} \qquad (3.8\text{–}3)$$

$$B = b_0b_2b_4 + 2\,b_1b_2b_3 - b_2^3 - b_0b_3^2 - b_4b_1^2$$

$$|M| = B^2$$

where B is the determinant of the third order matrices in the upper left and lower right corners of M.

As in the earlier work, the distribution of $x_1, \cdots, x_6$ is normal in six dimensions. The exponent is $-[2\,|\,M\,|\,]^{-1}$ times

$$\begin{aligned} &B_0(x_1^2 + x_4^2) + 2B_1(x_1x_2 - x_4x_5) - 2B_2(x_1x_3 + x_4x_6) \\ &\quad + B_{22}(x_2^2 + x_5^2) \quad - 2B_3(x_2x_3 - x_5x_6) \\ &\quad\quad + B_4(x_3^2 + x_6^2) \end{aligned} \qquad (3.8\text{–}4)$$

In line with the earlier work we set

$$x_1 = I_c = R\cos\theta \qquad x_4 = I_s = R\sin\theta$$

$$x_2 = I_s' = R'\sin\theta + R\cos\theta\theta'$$

$$x_5 = I_c' = R'\cos\theta - R\sin\theta\theta'$$

$$x_3 = I_c'' = R''\cos\theta - 2R'\sin\theta\theta' - R\cos\theta\theta'^2 - R\sin\theta\theta''$$

$$x_6 = I_s'' = R''\sin\theta + 2R'\cos\theta\theta' - R\sin\theta\theta'^2 + R\cos\theta\theta''$$

The angle θ varies from 0 to 2π and θ' and θ'' vary from $-\infty$ to $+\infty$. By forming the Jacobian it may be shown that

$$dx_1\,dx_2\cdots dx_6 = R^3\,dR\,dR'\,dR''\,d\theta\,d\theta'\,d\theta''$$

Also, the quantities in (3.8–4) are

$$x_1^2 + x_4^2 = R^2 \qquad x_1x_3 + x_4x_6 = RR'' - R^2\theta'^2$$

$$x_1x_2 - x_4x_5 = R^2\theta' \qquad x_2^2 + x_5^2 = R'^2 + R^2\theta'^2$$

$$x_2x_3 - x_5x_6 = RR''\theta' - 2R'^2\theta' - R'R\theta'' - R^2\theta'^3$$

$$x_3^2 + x_6^2 = R''^2 - 2RR''\theta'^2 + 4R'^2\theta'^2 + 4RR'\theta'\theta'' + R^2\theta'^4 + R^2\theta''^2$$

The expression for $p(R, 0, R'')$ is obtained when we set these values of the x's in (3.8–4) and integrate the resulting probability density over the ranges of θ, θ', θ'':

$$p(R, 0, R'') = \frac{R^3}{8\pi^3 B}\int_0^{2\pi} d\theta \int_{-\infty}^{+\infty} d\theta' \int_{-\infty}^{+\infty} d\theta'' \qquad (3.8\text{–}5)$$

$$\exp -\frac{1}{2B^2}\,[B_0R^2 + 2B_1R^2\theta' - 2B_2(RR'' - R^2\theta'^2) + B_{22}R^2\theta'^2 - 2B_3R\theta'(R'' - R\theta'^2) + B_4(R''^2 - 2RR''\theta'^2 + R^2\theta'^4 + R^2\theta''^2)]$$

The integrations with respect to θ and θ'' may be performed at once leaving $p(R, 0, R'')$ expressed as a single integral which, unfortunately, appears to be difficult to handle. For this reason we assume that $w(f)$ is symmetrical about the mid-band frequency f_m. From (3.8–2), b_1 and b_3 are zero and from (3.8–3), B_1 and B_3 are zero.

With this assumption (3.8–5) yields

$$p(R, 0, R'') = R^2(2\pi)^{-3/2} B_4^{-1/2} \int_{-\infty}^{+\infty} d\theta' \tag{3.8–6}$$

$$\exp -\frac{1}{2B^2}[B_0 R^2 + R([B_{22} + 2B_2]R\theta'^2 - 2B_2 R'') + B_4(R'' - R\theta'^2)^2]$$

The probability that a maximum occurs in the elementary rectangle $dR\ dt$ is, from (3.8–1), $p(t, R)\ dR\ dt$ where

$$p(t, R) = -\int_{-\infty}^{0} p(R, 0, R'')R''\ dR'' \tag{3.8–7}$$

We put (3.8–6) in this expression and make the following change of variables.

$$\begin{aligned} x &= \frac{B_4^{1/2}}{\sqrt{2}\,B} R\theta'^2, \qquad y = -\frac{B_4^{1/2}}{\sqrt{2}\,B} R'' \\ z &= -\frac{B_2}{\sqrt{2B_4}\,B} R = \frac{b_2^2}{\sqrt{2B_4}} R \\ b &= -\frac{(B_{22} + 2B_2)}{2B\,b_2^2} = \left[\frac{3}{2} - \frac{b_0 b_4}{2b_2^2}\right] = \tfrac{1}{2}(3 - a^2) \\ a^2 &= \frac{B_0}{2B^2}\frac{2B_4}{b_2^4} = \frac{b_0 b_4}{b_2^2} \end{aligned} \tag{3.8–8}$$

where we have used the expressions for the B's obtained by setting b_1 and b_3 to zero in (3.8–3). Thus

$$p(t, R) = \frac{1}{b_0 b_2^4}\left(\frac{Bz}{2\pi}\right)^{3/2} \int_0^\infty y\,dy \int_0^\infty x^{-1/2}\,dx \tag{3.8–9}$$

$$\exp\left[-a^2 z^2 + 2bzx + 2zy - (x + y)^2\right]$$

As was to be expected, this expression shows that $p(t, R)$ is independent of t.

A series for $p(t, R)$ may be obtained by expanding $\exp 2z(y + bx)$ and then integrating termwise. We use

$$\int_0^\infty dy \int_0^\infty dx\, x^\mu y^\gamma e^{-(x+y)^2} = \frac{\sqrt{\pi}}{2^{\mu+\gamma+2}} \frac{\Gamma(\gamma + 1)\Gamma(\mu + 1)}{\Gamma\left(\frac{\mu + \gamma + 3}{2}\right)}$$

which may be evaluated by setting

$$x = \rho^2 \cos^2 \varphi, \qquad y = \rho^2 \sin^2 \varphi$$

The double integral in (3.8–9) becomes

$$e^{-a^2z^2}\sqrt{\frac{\pi}{2}}\sum_{n=0}^{\infty}\frac{(2z)^n}{n!}\sum_{m=0}^{n}\frac{n!\,b^m}{m!\,(n-m)!}\frac{\Gamma(m+\frac{1}{2})\Gamma(n-m+2)}{2^{n+2}\,\Gamma\left(\frac{n}{2}+\frac{7}{4}\right)}$$

$$=\pi 2^{-5/2}\sum_{n=0}^{\infty}\frac{z^n e^{-a^2z^2}}{\Gamma\left(\frac{n}{2}+\frac{7}{4}\right)}A_n$$

where $A_0 = 1$ and

$$A_n = \sum_{m=0}^{n}\frac{(\frac{1}{2})(\frac{3}{2})\cdots(m-\frac{1}{2})}{m!}(n-m+1)b^m, \qquad 0 < n \tag{3.8–10}$$

$$A_n \sim (n+1)(1-b)^{-1/2} - \frac{b}{2}(1-b)^{-3/2}, \qquad n \text{ large}$$

The term corresponding to $m = 0$ in (3.8–10) is $n + 1$.

We thus obtain

$$\begin{aligned} p(t, R) &= \frac{e^{-a^2z^2}}{4b_0b_2^4}\frac{(Bz)^{3/2}}{\sqrt{\pi}}\sum_{n=0}^{\infty}\frac{z^n}{\Gamma\left(\frac{n}{2}+\frac{7}{4}\right)}A_n \\ &= \frac{e^{-a^2z^2}}{4\sqrt{\pi}}\frac{b_2^{1/2}}{b_0}(a^2-1)^{3/2}z^{3/2}\sum_{n=0}^{\infty}\frac{z^nA_n}{\Gamma\left(\frac{n}{2}+\frac{7}{4}\right)} \end{aligned} \tag{3.8–11}$$

We are interested in the expected number, N, of maxima per second. From the similar work for I, it follows that N is the coefficient of dt when (3.8–1) is integrated with respect to R from 0 to ∞. Thus from (3.8–7) and

$$\begin{aligned} dR &= \sqrt{2B_4}\,b_2^{-2}\,dz = (2b_0B)^{1/2}b_2^{-3/2}\,dz \\ &= [2b_0(a^2-1)]^{1/2}\,dz \end{aligned}$$

we find

$$\begin{aligned} N &= \int_0^{\infty} p(t, R)\,dR \\ &= \frac{(a^2-1)^2}{(2a)^{5/2}}\left(\frac{b_2}{\pi b_0}\right)^{1/2}\sum_{n=0}^{\infty}\frac{\Gamma\left(\frac{n}{2}+\frac{5}{4}\right)}{\Gamma\left(\frac{n}{2}+\frac{7}{4}\right)}\frac{A_n}{a^n} \end{aligned} \tag{3.8–12}$$

Equations (3.8–11) and (3.8–12) have been derived on the assumption that $w(f)$ is symmetrical about f_m, i.e. the band pass filter attenuation is

symmetrical about the mid-band frequency. We now go a step further and assume an ideal band pass filter:

$$\begin{aligned} w(f) &= w_0 \qquad f_a < f < f_b \\ w(f) &= 0 \qquad \text{otherwise} \\ 2f_m &= f_a + f_b \end{aligned} \tag{3.8-13}$$

Putting these in (3.8-2) we obtain zero for b_1 and b_3 and also

$$\begin{aligned} b_0 &= w_0(f_b - f_a) = \psi_0 \\ b_2 &= \frac{\pi^2 w_0}{3}(f_b - f_a)^3 \\ b_4 &= \frac{\pi^4 w_0}{5}(f_b - f_a)^5 \\ a^2 &= \tfrac{9}{5} \\ b &= \tfrac{1}{2}(3 - a^2) = \tfrac{3}{5} \\ R &= [2b_0(a^2 - 1)]^{1/2} z = [\tfrac{8}{5}\psi_0]^{1/2} z \\ \left(\frac{b_2}{\pi b_0}\right)^{1/2} &= \left[\frac{\pi}{3}\right]^{1/2}(f_b - f_a), \qquad a^2 z^2 = \frac{9R^2}{8\psi_0} \end{aligned} \tag{3.8-14}$$

n	A_n	n	A_n
0	1	4	6.775
1	2.3	5	8.333
2	3.735	6	9.9002
3	5.238	7	11.4736

$A_n \sim 1.5811\, n + .3953$

From (3.8-12) we find that the expected number of maxima per second of the envelope is

$$N = .64110\,(f_b - f_a) \tag{3.8-15}$$

assuming an ideal band pass filter.

The distribution of the maxima of R for an ideal band pass filter may be obtained by placing the results of (3.8-14) in (3.8-11). This gives

$$p(t, R)\,dR = \frac{dR}{\psi_0^{1/2}} \frac{(f_b - f_a)}{4} \sqrt{\frac{\pi}{3}} \left(\frac{4z}{5}\right)^{3/2} e^{-a^2 z^2} \sum_{n=0}^{\infty} \frac{z^n A_n}{\Gamma\left(\frac{n}{2} + \frac{7}{4}\right)}$$

It is convenient to define y as the ratio

$$y = \frac{R}{\text{r.m.s. } I(t)} = \frac{R}{\psi_0^{1/2}} = \left(\tfrac{8}{5}\right)^{1/2} z$$

where R is understood to correspond to a maximum of the envelope. Since the value of R corresponding to a maximum of the envelope selected at random is a random variable, y is also a random variable. Its probability density is $p_R(y)$, where

$$p_R(y)\, dy = \frac{p(t, R)\, dR}{0.64110(f_b - f_a)}$$

$p_R(y)$ has been computed and is plotted as a function of y in Fig. 3.

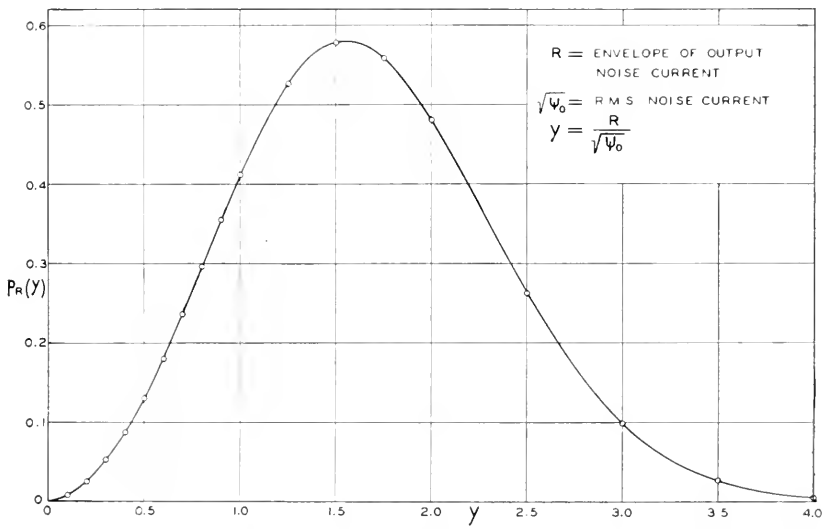


Fig. 3—Distribution of maxima of envelope of noise current. Noise through ideal band-pass filter.

$\frac{p_R(y)}{\sqrt{\psi_0}}\, dR$ = probability that a maximum of R selected at random lies between R and $R + dR$.

The distribution function $P(R_{\max} < y\sqrt{\psi_0})$ defined by

$$P(R_{\max} < y\sqrt{\psi_0}) = \int_0^y p_R(y)\, dy$$

and which gives the probability that a maximum of the envelope selected at random is less than a specified value $y\sqrt{\psi_0} = R$, is plotted in Fig. 4 together with other curves of the same nature.

When y is large, say greater than 2.5,

$$p_R(y) \sim \frac{\sqrt{\frac{\pi}{6}}}{.64110}(y^2 - 1)e^{-y^2/2}$$

$$P(R_{\max} < y\sqrt{\psi_0}) \sim 1 - \frac{\sqrt{\frac{\pi}{6}}}{.64110} ye^{-y^2/2}$$

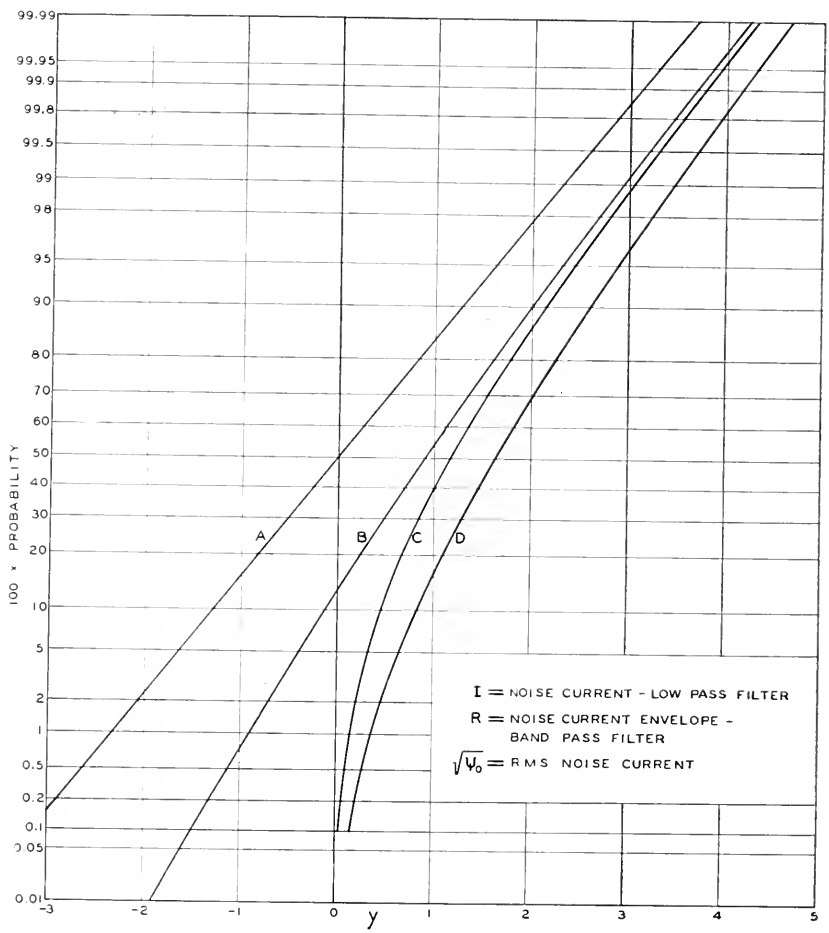


Fig. 4—Distribution of maxima

$A = P(I < y\sqrt{\psi_0})$ = probability of I being less than $y\sqrt{\psi_0}$. Similarly $C = P(R < y\sqrt{\psi_0})$.

$B = P(I \max < y\sqrt{\psi_0})$ = probability of random maximum of I being less than $y\sqrt{\psi_0}$. Similarly $D = P(R \max < y\sqrt{\psi_0})$.

The asymptotic expression for $p_R(y)$ may be obtained from the integral (3.8–9) for $p(t, R)$. Indeed, replacing the variables of integration x, y in (3.8–9) by

$$x' = x$$

$$y' = x + y,$$

integrating a portion of the y' integral by parts, and assuming $b < 1$ ($a^2 \geq 1$, by Schwarz's inequality, so that $b \leq 1$ always) leads to

$$p(t, R) \sim \left(\frac{b_2}{2\pi}\right)^{\frac{1}{2}} \frac{e^{-R^2/2\psi_0}}{\psi_0} \left(\frac{R^2}{\psi_0} - 1\right)$$

when R is large.

If, instead of an ideal band pass filter, we assume that $w(f)$ is given by

$$w(f) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(f-f_m)^2/2\sigma^2}, \qquad f_m \gg \sigma \tag{3.8–16}$$

we find that

$$b_0 = 1$$

$$b_2 = 4\pi^2\sigma^2$$

$$b_4 = 16\pi^4 \cdot 3\sigma^4$$

$$a^2 = 3, b = 0$$

$$A_n = (n + 1)$$

Some rough work indicates that the sum of the series in (3.8–12) is near 3.97. This gives the expected number of maxima of the envelope as

$$N = 2.52\sigma \tag{3.8–17}$$

per second.

The pass band is determined by σ. It appears difficult to compare this with an ideal band pass filter. If we use the fact that the filter given by

$$w(f) = w_0 \exp\left[-\pi\left(\frac{f - f_m}{f_b - f_a}\right)^2\right]$$

passes the same average amount of power as does an ideal band pass filter whose pass band is $f_b - f_a$, we have

$$f_b - f_a = \sigma\sqrt{2\pi}$$

and the expression for N becomes $1.006\,(f_b - f_a)$.

3.9 Energy Fluctuation

Some information regarding the statistical behavior of the random variable

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

where $I(t)$ is a noise current and t_1 is chosen at random, has been given in a recent article.[35] Here we study this behavior from a somewhat different point of view.

If we agree to use the representations (2.8–1) or (2.8–6) we may write, as in the paper, the random variable E as

$$E = \int_{-T/2}^{T/2} I^2(t)\, dt \tag{3.9–2}$$

where the randomness on the right is due either to the a_n's and b_n's if (2.8–1) is used or to the φ_n's if (2.8–6) is used.

The average value of E is m_T where, from (3.1–2),

$$\begin{aligned} \bar{E} = m_T &= \int_{-T/2}^{T/2} \overline{I^2(t)}\, dt = \int_{-T/2}^{T/2} \psi(0)\, dt = T\psi_0 \\ &= T\int_0^\infty w(f)\, df \end{aligned} \tag{3.9–3}$$

The second moment of E is

$$\overline{E^2} = \int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2\, \overline{I^2(t_1) I^2(t_2)} \tag{3.9–4}$$

If, for the time being, we set t_2 equal to $t_1 + \tau$, it is seen from section 3.2 that we have an expression for the probability density of $I(t_1)$ and $I(t_1 + \tau)$ and hence we may obtain the required average:

$$\begin{aligned} \overline{I_1^2 I_2^2} = \frac{1}{2\pi A} \int_{-\infty}^{+\infty} dI_1 \int_{-\infty}^{+\infty} dI_2\, I_1^2 I_2^2 \exp \\ \left(-\frac{1}{2A^2}\left(\psi_0 I_1^2 + \psi_0 I_2^2 - 2\psi_\tau I_1 I_2\right)\right) \end{aligned} \tag{3.9–5}$$

$$A^2 = \psi_0^2 - \psi_\tau^2, \quad I_1 = I(t_1), \quad I_2 = I(t_1 + \tau) = I(t_2)$$

The integral may be evaluated by (3.5–6) when we set

$$\begin{aligned} I_1 &= Ax\sqrt{\frac{2}{\psi_0}}, \qquad I_2 = Ay\sqrt{\frac{2}{\psi_0}} \\ \psi_\tau &= -\psi_0 \cos\varphi \\ A &= \psi_0 \sin\varphi \end{aligned} \tag{3.9–6}$$

[35] "Filtered Thermal Noise—Fluctuation of Energy as a Function of Interval Length", *Jour. Acous. Soc. Am.*, 14 (1943), 216–227.

Thus

$$\begin{aligned}\overline{I_1^2 I_2^2} &= \psi_0^2(1 + 2\cos^2\varphi) \\ &= \psi_0^2 + 2\psi_\tau^2\end{aligned} \tag{3.9-7}$$

Incidentally, this gives an expression for the correlation function of $I^2(t)$. Replacing τ by its value of $t_2 - t_1$ and returning to (3.9-4),

$$\overline{E^2} = T^2\psi_0^2 + 2\int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2 \psi^2(t_2 - t_1) \tag{3.9-8}$$

When we introduce σ_T, the standard deviation of E, and use

$$\sigma_T^2 = \overline{E^2} - m_T^2$$

we obtain

$$\begin{aligned}\sigma_T^2 = \overline{(E - \bar{E})^2} &= 2\int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2 \psi^2(t_2 - t_1) \\ &= 4\int_0^T (T - x)\psi^2(x)\, dx\end{aligned}$$

where the second line may be obtained from the first either by changing the variables of integration, as in (3.9-27), or by the method used below in dealing with $\overline{E^3}$. I am indebted to Prof. Kac for pointing out the advantage obtained by reducing the double integral to a single integral. It should be noted that the limits of integration $-T/2$, $T/2$ in the double integral may be replaced by 0, T by making the change of variable $t = t' - T/2$ for both t_1 and t_2.

When we use

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

we obtain the result stated in the paper, namely,

$$\sigma_T^2 = \int_0^\infty w(f_1)\, df_1 \int_0^\infty w(f_2)\, df_2 \left[\frac{\sin^2 \pi(f_1 + f_2)T}{\pi^2(f_1 + f_2)^2} + \frac{\sin^2 \pi(f_1 - f_2)T}{\pi^2(f_1 - f_2)^2}\right] \tag{3.9-9}$$

If this formula is applied to a relatively narrow band-pass filter and if $T(f_b - f_a) >> 1$ the contribution of the $f_1 + f_2$ term may be neglected and we have the approximation

$$\begin{aligned}\sigma_T^2 &\approx \int_{f_a}^{f_b} w_0\, df_1 \int_{-\infty}^{+\infty} w_0\, df_2 \frac{\sin^2 \pi(f_1 - f_2)T}{\pi^2(f_1 - f_2)^2} \\ &= w_0^2 T(f_b - f_a) \\ &= w_0 m_T\end{aligned} \tag{3.9-10}$$

where, from (3.9-3)

$$m_T = w_0 T(f_b - f_a) \tag{3.9-11}$$

The third moment $\overline{E^3}$ may be computed in the same way. However, in this case it pays to introduce the characteristic function for the distribution of $I(t_1)$, $I(t_2)$, $I(t_3)$. Since this distribution is normal its characteristic function is

$$\begin{aligned}\text{Average exp } [iz_1 I_1 + iz_2 I_2 + iz_3 I_3] \\ = \exp -\Big[\frac{\psi_0}{2}(z_1^2 + z_2^2 + z_3^2) + \psi(t_2 - t_1) z_1 z_2 \\ + \psi(t_3 - t_1) z_1 z_3 + \psi(t_3 - t_2) z_2 z_3\Big]\end{aligned} \tag{3.9-12}$$

From the definition of the characteristic function it follows that

$$\begin{aligned}\overline{I_1^2 I_2^2 I_3^2} &= -\text{coeff. of } \frac{z_1^2 z_2^2 z_3^2}{2!2!2!} \text{ in ch. f.} \\ &= \psi_0^3 + 2\psi_0(\psi_{21}^2 + \psi_{31}^2 + \psi_{32}^2) \\ &\quad + 8\psi_{21}\psi_{31}\psi_{32}\end{aligned} \tag{3.9-13}$$

where we have written ψ_{21} for $\psi(t_2 - t_1)$, etc. When (3.9-13) is multiplied by $dt_1\, dt_2\, dt_3$, the variables integrated from 0 to T, and the above double integral expression for σ_T^2 used, we find

$$\overline{(E - \bar{E})^3} = 2!2^2 \int_0^T dt_1 \int_0^T dt_2 \int_0^T dt_3 \psi_{21}\psi_{31}\psi_{32}\,.$$

Denoting the triple integral on the right by J and differentiating,

$$\begin{aligned}\frac{dJ}{dT} &= 3\int_0^T dt_1 \int_0^T dt_2 \psi(t_2 - t_1)\psi(T - t_1)\psi(T - t_2) \\ &= 3\int_0^T dx \int_0^T dy \psi(x - y)\psi(x)\psi(y) \\ &= 6\int_0^T dx \int_0^x dy \psi(x - y)\psi(x)\psi(y)\end{aligned}$$

In going from the first line to the second t_1 and t_2 were replaced by $T - x$ and $T - y$, respectively. In going from the second to the third use was made of the relations symbolized by

$$\begin{aligned}\int_0^T dx \int_0^T dy &= \int_0^T dx \int_0^x dy + \int_0^T dx \int_x^T dy \\ &= \int_0^T dx \int_0^x dy + \int_0^T dy \int_0^y dx\end{aligned}$$

and of the fact that the integrand is symmetrical in x and y. Integrating dJ/dT with respect to T from 0 to T_1, using the formula

$$\int_0^{T_1} dT \int_0^T f(x)\,dx = \int_0^{T_1} (T_1 - x)f(x)\,dx,$$

noting that J is zero when T is zero, and dropping the subscript on T_1 finally gives

$$\overline{(E - \bar{E})^3} = 48 \int_0^T dx \int_0^x dy(T - x)\psi(x)\psi(y)\psi(x - y).$$

$\overline{E^4}$ may be treated in a similar way. It is found that

$$\overline{(E - \bar{E})^4} - 3\overline{(E - \bar{E})^2}^2 = 312^3 \int_0^T dt_1 \int_0^T dt_2 \int_0^T dt_3 \int_0^T dt_1 \psi_{21}\psi_{31}\psi_{42}\psi_{43}$$

which may be reduced to the sum of two triple integrals. It is interesting to note that the expression on the left is the fourth semi-invariant of the random variable E and gives us a measure of the peakedness of the distribution (kurtosis). Likewise, the second and third moments about the mean are the second and third semi-invariants of E. This suggests that possibly the higher semi-invariants may also be expressed as similar multiple integrals.

So far, in this section, we have been speaking of the statistical constants of E. The determination of an exact expression for the probability density of E, in which T occurs as a parameter, seems to be quite difficult.

When T is very small E is approximately $I^2(t)T$. The probability that E lies in dE is the probability that the current lies in $-I$, $-I - dI$ plus the probability that the current lies in I, $I + dI$:

$$\frac{2dI}{\sqrt{2\pi\psi_0}} \exp -\frac{I^2}{2\psi_0} = (2\pi\psi_0 ET)^{-1/2} \exp -\frac{E}{2\psi_0 T}\, dE \qquad (3.9\text{–}14)$$

where E is positive,

$$I = \left(\frac{E}{T}\right)^{1/2}, \qquad dI = \frac{1}{2}(ET)^{-1/2}\, dE$$

and T is assumed to be so small that $I(t)$ does not change appreciably during an interval of length T.

When T is very large we may divide it into a number of intervals, say n, each of length T/n. Let E_r be the contribution of the r th interval. The energy E for the entire interval is then

$$E = E_1 + E_2 + \cdots + E_n$$

If the sub-intervals are large enough the E_r's are substantially independent random variables. If in addition n is large enough E is distributed nor-

mally, approximately. Hence when T is very large the probability that E lies in dE is

$$\frac{dE}{\sigma_T\sqrt{2\pi}} \exp - \frac{(E - m_T)^2}{2\sigma_T^2} \tag{3.9–15}$$

where

$$\begin{aligned} m_T &= T \int_0^\infty w(f)\, df \\ \sigma_T^2 &= T \int_0^\infty w^2(f)\, df \end{aligned} \tag{3.9–16}$$

the second relation being obtained by letting $T \to \infty$ in (3.9–9). The analogy with Campbell's theorem, section 1.2, is evident. When we deal with a band pass filter we may use (3.9–10) and (3.9–11).

Consider a relatively narrow band pass filter such that we may find a T for which $Tf_a >> 2\pi$ but $T(f_b - f_a) << .64$. Thus several cycles of frequency f_a are contained in T but, from (3.8–15), the envelope does not change appreciably during this interval. Thus throughout this interval $I(t)$ may be considered to be a sine wave of amplitude R. The corresponding value of E is approximately

$$E = T\frac{R^2}{2}$$

where the distribution of the envelope R is given by (3.7–10). From this it follows that the probability of E lying in dE is

$$\frac{dE}{\psi_0 T} \exp - \frac{E}{\psi_0 T} = \frac{dE}{m_T} e^{-E/m_T} \tag{3.9–17}$$

when E is small but not too small.

When we look at (3.9–14) and (3.9–17) we observe that they are of the form

$$\frac{a^{n+1}E^n}{\Gamma(n+1)} e^{-aE}\, dE \tag{3.9–18}$$

Moreover, the normal law (3.9–15), may be obtained from this by letting n become large. This suggests that an approximate expression for the distribution of E is given by (3.9–18) when a and n are selected so as to give the values of m_T and σ_T obtained from (3.9–3) and (3.9–9). This gives

$$a = \frac{m_T}{\sigma_T^2}, \qquad n + 1 = \frac{m_T^2}{\sigma_T^2} \tag{3.9–19}$$

and if we drop the subscript T and substitute the value of a in (3.9–18) we get

$$\frac{\left(\frac{mE}{\sigma^2}\right)^n}{\Gamma(n+1)} \exp\left(-\frac{mE}{\sigma^2}\right) d\left(\frac{mE}{\sigma^2}\right), \qquad n = \frac{m^2}{\sigma^2} - 1 \qquad (3.9\text{–}20)$$

An idea of how this distribution behaves may be obtained from the following table:

n	$T(f_b - f_a)$	$x_{.25}$	$x_{.50}$	$x_{.75}$	$\frac{x_{.25}}{x_{.50}}$	$\frac{x_{.75}}{x_{.50}}$
0	0	.29	.695	1.39	.415	2.00
1	1.45	.96	1.68	2.69	.572	1.60
2	2.4	1.73	2.67	3.94	.647	1.47
3	3.4	2.54	3.67	5.12	.692	1.39
5	5.4	4.22	5.67	7.42	.744	1.31
10	10.5	8.63	10.67	13.02	.808	1.22
24	25	21.47	24.67	28.17	.870	1.14
48	50	44.1	48.7	53.5	.905	1.10

where n is the exponent in (3.9–20). The column $T(f_b - f_a)$ holds only for a narrow band pass filter and was obtained by reading the curve y_A in Fig. 1 of the above mentioned paper. The figures in this column are not very accurate. The next three columns give the points which divide the distribution into four intervals of equal probability:

$$x_{.25} = \frac{mE_{.25}}{\sigma^2}, \quad E_{.25} = \text{energy exceeded } 75\% \text{ of time}$$

$$x_{.50} = \frac{mE_{.50}}{\sigma^2}, \quad E_{.50} = \text{energy exceeded } 50\% \text{ of time}$$

$$x_{.75} = \frac{mE_{.75}}{\sigma^2}, \quad E_{.75} = \text{energy exceeded } 25\% \text{ of time}$$

The values in these columns were obtained from Pearson's table of the incomplete gamma function. The last two columns show how the distribution clusters around the average value as the normal law is approached.

For the larger values of n we expected the normal law (3.9–15) to be approached. Since, for this law the 25, 50, and 75 per cent points are at $m - .675\sigma$, m, and $m + .675\sigma$ we have to a first approximation

$$\begin{aligned} x_{.50} &= \frac{m^2}{\sigma^2} = (n+1) \approx T(f_b - f_a) \\ x_{.25} &= \frac{m}{\sigma^2}(m - .675\sigma) = x_{.50} - .675\sqrt{x_{.50}} \\ x_{.75} &= x_{.50} + .675\sqrt{x_{.50}} \end{aligned} \qquad (3.9\text{–}21)$$

This agrees with the table.

Thiede[36] has studied the mean square value of the fluctuations of the integral

$$A(t) = \int_{-\infty}^{t} I^2(\tau) e^{-\alpha(t-\tau)}\, d\tau \tag{3.9-22}$$

The reading of a hot wire ammeter through which a current I is passing is proportional to $A(t)$. α is a constant of the meter. Here we study $A(t)$ by

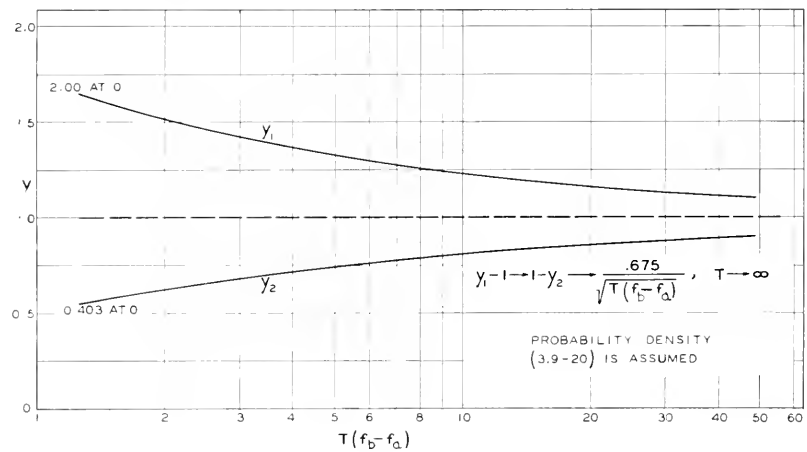


Fig. 5*—Filtered thermal noise—spread of energy fluctuation

$$E = \int_{t_1}^{t_1+T} I^2(t)\, dt, \qquad t_1 \text{ random}, \qquad I \text{ is noise current.}$$

$y_1 = E_{.75}/E_{.50}$, $y_2 = E_{.25}/E_{.50}$.
$f_b - f_a$ = band width of filter.

first obtaining its correlation function. This method of approach enables us to extend Thiede's results

The distributed portion of the power spectrum of $A(t)$ is given by (3.9–30). When the power spectrum $w(f)$ of $I(t)$ is zero except over the band $f_a < f < f_b$ where it is w_0 , the power spectrum of $A(t)$ is

$$\frac{2w_0^2(f_b - f_a - f)}{\alpha^2 + 4\pi^2 f^2} \quad \text{for} \quad 0 < f < f_b - f_a$$

and is zero from $f_b - f_a$ up to $2f_a$. The spectrum from $2f_a$ to $2f_b$ is not zero, and may be obtained from (3.9–34). The mean square fluctuation of $A(t)$ is given, in the general case, by (3.9–28) and (3.9–32). For the band pass case, when $(f_b - f_a)/\alpha$ is large,

$$\text{r.m.s.} \frac{A(t) - \bar{A}}{\bar{A}} = \left[\frac{\alpha}{2(f_b - f_a)}\right]^{1/2}$$

[36] *Elec. Nachr. Tek.*, 13 (1936), 84–95. This is an excellent article.

* Note added in proof. The value of y_2 at 0 should be .415 instead of .403.

We start by setting $\tau = t - u$ which transforms the integral for $A(t)$ into

$$A(t) = \int_0^\infty I^2(t - u)e^{-\alpha u}\, du \tag{3.9-23}$$

In order to obtain the correlation function $\Psi(\tau)$ for $A(t)$ we multiply $A(t)$ by $A(t + \tau)$ and average over all the possible currents

$$\begin{aligned}\Psi(\tau) &= \overline{A(t)A(t + \tau)} \\ &= \int_0^\infty e^{-\alpha u}\, du \int_0^\infty e^{-\alpha v}\, dv \text{ ave. } I^2(t - u)I^2(t + \tau - v)\end{aligned}$$

Just as in (3.9–4) the average in the integrand is the correlation function of $I^2(t)$, the argument being $t + \tau - v - t + u = \tau + u - v$. From (3.9–7) it is seen that this is

$$\psi_0^2 + 2\psi^2(\tau + u - v)$$

where $\psi(\tau)$ is the correlation function of $I(t)$. Hence

$$\Psi(\tau) = \frac{\psi_0^2}{\alpha^2} + 2\int_0^\infty du \int_0^\infty dv\, e^{-\alpha u - \alpha v}\psi^2(\tau + u - v) \tag{3.9-24}$$

From the integral (3.9–23) for $A(t)$ it is seen that the average value of $A(t)$ is

$$\bar{A} = \frac{\bar{I}^2}{\alpha} = \frac{\psi_0}{\alpha} \tag{3.9-25}$$

where we have used

$$\psi_0 = \psi(0) = \int_0^\infty w(f)\, df = \bar{I}^2$$

Using this result again, only this time applying it to $A(t)$, gives

$$\begin{aligned}\overline{A^2(t)} &= \Psi(0) \\ &= \bar{A}^2 + 2\int_0^\infty du \int_0^\infty dv\, e^{-\alpha u - \alpha v}\psi^2(u - v)\end{aligned} \tag{3.9-26}$$

The double integrals may be transformed by means of the change of variable $u + v = x$, $u - v = y$. Then (3.9–24) becomes

$$\begin{aligned}\Psi(\tau) &= \bar{A}^2 + \left[\int_0^\infty dy \int_y^\infty dx + \int_{-\infty}^0 dy \int_{-y}^\infty dx\right] e^{-\alpha x}\psi^2(\tau + y) \\ &= \bar{A}^2 + \frac{1}{\alpha}\int_0^\infty e^{-\alpha y}[\psi^2(\tau + y) + \psi^2(\tau - y)]\, dy\end{aligned} \tag{3.9-27}$$

When we make use of the fact that $\psi(y)$ is an even function of y we see, from (3.9–26), that the mean square fluctuation of $A(t)$ is

$$\overline{(A(t) - \bar{A})^2} = \overline{A^2(t)} - \bar{A}^2 = \frac{2}{\alpha}\int_0^\infty e^{-\alpha y}\psi^2(y)\,dy \qquad (3.9\text{–}28)$$

$\Psi(\tau)$ may be expressed in terms of integrals involving the power spectrum $w(f)$ of $I(t)$. The work starts with (3.9–24) and is much the same as in going from (3.9–8) to (3.9–9). The result is

$$\Psi(\tau) = \bar{A}^2 + \int_0^\infty df_1 \int_0^\infty df_2 w(f_1)w(f_2) \left[\frac{\cos 2\pi(f_1 + f_2)\tau}{\alpha^2 + [2\pi(f_1 + f_2)]^2} + \frac{\cos 2\pi(f_1 - f_2)\tau}{\alpha^2 + [2\pi(f_1 - f_2)]^2}\right]$$

It is convenient to define $w(-f)$ for negative frequencies to be equal to $w(f)$. The integration with respect to f_2 may then be taken from $-\infty$ to $+\infty$ and we get

$$\Psi(\tau) = \bar{A}^2 + \int_0^\infty df_1 \int_{-\infty}^{+\infty} df_2 w(f_1)w(f_2) \frac{\cos 2\pi(f_1 - f_2)\tau}{\alpha^2 + [2\pi(f_1 - f_2)]^2} \qquad (3.9\text{–}29)$$

The power spectrum $W(f)$ of $A(t)$ may be obtained by integrating $\Psi(\tau)$:

$$W(f) = 4\int_0^\infty \Psi(\tau)\cos 2\pi f\tau\,d\tau$$

Let us concern ourselves with the fluctuating portion $A(t) - \bar{A}$ of $A(t)$. Its power spectrum $W_c(f)$ is

$$W_c(f) = 4\int_0^\infty (\Psi(\tau) - \bar{A}^2)\cos 2\pi f\tau\,d\tau$$

The integration is simplified by using Fourier's integral formula in the form

$$\int_0^\infty d\tau \int_{-\infty}^{+\infty} df_2 F(f_2)\cos 2\pi(u - f_2)\tau = \tfrac{1}{2}F(u)$$

We get

$$W_c(f) = \frac{1}{\alpha^2 + 4\pi^2 f^2}\int_0^\infty df_1[w(f_1)w(f + f_1) + w(f_1)w(-f + f_1)] = \frac{1}{\alpha^2 + 4\pi^2 f^2}\int_{-\infty}^{+\infty} w(f_1)w(f - f_1)\,df_1 \qquad (3.9\text{–}30)$$

The simplicity of this result suggests that a simpler derivation may be found. If we attempt to use the result

$$\bar{w}(f) = \operatorname*{Limit}_{T\to\infty} \frac{\overline{2|S(f)|^2}}{T} \qquad (2.5\text{–}3)$$

where $S(f)$ is given by (2.1–2) we find that we need the result

$$\underset{T\to\infty}{\text{Limit}} \frac{2}{T}\int_0^T dt_1 \int_0^T dt_2\, e^{2\pi i f(t_2-t_1)}\, I^2(t_1)I^2(t_2) = \int_{-\infty}^{+\infty} w(f_1)w(f-f_1)\, df_1 \tag{3.9–31}$$

where $f > 0$ and $I(t)$ is a noise current with $w(f)$ as its power spectrum. This may be proved by using (3.9–7) and

$$8\int_0^\infty \psi^2(\tau)\cos 2\pi f\tau\, d\tau = \int_{-\infty}^{+\infty} w(x)w(f-x)\, dx$$

which is given by equation (4C-6) in Appendix 4C.

An expression for the mean square fluctuation of $A(t)$ in terms of $w(f)$ may be obtained by setting τ equal to zero in (3.9–29)

$$\overline{(A(t)-\bar{A})^2} = \Psi(0) - \bar{A}^2 = \int_0^\infty df_1 \int_{-\infty}^{+\infty} df_2 \frac{w(f_1)w(f_2)}{\alpha^2 + 4\pi^2(f_1-f_2)^2} \tag{3.9–32}$$

The same result may be obtained by integrating $W_c(f)$, (3.9–30), from 0 to ∞:

$$\int_0^\infty \frac{df}{\alpha^2+4\pi^2f^2}\int_{-\infty}^{+\infty} df_1 w(f_1)w(f-f_1) \tag{3.9–33}$$

Although this differs in appearance from (3.9–32) it may be transformed into that expression by making use of $w(-f) = w(f)$.

Suppose that $I(t)$ is the current through an ideal band pass filter so that $w(f)$ is zero except in the band $f_a < f < f_b$ where it is w_0. Then, if $3f_a > f_b$,

$$\bar{A} = \frac{w_0}{\alpha}(f_b - f_a) \tag{3.9–34}$$

$$\int_{-\infty}^{+\infty} w(x)w(f-x)\, dx = \begin{cases} 2w_0^2(f_b - f_a - f) & 0 < f \le f_b - f_a \\ w_0^2(f-2f_a) & 2f_a \le f \le f_b + f_a \\ w_0^2(2f_b - f) & f_b + f_a \le f \le 2f_b \end{cases}$$

and is zero outside these ranges. The power spectrum $W_c(f)$ may be obtained immediately from (3.9–30) by dividing these values by $\alpha^2 + 4\pi^2f^2$.

From (3.9–33)

$$\overline{(A(t)-\bar{A})^2} = 2w_0^2\int_0^{f_b-f_a}\frac{(f_b-f_a-f)\,df}{\alpha^2+4\pi^2f^2} + w_0^2\int_{2f_a}^{f_b+f_a}\frac{(f-2f_a)}{\alpha^2+4\pi^2f^2}\,df + w_0^2\int_{f_b+f_a}^{2f_b}\frac{(2f_b-f)}{\alpha^2+4\pi^2f^2}\,df$$

If an exact answer is desired the integrations may be performed. When we assume that $f_b - f_a << f_b + f_a$ we may obtain approximations for the last two integrals.

$$\overline{(A(t) - \bar{A})^2} \approx w_0^2 \left[\frac{f_b - f_a}{\pi\alpha} \tan^{-1} \frac{2\pi(f_b - f_a)}{\alpha} - \frac{1}{4\pi^2} \log \frac{\alpha^2 + 4\pi^2(f_b - f_a)^2}{\alpha^2} + \frac{(f_b - f_a)^2}{\alpha^2 + 4\pi^2(f_b + f_a)^2} \right]$$

Furthermore, if $2\pi(f_b - f_a)/\alpha$ is large we have

$$\overline{(A(t) - \bar{A})^2} \approx w_0^2 \frac{f_b - f_a}{2\alpha}$$

and the relative r.m.s. fluctuation is

$$\text{r.m.s. of } \left[\frac{(A(t) - \bar{A})}{\bar{A}} \right] \approx \left[\frac{\alpha}{2(f_b - f_a)} \right]^{1/2}$$

This result may also be obtained from (3.9–10) and (3.9–11) by assuming α so small that the integral for $A(t)$ may be broken into a great many integrals each extending over an interval T. αT is assumed so small that $e^{-\alpha u}$ is substantially constant over each interval.

3.10 Distribution of Noise Plus Sine Wave

Suppose we have a steady sinusoidal current

$$I_p = I_p(t) = P \cos(\omega_p t - \varphi_p) \tag{3.10–1}$$

We pick times $t_1, t_2, \cdots$ at random and note the corresponding values of the current. How are these values distributed? Picking the times at random in (3.10–1) is the same, statistically, as holding t constant and picking the phase angles φ_p at random from the range 0 to 2π. If I_p be regarded as a random variable defined by the random variable φ_p, its characteristic function is

$$\begin{aligned} \text{ave. } e^{izI_p} &= \frac{1}{2\pi} \int_0^{2\pi} e^{izP \cos(\omega_p t - \varphi)} \, d\varphi \\ &= J_0(Pz) \end{aligned} \tag{3.10–2}$$

and its probability density is

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-izI_p} J_0(Pz) \, dz = \begin{cases} \frac{1}{\pi}(P^2 - I_p^2)^{-1/2} & |I_p| < P \\ 0 & |I_p| > P \end{cases} \tag{3.10–3}$$

In this case it is simpler to obtain the probability density directly from (3.10–1) instead of from the characteristic function.

Now suppose that we have a noise current I_N plus a sine wave. By combining our representation (2.8–6) for I_N with the idea of φ_p being random mentioned above we are led to the representation

$$\begin{aligned} I(t) &= I = I_p + I_N \\ &= P\cos(\omega_p t - \varphi_p) + \sum_1^M c_n \cos(\omega_n t - \varphi_n), \qquad (3.10\text{–}4) \\ c_n^2 &= 2w(f_n)\Delta f \end{aligned}$$

where φ_p and $\varphi_1, \cdots \varphi_M$ are independent random angles.

If we note I at the random times $t_1, t_2 \cdots$ how are the observed values distributed? Since I_p and I_N may be regarded as independent random variables and since the characteristic function for the sum of two such variables is the product of their characteristic functions we have from (3.1–6) and (3.10–2)

$$\begin{aligned} \text{ave. } e^{izI} &= \text{ave. } e^{iz(I_p+I_N)} \\ &= J_0(Pz)\exp\left(\frac{-\psi_0 z^2}{2}\right) \end{aligned} \qquad (3.10\text{–}5)$$

which gives the characteristic function of I. The probability density of I is[37]

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-izI-(\psi_0 z^2/2)} J_0(Pz)\,dz = \frac{1}{\pi\sqrt{2\pi\psi_0}}\int_0^{\pi} e^{-(I-P\cos\theta)^2/2\psi_0}\,d\theta \qquad (3.10\text{–}6)$$

In the same way the two-dimensional probability density of (I_1, I_2), where $I_1 = I(t)$ is a sine wave plus noise (3.10–4) and $I_2 = I(t+\tau)$ is its value at a constant interval τ later, may be shown to be

$$\frac{(\psi_0^2 - \psi_\tau^2)^{-1/2}}{2\pi}\int_0^{2\pi} d\theta \exp\left[-\frac{B(\theta)}{2(\psi_0^2 - \psi_\tau^2)}\right] \qquad (3.10\text{–}7)$$

where

$$\begin{aligned} B(\theta) = \psi_0[(I_1 - P\cos\theta)^2 &+ (I_2 - P\cos(\theta + \omega_p\tau))^2] \\ &- 2\psi_\tau(I_1 - P\cos\theta)(I_2 - P\cos(\theta + \omega_p\tau)) \end{aligned}$$

The characteristic function for I_1 and I_2 is

$$\begin{aligned} \text{ave. } e^{iuI_1+ivI_2} &= J_0\left(P\sqrt{u^2 + v^2 + 2uv\cos\omega_p\tau}\right) \\ &\times \exp\left[-\frac{\psi_0}{2}(u^2 + v^2) - \psi_\tau uv\right] \end{aligned} \qquad (3.10\text{–}8)$$

[37] A different derivation of this expression is given by W. R. Bennett, *Jour. Acous. Soc. Amer.*, Vol. 15, p. 165 (Jan. 1944); *B.S.T.J.*, Vol. 23, p. 97 (Jan. 1944).

Sometimes the distribution of the envelope of

$$I = P \cos pt + I_N \tag{3.10-9}$$

is of interest. Here we have replaced ω_p by p and have set φ_p to zero. By the envelope we mean $R(t)$ given by

$$R^2(t) = R^2 = (P + I_c)^2 + I_s^2 \tag{3.10-10}$$

where I_c is the component of I_N "in phase" with cos pt and I_s is the component "in phase" with sin pt:

$$I_c = \sum c_n \cos [(\omega_n - p)t - \varphi_n]$$

$$I_s = \sum c_n \sin [(\omega_n - p)t - \varphi_n]$$

$$I_N = I_c \cos pt - I_s \sin pt$$

$$\overline{I_N^2} = \overline{I_c^2} = \overline{I_s^2} = \psi_0$$

Since I_c and I_s are distributed normally about zero with a variance of ψ_0, the probability densities of the variables

$$x = P + I_c$$

$$y = I_s$$

are

$$(2\pi\psi_0)^{-1/2} \exp - \frac{(x - P)^2}{2\psi_0}$$

$$(2\pi\psi_0)^{-1/2} \exp - \frac{y^2}{2\psi_0}$$

respectively. Setting

$$x = R \cos \theta$$

$$y = R \sin \theta$$

and using these distributions shows that the probability of a point (x, y) lying in the ring R, $R + dR$ is

$$\frac{R\,dR}{2\pi\psi_0} \int_0^{2\pi} \exp \left[-\frac{1}{2\psi_0} (R^2 + P^2 - 2RP \cos \theta) \right] d\theta$$

$$= \frac{R\,dR}{\psi_0} \exp \left[-\frac{R^2 + P^2}{2\psi_0} \right] I_0 \left(\frac{RP}{\psi_0} \right) \tag{3.10-11}$$

where I_0 is the Bessel function with imaginary argument.

$$I_0(z) = \sum_{n=0}^{\infty} \frac{z^{2n}}{2^{2n} n!\, n!}$$

and is a tabulated function. Thus (3.10-11) gives the probability density of the envelope R.

The average value of R^n may be obtained by multiplying (3.10-11) by R^n and integrating from 0 to ∞. Expansion of the Bessel function and term-wise integration gives

$$\overline{R^n} = (2\psi_0)^{n/2}\,\Gamma\left(\frac{n}{2}+1\right) e^{-P^2/2\psi_0}\,{}_1F_1\left(\frac{n}{2}+1;\,1;\,\frac{P^2}{2\psi_0}\right)$$

$$= (2\psi_0)^{n/2}\,\Gamma\left(\frac{n}{2}+1\right){}_1F_1\left(-\frac{n}{2};\,1;\,-\frac{P^2}{2\psi_0}\right) \qquad (3.10\text{-}12)$$

where ${}_1F_1$ is a hypergeometric function.[38] In going from the first line to the second we have used Kummer's first transformation of this function. A special case is

$$\overline{R^2} = P^2 + 2\psi_0 \qquad (3.10\text{-}13)$$

When only noise is present, $P = 0$ and

$$\overline{R} = (2\psi_0)^{1/2}\,\Gamma(\tfrac{3}{2}) = \left(\frac{\psi_0\pi}{2}\right)^{1/2}$$
$$\overline{R^2} = 2\psi_0 \qquad (3.10\text{-}14)$$

Before going further with (3.10-11) it is convenient to make the following change of notation

$$v = \frac{R}{\psi_0^{1/2}}, \qquad dv = \frac{dR}{\psi_0^{1/2}}, \qquad a = \frac{P}{\psi_0^{1/2}} \qquad (3.10\text{-}15)$$

"a" is the ratio (sine wave amplitude)/(r.m.s. noise current).
Instead of the random variable R we now have the random variable v whose probability density is

$$p(v) = v\exp\left[-\frac{v^2+a^2}{2}\right] I_0(av) \qquad (3.10\text{-}16)$$

Curves of $p(v)$ versus v are plotted in Fig. 6 for the values 0, 1, 2, 3, 5 of a. Curves showing the probability that v is less than a stated amount, i.e., distribution curves for v, are given in Fig. 7. These curves were obtained by integrating $p(v)$ numerically. The following useful expression for this probability has been given by W. R. Bennett in some unpublished work.

$$\int_0^v p(u)\,du = \exp\left[-\frac{v^2+a^2}{2}\right]\sum_{n=1}^{\infty}\left(\frac{v}{a}\right)^n I_n(av) \qquad (3.10\text{-}17)$$

[38] Curves of this function are given in "Tables of Functions", Jahnke and Emde (1938), p. 275, and some of its properties are stated in Appendix 4C.

This is obtained by integration by parts using

$$\int u^n I_{n-1}(au)\, du = u^n I_n(au)/a$$

When $av >> 1$ but $1 << a - v$, Bennett has shown that (3.10–17) leads to

$$\int_0^v p(u)\, du \approx \left(\frac{v}{2\pi a}\right)^{1/2} \frac{1}{a - v} \exp\left[-\frac{(v - a)^2}{2}\right] \left(1 - \frac{3(a + v)^2 - 4v^2}{8av(a - v)^2} \cdots\right) \tag{3.10–18}$$

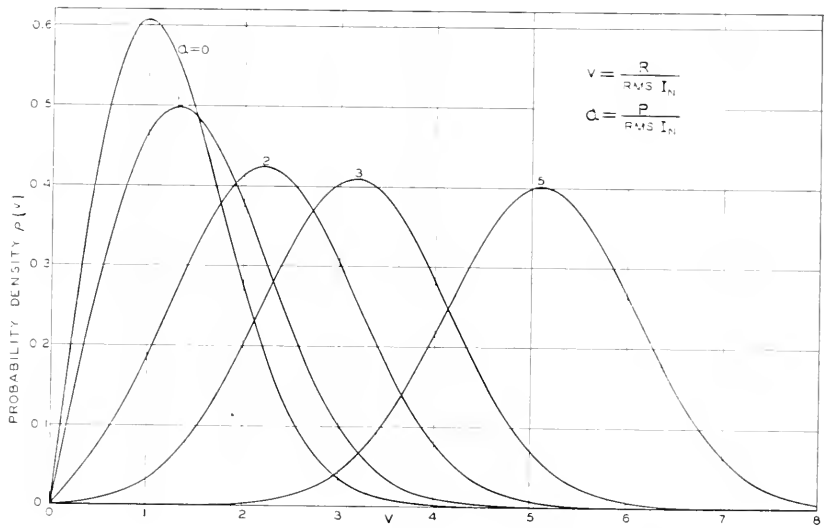


Fig. 6—Probability density of envelope R of $I(t) = P \cos pt + I_N$

This formula may also be obtained by putting the asymptotic expansion (3.10–19) for $p(v)$ in (3.10–17), integrating by parts twice, and neglecting higher order terms.

When av becomes large we may replace $I_0(av)$ by its asymptotic expression. The expression for $p(v)$ is then

$$p(v) \sim \left(1 + \frac{1}{8av}\right)\left(\frac{v}{2\pi a}\right)^{1/2} \exp\left[-\frac{(v - a)^2}{2}\right] \tag{3.10–19}$$

Thus when either a becomes large or v is far out on the tail of the probability density curve, the distribution behaves like a normal law. In terms of the original quantities, the normal law has an average of P and a standard deviation of $\psi_0^{1/2}$. This standard deviation is the same as the standard deviation

of the instantaneous values of I_N. When $av \gg 1$ and $a \gg v - a$ we may expand the coefficient of the exponential term in (3.10-19) in powers of

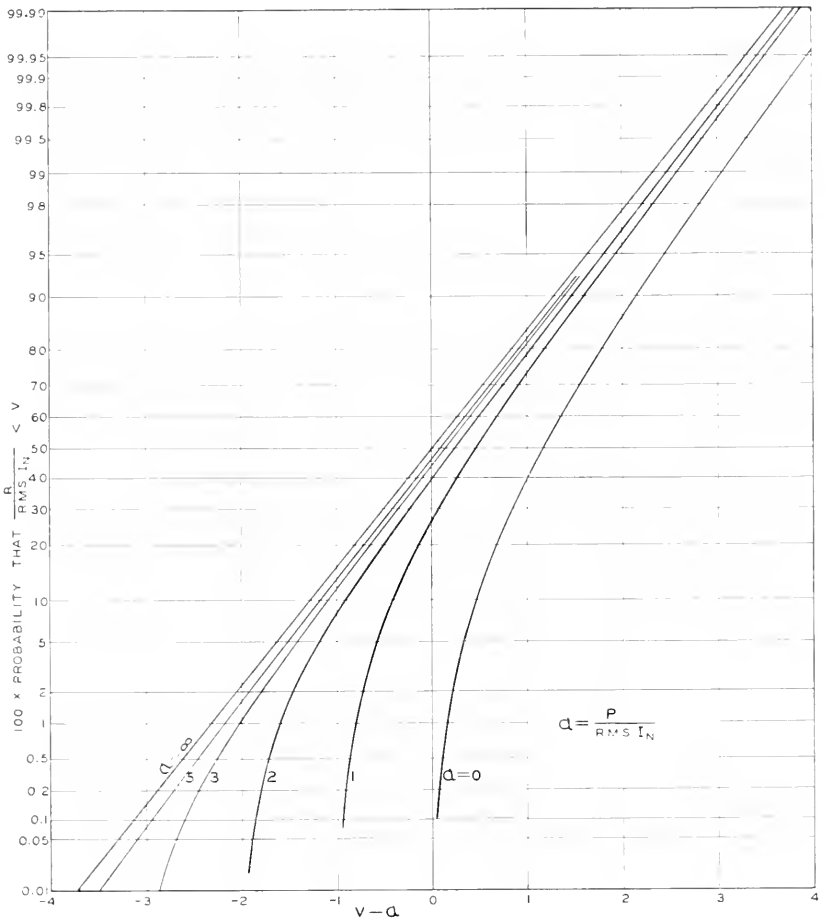


Fig. 7—Distribution function of envelope R of $I(t) = P \cos pt + I_N$

$(v - a)\sqrt{a}$. Integrating this expansion termwise gives, when terms of magnitude less than a^{-3} are neglected,

$$\int_0^v p(u)\,du \approx \frac{1}{2} + \frac{1}{2}\,erf\,\frac{v-a}{\sqrt{2}} - \frac{1}{2a\sqrt{2\pi}}\left[1 - \frac{v-a}{4a} + \frac{1+(v-a)^2}{8a^2}\right]\exp\left[-\frac{(v-a)^2}{2}\right]$$

When I consists of two sine waves plus noise

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

where the radian frequencies p and q are incommensurable, 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 ψ_0 is $\overline{I_N^2}$. When Q is zero the integral may be evaluated to give (3.10–11). When both P and Q are zero the probability density for R when only noise is present is obtained. If there are three sine waves instead of two then another Bessel function must be placed in the integrand, and so on. To define R it is convenient to think of the noise as being confined to a relatively narrow band and the frequencies of the sine waves lying within, or close to, this band. As in equations (3.7–2) to (3.7–4), we refer all terms to a representative mid-band frequency $f_m = \omega_m/2\pi$ by using equations of the type

$$\begin{aligned}\cos pt &= \cos\left[(p-\omega_m)t + \omega_m t\right]\\ &= \cos(p-\omega_m)t\cos\omega_m t - \sin(p-\omega_m)t\sin\omega_m t.\end{aligned}$$

In this way we obtain

$$V = A\cos\omega_m t - B\sin\omega_m t = R\cos(\omega_m t + \theta) \tag{3.10–22}$$

where A and B are relatively slowly varying functions of t given by

$$\begin{aligned}A &= P\cos(p-\omega_m)t + Q\cos(q-\omega_m)t\\ &\qquad + \sum_n c_n\cos(\omega_n t - \omega_m t - \varphi_n)\\ B &= P\sin(p-\omega_m)t + Q\sin(q-\omega_m)t\\ &\qquad + \sum_n c_n\sin(\omega_n t - \omega_m t - \varphi_n)\end{aligned} \tag{3.10–23}$$

and

$$\begin{aligned}R^2 &= A^2 + B^2, \qquad R > 0\\ \tan\theta &= B/A\end{aligned} \tag{3.10–24}$$

As might be expected, (3.10–21) is closely associated with the problem of random flights and may be obtained from Kluyver's result[39] by assuming

[39] G. N. Watson, "Theory of Bessel Functions" (Cambridge, 1922), p. 420.

the noise to correspond to a very large number of very small random displacements.

Another way of deriving (3.10-21) is to assume $(p - \omega_m)t$, $(q - \omega_m)t$, φ_1, φ_2, $\cdots$ are independent random angles. The characteristic function of A, B is

$$\text{ave. } e^{iuA+ivB} = J_0(P\sqrt{u^2 + v^2})J_0(Q\sqrt{u^2 + v^2})e^{-(\psi_0/2)(u^2+v^2)}$$

The probability density of A, B is

$$\left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} dv\, e^{-iuA-ivB} \text{ ave. } e^{iuA+ivB}$$

When the change of variables

$$A = R\cos\theta \qquad u = r\cos\varphi$$

$$B = R\sin\theta \qquad v = r\sin\varphi$$

is made the integration with respect to φ may be performed. The double integral becomes

$$\frac{1}{2\pi}\int_0^\infty rJ_0(Pr)J_0(Qr)J_0(Rr)e^{-(\psi_0/2)r^2}\,dr$$

This leads directly to (3.10-21) when we observe that $dA\,dB = R\,dR\,d\theta$.

Incidentally, if

$$I = Q(1 + k\cos pt)\cos qt + I_N$$

in which $p << q$, similar considerations show that the probability density of R is

$$\frac{R}{2\pi}\int_0^{2\pi} d\alpha \int_0^\infty rJ_0(Rr)J_0[Qr(1 + k\cos\alpha)]e^{-(\psi_0/2)r^2}\,dr$$

when ω_m is taken to be q. The integration with respect to r may be performed. This relation is closely connected with (3.10-11).

Returning now to the case in which I is the sum of two sine waves plus noise, we may show from (3.10-21) and

$$\int_0^\infty R^{n+1}J_0(Rr)\,dR = \frac{2^{n+1}\Gamma\left(1 + \frac{n}{2}\right)}{r^{n+2}\Gamma\left(-\frac{n}{2}\right)}$$

that the average value of R^n is, when $-2 < re\ (n) < -\frac{3}{2}$,

$$\overline{R^n} = \frac{2^{n+1}\Gamma\left(1 + \frac{n}{2}\right)}{\Gamma\left(-\frac{n}{2}\right)} \int_0^\infty r^{-n-1} J_0(Pr) J_0(Qr) e^{-\psi_0 r^2/2}\, dr$$

$$= (2\psi_0)^{n/2}\Gamma\left(\frac{n}{2} + 1\right) \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{\left(-\frac{n}{2}\right)_{k+m} (-x)^k (-y)^m}{k!\,k!\,m!\,m!} \tag{3.10-25}$$

$$= (2\psi_0)^{n/2}\Gamma\left(\frac{n}{2} + 1\right) \sum_{k=0}^{\infty} \frac{\left(-\frac{n}{2}\right)_k (y - x)^k}{k!\,k!} P_k\left(\frac{x + y}{x - y}\right)$$

It appears very probable that this result could be extended, by analytic continuation, to positive integer values of n. We have used the notation

$$(\alpha)_0 = 1, \qquad (\alpha)_k = \alpha(\alpha + 1) \cdots (\alpha + k - 1)$$
$$x = \frac{P^2}{2\psi_0}, \qquad y = \frac{Q^2}{2\psi_0} \tag{3.10-26}$$

and have denoted the Legendre polynomial by $P_k(z)$. The series converge for all values of P, Q, and ψ_0 and terminate when n is an even positive integer.

When x or y, or both, are large in comparison with unity we may use the integral for $\overline{R^n}$ to obtain the asymptotic expansion, assuming $Q < P$ so that $y < x$,

$$\overline{R^n} \sim P^n \sum_{k=0}^{\infty} \frac{\left(-\frac{n}{2}\right)_k \left(-\frac{n}{2}\right)_k}{k!\,x^k} {}_2F_1\left(k - \frac{n}{2},\ k - \frac{n}{2};\ 1;\ \frac{y}{x}\right) \tag{3.10-27}$$

When n is an even positive integer this series terminates and gives the same expression as (3.10-25). When n is an odd integer the ${}_2F_1$ may be expressed in terms of the complete elliptic functions E and K of modulus $y^{1/2}x^{-1/2}$:

$${}_2F_1\left(-\tfrac{1}{2},\ -\tfrac{1}{2};\ 1;\ \frac{y}{x}\right) = \frac{4}{\pi} E - \frac{2}{\pi}\left(1 - \frac{y}{x}\right) K$$
$${}_2F_1\left(\tfrac{1}{2},\ \tfrac{1}{2};\ 1;\ \frac{y}{x}\right) = \frac{2}{\pi} K \tag{3.10-28}$$

The higher terms may be computed from

$$a(1 - z)^2 {}_2F_1(a + 1, a + 1; 1; z) = (2a - 1)(1 + z){}_2F_1(a, a; 1; z)$$
$$+ (1 - a){}_2F_1(a - 1, a - 1; 1; z) \tag{3.10-29}$$

which is a special case of

$$ab(\gamma + 1)(1 - z)^2 {}_2F_1(a + 1, b + 1; c; z) = A\,{}_2F_1(a, b; c; z)$$
$$- (\gamma - 1)(c - a)(c - b)\,{}_2F_1(a - 1, b - 1; c; z) \quad (3.10\text{–}30)$$

where $\gamma = c - a - b$ and

$$A = (\gamma^2 - 1)\gamma + (1 - z)[(\gamma - 1)(c - b)(b - 1) + (\gamma + 1)a(c - a - 1)]$$

Although this expression does not show it, A is really symmetrical in a and b. A symmetrical form may be obtained by using the expression obtained by putting $z = 0$ in (3.10–30).

3.11 Shot Effect Representation

In most of the work in this part the representations (2.8–1) or (2.8–6) have been used as a starting point. Here we point out that the shot effect representation used in Part I may also be used as a starting point.

For example, suppose we wish to find the two dimensional distribution of $I(t)$ and $I(t + \tau)$ discussed in Section 3.2. This is a special case of the distribution of the two variables

$$\begin{aligned} I(t) &= \sum_{k=-\infty}^{+\infty} F(t - t_k) \\ J(t) &= \sum_{k=-\infty}^{+\infty} G(t - t_k) \end{aligned} \quad (3.11\text{–}1)$$

where we now assume

$$\int_{-\infty}^{+\infty} F(t)\,dt = \int_{-\infty}^{+\infty} G(t)\,dt = 0 \quad (3.11\text{–}2)$$

in order that the average values of I and J may be zero. In fact, to get $I(t + \tau)$ from $J(t)$ we set $G(t)$ equal to $F(t + \tau)$.

The distribution of I and J may be obtained in much the same manner as was the distribution of I alone in section 1.4. The characteristic function of the distribution is

$$\begin{aligned} f(u, v) &= \text{ave. } e^{iuI + ivJ} \\ &= \exp\ \nu \int_{-\infty}^{+\infty} [e^{iuF(t) + ivG(t)} - 1]\,dt \end{aligned} \quad (3.11\text{–}3)$$

where ν is the expected number of events (electron arrivals in the shot effect) per second. The probability density of I and J is

$$\frac{1}{4\pi^2} \int_{-\infty}^{+\infty} du \int_{-\infty}^{+\infty} dv\, e^{-iuI - ivJ} f(u, v) \quad (3.11\text{–}4)$$

The semi-invariants $\lambda_{m,n}$ are given by the generating function

$$\log f(u, v) = \sum_{m,n=1}^{k} \frac{\lambda_{m,n}}{m!\,n!} (iu)^m (iv)^n + o[(iu)^k, (iv)^k]$$

and are

$$\lambda_{m,n} = \nu \int_{-\infty}^{+\infty} F^m(t) G^n(t)\, dt \tag{3.11-5}$$

As $\nu \to \infty$ the distribution of I and J approaches a two dimensional normal law. The approximation to this normal law may be obtained in much the same manner as in section 1.6. From our assumption (3.11-2) it follows that λ_{10} and λ_{01} are zero. From the relation between the second moments and semi-invariants λ we have

$$\begin{aligned} \mu_{11} &= \lambda_{20} + \lambda_{10}^2 = \nu \int_{-\infty}^{+\infty} F^2(t)\, dt \\ \mu_{12} &= \lambda_{11} + \lambda_{10}\lambda_{01} = \nu \int_{-\infty}^{+\infty} F(t)G(t)\, dt \\ \mu_{22} &= \lambda_{02} + \lambda_{01}^2 = \nu \int_{-\infty}^{+\infty} G^2(t)\, dt \end{aligned} \tag{3.11-6}$$

where the notation in the subscripts of the μ's differs from that of the λ's, the change being made to bring it in line with sections 2.9 and 2.10 so that we may write down the normal distribution at once.

The formulas (3.11-6) are closely related to Rowland's generalization of Campbell's theorem mentioned just below equation (1.5-9).

PART IV

NOISE THROUGH NON-LINEAR DEVICES

4.0 Introduction

We shall consider two problems which concern noise passing through detectors or other non-linear devices. The first deals with the statistical properties of the output of a non-linear device, that is, with its average value, its fluctuation about this average and so on. The second problem may be stated more definitely: Given a non-linear device and an input consisting of noise alone, or of noise plus a signal. What is the power spectrum of the output?

There does not seem to be much published material on the first problem. However, from conversation with other people, I have learned that it has been studied independently by several investigators. The same is probably true of the second problem although here the published material is somewhat more plentiful. This makes it difficult to assign credit where credit is due. Much of the material given here had its origin in discussions with friends, especially with W. R. Bennett, J. H. Van Vleck, and David Middleton. Help was obtained from the recent paper[37] by Bennett, and also from the manuscript of a forthcoming paper by Middleton.[40]

4.1 Low Frequency Output of a Square Law Device

Let the output current I of the device be related to the input voltage V by

$$I = \alpha V^2 \tag{4.1-1}$$

where α is a constant. When the power spectrum of V is confined to a relatively narrow band, the power spectrum of I consists of two portions. One portion clusters around twice the mid-band frequency of V and the other around zero frequency. We are interested in the low frequency portion. The current corresponding to this portion will be denoted by $I_{t\ell}$, and is the current which would flow if a low pass filter were inserted in the output to remove the upper portion of the spectrum. It is convenient to divide $I_{t\ell}$ into two components:

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

[37] Loc. cit. (Section 3.10).

[40] Cruft Laboratory and the Research Laboratory of Physics, Harvard University, Cambridge, Mass. In the following sections references to Bennett's paper and Middleton's manuscript are made by simply giving the authors' names.

where the subscripts stand for "total low" frequency, "direct current." and "low frequency," respectively. We have

$$I_{dc} = \text{average } I_{t\ell} = \overline{I}_{t\ell} \tag{4.1-3}$$

$$\text{Mean Square } I_{\ell f} = \text{average } (I_{t\ell} - I_{dc})^2 = \overline{I_{t\ell}^2} - I_{dc}^2$$

Probably the simplest method of obtaining I_{dc} is to square the given expression for V and pick out the terms independent of time. Thus if

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

we have

$$I_{dc} = \alpha \left(\frac{P^2}{2} + \frac{Q^2}{2} + \overline{V_N^2} \right) \tag{4.1-5}$$

$I_{\ell f}$ may also be obtained by picking out the low frequency terms. However, here we wish to use the square law device, and the linear rectifier in the next section, to illustrate a general method of dealing with the statistical properties of the output of a non-linear device when the input voltage is restricted to a relatively narrow band.

If none of the low frequency spectrum is removed by filters,

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

where R is the envelope of V. The probability density and the statistical properties of $I_{t\ell}$ may be derived from this relation when the distribution function of R is known.[41] Before discussing these properties we shall establish (4.1-6).

Equation (4.1-6) is a special case of a more general result established in Section 4.3. However, its truth may be seen by taking the example

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

where $f_p = p/2\pi$ and $f_q = q/2\pi$ lie within, or close to, the band of the noise voltage V_N.

By using formulas of the type

$$\begin{aligned} \cos pt &= \cos [(p - \omega_m)t + \omega_m t] \\ &= \cos (p - \omega_m)t \cos \omega_m t - \sin (p - \omega_m)t \sin \omega_m t \end{aligned} \tag{4.1-7}$$

[41] When part of the low-frequency spectrum is removed, the problem becomes much more difficult. I_{dc} may be obtained as above, but to get $\overline{I_{\ell f}^2}$ it is necessary to first determine the power spectrum of I (Section 4.5) and then integrate over the appropriate portion of it. Concerning the distribution of $I_{\ell f}$, our present knowledge tells us only that it lies between the one given by (4.1-6) and the normal law which it approaches when only a narrow portion of the low frequency spectrum is passed by the audio frequency filter (Section 4.3).

we may refer all terms to the mid-band frequency $f_m = \omega_m/2\pi$, as is done in equations (3.7–2) to (3.7–4).

In this way we obtain

$$V = A \cos \omega_m t - B \sin \omega_m t = R \cos (\omega_m t + \theta), \tag{4.1–8}$$

where A and B are relatively slowly varying functions of t given by

$$A = P \cos (p - \omega_m)t + Q \cos (q - \omega_m)t + \sum_n c_n \cos (\omega_n t - \omega_m t - \varphi_n),$$

$$B = P \sin (p - \omega_m)t + Q \sin (q - \omega_m)t + \sum_n c_n \sin (\omega_n t - \omega_m t - \varphi_n)$$

and

$$\begin{aligned} R^2 &= A^2 + B^2, \qquad R > 0 \\ \tan \theta &= B/A. \end{aligned} \tag{4.1–9}$$

This definition of R has also been given in equations (3.10–22, 23, 24).

The envelope of V is R and the output current is

$$I = \alpha R^2 \left[\frac{1}{2} + \frac{1}{2} \cos (2\omega_m t + 2\theta) \right] \tag{4.1–10}$$

Since R is a slowly varying function of time, so is R^2. The power spectrum of R^2 is confined to frequencies much lower than $2f_m$ and consequently the power spectrum of $R^2 \cos (2\omega_m t + 2\theta)$ is clustered around $2f_m$. Thus the only term in I contributing to the low frequency output is $\alpha R^2/2$ which is what we wished to show.

We now return to the statistical properties of $I_{t\ell}$. First, consider the case in which V consists of noise only, $V = V_N$, so that the probability density of the envelope R is

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

where

$$\psi_0 = [rms\ V_N]^2 = \overline{V_N^2} \tag{4.1–11}$$

Hence

$$\begin{aligned} I_{dc} = \overline{I}_{t\ell} &= \frac{\alpha \overline{R^2}}{2} \\ &= \int_0^\infty \frac{\alpha R^2}{2} \frac{R}{\psi_0} e^{-R^2/2\psi_0}\, dR \\ &= \alpha \psi_0 \\ \overline{I_{fl}^2} = \overline{I_{t\ell}^2} - I_{dc}^2 &= \int_0^\infty \frac{\alpha^2 R^5}{4\psi_0} e^{-R^2/2\psi_0}\, dR - I_{dc}^2 \\ &= \alpha^2 \psi_0^2 \end{aligned} \tag{4.1–12}$$

Second, consider the case in which

$$V = V_N + P \cos pt \tag{4.1-13}$$

where $p/2\pi$ lies near the noise band of V_N. The probability density of the envelope R is

$$\frac{R}{\psi_0} \exp\left[-\frac{R^2 + p^2}{2\psi_0}\right] I_0\left(\frac{RP}{\psi_0}\right) \tag{3.10-11}$$

From this and equations (3.10–12), (3.10–13), we find

$$I_{dc} = \frac{\alpha \overline{R^2}}{2} = \alpha\psi_0 + \frac{\alpha P^2}{2} \tag{4.1-14}$$

$$\overline{I_{i\ell}^2} = \frac{\alpha^2}{4}\overline{R^4} = \alpha^2\left[2\psi_0^2 + 2P^2\psi_0 + \frac{P^4}{4}\right]$$

$$\overline{I_{\ell f}^2} = \overline{I_{i\ell}^2} - I_{dc}^2 = \alpha^2[\psi_0 + P^2]\psi_0 \tag{4.1-15}$$

In (4.1–14) ψ_0 is the mean square value of V_N and $P^2/2$ is the mean square value of the signal. These two equations show that I_{dc} and the rms value of $I_{\ell f}$ are independent of the distribution of the noise power spectrum in V_N as long as the input V is confined to a relatively narrow band. In other words, although this distribution does affect the power spectrum of the output, it does not affect the d.c. and rms $I_{\ell f}$ when ψ_0 and P are given. That the same is also true for a large class of non-linear devices was first pointed out by Middleton (see end of Section 4.9).

When the voltage is[42]

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

$p \neq q$, we obtain from equation (3.10–25)

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

$$\overline{I_{i\ell}^2} = \frac{\alpha^2}{4}\overline{R^4} \tag{4.1-16}$$

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

[42] These results are special cases, obtained by assuming no audio frequency filter, of formulas given by F. C. Williams, *Jour. Inst. of E. E.*, 80 (1937), 218–226. Williams also discusses the response of a linear rectifier to (4.1–4) when $P \gg Q + V_N$. An account of Williams' work is given by E. B. Moullin, "Spontaneous Fluctuations of Voltage," Oxford (1938), Chap. 7.

4.2 Low Frequency Output of a Linear Rectifier

In the case of the linear rectifier

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

the low frequency output current, assuming no audio frequency filter, is

$$I_{l\ell} = \frac{\alpha R}{\pi} \tag{4.2-2}$$

This formula, like its analogue (4.1–6) for the square law device, assumes that the applied signal and noise lie within a relatively narrow band. It may be used to compute the probability density and statistical properties of $I_{l\ell}$ when the corresponding information regarding the envelope R of the applied voltage is known.

The truth of (4.2–2) may be seen by considering the output I. It consists of the positive halves of the oscillations of αV. The envelope of I is the same as that of αV. However, the area under the loops of I is only about $1/\pi$ of the area under αR, this being the ratio of the area under a loop of $\sin x$ to the area of a rectangle of unit height and length 2π. From the low frequency point of view these loops of I merge into a current which varies as $\alpha R/\pi$.

When V is a sine wave plus noise,

$$V = V_N + P \cos pt \tag{4.1-13}$$

the average value of $I_{l\ell}$ is[43]

$$\begin{aligned} I_{dc} = \frac{\alpha}{\pi}\overline{R} &= \alpha\left(\frac{\psi_0}{2\pi}\right)^{1/2} {}_1F_1\left(-\frac{1}{2};\, 1;\, -\frac{P^2}{2\psi_0}\right) \\ &= \alpha\left(\frac{\psi_0}{2\pi}\right)^{1/2} e^{-x/2}\left[(1+x)I_0\left(\frac{x}{2}\right) + xI_1\left(\frac{x}{2}\right)\right] \end{aligned} \tag{4.2-3}$$

where I_0, I_1 are Bessel functions of imaginary argument and

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

[43] This result was discovered independently by several investigators, among whom we may mention W. R. Bennett and D. O. North. The latter has applied it to noise measurement work. He has found that the diode detector, when adapted to noise metering, is a great improvement over the thermocouple, and has used noise meters of this type satisfactorily since 1940. See D. O. North, "The Modification of Noise by Certain Non-Linear Devices", Paper read before I.R.E., Jan. 28, 1944.

ψ_0 being the average value of V_N^2. Equation (4.2–3) follows from the formulas (3.10–12) and (4B–9). When x is large the asymptotic expansion (4B–3) of the ${}_1F_1$ gives

$$I_{dc} \sim \frac{\alpha}{\pi}\left[P + \frac{\psi_0}{2P} + \frac{\psi_0^2}{8P^3} + \cdots\right] \tag{4.2–5}$$

Similarly, the mean square value of $I_{\ell\ell}$ is

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

and the mean square value of the low frequency current $I_{\ell f}$, excluding the d.c., is given by

$$\overline{I_{\ell f}^2} = \overline{I_{\ell\ell}^2} - I_{dc}^2$$

When x is large we have

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

and when $x = 0$,

$$\overline{I_{\ell f}^2} = \frac{\alpha^2}{\pi^2}\psi_0\left(2 - \frac{\pi}{2}\right) \tag{4.2–8}$$

Curves for I_{dc} are given in Figures 1, 2 and 3 of Bennett's paper. He also gives curves, in Fig. 4, showing $\overline{I_{\ell f}^2}$ versus x. These show that the effect of the higher order modulation terms is small when $I_{\ell f}$ is computed by adding low frequency modulation products.

When V consists of two sine waves plus noise,

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

the average value of $I_{\ell\ell}$ is, from (3.10–25), a sort of double ${}_1F_1$ function:

$$\begin{aligned} I_{dc} &= \frac{\alpha}{\pi}\overline{R} = \alpha\left(\frac{\psi_0}{2\pi}\right)^{1/2}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-\tfrac{1}{2})_{k+m}}{k!\,k!\,m!\,m!}(-x)^k(-y)^m \\ &= \alpha\left(\frac{\psi_0}{2\pi}\right)^{1/2}\sum_{k=0}^{\infty}\frac{(-\tfrac{1}{2})_k}{k!\,k!}(y-x)^k P_k\left(\frac{x+y}{x-y}\right) \end{aligned} \tag{4.2–9}$$

where

$$x = \frac{P^2}{2\psi_0}, \qquad y = \frac{Q^2}{2\psi_0}, \qquad P_k(z) = \text{Legendre polynomial} \tag{4.2–10}$$

If x is large and $y < x$, we have from (3.10–27) the asymptotic expression

$$I_{dc} \sim \frac{\alpha}{\pi}P\sum_{k=0}^{\infty}\frac{(-\tfrac{1}{2})_k(-\tfrac{1}{2})_k}{k!\,x^k}\,{}_2F_1\left(k - \tfrac{1}{2}, k - \tfrac{1}{2}; 1; \frac{y}{x}\right) \tag{4.2–11}$$

The ${}_2F_1$ may be expressed in terms of the complete elliptic functions E and K of modulus $y^{1/2}x^{-1/2}$. Thus

$$
\begin{aligned}
{}_2F_1\left(-\tfrac{1}{2}, -\tfrac{1}{2}; 1; \frac{y}{x}\right) &= \frac{4}{\pi} E - \frac{2}{\pi}\left(1 - \frac{y}{x}\right) K, \\
{}_2F_1\left(\tfrac{1}{2}, \tfrac{1}{2}; 1; \frac{y}{x}\right) &= \frac{2}{\pi} K
\end{aligned}
\tag{3.10-28}
$$

and the higher terms may be computed from the recurrence relation (3.10–29). The first term, $k = 0$, in (4.2–11) gives I_{dc} when the noise is absent.[44]

The mean square value of $I_{t\ell}$ is

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

From this expression and our expression for I_{dc}, the rms value of the low frequency current, $I_{\ell f}$, excluding the d.c., may be computed. For example, when the noise is small,

$$
\begin{aligned}
\overline{I_{\ell f}^2} \sim \frac{\alpha^2}{\pi^2}\Bigg[& P^2 + Q^2 - \left(P\,{}_2F_1\left(-\tfrac{1}{2}, -\tfrac{1}{2}; 1; \frac{y}{x}\right)\right)^2 \\
& + 2\psi_0\left(1 - {}_2F_1\left(-\tfrac{1}{2}, -\tfrac{1}{2}; 1; \frac{y}{x}\right)\frac{K}{\pi}\right)\Bigg]
\end{aligned}
\tag{4.2-15}
$$

The term independent of ψ_0 gives the mean square low frequency current in the absence of noise. As Q goes to zero (4.2–15) approaches the leading term in (4.2–7), as it should. When $P = Q$ our formula breaks down and it appears that we need the asymptotic behavior of[45]

$$
I_{dc} = \alpha\left(\frac{\psi_0}{2\pi}\right)^{1/2} \sum_{k=0}^{\infty} \frac{(-\tfrac{1}{2})_k (2k)!}{[k!]^4} (-x)^k
$$

In view of the questionable nature of the derivation given in Section 3.10 of equations (4.2–9) and (4.2–11) it was thought that a numerical check on their equivalence would be worth while. Accordingly, the values $x = 4$, $y = 3$ were used in the second series of (4.2–9). It was found that the largest term (about 130) in the summation occurred at $k = 11$. In all, 24 terms were taken. The result obtained was

$$
\frac{\overline{R}}{\sqrt{2\psi_0}} = 2.5502
$$

[44] See W. R. Bennett, *B.S.T.J.*, Vol. 12 (1933), 228–243.

[45] This may be done by the method given by W. B. Ford, Asymptotic Developments, Univ. of Mich. Press (1936), Chap. VI.

For the same values of x and y the asymptotic series (4.2–11) gave

$$2.40 + 0.171 + .075 + 0.52 + \cdots$$

If we stop just before the smallest term we get 2.57 for the sum. If we include the smallest term we get 2.65. This agreement indicates that (4.2–11) is actually the asymptotic expansion of (4.2–9).

When the voltage is of the form

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

we may use

$$\overline{R^n} = (2\psi_0)^{n/2} \Gamma\left(1 + \frac{n}{2}\right) \frac{1}{2\pi} \int_0^{2\pi} {}_1F_1\left[-\frac{n}{2}; 1; -y(1 + k \cos\theta)^2\right] d\theta \tag{4.2–16}$$

where R is the envelope with respect to the frequency $q/2\pi$ and y is given by (4.2–10). The integral may be evaluated by writing ${}_1F_1$ as a power series and integrating termwise using the result

$$\frac{1}{2\pi} \int_0^{2\pi} (1 + k \cos\theta)^\ell \cos m\theta \, d\theta = \frac{(-\ell)_m}{2^m m!} (-k)^m {}_2F_1\left[\frac{m - \ell}{2}, \frac{m - \ell + 1}{2}; m + 1; k^2\right] \tag{4.2–17}$$

where m is a non-negative integer, ℓ any number,

$$(\alpha)_m = \alpha(\alpha + 1) \cdots (\alpha + m - 1), \qquad (\alpha)_0 = 1, \qquad \text{and} \qquad (0)_0 = 1.$$

The integral may also be evaluated in terms of the associated Legendre function.

By applying the methods of Section 3.10 to (4.2–16) we are led to

$$\overline{R^2} = Q^2\left(1 + \frac{k^2}{2}\right) + 2\psi_0$$
$$\overline{R} \sim Q \sum_{s=0}^{\infty} \frac{(-\tfrac{1}{2})_s (-\tfrac{1}{2})_s}{s!\, y^s} {}_2F_1(s - \tfrac{1}{2}, s; 1; k^2) \tag{4.2–18}$$

where the asymptotic series holds when y is very large and k is not too close to unity. These expressions give

$$\overline{I_{lf}^2} \sim \frac{\alpha^2}{\pi^2}\left(Q^2 \frac{k^2}{2} + \psi_0[2 - (1 - k^2)^{-1/2}] + \cdots\right) \tag{4.2–19}$$

The reader might be tempted to associate the coefficient of ψ_0 in (4.2–19) with the continuous portion of the output power spectrum. However, this would not be correct. It appears that the principal contribution of the continuous portion of the power spectrum to $\overline{I_{\ell f}^2}$ is $\alpha^2\psi_0/\pi^2$, just as in (4.2–7) when k is zero. The difference between this and the corresponding term in (4.2–19) seems to arise from the fact that the amplitude of the recovered signal is not exactly $\alpha Qk/\pi$ but is modified by the presence of the noise. This general type of behavior might be expected on physical grounds since changing P, say doubling it, in (4.2–7) does not appreciably affect the $\overline{I_{\ell f}^2}$ in (4.2–7) (which is due entirely to the continuous portion of the noise spectrum). The modulating wave may be regarded as slowly making changes of this sort in P.

4.3 Some Statistical Properties of the Output of a General Non-Linear Device

Our general problem is this: Given a non-linear device whose output I is related to its input V by the relation

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

which is discussed in Appendix 4A. Let the input V contain noise in addition to the signal. Choose some frequency band in the output for study. What are the statistical properties of the current flowing in this band?

It seems to be difficult to handle this general problem. However, it appears that the two following results are true.

1. As the output band is chosen narrower and narrower the statistical properties of the corresponding current approach those of the random noise current discussed in Part III (provided no signal harmonic lies within the band). In particular, the instantaneous current values are distributed normally.

2. When the input V is confined to a relatively narrow band the power spectrum of the output I is clustered around the 0^{th} (d.c.), 1st, 2nd, etc. harmonics of the midband frequency of V. The low frequency output including the d.c. is

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

where R is the envelope of V.

The envelope of the nth harmonic of the output, when $n > 0$, is

$$A_n(R) = \frac{1}{\pi}\int_C F(iu)J_n(uR)\,du \tag{4.3–1}$$

The mathematical statement is

$$I = \sum_{n=0}^{\infty} A_n(R) \cos (n\omega_m t + n\theta) \tag{4.3-9}$$

where $f_m = \omega_m/(2\pi)$ is the representative mid-band frequency of V and θ is a relatively slowly varying phase angle. The results of Sections 4.1 and 4.2 are special cases of this.

Middleton's result that the noise power in each of the output bands (in the entire band corresponding to a given harmonic) depends only on $\overline{V_N^2} = \psi_0$ and not on the spectrum of V_N, where V_N is the noise voltage component of V, may also be obtained from (4.3-9). We note that the total power in the n^{th} band depends only on the mean square value of its envelope $A_n(R)$, and that the probability density of the envelope R of the input involves V_N only through ψ_0.

The argument we shall use in discussing the first result is not very satisfactory. It runs as follows. The output current I may be divided into two parts. One consists of sinusoidal terms due to the signal. The other consists of noise. We shall be concerned only with the latter which we shall call I_N. The correlation between two values of I_N separated by an interval of time approaches zero as the interval becomes large. Let τ be an interval long enough to ensure that the two values of I_N are substantially independent. Choose an interval of time T long enough to contain many intervals of length τ. Expand I_N as a Fourier series over this interval. We have

$$\begin{aligned} I_N &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos \frac{2\pi n t}{T} + b_n \sin \frac{2\pi n t}{T} \right] \\ a_n - ib_n &= \frac{2}{T} \int_0^T e^{-i2\pi n t/T} I_N(t)\, dt \end{aligned} \tag{4.3-2}$$

Let the band chosen for study be $f_0 - \frac{\beta}{2}$ to $f_0 + \frac{\beta}{2}$ and let

$$T\left(f_0 - \frac{\beta}{2}\right) = n_1, \qquad T\left(f_0 + \frac{\beta}{2}\right) = n_2 \tag{4.3-3}$$

where n_1 and n_2 are integers. The number of components in the band is $(n_2 - n_1)$. We suppose β is such that this is small in comparison with T/τ. The output of the band is

$$J_N = \sum_{n=n_1}^{n_2} \left[a_n \cos \frac{2\pi n}{T} t + b_n \sin \frac{2\pi n t}{T} \right] \tag{4.3-4}$$

where

$$a_n - ib_n = \frac{2}{T}\int_0^T e^{-i2\pi((n/T)-f_0)t} e^{-i2\pi f_0 t} I_N(t)\, dt$$

$$n = \frac{n_1 + n_2}{2} + n - \frac{n_1 + n_2}{2} = f_0 T + (n - f_0 T) \tag{4.3-5}$$

We choose the band so narrow that

$$n_2 - n_1 \ll T/\tau \quad \text{or} \quad \beta\tau \ll 1 \tag{4.3-6}$$

This enables us to write approximately

$$a_n - ib_n = \sum_{r=1}^{r_1} e^{-i2\pi((n/T)-f_0)r\tau} \frac{2}{T}\int_{(r-1)\tau}^{r\tau} e^{-i2\pi f_0 t} I_N(t)\, dt$$

$r_1 = T/\tau$, T being chosen to make r_1 an integer. Suppose we do this for a large number of intervals of length T. Then $I_N(t)$ will differ from interval to interval. The set of integrals for $r = 1$ gives us an array of values which we regard as defining the distribution of a complex random variable, say x_1. Similarly the set of integrals for $r = 2$ defines the distribution of a second random variable x_2, and so on to x_{r_1}. Because we have chosen τ so large that $I_N(t)$ in any one integral is practically independent of its values in the other integrals we may say that $x_1, x_2, \cdots x_{r_1}$ are independent.

We have

$$a_{n_1} - ib_{n_1} = \sum_{r=1}^{r_1} e^{-i2\pi((n_1/T)-f_0)r\tau} x_r$$

$$a_{n_1+1} - ib_{n_1+1} = \sum_{r=1}^{r_1} e^{-i2\pi((n_1+1/T)-f_0)r\tau} x_r$$

$$\vdots \qquad\qquad \vdots$$

$$a_{n_2} - ib_{n_2} = \sum_{r=1}^{r_1} e^{-i2\pi((n_2/T)-f_0)r\tau} x_r$$

and if $n_2 - n_1 \ll r_1$, as was assumed in (4.3-6), we may apply the central limit theorem to show that $a_{n_1}, b_{n_1}, a_{n_1+1}, \cdots a_{n_2}, b_{n_2}$ tend to become independent and normally distributed about zero as we let the band width $\beta \to 0$ and $T \to \infty$ (and hence $r_1 \to \infty$) in such a way as to keep $n_2 - n_1$ fixed. In this work we make use of the fact that $I_N(t)$ is such that the real and imaginary parts of $x_1, x_2, \cdots x_r$ all have the same average and standard deviation. It is convenient to assume $f_0 T$ is an integer.

Thus as the band width β approaches zero the band output J_N given by (4.3-4) may be represented in the same way, namely as (2.8-1), as was the random noise current studied in Part III. Hence J_N tends to have the

same properties as the random noise current studied there. For example, the distribution of J_N tends towards a normal law. In our discussion we had to assume that $\beta\tau \ll 1$. If the voltage V applied to the non-linear device is confined to a relatively narrow frequency band, say $f_b - f_a$, it appears that the interval τ (chosen above so that $I(t)$ and $I(t + \tau)$ are substantially independent) may be taken to be of the order of $1/(f_b - f_a)$. In this case J_N tends to behave like a random noise current if $\beta/(f_b - f_a)$ is much smaller than unity.

We now turn our attention to the second statement made at the beginning of this section. Let the applied voltage be confined to a relatively narrow band so that it may be represented by equation (4.1–8) of Section 4.1,

$$V = R \cos (\omega_m t + \theta), \qquad R \geq 0, \tag{4.1–8}$$

where $f_m = \omega_m/(2\pi)$ is some representative frequency within the band and R and θ are functions of time which vary slowly in comparison with $\cos \omega_m t$. We call R the envelope of V.

From equation (4A–1)

$$I = \frac{1}{2\pi} \int_C F(iu) e^{iuR \cos (\omega_m t + \theta)}\, du \tag{4.3–7}$$

We expand the integrand by means of

$$e^{ix \cos \varphi} = \sum_{n=0}^{\infty} \epsilon_n i^n \cos n\varphi J_n(x) \tag{4.3–8}$$

where ϵ_0 is 1 and ϵ_n is 2 when $n > 0$ and $J_n(x)$ is a Bessel function. Thus

$$I = \sum_{n=0}^{\infty} A_n(R) \cos (n\omega_m t + n\theta) \tag{4.3–9}$$

where

$$A_n(R) = \epsilon_n \frac{i^n}{2\pi} \int_C F(iu) J_n(uR)\, du \tag{4.3–10}$$

Since R is a relatively slowly varying function of time we expect the same to be true of $A_n(R)$, at least for moderately small values of n. Thus from (4.3–9) we see that the power spectrum of I will consist of a succession of bands, the n^{th} band being clustered around the frequency nf_m. If we eliminate all of the bands except the n^{th} by means of a filter we see that the output will have the envelope $A_n(R)$ when $n \geqq 1$. Taking n to be zero, shows that the low frequency output is simply

$$A_0(R) = \frac{1}{2\pi} \int_C F(iu) J_0(uR)\, du \tag{4.3–11}$$

Taking n to be one shows that the band around f_m is given by

$$\frac{A_1(R)}{R} V \tag{4.3-12}$$

The statistical properties of the low frequency output and of the envelopes of the output bands may be obtained from those of R. For example, the probability density of $A_n(R)$ is of the form

$$p(R) \Big/ \frac{dA_n(R)}{dR} \tag{4.3-13}$$

where $p(R)$ is the probability density of R. In this expression R is considered as a function of A_n.

It should be noted that we have been assuming that all of the band surrounding the harmonic frequency nf_m is taken. When we take only a portion of it, presumably the statistical properties will tend to approach those of a random noise current in accordance with the first statement made at the beginning of this section.

When we apply (4.3-11) to the square law device we have

$$F(iu) = \frac{2\alpha}{(iu)^3}$$

$$A_0(R) = -\frac{2\alpha}{2\pi i} \int^{(0+)} \frac{J_0(uR)}{u^3}\, du$$

$$= \frac{\alpha}{2} R^2$$

When we apply (4.3-11) to the linear rectifier:

$$F(iu) = -\frac{\alpha}{u^2}$$

$$A_0(R) = -\frac{\alpha}{2\pi} \int_{-\infty}^{+\infty} \frac{J_0(uR)}{u^2}\, du = \frac{\alpha R}{\pi}$$

where the path of integration passes under the origin. These two results agree with those obtained in Section 4.1 and 4.2 from simple considerations. As a final example we find the low frequency output of a biased linear rectifier in terms of the envelope R of the applied voltage. From the table of $F(iu)$ given in Appendix 4A we see that $F(iu)$ corresponding to

$$I = 0, \qquad V < B$$

$$I = V - B, \qquad V > B$$

is

$$F(iu) = -\frac{e^{-iuB}}{u^2}$$

Consequently, the low frequency output is

$$A_0(R) = -\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iuB}J_0(uR)u^{-2}\,du$$

where the path of integration is indented downwards at the origin. When $B > R$ the value of the integral is zero since then the path of integration may be closed in the lower half plane by an infinite semi-circle This value also follows at once from the physics of the problem. When $-R < B < R$ we may integrate by parts and get

$$\begin{aligned} A_0(R) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iuB}[iBJ_0(uR) + RJ_1(uR)]u^{-1}\,du \\ &= -\frac{B}{2} + \frac{1}{\pi}\int_0^{\infty} [B\sin uBJ_0(uR) + R\cos uBJ_1(uR)]u^{-1}\,du \\ &= -\frac{B}{2} + \frac{B}{\pi}\arcsin\frac{B}{R} + \frac{1}{\pi}\sqrt{R^2 - B^2} \\ &= -\frac{B}{2} + \frac{R}{\pi}F\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}; \frac{B^2}{R^2}\right), \quad -R < B < R \end{aligned} \tag{4.3-14}$$

This hypergeometric function turns up again in equation (4.7-6). Also in the range $-R < B < R$,

$$\frac{dA_0}{dR} = \frac{1}{\pi}\sqrt{1 - \frac{B^2}{R^2}}$$

When B is negative and $R < -B$, the path of integration may be closed by an infinite semicircle in the upper half plane and the value of the integral is proportional to the residue of the pole at the origin:

$$\begin{aligned} A_0(R) &= 2\pi i\left(-\frac{1}{2\pi}\right)(-iB) \\ &= -B \end{aligned}$$

Thus, to summarize, the low frequency output for our linear rectifier is, for $B > 0$, (R is always positive)

$$\begin{aligned} A_0(R) &= 0, \qquad R < B \\ A_0(R) &= -\frac{B}{2} + \frac{B}{\pi}\arcsin\frac{B}{R} + \frac{1}{\pi}\sqrt{R^2 - B^2}, \qquad B < R \end{aligned} \tag{4.3-15}$$

and for $B < 0$ it is

$$A_0(R) = |B|, \quad R < |B|$$
$$A_0(R) = +\frac{|B|}{2} + \frac{|B|}{\pi} \arcsin \frac{|B|}{R} + \frac{1}{\pi}\sqrt{R^2 - B^2}, \quad |B| < R \tag{4.3-16}$$

where the arc sines lie between 0 and $\pi/2$. $A_0(R)$ and its first derivative with respect to R are continuous.

From (4.3–15), the d.c. output current is, for $B > 0$,

$$I_{dc} = \int_B^\infty \left[-\frac{B}{2} + \frac{B}{\pi} \arcsin \frac{B}{R} + \frac{1}{\pi}\sqrt{R^2 - B^2}\right] p(R)\, dR \tag{4.3-15}$$

where $p(R)$ is the probability density of the envelope of the input V, e.g., $p(R)$ is of the form (3.7–10) for noise alone, and of the form (3.10–11) for noise plus a sine wave. Similarly, the rms value of the low frequency current I_{lf}, excluding d.c., may be computed from

$$\overline{I_{lf}^2} = \overline{I_{tl}^2} - I_{dc}^2$$

where, if $B > 0$,

$$\overline{I_{tl}^2} = \int_B^\infty \left[-\frac{B}{2} + \frac{B}{\pi} \arcsin \frac{B}{R} + \frac{1}{\pi}\sqrt{R^2 - B^2}\right]^2 p(R)\, dR \tag{4.3-16}$$

If V consists of a sine wave of amplitude P plus noise V_N, so it may be represented as (4.1–13), and if $P \gg$ rms V_N, the distribution of R is approximately normal. If, in addition, $P - B \gg$ rms $V_N > 0$, (4.3–15), (4.3–16), and (3.10–19) lead to the approximations

$$I_{dc} \approx -\frac{B}{2} + \frac{B}{\pi} \arcsin \frac{B}{P} + \frac{1}{\pi}\sqrt{P^2 - B^2} + \frac{\psi_0}{2\pi\sqrt{P^2 - B^2}}$$
$$\approx -\frac{B}{2} + \frac{P}{\pi} + \frac{B^2 + \psi_0}{2\pi P} \tag{4.3-17}$$
$$\overline{I_{lf}^2} \approx \frac{P^2 - B^2}{\pi^2 P^2}\,\psi_0$$

The second expression for I_{dc} assumes $P \gg B$. When $B = 0$, these reduce to the first terms of (4.2–5) and (4.2–7). By using a different method Middleton has obtained a more precise form of this result.

Incidentally, for a given applied voltage, $I_{dc}(+)$ for a positive bias $|B|$ is related to $I_{dc}(-)$ for a negative bias $-|B|$ by

$$I_{dc}(-) = |B| + I_{dc}(+) \tag{4.3-18}$$

Also r.m.s. $I_{lf}(+)$ is equal to r.m.s. $I_{lf}(-)$. Equation (4.3–18) follows from a physical argument based on the areas underneath a curve of I for

the two cases. Both of the above relations follow from formulas given by Middleton when V is the sum of a sine wave plus noise. They may also be derived from (4.3-15) and (4.3-16).

4.4 Output Power Spectrum

The remainder of Part IV will be concerned with methods of solving the following 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?

In some ways the answer to this problem gives us less information than the methods discussed in the first three sections. For example, beyond giving the rms value, it tells us very little about the probability density of the current corresponding to a given frequency band of the output. On the other hand, this rms value may be found (by integrating the power spectrum) for any band we choose to study. The methods described earlier depended on the input being confined to a relatively narrow band and gave information regarding only the entire band corresponding to a given harmonic (0th, 1st, 2nd, etc.) of the input. There was no way to study the output when part of a band was eliminated by filters except by obtaining the power spectrum of some function of the envelope.

At present there appear to be two general methods available for the determination of the output power spectrum each with its own advantages and disadvantages. First there is the direct method which has been used by W. R. Bennett*, F. C. Williams**, J. R. Ragazzini[46] and others. The noise is represented as the sum of a finite number of sinusoidal components. The typical modulation product is computed and the output power spectrum is obtained by considering the density and amplitude of these products. The chief advantage of this method lies in its close relation to the known theory of modulation in non-linear circuits. Generally, the lower order modulation products are the only ones which contribute significantly to the output power and when they are known, the problem is well along towards solution. The main disadvantage is the labor of counting the modulation products falling in a given interval. However, Bennett has developed a method for doing this.[47]

The fundamental idea of the second method is to obtain the correlation function for the output current. From this the output power spectrum may be obtained by Fourier's transform. The correlation function method and its variations are of more recent origin than the direct method. They have

* Cited in Section 4.0. Also much of this writer's work on interference in broad band communication systems may be carried over to noise theory without any change in the methods used.

** Cited in Section 4.1.

[46] *Proc. I.R.E.* Vol. 30, pp. 277–288 (June 1942), "The Effect of Fluctuation Voltages on the Linear Detector."

[47] *B.S.T.J.*, Vol. 19 (1940), pp. 587–610, Appendix B.

been discovered independently and at about the same time, by several workers. In a paper read before the I.R.E., Jan. 28, 1944, D. O. North described results obtained by using the correlation function. J. H. Van Vleck and D. Middleton have been using the two variations of the method which we shall describe in Sections 4.7 and 4.8, since early in 1943. A primitive form of the method of Section 4.8 had been used by A. D. Fowler and the writer in some unpublished material written in 1942. Recently, I have learned that a method similar to the one used by Fowler and myself had already been used by Kurt Fränz in 1941.[48]

The correlation function method avoids the problem of counting the modulation products. However, in some cases it becomes rather unwieldy. Probably it is best to have both methods in mind when investigating any particular problem. The direct method will be illustrated by applying it to the square law detector. Two approaches to the correlation function method will then be described and applied to examples.

4.5 Noise Through Square Law-Device

Probably the most direct method of obtaining the power spectrum $W(f)$ of I, where

$$I = \alpha V^2, \tag{4.1-1}$$

V being a noise voltage, is to square the expression

$$V = V_N = \sum_1^M c_m \cos(\omega_m t - \varphi_m) \tag{2.8-6}$$

in which c_m^2 is $2w(f_m)\Delta f$, $\omega_m = 2\pi f_m$, $f_m = m\Delta f$ and $\varphi_1, \varphi_2, \cdots \varphi_M$ are random phase angles.

Considerable simplification of the algebra results when we replace the representation (2.8-6) by

$$V_N = \frac{1}{2}\sum_{-\infty}^{\infty} c_m e^{imat - i\varphi_m} \tag{4.5-1}$$

Here we have added a term $c_0/2$ so as to not have any gaps in the summation and have introduced the definitions

$$\begin{aligned} c_{-m} &= c_m \\ \varphi_{-m} &= -\varphi_m \\ a &= 2\pi\Delta f \end{aligned} \tag{4.5-2}$$

[48] "Die Übertragung von Rauschspannung über den linearen Gleichrichter," *Hochfr. u. Elektroakyst.*, June 1941. Other articles by Fränz are (I am indebted to Dr. North for the following references) "Beitrage zur Berechnung des Verhaltnisses von Signal spannung zu Rauschspannung am Ausgang von Empfängern", *E.N.T.*, 17, 215, 1940 and 19, 285, 1942. "Die Amplituden von Geräuschspannungen", *E.N.T.*, 19, 166, 1942. The May 1944 (p. 237), issue of the Wireless Engineer contains an abstract of "The Influence of Carrier Waves on the Noise on the Far Side of Amplitude-Limiters and Linear Rectifiers" by Fränz and Vellat, *E.N.T.*, Vol. 20, pp. 183–189 (Aug. 1943).

Squaring (4.5–1) gives the double series

$$V_N^2 = \frac{1}{4} \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} c_m c_n e^{i(m+n)at - i\varphi_m - i\varphi_n}$$

$$= \frac{1}{4} \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} c_{k-n} c_n e^{ikat - i\varphi_{k-n} - i\varphi_n}$$

Suppose we wish to consider the component of V_N^2 of frequency $f_k = k\Delta f$. It is seen to be

$$A_k \cos(\omega_k t - \psi_k) = \frac{1}{2} \sum_{n=-\infty}^{+\infty} c_{k-n} c_n \cos(kat - \varphi_{k-n} - \varphi_n) \qquad (4.5\text{–}3)$$

The power spectrum $W(f)$ of I at frequency f_k is α^2 times the coefficient of Δf in the mean square value of (4.5–3) where the average is taken over the φ's. Thus

$$W(f_k)\Delta f = \frac{\alpha^2}{4} \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} c_{k-n} c_n c_{k-m} c_m$$

$$\times \text{ ave. } \cos(kat - \varphi_{k-n} - \varphi_n) \cos(kat - \varphi_{k-m} \overset{\bullet}{-} \varphi_m)$$

where the summations extend over m and n. Let n be fixed and consider those values of m which give an average different from zero. We see that $m = n$ and $m = k - n$ are two such values. The only other possibilities are $m = -n$ and $m = -k + n$, but these lead to terms containing (except when n or k equal zero) three different angles, φ_n, φ_{k-n}, and φ_{k+n} which average to zero. Using the fact that the average of cosine squared is one-half and that for a given n there are two such terms, we get

$$W(f_k)\Delta f = \frac{\alpha^2}{4} \sum_{n=-\infty}^{+\infty} c_{k-n}^2 c_n^2$$

$$= \alpha^2 \Delta f \sum_{n=-\infty}^{+\infty} w(f_k - f_n) w(f_n) \Delta f \qquad (4.4\text{–}5)$$

where in the last step we have used

$$f_{k-n} = (k - n)\Delta f = f_k - f_n$$

and have implied, from $c_{-n} = c_n$, that

$$w(f_{-n}) = w(-n\Delta f) = w(-f_n)$$

is equal to $w(f_n)$.

Thus, from (4.5–4), we get for the power spectrum of I

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

with the understanding that f is not zero and

$$w(-x) = w(x). \tag{4.5-6}$$

The result which is obtained by using (2.8–6), involving the cosines and only positive values of m, is

$$W(f) = \alpha^2 \int_0^f w(x)w(f - x)\,dx + 2\alpha^2 \int_0^\infty w(x)w(f + x)\,dx \tag{4.5-7}$$

This contains only positive values of frequency. (4.5–5) and (4.5–7) are equivalent and may readily be transformed into each other.

The first integral in (4.5–7) arises from second order modulation products of the sum type and the second integral from products of the difference type. This may be seen by writing the current as

$$\begin{aligned} I = \alpha V^2 &= \alpha \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} c_m c_n \cos(\omega_m t - \varphi_m) \cos(\omega_n t - \varphi_n) \\ &= \frac{\alpha}{2} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} c_m c_n \{\cos[(\omega_m - \omega_n)t - \varphi_m + \varphi_n] \\ &\qquad + \cos[(\omega_m + \omega_n)t + \varphi_m + \varphi_n]\} \end{aligned} \tag{4.5-8}$$

The power in the range $f_k, f_k + \Delta f$ is the power due to modulation products of the difference type, $\omega_{k+\ell} - \omega_\ell$, plus the power due to the modulation products of the sum type, $\omega_{k-\ell} + \omega_\ell$. In the first type ℓ runs from 1 to ∞ and in the second type ℓ runs from 1 to $k - 1$.

Consider the difference type first, and for the moment take both k and ℓ to be fixed. The two sets $m = k + \ell$, $n = \ell$ and $m = \ell$, $n = k + \ell$ are the only values of m and n in (4.5–8) leading to $\omega_{k+\ell} - \omega_\ell$. The two corresponding terms in (4.5–8) are equal because $\cos(-x)$ is equal to $\cos x$. The average power contributed by these two terms is

$$\begin{aligned} \left(\frac{\alpha}{2} c_{k+\ell} c_\ell\right)^2 \times \{&\text{Average of } (2\cos[(\omega_{k+\ell} - \omega_\ell)t - \varphi_{k+\ell} + \varphi_\ell])^2\} \\ &= \tfrac{1}{2}(\alpha c_{k+\ell} c_\ell)^2 \end{aligned} \tag{4.5-9}$$

The power contributed to $f_k, f_k + \Delta f$ by the difference modulation products is obtained by summing ℓ from 1 to ∞:

$$\begin{aligned} \frac{\alpha^2}{2} \sum_{\ell=1}^{\infty} c_{k+\ell}^2 c_\ell^2 &= 2\alpha^2 \sum_{\ell=1}^{\infty} w(f_{k+\ell})w(f_\ell)(\Delta f)^2 \\ &\to 2\alpha^2 \Delta f \int_0^\infty w(f_k + f)w(f)\,df \end{aligned}$$

This leads to the second term in (4.5–7).

Now consider the modulation products of the sum type. The terms of this type in (4.5–8) which give rise to the frequency ω_k are those for which $m + n$ is equal to k. Let n be 1 then $m = k - 1$. The phase of this term is random with respect to all the other terms except the one given by $n = k - 1$, $m = 1$ which has the same phase. The average power contributed by these two terms in (4.5–8) is, as in (4.5–9),

$$\tfrac{1}{2}(\alpha c_1 c_{k-1})^2$$

This disposes of two terms for which $m + n$ is equal to k. Taking n to be 2 and going through the same process gives two more. Thus, assuming for the moment that k is an odd number, the power contributed to the interval f_k, $f_k + \Delta f$ by the sum modulation products is

$$\frac{1}{2}\sum_{n=1}^{(k-1)/2}(\alpha c_n c_{k-n})^2 = \frac{1}{4}\sum_{n=1}^{k-1}(\alpha c_n c_{k-n})^2 \to \alpha^2 \Delta f \int_0^{f_k} w(f) w(f_k - f)\, df$$

and this leads to the second term in (4.5–7).

When the voltage V applied to the square law device is the sum of a noise voltage V_N and a sine wave:

$$V = P \cos pt + V_N, \tag{4.1–13}$$

we have

$$V^2 = P^2 \cos^2 pt + 2PV_N \cos pt + V_N^2 \tag{4.5–10}$$

From the two equations

$$\cos^2 pt = \frac{1}{2} + \frac{1}{2}\cos 2pt$$

$$\text{ave. } V_N^2 = \sum_1^M c_m^2 \frac{1}{2} \to \int_0^\infty w(f)\, df$$

we see that I, or αV^2, has a dc component of

$$\frac{\alpha P^2}{2} + \alpha \int_0^\infty w(f)\, df \tag{4.5–11}$$

which agrees with (4.1–14), and a sinusoidal component

$$\frac{\alpha P^2}{2}\cos 2pt \tag{4.5–12}$$

The continuous power spectrum $W_c(f)$ of the remaining portion of I may be computed from

$$2PV_N \cos pt + V_N^2.$$

Using the representation (2.8-6) we see

$$2PV_N \cos pt = P \sum_1^M c_m[\cos(\omega_m t + pt - \varphi_m) + \cos(\omega_m t - pt - \varphi_m)]$$

For the moment, we take $p = 2\pi r\Delta f$. The terms pertaining to frequency $f_n = n\Delta f$ are those for which

$$\omega_m + p = 2\pi f_n \qquad |\omega_m - p| = 2\pi f_n$$

$$m + r = n \qquad |m - r| = n$$

$$m = n - r \qquad m = r \pm n$$

where only positive values of m are to be taken: If $n > r$, then m is $n - r$ or $r + n$. If $n < r$, then m is $r - n$ or $r + n$. In either case the values of m are $|n - r|$ and $n + r$. The terms of frequency f_n in $2PV_N \cos pt$ are therefore

$$Pc_{|n-r|}\cos(2\pi f_n t - \varphi_{|n-r|}) + Pc_{n+r}\cos(2\pi f_n t - \varphi_{n+r})$$

and the mean square value of this expression, the average being taken over the φ's, is

$$\frac{P^2}{2}(c^2_{|n-r|} + c^2_{n+r}) = P^2\Delta f[w(f_{|n-r|}) + w(f_{n+r})]$$

$$= P^2\Delta f[w(|f_n - f_p|) + w(f_n + f_p)]$$

where f_p denotes $p/2\pi$.

By combining this with the expression (4.5–5) which arises from V_N^2 we see that the continuous portion $W_c(f)$ of the power spectrum of I is

$$W_c(f) = \alpha^2 P^2[w(f - f_p) + w(f + f_p)] + \alpha^2 \int_{-\infty}^{+\infty} w(x)w(f - x)\,dx \tag{4.5–13}$$

where $w(-f)$ has the same value as $w(f)$.

Equation (4.5–13) has been used to compute $W_c(f)$ as shown in Fig. 8. The input noise is assumed to be uniform over a band of width β centered at f_p, cf. Filter c, Appendix C. By noting the area under the low frequency portion of the spectrum we find

$$\int_0^\beta W_c(f)\,df = \alpha^2\beta w_0(P^2 + \beta w_0)$$

Since the mean square value of the input V_N is $\psi_0 = \beta w_0$, it is seen that this equation agrees with the expression (4.1–15) for the mean square value of $I\ell_f$, the low frequency current, excluding the d.c. If audio frequency

filters cut out part of the spectrum, $W_c(f)$ may be integrated over the remaining portion to give the mean square value of the corresponding output current. This idea is mentioned in the footnote pertaining to equation (4.1–6).

If V consists of V_N plus two sinusoidal voltages of incommensurable frequencies, say

$$V = P \cos pt + Q \cos qt + V_N ,$$

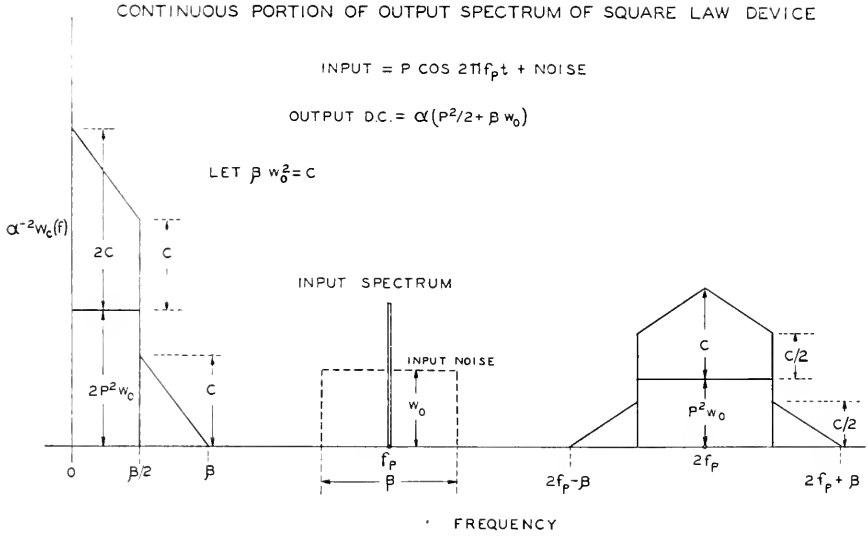


Fig. 8

the continuous portion $W_c(f)$ of the power spectrum of I may be shown to be (4.5–13) plus the additional terms

$$\alpha^2 Q^2 [w(f - f_q) + w(f + f_q)] \tag{4.5–14}$$

where f_q denotes $q/2\pi$.

When the voltage applied to the square law device (4.1–1) is[49]

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

$$= Q \cos qt + \frac{Qk}{2} \cos (p + q)t + \frac{Qk}{2} \cos (p - q)t + V_N$$

the resulting current contains the *dc* component

$$\frac{\alpha}{2} Q^2 \left(1 + \frac{k^2}{2}\right) + \alpha \int_0^\infty w(f)\, df \tag{4.5–16}$$

[49] A complete discussion of this problem is given by L. A. MacColl in a manuscript being prepared for publication.

The sinusoidal terms of I are obtained by squaring

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

and multiplying by α. The remaining portion of I has a continuous power spectrum given by

$$\begin{aligned} W_c(f) = \alpha^2 Q^2 \Big[& w(f - f_q) + w(f + f_q) \\ & + \frac{k^2}{4} w(f - f_p - f_q) + \frac{k^2}{4} w(f + f_p + f_q) \\ & + \frac{k^2}{4} w(f - f_p + f_q) + \frac{k^2}{4} w(f + f_p - f_q) \Big] \\ & + \alpha^2 \int_{-\infty}^{+\infty} w(x) w(f - x)\, dx \end{aligned} \tag{4.5-17}$$

where f_p denotes $p/2\pi$ and f_q denotes $q/2\pi$.

4.6 Two Correlation Function Methods

As mentioned in Section 4.4 these methods for determining the output power spectrum are based on finding the correlation function $\Psi(\tau)$ for the output current. From this the power spectrum, $W(f)$, of the output current may be obtained from (2.1–5), rewritten as

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

It will be recalled that $W(f)\Delta f$ may be regarded as the average power which would be dissipated by those components of I in the band $f, f + \Delta f$ if I were to flow through a resistance of one ohm.

The input of the non-linear device is taken to be a voltage $V(t)$. It may, for example, consist of a noise voltage $V_N(t)$ plus sinusoidal components. The output is taken to be a current $I(t)$. The non-linear device is specified by a relation between $V(t)$ and $I(t)$. In this work $I(t)$ at time t is assumed to be completely determined by the value of $V(t)$ at time t.

Two methods of obtaining $\Psi(\tau)$ will be described.

(a) Integrating the two-dimensional probability density of $V(t)$ and $V(t + \tau)$ over the values allowed by the non-linear device. This method, which is especially direct when applied to noise alone through rectifiers, was discovered independently by Van Vleck and North.

(b) Introducing and using the characteristic function, which for the sake of brevity will be abbreviated to ch. f., of the two-dimensional probability distribution of $V(t)$ and $V(t + \tau)$.

4.7 Linear Detection of Noise—The Van Vleck-North Method

The method due to Van Vleck and North will be illustrated by using it to determine the output power spectrum of a linear detector when the input consists of noise alone.

The linear detector is specified by

$$I(t) = \begin{cases} 0, & V(t) < 0 \\ V(t), & V(t) > 0, \end{cases} \tag{4.7-1}$$

which may be obtained from (4.2–1) by setting α equal to one, and the input voltage is

$$V(t) = V_N(t) \tag{4.7-2}$$

where $V_N(t)$ is a noise voltage whose correlation function is $\psi(\tau)$ and whose power spectrum is $w(f)$.

The correlation function $\Psi(\tau)$ is the average value of $I(t)I(t+\tau)$. This is the same as the average value of the function

$$F(V_1, V_2) = \begin{cases} V_1 V_2, & \text{when both } V_1, V_2 > 0 \\ 0, & \text{all other } V\text{'s}, \end{cases} \tag{4.7-3}$$

where we have set

$$V_1 = V(t)$$

$$V_2 = V(t+\tau)$$

The two-dimensional distribution of V_1 and V_2 is given by (3.2–4), and from this it follows that the average value of any function $F(V_1, V_2)$ is

$$\int_{-\infty}^{+\infty} dV_1 \int_{-\infty}^{+\infty} dV_2 \frac{F(V_1, V_2)}{2\pi |M|^{1/2}} \exp\left[-\frac{1}{2|M|}(\psi_0 V_1^2 + \psi_0 V_2^2 - 2\psi_\tau V_1 V_2)\right] \tag{4.7-4}$$

where

$$|M| = \psi_0^2 - \psi_\tau^2 .$$

For the linear rectifier case, where $F(V_1, V_2)$ is given by (4.7–3), the integral is

$$|M|^{-1/2} \frac{1}{2\pi} \int_0^\infty dV_1 \int_0^\infty dV_2 V_1 V_2 \exp\left[-\frac{1}{2|M|}(\psi_0 V_1^2 + \psi_0 V_2^2 - 2\psi_\tau V_1 V_2)\right]$$

$$= \frac{1}{2\pi}\left([\psi_0^2 - \psi_\tau^2]^{1/2} + \psi_\tau \cos^{-1}\left[\frac{-\psi_\tau}{\psi_0}\right]\right)$$

where we have used (3.5–4) to evaluate the integral. The arc cosine is taken to be between 0 and π. We therefore have for the correlation function of $I(t)$,

$$\Psi(\tau) = \frac{1}{2\pi}\left([\psi_0^2 - \psi_\tau^2]^{1/2} + \psi_\tau \cos^{-1}\left[\frac{-\psi_\tau}{\psi_0}\right]\right) \tag{4.7–5}$$

The power spectrum $W(f)$ may be obtained from this by use of (4.6–1). For this purpose it is convenient to write (4.7–5) in terms of a hypergeometric function. By expanding and comparing terms it is seen that

$$\begin{aligned}\Psi(\tau) &= \frac{\psi_\tau}{4} + \frac{\psi_0}{2\pi} F\left(-\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{1}{2}; \frac{\psi_\tau^2}{\psi_0^2}\right) \\ &= \frac{\psi_\tau}{4} + \frac{\psi_0}{2\pi} + \frac{\psi_\tau^2}{4\pi\psi_0} + \text{terms involving } \psi_\tau^4, \psi_\tau^6, \text{ etc.}\end{aligned} \tag{4.7–6}$$

As will be discussed more fully in Section 4.8, a constant term A^2 in $\psi(\tau)$ indicates a direct current component of $I(t)$ of A amperes. Thus $I(t)$ has a *dc* component equal to

$$\left[\frac{\psi_0}{2\pi}\right]^{1/2} = \frac{1}{\sqrt{2\pi}} \times \text{rms value of } V(t) \tag{4.7–7}$$

This agrees with (4.2–3) when the P of that equation is set equal to zero.

Integrals of the form

$$G_n(f) = \int_0^\infty \psi_\tau^n \cos 2\pi f\tau \, d\tau$$

which result when (4.7–6) is put in (4.6–1) and integrated termwise are discussed in Appendix 4C. From the results given there it is seen that if we neglect ψ_τ^4 and higher powers we obtain an approximation for the continuous portion $W_c(f)$ of $W(f)$:

$$\begin{aligned}W_c(f) &\approx G_1(f) + \frac{G_2(f)}{\pi\psi_0} \\ &= \frac{w(f)}{4} + \frac{1}{4\pi\psi_0}\cdot\frac{1}{2}\int_{-\infty}^{+\infty} w(x)w(f-x)\,dx\end{aligned} \tag{4.7–8}$$

where $w(-f)$ is defined as $w(f)$.

When $V_N(t)$ is uniform over a relatively narrow band extending from f_a to f_b so that $w(f)$ is equal to w_0 in this band and is zero outside it, we may use the results for Filter c of Appendix 4C. The f_0 and β given there are related to f_a and f_b by

$$f_a = f_0 - \frac{\beta}{2}, \qquad f_b = f_0 + \frac{\beta}{2}$$

and the value of w_0 taken there is the same as here and is ψ_0/β. The value of $G_2(f)$ given there leads to the approximation, for low frequencies:

$$\begin{aligned} W_c(f) &\approx \frac{1}{\pi\psi_0}\frac{\psi_0^2}{4\beta}\left(1 - \frac{f}{\beta}\right) \\ &= \frac{w_0}{4\pi}\left(1 - \frac{f}{f_b - f_a}\right) \end{aligned} \tag{4.7-9}$$

when $0 < f < f_b - f_a$, and to $W_0(f) \approx 0$ for $f_b - f_a < f < f_a$. By setting P equal to zero in the curve given in Fig. 8 for $W_c(f)$ corresponding to the square law detector, we see that the low frequency portion of the power spectrum is triangular in shape and is zero at $f = \beta$. Thus, looking at (4.7–9), we see that to a first approximation the shape of the output power spectrum is the same for a linear detector as for a square law detector when the input consists of a relatively narrow band of noise.

An approximate rms value of the low frequency output current may be obtained by integrating (4.7–9)

$$\begin{aligned} \overline{I_{lf}^2} &= \int_0^{f_b - f_a} W_c(f)\, df \\ &\approx \frac{w_0(f_b - f_a)}{8\pi} = \frac{\psi_0}{8\pi} \end{aligned}$$

$$\text{rms low freq. current} \approx \frac{1}{\sqrt{8\pi}} \times \text{rms applied voltage} \tag{4.7-10}$$

It is seen that this is half of the direct current. It must be kept in mind that (4.7–10) is an approximation because we have neglected ψ_τ^4 and higher powers. The true value may be obtained from (4.2–8). It is seen that the coefficient $(8\pi)^{-1/2} = 0.200$ should be replaced by

$$\frac{1}{\pi}\left(2 - \frac{\pi}{2}\right)^{1/2} = 0.209$$

$W_c(f)$ for other types of band pass filters may be obtained by using the corresponding G's given in appendix 4C. It turns out that (4.7–10) holds for all three types of filters. This is a special case of Middleton's theorem, mentioned several times before, that the total power in any modulation product (it will be shown later in Section 4.9 that the term ψ_τ^n in (4.7–6) corresponds to the n^{th} order modulation products) depends only on the total input power of the applied noise, not on its spectral distribution.

4.8 The Characteristic Function Method

As mentioned in the preceding parts, especially in connection with equation (1.4–3), the ch. f. of a random variable x is the average value of exp

(iux). This is a function of u. The ch. f. of two random variables x and y is the average value of $\exp(iux + ivy)$ and is a function of u and v. The ch. f. which we shall use here is the ch. f. of the two random variables $V(t)$ and $V(t+\tau)$ where $V(t)$ is the voltage applied to the non-linear device, and the randomness is introduced by t being selected at random, τ remaining fixed. We may write this characteristic function as

$$g(u, v, \tau) = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T \exp[iuV(t) + ivV(t+\tau)]\,dt \quad (4.8\text{–}1)$$

If $V(t)$ contains a noise voltage $V_N(t)$, as it always does in this section, and if we use the representation (2.8–1) or (2.8–6) a large number of random parameters (a_n's and b_n's or φ_n's) will appear in (4.8–1). In accordance with our use of such representations we may average over these parameters without changing the value of (4.8–1) and may thereby simplify the integration.

For example suppose

$$V(t) = V_s(t) + V_N(t) \quad (4.8\text{–}2)$$

where $V_s(t)$ is some regular voltage which may, e.g., consist of one or more sine waves. Substituting this in (4.8–1) and using the result (3.2–7) that the ch. f. of $V_N(t)$ and $V_N(t+\tau)$ is

$$\begin{aligned} g_N(u, v, \tau) &= \text{ave. } \exp[iuV_N(t) + ivV_N(t+\tau)] \\ &= \exp\left[-\frac{\psi_0}{2}(u^2 + v^2) - \psi_\tau uv\right] \end{aligned} \quad (4.8\text{–}3)$$

$\psi_\tau \equiv \psi(\tau)$ being the correlation function of $V_N(t)$, we obtain for the ch. f. of $V(t)$ and $V(t+\tau)$,

$$\begin{aligned} g(u, v, \tau) &= \exp\left[-\frac{\psi_0}{2}(u^2 + v^2) - \psi_\tau uv\right] \\ &\quad \times \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T \exp[iuV_s(t) + ivV_s(t+\tau)]\,dt \\ &= g_N(u, v, \tau) g_s(u, v, \tau) \end{aligned} \quad (4.8\text{–}4)$$

In the last line we have used $g_s(u, v, \tau)$ to denote the limit in the line above:

$$g_s(u, v, \tau) = \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T \exp[iuV_s(t) + ivV_s(t+\tau)]\,dt \quad (4.8\text{–}5)$$

The principal reason we use the ch. f. is because quite a few non-linear devices may be described by the integral

$$I = \frac{1}{2\pi}\int_C F(iu)e^{iVu}\,du \quad (4A\text{–}1)$$

where the function $F(iu)$ and the path of integration C are chosen to fit the device. Examples of such devices are given in Appendix 4A. The correlation function $\Psi(\tau)$ of $I(t)$ is given by

$$\begin{aligned}\Psi(\tau) &= \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T I(t)I(t+\tau)\,dt \\ &= \operatorname*{Limit}_{T\to\infty} \frac{1}{4\pi^2 T}\int_0^T dt \int_C F(iu)e^{iuV(t)}\,du \int_C F(iv)e^{ivV(t+\tau)}\,dv \\ &= \frac{1}{4\pi^2}\int_C F(iu)\,du \int_C F(iv)\,dv \\ &\qquad \operatorname*{Limit}_{T\to\infty} \frac{1}{T}\int_0^T \exp\left[iuV(t) + ivV(t+\tau)\right] dt \\ &= \frac{1}{4\pi^2}\int_C F(iu)\,du \int_C F(iv)g(u, v, \tau)\,dv\end{aligned} \tag{4.8-6}$$

This is the fundamental formula of the ch. f. method.

When $V(t)$ is the sum of a noise voltage and a regular voltage, as in (4.8-2), (4.8-6) becomes

$$\begin{aligned}\Psi(\tau) = \frac{1}{4\pi^2}\int_C F(iu)e^{-(\psi_0/2)u^2}\,du \int_C F(iv)e^{-(\psi_0/2)v^2} \\ e^{-\psi_\tau uv} g_s(u, v, \tau)\,dv\end{aligned} \tag{4.8-7}$$

where $g_s(u, v, \tau)$ is the ch. f. of $V_s(t)$ and $V_s(t+\tau)$ given by (4.8-5). This is a definite expression for $\Psi(\tau)$. All that follows is devoted to the evaluation of this integral and to the evaluation of

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

for the power spectrum of I.

Quite often $I(t)$ will contain *dc* and periodic components. It seems convenient to deal with these separately since they correspond to terms in $\Psi(\tau)$ which cause the integral (4.6-1) for $W(f)$ to diverge. In fact, from Section 2.2 it follows that a correlation function of the form

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

corresponds to a current

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

where the phase angle φ cannot be determined from (2.2–3) since it does not affect the average power.

Consider the correlation function for $V(t) = V_s(t) + V_N(t)$ given by (4.8–2). It is

$$\underset{T\to\infty}{\text{Limit}}\ \frac{1}{T}\left[\int_0^T V_s(t)V_s(t+\tau)\,dt + \int_0^T V_s(t)V_N(t+\tau)\,dt + \int_0^T V_N(t)V_s(t+\tau)\,dt + \int_0^T V_N(t)V_N(t+\tau)\,dt\right] \tag{4.8–8}$$

Since $V_s(t)$ and $V_N(t)$ are unrelated the contributions of the second and third integrals vanish leaving us with the result

$$\text{Correlation function of } V(t) = \text{Correlation function of } V_s(t) + \text{Correlation function of } V_N(t). \tag{4.8–9}$$

Now as $\tau \to \infty$ the correlation function of $V_N(t)$ becomes zero while that of $V_s(t)$ becomes of the type (2.2–3) given above. Hence the correlation function of the regular voltage $V_s(t)$ may be obtained from $V(t)$ by letting $\tau \to \infty$ and picking out the non-vanishing terms. Although we have been speaking of $V(t)$, the same results hold for $I(t)$ and this process may be used to pick out those parts of $\Psi(\tau)$ which correspond to the *dc* and periodic components of $I(t)$. Thus, if we look at (4.8–7) we see that as $\tau \to \infty$, $\psi_\tau \to 0$, while the $g_s(u, v, \tau)$ corresponding to $V_s(t)$ given by (4.8–5) remains unchanged in general magnitude. This last statement may be hard to see, but examination of the cases discussed later show that it is true, at least for these cases. Thus the portion of $\Psi(\tau)$ corresponding to the *dc* and periodic components of $I(t)$ is, setting $\psi_\tau = 0$ in (4.8–7),

$$\Psi_\infty(\tau) = \frac{1}{4\pi^2}\int_C F(iu)e^{-(\psi_0/2)u^2}\,du \int_C F(iv)e^{-(\psi_0/2)v^2}g_s(u, v, \tau)\,dv \tag{4.8–10}$$

where the subscript ∞ indicates that $\Psi_\infty(\tau)$ is that part of $\Psi(\tau)$ which does not vanish as $\tau \to \infty$.

We may write (4.8–9), when applied to $I(t)$, as

$$\Psi(\tau) = \Psi_\infty(\tau) + \Psi_c(\tau) \tag{4.8–11}$$

where $\Psi_c(\tau)$ is the correlation function of the "continuous" portion of the power spectrum of $I(t)$.

Incidentally, the separation of $\Psi(\tau)$ into the two parts shown in (4.8–11) may be avoided if one is willing to use the $\delta(f)$ functions in order to interpret the integral in (4.6–1) as explained in Section 2.2. This method gives the proper *dc* and sinusoidal components even though (4.6–1) does not converge (because of the presence of the terms leading to $\Psi_\infty(\tau)$).

4.9 Noise Plus Sine Wave Applied to Non-Linear Device

In order to illustrate the characteristic function method described in Section 4.8 we shall consider the case of a non-linear device specified by

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

when V consists of a noise voltage plus a sine wave:

$$V(t) = P \cos pt + V_N(t) \tag{4.1-13}$$

As usual, $V_N(t)$ has the power spectrum $w(f)$ and the correlation function $\psi(\tau)$. $\psi(\tau)$ is often written as ψ_τ for the sake of shortness. Comparing (4.1–13) with (4.8–2) gives

$$V_s(t) = P \cos pt \tag{4.9-1}$$

Our first task is to compute the ch. f. $g_s(u, v, \tau)$ for the pair of random variables $V_s(t)$ and $V_s(t + \tau)$. We do this by using the integral (4.8–5):

$$\begin{aligned} g_s(u, v, \tau) &= \operatorname*{Limit}_{T\to\infty} \frac{1}{T} \int_0^T \exp\left[iuP \cos pt + ivP \cos p(t + \tau)\right] dt \\ &= J_0(P\sqrt{u^2 + v^2 + 2uv \cos p\tau}) \end{aligned} \tag{4.9-2}$$

where J_0 is a Bessel function. The integration is performed by writing

$$\begin{aligned} u \cos pt + v \cos p\,(t + \tau) &= (u + v \cos p\tau) \cos pt - v \sin p\tau \sin pt \\ &= \sqrt{u^2 + v^2 + 2uv \cos p\tau}\, \cos\,(pt + \text{phase angle}) \end{aligned}$$

and using the integral

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{iz \cos t}\, dt$$

The correlation function for (4.1–13) has also been given in Section 3.10.

The correlation function $\Psi(\tau)$ for $I(t)$ may now be obtained by substituting the above expressions in (4.8–7)

$$\begin{aligned} \Psi(\tau) = \frac{1}{4\pi^2} \int_C du\, F(iu) e^{-(\psi_0/2)u^2} \int_C dv\, F(iv) e^{-(\psi_0/2)v^2} \\ e^{-\psi_\tau uv} J_0(P\sqrt{u^2 + v^2 + 2uv \cos p\tau}). \end{aligned} \tag{4.9-3}$$

$\Psi_\infty(\tau)$, the correlation function for the d.c. and periodic components of I, may, according to (4.8–10), be obtained from this by setting ψ_τ equal to zero.

When we have a particular non-linear device in mind the appropriate $F(iu)$ may often be obtained from Appendix 4A. For example, $F(iu)$ for a linear rectifier is $-u^{-2}$. Inserting this value in (4.9–3) gives a definite

double integral for $\Psi(\tau)$. If there were some easy way to evaluate this integral then everything would be fine. Unfortunately, no simple method of evaluation has yet been found. However, one method is available which is closely related to the direct method used by Bennett. It is based on the expansion

$$\begin{aligned} g_s(u, v, \tau) &= J_0(P\sqrt{u^2 + v^2 + 2uv \cos p\tau}) \\ &= \sum_{n=0}^{\infty} \epsilon_n (-)^n J_n(Pu) J_n(Pv) \cos np\tau \end{aligned} \tag{4.9-4}$$

$$\epsilon_0 = 1, \quad \epsilon_n = 2 \quad \text{for} \quad n \geq 1$$

This expansion enables us to write the troublesome terms in (4.9–3) as

$$\begin{aligned} &e^{-\psi_\tau uv} J_0(P\sqrt{u^2 + v^2 + 2uv \cos p\tau}) \\ &\quad = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} (-)^{n+k} \epsilon_n \cos np\tau \frac{(\psi_\tau uv)^k}{k!} J_n(Pu) J_n(Pv) \end{aligned} \tag{4.9-5}$$

The virtue of this double sum is that it simplifies the integration. Thus, putting it in (4.9–3) and setting

$$h_{nk} = \frac{i^{n+k}}{2\pi} \int_C F(iu) u^k J_n(Pu) e^{-(\psi_0/2)u^2} \, du \tag{4.9-6}$$

gives

$$\Psi(\tau) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{k!} \psi_\tau^k h_{nk}^2 \epsilon_n \cos np\tau \tag{4.9-7}$$

The correlation function $\Psi_\infty(\tau)$ for the dc and periodic components of I are obtained by letting $\tau \to \infty$ where $\psi_\tau \to 0$. Only the terms for which $k = 0$ remain:

$$\Psi_\infty(\tau) = \sum_{n=0}^{\infty} \epsilon_n h_{n0}^2 \cos np\tau \tag{4.9-8}$$

Comparing this with the known fact that the correlation function of

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

is

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

and remembering that ϵ_0 is one while ϵ_n is two for $n \geq 1$ shows that

$$\begin{aligned} &\text{Amplitude of dc component of } I = h_{00} \\ &\text{Amplitude of } \frac{np}{2\pi} \text{ component of } I = 2h_{n0} \end{aligned} \tag{4.9-9}$$

Incidentally, these expressions for the amplitudes follow almost at once from the direct method of solution. This will be shown in connection with equation (4.9–17).

Since the correlation function $\Psi_c(\tau)$ for the continuous portion $W_c(f)$ of the power spectrum for I is given by

$$\Psi_c(\tau) = \Psi(\tau) - \Psi_\infty(\tau), \tag{4.8–11}$$

we also have

$$\Psi_c(\tau) = \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} \frac{1}{k!} \psi_\tau^k h_{nk}^2 \epsilon_n \cos np\tau \tag{4.9–10}$$

When this is substituted in

$$W_c(f) = 4 \int_0^\infty \Psi_c(\tau) \cos 2\pi f\tau \, d\tau \tag{4.9–11}$$

we obtain

$$W_c(f) = \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} \frac{2\epsilon_n}{k!} h_{nk}^2 \left[G_k\left(f - \frac{np}{2\pi}\right) + G_k\left(f + \frac{np}{2\pi}\right)\right] \tag{4.9–12}$$

where

$$G_k(f) = \int_0^\infty \psi_\tau^k \cos 2\pi f\tau \, d\tau \tag{4.9–13}$$

is the function studied in Appendix 4C. $G_k(f)$ is an even function of f. The double series (4.9–12) for W_c looks rather formidable. However, when we are interested in a particular portion of the frequency spectrum often only a few terms of the series are needed.

It has been mentioned above that the direct method of obtaining the output power spectrum is closely related to the equations just derived. We now study this relation.

We start with the following result from modulation theory[50]: Let the voltage

$$\begin{aligned} V &= P_0 \cos x_0 + P_1 \cos x_1 + \cdots + P_N \cos x_N \\ x_k &= p_k t, \qquad k = 0, 1, \cdots N, \end{aligned} \tag{4.9–14}$$

where the p_k's are incommensurable, be applied to the device (4A–1). The output current is

$$I = \sum_{m_0=0}^{\infty} \cdots \sum_{m_N=0}^{\infty} \tfrac{1}{2} A_{m_0 \cdots m_N} \epsilon_{m_0} \cdots \epsilon_{m_N} \cos m_0 x_0 \cos m_1 x_1 \cdots \cos m_N x_N \tag{4.9–15}$$

[50] Bennett and Rice, "Note on Methods of Computing Modulation Products," *Phil. Mag.* S.7, V. 18, pp. 422–424, Sept. 1934, and Bennett's paper cited in Section 4.0.

where $\epsilon_0 = 1$ and $\epsilon_m = 2$ for $m \geq 1$. When the product of the cosines is expressed as a sum of cosines of the angles $m_0 x_0 \pm m_1 x_1 \cdots \pm m_N x_N$, it is seen that the coefficient of the typical term is $A_{m_0 \cdots m_N}$, except when all the m's are zero in which case it is $\frac{1}{2} A_{0 \cdots 0}$. Thus

$$\begin{aligned} \tfrac{1}{2} A_{00\cdots 0} &= \text{dc component of } I \\ | A_{m_0 \cdots m_N} | &= \text{amplitude of component of frequency} \\ & \qquad \frac{1}{2\pi} | m_0 p_0 \pm m_1 p_1 \pm \cdots \pm m_N p_N | \end{aligned} \tag{4.9-16}$$

For all values of the m's,

$$\begin{aligned} A_{m_0 \cdots m_N} &= \frac{i^M}{\pi} \int_C F(iu) \prod_{r=0}^{N} J_{m_r}(P_r u)\, du \\ M &= m_0 + m_1 + \cdots + m_N \end{aligned} \tag{4.9-17}$$

Following Bennett's procedure, we identify V as given by (4.9-14), with

$$V = P \cos pt + V_N \tag{4.1-13}$$

by setting $P_0 = P$, $p_0 = p$, and representing the noise voltage V_N by the sum of the remaining terms. Since this makes P_1, P_N all very small, Laplace's process indicates that in (4.9-17) we may put

$$\begin{aligned} \prod_{r=1}^{N} J_0(P_r u) &\approx \exp - \frac{u^2}{4} (P_1^2 + \cdots + P_N^2) \\ &\approx e^{-\psi_0 u^2/2} \end{aligned} \tag{4.9-18}$$

We have used the fact that ψ_0 is the mean square value of V_N. It follows from these equations that

$$\text{dc component of } I = \frac{1}{2\pi} \int_C F(iu) J_0(Pu) e^{(-\psi_0/2)u^2}\, du$$

$$\text{Component of frequency } \frac{np}{2\pi} = \frac{i^n}{\pi} \int_C F(iu) J_n(Pu) e^{-\psi_0 u^2/2}\, du$$

These results are identical with those of (4.9-9).

The equations just derived show that h_{n0} is to be associated with the n^{th} harmonic of p. In much the same way it may be shown that h_{nk} is to be associated with the modulation products arising from the n^{th} harmonic of p and k of the elementary sinusoidal components representing V_N. We consider only combinations of the form $p_1 \pm p_2 \pm p_3$, taking $k = 3$ for example, and neglect terms of the form $3p_1$ and $2p_1 \pm p_2$. The former type is much more numerous, there being about N^3 of them while there are only about N and N^2, respectively, of the latter type.

We again take $k = 3$ and consider m_1, m_2, m_3 to be one, and $m_4, \cdots m_N$ to be zero, corresponding to the modulation product $np \pm p_1 \pm p_2 \pm p_3$. By making the same sort of approximations as Bennett does we find

$$A_{n,1,1,1,0,0\cdots 0} = \frac{i^{n+3}}{\pi} \frac{P_1 P_2 P_3}{8} \int_C F(iu) J_n(Pu) u^3 e^{(-u^2/2)\psi_0} \, du$$

$$= \frac{P_1 P_2 P_3}{4} h_{n3}$$

When any other modulation product of the form $np \pm p_{r_1} \pm p_{r_2} \pm p_{r_3}$ is considered we get a similar expression in which $P_1P_2P_3$ is replaced by $P_{r_1}P_{r_2}P_{r_3}$. This may be done for any value of k. The result indicates that h_{nk}, and consequently also the $(n, k)^{\text{th}}$ terms in the double series (4.9–10) and (4.9–12) for $\Psi_c(\tau)$ and $W_c(f)$, are to be associated with the modulation products of order (n, k), the n referring to the signal and the k to the noise components.

We now may state a theorem due to Middleton regarding the total power in the modulation products of a given order. For a given non-linear device (i.e. $F(iu)$ is given), the total power which would be dissipated by all of the modulation products which are of order (n, k) if I were to flow through a resistance of one ohm is

$$\Psi_{nk}(0) = \frac{\epsilon_n[\psi(0)]^k}{k!} h_{nk}^2 = \frac{\epsilon_n[\overline{V_N^2}]^k h_{nk}^2}{k!} \tag{4.9–19}$$

The important feature of this expression is that it depends only on the r.m.s. value of V_N and on $F(iu)$. It depends not at all upon the spectral distribution of the noise power in the input.

The proof of (4.9–19) is based on the relation

$$\Psi_{nk}(0) = \int_0^\infty W_{nk}(f) \, df$$

between the total power dissipated by all the (n, k) order products and the corresponding correlation function obtained from (4.9–7).

This theorem has been used by Middleton to show that when the input is confined to a relatively narrow frequency band, so that the output spectrum consists of bands, the power in each band depends only on $\overline{V_N^2}$ and not on the spectrum of V_N.

4.10 Miscellaneous Results Obtained by Correlation Function Method

In this section a number of results which may be obtained from the theory given in the sections following 4.6 are given.

When the input to the square law device

$$I = \alpha V^2 \tag{4.1-1}$$

consists of noise only, so that $V = V_N$, the correlation function for I is

$$\Psi(\tau) = \alpha^2[\psi_0^2 + 2\psi_\tau^2] \tag{4.10-1}$$

where ψ_τ is the correlation function of V_N. This may be compared with equation (3.9–7). When V is general,

$$\begin{aligned}\Psi(\tau) &= \text{ave. } I(t)I(t+\tau)\\ &= \text{ave. } \alpha^2 V^2(t)V^2(t+\tau)\\ &= \alpha^2 \times \text{Coefficient of } \frac{(iu)^2}{2!}\frac{(iv)^2}{2!} \text{ in power series expansion}\\ &\qquad\qquad \text{of ch. f. of } V(t),\ V(t+\tau)\end{aligned} \tag{4.10-2}$$

where we have used a known property of the characteristic function. An expression for the ch. f., denoted by $g(u, v, \tau)$, is given by (4.8–4). For example, when V consists of a sine wave plus noise, (4.1–13), the ch. f. is obtainable from (4.9–3). Hence,

$$\begin{aligned}\Psi(\tau) &= \text{Coeff. of } \frac{u^2v^2}{4} \text{ in expansion of}\\ &\qquad \alpha^2 J_0\left(P\sqrt{u^2 + v^2 + 2uv\cos p\tau}\right)\\ &\qquad \times \exp\left[-\frac{\psi_0}{2}(u^2 + v^2) - \psi_\tau uv\right]\\ &= \alpha^2\left[\left(\frac{P}{2} + \psi_0\right)^2 + \frac{P^4}{8}\cos 2p\tau + 2P^2\psi_\tau \cos p\tau + 2\psi_\tau^2\right]\end{aligned} \tag{4.10-3}$$

The first two terms give the dc and second harmonic. The last two terms may be used to compute $W_c(f)$ as given by (4.5–13).

Expressions (4.10–1) and (4.10–3) are special cases of results obtained by Middleton who has studied the general theory of the quadratic rectifier by using the Van Vleck-North method, described in Section 4.7.

As an example to which the theory of Section 4.9 may be applied we consider the sine wave plus noise, (4.1–13), to be applied to the ν-law rectifier

$$\begin{aligned} I &= 0, \qquad V < 0\\ I &= V^\nu, \qquad V > 0\end{aligned} \tag{4.10-4}$$

From the table in Appendix 4.1 it is seen that

$$F(iu) = \Gamma(\nu + 1)(iu)^{-\nu-1}$$

and that the path of integration C runs along the real axis from $-\infty$ to ∞ with a downward indentation at the origin. The integral (4.9–6) for h_{nk} becomes

$$h_{nk} = \frac{i^{n+k-\nu-1}}{2\pi}\,\Gamma(\nu+1)\int_C u^{k-\nu-1}J_n(Pu)e^{-(\psi_0/2)u^2}\,du$$

$$= \frac{\left(\frac{\psi_0}{2}\right)^{(\nu-k)/2} x^{n/2}\,\Gamma(\nu+1)}{2\Gamma\left(\frac{2-k-n+\nu}{2}\right)n!}\,{}_1F_1\left(\frac{k+n-\nu}{2};\, n+1;\, -x\right) \qquad (4.10\text{–}5)$$

$$x = \frac{P^2}{2\psi_0}$$

where the integration has been performed by expanding $J_n(Pu)$ in powers of u and using

$$\int_C e^{-au^2}u^{2\lambda-1}\,du = ie^{-\lambda i\pi}a^{-\lambda}\sin\lambda\pi\Gamma(\lambda)$$

$$= \frac{a^{-\lambda}}{2}(1-e^{-2\lambda i\pi})\Gamma(\lambda) \qquad (4.10\text{–}6)$$

$$= \frac{i\pi e^{-\lambda i\pi}}{a^{\lambda}\,\Gamma(1-\lambda)}$$

it being understood that $\arg u = 0$ on the positive portion of C.

From (4.9–9), the dc component of I is

$$h_{00} = \frac{\Gamma(1+\nu)}{2\Gamma\left(1+\frac{\nu}{2}\right)}\left(\frac{\psi_0}{2}\right)^{\nu/2}{}_1F_1\left(-\frac{\nu}{2};\, 1;\, -x\right) \qquad (4.10\text{–}7)$$

which reduces to the expression (4.2–3) when $\nu = 1$ for the linear rectifier (aside from the factor α).

When the input (sine wave plus noise) is confined to a relatively narrow band, and when we are interested in the low frequency output, consideration of the modulation products suggests that we consider the difference products from the products of order (0, 0), (0, 2), (0, 4), $\cdots$ (1, 1), (1, 3), $\cdots$ (2, 0), (2, 2), $\cdots$ etc. where the typical product is of order (n, k). The orders (0, 0) and (2, 0) give the dc and second harmonic and hence are not considered in the computation of $W_c(f)$. Of the remaining terms, either (0, 2) or (1, 1) gives the greatest contribution to the series (4.9–12) and (4.9–10) for $W_c(f)$ and $\Psi_c(\tau)$. The remaining terms contribute less and less as n and

k increase. The low frequency portion of the continuous portion of the output power spectrum is then, from (4.9–12),

$$W_c(f) = \frac{4}{2!} h_{02}^2 G_2(f) + \frac{4}{4!} h_{04}^2 G_4(f) + \cdots$$
$$+ \frac{4}{1!} h_{11}^2 [G_1(f - f_0) + G_1(f + f_0)] + \frac{4}{3!} h_{13}^2 [G_3(f - f_0) \quad (4.10\text{–}8)$$
$$+ G_3(f + f_0)] + \frac{4}{2!} h_{22}^2 [G_2(f - 2f_0) + G_2(f + 2f_0)] + \cdots$$

From Table 2 of Appendix 4C we may pick out the low frequency portions of the G's. It must be remembered that $G_m(x)$ is an even function of x and that $0 < f \ll f_0$.

As an example we take the input noise W_N to have the same $w(f)$ and $\psi(\tau)$ as Filter a, the normal law filter, of Appendix 4C, so that

$$w(f) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(f-f_0)^2/2\sigma^2}$$

and assume that the sine wave signal is at the middle of the band, giving $p = 2\pi f_0$. Thus, from (4.10–8), for low frequencies and the normal law distribution of the input noise power,

$$W_c(f) = \frac{1}{4\sigma\sqrt{\pi}} h_{02}^2 \psi_0^2 e^{-f^2/4\sigma^2} + \frac{1}{64\sigma\sqrt{2\pi}} h_{04}^2 \psi_0^4 e^{-f^2/8\sigma^2}$$
$$+ \frac{2}{\sigma\sqrt{2\pi}} h_{11}^2 \psi_0 e^{-f^2/2\sigma^2} + \frac{1}{4\sigma\sqrt{6\pi}} h_{13}^2 \psi_0^3 e^{-f^2/6\sigma^2} \quad (4.10\text{–}9)$$
$$+ \frac{1}{4\sigma\sqrt{\pi}} h_{22}^2 \psi_0^2 e^{-f^2/4\sigma^2} + \cdots$$

Although we have been speaking of the ν-law rectifier, equation (4.10–9) gives the low frequency portion of $W_c(f)$, corresponding to a normal law noise power, for any non-linear device provided the proper h_{nk}'s are inserted.

When we set ν equal to one in the expression (4.10–5) for h_{nk} we may obtain the results given by Bennett. Middleton has studied the output of a biased linear rectifier, when the input consists of a sine wave plus noise, and also the special case of the unbiased linear rectifier. He has computed the output for a wide range of the ratios P^2/ψ_0, B^2/ψ_0 where B is the bias. In order to cover the entire range he had to derive two series for the corresponding h_{nk}'s, each series being suitable for its particular portion of the range.

A special case of (4.10–9) occurs when noise alone is applied to a linear rectifier. The low frequency portion of the output power spectrum is

$$W_c(f) = \frac{\psi_0}{\pi} \sum_{m=1}^{\infty} \frac{(-\tfrac{1}{2})_m (-\tfrac{1}{2})_m}{m!\,m!} \frac{1}{\sigma\sqrt{4m\pi}} e^{-f^2/4m\sigma^2}$$

$$= \frac{\psi_0 \pi^{-3/2}}{2\sigma} \left[\tfrac{1}{4} e^{-f^2/4\sigma^2} + \frac{1}{64\sqrt{2}} e^{-f^2/8\sigma^2} + \frac{1}{256\sqrt{3}} e^{-f^2/12\sigma^2} + \cdots \right] \tag{4.10–10}$$

where we have used (4.7–6) and Table 2 of Appendix 4C.

The correlation function of

$$V_s = P \cos pt + Q \cos qt,$$

where p and q are incommensurable, is

$$J_0(P\sqrt{u^2 + v^2 + 2uv \cos p\tau}) \times J_0(Q\sqrt{u^2 + v^2 + 2uv \cos q\tau})$$

From equations (4.9–16) and (4.9–17) it is seen immediately that

$$h_{000} = \frac{1}{2\pi} \int_C F(iu) J_0(Pu) J_0(Qu) e^{-(u^2/2)\psi_0}\, du \tag{4.10–11}$$

is the d.c. component of I when the applied voltage is

$$P \cos pt + Q \cos qt + V_N . \tag{4.1–4}$$

J. R. Ragazzini has obtained an approximate expression for the output power spectrum when the voltage

$$\begin{aligned} V &= V_s + V_N \\ V_s &= Q(1 + r \cos pt)\cos qt \end{aligned} \tag{4.10–12}$$

is impressed on a linear rectifier.[46] In terms of our notation his expression for the continuous portion of the power spectrum is (for low frequencies)

$$W_c(f) = \frac{1}{\pi^2 \alpha^2 (Q^2 + 2\psi_0)} \times \left[\begin{array}{l} W_c(f) \text{ given by equation} \\ (4.5\text{–}17) \text{ for square law device} \end{array} \right] \tag{4.10–13}$$

The α^2 is put in the denominator to cancel the α^2 in the expression (4.5–17). We take the linear rectifier to be

$$I = \begin{cases} 0, & V < 0 \\ V, & 0 < V \end{cases} \tag{4.10–14}$$

and replace the index of modulation, k, in (4.5–17) by r.

[46] Equation (12), "The Effect of Fluctuation Voltages on the Linear Detector," *Proc. I.R.E.*, V. 30, pp. 277–288 (June 1942).

Ragazzini's formula is quite accurate when the index of modulation r is small, especially when $y = Q^2/(2\psi_0)$ is large. To show this we put $r = 0$ in (4.10–13) and obtain

$$W_c(f) = \frac{1}{\pi^2(Q^2 + 2\psi_0)}\left[Q^2 w(f_q - f) + Q^2 w(f_q + f) + \int_{-\infty}^{+\infty} w(x)w(f - x)\,dx\right] \tag{4.10–15}$$

where $f_q = q/(2\pi)$. This is to be compared with the low frequency portion of $W_c(f)$ obtained by specializing (4.10–8) to obtain the output power spectrum of a linear rectifier when the input consists of a sine wave plus noise. The leading terms in (4.10–8) give

$$W_c(f) = h_{11}^2[w(f_q - f) + w(f_q + f)] + h_{02}^2 \frac{1}{4}\int_{-\infty}^{+\infty} w(x)w(f - x)\,dx \tag{4.10–16}$$

The values of the h's appropriate to a linear rectifier are obtained by setting $\nu = 1$ in (4.10–5) and noticing that Q now plays the role of P.

$$\begin{aligned} h_{11} &= \frac{1}{2}\left(\frac{y}{\pi}\right)^{1/2} {}_1F_1(\tfrac{1}{2};\, 2;\, -y) \\ h_{02} &= (2\pi\psi_0)^{-1/2}\, {}_1F_1(\tfrac{1}{2};\, 1;\, -y) \\ y &= Q^2/(2\psi_0) \end{aligned} \tag{4.10–17}$$

Incidentally, the first approximation to the output of a linear rectifier given by (4.10–16) is interesting in its own right. Fig. 9 shows the low frequency portion of $W_c(f)$ as computed from (4.10–16) when the input noise is uniformly distributed over a narrow frequency band of width β, f_q being the mid-band frequency. h_{11} and h_{02} may be obtained from the curves shown in Fig. 10. In these figures P and x replace Q and y of (4.10–17) in order to keep the notation the same as in Fig. 8 for the square law device. These curves may also be obtained from equations (33) to (43) of Bennett's paper.

The following values are useful for our comparison.

When $x = 0$	When x is large
$h_{11} = 0$	$h_{11} = 1/\pi$
$h_{02} = (2\pi\psi_0)^{-1/2}$	$h_{02} = 1/(\pi Q)$.

(4.10–18)

The values for large x are obtained from the asymptotic expansion (4B — 3) given in Appendix 4B.

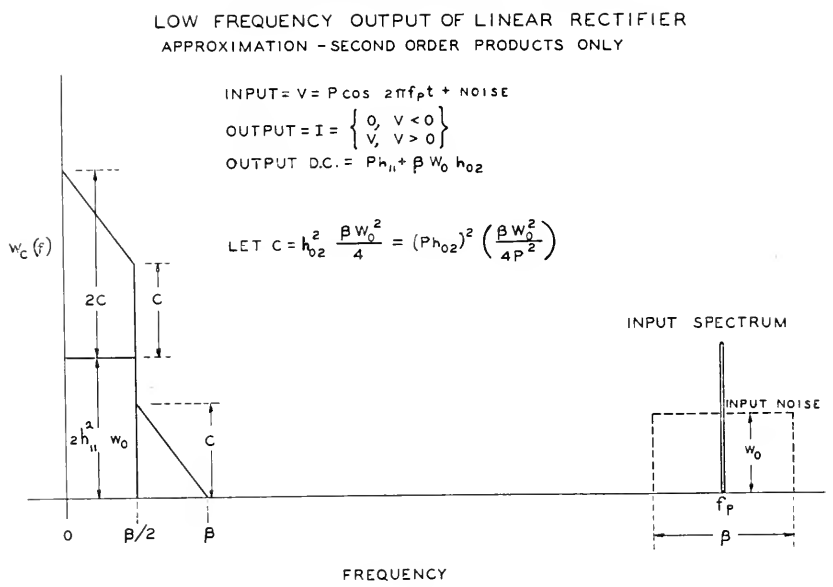


Fig. 9

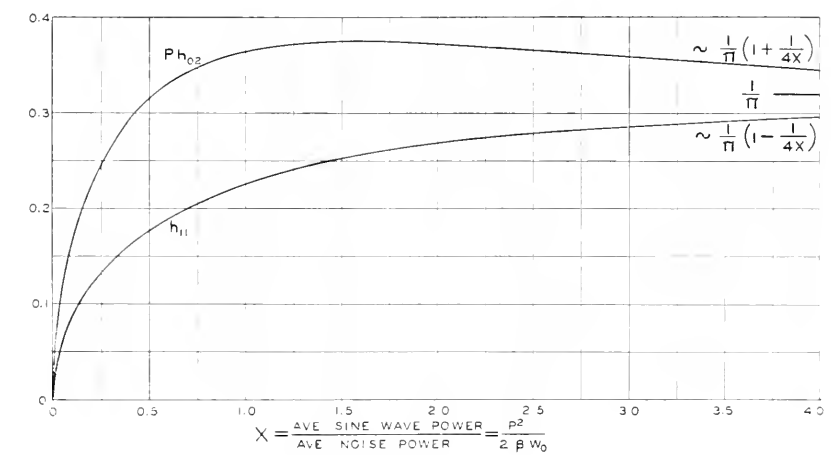


Fig. 10—Coefficients for linear detector output shown on Fig. 9

$$Ph_{02} = \sqrt{\frac{x}{\pi}}\,{}_1F_1(\tfrac{1}{2};\,1;\,-x) \qquad h_{11} = \frac{1}{2}\sqrt{\frac{x}{\pi}}\,{}_1F_1(\tfrac{1}{2};\,2;\,-x)$$

We make the first comparison between (4.10–15) and (4.10–16) by letting $Q \to \infty$. It is seen that both reduce to

$$W_c(f) = \frac{1}{\pi^2}\left[w(f_q - f) + w(f_q + f)\right] \qquad (4.10\text{–}19)$$

which shows that the agreement is perfect in this case. Next we let $Q = 0$. The two expressions then give

$$W_c(f) = \frac{1}{A 2\pi\psi_0} \int_{-\infty}^{+\infty} w(x)w(f - x)\, dx$$

where $A = \pi$ for Ragazzini's formula and $A = 4$ for (4.10–16). Thus the agreement is still quite good. The limiting value for (4.10–16) may also be obtained from (4.7–8).

Even if the index of modulation r is not negligibly small it may be shown that when $Q \to \infty$ $W_c(f)$ still approaches the value given by (4.10–19). Ragazzini's formula gives a somewhat larger answer because it includes the additional terms, shown in (4.5–17), which contain $k^2/4$, but this difference does not appear to be serious. If the $Q^2 + 2\psi_0$ in the denominator of (4.10–13) be replaced by $Q^2 + \frac{1}{2}Q^2k^2 + 2\psi_0$ the agreement is improved.

APPENDIX 4A

Table of Non-linear Devices Specified by Integrals

Quite a number of non-linear devices may be specified by integrals of the form

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

where the function $F(iu)$ and the path of integration C are chosen to fit the device.* The table gives examples of such devices. Some important cases cannot be simply represented in this form. An example is the limiter

$$\begin{aligned} I &= -\alpha D, && V < -D \\ I &= \alpha V, && -D < V < D \\ I &= \alpha D, && D < V \end{aligned} \tag{4A-2}$$

which may be represented as

$$\begin{aligned} I &= \frac{2\alpha}{\pi} \int_0^{\infty} \sin Vu \sin Du \frac{du}{u^2} \\ &= -\alpha D + \frac{2\alpha}{2\pi i} \int_C e^{iVu} \sin Du \frac{du}{u^2} \end{aligned} \tag{4A-3}$$

where C runs from $-\infty$ to $+\infty$ and is indented downward at the origin. This is not of the form assumed in the theory of Part IV. However it appears that it would not be difficult to extend the theory in the particular case of the limiter.

* Reference 50 cited in Section 4.9.

Non-Linear Devices Specified by Integrals

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

I	$F(iu)$	C	Type of Device
$I = \alpha V^n$, n integer	$\frac{\alpha\, n!}{(iu)^{n+1}}$	Positive Loop around $u = 0$	nth power device
$I = \alpha(V - B)^n$, n integer	$\frac{\alpha\, n!}{(iu)^{n+1}} e^{-iuB}$	Positive Loop around $u = 0$	nth power device with bias
$I = 0, \quad V < 0$ $I = \alpha V, \quad 0 < V$	$\frac{\alpha}{(iu)^2} = -\frac{\alpha}{u^2}$	Real u axis from $-\infty$ to $+\infty$ with downward indentation at $u = 0$	Linear rectifier cut-off at $V = 0$
$I = 0, \quad V < B$ $I = \alpha(V - B)^\nu$, $V > B$ ν any positive number	$\frac{\alpha\Gamma(\nu + 1)}{(iu)^{\nu+1}} e^{-iuB}$	"	νth power rectifier with bias
$I = 0, \quad V < 0$ $I = \alpha V, \quad 0 < V < D$ $I = \alpha D, \quad D < V$	$\frac{\alpha(1 - e^{-iuD})}{(iu)^2}$	"	Linear rectifier plus limiter
$I_+ = 0, \quad V < 0$ $I_- = \varphi(V), \quad V > 0$	$F(p) = \int_0^\infty e^{-pt}\varphi(t)\, dt$	"	

APPENDIX 4B

The Function ${}_1F_1(a; c; x)$

In problems concerning a sine wave plus noise the hypergeometric function

$${}_1F_1(a; c; z) = 1 + \frac{az}{c1!} + \frac{a(a + 1)}{c(c + 1)} \frac{z^2}{2!} + \cdots \tag{4B-1}$$

arises. Here we state some of its properties which are of use in the theory of Part IV. Curves of ${}_1F_1(a; c; z)$ are given for $a = -4, -3.5 \cdots, 3.5, 4.0$ and $c = -1.5, -.5, +.5, 1, 1.5, 2, 3, 4$ in the 1938 edition, page 275, of "Tables of Functions", by Jahnke and Emde. A list of properties of the function and other references are also given. In addition to these references we mention E. T. Copson, "Functions of a Complex Variable" (Oxford, 1935), page 260.

If c is not a negative integer or zero

$${}_1F_1(a; c; z) = e^z {}_1F_1(c - a; c; -z). \tag{4B-2}$$

When $R(z) > 0$ we have the asymptotic expansions

$$
\begin{aligned}
{}_1F_1(a; c; z) &\sim \frac{\Gamma(c)e^z}{\Gamma(a)z^{c-a}}\left[1 + \frac{(1-a)(c-a)}{1!z}\right. \\
&\quad \left. + \frac{(1-a)(2-a)(c-a)(c-a+1)}{2!z^2} + \cdots\right] \\
{}_1F_1(a; c; -z) &\sim \frac{\Gamma(c)}{\Gamma(c-a)z^a}\left[1 + \frac{a(1+a-c)}{1!z}\right. \\
&\quad \left. + \frac{a(a+1)(1+a-c)(2+a-c)}{2!z^2} + \cdots\right]
\end{aligned}
\tag{4B-3}
$$

Many of the hypergeometric functions encountered may be expressed in terms of Bessel functions of the first kind for imaginary argument. The connection may be made by means of the relation[51]

$${}_1F_1\left(\nu + \frac{1}{2}; 2\nu + 1; z\right) = 2^{2\nu}\Gamma(\nu+1)z^{-\nu}e^{z/2}I_\nu\left(\frac{z}{2}\right) \tag{4B-4}$$

together with the recurrence relations

	F_{a+}	F_{a-}	F_{c+}	F_{c-}	F
1.	a	$(a-c)$			$c - 2a - z$
2.	ac		$(c-a)z$		$-c(a+z)$
3.	a			$1-c$	$c - a - 1$
4.		$-c$	$-z$		c
5.		$a-c$		$c-1$	$1 - a - z$
6.			$(c-a)z$	$c(c-1)$	$c(1-c-z)$

For example, the first recurrence relation is obtained from line 1 as follows

$$
\begin{aligned}
&aF(a+1; c; z) + (a-c)F(a-1; c; z) \\
&\qquad + (c - 2a - z)F(a; c; z) = 0
\end{aligned}
\tag{4B-5}
$$

These six relations between the contiguous ${}_1F_1$ functions are analogous to the 15 relations, given by Gauss, between the contiguous ${}_2F_1$ hypergeometric functions and may be derived from these by using

$${}_1F_1(a; c; z) = \operatorname*{Limit}_{b\to\infty} {}_2F_1\left(a, b; c; \frac{z}{b}\right) \tag{4B-6}$$

A recurrence relation involving two ${}_1F_1$'s of the type (4B-4) may be obtained by replacing a by $a+1$ in the relation given by row four of the table

[51] G. N. Watson, "Theory of Bessel Functions" (Cambridge, 1922), p. 191.

and then eliminating ${}_1F_1(a+1;c;z)$ from this relation and the one obtained from row 3 of the table. There results

$$ {}_1F_1(a;c;z) = {}_1F_1(a;c-1;z) + \frac{za}{c(1-c)} F(a+1;c+1;z) \qquad (4B\text{–}7) $$

Setting ν equal to zero and one in (4B–4) and a equal to $\frac{1}{2}$, c equal to 2 in (4B–7) gives

$$ \begin{aligned} {}_1F_1\left(\frac{1}{2};1;z\right) &= e^{z/2} I_0\left(\frac{z}{2}\right) \\ {}_1F_1\left(\frac{3}{2};3;z\right) &= 4z^{-1}e^{z/2} I_1\left(\frac{z}{2}\right) \\ {}_1F_1\left(\frac{1}{2};2;z\right) &= e^{z/2}\left[I_0\left(\frac{z}{2}\right) - I_1\left(\frac{z}{2}\right)\right] \end{aligned} \qquad (4B\text{–}8) $$

Starting with these relations the relations in the table enable us to find an expression for ${}_1F_1(n+\frac{1}{2};m;z)$ where n and m are integers. A number of these are given in Bennett's paper. In particular, using (4B–2),

$$ {}_1F_1\left(-\frac{1}{2};1;-z\right) = e^{-z/2}\left[(1+z)I_0\left(\frac{z}{2}\right) + zI_1\left(\frac{z}{2}\right)\right]. \qquad (4B\text{–}9) $$

APPENDIX 4C

The Power Spectrum Corresponding to ψ_τ^n

Quite often we encounter the integral

$$ G_n(f) = \int_0^\infty [\psi(\tau)]^n \cos 2\pi f\tau \, d\tau \qquad (4C\text{–}1) $$

where $\psi(\tau)$ is the correlation function corresponding to the power spectrum $w(f)$. From the fundamental relation between $w(f)$ and $\psi(\tau)$ given by (2.1–5),

$$ G_1(f) = \frac{w(f)}{4} \qquad (4C\text{–}2) $$

The expression for the spectrum of the product of two functions enables us to write $G_n(f)$ in terms of $w(f)$. We shall use the following form of this expression: Let $F_r(f)$ be the spectrum of the function $\varphi_r(\tau)$ so that

$$ \begin{aligned} \varphi_r(\tau) &= \int_{-\infty}^{+\infty} F_r(f) e^{2\pi i f\tau} \, df, \quad r = 1, 2 \\ F_r(f) &= \int_{-\infty}^{+\infty} \varphi_r(\tau) e^{-2\pi i f\tau} \, dt \end{aligned} $$

Then

$$\int_{-\infty}^{+\infty} \varphi_1(\tau)\varphi_2(\tau)e^{-2\pi i f\tau}\,d\tau = \int_{-\infty}^{+\infty} F_1(x)F_2(f - x)\,dx \qquad (4C\text{–}3)$$

i.e., the spectrum of the product $\varphi_1(\tau)\varphi_2(\tau)$ is the integral on the right. If $\varphi_1(\tau)$ and $\varphi_2(\tau)$ are real even functions of τ, (4C–3) may be written as

$$\int_0^{\infty} \varphi_1(\tau)\varphi_2(\tau)\cos 2\pi f\tau\,d\tau = \frac{1}{2}\int_{-\infty}^{+\infty} F_1(x)F_2(f - x)\,dx \qquad (4C\text{–}4)$$

In order to obtain $G_2(f)$ we set $\varphi_1(\tau)$ and $\varphi_2(\tau)$ equal to $\psi(\tau)$. We may then use (4C-4) since $\psi(\tau)$ is an even real function of τ. When $\varphi_r(\tau)$ is an even real function of τ we see, from the Fourier integral for $F_r(f)$, that $F_r(f)$ must be an even real function of f. We therefore set

$$2F_r(f) = w(f), \qquad r = 1, 2$$

and define $w(f)$ for negative f by

$$w(-f) = w(f) \qquad (4C\text{–}5)$$

Equation (4C–4) then gives

$$\begin{aligned} G_2(f) &= \frac{1}{8}\int_{-\infty}^{+\infty} w(x)w(f - x)\,dx \\ &= \frac{1}{8}\int_0^{f} w(x)w(f - x)\,dx \\ &\quad + \frac{1}{4}\int_0^{\infty} w(x)w(f + x)\,dx \end{aligned} \qquad (4C\text{–}6)$$

where in the second equation only positive values of the argument of $w(f)$ appear.

In order to get $G_3(f)$ we set $\varphi_1(\tau)$ equal to $\psi(\tau)$, $2F_1(f)$ equal to $w(f)$, and $\varphi_2(\tau)$ equal to $\psi^2(\tau)$. Then

$$\begin{aligned} F_2(f) &= 2\int_0^{\infty} \varphi_2(\tau)\cos 2\pi f\tau\,d\tau \\ &= 2G_2(f) \end{aligned}$$

and from (4C–4) we obtain

$$\begin{aligned} G_3(f) &= \frac{1}{2}\int_{-\infty}^{+\infty} w(x)G_2(f - x)\,dx \\ &= \frac{1}{16}\int_{-\infty}^{+\infty} w(x)\,dx \int_{-\infty}^{+\infty} w(y)w(f - y)\,dy \end{aligned} \qquad (4C\text{–}7)$$

Equation (4C–7) suggests that we may write the expression for $G_2(f)$ as

$$G_2(f) = \frac{1}{2}\int_{-\infty}^{+\infty} w(x)G_1(f - x)\,dx \tag{4C–8}$$

This is seen to be true from (4C–2) and (4C–6). In fact it appears that

$$G_n(f) = \frac{1}{2}\int_{-\infty}^{+\infty} w(f - x)G_{n-1}(x)\,dx \tag{4C–9}$$

might be used for a step by step computation of $G_n(f)$.

We now consider $G_n(f)$ for the case of relatively narrow band pass filters. As examples we take filters whose characteristics give the following $w(f)$'s and $\psi(\tau)$'s

TABLE 1

Filter	$w(f)$ for $f > 0$	$\psi(\tau)$
a	$\frac{\psi_0}{\sigma\sqrt{2\pi}}\, e^{-(f-f_0)^2/2\sigma^2}$	$\psi_0 e^{-2(\pi\sigma\tau)^2} \cos 2\pi f_0\tau$
b	$\frac{\psi_0\,\alpha}{\pi}\,\frac{1}{\alpha^2 + (f - f_0)^2}$	$\psi_0 e^{-2\pi\alpha\lvert\tau\rvert} \cos 2\pi f_0\tau$
c	$w(f) = w_0 = \psi_0/\beta$ for $f_0 - \frac{\beta}{2} < f < f_0 + \frac{\beta}{2}$; $w(f) = 0$ elsewhere	$\psi_0 \frac{\sin \pi\beta\tau}{\pi\beta\tau} \cos 2\pi f_0\tau$

We shall refer to these filters as Filter a, Filter b, and Filter c, respectively. All have f_0 as the mid-frequency of the pass band. The constants have been chosen so that they all pass the same average power when a wide band voltage is applied:

$$\psi_0 = \int_0^\infty w(f)\,df = \text{mean square value of } I(t) \text{ or } V(t)$$

and it is assumed that $f_0 \gg \sigma$, $f_0 \gg \alpha$, $f_0 \gg \beta$ so that the pass bands are relatively narrow.

Expressions for $G_n(f)$ corresponding to several values of n are given in Table 2. When $n = 1$, $G_1(f)$ is simply $w(f)/4$. $G_2(f)$ is obtained by setting $n = 2$ in the definition (4C–1) for $G_n(f)$, squaring the $\psi(\tau)$'s of Table 1, and using

$$\cos^2 2\pi f_0\tau = \tfrac{1}{2} + \tfrac{1}{2}\cos 4\pi f_0\tau$$

TABLE 2

$G_n(f)$	Filter a	Filter b
$G_1(f)$	$\frac{\psi_0}{4\sigma\sqrt{2\pi}} e^{-(f-f_0)^2/2\sigma^2}$	$\frac{\alpha\psi_0}{4\pi} \frac{1}{\alpha^2 + (f-f_0)^2}$
$G_2(f)$	$\frac{\psi_0^2}{8\sigma\sqrt{4\pi}} [2e^{-f^2/4\sigma^2} + e^{-(f-2f_0)^2/4\sigma^2}]$	$\frac{2\alpha\psi_0^2}{8\pi}\left[\frac{2}{4\alpha^2+f^2} + \frac{1}{4\alpha^2+(f-2f_0)^2}\right]$
$G_3(f)$	$\frac{\psi_0^3}{16\sigma\sqrt{6\pi}} [3e^{-(f-f_0)^2/6\sigma^2} + e^{-(f-3f_0)^2/6\sigma^2}]$	$\frac{3\alpha\psi_0^3}{16\pi}\left[\frac{3}{9\alpha^2+(f-f_0)^2} + \frac{1}{9\alpha^2+(f-3f_0)^2}\right]$
$G_4(f)$	$\frac{\psi_0^4}{32\sigma\sqrt{8\pi}} [6e^{-f^2/8\sigma^2} + 4e^{-(f-2f_0)^2/8\sigma^2} + e^{-(f-4f_0)^2/8\sigma^2}]$	$\frac{4\alpha\psi_0^4}{32\pi}\left[\frac{6}{16\alpha^2+f^2} + \frac{4}{16\alpha^2+(f-2f_0)^2} + \frac{1}{16\alpha^2+(f-4f_0)^2}\right]$
$G_n(f)$ n odd f small	0	0
$G_n(f)$ n even f small	$\frac{\psi_0^n n!}{\left(\frac{n}{2}\right)!\left(\frac{n}{2}\right)!} \frac{e^{-f^2/2n\sigma^2}}{2^{n+1}\sigma\sqrt{2n\pi}}$	$\frac{\psi_0^n n!}{\left(\frac{n}{2}\right)!\left(\frac{n}{2}\right)! 2^n} \frac{1}{2^{n+1}\pi n\alpha} \frac{1}{1+\left(\frac{f}{n\alpha}\right)^2}$
$G_n(f)$ n even n large f small	$\frac{1}{2\pi\sigma n} e^{-f^2/2n\sigma^2}$	$\frac{2}{\alpha(2\pi n)^{3/2}} \frac{1}{1+\left(\frac{f}{n\alpha}\right)^2}$
Filter c $G_1(f)$	$\frac{\psi_0}{4\beta}$ when $f_0 - \frac{\beta}{2} < f < f_0 + \frac{\beta}{2}$ 0 elsewhere	
Filter c $G_2(f)$		$\frac{\psi_0^2}{4\beta}\left(1 - \frac{f}{\beta}\right)$ when $0 \le f \le \beta$ $\frac{\psi_0^2}{8\beta^2}(f - 2f_0 + \beta)$ " $2f_0 - \beta \le f \le 2f_0$ $\frac{\psi_0^2}{8\beta^2}(2f_0 + \beta - f)$ " $2f_0 \le f \le 2f_0 + \beta$

The expression for $G_2(f)$ given in Table 2 corresponding to Filter c is exact. The expressions for Filters a and b give good approximations around $f = 0$ and $f = 2f_0$ where $G_2(f)$ is large. However, they are not exact because terms involving $f + 2f_0$ have been omitted. It is seen that all three G_2's behave in the same manner. Each has a peak symmetrical about $2f_0$ whose width is twice that of the original $w(f)$, is almost zero between 0 and $2f_0$, and rises to a peak at 0 whose height is twice that at $2f_0$.

$G_3(f)$ is obtained by cubing the $\psi(\tau)$ given in Table 1 and using

$$\cos^3 2\pi f_0\tau = \tfrac{3}{4}\cos 2\pi f_0\tau + \tfrac{1}{4}\cos 6\pi f_0\tau.$$

From the way in which the cosine terms combine with $\cos 2\pi f\tau$ in (4C–1) we see that $G_3(f)$, for our relatively narrow band pass filters, has peaks at f_0 and $3f_0$, the first peak being three times as high as the second. The expressions given for $G_3(f)$ and $G_4(f)$ are approximate in the same sense as are those for $G_2(f)$. It will be observed that the coefficients within the brackets, for Filters a and b, are the binomial coefficients for the value of n concerned. Thus for $n = 2$, they are 2 and 1, for $n = 3$ they are 3 and 1, and for $n = 4$ they are 6, 4, and 1.

The higher $G_n(f)$'s for Filters a and b may be computed in the same way. The integrals to be used are

$$\int_0^\infty e^{-2n(\pi\sigma\tau)^2}\cos 2\pi f\tau\, d\tau = \frac{e^{-f^2/2n\sigma^2}}{2\sigma\sqrt{2n\pi}}$$

$$\int_0^\infty e^{-2n\pi\alpha\tau}\cos 2\pi f\tau\, d\tau = \frac{1}{2\pi}\frac{n\alpha}{n^2\alpha^2 + f^2}$$

In many of our examples we are interested only in the values $G_n(f)$ for f near zero, i.e., only in that peak which is at zero. It is seen that $G_n(f)$ has such a peak only when n is even, this peak arising from the constant term in the expansion

$$\cos^{2k} x = \frac{1}{2^{2k-1}}\left[\cos 2kx + 2k\cos 2(k-1)x + \frac{(2k)(2k-1)}{2!}\cos 2(k-2)x + \cdots + \frac{(2k)!}{(k-1)!(k+1)!}\cos 2x + \frac{(2k)!}{k!k!2}\right]$$

Abstracts of Technical Articles by Bell System Authors

Historical Background of Electron Optics.[1] C. J. Calbick. The discovery of electron optics resulted from studies of the action, upon electrons or other charged particles, of electric and magnetic fields employed for the purpose of obtaining sharply defined beams. The original Braun tube (1896) employed gas-focusing, as did the low-voltage cathode-ray oscilloscope developed by Johnson in 1920. It was early discovered that an axial magnetic field could be used to concentrate the electrons into a beam, and this method came into wide use in the field of high-voltage cathode-ray oscillography. In 1927 Busch published a theoretical study of the action of an axially-symmetric magnetic field upon paraxial electrons, showing that the equation of the trajectories of the electrons was similar to that of the paths of light rays through an axially symmetric optical system. He concluded that such magnetic fields constituted lenses for electrons and presented experimental confirmation. In 1931 Knoll and Ruska presented a large amount of additional experimental material and used the words "electron optics" to describe the analogy. In 1932 Bruche and Johannson published the first electron micrographs.

The Davisson and Germer electron diffraction experiments (1927) employed electron beams formed by electron guns consisting of a thermionic cathode emitting electrons which were accelerated by potentials applied to a series of plates containing aligned apertures. The resultant beam was quite divergent. Davisson and Calbick made a theoretical and experimental study of the forms of such beams. They concluded that the distorted electric field in the vicinity of an aperture in a charged plate constituted a lens for charged particles (1931). The optical analogy was either a cylindrical or a spherical lens, according as the aperture was a slit or a circular hole. The theory was confirmed by photographing the forms of electron beams, and by construction of an electrostatic electron microscope whose experimental magnification agreed with the theoretical.

Coaxial Cables and Associated Facilities.[2] J. J. Pilliod. (*Summary of Talk before St. Louis Electrical Board of Trade*, October 17, 1944.) Coaxial cables provide means of transmitting frequency bands several million cycles in width over a metal tube a little larger than a lead pencil, with a copper wire extending along its axis. Several of these tubes can be placed in a lead sheath.

The frequency band transmitted over coaxial cables may be split up so as to provide several hundred telephone circuits or, without such division,

[1] *Jour. Applied Physics*, October 1944.
[2] *FM and Television*, November 1944.

coaxial cables will provide for broad-band transmission service such as is required for television.

A cable is now being installed between Terre Haute and St. Louis which contains six coaxial tubes to provide telephone circuits, and which may, in the future, find use in connection with the provision of intercity television networks.

The structure of the tubes used with coaxial cables consists of a central copper conductor within a copper tube about $\frac{1}{4}$ in. in diameter, made from flat copper strip which is formed around the insulating discs. Around each copper tube are two steel tapes which supplement the shielding of the copper tube in preventing interference between tubes in close proximity. The central conductor is separated from the outer conductor by slotted insulating disks which are forced onto the wire. The cables are formed with an appropriate number of these tubes along with some small gauge pairs used for control and operating purposes.

In the case of underground cables buried directly in the earth, jute or plastic protective coverings are used to assist in reducing sheath corrosion. In some parts of the country it is essential to add a metal covering outside the lead sheath and the plastic or jute to protect the cables against the operations of ground squirrels or pocket gophers. In certain areas these animals have been found to carry away long sections of the jute covering and will chew holes in the lead sheath unless other metal protection is provided. Copper is sometimes used for this metal covering to assist in lightning protection.

Repeaters in the coaxial system are now located at intervals of about five miles. Power for repeaters in the auxiliary stations is supplied from the adjacent main stations located at something over 50 miles at 60 cycles over the coaxial conductors themselves.

Coaxial cables are in regular operation between New York and Philadelphia and between Minneapolis and Stevens Point, Wisconsin, a total distance of nearly 300 miles. A network of such cables totaling about 7,000 route miles and including a second transcontinental cable route is being planned over additional routes. The requirements of the armed forces, general business conditions, the volume and distribution of long distance telephone messages, the availability of the necessary manufactured cable and equipment, and other factors may modify the extent of this construction, the time of starting, and the routes which will be undertaken.

Western Electric Recording System—U. S. Naval Photographic Science Laboratory.[3] R. O. Strock and E. A. Dickinson. This paper describes the complete 35-mm film and $33\frac{1}{3}$ or 78 rpm. disk recording and re-recording equipment installed for the U. S. Navy at the Photographic Science Laboratory, Anacostia, D. C. Modern design, excellent performance, and ease of operation are features of the installation.

[3] *Jour. Soc. Motion Picture Engineers,* December 1944.

Contributors to this Issue

William A. Edson, Kansas University, B.S. 1934; M.S. 1935. Harvard University, D.S. 1937. Bell Telephone Laboratories, 1937–1941 and 1943–. Assistant Professor of Electrical Engineering 1941–1942 at Illinois Institute of Technology, Chicago. Prior to 1941 Dr. Edson was concerned with carrier telephone terminal devices. At the present time he is engaged full time on war projects.

Richard C. Eggleston, Ph.B. 1909 and M.F. 1910, Yale University; U. S. Forest Service, 1910–1917; Pennsylvania Railroad, 1917–1920; First Lieutenant, Engineering Div., Ordnance Dept., World War I, 1918-1919; American Telephone and Telegraph Company, 1920–1927; Bell Telephone Laboratories, 1927–. Mr. Eggleston has been engaged chiefly with problems relating to the strength of timber and with statistical investigations in the timber products field.

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.

VOLUME XXIV APRIL, 1945 NUMBER 2

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per copy *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are 50 cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

CORRECTIONS FOR ISSUE OF APRIL, 1945

Page 207: Abscissa for Fig. 12.29, $P = \dfrac{M}{1 - c_t/c_0}$

should be $P = \dfrac{M}{1 + c_t/c_0}$

Page 240: Equation (15.60), $W_0 = \dfrac{\beta}{\delta}\left[\dfrac{\alpha}{\beta} - \dfrac{\bar{\gamma}}{\bar{\delta}}\right] \Big/ \left[\dfrac{\gamma}{\delta} + \dfrac{\bar{\gamma}}{\bar{\delta}}\right]$

should be $W_0 = \dfrac{\beta}{\delta}\left[\dfrac{\alpha}{\beta} + \dfrac{\bar{\gamma}}{\bar{\delta}}\right] \Big/ \left[\dfrac{\gamma}{\delta} + \dfrac{\bar{\gamma}}{\bar{\delta}}\right]$

The Bell System Technical Journal

Vol. XXIV *April, 1945* *No. 2*

Piezoelectric Crystals in Oscillator Circuits

By I. E. FAIR

12.00 Introduction

A STUDY or an explanation of the performance of a piezoelectric crystal in an oscillator circuit involves a study or explanation of oscillator circuits in general and a study of the crystal as a circuit element. Nicolson[1] appears to have been the first to discover that a piezoelectric crystal had sufficient coupling between electrical electrodes and mechanical vibratory movement so that when the electrodes were suitably connected to a vacuum tube circuit, sustained oscillations were produced. In such an oscillator the mechanical oscillatory movement of the crystal functions as does the electrical oscillatory circuit of the usual vacuum tube oscillator. His circuit is shown in Fig. 12.1. Cady[2] independently though later made the same discovery, but he utilized it somewhat differently and expressed it differently. He found that when the electrodes of a quartz crystal are connected in certain ways to an electric oscillator circuit, the frequency is held very constant at a value which coincides with the period of the vibrating crystal. He made the further discovery that due to the very sharp resonance properties of the quartz crystal, the constancy in frequency to be secured was far greater than could be obtained by any purely electric oscillator.

The development of analytical explanations of the crystal controlled oscillator came along rather slowly. Cady explained the control in terms of operation upon the electrical oscillator to which the crystal was attached. He said that the "capacity" of the crystal changes rapidly with frequency in the neighborhood of mechanical resonance, even becoming negative. This "capacity" connected across the oscillator tuned circuit or in other places prevented the frequency from changing to any extent, as any frequency change caused such a "capacity" change in the crystal as to tend to tune the circuit in the other direction. Cady, however, devised one circuit, Fig. 12.2, in which no tuned electrical circuit was used, but he confined his explanation to "a mechanically tuned feedback path from the plate to the grid of the amplifier". Pierce[3] came along later with a two-electrode crystal connected between plate and grid, and no tuned circuit, and also with a

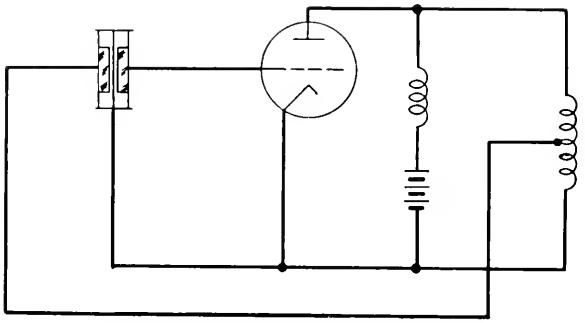


Fig. 12.1—Nicolson's crystal oscillator circuit

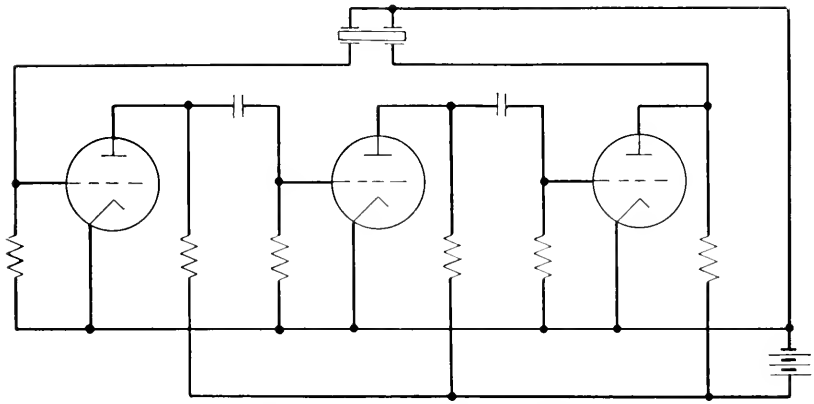


Fig. 12.2—Cady's oscillator circuit using a crystal as a "mechanically tuned feedback path"

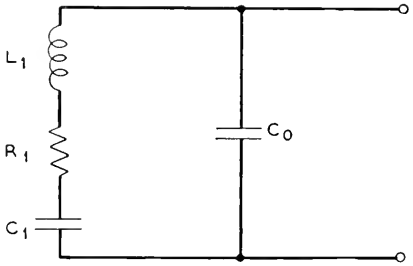


Fig. 12.3—Equivalent electrical circuit of a piezoelectric crystal near its resonant frequency

two-electrode crystal connected between grid and cathode and no tuned circuit, where the operation would not be satisfactorily explained by Cady's method. His circuits would require the crystal to exhibit inductive react-

ance, rather than the capacitance Cady spoke of. Miller[4] also produced a circuit with a two-electrode crystal connected between grid and cathode but with a tuned circuit in the plate lead, which circuit required the crystal to provide inductive reactance.

It was not until after Van Dyke[5] showed that the crystal could be represented by the circuit network of Fig. 12.3 that it was possible to explain these various phenomena. With this view of the crystal, and using the differential equation method of circuit analysis, Terry[6] pointed out that, as with electrical oscillators, the frequency is not completely governed by the resonant element, in this case the crystal, but is influenced somewhat by the circuit elements. The circuit as a whole is quite complex and the equations are difficult to use. Wright[7] and Vigoureux[8] also made analyses of the Pierce type oscillator. Because of the complexity of the equations, the frequency, amplitude, or activity are not computed directly, but the effects of the circuit variables are analyzed in a qualitative manner and the results compared with experimental data.

Oscillators employing crystals may be classified in a number of ways. One classification is based upon whether or not the circuit without the crystal is in itself an oscillator. If it is, the oscillator is called a "crystal controlled" oscillator. If it is not, it is called a "crystal" oscillator. All of Cady's oscillator circuits, except the one shown in Fig. 12.2, are of the first named class. This type of circuit will oscillate at a frequency determined by the tuned circuit if the crystal becomes broken or disconnected, or if high resistance develops in the crystal, or if the electric tuned circuit should become tuned too far from the resonant frequency of the crystal. This property at times is an advantage and at other times a disadvantage. This type of circuit will oscillate under control of the crystal with much less active crystals than most of the other types.

Nicolson's, Pierce's, Cady's of Fig. 12.2 and Miller's oscillators belong to the second named class. They will cease oscillating if the crystal breaks, develops high resistance or is disconnected. Failure of the oscillator to function at all then serves as a warning that something has happened to the crystal.

This second named class of crystal oscillators has been used much more than the first named. The crystal is the principal frequency determining element in the circuit. Often there are required only resistances, or resistances and an inductance, as the other elements to embody along with the vacuum tube and crystal. The simplicity, low costs, and usually no tuning, have made this class attractive. Most analytical studies of oscillator circuits have been made upon this class. For that reason the discussion in this chapter will be limited to this class.

An analytic study of the crystal oscillator can readily start by looking

upon the oscillator as consisting only of inductances, capacitances, and resistances, along with the vacuum tube. The crystal is replaced by the proper circuital elements arranged as in Fig. 12.3. This circuit or equivalent of the crystal is that of a series resonant circuit having capacitance paralleling it. The circuit will show both phenomena of series resonance and parallel resonance, the two frequencies being very close together. By making suitable measurements on a crystal, the magnitudes of the inductance, resistance, and the two capacitances can be determined. It is usually found that the series inductance is computed as hundreds or thousands of henries, and the series capacitance is a small fraction of a micromicrofarad. The magnitudes of the inductance and capacitance are beyond what it is possible to construct in the usual forms of building inductances and capacitances. This accounts for its superior frequency control properties.

Although reducing the crystal to an equivalent electrical circuit provides one notable step in understanding the performance of the crystal oscillator, it does not readily lead to a full understanding. The electric oscillator in itself is not fully and completely analyzed in all its ramifications, although it has been under study for over 25 years. These studies have been mathematical and experimental in character, but in all cases it appears there have been approximations of some kind, made because the variable impedance characteristics both of the plate circuit and the grid circuit of the tubes did not lend themselves readily to a rigorous analysis. The earlier investigations assumed a linear relation between grid voltage and plate current and assumed constant plate impedance. Later investigations brought in further elements and further variables, the different investigators attacking the problem in different ways and attempting to prove different points. By this means a large number of factors in oscillators have been ascertained to a first degree of approximation so that a qualitative review of the performance of the electric oscillator is very well known. It is the quantitative view upon the first order magnitude which is still difficult or uncertain. This is particularly true of the crystal oscillator because of the slightly different circuit.

It is proposed, therefore, in this paper to cover briefly a number of the studies on crystal oscillators so as to point out the different modes of attack and the different behavior points in the oscillators which the various investigators have studied. After covering these points, there will be discussed the frequency control properties of the crystal and the frequency stability of crystal oscillators. The performance of the crystal in the oscillator with respect to activity is then treated. There will be introduced two new yardsticks for measuring or indicating crystal quality, one called "figure of merit" and the other called "performance index." These are related to

the crystal constants and paralleling capacitances which are usually involved. They will be defined and their method of use and application in oscillators will be pointed out.

12.10 Solution by Differential Equations

The most direct method of determining the oscillating conditions in a circuit is to analyze the differential equation for the current in some particular branch of the circuit. The relations existing between the coefficients determine whether the current builds up, dies out, or is maintained at a constant value and frequency. Unfortunately the equations resulting from the application of this method to the crystal oscillator circuit are quite complicated. However, lower order differential equations result from the

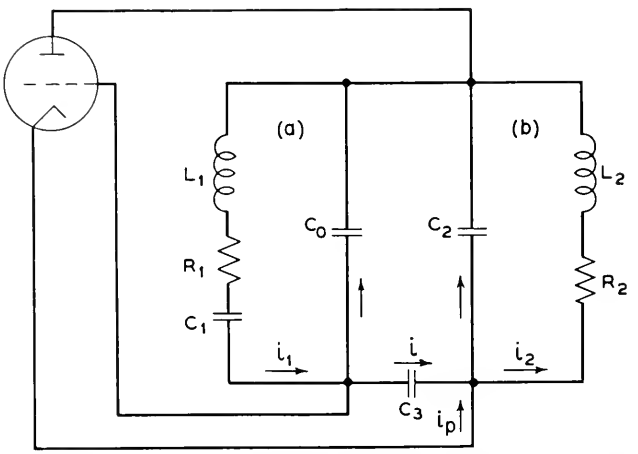


Fig. 12.4—Equivalent circuit of oscillator with crystal connected between grid and plate

application of this method to similar electric oscillator circuits, and certain qualitative information obtained from the latter is applicable to crystal oscillators. Thus Heising's[9] analysis of the Colpitts and Hartley circuits gives much information directly applicable to the Pierce and Miller types of crystal oscillators. From this the circuit conditions necessary for oscillations to exist and the effect of certain circuit variables upon the frequency are ascertained. The more complex qualitative view is given by Terry[6] who shows the relations of the coefficients of linear differential equations of the 2nd, 3rd, and 4th orders, and applies them to the analysis of three common types of crystal oscillator circuits. The resulting equations, together with certain qualitative information regarding their interpretation, are repeated here. In making this analysis the grid current is disregarded and the static tube characteristic is considered linear.

The equation is the same for the three types of circuits considered and is derived for the current i_1, in Figs. 12.4 and 12.5, although it may be set up in terms of any of the currents or voltages existing in the circuit. It is of the form

$$\frac{d^4 i_1}{dt^4} + P_1 \frac{d^3 i_1}{dt^3} + P_2 \frac{d^2 i_1}{dt^2} + P_3 \frac{di_1}{dt} + P_4 i_1 = 0 \qquad (12.1)$$

The P coefficients are functions of the circuit elements and are defined for each type of circuit in the following sections.

The solution of (12.1) normally represents a doubly periodic function arising from the two coupled antiresonant meshes (a) and (b). The normal

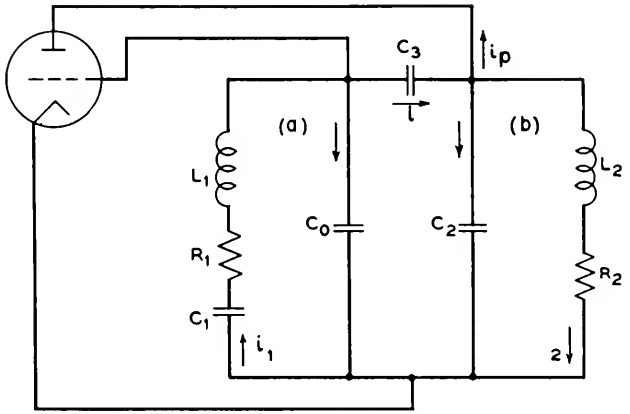


Fig. 12.5—Equivalent circuit of oscillator with crystal connected between grid and cathode

modes of oscillation consist of two currents in each mesh with frequency and damping factors β_1 and α_1, β_2 and α_2 respectively.

The conditions for undamped oscillation as derived from the general equation (12.1) are expressed in terms of the coefficients by

$$\frac{P_3}{P_1} = \frac{P_2 \pm \sqrt{P_2^2 - 4P_4}}{2} \qquad (12.2)$$

and the angular frequencies are

$$\beta^2 = \frac{P_2 \pm \sqrt{P_2^2 - 4P_4}}{2} \qquad (12.3)$$

where the plus sign gives the condition for one damping factor to be zero and the minus sign that for the other to be zero.

The frequency at which oscillations are maintained is determined by the required phase relation of voltages applied to the tube. With crystal from grid to plate, as in Fig. 12.4, the phase difference of grid and plate voltages is such that the circuit oscillates at only one of the normal modes, and with crystal connected between grid and cathode, as in Fig. 12.5, it oscillates at the other only.

12.11 Crystal Between Grid and Plate

With the crystal connected between the grid and plate of the tube, as in Fig. 12.4, the coefficients of the general equation (12.1) are

$$\left.\begin{aligned}
P_1 &= \frac{R_1}{L_1} + \frac{R_2}{L_2} + \frac{1}{R_p C_b'} \\
P_2 &= \frac{1}{L_1 C_a} + \frac{R_1 R_2}{L_1 L_2} + \frac{1}{L_2 C_b} + \frac{1}{R_p}\left(\frac{R_1}{L_1} + \frac{R_2}{L_2}\right)\frac{1}{C_b'} \\
P_3 &= \frac{R_2}{L_1 L_2 C_a} + \frac{R_1}{L_1 L_2 C_b} + \frac{1}{R_p}\left(\frac{1}{L_1 C_a C_b'} + \frac{R_1 R_2}{L_1 L_2 C_b'} - \frac{1}{L_1 C_m C_m'}\right) \\
P_4 &= \frac{1}{L_1 L_2 C_a C_b} - \frac{1}{L_1 L_2 C_m^2} + \frac{R_2}{R_p}\left(\frac{1}{L_1 L_2 C_a C_b'} - \frac{1}{L_1 L_2 C_m C_m'}\right)
\end{aligned}\right\} \quad (12.4)$$

where

$$\frac{1}{C_a} = \frac{1}{C_1} + \frac{1}{C_0} - \frac{C_x}{C_0^2} \qquad \frac{1}{C_m} = \frac{C_x}{C_2 C_0}$$

$$\frac{1}{C_b} = \frac{1}{C_2} - \frac{C_x}{C_2^2} \qquad \frac{1}{C_m'} = \frac{1}{C_m} - \frac{\mu C_x}{C_0 C_3}$$

$$\frac{1}{C_b'} = \frac{1}{C_b} + \frac{\mu C_x}{C_2 C_3} \qquad \frac{1}{C_x} = \frac{1}{C_0} + \frac{1}{C_2} + \frac{1}{C_3}$$

μ = the amplification factor of the tube.

$$R_p = \frac{\partial e_p}{\partial i_p} \quad (e_g \text{ constant})$$

The uncoupled damping factors, α_a and α_b, the uncoupled undamped angular frequencies, β_a and β_b, and the coupling coefficient τ may be introduced as follows:

$$\alpha_a = \frac{R_1}{2L_1}, \qquad \beta_a^2 = \frac{1}{L_1 C_a} \qquad \tau^2 = \frac{C_a C_b}{C_m^2}$$

$$\alpha_b = \frac{R_2}{2L_2}, \qquad \beta_b^2 = \frac{1}{L_2 C_b}$$

Note that C_a is the total capacitance across L_1 and R_1, and C_b is the total capacitance across L_2 and R_2.

The coefficients of (12.4) become

$$\left.\begin{aligned} P_1 &= 2(\alpha_a + \alpha_b) + \frac{1}{R_p C_b'} \\ P_2 &= \beta_a^2 + 4\alpha_a\alpha_b + \beta_b^2 + \frac{1}{R_p}(\alpha_a + \alpha_b)\frac{2}{C_b'} \\ P_3 &= 2(\alpha_b\beta_a^2 + \alpha_a\beta_b^2) + \frac{1}{R_p}\left[(\beta_a^2 + 4\alpha_a\alpha_b)\frac{1}{C_b'} - \frac{1}{L_1 C_m C_m'}\right] \\ P_4 &= \beta_a^2\beta_b^2\left[1 - \tau^2 + \frac{R_2}{R_p}\left(\frac{C_b}{C_b'} - \frac{C_m}{C_m'}\tau^2\right)\right] \end{aligned}\right\} \tag{12.5}$$

The coefficients as given by (12.5) satisfy (12.2) and (12.3) only when the plus sign is used.

The equations are simplified by dividing through by β_a^2 thus

$$\frac{P_3}{\beta_a^2 P_1} = \frac{\dfrac{P_2}{\beta_a^2} + \sqrt{\left(\dfrac{P_2}{\beta_a^2}\right)^2 - \dfrac{4P_4}{\beta_a^4}}}{2} \tag{12.6}$$

$$\frac{\beta^2}{\beta_a^2} = \frac{\dfrac{P^2}{\beta_a^2} + \sqrt{\left(\dfrac{P_2}{\beta_a^2}\right)^2 - \dfrac{4P_4}{\beta_a^4}}}{2} \tag{12.7}$$

which gives the ratio of driven frequency of the crystal to its undriven value. The common variable R_p must satisfy both (12.6) and (12.7). The method of computing the frequency would be to solve for R_p in (12.6) and substitute in (12.7). However, the equations are too complicated a function of R_p for this to be practical. Terry solved them graphically by plotting (12.6) and (12.7) as functions of R_p for assigned values of the circuit, and the intersection of these curves gave the frequency for the different circuit conditions. The results are shown in Fig. 12.6. The G-P curves show the frequency change as a function of plate circuit tuning for the grid to plate connection of the crystal.

12.12 Crystal Between Grid and Cathode

With the crystal connected between the grid and cathode of the tube, the circuit is as shown in Fig. 12.5. The coefficients of equation (12.1) are as follows:

$$\left.\begin{aligned}
P_1 &= \frac{R_1}{L_1} + \frac{R_2}{L_2} + \frac{1}{R_p C_b''} \\
P_2 &= \frac{1}{L_1 C_a} + \frac{R_1 R_2}{L_1 L_2} + \frac{1}{L_2 C_b} + \frac{1}{R_p}\left(\frac{R_1}{L_1} + \frac{R_2}{L_2}\right)\frac{1}{C_b''} \\
P_3 &= \frac{R_2}{L_1 L_2 C_a} + \frac{R_1}{L_1 L_2 C_b} + \frac{1}{R_p}\left(\frac{1}{L_1 C_a C_b''} + \frac{R_1 R_2}{L_1 L_2 C_b''} - \frac{1}{L_1 C_m C_m''}\right) \\
P_4 &= \frac{1}{L_1 L_2 C_a C_b} - \frac{1}{L_1 L_2 C_m^2} + \frac{R_2}{R_p}\left(\frac{1}{L_1 L_2 C_a C_b''} - \frac{1}{L_1 L_2 C_m C_m''}\right)
\end{aligned}\right\} \tag{12.8}$$

With the substitution of uncoupled frequencies, damping factors and coupling coefficient as described in the previous section, they become

$$\left.\begin{aligned}
P_1 &= 2(\alpha_a + \alpha_b) + \frac{1}{R_p C_b''} \\
P_2 &= \beta_a^2 + 4\alpha_a\alpha_b + \beta_b^2 + \frac{1}{R_p}(\alpha_a + \alpha_b)\frac{2}{C_b''} \\
P_3 &= 2(\alpha_b\beta_a^2 + \alpha_a\beta_b^2) + \frac{1}{R_p}\left[(\beta_a^2 + 4\alpha_a\alpha_b)\frac{1}{C_b''} - \frac{1}{L_1 C_m C_m''}\right] \\
P_4 &= \beta_a^2\beta_b^2\left[1 - \tau^2 + \frac{R_2}{R_p}\left(\frac{C_b}{C_b''} - \frac{C_m}{C_m''}\tau^2\right)\right]
\end{aligned}\right\} \tag{12.9}$$

Where

$$\frac{1}{C_m''} = \frac{1}{C_m} + \frac{\mu}{C_d}$$

$$\frac{1}{C_b''} = \frac{1}{C_b} + \frac{\mu}{C_m}$$

$$\frac{1}{C_m} = \frac{C_x}{C_2 C_0}$$

$$C_b = C_2 + \frac{C_0 C_3}{C_0 + C_3}$$

$$C_d = C_0 + \frac{C_2 C_3}{C_2 + C_3}$$

μ = amplification factor of tube

$R_p = \dfrac{\partial e_p}{\partial i_p}$ (e_g constant)

$$\frac{1}{C_x} = \frac{1}{C_0} + \frac{1}{C_2} + \frac{1}{C_3}$$

These equations of conditions for oscillation in this case satisfy (12.4) and (12.5) only when the minus sign is used. That is

$$\frac{P_3}{P_1} = \frac{P_2 - \sqrt{P_2^2 - 4P_4}}{2} \tag{12.10}$$

$$\beta^2 = \frac{P_2 - \sqrt{P_2^2 - 4P_4}}{2} \tag{12.11}$$

Again dividing by β_a^2 to obtain the frequency as a ratio of driven to undriven crystal frequency, we have

$$\frac{P_3}{\beta_a^2 P_1} = \frac{\frac{P_2}{\beta_a^2} - \sqrt{\left(\frac{P_2}{\beta_a^2}\right)^2 - \frac{4P_4}{\beta_a^4}}}{2} \tag{12.12}$$

$$\frac{\beta^2}{\beta_2^2} = \frac{\frac{P_2}{\beta_2^2} - \sqrt{\left(\frac{P_2}{\beta_a^2}\right)^2 - \frac{4P_4}{\beta_a^4}}}{2} \tag{12.13}$$

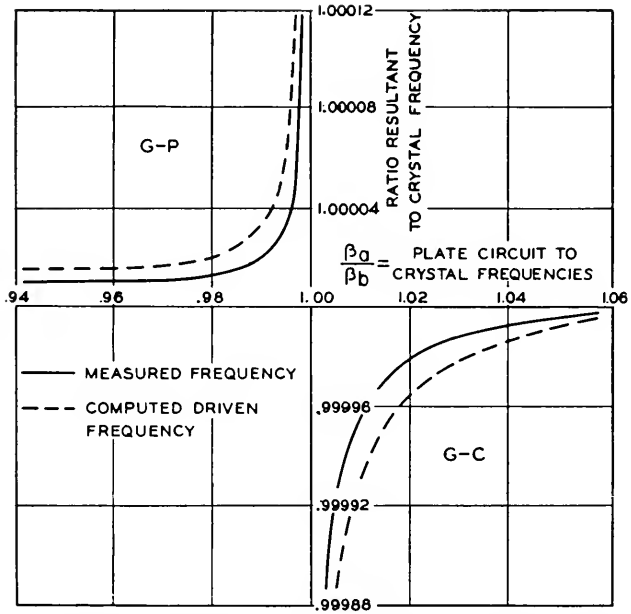


Fig. 12.6—The oscillating frequency as a function of the plate circuit frequency for the crystal connected grid to plate (G-P) and grid to cathode (G-C)

The frequency change as a function of plate circuit tuning was determined graphically in the manner described in section (12.11) and the curves are shown in Fig. 12.6 as the G-C curves.

12.13 Resistance Load Circuit

This is a special case of Plate-Grid connection of the crystal described in section (12.11) in which the plate circuit consists of a capacitance and resistance in parallel. This is a very common Pierce type of oscillator circuit and has the advantage that no tuning adjustment is necessary when using crystals of different frequencies.

Since this circuit is singly periodic, the differential equation for i_1 is of the third order and is derived from (12.1) by setting the plate inductance L_2 of the P coefficients equal to zero. The general equation then becomes

$$\frac{d^3 i_1}{dt^3} + P_1 \frac{d^2 i_1}{dt^2} + P_2 \frac{di_1}{dt} + P_3 i_1 = 0 \tag{12.14}$$

where

$$\left.\begin{aligned} P_1 &= \frac{R_1}{L_1} + \frac{1}{R_2 C_b} + \frac{1}{R_p C_b'} \\ P_2 &= \frac{1}{L_1 C_a} + \frac{R_1}{R_2 L_1 C_b} + \frac{R_1}{R_p L_1 C_b'} \\ P_3 &= \frac{1}{R_2 L_1 C_a C_b} - \frac{1}{R_2 L_1 C_m^2} + \frac{1}{R_p}\left(\frac{1}{L_1 C_a C_b'} - \frac{1}{L_1 C_m C_m'}\right) \end{aligned}\right\} \tag{12.15}$$

With the substitution of the uncoupled damping factors and frequencies, (12.15) becomes

$$\left.\begin{aligned} P_1 &= 2\alpha_a + \frac{1}{R_2 C_b} + \frac{1}{R_p C_b'} \\ P_2 &= \beta_a^2 + \frac{2\alpha_a}{R_2 C_b} + \frac{2\alpha_a}{R_p C_b'} \\ P_3 &= \frac{\beta_a^2}{R_2 C_b} - \frac{1}{R_2 L_1 C_m^2} + \frac{1}{R_p}\left(\frac{\beta_a^2}{C_b'} - \frac{1}{L_1 C_m C_m'}\right) \end{aligned}\right\} \tag{12.16}$$

The frequency as obtained from (12.14) is

$$\beta^2 = P_2 \tag{12.17}$$

with the conditions for oscillation

$$P_2 = \frac{P_3}{P_1} \tag{12.18}$$

obtained by setting the damping factor α equal to zero. The ratio of driven to undriven frequency is obtained by dividing (12.17) and (12.18) by β_a^2. That is

$$\frac{\beta^2}{\beta_a^2} = \frac{P_2}{\beta_a^2} = \frac{P_3}{\beta_a^2 P_1} \tag{12.19}$$

12.14 Interpretation of the Equations

It is learned from this analysis that the frequency of oscillation while governed principally by the frequency of the crystal also depends upon all the constants of the circuit. The effect of the plate circuit impedance is

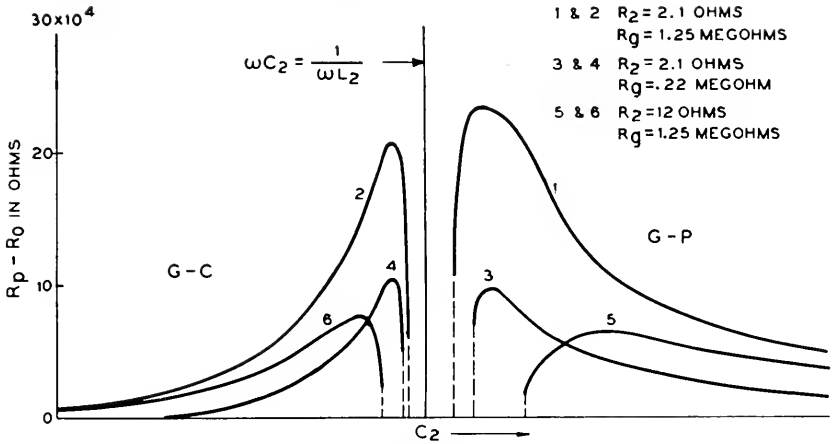


Fig. 12.7—Calculated increase in mean plate resistance against capacitance of the oscillatory circuit

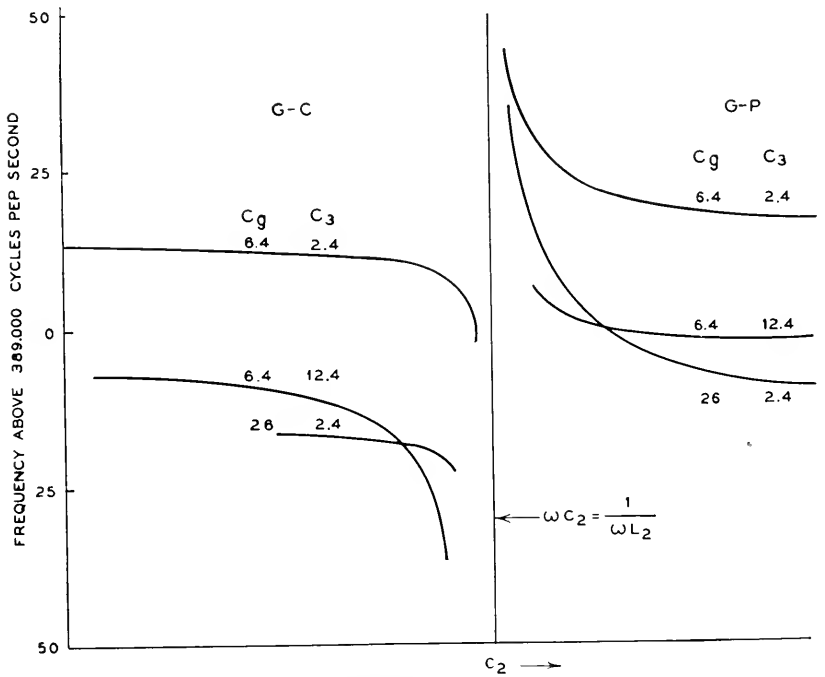


Fig. 12.8—Experimental curves, showing the influence of interelectrode capacitances on the frequency

shown in Fig. 12.6. It is pointed out that the effect of the crystal resistance R_1 is to decrease the frequency for the G-C connection and increase the frequency for the G-P connection. The discrepancy between the measured

and experimental values shown on the curves is attributed to the difference between chosen and actual value of R_1. The effect of the input loss of the tube is not shown because the grid current was disregarded; however, this loss may be reduced to an equivalent R_1. The resistance of the plate circuit R_2 affects the frequency in a similar manner. The effects of these resistances on frequency are less for low values of plate circuit impedances.

The required value of R_p gives a measure of amplitude of oscillation because it is necessary for oscillations to build up until the internal plate resistance is equal to the calculated value. It is found that R_p increases

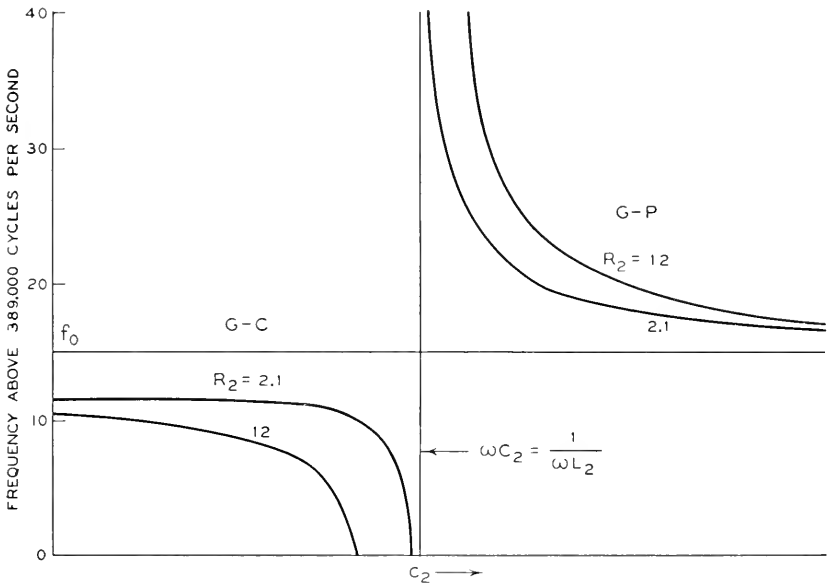


Fig. 12.9—Experimental curves, showing the relation between the frequency and the resistance of the oscillatory circuit

gradually to a maximum as the common frequency for the two types of circuits is approached then abruptly drops.

Vigoureux[8] analyzes the crystal oscillator in a manner similar to Terry and correlates his interpretations of the equations with considerable experimental data, some of which are shown in Figs. 12.7, 12.8, 12.9 and 12.10. He points out that there is an optimum value of grid capacitance with the crystal connected between grid and plate and a certain amount of grid-plate capacitance is required when the crystal is connected between grid and cathode.

Wheeler[10] does not assume a linear static tube characteristic but represents it by a three-term nonlinear expression. The results are more complex and it is necessary in the end to disregard certain resistance terms.

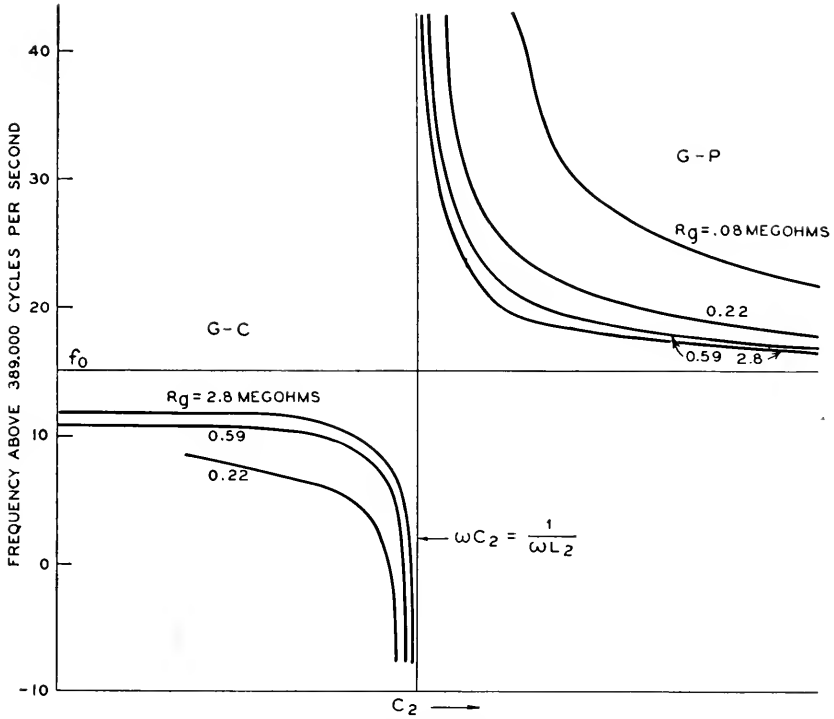


Fig. 12.10—Experimental curves, showing the relation between the frequency of a quartz oscillator and the capacitance of the oscillatory circuit for various values of the grid leak

12.20 Solution by Complex Functions

The analysis of oscillator circuits may be simplified when only steady state conditions are of interest, all circuit elements are considered linear, and certain requirements which define the conditions necessary for oscillations are known. Under these conditions the common circuit equations of complex numbers give the information desired. In this method the voltage induced in the plate circuit is considered the driving voltage which produces a current in the grid circuit (see Fig. 12.11). The network be-

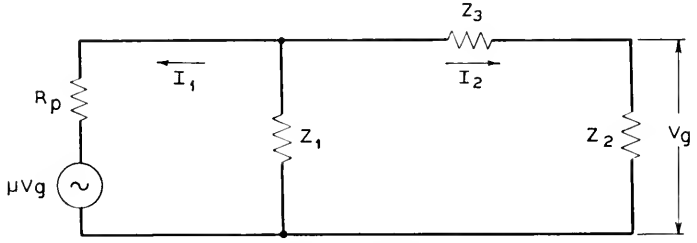


Fig. 12.11—Equivalent circuit of Pierce and Miller types of oscillators shown in Fig. 12.12

tween plate and grid may be of any type and oscillations are maintained when the total gain through the circuit is unity (gain of tubes = attenuation through circuit) and the phase relation between the induced plate voltage (μV_g) and the grid voltage (V_g) is 180° (the phase shift is zero when μ is considered negative). The expression $\mu\beta = 1$ defines these requirements. Llewellyn[11] applies this method to oscillator circuits in general and Koga[12] uses it to study the crystal oscillator in particular.

The equations are developed on the assumption that the grid-voltage vs. plate-current characteristic of the tube is linear. The fundamental equation of $\mu\beta$ is given by the ratio of the voltage developed across the grid

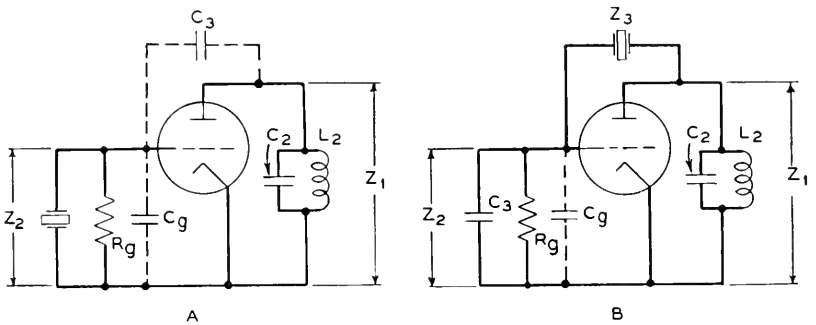


Fig. 12.12—Circuit diagrams of crystal oscillators with crystal connected from grid to cathode (A) and grid to plate (B)

circuit by the fictitious driving voltage μV_g to the voltage V_g. For the general circuit, Fig. 12.11, it is

$$\mu\beta = \frac{I_2 Z_2}{V_g} = \frac{-\mu Z_1 Z_2}{R_p Z_s + Z_1(Z_2 + Z_3)} \tag{12.20}$$

where

$$Z_s = Z_1 + Z_2 + Z_3$$

It is more convenient to write this in the reciprocal form

$$\frac{1}{\mu\beta} = \frac{R_p Z_s + Z_1(Z_2 + Z_3)}{-\mu Z_1 Z_2} = 1 \tag{12.21}$$

In applying this to the crystal oscillator, the additional assumptions made are that the grid current is negligible and the resistance in the plate impedance Z_1 is zero.

12.21 Crystal Grid to Cathode

With the assumptions made above and the crystal connected from grid to cathode of the tube according to Fig. 12.12A, the impedances are

$$Z_1 = jX_1 \qquad Z_2 = R_{cg} + jX_{cg} \qquad Z_3 = jX_3$$

where R_{cg} is the effective resistance and X_{cg} the effective reactance of the crystal, the grid resistance R_g and the circuit capacitance C_g in parallel at the oscillating frequency. Upon substitution of these in (12.21)

$$\frac{1}{\mu\beta} = \frac{[R_{cg}R_p - X_1(X_{cg} + X_3)] + j(X_1R_{cg} + R_pX_s)}{\mu X_1 X_{cg} - j\mu X_1 R_{cg}} = 1 \tag{12.22}$$

where

$$X_s = X_1 + X_{cg} + X_3$$

Thus $\frac{1}{\mu\beta}$ is of the form

$$\frac{1}{\mu\beta} = P + jQ$$

which means that $P = 1$ and $Q = 0$.
This results in the following two equations obtained from the real and imaginary parts of (12.22) both of which must be satisfied for oscillations to be maintained.

The real part of (12.22) gives

$$-R_p = \frac{X_1(\mu + 1)(R_{cg}^2 + X_{cg}^2) + X_{cg}X_3}{R_{cg}(X_1 + X_3)} \tag{12.23}$$

and from the imaginary part is obtained

$$X_s = \frac{X_1X_3 - R_{cg}R_p}{R_p\phi_{cg}} \tag{12.24}$$

where $\phi_{cg} = \frac{X_{cg}}{R_{cg}}$ (This ratio of reactance to resistance of the crystal circuit will appear in various equations later.)

Equation (12.24) may be said to define the oscillating frequency and is in a convenient form to examine the effect of the various circuit variables upon the frequency. The impedances X_1, R_{cg}, X_{cg} and X_3 may be thought of as forming an oscillating loop (See Fig. 12.11). For oscillations to be maintained in such a loop the sum of the reactances must equal zero and the sum of the resistances must equal zero. But the sum of the resistances cannot equal zero since R_{cg} is the only resistance in the loop and it is positive. It is therefore necessary for the driving voltage μV_g to act upon the circuit and supply the energy dissipated by the resistance R_{cg} (and also R_p through which the energy is supplied). This alters the frequency somewhat and it is no longer determined by setting the three reactances equal to zero as may be seen by equation (12.24). Nevertheless, the right side of this equation is small and approaches zero when R_{cg} approaches zero. It also becomes very small when the reactance X_1 becomes small and R_{cg} is not too great. This is the same condition as found by the differential

equation method and illustrated in Fig. 12.6 by the G-C curves. As the plate reactance X_1 is made small the frequency increases and approaches a limiting value but does not quite reach it. This limiting value is the frequency at which $X_s = 0$. The dotted G-C curve shows that R_{cg} tends to lower the frequency and determines how close the limiting frequency is approached. The plate circuit resistance R_2 (component of Z_1), if considered, would have a similar effect as shown by the experimental curves 12.9. The grid resistance R_g (component of Z_2) has an opposite effect as shown in Figure 12.10 because increasing R_g is equivalent to decreasing the effective resistance R_{cg}.

The effect of the various constants of the crystal and circuit upon the oscillating frequency may be obtained from (12.24) upon substitution of these constants for the reactances and resistance R_{cg}. The equation is put in a more convenient form for this purpose by Koga.[12] Equation (12.21) is written,

$$\frac{1}{Z_2} + \frac{1}{Z_3} + \frac{\mu}{Z_3\left(1 + R_p/Z_1 + \dfrac{R_p}{Z_2 + Z_3}\right)} = 0 \tag{12.25}$$

It is assumed that the current in the grid branch is small compared to the plate current. This reduces the equation to

$$\frac{1}{Z_2} + \frac{1}{Z_3} + \frac{\mu}{Z_3(1 + R_p/Z_1)} = 0 \tag{12.26}$$

The admittance expression for the crystal is

$$\frac{1}{Z_c} = \frac{R_1 - j\left[\omega L_1 - \dfrac{1}{\omega C_1} - \dfrac{1}{\omega(C_0 + C_4)}\right]}{R_1^2 + \left[\omega L_1 - \dfrac{1}{\omega C_1} - \dfrac{1}{\omega(C_0 + C_4)}\right]^2}\left(\frac{C_4}{C_0 + C_4}\right)^2 + j\omega\frac{C_0 C_4}{C_0 + C_4} \tag{12.27}$$

Note that Koga considers the air gap capacitance C_4 as a separate factor but it may be included in the other constants of the crystal in which case the equivalent circuit is as shown in Fig. 12.3. With the crystal connected between grid and cathode the various circuit admittances are:

$$\frac{1}{Z_1} = \frac{1}{j\omega L_2} + j\omega C_2$$

$$\frac{1}{Z_2} = \frac{1}{Z_c} + \frac{1}{R_g} + j\omega C_g$$

$$\frac{1}{Z_3} = j\omega C_3$$

After substitution of these values of the admittances in (12.26) and setting the real and imaginary parts equal to zero, the following two equations are obtained:

$$\frac{R_1}{R_1^2 + \left[\omega L_1 - \frac{1}{\omega C_1} - \frac{1}{\omega(C_0 + C_4)}\right]^2}\left(\frac{C_4}{C_0 + C_4}\right)^2 + \frac{1}{R_g} - \mu\omega C_3 \frac{R_p\left(\frac{1}{\omega L_2} - \omega C_2\right)}{1 + R_p^2\left(\frac{1}{\omega L_2} - \omega C_2\right)^2} = 0 \tag{12.28}$$

and

$$\frac{\omega L_1 - \frac{1}{\omega C_1} - \frac{1}{\omega(C_0 + C_4)}}{R_1^2 + \left[\omega L_1 - \frac{1}{\omega C_1} - \frac{1}{\omega(C_0 + C_4)}\right]^2} = \omega\left(\frac{C_0 + C_4}{C_4}\right)^2 \left[C_g + \frac{C_0 C_4}{C_0 + C_4} + C_3 + \frac{\mu C_3}{1 + R_p^2\left(\frac{1}{\omega L_2} - \omega C_2\right)^2}\right] \tag{12.29}$$

Equation (12.28) gives the conditions necessary for oscillations and (12.29) gives the oscillating frequency as explained below:

12.22 Frequency of Oscillations for G-C Connection of Crystal

Equation (12.29) for frequency is simplified by the fact that over the narrow frequency range considered, the reactances of L_2 and C_2 do not change appreciably. Also at the oscillating frequency,

$$R_1^2 \ll \left[\omega L_1 - \frac{1}{\omega C_1} - \frac{1}{\omega(C_0 + C_4)}\right]^2$$

With these approximations (12.29) may be written

$$\omega^2 = \frac{1}{L_1 C_1} + \frac{1}{L_1 C_0}\left[1 - \frac{1}{\frac{C_0 + C_t}{C_t} + \frac{C_0}{C_4}}\right] \tag{12.30}$$

where

$$C_t = C_g + C_3 + \frac{\mu C_3}{1 + R_p^2\left(\frac{1}{\omega_0 L_2} - \omega_0 C_2\right)^2}$$

and ω_0 is a constant approximating the oscillating frequency.

Since the frequency is a function of the internal plate resistance of the tube (R_p) and this is in turn a function of the other circuit variables, the frequency equation (12.30) is not sufficient to calculate the frequency. However, qualitative effects of the various circuit components upon frequency are obtained by assuming R_p an independent variable. It is readily seen that an increase in R_p increases the frequency. The effect of the air gap between crystal and electrodes, which is represented by the capacitance C_4, and the effect of the capacitance across the crystal C_g are illustrated in Fig. (12.13).* To determine the frequency change caused by tuning of

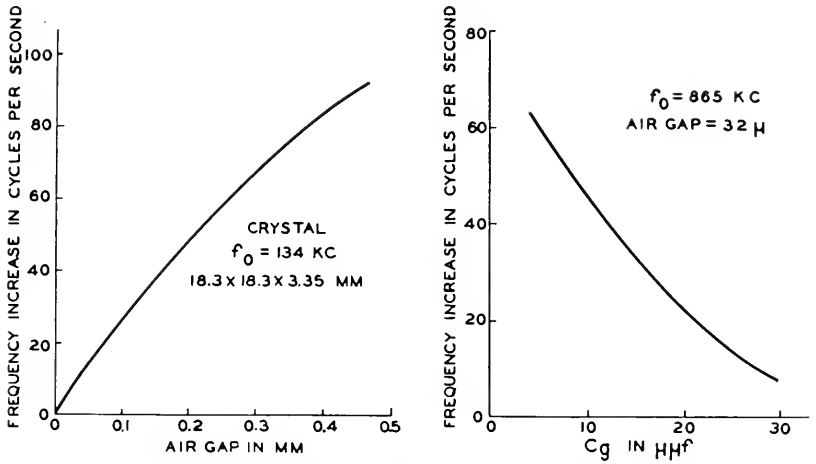


Fig. 12.13—Experimental curves, showing the effect of crystal air gap and grid capacitance on the frequency of oscillations

the plate circuit (variations of C_2) requires the calculation of the change of the variable part of C_t. This quantity is

$$C_v = \frac{\mu C_3}{1 + R_p^2\left(\frac{1}{\omega_0 L_2} - \omega_0 C_2\right)^2} \tag{12.31}$$

The plot of C_v is shown in Fig. (12.14A). The frequency decrease is proportional to the increase in C_v. This is indicated in Fig. (12.14B). Oscillations stop before the point $\omega_0 C_2 = \frac{1}{\omega_0 L_2}$ is reached. The frequency thus varies in the same manner as shown in Fig. (12.6) but the curve is reversed because of the fact that the independent variable is taken as C_2 instead of the frequency function of C_2.

The frequency change resulting from variations in the grid-plate capacitance C_3 depends also upon the value of C_v as seen from (12.31). It is also

* See also: "The Piezoelectric Resonator and the Effect of Electrode Spacing upon Frequency," Walter G. Cady, *Physics*, Vol. 7, July 1936.

seen that the smaller the value of C_2 (lower the plate reactance) the less effect will the tube constants μ, R_p and C_3 have upon the frequency. The circuit is therefore more stable. For this reason it has become customary to measure the frequency of crystals with the capacitance C_2 reduced to a value below that which gives maximum amplitude of oscillations.

12.23 Amplitude of Oscillations for G-C Connection of Crystal

A measure of the amplitude of oscillations is obtained from (12.28) which expresses the necessary conditions for oscillations to be maintained. In order for oscillations to start the expression must be negative, and, as the amplitude builds up, R_p increases which reduces the negative terms

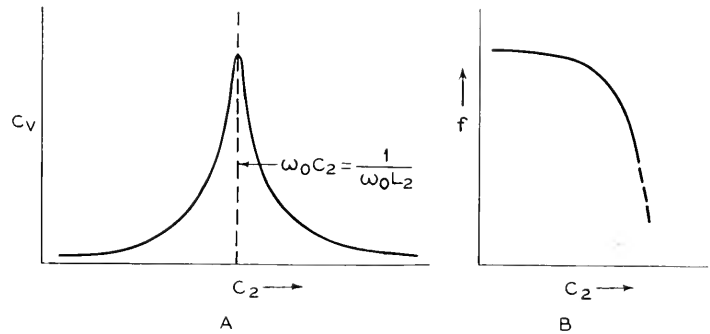


Fig. 12.14—The variation of grid to cathode capacitance (A) and oscillator frequency (B) with change in plate circuit capacitance. Crystal connected grid to cathode

until the equality is satisfied. The difference between the positive and negative terms is therefore a measure of the amplitude of oscillations.

Equation (12.28) may be written

$$\psi - \left[\Phi_0 + \frac{1}{R_g}\right] = A \tag{12.32}$$

where A is a measure of the amplitude,

$$\psi = \mu C_3 \omega_0 \frac{R_p\left(\frac{1}{\omega_0 L_2} - \omega_0 C_2\right)}{1 + R_p^2\left(\frac{1}{\omega_0 L_2} - \omega_0 C_2\right)^2} \tag{12.33}$$

and

$$\Phi_0 = R_1 \omega_0^2 \left(\frac{C_0 + C_4}{C_4}\right)^2 \left[\frac{C_0 C_4}{C_0 + C_4} + C_g + C_3 + \frac{\mu C_3}{1 + R_p^2\left(\frac{1}{\omega_0 L_2} - \omega_0 C_2\right)^2}\right]^2 \tag{12.34}$$

where again R_1^2 is assumed small compared to

$$\left[\omega_0 L_1 - \frac{1}{\omega_0 C_1} - \frac{1}{\omega_0(C_0 + C_4)}\right]^2$$

and ω_0 is considered a constant.

Equation (12.32) shows that in order to obtain a large amplitude ψ should be large and Φ_0 should be small. With this in mind equations (12.33)

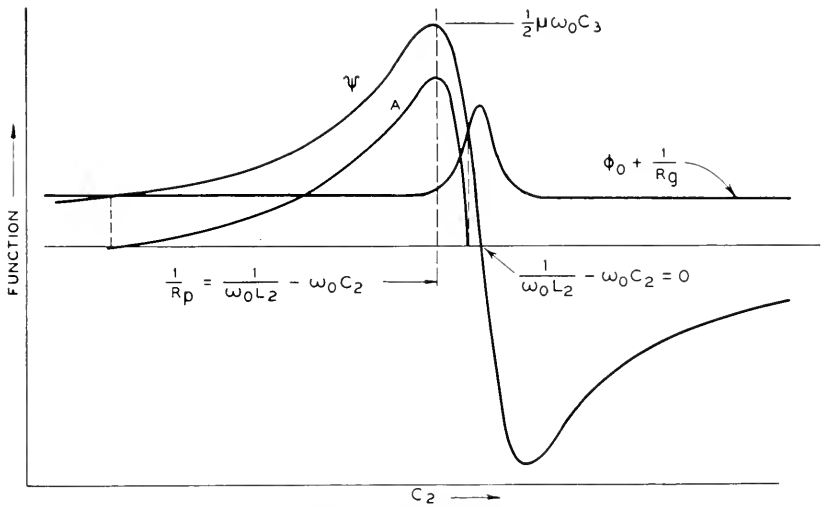


Fig. 12.15—Functions from which the activity variations (A) are determined as the plate circuit capacitance is varied. Crystal connected grid to cathode

and (12.34) may be analyzed to determine the relation between the circuit components and amplitude. It is found that for maximum amplitude

C_g and R_1 should be small,
C_4 should be large,
C_3 has an optimum value, and
R_g should be large.

As to the plate circuit, the amplitude is maximum when $\frac{1}{R_p} = \frac{1}{\omega_0 L_2} - \omega_0 C_2$. A plot of ψ and $\Phi_0 + \frac{1}{R_g}$ is shown in Fig. 12.15. The difference between these two curves is a measure of the amplitude and is shown by curve A. Oscillations can exist only where ψ lies over $\Phi_0 + \frac{1}{R_g}$. The sharpness of ψ varies considerably with the value of R_p and the resistance of the $L_2 - C_2$ circuit. The latter is disregarded for simplicity. Here again the

results can only be considered a first approximation, but agree with actual conditions sufficiently to be of considerable interest.

12.24 Crystal Grid to Plate

The equation (12.20) is general and for the condition of crystal connected between grid and plate of the tube (See Figure 12.12B) Z_3 represents the crystal impedance which will be called $Z_c = R_c + jX_c$, also: $Z_1 = jX_1$, $Z_2 = jX_2$ and $X_s = X_1 + X_2 + X_c$.

Note that R_g and C_3 are disregarded in this case because their effects are similar to those determined for the foregoing case of crystal connected grid to cathode.

After substitution of these values in (12.20) the real part is found to be

$$R_p = \frac{(\mu + 1)X_1X_2 + X_1X_c}{R_c} \tag{12.35}$$

and the imaginary part is

$$X_s = -\frac{R_cX_1}{R_p} \tag{12.36}$$

which shows the effect of the various variables on the frequency. The right hand side of equation (12.36) is comparatively small and the frequency is therefore close to a value f_0 which makes $X_s = 0$. In this case the frequency is above the limiting frequency f_0 because the right hand side is positive since X_1 is negative, whereas it was found that the frequency was below f_0 for the crystal connected between grid and cathode. As R_c and X_1 are increased the frequency will increase and as R_p increases the frequency decreases. These interpretations are verified by the *G-P* curves of Figures 12.6, 12.9 and 12.10.

The effects of the various circuit and crystal constants are determined by Koga[12] by writing the general $\mu\beta$ equation as

$$Z_3 + Z_2 + \frac{\mu Z_2}{1 + R_p/Z_1} = 0 \tag{12.37}$$

After substitution for the Z's, the real and imaginary parts are respectively,

$$\left(\frac{1}{\omega C_0}\right)^2 \frac{R_1}{R_1^2 + \left[\omega L_1 - \frac{1}{\omega C_1} - \frac{1}{\omega C_0}\right]^2} + \frac{\mu}{\omega C_g} \cdot \frac{R_p\left(\frac{1}{\omega L_2} - \omega C_2\right)}{1 + R_p^2\left(\frac{1}{\omega L_2} - \omega C_2\right)^2} = 0 \tag{12.38}$$

and

$$\frac{\frac{1}{\omega C_1}+\frac{1}{\omega C_0}-\omega L_1}{R_1^2+\left[\omega L_1-\frac{1}{\omega C_1}-\frac{1}{\omega C_0}\right]^2} = \omega C_0^2\left[\frac{1}{C_0}+\frac{1}{C_4}+\frac{1}{C_g}+\frac{\mu}{C_g}\cdot\frac{1}{1+R_p^2\left(\frac{1}{\omega L_2}-\omega C_2\right)^2}\right] \tag{12.39}$$

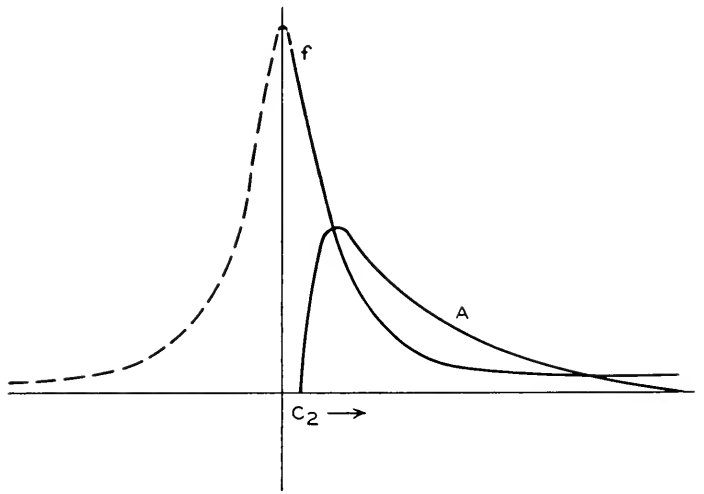


Fig. 12.16—Frequency and activity change for variations in the plate circuit capacitance. Crystal connected grid to plate

There are two values of ω which satisfy (12.39) but only one of these ω_n will satisfy (12.38). At this value of ω_n

$$R_1^2 \ll \left[\omega L_1-\frac{1}{\omega C_1}-\frac{1}{\omega C_0}\right]^2 \tag{12.40}$$

By introduction of this and the assumption that ω_0 is essentially constant, (12.38) may be written

$$R_1 C_0^2\left[\frac{1}{C_0}+\frac{1}{C_4}+\frac{1}{C_g}+\frac{\mu}{C_g}\cdot\frac{1}{1+R_p^2\left(\frac{1}{\omega_0 L_2}-\omega_0 C_2\right)^2}\right]^2 + \frac{\mu}{\omega_0 C_g}\cdot\frac{R_p\left(\frac{1}{\omega_0 L_2}-\omega_0 C_2\right)}{1+R_p^2\left(\frac{1}{\omega_0 L_2}-\omega_0 C_2\right)^2} = 0 \tag{12.41}$$

This is an approximation for the conditions for oscillation and relative amplitude.

The frequency equation (12.39) becomes

$$\omega_n^2 = \frac{1}{L_1}\left[\frac{1}{C_1} + \frac{1}{C_0} - \frac{1}{G}\right] \tag{12.42}$$

where

$$G = C_0^2\left[\frac{1}{C_0} + \frac{1}{C_4} + \frac{1}{C_g} + \frac{\mu}{C_g}\cdot\frac{1}{1 + R_p^2\left(\frac{1}{\omega_0 L_2} - \omega_0 C_2\right)^2}\right]$$

and ω_0 is a fixed value written in place of ω_n. Figure 12.16 shows the frequency and amplitude changes as a function of C_2 for the crystal connected between grid and plate.

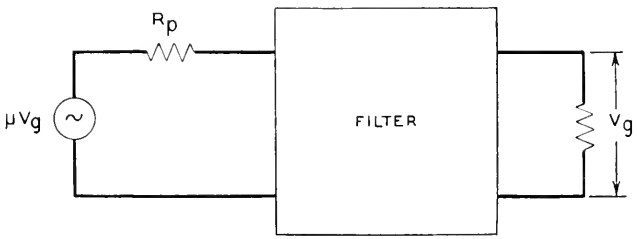


Fig. 12.17—Generalized oscillator circuit in the form of a filter network

12.25 Condition $\mu\beta = 1$ for Circuits in General

It is convenient to apply the rule $\mu\beta = 1$ as the condition for sustained oscillations to more complex oscillator circuits. The circuits may be drawn as shown in Figure 12.17 and the characteristics of the filter network between transmitting and receiving end may be analyzed by conventional filter theory to determine the conditions which fulfill the oscillation requirements. An example of this is the oscillator shown in Figure 12.18A. The equivalent configuration, Figure 12.18B, indicates that the crystal is part of a low pass filter and the frequency of operation is that at which the total phase shift is 180°.

Oscillators involving more than one tube may also be inspected in this manner. Figure 12.19 is a two tube oscillator designed to operate at a frequency close to the resonant frequency of the crystal. The proper phase shift is obtained by a two-stage amplifier and, therefore, no phase shift is required through the crystal network. The crystal thus must operate as a resistance which it can only do at its resonant or antiresonant frequency. Since the transmission through the crystal branch is very low at the antiresonant frequency of the crystal, it will oscillate only at the resonant

frequency. Heegner[13] explains a number of crystal oscillator circuits by the method briefly outlined above.

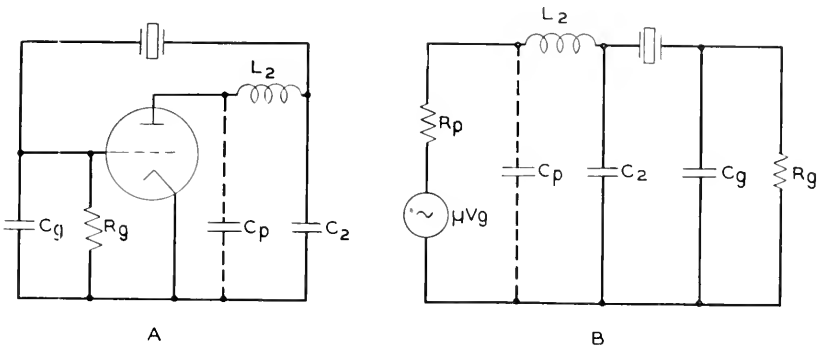


Fig. 12.18—The oscillator circuit (A) is equivalent to the filter circuit (B)

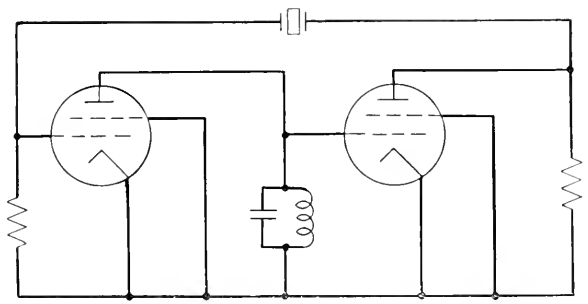


Fig. 12.19—Oscillator circuit in which the crystal operates at its series resonant frequency

12.30 Vector Method of Oscillator Analysis

A convenient method of examining the effect of certain circuit variables on frequency and the necessary conditions for oscillation is by the vector representation of the voltages and currents in the circuit. Much of Heising's[14] early work on the analysis of electric oscillators by vector methods is directly applicable to crystal oscillators. Boella[15] analyzed the crystal oscillator circuit by this method and treated in detail the effect of the decrement of the crystal on the oscillating frequency. Since some engineers prefer this method of qualitative analysis to approximate equations it will be briefly explained.

The vector diagrams for the two conditions, crystal between grid and plate and between grid and cathode, are shown in Figure 12.20A and B as applied to the circuit diagrams, Figure 12.12A and B, respectively when in the simplified form of Figure 12.11. The necessary conditions for oscilla-

tions are that V_g is in phase with and equal to μV_g (note that μ is considered negative). Like Koga, Boella assumes the current I_2 small compared to I_1, hence the voltage drop across Z_1 is approximately $Z_1 I_1$. The angle this makes with V_g is determined by the value of Z_1 and the internal plate impedance R_p. Any change in either of these requires a change in the angles ψ and ψ' in order that V_g shall be in phase with μV_g. This means that the frequency must vary to produce this change in ψ and ψ'. Because of the rapid change in the reactance and resistance of the crystal with frequency, these requirements are met with very little change in frequency, which accounts for the high degree of frequency stability obtained with crystals. This is described more in detail in a later section.

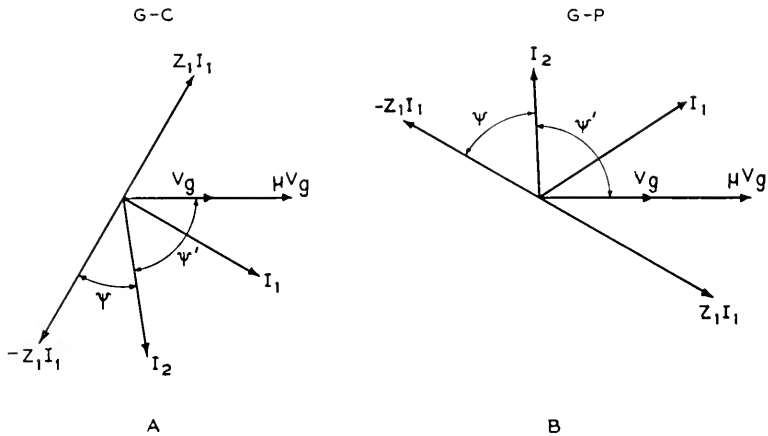


Fig. 12.20—Vector diagrams of currents and voltages in the oscillator circuit Figure 12.11 with crystal connected grid to cathode (A) and grid to plate (B)

12.31 Change in Frequency with Decrement of Crystal

It has been found that for the crystal connected from grid to cathode there is a maximum theoretical frequency at which the circuit can be made to oscillate by reducing the plate circuit impedance. This also corresponds to the minimum frequency which can be obtained with the crystal connected between grid and plate. This was called the limiting frequency f_0. It is interesting to note that f_0 is determined by the intersection of the reactance curve of the crystal plotted as a function of frequency and the reactance curve of the capacitance in series with the crystal. This series capacitance is the grid-plate capacitance for one case and the grid-cathode capacitance for the other. As illustrated in the curves Figure 12.21, the limiting frequency f_0 increases as the decrement of the crystal increases.

The difference between the true frequency of oscillations and f_0 increases

as the plate impedance is increased and as the losses in any of the circuit elements increase. This is necessary for the proper angle of $\psi + \psi'$ in the vector diagram. With the *G-P* connections, the departure from f_0 and change in f_0 as the decrement of the quartz varies are in the same direction, while for the grid-cathode connection they vary in opposite directions, and the net result will depend upon the value of the internal plate resistance and plate circuit impedance. The curves of Figure 12.21 show that the

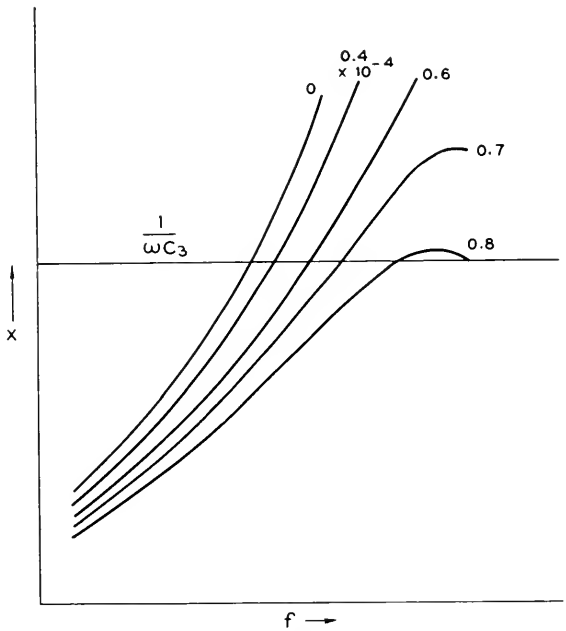


Fig. 12.21—The change in reactance characteristic of a crystal resulting from a change in decrement

change in f_0 for a given change in decrement is less for smaller values of $\frac{1}{\omega C_3}$ (larger values of series capacitance C_3). That is, the effect of the decrement of the crystal upon the oscillating frequency is small when the crystal is operated near its frequency of resonance.

12.40 Negative Resistance Method of Analysis

The methods of analyzing oscillator circuits described in the previous sections define the operation in terms of the individual circuit elements and the crystal is treated as one of the circuit elements. Certain advantages result, however, by grouping all the circuit elements, except the crystal,

into a single impedance as shown in Fig. 12.22A. Here Z_t represents the impedance looking into the oscillator from the crystal terminals.

The requirements for sustained oscillations are that the sum of the reactances around the loop equal zero and the sum of the resistances equal zero as previously stated in section 12.21. These conditions are obtained when Z_t is a negative resistance ρ in parallel with (or in series with) a capacitance C_t as shown in Fig. 12.22C. The crystal is considered to be operating as an inductance L_c and resistance R_c as determined in the pre-

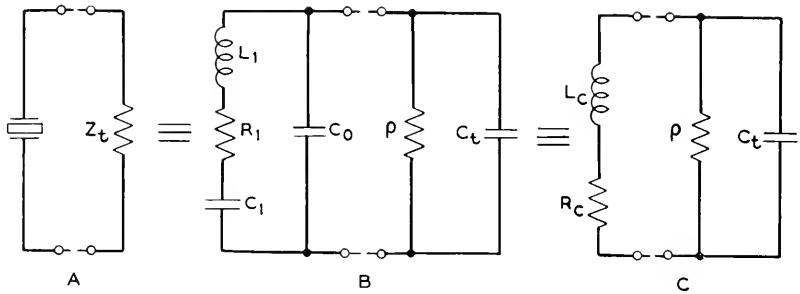


Fig. 12.22—Equivalent representations of crystal and oscillator circuit

vious sections. The frequency equation has been derived by Reich[16] from the differential equation for the current in the loop. It is

$$\omega = \sqrt{\frac{\rho + R_c}{\rho} \cdot \frac{1}{L_c C_t}} \tag{12.43}$$

and the condition for oscillation is shown to be

$$\left| C_t \rho \right| \leqq \left| \frac{L_c}{R_c} \right| \tag{12.44}$$

We shall consider the crystal connected between the grid and cathode of the tube, in which case Z_t is the input impedance of the vacuum tube. The expression for $\frac{1}{Z_t}$ was developed by Chaffee[17] from which it is possible to determine the circuit conditions necessary for the input resistance and reactance to be negative. The effect of the circuit variables upon the absolute values of ρ and C_t determines their effect upon the frequency and activity according to equations (12.43) and (12.44).

12.41 Input Admittance of the Vacuum Tube

With the assumption that the grid current is negligible and the static tube capacitances C_p and C_g are part of the external circuit, Chaffee's equation for input conductance becomes

$$g = \frac{C_3^2 \omega (K + G_1) + C_3 \omega \mu K (C_3 \omega - B_1)}{(K + G_1)^2 + (C_3 \omega - B_1)^2} \tag{12.45}$$

and for the input susceptance

$$b = -C_3\omega - \frac{C_3\omega\mu K(K + G_1) - C_3^2\omega^2(C_3\omega - B_1)}{(K + G_1)^2 + (C_3\omega - B_1)^2} \tag{12.46}$$

where K and μ are defined as follows:

$$K = \left(\frac{\partial i_p}{\partial e_p}\right) \quad (e_g \text{ constant})$$

$$\mu = -\left(\frac{\partial e_p}{\partial e_g}\right) \quad (i_p \text{ constant})$$

and G_1 and B_1 are the conductance and susceptance of the plate circuit. If we let

$$h = \frac{G_1}{|B_1|}$$

$$A = \frac{\omega C_3}{K} \quad \text{and} \quad B = \frac{B_1}{K}$$

(12.45) becomes

$$g = C_3\omega A \frac{(1 + hB) + \mu\left(1 + \frac{B}{A}\right)}{(1 + hB)^2 + A^2\left(1 - \frac{B}{A}\right)^2} \tag{12.47}$$

and (12.46) becomes

$$b = -C_3\omega\left[1 + \frac{\mu(1 + hB) - A^2\left(1 - \frac{B}{A}\right)}{(1 + hB)^2 + A^2\left(1 - \frac{B}{A}\right)^2}\right] \tag{12.48}$$

When the resistance of the plate circuit is neglected (i.e. $h = 0$), and $\mu \gg 1$ we may write

$$\frac{g}{K} = A\frac{\mu(A - B)}{1 + (A - B)^2} \tag{12.49}$$

and

$$\frac{b}{K} = -A\left[\frac{\mu - B(A - B)}{1 + (A - B)^2}\right] \tag{12.50}$$

These equations are in a convenient form to determine the effect of the plate tuning $f(B)$ and grid-plate capacitance $f(A)$ upon the resistance ρ

and capacitance C_t with the assumptions of no grid current, low plate circuit resistance, and $\mu \gg 1$.

From (12.49) it is seen that in order for g to be negative, B must be positive and greater than A, since A is normally positive. That is, the plate circuit reactance must be positive and less than the grid-plate reactance when the latter is a capacitance. Under these conditions b/K and hence the input reactance will be negative according to (12.50).

Curves of b/K are shown in Fig. 12.23 with B as independent variable and A as parameter. These curves indicate frequency change. On the

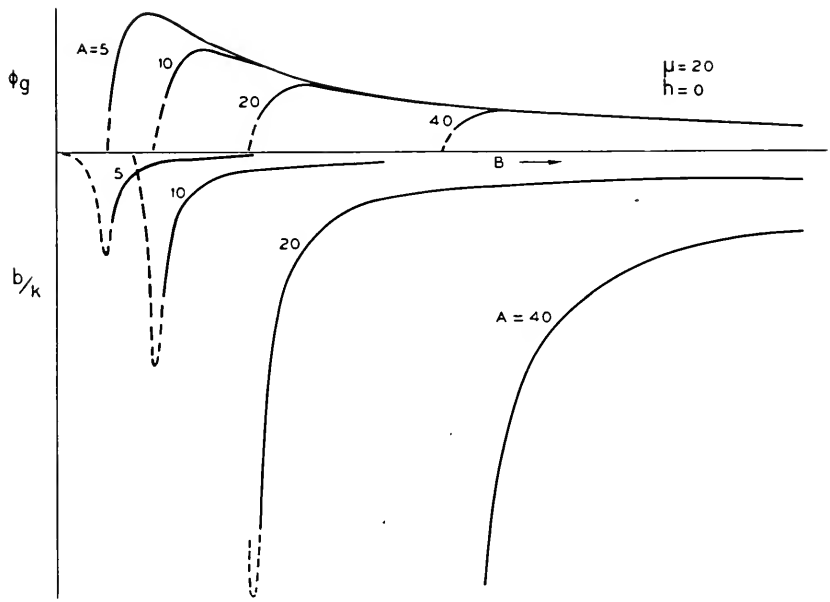


Fig. 12.23—Variations in the input impedance functions of an oscillator circuit for changes in plate circuit tuning

same figure is plotted b/g called ϕ_g. This may be considered the sensitivity of the oscillator or, for a given value of $\omega L_c/R_c$ of the crystal, it represents the activity. The similarity between these curves and the actual change in frequency and activity normally experienced is apparent.

It should be pointed out here that the presence of harmonics is effective in changing the input impedance of the vacuum tube and hence the frequency and activity of the oscillator. The presence of harmonics results from the non-linear characteristics of the vacuum tube. Llewellyn[11] explains that a non-linear resistance may be represented by a linear resistance plus a linear reactance. From what has been said concerning the

frequency of the oscillating loop, it is apparent that this effective reactance will alter the frequency. However, this reactance is small when the impedance of the circuit is low at the harmonic frequencies and is zero when the external circuit is a pure resistance.

12.50 Efficiency and Power Output of Oscillators

In many applications of crystal oscillators the efficiency and power output are important factors. These are not treated here but reference is made to the work of Heising[14] which covers this aspect for various electric oscillator circuits. Much of the analysis is directly applicable to crystal oscillators.

12.60 Frequency Stability of Crystal Oscillators

The equations for frequency show that the frequency is governed somewhat by the amplification factor, the grid resistance and internal plate resistance of the vacuum tube. Since these factors are functions of voltages applied to the tube and amplitude of oscillation, they cannot be considered fixed. If the frequency change resulting from these variables is great, the frequency stability is said to be low, and if very little frequency change takes place the frequency is determined principally by the circuit constants and the frequency stability is said to be high.

Llewellyn[11] shows how it is possible to compensate for the change in plate resistance by the proper value of circuit elements. This was done by determining the relations necessary for R_p to be eliminated from the frequency equation. It is sometimes helpful in designing very stable oscillators for frequency standards to select circuit elements which will reduce the effect of plate voltage changes on the frequency. It is more the purpose of this section, however, to show Llewellyn's derivation of the equations for frequency stability which have not heretofore been published and from them point out the characteristic of crystals which enable them to stabilize oscillators.

12.61 The Frequency Stability Equation

The steady state oscillating condition is

$$\mu\beta = 1 \tag{12.51}$$

In general β is a function of the frequency, the amplitude of oscillations, and of some independent variable V. This independent variable is the one for which it is desired to stabilize the frequency. It may be the potential applied to the tube, or it may be a capacitance located somewhere in the circuit. β depends upon these three variables thus:

$$\mu\beta = f(p\ ,\ a\ ,\ V) \tag{12.52}$$

Instead of the frequency, a more general symbol p is used and may be thought of as the differential operator d/dt which occurs in the fundamental linear differential equations taken as describing the oscillatory system. That is

$$p = \frac{d}{dt} = \alpha + i\omega \tag{12.53}$$

The function $\mu\beta$ may have the form

$$\mu\beta = Ae^{i\theta} \tag{12.54}$$

The result of taking a general variation δ of (12.54) is then

$$\frac{\delta A}{A} + i\delta\theta = 0 \tag{12.55}$$

Since (12.54) is a function of the three variables p, a, and V the variational equation (12.55) may be expressed in terms of partial derivatives with respect to these three variables. That is

$$\frac{1}{A}\left[\frac{\partial A}{\partial p}\delta p + \frac{\partial A}{\partial a}\delta a + \frac{\partial A}{\partial V}\delta V\right] + i\left[\frac{\partial \theta}{\partial p}\delta p + \frac{\partial \theta}{\partial a}\delta a + \frac{\partial \theta}{\partial V}\delta V\right] = 0 \tag{12.56}$$

The solution of (12.56) for the variation in p is

$$\delta p = -\frac{\frac{1}{A}\left(\frac{\partial A}{\partial V}\delta V + \frac{\partial A}{\partial a}\delta a\right) + i\left(\frac{\partial \theta}{\partial V}\delta V + \frac{\partial \theta}{\partial a}\delta a\right)}{\frac{1}{A}\frac{\partial A}{\partial p} + i\frac{\partial \theta}{\partial p}} \tag{12.57}$$

It is a property of functions of complex variables that, provided they possess derivatives at all, then the value of the derivative is the same regardless of the direction in which the limiting point is approached. This fact is expressed by

$$\left.\begin{aligned} \frac{\partial A}{\partial p} &= \frac{\partial A}{\partial \alpha} = i\frac{\partial A}{\partial \omega} \\ \frac{\partial \theta}{\partial p} &= \frac{\partial \theta}{\partial \alpha} = i\frac{\partial \theta}{\partial \omega} \\ \text{and } \delta p &= \delta\alpha + i\delta\omega \end{aligned}\right\} \tag{12.58}$$

and provides means by which the real and imaginary parts of (12.57) may be separated to yield the two equations

$$\delta\alpha = \frac{\left[\frac{1}{A}\frac{\partial A}{\partial \omega}\left(\frac{\partial \theta}{\partial V}\delta V + \frac{\partial \theta}{\partial a}\delta a\right) - \frac{\partial \theta}{\partial \omega}\left(\frac{1}{A}\frac{\partial A}{\partial V}\delta V + \frac{1}{A}\frac{\partial A}{\partial a}\delta a\right)\right]}{\left(\frac{1}{A}\frac{\partial A}{\partial \omega}\right)^2 + \left(\frac{\partial \theta}{\partial \omega}\right)^2} \tag{12.59}$$

and

$$\delta\omega = \frac{-\left[\frac{1}{A}\frac{\partial A}{\partial \omega}\left(\frac{1}{A}\frac{\partial A}{\partial V}\delta V + \frac{1}{A}\frac{\partial A}{\partial a}\delta a\right) + \frac{\partial \theta}{\partial \omega}\left(\frac{\partial \theta}{\partial V}\delta V + \frac{\partial \theta}{\partial a}\delta a\right)\right]}{\left(\frac{1}{A}\frac{\partial A}{\partial \omega}\right)^2 + \left(\frac{\partial \theta}{\partial \omega}\right)^2} \tag{12.60}$$

The variable p in general may be written as the sum of α and $i\omega$. With the remembrance that p is the differential operator d/dt and that a set of linear equations expresses the transient condition, it is evident that the current will have the form Ie^{pt} which is equivalent to $Ie^{(\alpha+i\omega)t}$. Inspection of this shows that the real part of p, namely α, determines whether the currents in the system are going to build up with time, or die away with time, or remain constant, depending respectively upon whether α is greater than zero, is less than zero, or is actually equal to zero. With this in mind we see that (12.59) and (12.60) state the change in α and ω respectively which would result from some change in the circuit condition. Initially the circuit was oscillating in a steady manner so that α was zero and ω had some particular value. A change in V then occurred. This produced a change in the amplitude accompanied by a change in the frequency as expressed by (12.60) and a change in the transient term α. Suppose now that the change in V were very small. Then in order for oscillations again to assume a steady value it is necessary for the amplitude "a" to change a sufficient amount to cause α to become zero. Thus in (12.59) we put $\delta\alpha$ equal to zero and solve for the required amplitude change. This may then be eliminated from (12.60) resulting in the final expression

$$\delta\omega = \frac{\frac{1}{A}\frac{\partial A}{\partial V}\frac{\partial \theta}{\partial a} - \frac{1}{A}\frac{\partial A}{\partial a}\frac{\partial \theta}{\partial V}}{\frac{1}{A}\frac{\partial A}{\partial \omega}\frac{\partial \theta}{\partial a} - \frac{1}{A}\frac{\partial A}{\partial a}\frac{\partial \theta}{\partial \omega}}\delta V \tag{12.61}$$

which gives the frequency change $\delta\omega$ in terms of the change of the independent variable δV.

12.62 Frequency Stability of Conventional Oscillator

In applying this equation to the oscillator circuit, Fig. 12.24, we must first set up the conditions for oscillations. The $\mu\beta$ equation is

$$\mu\beta = \frac{\mu X_1 X_2 R_g}{i[X_s R_p R_g - X_1 X_2 X_3] - [R_p X_2(X_1 + X_3) + R_g X_1(X_2 + X_3)]} \tag{12.62}$$

The oscillating conditions $\mu\beta = 1$ requires

$$X_s R_p R_g = X_1 X_2 X_3$$

and

$$\mu X_1 X_2 R_g + R_p X_2 (X_1 + X_3) + R_g X_1 (X_2 + X_3) = 0 \qquad (12.63)$$

It will be assumed that the following relations exist:

$$\mu = f_1(V),\ R_g = f_2(a),\ X_s = X_1 + X_2 + X_3 = f_3(\omega),\ R_p = \text{a constant}$$

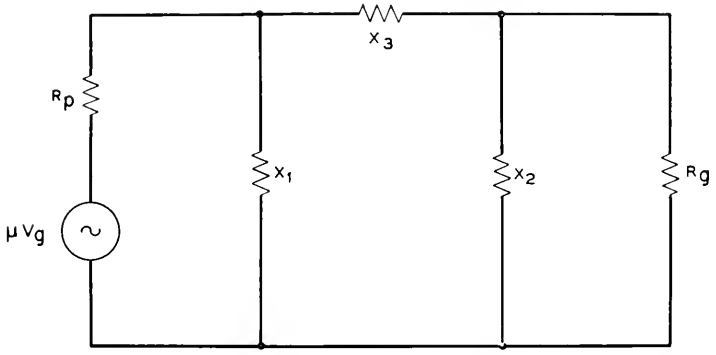


Fig. 12.24—Equivalent oscillator circuit analyzed for frequency stability

Then we obtain from (12.63)

$$\frac{1}{A}\frac{\partial A}{\partial V} = \frac{1}{\mu}\frac{\partial \mu}{\partial V} \qquad \frac{\partial \theta}{\partial V} = 0$$

$$\frac{1}{A}\frac{\partial A}{\partial a} = \frac{1}{R_g}\frac{\partial R_g}{\partial a}\left[1 + \frac{X_2 + X_3}{\mu X_2}\right] \qquad \frac{\partial \theta}{\partial a} = -\frac{1}{R_g}\frac{\partial R_g}{\partial a}\frac{X_0 R_p}{\mu X_1 X_2}$$

$$\frac{1}{A}\frac{\partial A}{\partial \omega} = \frac{1}{X_1}\frac{\partial X_1}{\partial \omega}\left[1 + \frac{R_p X_2 + R_g(X_2 + X_3)}{\mu R_p X_2}\right]$$

$$+ \frac{1}{X_2}\frac{\partial X_2}{\partial \omega}\left[1 + \frac{R_p(X_2 + X_3) + R_g X_1}{\mu R_p X_1}\right] \qquad (12.64)$$

$$+ \frac{1}{X_3}\frac{\partial X_3}{\partial \omega}\left[\frac{X_3(R_p X_2 + R_g X_1)}{\mu R_p X_1 X_2}\right]$$

$$\frac{\partial \theta}{\partial \omega} = -\frac{R_p}{\mu X_1 X_2}\left[(X_1 - X_s)\frac{1}{X_1}\frac{\partial X_1}{\partial \omega} + (X_2 - X_s)\frac{1}{X_2}\frac{\partial X_2}{\partial \omega}\right.$$

$$\left. + (X_3 - X_s)\frac{1}{X_3}\frac{\partial X_3}{\partial \omega}\right]$$

By substitution of these values in (12.61) and disregard of X_s in comparison with all other X's the equation for frequency stability is obtained as

$$\frac{d\omega}{dV} = \frac{\frac{1}{\mu}\frac{\partial \mu}{\partial V} X_1 X_2 X_3}{\left(1 - \frac{X_1}{\mu X_2}\right)\left(\frac{\partial X_1}{\partial \omega} + \frac{\partial X_2}{\partial \omega} + \frac{\partial X_3}{\partial \omega}\right) R_p R_g} \tag{12.65}$$

From this we learn that the values of the reactances X_1, X_2, and X_3 should be small and the values of R_p and R_g large to give small changes in ω when V is varied. These variables are more or less limited, however, by the conditions necessary for sustained oscillations according to equation (12.63). It is important to notice that the denominator of (12.65) contains functions which do not appear in equation (12.63) and hence may be of any value. These factors are the rates of change of the various reactances with frequency. For given values of circuit constants, the equation shows that the *frequency stability increases as these rates of change increase.*

12.63 Frequency Stability Coefficient of Crystals

The rate of change of the reactance of an element is referred to as the "frequency stability coefficient"* of the element. Expressed in per cent, we have for the frequency stability coefficient of a reactance

$$F(X) = \frac{dX}{d\omega} \cdot \frac{\omega}{X} \tag{12.66}$$

Let us now examine the frequency stability coefficient of a crystal which is used as the reactance X_2 when connected between grid and cathode of the tube and as X_3 when connected between grid and plate (See Fig. 12.24). The resistance of the crystal will be assumed to equal zero due to the negligible effect of the resistance variations upon the reactance for crystals with average Q and operated at a frequency not too near the anti-resonant frequency. (This may be observed in Fig. 12.21.)

The reactance of the crystal then is

$$X_c = -\frac{j}{\omega C_0}\frac{\omega^2 - \omega_1^2}{\omega^2 - \omega_2^2} \tag{12.67}$$

where

$\omega = 2\pi \times$ frequency
$\omega_1 = 2\pi \times$ resonant frequency
$\omega_2 = 2\pi \times$ anti-resonant frequency

* First suggested by N. E. Sowers.

By substitution of the relations

$$\frac{C_1}{C_0} = \frac{\omega_2^2 - \omega_1^2}{\omega_1^2} \quad \text{and} \quad -\frac{j}{\omega C_0} = X_0$$

into (12.67) we obtained

$$X_c = X_0\left[1 - \frac{C_1}{C_0}\frac{\omega_1^2}{\omega_2^2 - \omega^2}\right] \tag{12.68}$$

and by differentiation

$$\frac{dX_c}{d\omega} = -\frac{X_0}{\omega}\left[1 - \frac{C_1}{C_0}\frac{\omega_1^2}{\omega_2^2 - \omega^2}\right] - X_0\frac{C_1}{C_0}\left[\frac{\omega_1^2 2\omega}{(\omega_2^2 - \omega^2)^2}\right] \tag{12.69}$$

Multiply by $\frac{\omega}{X_c}$ to obtain the stability coefficient

$$F(X_c) = \frac{\omega}{X_c}\frac{dX_c}{d\omega} = -\frac{X_0}{X_c}\left[1 - \frac{C_1}{C_0}\cdot\frac{\omega^2}{\omega_2^2 - \omega^2}\right] - \frac{X_0}{X_c}\cdot\frac{C_1}{C_0}\cdot\frac{\omega^2}{\omega_1^2}\cdot\frac{2\omega_1^4}{(\omega_2^2 - \omega^2)^2} \tag{12.70}$$

and eliminate ω by substituting in (12.70) the relations obtained from equation (12.68). These are

$$\left.\begin{aligned} \frac{\omega_1^2}{\omega_2^2 - \omega^2} &= \left(1 - \frac{X_c}{X_0}\right)\frac{C_0}{C_1} \\ \text{and} \\ \frac{\omega^2}{\omega_1^2} &= \frac{C_1}{C_0} + 1 - \frac{C_1}{C_0}\frac{1}{1 - \frac{X_c}{X_0}} \end{aligned}\right\} \tag{12.71}$$

Thus

$$F(X_c) = -1 - 2\frac{X_0}{X_c}\left(1 - \frac{X_c}{X_0}\right)^2\left[1 + \frac{C_0}{C_1} - \frac{1}{1 - \frac{X_c}{X_0}}\right] \tag{12.72}$$

which may be written

$$F(X_c) = -\frac{X_c}{X_0}\left[1 + \left(1 - \frac{X_0}{X_c}\right) + 2\frac{C_0}{C_1}\left(1 - \frac{X_0}{X_c}\right)^2\right] \tag{12.73}$$

The stability coefficient of a coil and condenser $F(X)'$ may be obtained from (12.73) by letting $C_1 = \infty$. Then

$$F(X)' = -\frac{X_c}{X_0}\left[1 + \left(1 - \frac{X_0}{X_c}\right)\right] \tag{12.74}$$

The comparative stability of the crystal and tuned circuit is given by the ratio

$$\frac{F(X_c)}{F(X)'} = 1 + 2\frac{C_0}{C_1}\cdot\frac{\left(1 - \frac{X_0}{X_c}\right)^2}{1 + \left(1 - \frac{X_0}{X_c}\right)} \tag{12.75}$$

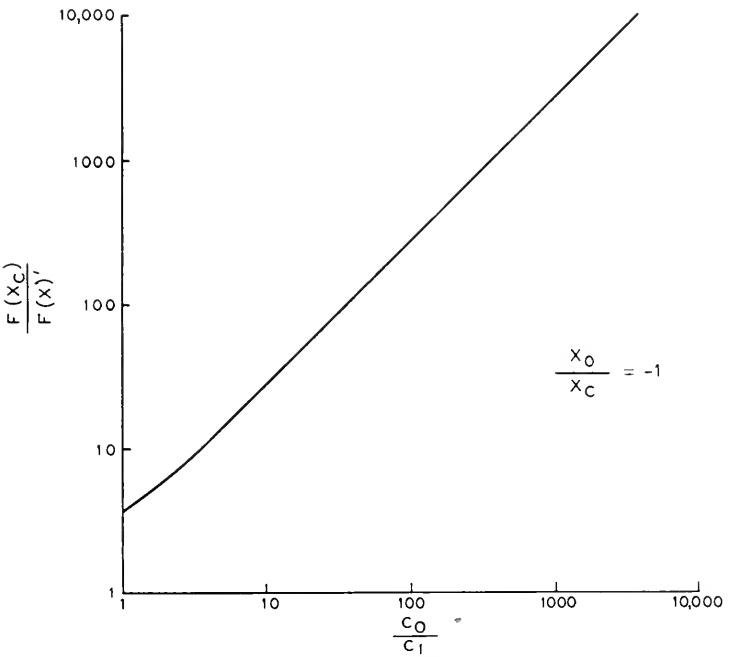


Fig. 12.25—The stability coefficient of a crystal as compared to a coil and condenser for variations of the ratio of capacitances

This ratio is plotted in Fig. 12.25 for $\frac{X_0}{X_c} = -1$ and with $\frac{C_0}{C_1}$ the independent function. It is apparent that the value of $\frac{C_0}{C_1}$ of a crystal is the factor which determines its frequency stability for given operating conditions. For an AT crystal in an air gap holder, the ratio of capacitances is of the order of 10^3 and its stability coefficient is therefore 2.6×10^3 greater than for a simple anti-resonant circuit. Since this is so much greater than the stability coefficients for the other reactances which appear in the denominator of equation (12.65) it represents the order of magnitude of improvement of the frequency stability of an oscillator obtained by the use of a crystal.

The fact that the frequency stability of a crystal oscillator is a function of $\frac{C_0}{C_1}$ explains why a BT cut crystal is in general more stable than an AT cut. The two may be made equal, however, by adding capacitance across the AT cut crystal.

Actually we have compared the frequency stability obtained by the use of one type of circuit (the equivalent crystal circuit) with one of a different configuration obtained by making $C_1 = \infty$. In practice this is usually the case since C_1 must be large to obtain oscillations when using coils and condensers. The limiting factor is therefore the value of $\frac{C_0}{C_1}$ at which oscillations stop and this is determined by the Q of the circuit elements as shown in the next section which deals with activity. It will be shown that the Q required is proportional to $\frac{C_0}{C_1}$ and therefore the maximum frequency stability that can be obtained is directly related to Q.

12.70 Relation Between Crystal Quality and Amplitude of Oscillations

The activity of a crystal is usually thought of as the relative amount of grid current produced in an oscillator circuit. This method of defining activity affords a means of comparing the quality of one crystal with another for a particular set of conditions. The disadvantages are first; it is only a relative measure, and second; it is not possible to compute the activity as thus defined by any method of oscillator analysis so far presented. Curves have been shown of amplitude of oscillations as a function of certain circuit variables, but these represent only qualitative changes associated with plate resistance variations. The first objection has been somewhat rectified by the use of reference oscillators* in which all the circuit elements including the tubes have been carefully matched. There is still the difficulty, however, of comparing crystals of different frequencies for it cannot be assumed that the measurements are independent of this variable. It would be more desirable to have some absolute measure of activity and particularly one which would lend itself to convenient computation from readily measurable constants of the crystal.

12.71 Definition of Crystal Quality for Oscillator Purposes

In deriving an expression for the quality of a crystal, it is convenient to use the negative resistance concept of the oscillator as described in section 12.40. The equations are general and in a form which admit of separating

* Developed by G. M. Thurston.

the crystal from the oscillator circuit. Equation (12.44) which gives the conditions necessary for oscillations to exist, may be written in the form

$$\omega C_t \rho \leqq \frac{\omega L_c}{R_c} \tag{12.76}$$

In order for oscillations to start, the right side of this equation must be equal to or greater than the left side. If it is greater, oscillations build up causing ρ to increase until the equality is satisfied. The difference between these two terms before oscillations start is therefore a relative measure of the final amplitude for a particular oscillator. The absolute value of amplitude cannot be obtained from equation (12.76) since we do not know the relation between ρ and amplitude. However the greater the magnitude of $\frac{\omega L_c}{R_c}$ the greater will be the amplitude of oscillations for a given set of oscillator conditions. This term may therefore be considered a measure of crystal quality. It is the effective Q of the crystal unit as measured at its two terminals and at the operating frequency. To distinguish this from the Q of the crystal as usually spoken of, it will be called φ_c.

In the same respect the left side of equation (12.76) may be thought of as a measure of quality of the oscillator circuit, that is, $\rho\omega C_t = \frac{1}{\varphi_g}$, then (12.76) becomes

$$\varphi_c \varphi_g \geqq 1 \tag{12.77}$$

12.72 Figure of Merit of Crystals $-M$

It is very inconvenient to use φ_c as a figure of merit of the crystal because it is a complex function of the constants of the crystal circuits, Figure 12.3, and also the frequency. The computation of φ_c from such measurable characteristics as frequency of resonance f_1, frequency of anti-resonance f_2, resonant resistance R_1, and static capacity C_0, requires considerable time and effort.

It is highly desirable that a simple, easily determined expression for a figure of merit be found. The steps to indicate a suitable one are as follows:

The equation for φ_c in terms of measurable quantities for computing it is derived from equation (12.27) and given by the formula

$$\varphi_c = \frac{\omega L_c}{R_c} = \frac{-1 - \dfrac{\omega L_1^2}{R_1^2}\dfrac{(\omega^2 - \omega_1^2)(\omega^2 - \omega_2^2)}{\omega^4}}{\dfrac{\omega L_1}{R_1}\dfrac{\omega_2^2 - \omega_1^2}{\omega^2}} \tag{12.78}$$

By letting

$$M = \frac{\omega L_1}{R_1}\frac{\omega_2^2 - \omega_1^2}{\omega^2} \tag{12.79}$$

and

$$n = \frac{\omega_2^2 - \omega^2}{\omega_2^2 - \omega_1^2} \tag{12.80}$$

and with the assumption that over the narrow frequency range that the crystal will operate

$$\frac{\omega_2^2 - \omega_1^2}{\omega_1} \cong \frac{\omega_2^2 - \omega_1^2}{\omega_1 \omega} \tag{12.81}$$

equation (12.78) is reduced to

$$\varphi_c = -\frac{1 + nM^2(n - 1)}{M} \tag{12.82}$$

In this equation it will be observed are only two variables, namely, n which varies widely as the frequency is varied between f_1 and f_2 and M which is substantially constant over the same range.

A set of curves is plotted in Fig. 12.26 for a hypothetical set of crystals having values of M of 1, 2, 5, and 10 with n varied over a range that falls between measured frequencies f_1 and f_2. Studies will show that whenever M increases φ_c will increase. M is readily calculated from measured constants as seen from the following: From (12.79)

$$M = \frac{\omega L_1}{R_1} \frac{\omega_2^2 - \omega_1^2}{\omega^2} = \frac{\omega_1 L_1}{R_1} \frac{\omega_2^2 - \omega_1^2}{\omega_1 \omega} \tag{12.83}$$

With the assumption in (12.81)

$$M = \frac{\omega_1 L_1}{R_1} \frac{C_1}{C_0} = \frac{1}{\omega_1 C_0 R_1} \tag{12.84}$$

which is a simple expression containing three of the four measured quantities mentioned above, and which bears a direct relation to activity for a given value of the frequency variable n. M is the new figure of merit of the crystal.

Figure 12.26 contains a further indication which is useful on occasions. Here φ_c is not positive at any frequency unless M is greater than 2. But φ_c must be positive for the crystal to oscillate in the two general types of circuits considered here.* It provides a measurable index to separate completely non-useful crystals from those that can oscillate in a given circuit.

Equation (12.84) may be written

$$M = \frac{Q}{r} \tag{12.85}$$

* For a description of oscillator circuits which do not require the crystal to exhibit a positive reactance see: "A New Direct Crystal-Controlled Oscillator for Ultra-Short-Wave Frequencies" by W. P. Mason and I. E. Fair, *Proc. I.R.E.*, Vol. 30, p. 464, Oct. 1943.

where Q is the Q of the crystal and r its ratio of capacitances. Thus the figure of merit involves the dissipation in the crystal determined by Q and the piezo-electric effect determined by r.[13]

It was pointed out in the preceding section that the frequency stability increases as r is increased. The above equation shows that Q must increase

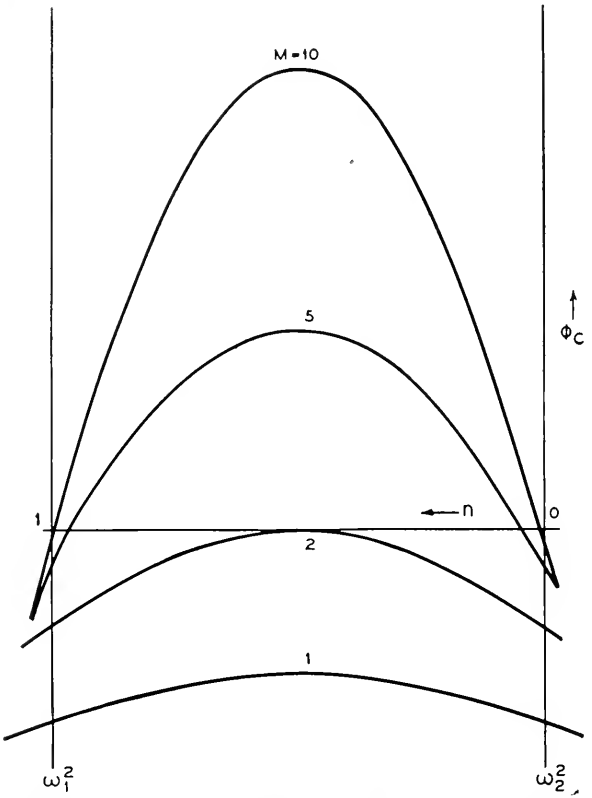


Fig. 12.26—The dependence of the quality function ϕ_c of a crystal upon frequency and figure of merit M

proportionately if the same figure of merit is to be maintained. The frequency stability obtainable in a particular oscillator is therefore limited by the Q of the crystal and the frequency stability coefficients should be compared on this basis.

12.80 Activity of Crystals

In deriving a figure of merit for crystals as oscillators, it was found that the amplitude of oscillations in a given circuit not only depends upon M but also it is a function of frequency relative to the resonant frequency

of the crystal. This may be explained by referring to Fig. 12.27 which shows curves of the reactance X_c of the crystal plotted as a function of frequency. The frequency at which oscillations occur depends principally upon the value of circuit capacitance C_t. Equation (12.43) shows that the frequency must adjust itself to a value at which C_t resonates with the reactance of the crystal. This value of reactance is represented on the curve as X_{co} and the corresponding frequency of oscillations as f_o. The

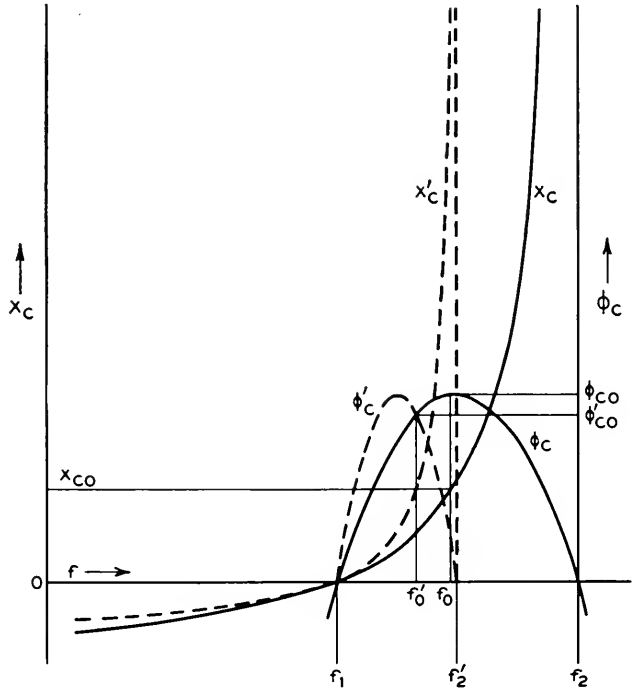


Fig. 12.27—Two crystals having the same figure of merit but with different reactance characteristics X_c and X_c' will operate with different amplitudes according to the relative values of φ_{co} and φ_o' respectively

circuit capacitance (or X_{co}) has been so chosen in the illustration that f_o lies equidistant between the resonant frequency f_1 and the anti-resonant frequency f_2. Then at this value of f_o, φ_c is a maximum, as shown by the φ_c curve, and for a given value of φ_g this will result in the greatest activity. Now let the capacitance C_0 of the crystal be increased. The frequencies f_1 and f_2 will then become closer and at the same time the height of the φ_c curve is reduced. Assume also that the Q is increased in order to maintain the same value of M and hence the same maximum value for φ_c. The reactance-frequency curve for the modified crystal and the corresponding

curve for φ_c' are represented by the dotted curves. Note that the oscillating frequency f_o' for the new curve is closer to f_2' than it is for f_1, therefore φ_{co}' is less than φ_{co} and the amplitude of oscillations will be less. Thus two crystals may have the same value of M but will not give the same output unless operated at the same relative frequency with respect to their resonant and anti-resonant frequencies. It would not have been possible to increase the oscillator output by increasing C_t so as to lower the frequency f_o' because by doing so φ_g is decreased more rapidly than φ_c' is increased, and the result would be a further decrease of activity. It is therefore necessary in deriving an expression for activity to include the variable of relative frequency or n which we have shown to be a function of the reactance of the circuit and the crystal constants.

12.81 Derivation of Performance Index of Crystals—*PI*

It will be assumed in the first derivation that the negative resistance ρ of the circuit is much greater than the effective resistance of the crystal R_c under stable oscillating conditions. Equation (12.43) which expresses the frequency then becomes

$$\omega \cong \sqrt{\frac{1}{L_c C_t}} \tag{12.86}$$

This leads to a very simple solution from which a more exact expression is later obtained.

Equation (12.44), which expresses conditions necessary for oscillations, may be written

$$|\rho| \leqq \left| \frac{\omega L_c}{\omega C_t R_c} \right| \tag{12.87}$$

As before, the numerical difference between ρ and the right side of the equation is a measure of activity. In fact, the right side of the equation may be considered to be an expression of the activity performance provided the terms are themselves independent of ρ. This is not quite true, since previous sections show that the capacitance C_t is not entirely independent of the activity. (See equations (12.30) and (12.46).) However, this effect may be considered negligible for most practical purposes and the value of the right side of (12.87) called the Performance Index (*PI*) of the crystal. From equations (12.86) and (12.87) the performance index is found to be

$$PI = \frac{1}{R_c \omega^2 C_t^2} \tag{12.88}$$

This equation may be greatly in error under operating condition which makes R_c large compared to ρ. Also R_c varies rapidly with frequency and is

difficult to evaluate. R_c is most readily eliminated from the equation by revising the picture slightly. With reference to the simplified oscillator circuit, Fig. 12.22B, it is apparent that the static crystal capacitance C_0 and the circuit capacitance C_t may be combined. This leaves for the crystal branch the inductance L_c' (different from L_c) which is a function of frequency and the resonant resistance of the crystal R_1 which is not a function of frequency. Now R_1 may be assumed small compared to ρ with considerable accuracy. It is only necessary, then, to replace C_t in equation (12.88) with $(C_0 + C_t)$ and R_c by R_1. This equation then becomes

$$PI = \frac{1}{R_1 \omega^2 (C_0 + C_t)^2} \tag{12.89}$$

An exact equation for PI is derived in section 12.83 and it is shown that the error in the simple expression above will in most cases be very small.

An approximate equation for the relation between R_1 and R_c is obtained by dividing (12.88) by (12.89). We thus find

$$1 = \frac{R_1 (C_0 + C_t)^2}{R_c C_t^2} \tag{12.90}$$

or the effective resistance of the crystal at the operating frequency is

$$R_c = R_1 \left(\frac{C_0}{C_t} + 1 \right)^2 \tag{12.91}$$

Because of the approximation in equation (12.88) the equation for R_c above is accurate only when $\left(\frac{C_0}{C_t} + 1 \right)^2 \ll M^2$ as will be shown in section 12.83.

The expression for PI as given by (12.89) may be written

$$PI = \frac{1}{\omega^2 C_0^2 R_1 \left(1 + \frac{C_t}{C_0} \right)^2} \tag{12.92}$$

which is the most convenient form for calculating PI from the constants of the crystal and the oscillator circuit.

12.82 Relation Between M and PI

It was found that

$$M = \frac{1}{\omega_1 C_0 R_1} \tag{12.93}$$

and this is essentially equal to $\frac{1}{\omega C_0 R_1}$ over the narrow frequency range considered. Therefore,

$$PI = \frac{M}{\omega C_0 \left(1 + \frac{C_t}{C_0} \right)^2} \tag{12.94}$$

This gives a relation between the performance index and the figure of merit of the crystal.

Another useful relation between M and PI is obtained from (12.89) and (12.93). Equation (12.93) may be written

$$M = \frac{X_0}{R_1} \tag{12.95}$$

This is of the same form as the Q of a coil when X_0 is considered to be the reactance of the coil and R_1 its resistance. Like the Q of a coil M is essen-

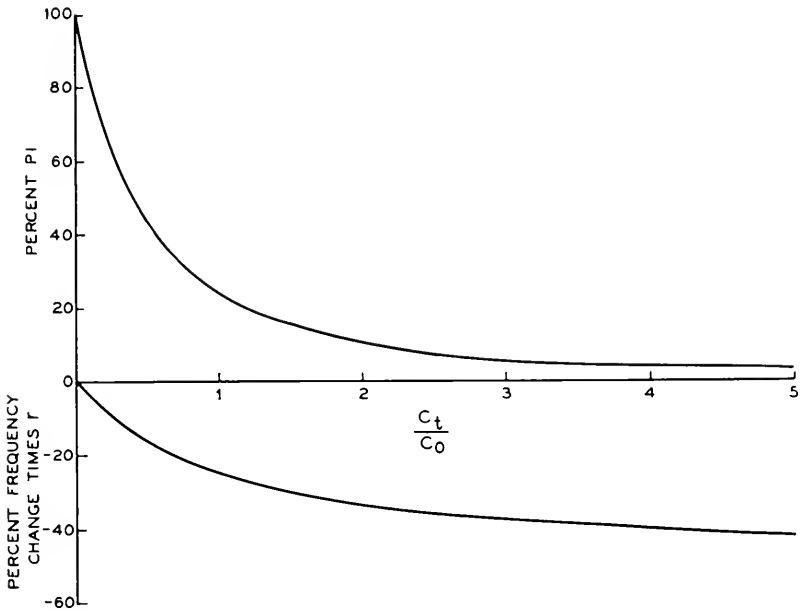


Fig. 12.28—The change in PI and oscillating frequency of a crystal as the shunt capacitance is increased

tially constant over a wide frequency range. Now if we let C_t approach zero in (12.89) it becomes

$$PI = \frac{1}{R_1 \omega^2 C_0^2} = \frac{X_0^2}{R_1} \tag{12.96}$$

This equation for PI is of the same form as the anti-resonant impedance of a coil and condenser in parallel, and like this impedance it changes rapidly with frequency. The maximum value of PI is therefore X_0 times M and is obtained when $C_t = 0$. Figure 12.28 shows a curve of % PI plotted as a function of $\frac{C_t}{C_0}$. This curve represents the change in activity as capacitance is added across the crystals (increase in C_t).

12.83 Exact Expressions for PI and R_e

The error in PI caused by the assumption that the frequency is independent of the crystal resistance R_1, that is, by use of approximate equation (12.86) for the frequency, may be investigated as follows:

The impedance of the crystal and C_t in parallel is given by

$$Z = \frac{1}{\omega(C_0 + C_t)(1 + m^2 P^2)} [P - j(1 + mP^2(m - 1))] \quad (12.97)$$

where

$$P = \frac{MC_0}{C_0 + C_t} \qquad m = \frac{\omega_3^2 - \omega^2}{\omega_3^2 - \omega_1^2}$$

ω_3 = 2π times frequency of anti-resonance of the crystal and C_t combination when $R_1 = 0$

ω_1 = 2π times frequency of resonance of the crystal and C_t combination when $R_1 = 0$

ω = 2π times operating frequency

(Note that P is the figure of merit of the crystal and C_t in parallel.) The imaginary part of Z is

$$X = -\frac{1 + mP^2(m - 1)}{\omega(C_0 + C_t)(1 + m^2 P^2)} \quad (12.98)$$

The condition for stable oscillations requires $X = 0$. For this condition

$$m = \tfrac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{1}{P^2}} \quad (12.99)$$

which defines the exact frequency of oscillation. The negative sign before the radical is used since the effective resistance is greater at this frequency, thus requiring less negative conductance for oscillation.

With P large ($m \to 0$) the frequency of oscillations coincides with ω_3 which is the same as given by the approximate frequency equation (12.86). The real part of (12.97) is

$$R = \frac{P}{\omega(C_0 + C_t)(1 + m^2 P^2)} \quad (12.100)$$

and when $m = 0$

$$R = \frac{P}{\omega(C_0 + C_t)} = \frac{M}{\omega C_0 \left(1 + \frac{C_t}{C_0}\right)^2} \quad (12.101)$$

This is identical to the expression for PI of equation (12.94). *PI is therefore the anti-resonant resistance of the crystal and capacitance C_t in parallel.* Substitution of the value of m as given by (12.99) into (12.100) gives the anti-resonant resistance at the oscillating frequency. Thus

$$R_o = \frac{P}{\omega(C_0 + C_t)}\left[1 - \frac{1 - \sqrt{1 - \frac{4}{P^2}}}{2}\right] \tag{12.102}$$

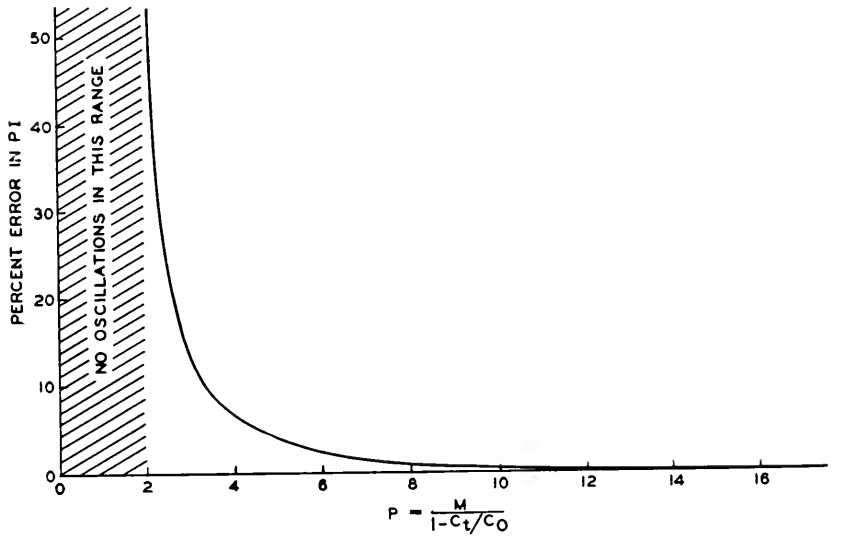


Fig. 12.29—The error in PI resulting from the use of approximate equation (12.94)

which is the exact expression for PI. The differential error resulting from the use of approximate equation (12.94) is then

$$\frac{PI - R_o}{R_o} = \frac{1 - \sqrt{1 - \frac{4}{P^2}}}{1 + \sqrt{1 - \frac{4}{P^2}}} \tag{12.103}$$

The per cent error as a function of P is shown in Fig. (12.29). The error diminishes rapidly with increase in P and is negligible for crystals that are of sufficient quality for most oscillator purposes.

Equation (12.91) for R_c is also approximate because of the assumption that the frequency is independent of R_c. A more exact expression will be derived.

The impedance of the crystal alone is

$$Z_c = \frac{1}{\omega C_0(1 + n^2 M^2)}\left[M - j(1 + nM^2(n - 1))\right] \tag{12.104}$$

where

M = the figure of merit of the crystal

$n = \dfrac{\omega_2^2 - \omega^2}{\omega_2^2 - \omega_1^2}$

ω_1 = 2π times frequency of resonance of the crystal
ω_2 = 2π times frequency of anti-resonance of the crystal
ω = the independent variable, 2π times frequency
C_0 = the static capacitance of the crystal

The resistive component of Z_c is

$$R_c = \frac{M}{\omega C_0(1 + n^2 M^2)} = \frac{R_1}{\dfrac{1}{M^2} + n^2} \tag{12.105}$$

In order to express R_c in terms of C_0 and C_t, the quantity n must be expressed in terms of these variables which define the oscillating frequency. This is accomplished as follows:

The equation of ratio of capacitances of a crystal is

$$\frac{C_1}{C_0} = \frac{\omega_2^2 - \omega_1^2}{\omega_1^2} \tag{12.106}$$

Similarly, when the capacitance C_t is placed across the crystal

$$\frac{C_1}{C_0 + C_t} = \frac{\omega_3^2 - \omega_1^2}{\omega_1^2} \tag{12.107}$$

where ω_3 is 2π times the anti-resonant frequency of the crystal and C_t in parallel. The ratio of (12.107) and (12.106) is

$$\frac{C_0}{C_0 + C_t} = \frac{\omega_3^2 - \omega_1^2}{\omega_2^2 - \omega_1^2} \tag{12.108}$$

The oscillating frequency is given by (12.99) in which m is as defined under (12.97). The oscillating frequency ω is therefore given by

$$\frac{\omega_3^2 - \omega^2}{\omega_3^2 - \omega_1^2} = \frac{1 - \sqrt{1 - \dfrac{4}{P^2}}}{2} \tag{12.109}$$

or

$$-\frac{\omega_1^2 - \omega^2}{\omega_3^2 - \omega_1^2} = \frac{1 + \sqrt{1 - \dfrac{4}{P^2}}}{2} \tag{12.110}$$

The angular frequency ω_3 is eliminated by multiplying this by (12.108). Thus

$$\frac{\omega_1^2 - \omega^2}{\omega_2^2 - \omega_1^2} = \frac{-C_0}{C_0 + C_t}\left[\frac{1 + \sqrt{1 - \frac{4}{P^2}}}{2}\right] \tag{12.111}$$

or

$$\frac{\omega_2^2 - \omega^2}{\omega_2^2 - \omega_1^2} = \frac{-C_0}{C_0 + C_t}\left[\frac{1 + \sqrt{1 - \frac{4}{P^2}}}{2}\right] + 1 \tag{12.112}$$

This is the value for n at the oscillating frequency and may be reduced to the form,

$$n = \left[\frac{1 - \sqrt{1 - \frac{4}{P^2}}}{2} + \frac{1 + \sqrt{1 - \frac{4}{P^2}}}{2\left(\frac{C_0}{C_t} + 1\right)}\right] \tag{12.113}$$

When this value of n is substituted in (12.105), the value of R_c is found to be

$$R_c = \frac{R_1}{\frac{1}{M^2} + \left[\frac{1 - \sqrt{1 - \frac{4}{P^2}}}{2} + \frac{1 + \sqrt{1 - \frac{4}{P^2}}}{2\left(\frac{C_0}{C_t} + 1\right)}\right]^2} \tag{12.114}$$

For crystals of usable quality $\frac{4}{P^2} << 1$ and by this assumption the equation reduces to

$$R_c = \frac{R_1}{\frac{1}{M^2} + \frac{1}{\left(\frac{C_0}{C_t} + 1\right)^2}} \tag{12.115}$$

This again reduces to (12.91) when $M^2 >> \left(\frac{C_0}{C_t} + 1\right)^2$.

12.84 Frequency Change Resulting from Paralleling Capacitance

It is often desirable to know how much the frequency of an oscillator may be changed by varying the capacitance C_t across the crystal. This is determined from (12.112) which gives the oscillating frequency as a function

of C_t. For practical considerations we may assume $\frac{4}{P^2} \ll 1$ which reduces the equation to

$$\frac{\omega_2^2 - \omega^2}{\omega_2^2 - \omega_1^2} = \frac{C_t}{C_0 + C_t} \tag{12.116}$$

From this and (12.106) we obtain

$$\frac{\omega_2^2 - \omega^2}{\omega_1^2} = \frac{C_1 C_t}{C_0 (C_0 + C_t)} \tag{12.117}$$

Since

$$\frac{\omega_2^2 - \omega^2}{\omega_1^2} = \frac{(\omega_2 - \omega)(\omega_2 + \omega)}{\omega_1^2} \cong \frac{2(\omega_2 - \omega)}{\omega_1}$$

then

$$\frac{\omega - \omega_2}{\omega_1} \cong \frac{-1}{2r\left(\frac{C_0}{C_t} + 1\right)} \tag{12.118}$$

where r is the ratio of the capacitances of the crystal.

A curve of per cent frequency change multiplied by r as a function of $\frac{C_t}{C_0}$ is shown on Fig. 12.28 for comparison with the associated PI change.

12.85 Relation between PI and Oscillator Activity

The relation between PI and activity obtained in a particular oscillator will now be examined. Let the curves of Fig. 12.30 represent the variations of ρ with amplitude for two oscillators A and B, or they might be for the same oscillator at widely different frequencies. These are characteristics of the oscillator circuits and may be of any shape. However, for oscillators with grid leak bias, the curves normally have no negative slopes. The rate of change of ρ depends upon the rate of change of μ and plate resistance of the vacuum tube as shown by (12.45) for input conductance.* Since ρ builds up to a value equal to PI we may plot PI for ρ. The grid current I_g is usually taken as a measure of amplitude. Therefore, Fig. 12.30 may be plotted as shown in Fig. 12.31 where PI is the independent variable. These curves are the characteristics of the oscillator circuits A and B with PI defining the quality of the crystal when used with a particular value of C_t. It is characteristic of oscillators to "saturate" as shown by the curves.

* It is also a function of grid resistance but this does not appear in the approximate equation (12.45) because of the assumption of no grid current. See Chaffee's[17] complete equation for input admittance.

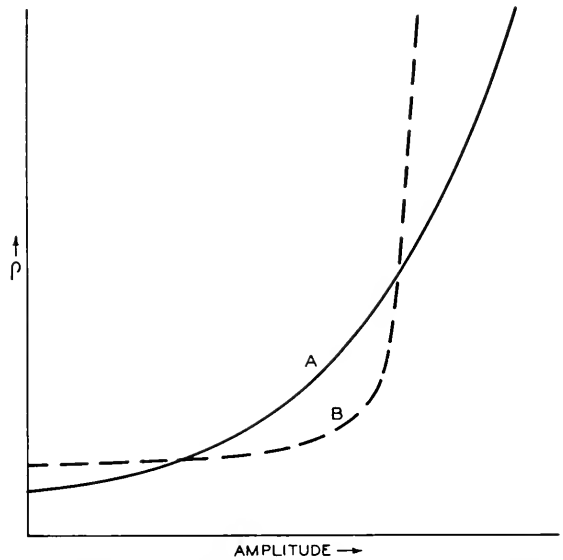


Fig. 12.30—Hypothetical curves illustrating the normal relation between the negative resistance of oscillator circuits and the amplitude of oscillations

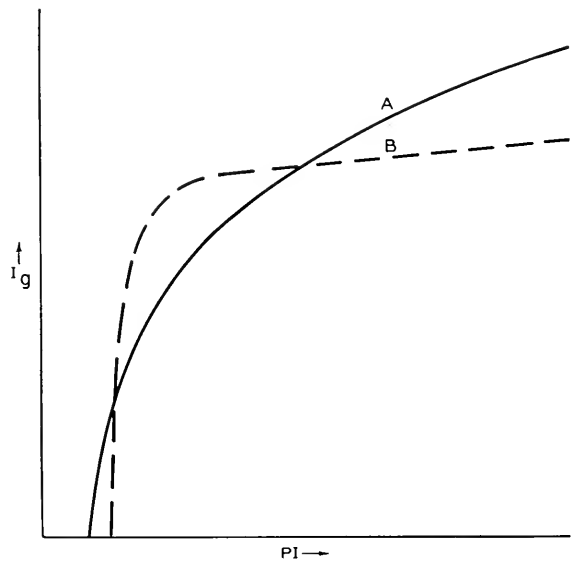


Fig. 12.31—By interchanging the coordinates of Figure 12.30 the curves will represent the relation between *PI* and oscillator grid current

Some oscillators saturate very rapidly and completely according to curve B and no further output is obtained regardless of the improvement in the crystal quality. For this reason it has not been possible in the past to separate the performance of the oscillator and the crystal since both were based upon the grid current as a measure of quality. By defining crystal activities and oscillator sensitivity in the manner outlined, the crystal and circuit can be studied separately. The per cent of crystals obtainable with PI above a certain value will be known and the design and improvement of oscillator circuits will be facilitated.

12.86 Use of PI in Crystal Design

The expression of PI in terms of the crystal constants and C_t as given by equations (12.89) or (12.92) assists in the design of crystals. As an example, the effect of changing the area of the crystal electrodes will be computed. The Q of a crystal is defined as

$$Q = \frac{1}{\omega_1 C_1 R_1} \tag{12.119}$$

By introduction of the ratio of capacitances of the crystal $r = \frac{C_0}{C_1}$ equation (12.119) becomes

$$Q = \frac{r}{\omega_1 C_0 R_1} \tag{12.120}$$

or

$$\frac{Q}{r} = \frac{1}{\omega_1 C_0 R_1} = M \tag{12.121}$$

Assuming Q and r do not vary, that is, disregarding effects such as secondary modes, change in damping produced by the mounting etc., and substituting (12.121) in (12.94) we obtain

$$PI = \frac{Q}{r} \cdot \frac{C_0}{\omega (C_0 + C_t)^2} \tag{12.122}$$

where $\frac{Q}{r}$ is considered constant. Differentiating (12.122) with respect to C_0 we find that PI is a maximum when $C_0 = C_t$. Since C_0 is proportional to the area of the electrodes this establishes the optimum area for a particular value of circuit capacitance.

The capacitance of BT-cut plates is 1.68 mmf. per square centimeter per megacycle.* Substitution of this for C_0 in (12.122) gives

$$PI = \frac{.268 \times 10^6 MA}{(1.68\, Af + C_t)^2} \tag{12.123}$$

* All frequencies are referred to the time interval of one second throughout this paper, i.e. *megacycles per second* is called simply *megacycles* as is customary in the radio field.

where A = the area of the crystal in square centimeters.
f = the frequency in megacycles
C_t = circuit capacitance in mmf.
M = figure of merit of the crystal (assumed constant)

Thus for crystals of a given area, the performance index should decrease as the frequency increases. Figure (12.32) shows the theoretical variations of PI as the function of the diameter of the electrodes of three frequencies and

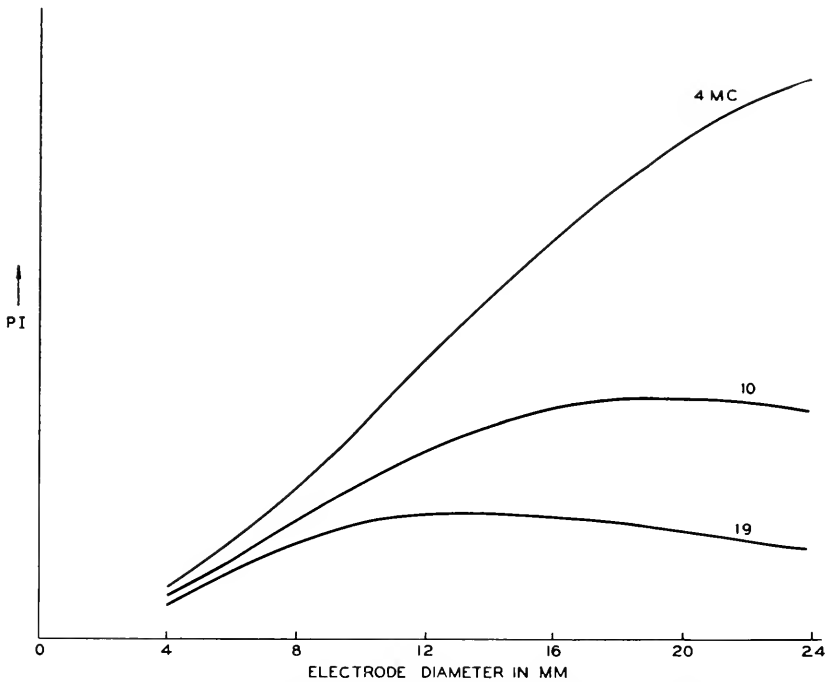


Fig. 12.32—Theoretical curves showing the relations of PI, electrode diameter, and crystal frequency for BT crystals and a circuit capacitance of 50 $\mu\mu$f

for a circuit capacitance of 50 mmf. The activity of a 4-megacycle crystal with 11-mm. diameter electrodes is about the same as a 10-megacycle crystal with 18 mm. electrodes. It must be remembered in making this comparison that it is assumed that the damping introduced by the mounting is the same in both cases. Actually the damping is much greater for low-frequency crystals of this type than for high-frequency ones and maximum PI occurs at some intermediate frequency as shown by the curves of Fig. 12.33. These curves show that the damping caused by the particular mounting used was small for frequencies above 6 megacycles but increases rapidly below this value.

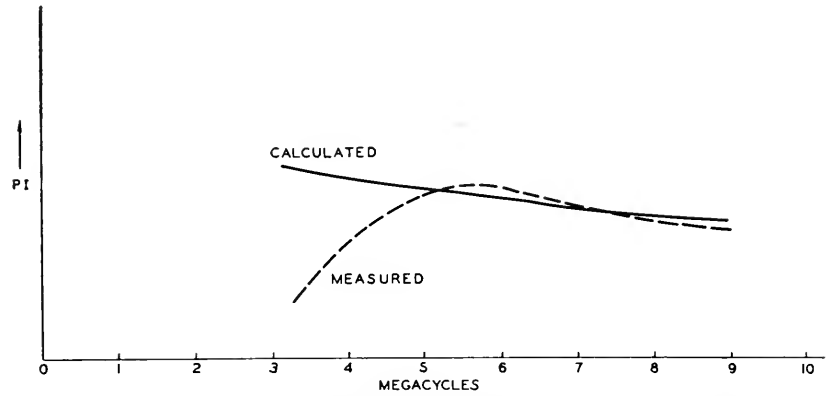


Fig. 12.33—Calculated and measured values of *PI* for *BT* crystals. The discrepancy is a measure of mounting loss

12.87 Measurement of *PI* and *M*

In all the discussions so far regarding the performance of crystals in oscillator circuits, the crystal has been represented by the equivalent circuit of Fig. 12.3 in which all the elements were considered constant. It is possible to obtain crystals in which this is essentially the case, but in general there are three secondary effects which complicate the picture. These are, first, the effect of other modes of vibration of the crystal, second, variations in the crystal constants resulting from variations in the amplitude of vibration, and third, the leakage or dielectric loss in the crystal holder. These factors will be considered in the order named.

Secondary modes of vibration affect the crystal for oscillator purposes only when the frequencies of these modes are sufficiently close to the principal one to alter its impedance characteristic in the frequency range of oscillation; that is, to alter the reactance as shown in Fig. 12.27 between the frequency f_1 and f_2 and the corresponding effective resistance between these two frequencies. With interfering modes present, the equivalent crystal circuit is so complicated as to make it impractical to compute *PI* or *M* from such measurable quantities as resonant resistance R_1, series resonant frequency f_1, anti-resonant frequency f_2, etc. For this reason it is necessary to measure the reactance and effective resistance of the crystal at the operating frequency in order to obtain a measure of crystal quality which will correlate with the crystal performance. For the same reason it is important when comparing oscillator circuits that the crystal should be operated at the same frequency in each case.

It is believed that the non-linear effect noticed in crystals when used as oscillators is produced by the changes in the mounting as the amplitude of

vibration varies. The *PI* of some clamped or pressure-mounted crystals has been found to vary as much as 50% with change in drive. Noticeable frequency change also occurs. A change in the nature of secondary modes as the amplitude is varied has also been observed. Some secondary modes which interfere with large amplitude of vibrations practically disappear when the amplitude is reduced. This may be explained by the fact that certain modes are damped out by the pressure of the mounting and with large amplitude of vibration the effect of the pressure is reduced.

The dielectric loss of the holder was considered negligible in the theory but it is found that certain phenolic holders have equivalent high-frequency leakage resistances less than 100,000 ohms. This resistance is in parallel with the crystal and will therefore reduce the *PI* according to the equation

$$PI = \frac{PI_c R_L}{PI_c + R_L} \tag{12.124}$$

where

PI = resulting PI
PI_c = calculated PI
R_L = equivalent high-frequency leakage resistance

Because of these secondary effects which are not negligible it is essential in measuring crystal activity that the frequency and voltage across the crystal be known. Standard test circuits should simulate operating conditions in this respect. With these considerations, a crystal *PI* meter has been developed in which the frequency and amplitude may be adjusted to correlate with various oscillators. The principle of operation and performance of this meter is described by C. W. Harrison.*

Bibliography

1. A. McL. Nicolson, U. S. Patent Nos. 1495429 and 2212845, filed in 1918.
2. W. G. Cady, The Piezo-Electric Resonator. I.R.E., Vol. 10, p. 83, April, 1922.
3. G. W. Pierce, Piezo-Electric Crystal Resonators and Crystal Oscillators Applied to the Precision Calibration of Wavemeters. Proc. Amer. Acad. Arts & Sci., Vol. 59, p. 81, 1923.
4. A. Crossley, Piezo-Electric Crystal Controlled Oscillators. I.R.E., Vol 15, p. 9, Jan., 1927.
5. K. S. Van Dyke, The Electrical Network Equivalent of a Piezo-Electric Resonator. Physical Rev., Vol. 25, p. 895, 1925.
   The Piezo-Electric Resonator and Its Equivalent Network. I.R.E., Vol. 16, p. 742, June, 1928.
6. E. M. Terry, The Dependence of the Frequency of Quartz Piezo-Electric Oscillators Upon Circuit Constants. I.R.E., Vol. 16, p. 1486, Nov., 1928.
7. J. W. Wright, The Piezo-Electric Crystal Oscillator. I.R.E., Vol. 17, p. 127, Jan. 1929.

* "The Measurement of the Performance Index of Quartz Plates," this issue of the *B.S.T.J.*

8. P. Vigoureux, Quartz Resonators and Oscillators. Published by H. M. Stationery Office, Adastral House, Kingsway, London, W.C. 2.
9. R. A. Heising, The Audion Oscillator. The Physical Review, N. S., Vol. XVI, No. 3, Sept. 1920.
10. L. P. Wheeler, An Analysis of a Pieze-Electric Oscillator Circuit, I.R.E., Vol. 19, p. 627, April 1931.
11. F. B. Llewellyn, Constant Frequency Oscillators. I.R.E., Vol. 19, p. 2063, Dec. 1931; B.S.T.J., Jan. 1932.
12. Issac Koga, Characteristics of Piezo-Electric Quartz Oscillators. I.R.E., Vol. 18, p. 1935, Nov. 1930.
13. K. Heegner, Gekoppelte Selbsterregte Kreise und Kristallozillatorën. E.N.T., Vol. 15, p. 364, 1938.
14. R. A. Heising, The Audion Oscillator, Journal of the American Institute of Electrical Engineers. April and May, 1920.
15. M. Boella, Performance of Piezo-Oscillators and the Influence of the Decrement of the Quartz on the Frequency of Oscillations. I.R.E., Vol. 19, p. 1252, July 1931.
16. H. J. Reich, Theory and Application of Electron Tubes. Page 313.
17. E. L. Chaffee, Equivalent Circuits of an Electron Triode and the Equivalent Input and Output Admittances, I.R.E., Vol. 17, p. 1633, Sept. 1929.
18. W. P. Mason, An Electromechanical Representation of a Piezo-Electric Crystal Used as a Transducer. I.R.E., Vol. 23, p. 1252, Oct. 1935.

The Measurement of the Performance Index of Quartz Plates

By C. W. HARRISON

15.00 Introduction

THE theory of the general behavior of crystals in oscillator circuits has been described by I. E. Fair[1]. In Fair's paper as well as in others[2], it has been pointed out that in the neighborhood of the operating frequency a crystal is equivalent to the circuit shown in Fig. 15.1A. The crystal possesses two resonant frequencies, a series resonant frequency determined by the effective inductance, L_1, and effective capacitance, C_1, and an anti-resonant frequency determined by these same elements plus the paralleling capacitance, C_0. This paralleling capacitance is the static capacitance between electrodes of the crystal and any capacitance connected thereto by the crystal holder and lead wires within the holder. The dotted resistor, R_L, shunting the equivalent crystal circuit represents the effective shunt loss of the holder. In the ideal case and in many practical instances this loss is negligible.

It is rather difficult to express the circuital merit of a crystal quantitatively in a single term such as has been found useful for inductances and capacitances. It is customary to express the circuital merit of these two elements in the form of the ratio of reactance to resistance. That is, for an inductance

$$Q = \frac{\omega L}{R} \tag{15.1}$$

and for a capacitance

$$Q = \frac{1}{\omega CR}. \tag{15.2}$$

For filter purposes, the Q of a crystal involving only the inductance, L_1, and resistance, R_1, of Fig. 15.1A is adequate to express its usefulness in certain parts of a filter network, but for oscillator purposes it is insufficient. At frequencies other than the series resonant frequency the paralleling capacitor C_0 together with the associated shunt loss of the holder enters into the determination of a crystal's performance. The term Q therefore is not completely indicative of the crystal performance. There has been devised, as pointed out in Fair's paper[1], a term called "figure of merit" for a crystal

[1] I. E. Fair, "Piezoelectric Crystals in Oscillator Circuits," this issue of the *B.S.T.J.*

[2] K. S. Van Dyke, "The Electrical Network of a Piezo-Electric Resonator", *Physical Review*, Vol. 25, pp. 895, 1925.

which involves all the elements in the effective crystal circuit, and this term is much more expressive of the quality of a crystal. The figure of merit is:

$$M = \frac{\omega L_1}{R_1}\frac{C_1}{C_0} = \frac{Q}{r} \tag{15.3}$$

where "r" is the ratio of the paralleling capacitance to the series branch capacitance. Figure of merit is useful for expressing the quality of a crystal

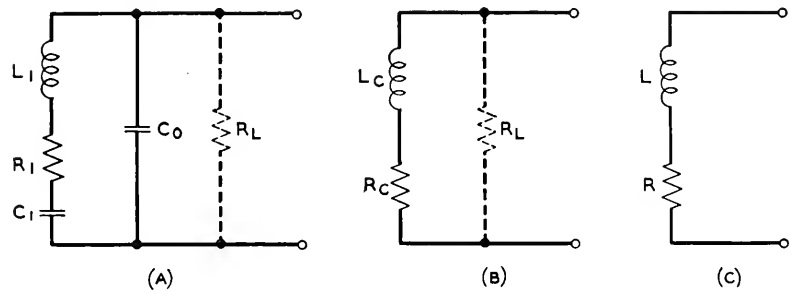


Fig. 15.1—Electrical equivalent circuits of a piezoelectric crystal—(A) At any frequency between the resonant frequency ω_1, and anti-resonant frequency ω_2; (B) and (C) at any specific frequency between ω_1 and ω_2.

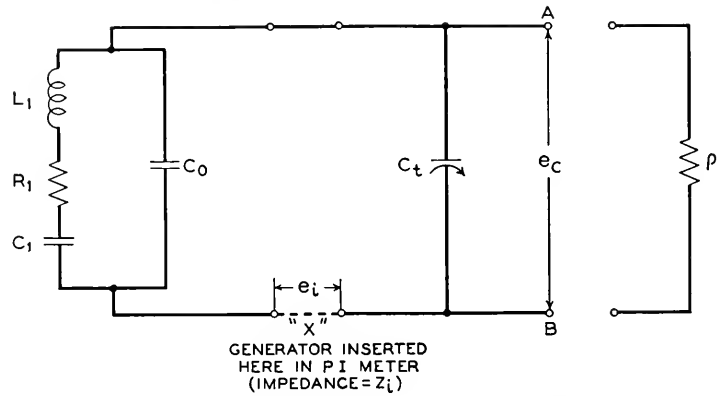


Fig. 15.2—Generalized oscillator circuit of the Pierce or Miller type.

in its holder or mount; however by definition it is independent of the value of R_L, and does not permit a ready evaluation of the performance of the crystal in an oscillator of the type that may be represented by Fig. 15.2. Any oscillator that operates the crystal in the positive region of the reactance vs. frequency characteristic exhibits capacitive reactance and negative resistance paralleled across the terminals to which the crystal is connected. The operation of the crystal when connected to an oscillator will be influenced by the magnitude of these two terms, and the combination must operate at

such a frequency that the total reactance is zero and at such an amplitude that the total resistance is zero. The performance of the crystal will therefore not depend solely upon its figure of merit, but will involve the impedance of the remainder of the oscillator. Up to the present time circuit design engineers have not devised standards or units to express the quality of their oscillator circuits without the crystal, so there are no corresponding circuital units of quality with which to correlate figures of merit of crystals to ascertain the suitability of one for the other.

It was a practice for many years for manufacturers to test crystals in a model of the oscillator in which the crystal was to be used. This required manufacturers to keep on hand models of all oscillators for which they expected to make crystals. To avoid the mounting number of such test oscillators a special test set was developed which could be adjusted to simulate any oscillator. By correlating various oscillator circuits to a set of adjustments on the test set, the actual model of the oscillator can be dispensed with. This special test set usually referred to as the '"D" spec. test set', eliminated the "file" of oscillators, and substituted a file of adjustment readings that would be their equivalent. However, the '"D" spec. test set' is still inadequate to the development engineer since it defines "activity" in terms of oscillator grid current rather than in terms of the electrical equivalent circuit of the crystal. The activity as expressed by grid current is a purely arbitrary standard and serves only as a means of determining the relative activity as against other crystals of the same frequency operated under the same circuit conditions.

The need for a system of measurement using units that are fundamental and not empirical has led to the proposal of "Performance Index". An instrument to make such measurements is to be described in this paper.

Specifically the Performance Index is

$$PI = \frac{\omega L}{\omega C_t R} \tag{15.4}$$

where C_t is the paralleling capacitance that is found in the oscillator circuit to which the crystal is attached, and L and R represent the effective inductance and resistance of the crystal as measured at the operating frequency indicated in Fig. 15.1C which is its equivalent *at that frequency*. If the loss in the holder is so low that the resistance, R_L, may be neglected, then PI may be expressed in other relations that are more useful such as,

$$\left.\begin{aligned} PI &= \frac{M}{\omega C_0 \left(1 + \frac{C_t}{C_0}\right)^2} \\ \text{or} \quad PI &= P^2 R_1 \end{aligned}\right\} \tag{15.5}$$

where the symbols R_1 and C_0 are as shown in Figs. 15.1A and 15.2 and P is expressed as

$$P = \frac{M}{1 + \frac{C_t}{C_0}} \tag{15.6}$$

With the effective capacitance, C_t, of the remainder of the oscillator added to the paralleling capacitance, C_0, in Fig. 15.2, the operating frequency will be that frequency at which the combination will exhibit a pure resistance at the terminals AB (excluding the generator "X" which is involved in the measuring technique). This leads to the definition:

The Performance Index is the anti-resonant resistance of the crystal and holder having in parallel with it the capacitance introduced by the remainder of the oscillator.

The Performance Index is therefore a term to express performance not in terms of the grid current of some particular oscillator, but in fundamental circuital units—impedance. The Performance Index is a term that may be used to compare performance of crystals at different frequencies. Its value is independent of plate voltage, grid leak resistance, or of plate impedance. It provides a measuring stick that should replace the "activity" figures of grid current in so far as the crystal is concerned. It paves the way for the oscillator circuit designers to come forth with standards of measurement for the oscillator circuit without the crystal in the hope that the two may be quantitatively associated and lend themselves to theoretical calculation of full oscillator performance.

15.10 Theory of Measurement

The problems of measurement are most readily explained by reference to Fig. 15.2. The crystal provides elements L_1, C_1, R_1 and C_0. The circuit of the oscillator provides an effective capacitance, C_t, which is composed of grid and lead wire capacitances plus capacitance introduced from the plate circuit. The frequency at which this combination exhibits anti-resonance as measured at AB is the oscillating frequency. The resistance when added to negative resistance, ρ, will be zero. Oscillations will start with ρ numerically smaller than the anti-resonant resistance measured at AB, but the amplitude of oscillations will increase causing ρ to increase until ρ and Z_{AB} are equal numerically. The primary problem is to measure the anti-resonant resistance at AB at the anti-resonant frequency with ρ disconnected.

Measurement of anti-resonant resistance directly is very difficult. The current flowing into an anti-resonant circuit is too small to measure with the usual meters. Other devices for measuring the current are likely to introduce paralleling capacitance that will vitiate the readings. The sug-

gested method of measurement utilizes a suitable driving voltage at "X" (Fig. 15.2) and a means to indicate the voltage at "X" as well as at points AB. From these and other measured constants, the anti-resonant impedance can be computed.

This method of measurement has its own difficulties, but it is believed corrections can be made to allow for errors introduced. Fundamentally, the series resonant frequency and the anti-resonant frequency are the same only when the resistances in the inductive and capacitive branches are equal. When the resistance is practically all in the inductive branch, which is true in this case, the impedance between terminals AB, at the series resonant frequency will exhibit capacitive reactance, though the total impedance will scarcely be different from that at the anti-resonant frequency. In the Performance Index meter, although the voltage is introduced in series with the circuit, the frequency is adjusted to the point of maximum voltage across AB, which further minimizes this frequency difference. A second error is inherently introduced by the loss in the crystal holder. This means that the series resonant frequency is also altered by the presence of this loss. Errors of any seriousness will result from the assumption that the series resonant and anti-resonant frequencies are identical only when the resistance in the inductive branch and loss in the crystal holder approach the effective crystal reactance in magnitude. These errors will be discussed in greater detail in a succeeding section.

The development and operation of a satisfactory meter to measure PI (Performance Index) depends upon a number of factors such as:

1. A method to determine capacitance, C_t, of the circuit (Fig. 15.2).
2. A generator "X" to produce the driving voltage e_i having variability in frequency and negligible internal impedance.
3. A current indicator that introduces a minimum of reactance and resistance.
4. A circuit or method to indicate PI directly, or with a minimum of calculations.
5. A number of other factors associated with the above which will be mentioned at the logical times.

To construct a measuring circuit to determine the anti-resonant impedance by means of a series circuit so as to avoid any unnecessary measurements and computations involves the following basic principle. Excluding ρ, Fig. 15.2 is essentially equivalent to the circuit used in Q meters. The ratio of voltage e_c to the driving voltage e_i is the voltage stepup or the Q of that part of the circuit containing the resistance. In this case, the resistance is in the crystal which at the operating frequency has an effective Q of

$$Q_1 = \frac{\omega L}{R} \tag{15.7}$$

where L and R represent the effective values of the crystal. Equation (15.4) will be found to embody the relation

$$PI = \frac{\omega L}{\omega C_t R} = Q_1 X_t \tag{15.8}$$

where X_t is the reactance of C_t, the capacitance introduced by the circuit at the operating frequency. If e_i is kept constant, e_c will at all times be proportional to Q_1. By insertion of an attenuator network, whose attenuation varies with frequency in the same manner as does the reactance of C_t, between terminals AB and the voltmeter, the meter indication will be proportional to the product of these quantities or *proportional to PI*. With

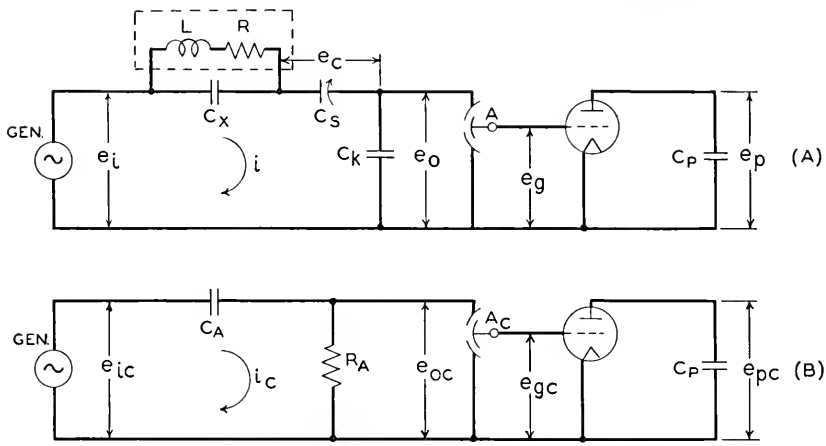


Fig. 15.3—Measuring circuits of the performance index meter.

suitable calibrations, therefore, it should be possible to get indications of PI as readily as is now done for Q.

The circuit shown in Fig. 15.2 is now best redrawn as in Fig. 15.3A. The crystal embodying elements L_1, C_1, R_1, and C_0 of Fig. 15.2 is now represented in Fig. 15.3A by the dotted rectangle and as having effective inductance, L, and effective resistance, R, both of which are functions of frequency. Capacitance, C_t, is simulated by capacitors C_x plus C_s and C_k in series where C_x represents the capacitance of the crystal socket. Zero internal impedance of the generator is simulated by maintaining the driving voltage constant at all times and at all frequencies. To facilitate explanation, the measured voltage e_i at the place shown is considered to be the driving voltage from a zero internal impedance generator.

Instead of using an ammeter to indicate current in the circuit, a voltmeter is utilized to measure voltage across an element under such conditions as

not to introduce disturbing capacitance. Splitting the series capacitance into two parts, C_s and C_k, the latter fixed and large compared to C_s, provides the impedance element across which the voltmeter is connected. The input capacitance of the voltmeter is incorporated in the magnitude of C_k. A capacitance attenuator, A, of known or calibrated values interposed on the input of the voltmeter enables the voltmeter to be used to indicate voltage ratios in terms of the attenuator calibration.

The measuring voltmeter and a shunting capacitance, C_p, are connected in the plate circuit of the amplifier tube, V-1. This circuit provides sufficient gain to furnish an output voltage of measurable magnitude and also provides an output voltage inversely proportional to frequency. The indication of the output voltage is proportional to PI.

The utilization of a vacuum tube in a circuit leading to a quantitative measuring instrument such as the voltmeter across C_p involves determination of tube constants or calibration. The determination of these constants is best evaluated experimentally. A calibrating circuit for that purpose is shown in Fig. 15.3B. A capacitance, C_A, of high impedance in series with comparatively negligible resistance, R_A, is connected across the driving voltage terminals of e_i with a voltmeter measuring e_i giving a reading e_{ic}. The second subscript "c" indicates calibration conditions. By connecting the input circuit of V-1 across this resistance, the attenuation variation with frequency of the $R_A - C_A$ network cancels the attenuation variation with frequency in the plate circuit of V-1. The ratio of e_{ic} to e_{pc} will then be independent of frequency. In the "calibrate" circuit (Fig. 15.3B), the capactior attenuator, A_c, interposed in the grid circuit is set at unity (minimum insertion loss) for a given deflection of the meter indicating e_p. In the operate circuit (Fig. 15.3A), the attenuator is readjusted so that voltage e_0 produces the reading of e_p as obtained in the calibrate position. The quantitative action of the amplifier then may be expressed in terms of C_p, R_A, C_A and a reading from the attenuator A, as will be shown later, and it is constant and independent of frequency. By placing this resulting constant in an equation, which will also be derived later, the value of PI may be determined in terms of such constant, of the reading of attenuator A, and of a reading on the scale of C_s that has been calibrated in terms of C_t.

To facilitate still further the operation of the PI meter, the voltage e_i is produced as shown in Fig. 15.4 by arranging for the oscillator to have its frequency controlled by the crystal through feedback from capacitor, C_k. Automatic volume control is provided such that the amplitude of e_i is essentially constant at all times and at all frequencies. The circuit is constructed to oscillate at the desired frequency, and adjustment for insuring this operation is provided in the form of a phase shifting circuit with variable

capacitor, C_r. After a crystal has been inserted in its proper place, oscillations will begin, but may be slightly above or below the resonant frequency of the crystal plus C_t. By adjustment of C_r the frequency can be shifted the slight amount necessary for resonance. This is observed by placing switch S in the PI position and making the adjustment to give maximum deflection of e_p.

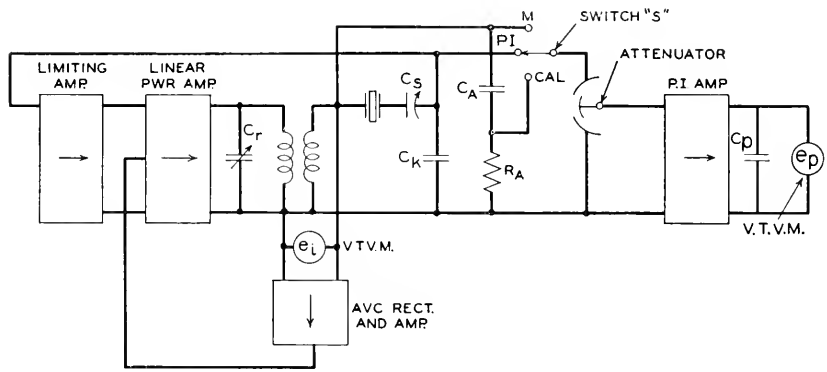


Fig. 15.4—Diagram of Performance Index meter.

15.20 Derivation of PI Circuit Equation

The following circuit relations derived from Figure 15.3 show first, that the ratio of e_p to e_i is a function of the Performance Index of the crystal, and second, that the calibration circuit permits an absolute evaluation of its magnitude.

At resonance, the effective circuit Q, designated as Q_2, is determined from

$$Q_2 = \frac{|e_c + e_0|}{e_i} = \frac{e_0}{e_i}\left(1 + \frac{C_k}{C_s}\right) \tag{15.9}$$

Since the circuit Q includes the capacitance of C_x as a part of the crystal, it is necessary to express Q_2 in terms of the crystal's properties (see Fig. 15.1). Since $Q_1\left(=\frac{X}{R}\right)$ of the crystal is independent of C_x, the relationship between Q_1 and Q_2 is readily obtained by equating the expressions for the anti-resonant impedance first, when C_x is considered to be in shunt with the series capacitor, C_t, and second, when C_x is considered as part of the crystal. This results in

$$Q_2 = Q_1 \frac{C_t - C_x}{C_t} \tag{15.10}$$

where

$$C_t = C_x + \frac{C_s C_k}{C_s + C_k} \tag{15.11}$$

enabling Q_1 to be expressed as

$$Q_1 = \frac{e_0}{e_i} \frac{C_t C_k}{(C_t - C_x)^2} \tag{15.12}$$

Expressing e_0 in terms of e_p, we have

$$|\mu e_g| = |i_p| \sqrt{r_p^2 + X_{C_p}^2} \tag{15.13}$$

$$|i_p| = \frac{e_p}{X_{C_p}} \tag{15.14}$$

hence

$$e_g = \frac{e_p \omega C_p}{G_m} \sqrt{1 + \left[\frac{X_{C_p}}{r_p}\right]^2} \tag{15.15}$$

and

$$\frac{e_0}{A} = e_g \tag{15.16}$$

With the above equations substituted in (15.12), we may express Q_1 as

$$Q_1 = \frac{e_p}{e_i} \omega \frac{C_p}{G_m} A \frac{C_k C_t}{(C_t - C_x)^2} \sqrt{1 + \left[\frac{X_{C_p}}{r_p}\right]^2} \tag{15.17}$$

Now

$$PI = \frac{Q_1}{\omega C_t} = Q_1 X_t \tag{15.18}$$

Therefore

$$PI = \frac{e_p}{e_i} A \frac{C_p}{G_m} \frac{C_k}{(C_t - C_x)^2} \sqrt{1 + \left[\frac{X_{C_p}}{r_p}\right]^2} \tag{15.19}$$

If

$$\sqrt{1 + \left[\frac{X_{C_p}}{r_p}\right]^2} \cong 1 \tag{15.20}$$

$$PI = \frac{e_p}{e_i} A \frac{C_p}{G_m} \frac{C_k}{(C_t - C_x)^2} \tag{15.21}$$

The simplified PI expression (15.21) assumes that the reactance of C_p is small compared to the plate resistance and plate load resistance of V-1. The evaluation of PI from this expression has three obvious difficulties:

(1) C_p/G_m is a quantity that is difficult to evaluate numerically, (2) the magnitude of PI is measured in terms of the ratio of the two voltages, e_i and e_p, and (3) the measurement is dependent upon the gain of a vacuum tube amplifier, V-1. These difficulties may be materially reduced in their consequence by an internal calibration circuit.

The internal calibration circuit (Fig. 15.3B) consists of a capacitor, C_A, and resistor, R_A, in series. If the reactance of C_A is very much greater than R_A, and the plate resistance of V-1 is very much greater than the reactance of C_p, the calibration is essentially independent of frequency.

The internal calibration circuit (Figure 15.3B) enables the evaluation of C_p/G_m to be carried out. The additional subscript, c, indicates "calibrate" conditions.

$$i_c = \frac{e_{ic}}{\sqrt{R_A^2 + X_{C_A}^2}} \tag{15.22}$$

$$e_{0c} = i_c R_A = \frac{e_{ic}\,\omega C_A R_A}{\sqrt{1 + \left[\frac{R_A}{X_{C_A}}\right]^2}} \tag{15.23}$$

$$\frac{e_{0c}}{A_c} = e_{gc} \tag{15.24}$$

Equation (15.15) remains the same for both "operate" and "calibrate" conditions with the exception of the second subscript reserved for the "calibrate" operation. Therefore, by solving for C_p/G_m we find

$$\left|\frac{C_p}{G_m}\right| = \frac{e_{gc}}{e_{pc}\,\omega \sqrt{1 + \left[\frac{X_{C_p}}{r_p}\right]^2}} \tag{15.25}$$

Equation (15.25) may be rewritten as (15.26), if (15.23) and (15.24) are substituted in (15.25)

$$\left|\frac{C_p}{G_m}\right| = \frac{e_{ic} R_A C_A}{e_{pc} A_c}\left[\frac{1}{\sqrt{1 + \left[\frac{X_{C_p}}{r_p}\right]^2}}\right]\left[\frac{1}{\sqrt{1 + \left[\frac{R_A}{X_{C_A}}\right]^2}}\right] \tag{15.26}$$

If (15.26) is substituted in (15.19), it is found that PI may be expressed as follows:

$$PI = \frac{e_p}{e_i}\frac{e_{ic}}{e_{pc}}\frac{A}{A_c} R_A C_A \left[\frac{C_k}{(C_t - C_x)^2}\right] \frac{1}{\sqrt{1 + \left[\frac{R_A}{X_{C_A}}\right]^2}} \tag{15.27}$$

The above equation involves only the original approximation that maximum current indicates resonance. If R_A and C_A are selected such that $R_A << X_{C_A}$ and if A_c equals unity, then

$$PI = \left[\frac{e_p}{e_i}\frac{e_{ic}}{e_{pc}}\right] A \frac{C_k}{(C_t - C_x)^2} R_A C_A \tag{15.28}$$

From this expression it can be seen that the PI measurement is independent of calibration of both the e_p and e_i vacuum tube voltmeters, provided that the same voltmeter scale factors are used for the "operate" and "calibrate" conditions. The absolute calibration then depends on the magnitude of A, R_A, C_A, C_k, C_t and C_x. The "multiply-by" factor that is to appear on the C_s dial is determined by the magnitude of $\frac{R_A C_A C_k}{(C_t - C_x)^2}$.

Accurate evaluation of this quantity by capacitance and resistance measurements is a little difficult since the denominator represents the square of the difference of two small capacitances. When C_t is large, the evaluation of this factor is helped considerably. This "multiply-by" factor may be experimentally determined by a voltage measuring means which permits an evaluation of this factor to a higher degree of accuracy. Substituting (15.11) in (15.28) we have

$$PI = \left[\frac{e_p}{e_{pc}}\frac{e_{ic}}{e_i}\right] A \frac{R_A C_A}{C_k}\left(1 + \frac{C_k}{C_s}\right)^2 \tag{15.29}$$

Now by shorting the crystal socket terminals (Fig. 15.3A) and applying a voltage e_1 at the e_i generator terminals of external origin (the crystal oscillator circuit itself may be used if self-excitation is provided), the current i_1 through the capacitors C_k and C_s is given as

$$i_1 = \frac{e_1 \omega C_s C_k}{C_s + C_k} = \frac{e_1 \omega C_k}{\left(1 + \frac{C_k}{C_s}\right)} \tag{15.30}$$

Now the voltage, e_2, across the series capacitor, C_k, is

$$e_2 = \frac{i_1}{\omega C_k} \tag{15.31}$$

The ratio of e_1/e_2 may be expressed as given in (15.32) when (15.31) is substituted in (15.30)

$$\frac{e_1}{e_2} = \left(1 + \frac{C_k}{C_s}\right) \tag{15.32}$$

If (15.32) is substituted in (15.29), we find

$$PI = \left[\frac{e_p}{e_{pc}}\frac{e_{ic}}{e_i}\right] A \frac{R_A C_A}{C_k}\left[\frac{e_1}{e_2}\right]^2 \tag{15.33}$$

The quantity $\frac{e_1}{e_2}$ is readily determined by the attenuator, A, when the switch, S, (Fig. 15.4) is operated between "M" and "PI" for the above described conditions. The absolute calibration then depends upon A, R_A, C_A and C_k. All four of these quantities may be determined within a few per cent.

15.30 Oscillator Correlation

The equivalent crystal circuit has been discussed in so far as the measurement of PI is concerned; however, for correlation with an oscillator, the behavior of the crystal in that oscillator must be duplicated. Correlation of the PI meter with an oscillator is a function of both amplitude and frequency. It is obviously necessary from the derivation of (15.28) that the frequency of operation be duplicated, but the necessity for amplitude correlation can only be explained from the practical consideration that the equivalent circuit components of Fig. 15.1 are parameters that may be functions of amplitude. Crystals having nonlinear characteristics of the type that necessitate amplitude correlation may in part be attributed to either the method of mounting the crystal or couplings to other modes of vibration whose coupling coefficients are functions of amplitude.

In most oscillators the voltage across the terminals of a crystal is a function of many parameters such as plate voltage, vacuum tubes, etc. With an average set of conditions, however, reasonable correlation is obtained with the PI meter for a single adjustment of the generator voltage, e_i, for all crystals. The magnitude of e_i must, of course, be chosen to produce a voltage across the crystal equal to the average value obtained in the oscillator circuit for which the crystal is intended.

Frequency correlation with an external oscillator is a function of the effective capacitance, C_t, in shunt with the crystal. In order to duplicate the oscillator frequency with the PI meter, the capacitance, C_s, (Fig. 15.4) must be adjusted until the frequency of oscillation in the PI meter is the same as that in the oscillator for a crystal having average activity. In Fig. 15.4, the capacitance, C_s, is variable, and its dial is calibrated in terms of both the total effective capacitance across the crystal, C_t, and the resulting multiplying factor $\frac{R_A C_A C_k}{(C_t - C_x)^2}$. The magnitude of C_t may be measured by means of a capacitance bridge connected across the crystal socket terminals with the generator impedance shorted.

The determination of the dynamic or effective capacitance, C_t, across the crystal for an oscillator may similarly be obtained by adjusting the magnitude of C_s in the PI meter until the frequencies of oscillation in the PI

meter and in the oscillator under test are identical for the same amplitude of oscillation. By this means, the *PI* meter directly indicates the effective oscillator capacitance, C_t. The amplitude of oscillation must be duplicated in as much as C_t is not independent of the amplitude in most oscillators.

15.40 Description of Oscillator Generating "e_i"

The generator plays no part in the theory of *PI* measurement as it could be replaced by a signal generator or any other suitable source of radio frequency energy. It is convenient, however, to utilize the voltage appearing across C_k as an input to an amplifier whose output represents the generator. This in effect constitutes a feedback oscillator whose frequency is controlled by the crystal under test. Initial consideration of the over-all characteristics of the *PI* meter oscillator leads to the following requirements. The oscillator must,

1. Be capable of oscillating all crystals usable in other oscillator circuits.
2. Be capable of operating the crystal over a wide range of shunting capacitances in order to duplicate all the frequencies of oscillators now in the field.
3. Be capable of permitting high degrees of *AVC* control in order to maintain the generator voltage constant while the frequency is adjusted for reasonance.

If the generator voltage, e_i, is constant, resonance of the crystal circuit is essentially indicated by maximum crystal current, and oscillation is maintained at that resonant frequency. The adjustment to obtain maximum current is such that the phase shift throughout the oscillator loop is $2\pi n$ where $n = 0, 1, 2, 3$, etc. As previously described the phase shift and resulting frequency of oscillation are varied by a tuned circuit. The generator voltage, e_i, is held constant by an automatic amplitude control similar to the automatic volume control which is often applied to amplifiers. The manual control of the magnitude of the generator, e_i, is provided by an adjustment of the bias voltage of the automatic amplitude control circuit. In this way the maximum or start gain is independent of the setting of the amplitude control.

Automatic amplitude control (commonly referred to as automatic volume control, *AVC*) of an oscillator may be applied by the separation of the limiter from the linear amplifier. This means that in order to apply a high degree of *AVC* to the *PI* oscillator (Fig. 15.4), the input voltage of the limiter must be above the threshold of limiting by an amount exceeding the variation in the β path caused by the *AVC* control. This enables the limiter to absorb the changes in the gain of the linear amplifier such that $\mu\beta = 1$ at all times.

The time constant of the limiter is fast compared to that of the *AVC*

circuit, a condition which permits damping of transients set up by changes in gain occurring from AVC action. The input of the linear amplifier is held constant by the limiter. Gain changes in the linear amplifier produced by the variation of C_r (Fig. 15.4) are absorbed by AVC, while the variation of activity in the crystal is absorbed by the limiting amplifier.

15.50 Evaluating Performance Index

From (15.28) it can be seen that the attenuator, A, the effective variable capacitance, C_t, together with the vacuum tube voltmeters, e_i and e_p, offer a number of possible variations in the method of evaluating the constants used to determine the PI of quartz crystals. There are, however, two principal methods—the first provides direct reading, while the second is more accurate but requires an indirect evaluation.

The first method utilizes a means of calibration of the meter scales directly in terms of PI. The attenuator is adjusted such that its indicator reading times the multiplying factor associated with the dial attached to C_s is some multiple of 10. If in the calibrate position, A_c is set at unity, and e_{pc} and e_{ic} are adjusted by varying the capacitive load to some reference deflection, then the expression for PI becomes

$$PI = \frac{e_p}{e_i} K_1 \tag{15.34}$$

where

$$K_1 = \left[\frac{e_{ic}}{e_{pc}} A \frac{C_k R_A C_A}{(C_t - C_x)^2} \right]$$

The Performance Index then is indicated by the two readings of e_p and e_i. The absolute magnitudes of e_i and e_p need not be known since it is possible to use as a reference, the arbitrary calibrating deflections of e_{ic} and e_{pc}. The magnitude of e_p indicates the significant figures while e_i is a multiplying factor.

The second method of evaluating Performance Index eliminates any calibration errors in the two vacuum tube voltmeters, e_i and e_p. This method utilizes the attenuator to adjust $\left[\frac{e_p e_{ic}}{e_i e_{pc}} \right] = 1$. In the "calibrate" operation, e_{ic} is set to equal e_i and then e_{pc} is adjusted for full scale or a convenient deflection. In the "operate" position, the attenuator is varied until e_p equals e_{pc}. In this manner, the two readings of the attenuator are used to determine the ratio of $\frac{e_p}{e_{pc}}$ and the measurement is independent

of the voltmeter calibration. The factor $R_A C_A$ is a constant; therefore, the P.I. equation (15.28) simplifies to

$$P.I. = (\Delta A) K_2 \tag{15.35}$$

where

ΔA = change in attenuator insertion loss between the "operate" and "calibrate" conditions in terms of output voltage ratio $\frac{e_p}{e_{pc}}$, given as $\frac{A}{A_c}$.

$K_2 = \frac{C_k R_A C_A}{(C_t - C_x)^2}$ where C_t is the effective capacitance in series with the crystal, C_k is the fixed series capacitance and C_x is the crystal socket capacitance. (See Fig. 15.3)

15.60 P.I. Meter Applications

The application of this instrument can be extended to determine other properties of both crystal and oscillator. With the aid of a frequency measuring means and a capacitance bridge, the P.I. meter may be used to determine all the circuit constants designated in the electrical equivalent circuit of Fig. 15.1. If the loss in the holder is negligible then the equations are considerably simplified; however, in those instances where holder loss must be considered, the approximation that $X_t \ll R_L$ which may be allowed for most cases enables an evaluation of M and Q_c that is readily computed.

The dial controlling C_s, that is calibrated in terms of the total capacitance, C_t, makes possible the calculation of the magnitude of the input impedance to the crystal circuit, R, as well as Q_1.

$$\left.\begin{aligned} R &= \frac{X_t^2}{P.I.} \\ Q_1 &= \frac{P.I.}{X_t} \end{aligned}\right\}. \tag{15.36}$$

The magnitude of Q_1 may also be measured directly from equation (15.12) where Q_1 was given as

$$Q_1 = \frac{e_0}{e_i} \frac{C_k C_t}{(C_t - C_x)^2} \tag{15.12}$$

As may be seen from Fig. 15.4, e_0/e_i can be evaluated in terms of the attenuator calibration by enabling switch, S, to select the "PI" and "M" positions respectively and adjusting the attenuator such that the same output meter indication is obtained in the two cases.

The quantity Q_1 makes possible the calculation of Q_c where Q_c is defined as

$$Q_c = \frac{\omega L_c}{R_c} \cong \frac{X_t}{R_1\left(1 + \frac{C_0}{C_t}\right)^2} \tag{15.37}*$$

It can be shown that Q_c in terms of Q_1 is given by the following equation if $R_L \gg X_t$.

$$Q_c = \frac{Q_1}{1 - Q_1\frac{X_t}{R_L}} = \frac{Q_1}{1 - \frac{P.I.}{R_L}} \tag{15.38}$$

The Figure of Merit, M, defined by (15.3) at the series resonant frequency of the crystal, ω_1, becomes

$$M \cong \frac{1}{\omega_1 C_0 R_1} = \frac{X_{c_0}}{R_1} \tag{15.39}$$

The measurement of M can be determined from Q_c provided C_t and C_0 are known. M may be determined from the following expression

$$M = \frac{Q_1}{1 - \frac{P.I.}{R_L}} \frac{C_t}{C_0}\left(1 + \frac{C_0}{C_t}\right)^2 \tag{15.40}$$

If C_t is selected such that $C_t \gg C_0$, then for most cases R_L is large compared to $Q_1 X_t$. This means that $Q_1 \cong Q_c$ and $R \cong R_c$. With this approximation, (15.40) becomes

$$M = Q_1 \frac{C_t}{C_0}\left(1 + \frac{C_0}{C_t}\right)^2 \tag{15.41}$$

The relationship between R_c and R_1 as a function of frequency may be expressed directly from the input impedance expression of the equivalent circuit. This is given as

$$R_c = \frac{R_1}{\left[\frac{\omega_2^2 - \omega^2}{\omega_2^2 - \omega_1^2}\right]^2 + \frac{1}{M^2}} \tag{15.42}$$

where ω is the unity power factor frequency in the P.I. meter, neglecting R_L. If

$$\frac{1}{M} \ll \left[\frac{\omega_2^2 - \omega^2}{\omega_2^2 - \omega_1^2}\right]$$

---

* The relationship $R_c \cong R_1\left(1 + \frac{C_0}{C_t}\right)^2$ was derived in Fair's paper*[1].

which is a plausible assumption when $C_l \geq C_0$, and we assume that $\frac{\omega_1 + \omega_2}{2} = \omega_c$ then,

$$R_c = \frac{R_1}{\left[\frac{\omega_2 - \omega}{\omega_2 - \omega_1}\right]^2} \tag{15.43}$$

The relationship between R_1 and R_c does not involve R_L and may be expressed in terms of C_0 and C_l instead of frequency. Neglecting R_L to determine the relationship between capacitance and frequency in the P.I. meter as derived in Section 15.92, we find

$$\frac{\omega_2 - \omega}{\omega_2 - \omega_1} = \frac{1}{1 + \frac{C_0}{C_l}} \frac{\left[1 + \sqrt{1 + \frac{4}{P_1^2}}\right]}{2} \tag{15.44}$$

where,

$$P_1 = \frac{M}{\left(1 + \frac{C_0}{C_l}\right)} \tag{15.45}$$

If $P_1 \gg 2$ then the expression between R_c and R_1 may be written

$$R_c = R_1 \left(1 + \frac{C_0}{C_l}\right)^2 \tag{15.46}$$

The restriction that $C_l \geq C_0$ may be removed if the error between (15.42) and (15.46) is taken into account. This error may be expressed as

$$\text{Per Cent Error in } R_c = 100 \left[\frac{1}{P_1^2}\right]\left[\frac{2}{1 + \sqrt{1 + 4/P_1^2}}\right] \tag{15.47}$$

The resonant resistance, R_1, together with (15.46) provides a means of checking the P.I. meter. The magnitude of R_1 may be determined by the substitution method and from this, the value of R_c calculated. Fig. 15.5 represents the agreement between the P.I. meter and those expected from resonant frequency measurements.

The remaining crystal constants L_1 and C_1 (Fig. 15.1) may be evaluated from the measurement of ω_1, ω_2 and C_0. The resonant frequency, ω_1, is defined as

$$\omega_1^2 = \frac{1}{L_1 C_1} \tag{15.48}$$

The anti-resonant frequency, ω_2, is defined as

$$\omega_2^2 = \frac{1}{L_1}\left(\frac{1}{C_1} + \frac{1}{C_0}\right) \tag{15.49}$$

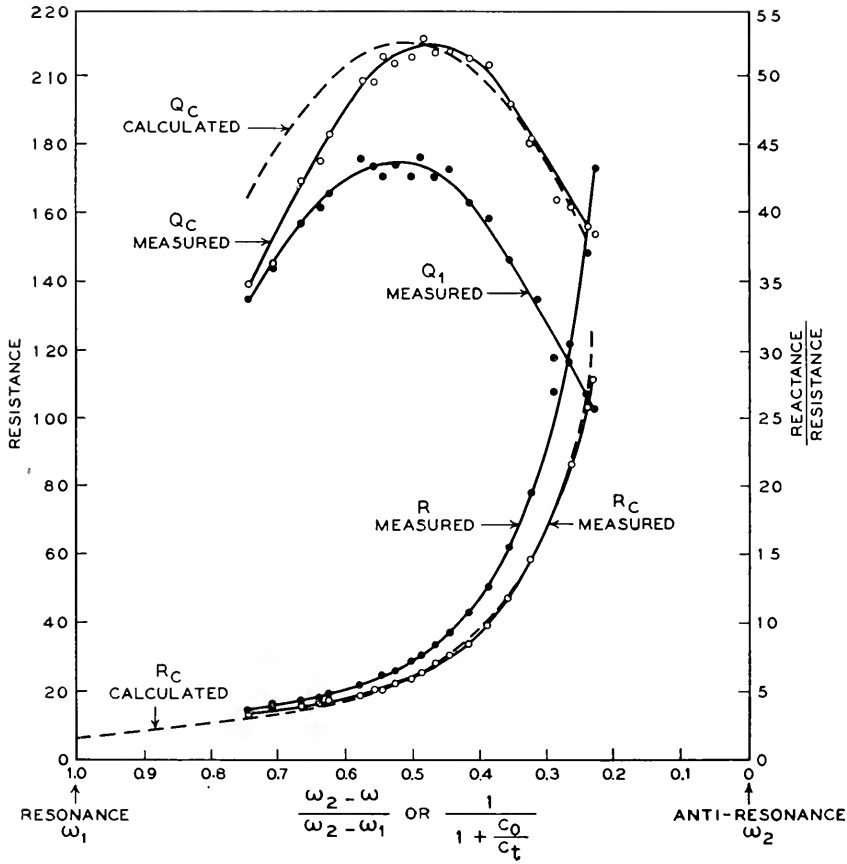


Fig. 15.5—Typical characteristic of a quartz crystal measured by the Performance Index meter.

Solving these equations simultaneously, it is found that

$$\left.\begin{aligned} L_1 &= \frac{1}{C_0(\omega_2^2 - \omega_1^2)} \backsimeq \frac{1}{2\omega_c C_0(\omega_2 - \omega_1)} \\ C_1 &= \frac{C_0(\omega_2^2 - \omega_1^2)}{\omega_1^2} \backsimeq \frac{2C_0(\omega_2 - \omega_1)}{\omega_1} \end{aligned}\right\} \tag{15.50}$$

15.70 Experimental Data

The performance of the P.I. meter may best be illustrated by experimental data. The following data indicate the correlation which may be obtained between the P.I. meter and various types of oscillator circuits. Experimental considerations are extended to

(1) Frequency and amplitude correlation with a "Pierce" and "Tuned-Plate" oscillator

(2) The measurement of the effective capacitance, C_t, of an oscillator as a function of tuning, and

(3) The variation of P.I. as a function of voltage across the crystal.

The results presented are not to be considered as generalized data, but are intended only to show a set of measurements obtained for a specific set of operating conditions for each type of circuit.

It has been pointed out that the frequency of oscillation is a function of R_1, ω_1, ω_2, C_0 and C_t. Since R_1, ω_1, ω_2 and C_0 are explicit parameters of the crystal, the capacitor, C_t, becomes the only frequency determining element in the P.I. meter. From analytical methods to be described in

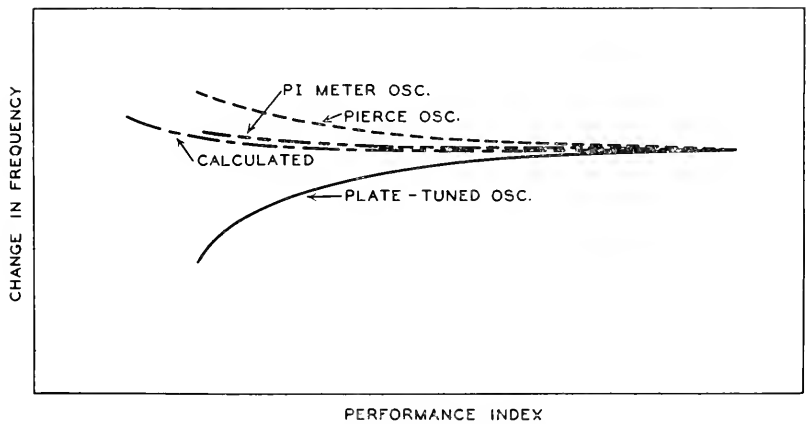


Fig. 15.6—Frequency variations in oscillators as a function of Performance Index.

Sections 15.80 and 15.92 the frequency difference between the P.I. meter and the generalized oscillator (neglecting R_L) is given as

$$\Delta\omega \leqq \frac{(\omega_2 - \omega_1)}{M^2}\left(1 + \frac{C_0}{C_t}\right)\frac{2}{\left[1 + \sqrt{1 + \frac{4}{P_1^2}}\right]} - \frac{(\omega_2 - \omega_1)}{M^2}\left(1 + \frac{C_t}{C_0}\right) \tag{15.51}$$

Since the magnitude of frequency change in the above equation is small compared to the variations caused by changes in operating conditions, the P.I. meter may be used as a frequency correlation medium. Fig. 15.6 is an example of the correlation between the "Pierce" and "Tuned-Plate" oscillator and the P.I. meter. The P.I. meter falls between these two oscillators in frequency for any crystal activity. It must be recognized that Fig. 15.6 is not conclusive to the extent of generalization; however, it is indicative of possible correlation with these two popular oscillator circuits.

Figures 15.7 (A) and (B) show the amplitude correlation between the Performance Index meter and the grid current of the "Tuned-Plate" and "Pierce" oscillators, respectively. The P.I. vs. grid current characteristic was arbitrarily taken at three frequencies—4.5 mcs, 5.66 mcs and 7.81 mcs. The change in grid current of the oscillator shown in Fig. 15.7 (A) with frequency is caused by the varying L-C ratio in the plate circuit. (See curves A, B and C for constant crystal activity.) The curves A and D represent the effect of changing bands by switching coils, varying the L-C ratio 2 to 1 in the plate circuit for the same crystal frequency. Fig. 15.7

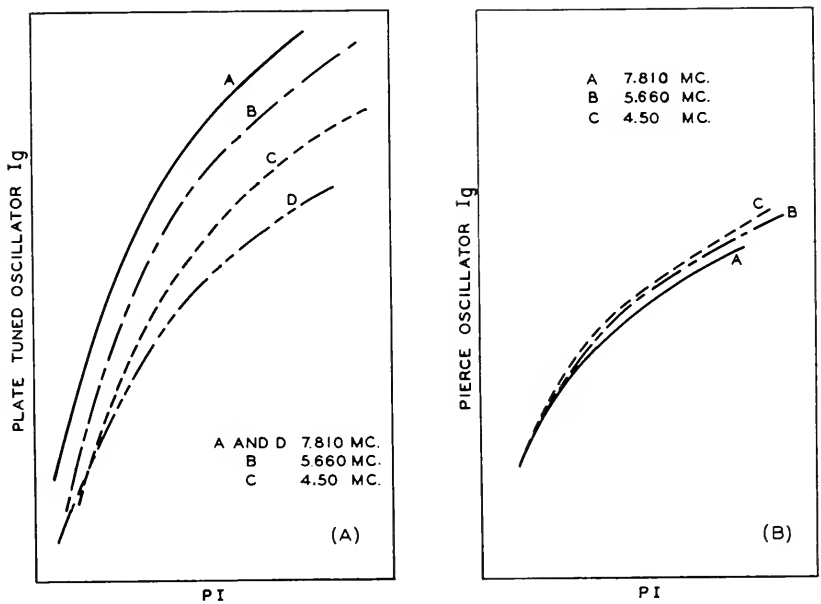


Fig. 15.7—Typical oscillator characteristics.

(B) represents the correlation between P.I. and grid current of the "Pierce" type oscillator. The results of this correlation indicate that the grid current is essentially independent of the operating frequency.

The measurement of P.I. is independent of the level of crystal vibration, provided that the electrical equivalent circuit parameters of Fig. 15.1 become constants; however, in actual practice these are not constants, particularly R_1. Variations of this type, as previously discussed, make it necessary to duplicate the amplitude of oscillation of the P.I. meter with the oscillator. Fig. 15.8 represents the variation of P.I. as a function of voltage across the crystal terminals for five crystals arbitrarily selected. It is readily observed that P.I. may be a random function of amplitude.

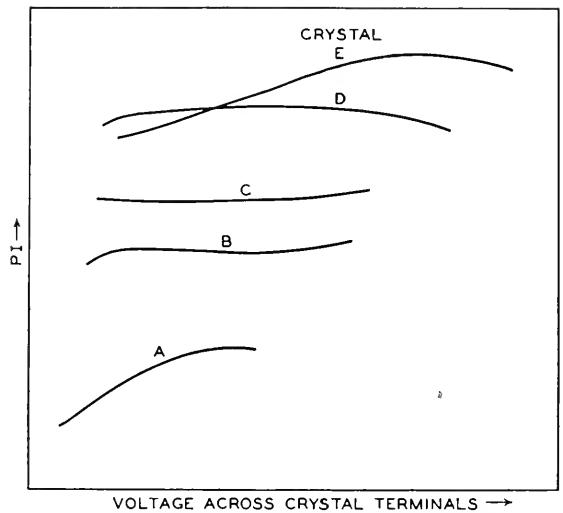


Fig. 15.8—Observed variation of Performance Index as a function of voltage across the crystal terminals.

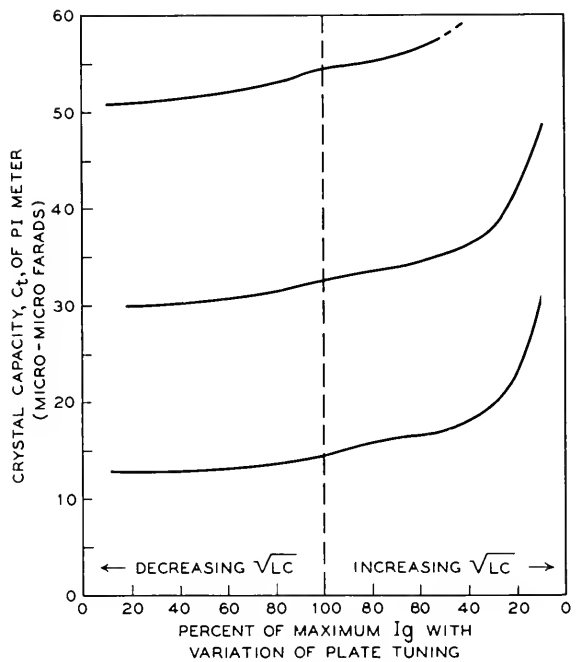


Fig. 15.9—Observed variation of effective crystal capacitance in a Miller oscillator.

Equation (15.44) indicates that the capacitance, C_t, determines the operating frequency between ω_1 and ω_2 of any given crystal. The capacitance, however, may include reflected reactances from associated circuits or possibly from circuits unintentionally coupled to the oscillator. Normally, crystals are adjusted to frequency for a specified value of C_t. This makes it of interest to measure the magnitude of C_t over the range of the manual oscillator tuning adjustments, as well as over the frequency range of the oscillator. Fig. 15.9 shows the circuit capacitance, C_t, plotted as a function of tuning of the plate circuit of a Tuned-Plate oscillator. Tuning of the plate circuit is expressed in terms of percentage of maximum grid current.

15.80 Circuit Analysis Involving the Accuracy of P.I. Measurements

The method of P.I. measurement just described involved a number of unverified approximations. These approximations under the majority of conditions will be proven to be justified and the resulting expressions for percentage error will be obtained.

It is necessary to apply a method of analysis that is most readily adaptable to the crystal circuit. The analysis described in this section involves the use of Conformal representation as a means of determining (1) the behavior of the equivalent crystal circuit, (2) the error resulting in P.I. from assuming operation at the resonant frequency rather than the frequency for minimum impedance, and (3) the comparison of frequency of oscillation in the P.I. meter with other oscillator circuits.

Generally, the variations of reactance, X, and resistance, R, of the equivalent circuit (Fig. 15.1) are plotted as a function of frequency, and the analysis of the impedance, $R + jX$, between the resonant frequency, ω_1, and anti-resonant frequency, ω_2, are handled in precisely the same way as any linear passive element. This was essentially the procedure used to derive the equations in section 15.60. The analysis required to evaluate the errors leads to rather an elaborate study; however, in Section 15.81, it will be shown that it is very helpful analytically if the impedance of the crystal is plotted in the form of a circle diagram, that is, with the ordinate representing reactance, X, and the abscissa representing resistance, R.

15.81 *Conformal Representation*

Conformal Representation or Mapping is a convenient tool which for this application enables the physical operating condition to be expressed quantitatively from its graphical counterpart. Physical interpretation also makes it possible to draw many other conclusions that by other methods prove clumsy and laborious.

The basis for this analysis depends upon the ability to utilize the following equation to represent any impedance whose frequency of operation is controlled by a crystal. This equation is known as a linear fractional transformation

$$W = \frac{\alpha + \beta Z}{\gamma + \delta Z} \tag{15.52}$$

The terms α, β, γ, and δ represent complex constants and Z represents a complex variable later to be chosen to represent a linear function of frequency. Since W and Z represent the dependent and independent variable, they may also be considered as representing two separate planes. The abscissa and ordinate of these two planes represent their real and imaginary components respectively. The planes are linked by (15.52), that is, this equation will transform a specific point from one plane to the other.

The constants α, β, γ and δ for the equivalent crystal circuit are determined by writing the expression for Z_c in the form of (15.52). For example (neglecting R_L), Z_c from Fig. 15.1 may be written as

$$Z_c = \frac{1}{j\omega C_0} \frac{\left[R_1 + j\left(\omega L_1 - \frac{1}{\omega C_1}\right)\right]}{\left[R_1 + j\left(\omega L_1 - \frac{1}{\omega}\left(\frac{1}{C_1} + \frac{1}{C_0}\right)\right)\right]} \tag{15.53}$$

By substituting (15.48) and (15.49) in (15.53), this impedance may be written as

$$Z_c = \frac{1}{j\omega C_0} \frac{[\omega R_1 + jL_1(\omega - \omega_1)(\omega + \omega_1)]}{[\omega R_1 + jL_1(\omega - \omega_2)(\omega + \omega_2)]} \tag{15.54}$$

Since the operating frequency, ω, represents some frequency between ω_1 and ω_2, and $\omega \gg \omega_2 - \omega_1$, we can make the following approximations in this operating range. The symbol ω_e is defined as the average operating radian frequency.

$$\left.\begin{aligned} \omega_e &= \frac{\omega_1 + \omega_2}{2} \\ \omega_e &\cong \omega \end{aligned}\right\} \tag{15.55}$$

If (15.55) is substituted in (15.54), factor ω_e out, add and subtract ω_2 from the imaginary component in both numerator and denominator, we may write Z_c as

$$Z_c = \frac{\left[\frac{1}{\omega_e C_0} + j\,\frac{1}{\omega_e C_0}\,\frac{2L_1}{R_1}\,(\omega_2 - \omega_1)\right] + \left[\frac{1}{\omega_e C_0}\right] j\,\frac{2L_1}{R_1}\,(\omega - \omega_2)}{[j] + [j]\,j\,\frac{2L_1}{R_1}\,(\omega - \omega_2)} \tag{15.56}$$

The constants α, β, γ and δ may be written immedately from (15.56); however, for purposes of simplification let

$$\left.\begin{aligned} \tau &= \frac{1}{\omega_e C_0} \\ Z &= j\frac{2L_1}{R_1}(\omega - \omega_2) \\ M &= \frac{2L_1}{R_1}(\omega_2 - \omega_1) \quad \text{(See equation 15.50)} \end{aligned}\right\} \tag{15.57}$$

then,

$$Z_c = \frac{[\tau + jM\tau] + [\tau]Z}{j + jZ} \tag{15.58}$$

Now if Z_c may be represented by W in (15.52) the remaining constants must be

$$\left.\begin{aligned} \alpha &= \tau + j\tau M \qquad \gamma = j \\ \beta &= \tau \qquad\qquad \delta = j \end{aligned}\right\} \tag{15.59}$$

Now Z represents the frequency variable and graphically represents a line coincident with the Y-axis in the Z-plane, and W, the corresponding impedance variable, represents a circle in the W-plane. The coordinates of the center of the circle, W_0, in the W-plane is given by

$$W_0 = \frac{\dfrac{\beta}{\delta}\left[\dfrac{\alpha}{\beta} - \dfrac{\bar{\gamma}}{\bar{\delta}}\right]}{\left[\dfrac{\gamma}{\delta} + \dfrac{\bar{\gamma}}{\bar{\delta}}\right]} \tag{15.60}*$$

when $\bar{\gamma}$ and $\bar{\delta}$ represent the conjugate functions of γ and δ respectively. The radius of the circle is given by

$$\sigma = \left|\frac{\dfrac{\beta}{\delta}\left[\dfrac{\alpha}{\beta} - \dfrac{\gamma}{\delta}\right]}{\left[\dfrac{\gamma}{\delta} + \dfrac{\bar{\gamma}}{\bar{\delta}}\right]}\right| \tag{15.61}*$$

* E. C. Titchmarch, "Theory of Functions," Oxford 1932, pp. 191–192. Note: These equations are not derived in Titchmarch; however, by taking the limit as the diameter of the circle in the Z-plane approaches infinity, (15.60) and (15.61) result.

Now substituting the values given in (15.59) in (15.60) and (15.61) to determine the coordinates of the center of the circle and the radius respectively, we find

$$\left.\begin{aligned} W_0 &= \frac{\tau M}{2} - j\tau \\ \sigma &= \frac{M\tau}{2} \end{aligned}\right\} \tag{15.62}$$

As might be expected, the radius of the circle is equal to the real component of the coordinates of the center of the circle. This indicates that Fig. 15.10

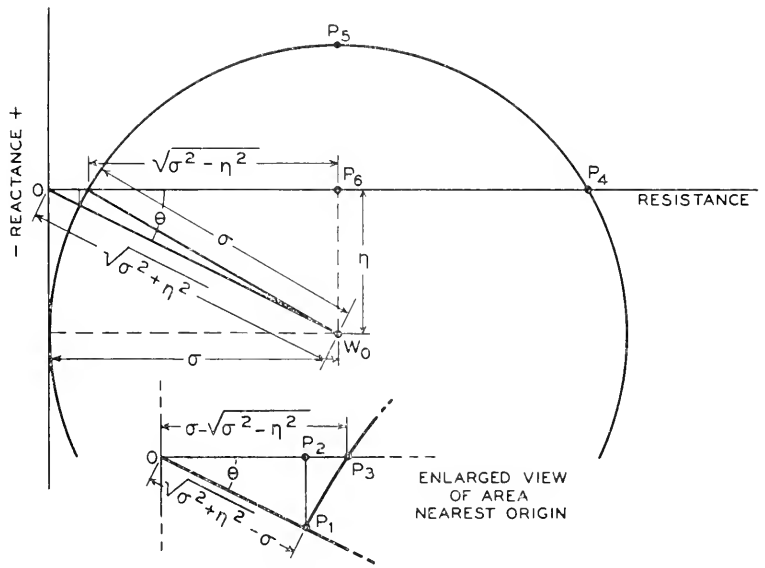


Fig. 15.10—Circle diagram for crystal circuit combinations.

graphically represents the circle diagram where $\eta \cong \tau$. The impedance, Z_c, is represented for a given frequency, as a vector from the origin to the corresponding point on the perimeter of the circle. From Fig. 15.10 the following crystal properties may be deduced.

1. The anti-resonant impedance designated as $O - P_4$ is simply $\sigma + \sqrt{\sigma^2 - \eta^2}$, and when evaluated equals $R_1(M^2 - 1) \cong R_1 M^2$.
2. The resonant resistance, $O - P_3$ given by $\sigma - \sqrt{\sigma^2 - \eta^2}$ is equal to R_1.
3. The maximum positive reactance between ω_1 and ω_2 is represented by the distance from P_5 to P_6. It is given as $\sigma - \eta$ which equals $\left(\frac{R_1 M^2}{2} - X_{c_0}\right)$.

4. The condition which must exist when the crystal reactance is zero between ω_1 and ω_2 occurs when $\sigma = \eta$ or $M = 2$.

It follows that the error of measuring, say the series resistance R_1, by varying the frequency for maximum transmission and assuming true resonance when the crystal is between two low non-inductive resistors is associated with the difference between the length of the two vectors $O - P_3$ and $O - P_1$. The per cent error caused by the crystal capacitor, C_0, by this method of measurement of R_1 is given as

$$\begin{aligned} \text{Per cent error of } R_1 &= 100\left[1 - \frac{O - P_1}{O - P_3}\right] \\ &= 100\left[1 - \frac{\sqrt{\sigma^2 + \eta^2} - \sigma}{\sigma - \sqrt{\sigma^2 - \eta^2}}\right] \end{aligned} \tag{15.63}$$

If σ and η are substituted in (15.63), we find

$$\text{Per cent error of } R_1 = 100\left[1 - \frac{1 + \sqrt{1 - \frac{4}{M^2}}}{1 + \sqrt{1 + \frac{4}{M^2}}}\right] \tag{15.64}$$

This difference in amplitude was caused by the difference in frequency between resonance and minimum impedance. This frequency difference may be determined by transforming the points P_1 and P_3 into the Z plane by (15.52) and subtracting them arithmetically.

In order to express the coordinates of any point in the W-plane by its real and imaginary components let

$$W = R_0 + jX_0 \tag{15.65}$$

Now the coordinates of P_3 may be expressed as,

$$R_0 = \sigma\left[1 - \sqrt{1 - \left(\frac{\eta}{\sigma}\right)^2}\right]; \qquad X_0 = 0 \tag{15.66}$$

The coordinates of P_1 are similarly given by

$$\left.\begin{aligned} R_0 &= \sigma\left[1 - \frac{1}{\sqrt{1 + \left(\frac{\eta}{\sigma}\right)^2}}\right] \\ X_0 &= -\eta\left[1 - \frac{1}{\sqrt{1 + \left(\frac{\eta}{\sigma}\right)^2}}\right] \end{aligned}\right\} \tag{15.67}$$

The coordinates of these two points represent values of W; now (15.52) may be solved for Z.

$$Z = \frac{\alpha - \gamma W}{\delta W - \beta} \tag{15.68}$$

Substituting in values for α, β, γ and δ given by (15.59), we find

$$Z = j\frac{2L_1}{R_1}(\omega - \omega_2) = j\frac{M(\omega - \omega_2)}{(\omega_2 - \omega_1)} = -j2\frac{\sigma(\eta + X_0)}{(\eta + X_0)^2 + R_0^2} \tag{15.69}$$

The real component of Z must be zero since the function of Z is coincident with the Y-axis. By substituting values of R_0 and X_0 from (15.66) and (15.67), we find

$$\omega - \omega_2 = -\frac{(\omega_2 - \omega_1)}{M}\frac{\eta/\sigma}{[1 - \sqrt{1 - (\eta/\sigma)^2}]} \tag{15.70}$$

also

$$\omega - \omega_2 = -\frac{(\omega_2 - \omega_1)}{M}\frac{\eta/\sigma}{\left[\sqrt{1 + \left(\frac{\eta}{\sigma}\right)^2} - 1\right]} \tag{15.71}$$

Subtracting (15.71) from (15.70) to get $\Delta\omega$ we have

$$\Delta\omega = \frac{\eta}{\sigma}\frac{(\omega_2 - \omega_1)}{M}A \tag{15.72}$$

where

$$A = \frac{\sqrt{1 - (\eta/\sigma)^2} - \sqrt{1 + (\eta/\sigma)^2}}{(\eta/\sigma)^2} \tag{15.73}$$

Now the $\lim_{\eta/\sigma \to 0} A = 1$. When $\frac{\eta}{\sigma} = \frac{2}{M}$ then $\Delta\omega = \frac{\omega_2 - \omega_1}{M}A$

15.82 *Circuit Analysis Involving Crystals*

The same procedure could be followed for the impedance, Z_i, in Fig. 15.3; however, the impedance expression conforming to (15.52) may be written directly if a more general expression is derived for impedances added in parallel or series.

Paralleling the impedance, W, with an impedance, T, (Fig. 15.11) modifies the constants in (15.52) but not its form providing T is essentially constant between ω_1 and ω_2. For parallel impedances the impedance equation becomes,

$$\frac{WT}{W + T} = \frac{\alpha + \beta Z}{\left(\gamma + \frac{\alpha}{T}\right) + \left(\delta + \frac{\beta}{T}\right)Z} = \frac{\alpha' + \beta' Z}{\gamma' + \delta' Z} \tag{15.74}$$

The numerator remains the same; however, the denominator has an additional term added to both γ and δ.

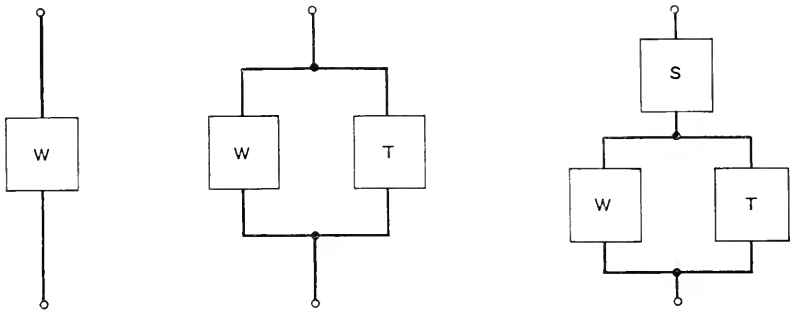


Fig. 15.11—Various crystal circuit combinations.

In the same way, a series element, S, may be added as shown in Fig. 15.11. The impedance of this combination is similarly modified and the input impedance may be expressed as

$$S + \frac{WT}{W+T} = \frac{\left[\alpha + S\left(\gamma + \frac{\alpha}{T}\right)\right] + \left[\beta + S\left(\delta + \frac{\beta}{T}\right)\right] Z}{\left(\gamma + \frac{\alpha}{T}\right) + \left(\delta + \frac{\beta}{T}\right) Z} = \frac{\alpha'' + \beta'' Z}{\gamma'' + \delta'' Z} \tag{15.75}$$

The addition of a series element leaves the denominator unchanged but adds a term to the numerator. The form of (15.75) is exactly the same as (15.52) except that the magnitudes of the constants, α, β, γ and δ have been modified.

As an example, apply (15.74) to obtain an expression for the impedance Z_{AB} (Fig. 15.2). This expression for impedance may be written directly if the following values are substituted in (15.74).

$$\left.\begin{aligned} \alpha &= \tau + jM\tau \\ \beta &= \tau \\ \gamma &= \delta = j \\ T &= \frac{1}{j\omega C_t} \end{aligned}\right\} \tag{15.76}$$

$$Z_{AB} = \frac{WT}{W+T} = \frac{[\tau + jM\tau] + [\tau]Z}{\left[-M\frac{C_t}{C_0} + j\left(1 + \frac{C_t}{C_0}\right)\right] + \left[j\left(1 + \frac{C_t}{C_0}\right)\right] Z} \tag{15.77}$$

From this expression we can again see the same form as (15.52) except that the coefficients of α, β, γ and δ are modified by the capacitance, C_t, in parallel with the equivalent crystal circuit. It has been previously explained that the frequency of oscillation in the generalized oscillator is determined by the anti-resonant frequency of the impedance, Z_{AB}, and that this impedance represents the exact definition of P.I. P.I. therefore is represented in magnitude by $O - P_4$ in Fig. 15.10.

Equation (15.75) enables us to write the expression for the input impedance, Z_i, for the P.I. meter directly and furthermore, it enables us to compute the error produced by the adjustment of C_r to minimum impedance rather than unity power factor as assumed in the derivation of (15.28). This error obviously will be a function of M, C_0, C_t and R_L. Writing the impedance expression for Z_i directly, requires a little modification in that the crystal is shunted by two elements, C_x (the crystal socket capacitance) and R_L (the holder loss) as well as having the combination in series with the capacitor, C_3, where $C_3 = \dfrac{C_s C_k}{C_s + C_k}$. Shunting the crystal with C_x modifies (15.77) only in that $C_t = C_x$; from this we can use (15.75) directly in a form readily adaptable to the determination of our original coefficients, α'', β'', γ'' and δ''. Adding a series capacitor, C_3, and a shunt resistor, R_L, modifies (15.77) in a manner specified by (15.75). Here,

$$\left.\begin{aligned} \alpha &= \tau + jM\tau & \delta &= j\left(1 + \frac{C_x}{C_0}\right) \\ \beta &= \tau & T &= R_L \\ \gamma &= -M\frac{C_x}{C_0} + j\left(1 + \frac{C_x}{C_0}\right) & S &= \frac{1}{j\omega C_3} \end{aligned}\right\} \qquad (15.78)$$

Substitute these constants given by (15.78) in (15.75) then α'', β'', γ'' and δ'' become,

$$\left.\begin{aligned} \alpha'' &= \alpha + S\left(\gamma + \frac{\alpha}{T}\right) = \left[\tau + X_{c_3}\left(1 + \frac{C_x}{C_0}\right) + \frac{M\tau X_{c_3}}{R_L}\right] \\ &\qquad + j\left[M\tau + \frac{MC_x X_{c_3}}{C_0} - \frac{\tau}{R_L}X_{c_3}\right] \\ \beta'' &= \beta + S\left(\delta + \frac{\beta}{T}\right) = \left[\tau + X_{c_3}\left(1 + \frac{C_x}{C_0}\right)\right] - j\left[\frac{X_{c_3}\tau}{R_L}\right] \\ \gamma'' &= \left(\gamma + \frac{\alpha}{T}\right) = \left[-M\frac{C_x}{C_0} + \frac{\tau}{R_L}\right] + j\left[1 + \frac{C_x}{C_0} + \frac{M\tau}{R_L}\right] \\ \delta'' &= \left(\delta + \frac{\beta}{T}\right) = \left[\frac{\tau}{R_L} + j\left(1 + \frac{C_x}{C_0}\right)\right] \end{aligned}\right\} \qquad (15.79)$$

Substitute these values in (15.60) and (15.61) to obtain the coordinates of the center of the circle, W_0, and the radius, σ.

$$W_0 = \tau \frac{\left[\left(M\frac{C_0}{C_0+C_x}+2\frac{\tau}{R_L}\right)-2j\left(1+\frac{C_0+C_x}{C_3}+\frac{\tau X_{c_3}}{R_L{}^2}+\frac{M\tau C_0}{R_L C_3}\right)\right]}{2\left[1+\left(\frac{\tau}{R_L}\right)^2+\frac{MC_0}{C_0+C_x}\frac{\tau}{R_L}\right]} \tag{15.80}$$

$$\sigma = \tau \frac{M\frac{C_0}{C_0+C_x}}{\left[1+\left(\frac{\tau}{R_L}\right)^2+\frac{MC_0}{C_0+C_x}\frac{\tau}{R_L}\right]} \tag{15.81}$$

From the above expressions, σ is not exactly equal to the real component of W_0. If however, $2\frac{\tau}{R_L} \ll \frac{MC_0}{C_0+C_x}$, then the above equations reduce to

$$W_0 = \tau \frac{\left[M\frac{C_0}{C_0+C_x}-2j\left(1+\frac{C_0+C_x}{C_3}+\frac{\tau MC_0}{R_L C_3}\right)\right]}{\left[2\left(1+\frac{MC_0}{C_0+C_x}\frac{\tau}{R_L}\right)\right]} \tag{15.82}$$

and

$$\sigma = \frac{\tau\left[\frac{MC_0}{C_0+C_x}\right]}{2\left[1+\frac{MC_0}{C_0+C_x}\frac{\tau}{R_L}\right]} \tag{15.83}$$

Now the radius, σ, equals the real component of W_0. The per cent error of P.I. resulting from tuning to minimum impedance rather than unity power-factor in terms of Fig. 15.10 is given as,

$$\begin{aligned}\text{Per Cent Error of P.I.} &= 100\left[1-\frac{O-P_3}{O-P_1}\right] \\ &= 100\left[1-\frac{\left[1-\sqrt{1-\left(\frac{\eta}{\sigma}\right)^2}\right]}{\left[\sqrt{1+\left(\frac{\eta}{\sigma}\right)^2}-1\right]}\right]\end{aligned} \tag{15.84}$$

Now since η represents the imaginary component of (15.82) then

$$\frac{2\sigma}{\eta} = P_2 = \frac{M}{\left(1+\frac{C_x}{C_0}\right)\left(1+\frac{C_0+C_x}{C_3}+M\frac{X_{c_t}}{R_L}\right)} \tag{15.85}$$

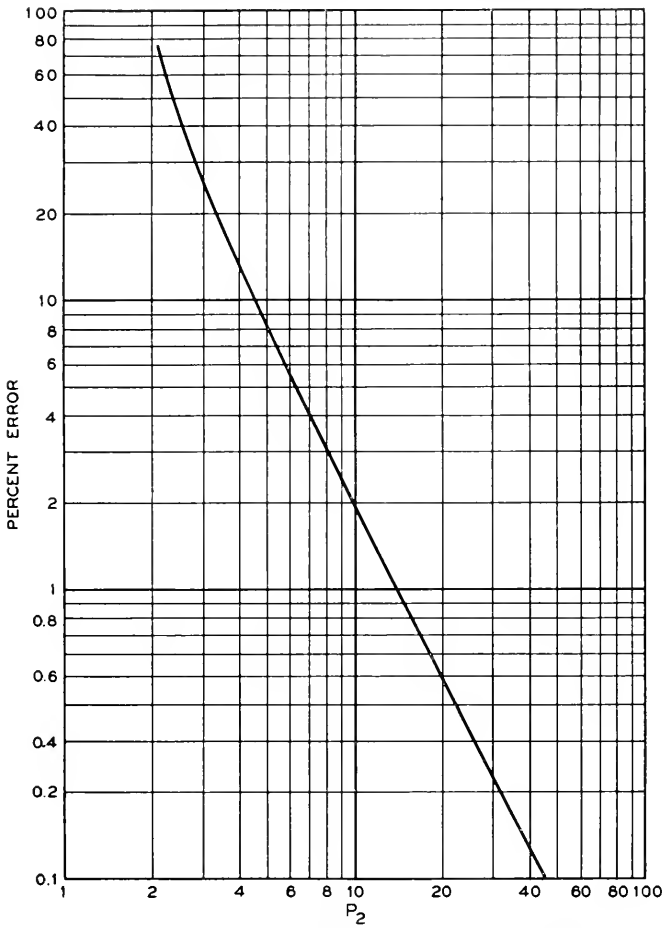


Fig. 15.12—Inherent error of the P.I. meter due to tuning for an indication of minimum impedance rather than unity power factor.

Rewriting (15.84) we have

$$\text{Per cent Error of P.I.} = 100\left[1 - \frac{\left[1 - \sqrt{1 - \frac{4}{P_2^2}}\right]}{\left[\sqrt{1 + \frac{4}{P_2^2}} - 1\right]}\right] \tag{15.86}$$

If $C_0 \gg C_x$ then

$$P_2 = \frac{M}{\left(1 + \frac{C_0}{C_t} + M\frac{X_{c_t}}{R_L}\right)} \tag{15.87}$$

The error given by (15.86) is plotted in Fig. 15.12 as a function of P_2.

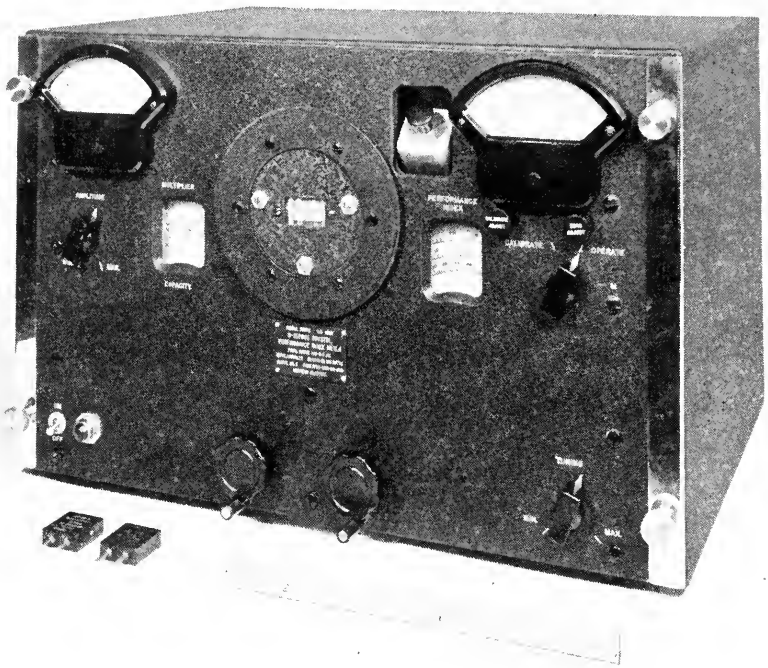


Fig. 15.13—Exterior view of crystal Performance Index meter.

15.83 *Frequency Errors*

As suggested in Section 15.81, conformal representation simplifies the mathematics required for the determination of frequency errors. In Section 15.81, the difference between the resonant frequency and the minimum impedance frequency was computed for the equivalent crystal circuit. This same procedure could be used to compute the frequency difference between the antiresonant frequency of the generalized oscillator circuit (Fig. 15.2) and the minimum impedance frequency of the P.I. meter for the same value of C_t. Comparison of the frequency of these two oscillators is plotted in Fig. 15.6 together with the measured values obtained from a "Pierce" and a "Tuned Plate" oscillator. This frequency comparison involves setting up two circle diagrams, one for Z_{AB} (Fig. 15.2) and one for the impedance, Z_i (Fig. 15.3) similar to Fig. 15.10. The impedance equations for both Z_{AB} and Z_i would be arranged such that they have the same function of Z in

order to have a common Z plane. In this way, transformation of operating points such as the "anti-resonant frequency" operating point (P_4) for the Z_{AB} impedance circle, could be subtracted from the "minimum impedance frequency" operating point (P_1) of the P.I. meter impedance circle in the common Z-plane. As in section 15.81, the frequency difference represents the arithmetic difference between P_4 and P_1 in the Z-plane in terms of $\frac{2L_1}{R_1}(\omega - \omega_2)$. As an example, look at the calculated curve in Fig. 15.6. This curve was computed for the case when R_L is negligible. The derivation of (15.51) given in section 15.92 precisely follows the procedure just described.

It is of interest to note that (15.27) may also be derived by Conformal means. It is more laborious than the usual circuit equations of section 15.2; however, it does provide a check of the methods used.

15.84 *Errors of Other Approximations*

Further consideration of P.I. meter errors leads to the assumptions made in the derivation of (15.27). The derivation of (15.27) assumed that the resistor, R_A, was non-reactive. While actually it can be made essentially noninductive, we have neglected the effect of the input capacitance of the attenuator that is shunted across its terminals. The error from neglecting this capacitance in (15.28) is given by the following expression

$$\text{Per Cent Error} = 100\left[1 - \frac{\sqrt{1 + Q_A^2}}{\omega R_A(C_u Q_A^2 + C_A)}\right] \qquad (15.88)$$

Where Q_A equals the reactance of the shunt capacitance of the attenuator, C_u, divided by the magnitude of the calibration resistor, R_A.

It is interesting to note in the derivation of (15.28) that R_A was assumed to be very much less than X_{C_A} which introduces an error of,

$$\text{Per Cent Error} = 100\left[1 - \sqrt{1 + \left[\frac{R_A}{X_{C_A}}\right]^2}\right] \qquad (15.89)$$

15.90 *Derivation of Circuit Equations*

15.91 *Derivation of Equation* (15.42)

Other equations used in this paper may best be developed from Fig. 15.3. By analysis of the input impedance, Z_i, the basis for the development of (15.42) is as follows:

From Fig. 15.2

$$Z_i = \frac{1}{j\omega C_t} + \frac{1}{j\omega C_0}\frac{\left[R_1 + j\left(\omega L_1 - \frac{1}{\omega C_1}\right)\right]}{\left[R_1 + j\left[\omega L_1 - \frac{1}{\omega}\left(\frac{1}{C_0} + \frac{1}{C_1}\right)\right]\right]} \tag{15.90}$$

By substituting (15.48) and (15.49) in (15.90), we find

$$Z_i = \frac{1}{j\omega}\left[\frac{1}{C_t} + \frac{1}{C_0}\left[\frac{\omega R_1 + jL_1(\omega^2 - \omega_1^2)}{\omega R_1 + jL_1(\omega^2 - \omega_2^2)}\right]\right] \tag{15.91}$$

This may be expressed in the form,

$$Z_i = \frac{\frac{\omega R_1}{H} + jL_1\left[\frac{(\omega^2 - \omega_2^2)}{C_t} + \frac{(\omega^2 - \omega_1^2)}{C_0}\right]}{-\omega L_1(\omega^2 - \omega_2^2) + j\omega^2 R_1} \tag{15.92}$$

where

$$H = \frac{C_0 C_t}{C_0 + C_t}$$

Now adding and subtracting ω_2^2 to the $\frac{\omega_2^2 - \omega_1^2}{C_0}$ term we have

$$Z_1 = \frac{\frac{\omega R_1}{H} + jL_1\left[\frac{(\omega^2 - \omega_2^2)}{C_t} + \frac{(\omega^2 - \omega_2^2)}{C_0} + \frac{(\omega_2^2 - \omega_1^2)}{C_0}\right]}{-\omega L_1(\omega^2 - \omega_2^2) + j\omega^2 R_1} \tag{15.93}$$

Rationalizing (15.93) and equating it to R_e and substituting in $L_1 = \frac{1}{C_0(\omega_2^2 - \omega_1^2)}$ (obtained from (15.50)), we find

$$R_e = \frac{R_1}{\left[\frac{\omega^2 - \omega_2^2}{\omega_2^2 - \omega_1^2}\right]^2 + \left[\frac{\omega}{\omega_1}\frac{1}{M}\right]^2} \tag{15.94}$$

15.92 *Derivation of Equations* (15.51) *and* (15.44)

Equation (15.51) makes possible the theoretical computation of the frequency difference between the generalized oscillator and the minimum impedance frequency adjustment of the *PI* meter. The derivation assumes that R_L is negligible and that the total capacitance across the crystal terminals is lumped in series with the crystal.

The impedance, Z_{AB}, in the generalized oscillator, Fig. 15.2, was given by (15.77). This equation may be expressed as follows:

$$Z_{AB}\omega_e(C_0 + C_t) = \frac{[1 + jM] + Z}{\left[-\frac{MC_t}{C_0 + C_t} + j\right] + jZ} \tag{15.95}$$

From this expression, as previously explained, (see Section 15.81), the following values for σ and η may be determined.

$$\left.\begin{aligned} \sigma &= \frac{M}{2}\frac{C_0}{C_0 + C_t} \\ \eta &= 1 \end{aligned}\right\} \tag{15.96}$$

The anti-resonant impedance of Z_{AB} is represented by $O - P_4$ in Fig. 15.10. The left hand term of (15.95) for the P_4 operating point becomes

$$Z_{AB}\omega_c(C_0 + C_t) = \sigma + \sqrt{\sigma^2 - 1} \tag{15.97}$$

Substituting this value in (15.95) and solving for Z, we find

$$Z = -jM\left[K_3\frac{C_0}{C_0 + C_t} - 1\right] \tag{15.98}$$

where

$$K_3 = \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{1}{\sigma^2}}$$

If K_3 is expanded and all except the first two terms are neglected, Z may be expressed as

$$Z = -j\left[\frac{M}{1 + \dfrac{C_0}{C_t}} + \frac{\left(1 + \dfrac{C_t}{C_0}\right)}{M}\right] \tag{15.99}$$

The next step is to obtain a similar expression to (15.99) only for the minimum frequency impedance of Z_i in Fig. 15.3 with ρ disconnected. For this application $S = \dfrac{1}{j\omega C_t}$ and $T = \infty$. Substituting these values, as well as those in (15.59), in (15.75), we have

$$Z_i = \frac{\left[\tau + jM\tau + \dfrac{1}{\omega C_t}\right] + \left[\tau + \dfrac{1}{\omega C_t}\right]Z}{j + jZ} \tag{15.100}$$

From this equation, values for σ and η may be determined as described in Section 15.81.

$$\left.\begin{aligned} \sigma &= \frac{M\tau}{2} \\ \eta &= \frac{C_0 + C_t}{C_t}\tau \end{aligned}\right\} \tag{15.101}$$

By the same procedure just described for evaluating Z from the impedance expression Z_{AB}, the value of Z corresponding to the minimum impedance operating point, P_1, (Fig. 15.10) must be determined. For this operating condition Z_1 may be expressed by (15.65). The coefficients of this operating point are given by (15.67) with the above values of σ and η (Equation 15.101). Utilizing (15.68) to solve for Z, we get

$$Z_i' = -j \frac{M}{\left(1 + \frac{C_0}{C_t}\right)} \frac{\left[1 + \sqrt{1 + \frac{4}{P_1^2}}\right]}{2} \text{ (minimum impedance)} \quad (15.102)$$

now

$$Z_i - Z_i' = \frac{2L_1}{R_1} (\Delta\omega)$$

where $\Delta\omega$ = the difference in radian frequency between the frequency of oscillation in the generalized oscillator (anti-resonant frequency of the impedance, Z_{AB}) and the frequency of oscillation in the PI meter (minimum impedance frequency of the impedance, Z_i')

$$\Delta\omega = \left[\frac{\omega_2 - \omega_1}{M^2}\left(1 + \frac{C_0}{C_t}\right) \frac{2}{\left[1 + \sqrt{1 + \frac{4}{P_1^2}}\right]} - \frac{\omega_2 - \omega_1}{M^2}\left(1 + \frac{C_t}{C_0}\right)\right] \quad (15.51)$$

It is of interest to note that (15.102) becomes (15.44) when the value of Z (15.69) is introduced.

Lightning Protection of Buried Toll Cable

By E. D. SUNDE

A theoretical study of lightning voltages in buried telephone cable, of the liability of such cable to damage by lightning and of remedial measures, together with the results of simulative surge tests, oscillographic observations of lightning voltages and lightning trouble experience.

Introduction

PRACTICALLY all of the toll cable installed since 1939 has been of the carrier type and most of it has been buried in order to secure greater immunity from mechanical damage. It was realized, however, that burying the cable would not prevent damage due to lightning and that, on account of their smaller size, more damage was to be expected on the new carrier cables than on the much larger voice-frequency underground cables then in use. Moreover, when damage by lightning does occur, such as fusing of cable pairs or holes in the sheath, it is not so easy to locate and repair as on aerial cables, since excavations may have to be made at a number of points. Studies were therefore made of the factors affecting damage of buried cables by lightning and remedial measures were devised and put into effect in cases where a high rate of lightning failures was anticipated on new installations, or was experienced with cable already installed. Most of the cable installed was thus provided with extra core insulation, and shield wires were plowed in on many of the new routes.

It was recognized early in these studies that more effective lightning protection might be secured by providing the lead sheath with a thermoplastic coating of adequate dielectric strength and an outside copper shield, and that such cable might be required in territory where the earth resistivity is very high. This type of cable has recently been installed on a route in high-resistivity territory where experience has indicated that other types of construction would probably be inadequate and, since it has advantages also from the standpoint of corrosion and mechanical protection, it may be used also where lightning is not of such decisive importance.

When lightning strikes, the current spreads in all directions from the point where it enters the ground. If there are cables in the vicinity they will provide low resistance paths, so that much of the current will flow to the cables near the lightning stroke and in both directions along the sheath to remote points. The flow of current in the ground between the lightning channel and the cables may give rise to such a large voltage drop that the breakdown voltage of the soil is exceeded, particularly when the earth

resistivity is high. The lightning stroke will then arc directly to the cables from the point where it enters the ground, often at the base of a tree. Furrows longer than 100 feet have been found in the ground along the path of such arcs.

The current entering the sheath near the stroke point is attenuated as it flows towards remote points. Since a high earth resistivity is accompanied by a small leakage conductance between sheath and ground, the current will travel farther the larger the earth resistivity. The current along the sheath produces a voltage between the sheath and the core conductors, which is largest at the stroke point. This voltage is equal to the resistance drop in the sheath, between the stroke point and a point which is sufficiently remote so that the current in the sheath is negligible. Since the current travels farther along the sheath the higher the earth resistivity, this resistance drop will also increase with the earth resistivity. The maximum voltage between sheath and core is thus proportional to the sheath resistance and, as it turns out, to the square root of the earth resistivity. Carrier cables now being used are of smaller size and have a higher sheath resistance than full-size voice-frequency cables, and for this reason they are liable to have more lightning damage, particularly when the earth resistivity is high.

To secure experimental verification of certain points of the theory presented here, staged surge tests were made on the Stevens Point-Minneapolis cable, one of the first small-size buried toll cables to be installed. The results of these tests, which have already been published,[1] are here compared with those obtained theoretically, on the basis of the earth resistivity measured at the test location. Lightning voltages on this cable route were also recorded by automatic oscillographs and the results of these observations are also briefly discussed together with the rate of lightning failures experienced on this and other routes.

The first part of the paper deals with voltages between the cable conductors and the sheath due to sinusoidal currents and surge currents. The second part deals with the liability of damage due to excessive lightning voltages and with certain characteristics of lightning discharges of importance in connection with the present problem, such as the impedance encountered by the lightning channel in the ground, the rate of lightning strokes to ground and to buried structures and the crest current distribution for such strokes. In the third part remedial measures are discussed, together with lightning-resistant cable.

I. Voltages Between Cable Conductors and Sheath

1.1 *General*

Cable installed in the ground is designated "underground" when placed in duct, and "buried" when not in duct. Buried cable is sometimes pro-

vided with steel tape armor for protection against mechanical damage. While such armor may also reduce voltages due to low-frequency induction, mainly because of the high permeability of the steel, this is not true in the case of lightning voltages. The magnetic field in the armor due to lightning current in the cable is rather high, and the corresponding permeability fairly low. The armor resistance is, furthermore, quite high compared to that of the sheath, so that the effect of the armor may be neglected in considering lightning voltages. The tape or armor is usually separated from the sheath by paper and asphalt, but is bonded to the sheath at every splice point. Strokes to ground, or to the cable, may give rise to large currents in the armor and thus to excessive voltages between the armor and the sheath some distance from bonding points. The resulting arc may fuse a hole in the sheath or dent it, due to the explosive effect of the confined arc, and insulation failures may be experienced on this account. Such failures are not considered here since they are usually confined to a single point and are thus of less importance than insulation failures due to excessive voltages between the core conductors and the sheath, which may be spread for a considerable distance along the cable.

For protection against corrosion, buried cables are usually jute-covered (asphalt, paper and jute) and in some cases have thermoplastic or rubber coating. The leakance of jute-covered sheaths is usually large enough so that the cable may be assumed to be in direct contact with the earth and the leakance is, furthermore, increased at the time of lightning strokes by numerous punctures due to excessive voltage between sheath and ground. This effect is large enough so that even rubber-covered cable may be regarded as in direct contact with the soil in the case of direct strokes and sometimes also for strokes to ground in the vicinity of the cable, as discussed later.

In order to calculate the voltage between the sheath and the core conductors of a buried cable, due to a surge current entering the sheath or the ground in the vicinity of the cable, it is convenient to consider at first a sinusoidal current. The voltage due to a unit step current may then be obtained by operational solution and, in turn the voltage for a current $J(t)$ of arbitrary wave shape, by means of either one of the integrals:

$$\begin{aligned} V(t) &= \int_0^t J'(t - \tau)S(\tau)\, d\tau \\ &= \int_0^t J(t - \tau)S'(\tau)\, d\tau \end{aligned} \tag{1}$$

where $S(t)$ is the voltage due to unit step current and $S'(t)$ the time derivative of this voltage. The second of the above integrals is more convenient in the present case.

Photographic observations[2] indicate that a lightning discharge is usually initiated in the cloud by a so-called "stepped leader," except in the case of discharges to sufficiently tall structures[3] where this leader is initiated at the ground end. After the leader reaches the ground, or the cloud in the case of a tall structure, a heavy current "return stroke" proceeds from the ground toward the cloud at about $\frac{1}{10}$ the velocity of light. The main surge of current in the return stroke, which usually lasts for less than 100 microseconds, may be followed by a low current lasting for $\frac{1}{10}$ second or so.

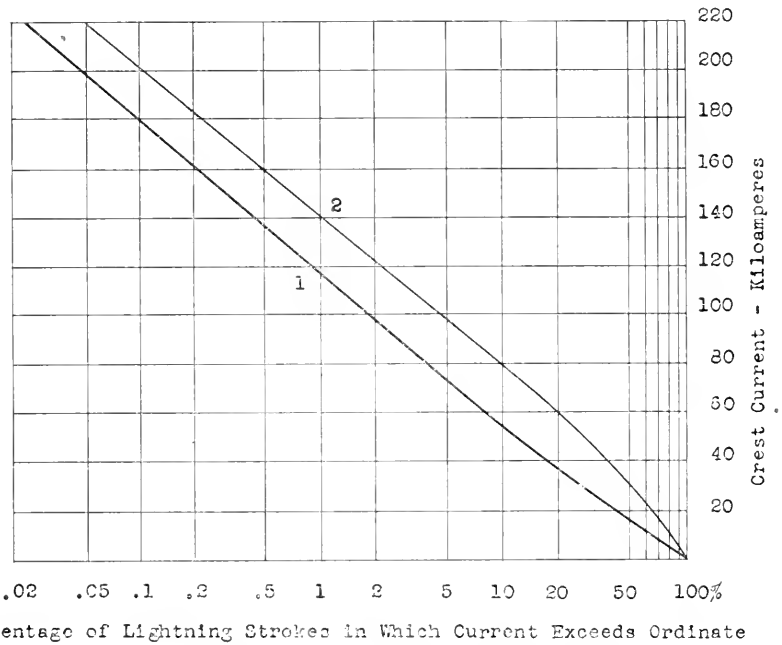


Figure 1—Distribution of crest currents in lightning strokes.

Curve 1: Currents in strokes to transmission line ground structures, based on 4410 measurements, 2721 in U. S. and 1689 in Europe.

Curve 2: Currents in strokes to buried structures, derived from curve 1.

There may then be a second leader, which does not exhibit the stepped character of the first leader and always proceeds from the cloud, and a second return stroke. This may be followed by a third leader and so on, the average number of strokes in multiple discharges being about 4 and the average time interval between strokes about $\frac{1}{10}$ second. Single-stroke discharges are, however, most common, discharges having more than 6 strokes being quite rare although discharges with as many as 40 strokes have been observed.

The crest value of currents in lightning discharges varies over wide limits.

Measurements of current in the ground structure of transmission lines[4,5] indicate that a relationship as shown in Fig. 1 exists between the crest currents and the percentage of discharges in which they occur. In the same figure is shown the crest current distribution for strokes to buried structures, which is derived in Part II from the curve for strokes to transmission line ground structures. Although measurements of wave shape are not extensive, they indicate that the current reaches its crest value in 5 to 10 microseconds and that it decays to half its maximum in 25 to 100 microseconds, the average being about 50 microseconds.[6] An average wave shape is assumed in this investigation.

In some 80 per cent of all lightning discharges the cloud is negative, so that the flow of current is from the earth toward the cloud. Further details about lightning discharges are summarized in recent surveys[6,7] which also contain an extensive list of references.

1.2 *Direct Strokes—Current Propagation Along Sheath*

As mentioned in the introduction and discussed further in Part II, a lightning stroke to ground may arc to a buried cable in the vicinity, in which case virtually all of the current will enter the sheath near the stroke point.

When a sinusoidal current J enters the sheath at $x = 0$ and the sheath is assumed to extend indefinitely in opposite directions from this point, the sheath current at the distance x is given by the following approximate expression

$$I(x) = \frac{J}{2} e^{-\Gamma x} \tag{2}$$

where Γ is the propagation constant of the sheath-earth circuit and is given by the following expression, derived in a previous paper.[8]

$$\Gamma = \frac{1}{v} [i\omega(i\omega + 1/\rho\kappa)]^{\frac{1}{2}} \tag{3}$$

where: v = Velocity of propagation along sheath
$= (2/\nu\kappa)^{\frac{1}{2}}$ meters per second
ν = Inductivity of earth = $1.256 \cdot 10^{-6}$ hy/meter
κ = Capacitivity of earth = $\epsilon \cdot 8.858 \cdot 10^{-12}$ fd/meter
ϵ = Dielectric constant of earth
ρ = Earth resistivity, meter-ohms

In deriving the above formula the resistance of the sheath is neglected in comparison with its external reactance, which is permissible for frequencies in the range of importance, and the sheath is assumed to be half buried, that is, with its axis in the plane of the earth's surface. The latter assumption gives rise to a comparatively small error when the formula is applied

to cables buried at depths up to one meter or so, the propagation constant for a cable buried at infinite depth being larger than that given above by a factor of $\sqrt{2}$.

It is assumed that the current is propagated from the cable up the lightning channel with infinite velocity. The voltage between the cable conductors and the sheath obtained in this manner reaches a crest value after some 50 to 100 microseconds, or after the current in an actual lightning channel has traveled from the ground to the cloud. The error due to this assumption is thus probably quite small as regards the crest voltage, although the wave front will be somewhat slower when the actual velocity of propagation is considered.

1.3 *Direct Strokes—Voltage for Sinusoidal Current*

The current along the sheath gives rise to an electric force along the latter. The electric force along the inner surface of the sheath is given by:

$$E(x) = zI(x) = \frac{J}{2} ze^{-\Gamma x} \tag{4}$$

where z is the mutual-impedance of the sheath-earth and core-sheath circuits.

The latter mutual impedance is equal to the ratio of electric force along the inner surface of the sheath at any point, to the total current along the sheath at the same point, and for low frequencies equals the direct-current resistance of the sheath. It is given by the following slightly approximate formula:[9]

$$Z = R(i\omega\gamma)^{\frac{1}{2}}/\sinh\,(i\omega\gamma)^{\frac{1}{2}} \tag{5}$$

where: $\omega = 2\pi f$ and
R = Unit length d-c resistance of sheath, ohms/meter
$\gamma = \nu\delta/2\pi aR$
δ = Thickness of sheath, meter
a = Radius of sheath, meter
ν = Intrinsic Inductivity of sheath
$= 1.256 \times 10^{-6}$ henrys/meter

The current in the core-sheath circuit and the voltage between core and sheath due to an impressed field $E(x)$ along the core (inner surface of sheath) are obtained from the following equations, which are the general solutions of the transmission line equation for the core-sheath circuit.

$$J(x) = [A + P(x)]e^{-\Gamma_0 x} - [B + Q(x)]e^{\Gamma_0 x} \tag{6}$$

$$U(x) = K_0[A + P(x)]e^{-\Gamma_0 x} + K_0[B + Q(x)]e^{\Gamma_0 x} \tag{7}$$

where Γ_0 and K_0 are the propagation constant and the characteristic impedance of the core-sheath circuit and

$$P(x) = \frac{1}{2K_0}\int_0^x E(x)e^{\Gamma_0 x}\,dx = \frac{z}{2K_0}\,\frac{1 - e^{-(\Gamma-\Gamma_0)x}}{\Gamma - \Gamma_0} \tag{8}$$

$$Q(x) = \frac{1}{2K_0}\int_0^x E(x)e^{-\Gamma_0 x}\,dx = \frac{z}{2K_0}\,\frac{1 - e^{-(\Gamma+\Gamma_0)x}}{\Gamma - \Gamma_0} \tag{9}$$

Since the current must be zero at $x = 0$, it is necessary that $A = B$. To make the curent vanish when x becomes infinity, B must equal $-Q(\infty) = -\frac{z}{2K_0}\,\frac{1}{\Gamma + \Gamma_0}$. With these boundary conditions the voltage between core and sheath becomes:

$$U(x) = \frac{J}{2}\,\frac{z}{\Gamma^2 - \Gamma_0^2}\,(\Gamma e^{-\Gamma x} - \Gamma_0 e^{-\Gamma_0 x}) \tag{10}$$

The propagation constant Γ_0 is much smaller than Γ, and may be taken as:

$$\begin{aligned}\Gamma_0 &= [(R_0 + i\omega L_0)i\omega C_0]^{\frac{1}{2}} \\ &= \frac{1}{v_0}\,[i\omega(i\omega + R_0 L_0)]^{\frac{1}{2}}\end{aligned} \tag{11}$$

R_0 = Unit length resistance of core-sheath circuit, ohms/meter
L_0 = Unit length inductance of core-sheath circuit, hy/meter
C = Unit length capacitance of core-sheath circuit, fd/meter
$v_0 = (1/L_0C_0)^{\frac{1}{2}}$

1.4 *Direct Strokes—Lightning Voltage at Stroke Point*

The largest voltage between sheath and core conductors is obtained for $x = 0$, and for this case (10) becomes:

$$\begin{aligned}U(0) &= \frac{J}{2}\,\frac{z}{\Gamma + \Gamma_0} \\ &= \frac{J}{2}\,\frac{R\gamma^{\frac{1}{2}}/\sinh\,(i\omega\gamma)^{\frac{1}{2}}}{\frac{1}{v}\,(i\omega + 1/\rho\kappa)^{\frac{1}{2}} + \frac{1}{v_0}\,(i\omega + R_0 L_0)^{\frac{1}{2}}}\end{aligned} \tag{12}$$

For sufficiently high frequencies, so that $(i\omega\gamma)^{\frac{1}{2}} \gg 1$, $i\omega > 1/\rho\kappa$, $i\omega > R_0L_0$, and $\sinh\,(i\omega\gamma)^{\frac{1}{2}} \leqq \frac{1}{2}\exp\,(i\omega\gamma)^{\frac{1}{2}}$ expression (12) becomes

$$U(0) = JR\gamma^{\frac{1}{2}}\,\frac{vv_0}{v + v_0}\,(i\omega)^{-\frac{1}{2}}e^{-(i\omega\gamma)^{\frac{1}{2}}} \tag{13}$$

For sufficiently low frequencies, so that $(i\omega\gamma)^{\frac{1}{2}} < 1$, $i\omega < i/\rho\kappa$, $i\omega < R_0/L_0$, and $\sinh (i\omega\gamma)^{\frac{1}{2}} \leqq (i\omega\gamma)^{\frac{1}{2}}(1 + i\omega\gamma/6)$, expression (12) becomes

$$U'(0) = \frac{J}{2} \frac{R}{\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}}} \frac{1}{(i\omega)^{\frac{1}{2}}(i + \omega\gamma\ 6)^{\frac{1}{2}}} \tag{14}$$

where

$$\alpha = \nu/2\rho, \qquad \beta = R_0C_0$$

For small values of time, corresponding to large values of $i\omega$, the function $S'(t)$ defined before, as obtained by operational solution of (13) is [10]

$$S'(t) = R\gamma^{\frac{1}{2}} \frac{vv_0}{v + v_0} \left(\frac{1}{\pi t}\right)^{\frac{1}{2}} e^{-\gamma\ 4t} \tag{15}$$

For large values of time, corresponding to small values of $i\omega$, the function is obtained by operational solution of (14) and equals: (Reference 10, pair 542)

$$S'(t) = \frac{R}{2(\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}})} \left(\frac{6}{\gamma}\right)^{\frac{1}{2}} h(\sqrt{6t/\gamma}) \tag{16}$$

where, with $(6t/\gamma)^{\frac{1}{2}} = u$:

$$h(u) = -ie^{-u^2} \operatorname{erf} (iu) = \frac{2}{\sqrt{\pi}} e^{-u^2} \int_0^u e^{\tau^2}\, d\tau \tag{17}$$

erf being the error function.

Values of the function $h(u)$ are given in Table I.

In Fig. 2, curve 1 shows the function $S'(t)$ calculated from (15), and curve 2 that calculated from (16), for a cable of 1.4″ diameter, using constants as indicated in figure. The constants apply to a cable on which measurements have been made of the voltage between sheath and core conductors, at a location where the measured earth resistivity was 400 meter-ohms. The function $S'(t)$ corresponding to equation (12) is obtained with sufficient accuracy by drawing a transition curve, 3, between curves 1 and 2.

If the impedance z is taken equal to the direct-current resistance R of the sheath and if the velocity of propagation along the sheath and along the core are assumed to be infinite, so that $\Gamma = (i\omega\alpha)^{\frac{1}{2}}$ and $\Gamma_0 = (i\omega\beta)^{\frac{1}{2}}$, the following expression is obtained

$$S'(t) = \frac{R}{2(\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}})} \left(\frac{1}{\pi t}\right)^{\frac{1}{2}} \tag{18}$$

In the following it will be shown that (18) is accurate enough for practical purposes.

The wave shape of the current in lightning strokes may be approximated by an expression of the form:

$$J(t) = I(e^{-at} - e^{-bt}). \tag{19}$$

With $a = .013 \cdot 10^6$, $b = .5 \cdot 10^6$, a current of the wave shape used in the measurements referred to above is obtained. This current reaches its crest in 10 microseconds and decays to its half-value in 65 microseconds, and is fairly representative of the average wave shape of lightning stroke currents. In the following, the voltages are for convenience referred to a crest value of 1000 amperes, which is obtained when $I = 1150$ amperes, the latter current being the initial value of each of the two exponential component currents included in (19).

TABLE I

$$\sqrt{\frac{\pi}{2}}\, h(u) = e^{-u^2} \int_0^u e^{\tau^2}\, d\tau = -\sqrt{\frac{\pi}{2}}\, ie^{-u^2}\, erf(iu)$$

$$\cong u \text{ when } u < .1$$

$$\cong \frac{1}{2u} \text{ when } u > 10$$

u	$\sqrt{\frac{\pi}{2}}\, h(u)$	u	$\sqrt{\frac{\pi}{2}}\, h(u)$
0	0	1.5	.4283
.1	.0993	2.	.3014
.2	.1948	2.5	.2232
.3	.2826	3.0	.1782
.4	.3599	3.5	.1496
.5	.4244	4.	.1293
.6	.4748	4.5	.1141
.7	.5105	5.	.1021
.8	.5321	6.	.0845
.9	.5407	8.	.0630
1.0	.5381	10.	.0503

A more complete table for the range between $u = 0$ and $u = 4$ is published in *Berichteverhandlungen Akademie der Wissenschaften, Leipzig, Math-Phys. Klasse*, Vol. 80, 1928, pages 217 to 223.

In Fig. 3, the dashed curve shows the measured voltage and curves 1 and 2 that calculated for the above surge current for two conditions. In calculating curve 1, $S'(t)$ was taken as curve 3 of Fig. 2, the voltage being obtained by numerical integration in accordance with (1); in calculating curve 2, z is taken as the direct-current resistance of the sheath and the velocities of propagation are assumed to be infinite, so that $S'(t)$ is given by (18). In the latter case the following expression is obtained for the voltage by solution of (1):

$$V(t) = \frac{JR}{2(\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}})} \left[a^{-\frac{1}{2}} h(\sqrt{at}) - b^{-\frac{1}{2}} h(\sqrt{bt})\right] \tag{20}$$

where the function $h(u)$, $u = \sqrt{at}$ or $\sqrt{bt}$, is defined as before.

Comparison of curves 1 and 2 of Fig. 3 shows that (20) is accurate enough for practical purposes, so that the voltage may be taken proportional to the direct-current resistance of the sheath. Since $\beta^{\frac{1}{2}}$ is only 2.5 per cent of $\alpha^{\frac{1}{2}}$, propagation in the core-sheath circuit may be neglected in comparison with propagation along the sheath-earth circuit, so that it is permissible to take the voltage proportional to the square root of the earth resistivity.

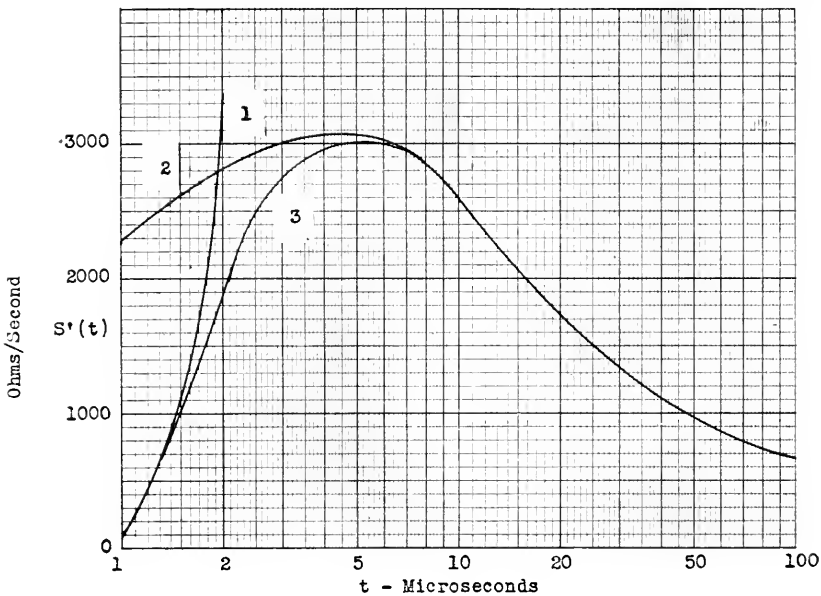


Figure 2—Approximate solution for $S'(t)$.

1: Calculated from formula for small times.
2: Calculated from formula for large times.
3: Transition curve giving approximate solution for $S'(t)$.
Earth resistivity, $\rho = 400$ meter-ohms.
Radius of cable, $a = 1.75$ cm.
Sheath thickness, $\delta = 2.4$ mm.
Sheath resistance, $R = .92 \cdot 10^{-3}$ ohms meter.
Core-sheath cap. $C_0 = .96 \cdot 10^{-9}$ fd meter.
Core-sheath resist. $R_0 = R \cong .92 \cdot 10^{-3}$ ohm meter.
Velocity $v_0 = 2 \cdot 10^8$ meter sec.
Velocity $v = 1 \cdot 10^8$ meter sec.

Furthermore, from (20) it is seen that when a and b are divided by the same factor k, so that the wave shape of the current remains the same but the duration of the current is increased k times, the voltage is increased $\sqrt{k}$ times. Thus, if the surge current had reached its crest value in 20 microseconds and its half-value in 130 microseconds, the voltage would be increased by $\sqrt{2}$, and the crest voltage would have been reached after 120 rather than 60 microseconds.

If the breakdown voltage of the core insulation is assumed to be 2000 volts, the above cable would be able to withstand a stroke current of about 30,000 amperes before the insulation is punctured. From Fig. 1 it is seen

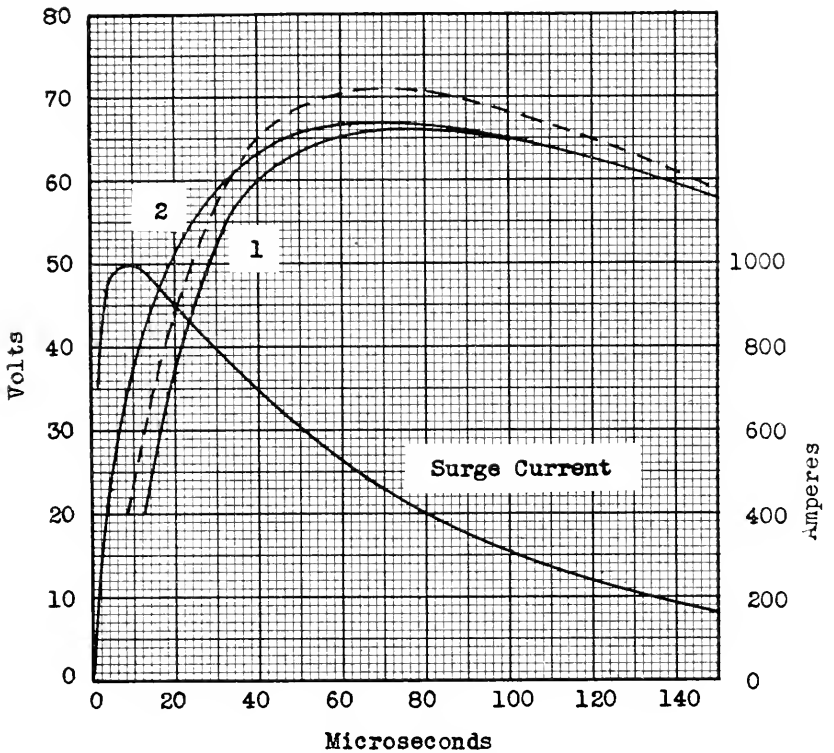


Figure 3—Comparison of measured, shown by dashed curve, and calculated voltage between sheath and core conductors, shown by curves 1 and 2, for surge current as shown and cable constants as given in Fig. 2.

1: Calculated from formula including skin-effect in sheath and finite velocity of propagation.

2: Calculated from formula based on d-c resistance of sheath and assuming infinite velocity of propagation.

that in about 50 per cent of all strokes the crest current exceeds 30,000 amperes.

When there are two cables, each will provide shielding for the other, and the shielding effect may be calculated as for shield wires (Sec. 3.3). Frequently the cables are of equal or nearly equal size and are close together. It is then accurate enough to use the parallel resistance of the two sheaths in calculating the voltage, which will be practically the same in both cables.

1.5 *Direct Strokes—Lightning Voltages Along Cable*

It was shown above that with only a minor error the impedance z may be taken equal to the direct-current resistance of the sheath and that the propagation constants may be taken as

$$\Gamma = (i\omega\alpha)^{\frac{1}{2}}, \qquad \Gamma_0 = (i\omega\beta)^{\frac{1}{2}}$$

With this modification expression (10) becomes:

$$U(x) = \frac{J}{2}\frac{R}{\alpha - \beta}[\alpha^{\frac{1}{2}}e^{-(i\omega\alpha)^{\frac{1}{2}}x} - \beta^{\frac{1}{2}}e^{-(i\omega\beta)^{\frac{1}{2}}x}](i\omega)^{-\frac{1}{2}} \tag{21}$$

The corresponding function S' is:[10]

$$S'(x, t) = \frac{R}{2(\alpha - \beta)}\left(\frac{1}{\pi t}\right)^{\frac{1}{2}}(\alpha^{\frac{1}{2}}e^{-\alpha x^2/4t} - \beta^{\frac{1}{2}}e^{-\beta x^2/4t}) \tag{22}$$

The voltage due to a surge current $J(t)$, as obtained from (1), may be expressed as:

$$V(x, t) = \frac{R}{2(\alpha - \beta)}[\alpha^{\frac{1}{2}}g(\alpha^{\frac{1}{2}}x, t) - \beta^{\frac{1}{2}}g(\beta^{\frac{1}{2}}x, t)] \tag{23}$$

where, with $\sigma = \alpha^{\frac{1}{2}}x$ or $\beta^{\frac{1}{2}}x$,

$$g(\sigma, t) = \int_0^t J(t - \tau)\left(\frac{1}{\pi\tau}\right)^{\frac{1}{2}}e^{-\sigma^2/4\tau}\,d\tau \tag{24}$$

For a current as given by (19), the latter integral may be expressed in terms of error functions of complex arguments, for which, however, no tables are available at present. Curves for the function g, as obtained by numerical integration are shown in Fig. 4.

When the core conductors are connected to the sheath at $x = 0$, the constants A and B of (6) and (7) are obtained from the following boundary conditions: At $x = 0$, $V(0) = 0$ so that $A = -B$. As before, $B = -Q(\infty)$. The voltage between core and sheath is then given by:

$$\begin{aligned} U_0(x) &= \frac{J}{2}\frac{R\Gamma}{\Gamma^2 - \Gamma_0^2}(e^{-\Gamma x} - e^{-\Gamma_0 x}) \\ &= \frac{JR\alpha^{\frac{1}{2}}}{\Gamma^2 - \Gamma_0^2}[e^{-(i\omega\alpha)^{\frac{1}{2}}x} - e^{-(i\omega\beta)^{\frac{1}{2}}x}] \end{aligned} \tag{25}$$

In this case the derivative of the voltage due to unit step current is:

$$S_0'(x, t) = \frac{R\alpha^{\frac{1}{2}}}{2(\alpha - \beta)}\left(\frac{1}{\pi t}\right)^{\frac{1}{2}}[e^{-\alpha x^2/4t} - e^{-\beta x^2/4t}] \tag{26}$$

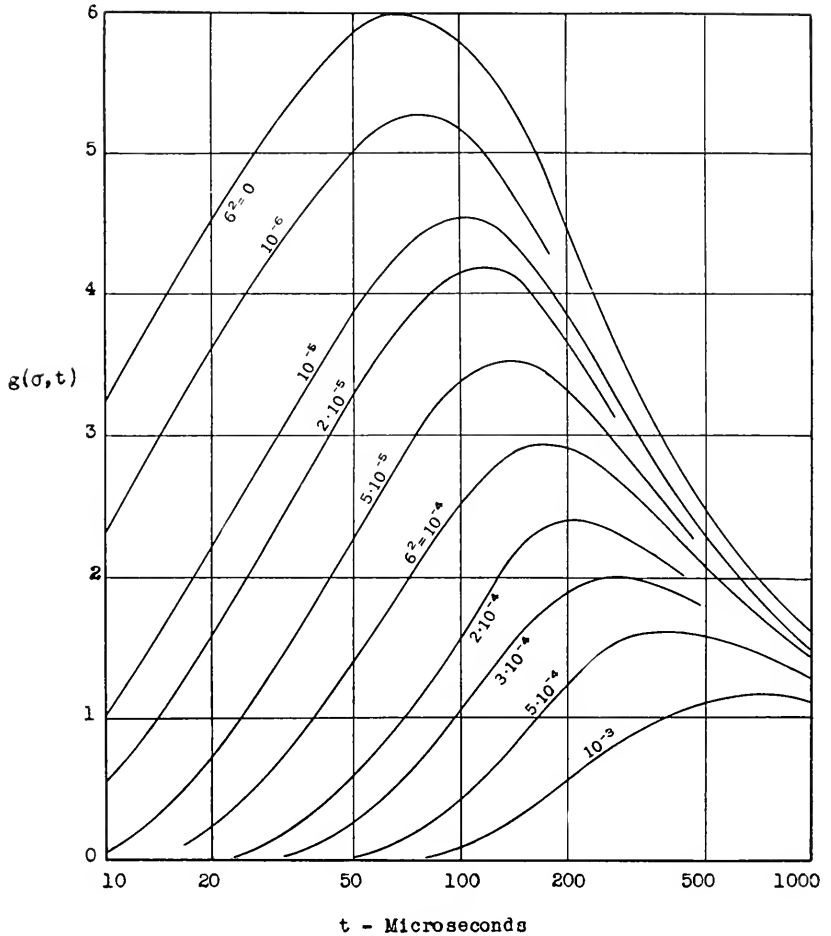


Figure 4—Function $g(\sigma, t) = \int_0^t J(t-\tau)\left(\frac{1}{\pi\tau}\right)^{\frac{1}{2}} e^{-\sigma^2/4\tau}\, d\tau$

$J(t) = I(e^{-at} - e^{-bt})$

$a = 1.3 \cdot 10^4$, $b = 5 \cdot 10^5$ $I = 1150$ amperes

and

$$V_0(x, t) = \frac{R\alpha^{\frac{1}{2}}}{2(\alpha - \beta)} [g(\alpha^{\frac{1}{2}}x, t) - g(\beta^{\frac{1}{2}}x, t)] \tag{27}$$

In Fig. 5 the crest values of $V(x, t)$ and $V_0(x, t)$, calculated for the cable considered before, are plotted against x, together with those observed in the tests. When the voltage $V(0, t)$ at the point where current enters the sheath is great enough to break down the insulation of a core conductor, the

latter will be in contact with the sheath by virtue of arcing. Under this condition, the voltage $V_0(x, t)$ between this conductor and the sheath will increase with distance along the cable as shown in Fig. 5. A maximum is reached a fairly short distance from the original fault, and beyond this point the voltage slowly decreases. After a puncture of the insulation where current enters the sheath, other failures may therefore occur, not necessarily at the point where $V_0(x, t)$ is largest, but sometimes at points nearer or much farther away where the insulation may be weaker. A single lightning stroke may thus cause insulation failures over a considerable distance along the cable.

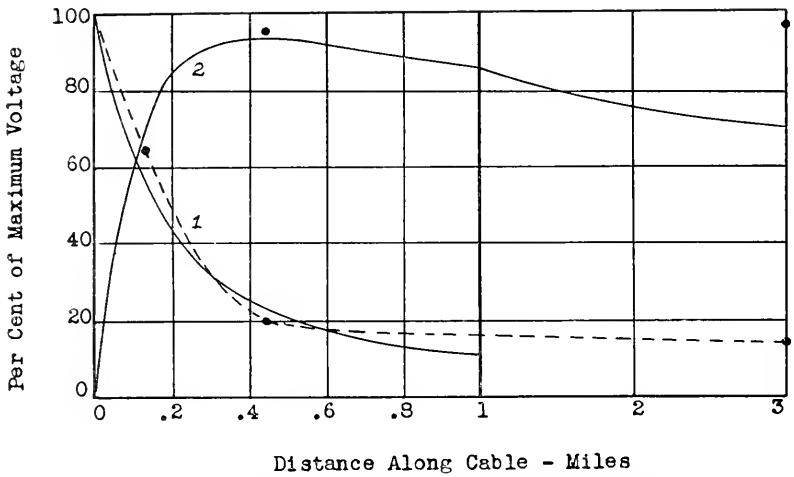


Figure 5—Comparison of measured variation of voltage between sheath and core, along cable, as shown by points and dashed curve, with calculated variation shown by curves 1 and 2.
1: Conductor not connected to sheath.
2: Conductor connected to sheath at point where surge current enters sheath.

1.6 *Direct Strokes—Voltage Due to Long Duration Current*

As mentioned before, a current of low value and long duration may exist on the lightning channel after the main discharge. This current is usually of such long duration that the resistance of the sheath must be considered in calculating the current propagation along the cable. The propagation constant in that case becomes

$$\Gamma = [(R + i\omega L)G]^{\frac{1}{2}} = \left(\frac{R}{L} + p\right)^{\frac{1}{2}} \left(\frac{\nu}{2\rho}\right)^{\frac{1}{2}} \tag{28}$$

R, L and G being the unit length resistance, inductance and leakance of the sheath-earth circuit. Neglecting propagation in the core-sheath circuit,

the voltage between core and sheath at the stroke point due to a sinusoidal current J_1 becomes,

$$U_1(0) = \frac{J_1 R}{2\Gamma} = J_1\left(\frac{\rho}{2\nu}\right)^{\frac{1}{2}} R(p + R L)^{-\frac{1}{2}} \tag{29}$$

The corresponding voltage for a unit step current J_1 is:

$$V_1(0, t) = J_1 R\left(\frac{\rho L}{2\nu R}\right)^{\frac{1}{2}} \operatorname{erf} (Rt L)^{\frac{1}{2}} \tag{30}$$

where erf is the error function.

For large values of time, when $Rt/L > 1$, (30) becomes

$$V_1(0, t) = J_1 R\left(\frac{\rho L}{2\nu R}\right)^{\frac{1}{2}} \tag{31}$$

The latter expression is valid when t exceeds about 2 milliseconds and thus applies for the long duration current of a lightning stroke, since the latter usually lasts for about 100 milliseconds. For a current of 1000 amperes, the core-sheath voltage for a cable of 1.4″ diameter is about 700 volts. In many strokes the long duration current may be several hundred amperes, and a substantial voltage may then exist between core and sheath for .1 second or so. Thus, while this current component does not increase the crest voltage, it substantially increases the likelihood of permanent failure when the insulation is punctured by prolonging the current through the puncture.

1.7 *Strokes to Ground Not Arcing to Cable*

Let it be assumed that the current enters the ground at the distance y from a buried cable and that conditions are such that it does not arc to the latter. The flow of current in the ground gives rise to an electric force in the ground along the cable, and thus to currents in the sheath and to voltages between core and sheath. When the earth is assumed to have uniform conductivity, the earth potential at the distance r from the point where current enters the ground is given by:

$$Ve = JQ_0(r) = J\rho/2\pi r \tag{32}$$

where

ρ = Earth resistivity in meter-ohms
$r = (x^2 + y^2)^{\frac{1}{2}}$ = Distance in meters

The sheath current and the voltage between sheath and core may in this case be obtained from published formulas,[11] provided propagation along

the core-sheath circuit is neglected in comparison with propagation along the sheath-earth circuit, which is permissible. The voltage between sheath and core conductor differs by the factor z/Z from the voltage between sheath and ground as given in Table II, case 3 of the paper referred to, z being defined as before and Z being the unit length self-impedance of the sheath-earth circuit. At a point opposite the lightning stroke, $x = 0$, the voltage between core and sheath is in this case given by:

$$U(0, y) = J\frac{R\Gamma}{Z}\int_0^\infty Q_0(r)e^{-x}\,dx = J\frac{R}{Z}\frac{\rho}{2\pi}\Phi(\Gamma y) \tag{33}$$

where $Z = \Gamma^2/G$ and G is the unit length leakance of the sheath-earth circuit. The leakance is given by the approximate expression:

$$G = \left(\frac{\rho}{\pi}\log\frac{1}{\Gamma a}\right)^{-1} \tag{34}$$

a being the radius of the sheath and $\log = \log_\epsilon$.

The function $\Phi(\Gamma y)$ is given by the approximate formula:

$$\Phi(\Gamma y) \leqq \log\frac{1 + \Gamma y}{\Gamma y} \tag{35}$$

Inserting (34) and (35) in (33), the latter expression may be written:

$$U(0, y) = U(0, a)\lambda(\Gamma y) \tag{36}$$

where $U(0, a) = U(0)$ is the voltages when the current enters the sheath directly ($y = a$) and:

$$\lambda(\Gamma y) = \left(\log\frac{1 + \Gamma y}{\Gamma y}\right)\Big/\log 1/\Gamma a \tag{37}$$

where $\Gamma = (i\omega\nu/2\rho)^{\frac{1}{2}} = (i\omega\alpha)^{\frac{1}{2}}$.

The rigorous solution of the time function corresponding to (36) would be rather complicated. Since, however, λ is the ratio of two functions, each of which varies logarithmically with Γ, and thus varies only slightly with $i\omega$, an approximate solution is obtained by replacing $i\omega$ with $1/t$ in (37). For instance, the solution of an operational expression p^{-n} is $t^n/n!$ while the solution of $p^{-n}\log p$ is $[\psi(1 + n) + \log 1/t]t^n/n!$, ψ being the logarithmic derivative of the gamma function. For representative values of n and $t(n < 1, t < 10^{-4})$, ψ is less than 5% of $\log 1/t$, so that a good approximation is obtained by replacing p by $1/t$ in $\log p$, which in this illustration simulates the factor $\lambda(\Gamma y)$. With this approximation:

$$U(0, y, t) = U(0, a, t)\lambda[y(\alpha/t)^{\frac{1}{2}}] \tag{38}$$

In Fig. 6 is shown the variation in the voltage calculated from (38), together with that observed in the tests referred to before. That the measured decrease in the voltage is smaller than calculated is due to the fact that the earth resistivity at the test location increases with depth. Earth resistivity measurements made by the four-electrode method show that the resistivity is about 400 meter-ohms for electrode spacings up to about 20 feet and then gradually increases, reaching about 700 meter-ohms at 300 feet, 1200 meter-ohms at 1000 feet and approaching 1500 meter-ohms for

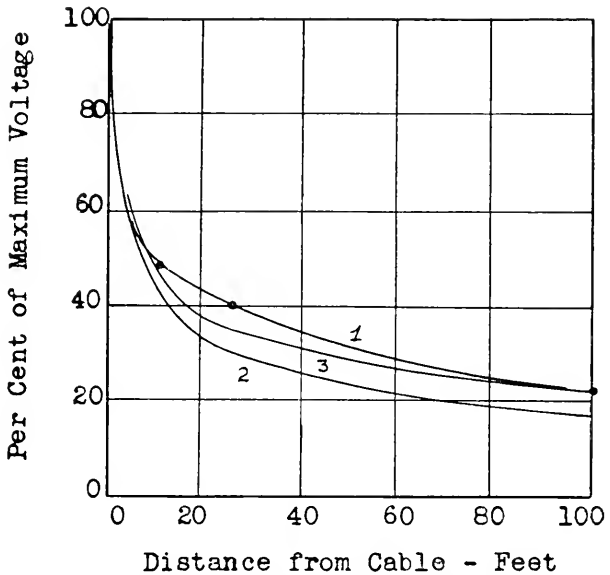


Figure 6—Reduction in voltage between sheath and core with increasing distance from cable to point where current enters the ground.
1: Measured when remote ground representing cloud is at a distance of 1000 ft.
2: Calculated for uniformly conducting earth with remote ground at distance of 1000 ft.
3: Calculated for uniformly conducting earth with remote ground at infinity.

large electrode spacings. The measured variation in voltage with separation is in substantial agreement with that calculated for an earth structure of this type in the manner outlined in Section 1.9.

1.8 *Discharges Between Clouds*

In considering voltages due to discharges between clouds, the lightning channel is assumed to parallel the cable. Due to magnetic induction, the lightning current will give rise to an impressed electric force along the cable sheath. Without much error it may be assumed that there is no impressed force outside the exposed section of the sheath and that the electric force in the exposed section due to a sinusoidal current J is $E^0(x) =$

JM, where M is the unit length mutual impedance of the lightning channel and the sheath. The resulting sheath current $I(x)$ is obtained from (6) and (7), when E is replaced by E^0, Γ_0 by Γ and K_0 by K, the characteristic impedance of the sheath-earth circuit. The constants A and B are found by observing the voltage between sheath and ground is zero at $x = s/2$. The electric force along the core is given by $E(x) = RI(x)$, and the voltage between the core conductors and the sheath is obtained by a second application of (6) and (7), the constants A and B being determined from the condition that the latter voltage must equal zero at $x = s/2$. The voltage between the core conductors and the sheath at the distance x along the cable beyond one end or the other of the lightning channel projection on the cable is then:

$$U(x) = \frac{JRM\Gamma^2 s}{2Z(\Gamma^2 - \Gamma_0^2)}\left[\frac{e^{-\Gamma_0 x}}{\Gamma_0 s}(1 - e^{-\Gamma_0 s}) - \frac{e^{-\Gamma x}}{\Gamma s}(1 - e^{-\Gamma s})\right] \quad (39)$$

the sign of the voltage beyond one end of the channel being opposite to that beyond the other end.

Since $\Gamma \gg \Gamma_0$, the last bracket term may be neglected. It was shown previously, that attenuation along the core-sheath circuit within a distance of one mile, which is representative of the length s, is quite small, so that $1 - e^{-\Gamma_0 s} \cong \Gamma_0 s$. With these modifications:

$$U(x) = \frac{JRM\Gamma^2 s}{2Z(\Gamma^2 - \Gamma_0^2)} e^{-\Gamma_0 x} \quad (40)$$

The earth-return impedances M and Z are given by the following approximate expressions:[12]

$$M = \frac{i\omega\nu}{2\pi} \frac{\sqrt{2}h}{(h^2 + y^2)^{\frac{1}{2}}(i\omega\alpha)^{\frac{1}{2}}} \quad (41)$$

$$Z = \frac{i\omega\nu}{2\pi} \log \frac{\sqrt{2}}{a(i\omega\alpha)^{\frac{1}{2}}} \quad (42)$$

where α is defined as before, log = $\log_e$ and:

h = height of lightning channel above ground
y = horizontal separation of lightning channel from cable

The expression for M holds when $\alpha(h^2 + y^2)^{\frac{1}{2}} > 5$, a condition which is satisfied in the important part of the frequency range.

Inserting (41) and (42) in (40):

$$U(x) = \frac{JR\alpha^{\frac{1}{2}}}{2(\alpha - \beta)} \mu \left(\frac{1}{i\omega}\right)^{\frac{1}{2}} e^{-(i\omega\beta)^{\frac{1}{2}}x} \quad (43)$$

where

$$\mu = \frac{\sqrt{2}hs}{(h^2 + y^2) \log \frac{\sqrt{2}}{a(i\omega\alpha)^{\frac{1}{2}}}} \tag{44}$$

Comparison with (21) and (23) shows that in this case:

$$V(x, t) = \frac{R\alpha^{\frac{1}{2}}}{2(\alpha - \beta)^{\frac{1}{2}}} \mu g(\beta^{\frac{1}{2}} x, t) \tag{45}$$

In the above solution, μ was assumed constant. Actually it changes slightly with frequency and, for reasons mentioned before, it is accurate enough for practical purposes to replace $i\omega$ with $1/t$ when calculating μ.

The maximum voltage is obtained at $x = 0$, i.e., at a point opposite one end or the other of the lightning channel, and comparison with (22) shows that this voltage differs from that obtained in the case of a direct stroke by the factor μ, since $\beta^{\frac{1}{2}} \ll \alpha^{\frac{1}{2}}$ so that the second term may be neglected in (23). The above factor has the following approximate value:

$$\mu = .14 \frac{sh}{h^2 + y^2} \tag{46}$$

Since each cloud has an equal and oppositely charged image at the distance h below the surface of the ground, the electric field between cloud and ground is substantially equal to that between the clouds when $s = 2h$. For a discharge to take place between clouds, rather than to the earth, the length s would, therefore, have to be less than $2h$; so that with $y = 0$ the factor would not be expected to exceed $\mu = .28$. Thus, for the cable previously considered, failures due to discharges between clouds would not be expected except for currents in excess of 100,000 amperes in a lightning channel approximately above and parallel to the cable.

Maximum voltages of opposite signs are obtained at the two ends of the lightning channel, the voltage at the mid-point being zero. As the distance x from one end of the lightning channel increases, the voltage diminishes rather slowly in the same manner as shown in Fig. 5 for $V_0(x, t)$.

1.9 *Stratified Earth Structures*

In the foregoing, the earth was assumed to have a uniform resistivity ρ. In many cases the average resistivity near the surface along a route may be substantially greater or smaller than the resistivity at greater depths, and this condition may affect the nature of lightning troubles, as will be shown below.

The function $Q_0(r)$ appearing in equation (33) represents the earth

potential at a point due to unit current entering the earth at a distance r from the point; i.e. $Q_0(r)$ is the mutual resistance between two points on the earth's surface separated by the distance r. In the case of a two-layer horizontally stratified earth the function $Q_0(r)$ may be approximated by the following expression:

$$Q_0(r) = \frac{1}{2\pi r}\left[\rho_2 + (\rho_1 - \rho_2)e^{-\gamma_0 r}\right] \tag{47}$$

where

ρ_1 = Resistivity of upper layer, meter-ohms
ρ_2 = Resistivity of lower layer, meter-ohms
$\gamma_0 = k/2d$
d = Depth of upper layer, meters
k = Constant depending on the ratio ρ_1/ρ_2

When r is sufficiently small compared to d, the above expression approaches the limit $Q_0(r) = \rho_1/2\pi r$ and when r is sufficiently large compared to d the expression becomes $Q_0(r) = \rho_2/2\pi r$. Expression (47) gives a fair approximation to the function $Q_0(r)$ as given by curves calculated from rather complicated integrals.[13] Earth resistivity measurements by the so-called "four-electrode measurements" are based on measurements of $Q_0(r)$ and the results of such measurements may usually be approximated to the same degree of accuracy by (47) as by the curves applying accurately for two-layer earth, for the reason that the earth structure usually departs considerably from an ideal two-layer earth. For various ratios ρ_1/ρ_2 the constant k is about as follows:

ρ_1/ρ_2 =	100	10	1	.1	.02
k =	2	1.84	1.16	.4	.12

Inserting (47) in (33) and proceeding as before, the voltage between core and sheath due to a stroke at the distance y may be written:

$$V(0, y, t) = V(0, a, t)\,\frac{\rho_2\lambda(y) + (\rho_1 - \rho_2)\mu(y)}{\rho_2\lambda(a) + (\rho_1 - \rho_2)\mu(a)} \tag{48}$$

where $V(0, a, t)$ is the voltage for a direct stroke calculated for an equivalent earth-resistivity:

$$\rho_e = \rho_2\lambda(a) + (\rho_1 - \rho_2)\mu(a) \tag{49}$$

and where

$$\lambda(y) = \log\left[(1 + \Gamma y)/\Gamma y\right] \tag{50}$$

$$\mu(y) = \int_0^\infty \frac{1}{r} e^{-\alpha r} e^{-\Gamma x} dx \tag{51}$$

$$\leqq \log \frac{1 + y(\gamma_0 + \Gamma)}{y(\gamma_0 + \Gamma)} \quad \text{when} \quad \gamma_0 y \ll 1 \tag{52}$$

$$\leqq e^{-\gamma_0 y} \lambda(y) \quad \text{when} \quad \gamma_0 y \gg 1 \tag{53}$$

In applying the above expressions, a rough value of ρ_e is first assumed in calculating $\lambda(a)$ and $u(a)$ and a more accurate value next obtained from

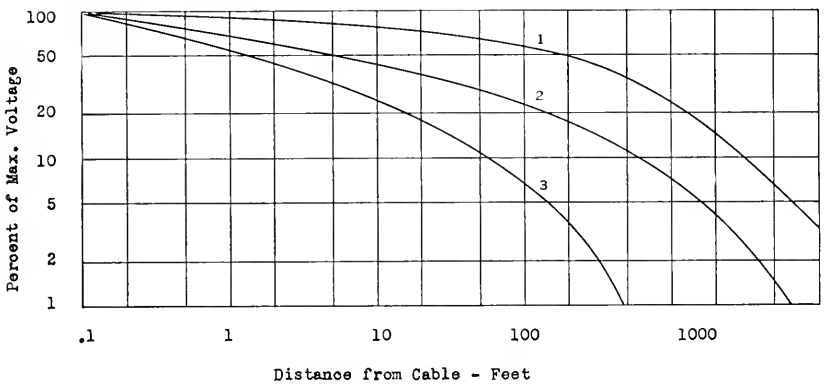


Figure 7—Reduction in voltage between sheath and core with increasing distance from cable to point where current enters ground.
1: Upper layer of 400 meter-ohms and 30 ft. depth. Lower layer of 4000 meter-ohms and infinite depth.
2: Uniformly conducting earth.
3: Upper layer of 1500 meter-ohms and 30 ft. depth.
Lower layer of 150 meter-ohms and infinite depth.

(49). If the value of ρ_e thus obtained differs materially from the assumed value, a second calculation may be required. In the expressions for λ and μ the resistivity ρ_e is to be used in calculating Γ, the latter being taken as $(\nu/2\rho_e t)^{\frac{1}{2}}$ where t is the time to crest value of the voltage as before.

In Fig. 7 is shown the manner in which the voltage decreases with increasing separation for three assumed earth structures. The resistivities and the depth of the upper layer were selected such that the equivalent earth-resistivity, and thus the voltage in the case of a direct stroke, is the same in all cases and equal to 1000 meter-ohms. It will be noticed that in the case where the resistivity of the lower layer is high, the voltage due to a stroke at a distance of 200 ft. is 50% and at a distance of 1000 ft., 25% of the voltage due to a direct stroke. When the cable is small, insulation failures may thus be occasioned by strokes to ground at considerable

distances from the cable, although the resistivity near the surface up to depths of say 50 ft. is only moderately high.

It is seen, however, that when the earth resistivity of the lower layer is low, failures due to strokes to ground not arcing to the cable are rather unlikely, even when the earth resistivity near the surface is rather high. On account of the higher surface resistivity, however, a greater number of strokes would be expected to arc to the cable for a given equivalent resistivity, than when the conductivity is uniformly distributed. On the other hand, many strokes which would arc to the cable if the earth were uniformly conducting may channel through the surface layer to the good conducting lower layer, so that the incidence of direct strokes is reduced on this account. Experience indicates that the latter factor tends to predominate, so that lightning damage is not ordinarily severe when the resistivity is low at depths beyond 20 ft. or so.

In the case of discharges between clouds the coupling between the lightning channel and the cable depends, in the frequency range of importance, to a great extent on the resistance of the lower layer. Thus, when the resistivity of the lower layer is very high the voltages may possibly give rise to insulation failures in the case of small cables, while this is not likely to occur when the resistivity of the lower layer is small or when the earth structure is uniform and of moderately high resistivity.

1.10 *Cables with Insulated Sheaths*

Assume that a short length Δx of insulated sheath is placed on the ground and that a voltage is applied between the sheath and a remote ground. When the applied voltage is greater than the breakdown voltage of the insulation, arcing to ground will take place at numerous equidistant points, provided the insulation and the earth are assumed to be uniform. The voltage between the sheath and adjacent ground increases from zero at a point where arcing takes place to a maximum value midway between two points at which arcing occurs, the maximum value being equal to the breakdown voltage of the insulation. Midway between two arcing points the potential in the earth (referred to infinity) may with negligible error be calculated as though the leakage current through the numerous arcs were uniformly distributed along the sheath. This potential in the ground would then be $\Delta I/G\Delta x$, where ΔI is the total leakage current and $1/G\Delta X$ the resistance to ground of the sheath without insulation, G being the unit length leakage conductance. Midway between the arcing points the potential of the sheath to a remote ground is then:

$$V = V_0 + \Delta I/\Delta x G \tag{54}$$

Let dI_0/dx be the leakage current through the arcs and dI_1/dx the leakage current due to capacity C between the sheath and the adjacent ground. For sinusoidal currents the following equations then hold, when Z is the unit length impedance and G the unit length leakance for a sheath in direct contact with the earth:

$$-\left(\frac{dI_0}{dx} + \frac{dI_1}{dx}\right)\frac{1}{G} + V_0 = V \tag{55}$$

$$-(I_0 + I_1)Z = \frac{dV}{dx} \tag{56}$$

$$-\frac{1}{i\omega C}\frac{dI_1}{dx} = V_0 \tag{57}$$

In the last equation it is assumed that the voltage between sheath and ground is equal to the breakdown voltage of the insulation, although this is not true in the immediate vicinity of the arcs.

Eliminating V the following equation is obtained:

$$\left(\frac{d^2 I_0}{dx^2}\frac{1}{G} - I_0 Z\right) + \left(\frac{d^2 I_1}{dx^2}\frac{1}{Y} - I_1 Z\right) = 0 \tag{58}$$

where:

$$\frac{1}{Y} = \frac{1}{G} + \frac{1}{i\omega C} \quad \text{or:} \quad Y = \frac{i\omega CG}{G + i\omega C}$$

Equation (58) is satisfied when:

$$\begin{aligned} I_0 &= A_0 e^{-\Gamma x} + B_0 e^{\Gamma x} \\ I_1 &= A_1 e^{-\Gamma_1 x} + B_1 e^{\Gamma_1 x} \end{aligned} \tag{59}$$

where Γ and Γ_1 are the propagation constants for a sheath in direct contact with the ground and for an insulated sheath without breakdown, respectively.

$$\Gamma = (GZ)^{\frac{1}{2}}, \qquad \Gamma_1 = (YZ)^{\frac{1}{2}}$$

For a sheath of infinite length the B_0 and B_1 terms vanish, so that:

$$I(x) = I_0 + I_1 = A_0 e^{-\Gamma x} + A_1 e^{-\Gamma_1 x} \tag{60}$$

The constants A_0 and A_1 are obtained from the following boundary conditions:

$$\text{At } x = 0 \qquad I(x) = I(0) = A_0 + A_1 \tag{61}$$

$$\text{As } x \to \infty \qquad I(x) \to A_1 e^{-\Gamma_1 x} = \frac{V_0}{K_1} e^{-\Gamma_1 x} \tag{62}$$

where $K_1 = (Z/Y)^{\frac{1}{2}}$ is the characteristic impedance of the insulated sheath without breakdown.

From (61) and (62)

$$A_1 = \frac{V_0}{K_1} \qquad A_3 = I(0) - \frac{V_0}{K_1} \tag{63}$$

So that:

$$I(x) = I(0)e^{-\Gamma x} + \frac{V_0}{K_1}(e^{-\Gamma_1 x} - e^{-\Gamma x}) \tag{64}$$

For a rubber insulated cable the breakdown voltage V_0 would be in the order of 30,000 volts and the characteristic impedance K_1 would be in the order of 100 ohms. The maximum current which could flow on the sheath without breakdown, V_0/K_1, is then about 300 amperes, while in the case of an average lightning stroke the current $I(0)$ would be about 15,000 amperes (i.e. 30,000 amperes total). The first term in (64) gives the attenuation along a sheath in direct contact with the ground. The current given by this term would diminish from 15,000 amperes to about 2,000 amperes within a distance of $\frac{1}{2}$ mile or so, for a typical lightning stroke wave shape and an earth resistivity of 1000 meter-ohms. At distances of several miles from the stroke point the first term will vanish and the current will be determined by the second term, since $\exp(-\Gamma_1 x)$ will vanish much more slowly than $\exp(-\Gamma x)$.

In the case of a stroke to ground the impressed electric force in the ground along the sheath is

$$E_0(x) = -dV_e(x)/dx \tag{65}$$

where V_e is the earth potential due to the lightning stroke current and is given by

$$V_e(x) = \frac{J\rho}{2\pi(x^2 + y^2)^{\frac{1}{2}}} \tag{66}$$

Instead of equation (58), the following equation is obtained for the currents in the sheath

$$\left(\frac{d^2 I_0}{dx^2}\frac{1}{G} - I_0 Z\right) + \left(\frac{d^2 I_1}{dx^2}\frac{1}{Y} - I_1 Z\right) = -E_0(x) \tag{67}$$

Writing $E_0(x) = cE_0(x) + (1 - c)E_0(x)$, the solution of the latter equation may be written as the sum of two solutions of the form given by (6) and (7). After the constants A_0, B_0 applying to the current I_0 and the constants A_1 and B_1 applying to the current I_1 have been determined from the boundary

conditions, in the same manner as before, the total sheath current may be written in the form:

$$I(x) = cI_g(x) + (1 - c)I_i(x) \tag{68}$$

where I_g is the current for a grounded sheath, I_i the current entering a perfectly insulated sheath by virtue of its capacity to ground and c is given by:

$$c = V_0/V^0 \cong V_0/V_c(0) \tag{69}$$

where V^0 is the potential difference between sheath and adjacent ground without breakdown at $x = 0$, which is substantially equal to the earth potential. The above relationship for the constant c is obtained by applying equation (57) at $x = 0$ to the general solution for I_1, V^0 being given by:

$$V^0 = \frac{G}{G + i\omega c} \int_0^\infty E_0(x)e^{-\Gamma_1 x}\, dx \tag{70}$$

The voltage between core and sheath of an insulated cable may be written in a similar manner when I_g is replaced by V_g, the voltage for a cable in direct contact with the ground, and I_i replaced by V_i, the voltage for a cable insulated from ground.

From (68) and (69) it will be seen that when $V_c(0)$ is much greater than the breakdown voltage of the insulation, the current entering the sheath is nearly the same as for a sheath in direct contact with the ground. Thus, when the earth resistivity is 1000 meter-ohms, and the stroke current 30,000 amperes, the earth-potential at a distance of 30 meters (100 feet) is 160,000 volts. The impulse breakdown of the insulation may be in the order of 30,000 volts, so that the current entering the sheath will be substantially the same as for a cable in direct contact with the soil. When the earth-potential at $x = 0$ is only slightly larger than the breakdown voltage of the insulation, however, the current entering the sheath through punctures in the insulation is fairly small.

In the above derivation the voltages and currents were assumed to vary sinusoidally which, of course, is a rather rough approximation in a phenomenon where breakdown occurs after the voltage reaches a certain instantaneous value. While the derivation is not accurate, it does indicate under what conditions an insulated cable behaves like a cable in direct contact with the ground.

1.11 *Oscillographic Observations of Lightning Voltages*

To obtain data on the characteristics of lightning voltages in buried cable, five magnetic string oscillographs were installed for one lightning

season along a 50-mile section of the Stevens Point-Minneapolis route. The oscillographs, which were arranged to trip where the voltage exceeded 100 volts, recorded lightning voltages due to some 600 strokes on 38 days, the Weather Bureau average being 33 thunderstorm days for this region in the same months. The character of the voltages varied widely from sharp transients of a few millisecond duration to slowly changing voltages lasting .2 seconds from one zero value to the next, voltages due to multiple discharges being quite common, the interval between voltage peaks in such cases being in the order of .1 second. Of the disturbances, 90% lasted for more than .1 second, 50% for more than .4 and 10% for more than 1.25 second, the maximum duration being 2.3 seconds. By way of comparison, the observed duration of discharges to a tall structure (3) were, in respectively 90, 50 and 10% of all cases, in excess of .08, .3 and .6 second, the maximum duration being 1.5 second. The maximum voltage recorded was 940 volts and was probably due to a stroke to ground near the cable. About 2% of the voltages were in excess of 500 volts, most of these and the lower voltages being due to discharges between clouds, as indicated by the opposite polarity of the voltages at the two ends of the test section. The wave shape of the voltages at the ends of the section were much the same as at intermediate points, even for the sharpest surges recorded, the attenuation along the core-sheath circuit being quite small. It is possible that substantially higher voltages than observed may obtain in the case of severe discharges along a path parallel to and directly above the cable, and that such voltages may produce cable failure if the core insulation is below normal.

While the oscillographs were arranged to trip on 100 volts, a smaller voltage was recorded in 40% of all cases, as the peaks were too fast to be recorded by the type of oscillograph used. It is also possible that for this reason fast voltage peaks in excess of the maximum given above may have escaped measurement.

II. Lightning Trouble Expectancy

2.1 *General*

In estimating the liability of a cable to lightning damage, it is assumed below that once the core insulation is punctured, as it is likely to be at several points, at least one permanent failure will occur. The lightning trouble expectancy curves presented here thus give the number of times lightning damage is likely to occur, without consideration of the extent of the damage on each occasion. Each case of lightning damage usually involves several pairs and, based on experience, repair of each such case would require about four sheath openings. Damage due both to direct strokes and strokes to ground is included. Discharges between clouds have been neglected as a

source of lightning damage, however, as the voltages are likely to be insufficient unless the insulation is below normal.

The curves of lightning trouble expectancy calculated here are significant only if troubles on a long cable route are considered over a period of several years, so that the mile-years covered are in the order of 1000 or more.

2.2 *Incidence of Strokes to Ground*

To estimate the lightning trouble expectancy it is necessary to consider the incidence of strokes to ground in the vicinity of the cable, the number

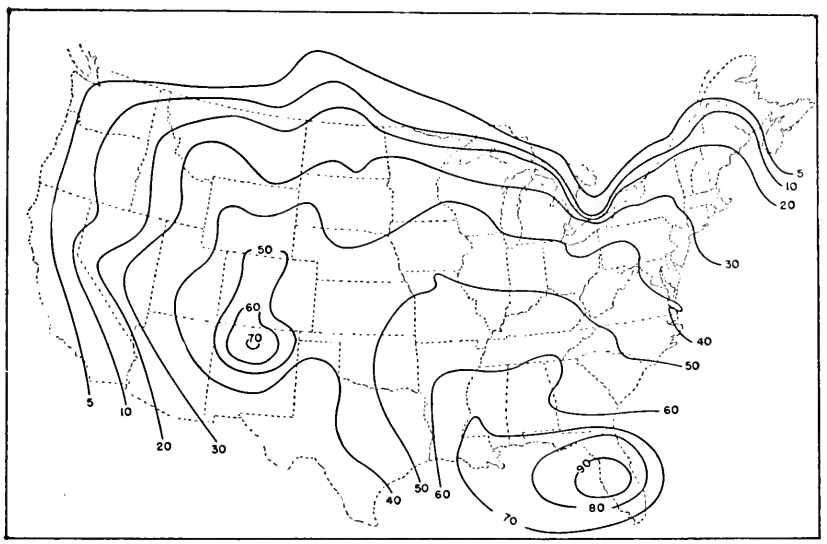


Figure 8—Map showing the average number of thunderstorm days per year.

of such strokes that will arc to the cable and cause damage in this manner and the number that will give rise to failure without arcing to the cable.

Magnetic link measurements (14) indicate that high tension transmission lines will be struck by lightning about 113 times per 100 miles per year, on the average, the minimum incidence in one year being about one half and the maximum about 1.6 times the average value. The above average incidence is based on observations covering about 1600 mile-years and applies for lines traversing areas where some 35 thunderstorm days are expected per year, as indicated by data issued by the U. S. Weather Bureau and collected over a period of 30 years,[15] and shown in Fig. 8. The above data on strokes to transmission lines may be used to estimate the rate of lightning strokes to ground, provided the width of the zone within which a transmission line will attract lightning can be determined.

Based on laboratory observations on small-scale models[16, 17, 18, 19], a line above ground will attract lightning strokes within an average distance on each side of the line which is about 3.5 times the height of the line when the cloud is positive and about 5.5 times the height when the cloud is negative. When the average height of a transmission line ground structure above ground is taken as 70 feet, a 100-mile line will thus attract positive strokes (i.e., strokes originating from a positive cloud) within an area of about 9.3 square miles and negative strokes within an area of approximately 14.5 square miles.

About 15% of the strokes to transmission line ground structures have positive polarity,[4, 5] so that the average rate of positive strokes to ground would be about 1.8 and that of negative strokes about 6.6 per square mile per year. The rate of positive and negative strokes to ground would thus be about 8.4 per square mile per year in areas where the yearly number of thunderstorm days is about 35, corresponding to about 2.4 strokes per square mile per 10 thunderstorm days.

Based on the above date, the ratio of negative to positive strokes to ground in open country would be 3.6. The ratio of negative to positive strokes has been determined by various investigators in different ways.[7] The ratio derived from measurement of field changes during thunderstorms varies between 2.1 and 6.5, that obtained from voltages observed in antennas is about 2.9, while point discharge recorder measurements indicate a ratio of 3.5 and the magnetization of basalt rocks struck by lightning indicates a ratio of 2.25. The above data were obtained in the temperate zone; in the tropics nearly all strokes have negative polarity.

2.3 *Arcing to Cable of Strokes to Ground*

As the stepped leader of a lightning discharge approaches the earth, charges accumulate in the ground under the leader and the resulting flow of current in the ground will give rise to a potential difference between points in the ground under the leader and remote points. Thus, when the tip of the leader has approached within 10 meters off the ground and the leader current is assumed to be as high as 500 amperes and the earth resistivity to be 1000 meter-ohms, a point directly under the leader will have a potential of 8000 volts with respect to a remote ground. The potential gradient along the surface of the ground would, of course, be affected by the presence of a buried cable. The total potential involved is, however, so small that the effect of a buried cable on the path of the leader would be entirely negligible compared to the effect of irregularities in the surface of the earth. As the tip of the leader contacts the ground, the potential may be large enough so that the leader may arc to a cable located within 2 ft. or so of the leader. Only

after the return stroke is initiated, however, is the potential large enough so that arcing will occur for appreciable distances.

When a stroke current J enters the ground, the electric force in the ground at the distance r from the stroke point is given by

$$E(r) = J\rho/2\pi r^2 \tag{71}$$

ρ being the earth resistivity. If the breakdown voltage gradient of the soil is e_0, breakdown of the soil will take place until $E(r) = e_0$, or for a distance:

$$r_0 = (J\rho/2\pi e_0)^{\frac{1}{2}} \tag{72}$$

For a distance r_0 around the stroke point the soil may then be regarded as a conductor of negligible resistivity.

The resistance encountered by the lightning channel in the ground is then

$$R_0 = \frac{\rho}{2\pi r_0} = \left(\frac{\rho e_0}{2\pi J}\right)^{\frac{1}{2}} \tag{73}$$

Measurements of the surge characteristics of grounds of fairly small dimensions[20, 21] indicate that the breakdown voltage of the soil may vary from roughly 1000 volts per cm to some 5000 volts/cm. The data are, however, quite limited and there is no assurance that the above values represent the limits. In the first of the papers referred to, measurement was made of the resistance encountered when a current of 10,000 amperes crest value was discharged into the ground from an electrode suspended above the ground. The measured resistance is in satisfactory agreement with that obtained from (73) using the earth resistivity and breakdown voltage determined from other measurements in the paper referred to. In connection with the present study some small scale measurements were made of the breakdown voltage of sand between plane electrodes, and of the resistance encountered in discharges into the sand over point electrodes. These measurements indicated a breakdown voltage of 5000 volts per cm for dry sand having a resistivity of 3700 meter-ohms and 2400 volts/cm when sufficient water was added to reduce the resistivity to 100 meter-ohms. The measured resistance of point electrodes was a satisfactory agreement with that calculated from (73) on the basis of the measured resistivities and breakdown voltages. These experiments were made with currents having crest values from 1 to 50 amperes.

Measurements of the breakdown voltage of various types of soil between completely buried spherical electrodes indicate substantially higher voltages than those given above, from about 11 to 23 kv/cm.[22] It is possible that for surface electrodes, breakdown at the lower voltage gradients occurs along the surface of the ground, so as to form a conducting plane of radius

r_0. This circumstance would change the preceding formulas only to the extent that 2π would be replaced by 4. For a given e_0, the resistance would then be about 25% higher.

The radius r_0 is not necessarily the same as the distance across which a lightning stroke would arc to a cable in the vicinity. Streamers will extend in various directions beyond r_0 so that ionization of the soil increases the conductivity for a greater distance, r_0 being the effective radius of an equivalent hemisphere of infinite conductivity.

The potential difference between the conducting sphere of radius r_0 and a point at the distance r_1 is

$$V_{01} = \frac{J\rho}{2\pi}\left(\frac{1}{r_0} - \frac{1}{r_1}\right) = \frac{J\rho}{2\pi}\frac{r_1 - r_0}{r_1 r_0} \tag{74}$$

Let it be assumed that streamers extend beyond r_0 until the average voltage gradient between r_0 and r_1 equals e_1. The potential difference is then:

$$V_{01} = e_1(r_1 - r_0) \tag{75}$$

From (74) and (75):

$$r_1 = \frac{J\rho}{2\pi r_0 e_1} = r_0\frac{e_0}{e_1} \tag{76}$$

The latter expression applies when the field is assumed to have a radial symmetry about the lightning channel. When a buried cable is present, however, this symmetry is disturbed. Calculations indicate that the potential of the cable at the point nearest to the lightning stroke will be less than 20% of the earth potential at the same point if the cable were absent. As a first approximation the cable may, therefore, be considered to have zero potential. The potential difference between the sphere considered above and the cable is then:

$$V_{02} = \frac{J\rho}{2\pi}\frac{1}{r_0} \tag{77}$$

If r_2 is the distance from the channel to the cable, the potential difference is also given by:

$$V_{02} = e_1(r_2 - r_0) \tag{78}$$

From (77) and (78):

$$r_2 = r_0\left(1 + \frac{e_0}{e_1}\right) \tag{79}$$

The arcing distance calculated in this manner will be the maximum, while that calculated from (76) will be the minimum distance.

From measurements made of the effective corona radius of a conductor in air,[22] it is found that the latter may be determined on the assumption that

breakdown occurs until $e_0 \leqq 14000$ volts/cm. On the other hand it is known that arcing between two conductors a considerable distance apart will occur when the average gradient is about 10,000 volts/cm. For air the ratio e_0/e_1 is thus about 1.4. For breakdown in the air, the ionization need not be very dense in order that the corona envelope may be regarded as conducting as regards displacement currents. For breakdown in the earth, however, the ionization must be much more complete in order to provide substantially better conductivity than the soil. The ratio e_0/e_1 is, therefore, likely to be greater, and has been taken as 2 in the following.

If the latter ratio is assumed for soil, it is seen by comparison of (76) and (79) that the presence of a cable may increase the arcing distance as much as 50%.

The arcing distance may be written in the form

$$r_1 = (J\rho)^{\frac{1}{2}} q_1, \qquad r_2 = (J\rho)^{\frac{1}{2}} q_2 \tag{80}$$

where:

$$q_1 = \left(\frac{1}{2\pi e_0}\right)^{\frac{1}{2}} \frac{e_0}{e_1}, \qquad q_2 = \left(\frac{1}{2\pi e_0}\right)^{\frac{1}{2}} \left(1 + \frac{e_0}{e_1}\right) \tag{81}$$

With $e_0/e_1 = 2$ the following values are obtained

$e_0 =$	250,000	500,000 volts/meter
$q_1 =$	$1.6 \cdot 10^{-3}$	$1.13 \cdot 10^{-3}$
$q_2 =$	$2.4 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$

Values of q_1 and q_2 about 25% greater than those given above are obtained by assuming that breakdown does not take place in the soil, but that a conducting plane of radius r_0 is formed by ionization of the air near the ground. Expression (71) is in this case replaced by $E(r) = J\rho/4r^2$ and the expressions for q become

$$\begin{aligned} q_1 &= \left(\frac{1}{4e_0}\right)^{\frac{1}{2}} \frac{e_0}{e_1} \\ q_2 &= \left(\frac{1}{4e_0}\right)^{\frac{1}{2}} \left(1 + \frac{e_0}{e_1}\right) \end{aligned} \tag{82}$$

In the following $q = 2.5 \cdot 10^{-3}$ has been taken as representative for low-resistivity soil and $q = 1.5 \cdot 10^{-3}$ for high-resistivity soil. When the current is expressed in kiloamperes, the corresponding values are .08 and .047.

The distance to which a stroke may arc is accordingly taken as:

$\rho \leq 100$ meter-ohms	$\rho \geq 1000$ meter-ohms
$r = .08\ (J\rho)^{\frac{1}{2}}$ meter	$r = .047\ (J\rho)^{\frac{1}{2}}$ meter
$= .26\ (J\rho)^{\frac{1}{2}}$ feet	$= .15\ \ (J\rho)^{\frac{1}{2}}$ feet

where J is in kiloamperes.

These values are, of course, of an approximate nature, and are only indicative of what may be expected under average conditions. In some cases the breakdown voltage of high-resistivity soil may be substantially lower than assumed, while that of low-resistivity soil may be noticeably higher.

2.4 *Crest Current Distribution for Strokes to Ground*

When the earth resistivity is taken as high as 5000 meter-ohms and the breakdown voltage of the soil is taken as high as 5000 volts/cm, the resistance encountered by the channel on the ground for a current of 25,000 amperes is about 250 ohms. If the lightning channel were a long conductor already in existence at the initiation of the return stroke and capable of carrying the stroke current without being fused, the current would be propagated upward with the velocity of light and the surge impedance of the channel would be in the order of 500 ohms. Due to the resistance in the ground the current would then be some 30% smaller than for a stroke to an object of zero resistance to ground. However, the lightning channel may not be regarded in the above manner, but as a conductor which is gradually prolonged at about 1/10 the velocity of light, and the impedance of the channel is then much larger, perhaps 5000 ohms. The surge impedance of a long insulated conductor having unit length capacitance C is $(1/Cv)$, v being the velocity of propagation. When energy is required to create the conductor, so that the velocity of propagation is reduced, the impedance is increased. Because of the high impedance of the channel, the resistance encountered in the ground may, therefore, be neglected as regards the effect on the crest current. The crest current distribution curve for strokes to transmission line ground structures may thus be used also in the case of strokes to ground, although a different distribution curve is obtained for those of the strokes to ground which are to buried cable (Section 2.7).

2.5 *Failures Due to Direct Strokes and Strokes to Ground*

In calculating the number of failures due to direct strokes and strokes to ground, the earth is assumed to be a plane surface. A tree placed at random may attract a lightning stroke toward a cable or it may divert it from the cable and the net effect of a large number of trees along a route of substantial length is likely to be small. This is also true for variations in the terrain.

When N is the number of lightning strokes to ground per unit of area, and s the length of the cable, the number of lightning strokes on both sides of the cable within y and $y + dy$ is:

$$dN = 2Ns\,dy \tag{83}$$

A lightning stroke at the distance y will cause cable failure when the crest current i exceeds a certain value which depends on the distance:

$$i = f(y) \tag{84}$$

The fraction of all lightning strokes which has a crest current larger than i will be designated $P_0(i)$. The fraction of the lightning strokes dN which will cause cable failures is then:

$$dn = dNP_0(i) = 2NsP_0(i)\,dy \tag{85}$$

The number of cable failures along the length s due to all lightning strokes to ground up to the maximum distance Y that need to be considered is:

$$n = 2Ns \int_0^Y P_0(i)\,dy \tag{86}$$

For the purpose of computation it is convenient to change the variable in the latter integral from y to i. With $y = f^{-1}(i) = y(i)$, $di = dy \cdot f^1(y)$, $i_0 = f(0)$ and $I = f(Y)$, the following integral is obtained:

$$n = 2Ns\left[YP_0(i) - \int_{i_0}^{I} y(i)P_0'(i)\,di\right] \tag{87}$$

In (87), I is the maximum stroke current that needs to be considered and may actually be replaced by infinity, as will be evident later on. The current i_0, which is the minimum current that will cause insulation puncture in the case of a direct stroke, may readily be determined from the breakdown voltage of the insulation and the calculated voltage between core and sheath per kiloampere, in the manner illustrated in Section 2.4.

In order to evaluate the integral of (87), it is divided as follows:

$$n = 2Ns\left[YP_0(I) - \int_{i_0}^{i_1} y_1(i)P_0'(i)\,di - \int_{i_1}^{I} y_2(i)P_0'(i)\,di\right] \tag{88}$$

When $i < i_1$, failures of the cables will be due to arcing of the stroke to the cable and when $i > i_1$, failures will occur before arcing takes place, due to the leakage current entering the sheath. Within each of the above two ranges the relationship of y to i is different and is designated $y_1(i)$ and $y_2(i)$, respectively.

As already shown, failures due to arcing will take place when:

$$y_1(i) \leq q(\rho i)^{\frac{1}{2}} \tag{89}$$

where q is defined as before and ρ is the earth resistivity in meter-ohms.

In Section 1.7 it was shown that failures due to leakage current will occur when

$$i \geq i_0/\lambda(y)$$

where

$$\lambda(y) = \log\left(\frac{1 + \Gamma y}{\Gamma y}\right) \Big/ \log\,(1/\Gamma a)$$

Γ being the propagation constant of the sheath-earth circuit and a the radius of the sheath. The solution of the latter equation for y is:

$$y = y_2(i) = \frac{1}{\Gamma(e^{mi_0/i} - 1)} \cong \left(\frac{2\rho t}{\nu}\right)^{\frac{1}{2}} \frac{1}{e^{mi_0/i} - 1} \tag{90}$$

where

$m = \log\,(1/a\Gamma)$
$\Gamma \cong (\nu/2\rho t)^{\frac{1}{2}}$ per meter
$\nu = 1.256 \cdot 10^{-6}$ henries per meter
$t \cong 10^{-4}$ sec. $\cong$ time to crest of core-sheath voltage.

When (89) and (90) are equated, the following expression is obtained:

$$i^{\frac{1}{2}}(e^{mi_0/i} - 1) = \left(\frac{2t}{\nu q^2}\right)^{\frac{1}{2}} \tag{91}$$

The value of i which satisfies the latter equation is the value i_1 defined above, and is shown in Fig. 9 as a function of mi_0 for various values of $(2t/\nu q^2)^{\frac{1}{2}}$.

When (89) and (90) are inserted in (88), the latter integral may be expressed as follows:

$$n = 2Ns\rho^{\frac{1}{2}}\left(q[H(i_0) - H(i_1)] + \left(\frac{2t}{\nu}\right)^{\frac{1}{2}} G(i_1, mi_0)\right) \tag{92}$$

where the current is in kiloamperes and:

N = Number of strokes to ground per square meter
s = Length of cable in meters
$q \cong .08$ when $\rho \leq 100$, $q \cong .047$ when $\rho \geq 1000$.

$$H(i) = -\int_i^I i^{\frac{1}{2}} P_0'(1)\, di \tag{93}$$

$$G(i, mi_0) = \frac{P_0(I)}{e^{mi_0/I} - 1} - \int_i^I \frac{P_0'(i)\, di}{e^{mi_0/i} - 1} \tag{94}$$

The first term in (94) equals $\Gamma P_0(I) = y_2(I)P_0(I)$. Since $P_0'(i)$ is negative, the above integrals will have positive values.

The term $q[H(i_0) - H(i_1)]$ of (92) gives the portion of failures due to direct strokes while the term involving the function G gives the portion of failures due to ground strokes not necessarily arcing to the cable although

they may do so (i.e. many of the currents in excess of i_1 may arc to the cable, although this is not essential in order to produce cable failure).

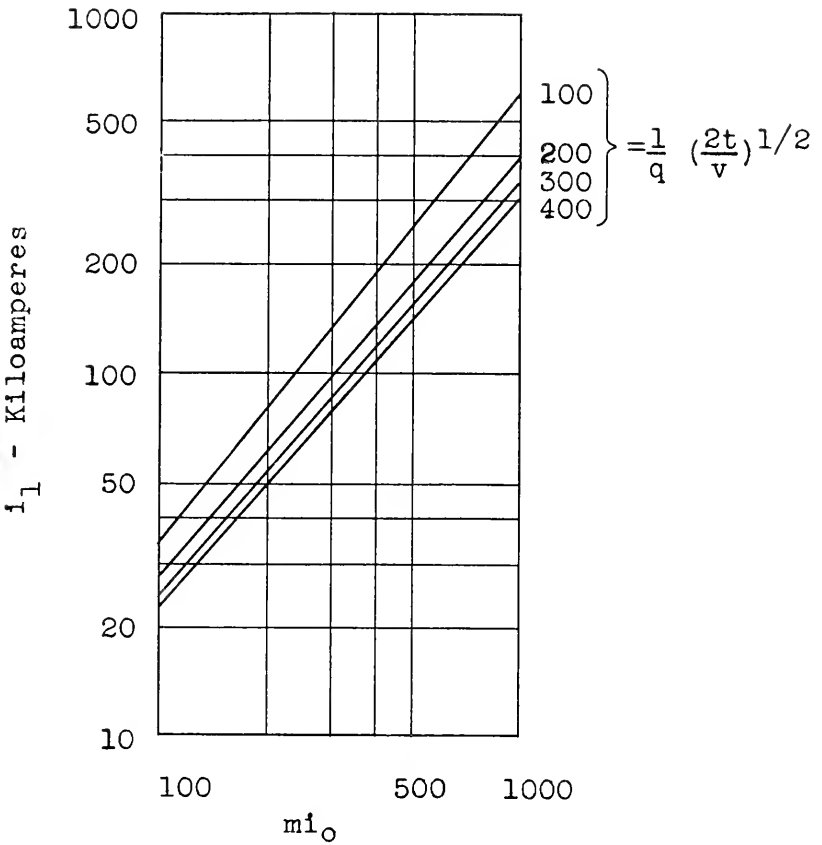


Figure 9—Solution of the equation:

$$i_1^{\frac{1}{2}}(e^{mi_0/i_1} - 1) = \frac{1}{q}\left(\frac{2t}{\nu}\right)^{\frac{1}{2}}$$

If strokes to ground not arcing to the cable were neglected as a source of failures, the number of failures would equal:

$$n_d = 2Ns\rho^{\frac{1}{2}}qH(i_0) \tag{95}$$

If, on the other hand, the dielectric strength of the earth were assumed to be infinite, so that none of the strokes to ground would arc to the cable, the number of failures would equal

$$n_g = 2Ns\rho^{\frac{1}{2}}\left(\frac{2t}{\nu}\right)^{\frac{1}{2}} G(i_1, mi_0) \tag{96}$$

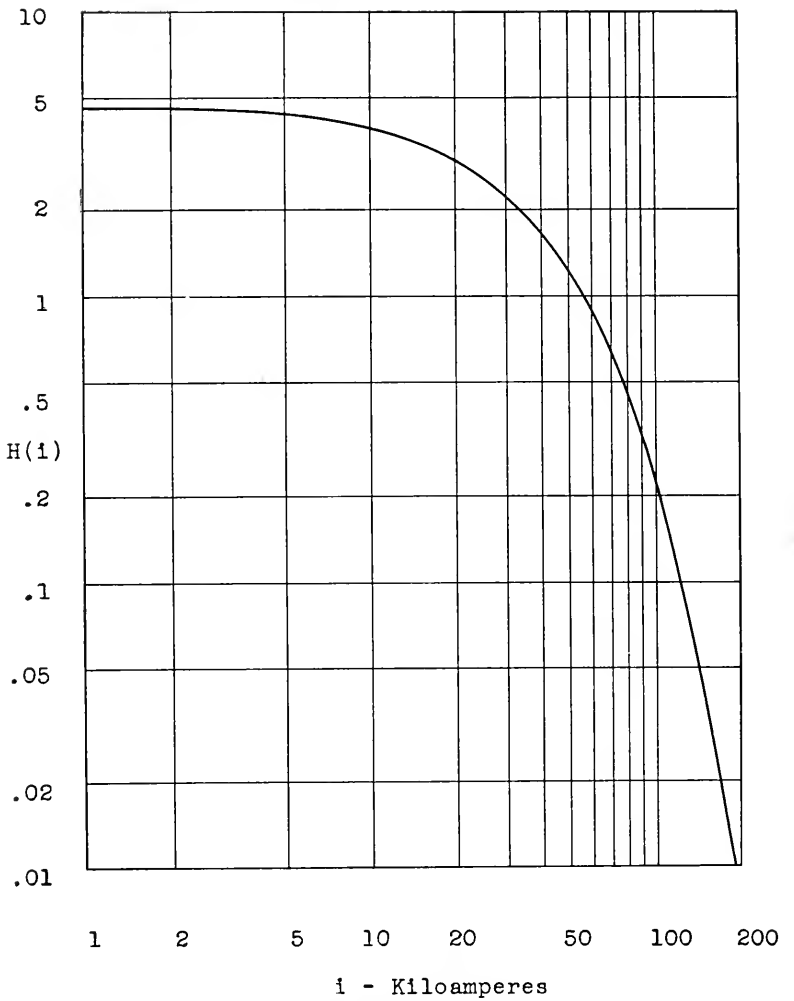


Figure 10—Function $H(i)$ when $P_0(i)$ is approximated by:
$P_0(i) = \exp(-ki)$, $k = .038$ per kiloampere.
$H(i) = k^{-\frac{1}{2}}[(ki)^{\frac{1}{2}}e^{-ki} + \frac{1}{2}\pi^{\frac{1}{2}}\operatorname{erfc}(ki)^{\frac{1}{2}}]$

From Fig. 1, it is seen that $P_0(i)$, as represented by curve 1, is nearly a straight line on semi-log paper and may, therefore, be approximated by:

$$P_0(i) \backsimeq e^{-ki} \tag{97}$$

With $k = .038$ per kiloampere, a straight line is obtained which coincides with curve 1 at $i = 0$ and $i = 100$ kiloamperes, and this value of k has been used in the following. The functions H and G obtained with this approximation are shown in Figs. 10 and 11. In obtaining these integrals, the up-

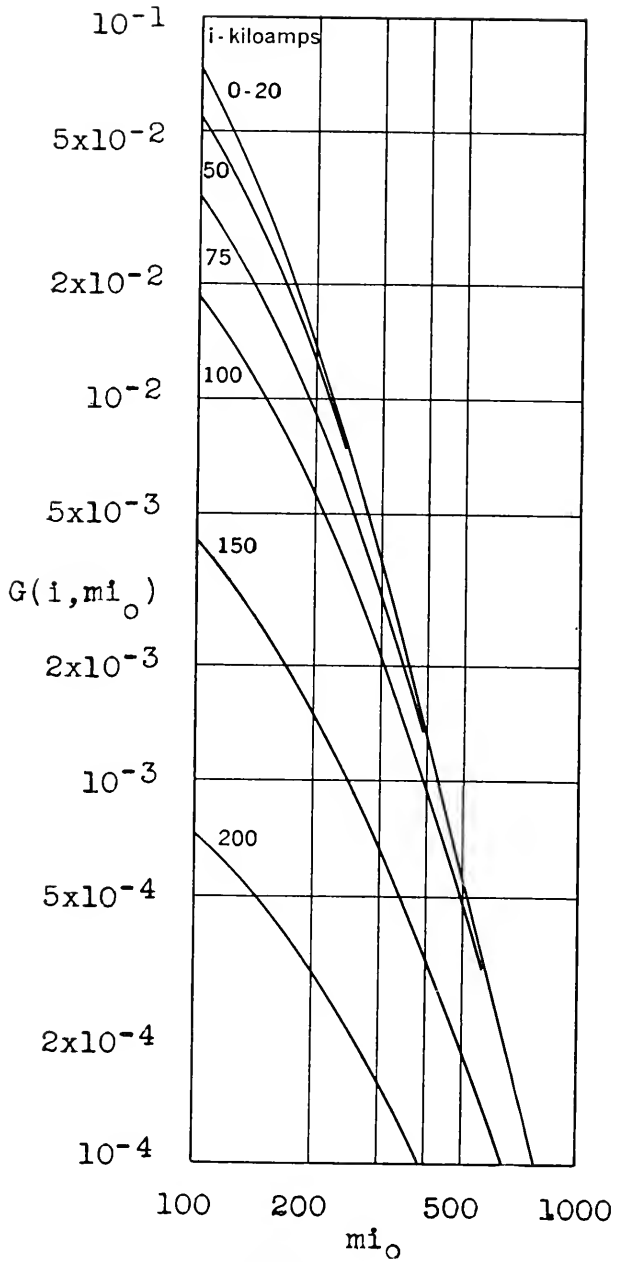


Figure 11—Function $G(i, mi_0)$ when $P_0(i)$ is approximated by: $P_0(i) = \exp(-ki)$; $k = .038$ per kiloampere.

$$G(i, mi_0) = k \int_i^\infty \frac{e^{-ki}\, di}{e^{mi_0/i} - 1}$$

per limit I may be replaced by infinity, and it is also seen that the first term in (94) then vanishes.

The lightning trouble expectancy as calculated from (92) is shown in Fig. 12 as a function of the earth resistivity for various sheath resistances. The curves are based on 2.4 strokes to ground per square mile, which is approximately the number of strokes per square mile during 10 thunderstorm days. The number of thunderstorm days per year along a given route is obtained from Fig. 8 and thus the number of times lightning failures would be expected during one year.

2.6 *Expectancy of Direct Strokes*

The incidence of direct strokes to the cable may be obtained from (96) with $i_0 = 0$ kiloamperes.

The number of strokes arcing to the cable is thus

$$n_a = 2Ns\rho^{\frac{1}{2}}qH(0) \tag{98}$$

The cable will thus attract strokes within an effective distance.

$$y = \rho^{\frac{1}{2}}qH(0) \tag{99}$$

$\rho \leq 100$ meter-ohms	$\rho \geq 1000$ meter-ohms
$y = .365\ \rho^{\frac{1}{2}}$ meters	$y = .22\ \rho^{\frac{1}{2}}$ meters
$= 1.2\ \rho^{\frac{1}{2}}$ feet	$= .7\ \rho^{\frac{1}{2}}$ feet

2.7 *Crest Current Distribution for Direct Strokes*

The fraction of the strokes to the cable having crest values in excess of i is given by:

$$\begin{aligned} P(i) &= H(i)/H(0) \\ &= 2\left(\frac{ik}{\pi}\right)^{\frac{1}{2}} e^{-ik} + \operatorname{erfc}(ik)^{\frac{1}{2}} \end{aligned} \tag{100}$$

and is shown by curve 2 in Fig. 1. It will be noticed that a buried cable attracts a greater proportion of heavy currents than a transmission line, because of the circumstance that heavy currents to ground arc for greater distances.

2.8 *Lightning Trouble Experience*

As mentioned before, lightning damage may be due to denting or to fusing of holes in the sheath, or to excessive voltages between the sheath and the cable conductors. Only the latter form of lightning failures have been considered here, since they predominate for cable of the size now being used, particularly in high-resistivity areas, and are likely to extend for a considerable distance to both sides of the point struck by lightning and are

thus more difficult to repair. For full-size cable in low-resistivity areas, however, insulation failures are more likely to occur as a result of sheath denting. For instance, along the 300-mile Kansas City-Dallas full-size cable route, where the earth resistivity is in the order of 100 meter-ohms, and where there are some 50 thunderstorm days per year, failures over a period of about 15 years have occurred about .5 times per 100 miles per year. Of these troubles 85% were due to sheath denting as a result of arcing between the tape armor and the sheath. Based on (99), 100 miles of cable would attract lightning strokes within an area of .5 square mile, so that the cable would be struck about six times per 100 miles per year, when the number of strokes per square mile per year is 2.4 per 10 thunderstorm days (Section 2.2). The rate of lightning failures experienced on this route may thus be accounted for by assuming that about 7% of the strokes, i.e. currents in excess of 90 kiloamperes as obtained from curve 2, Fig. 1, will produce sheath denting severe enough to cause insulation failure, while about 1%, i.e. currents in excess of 140 kiloamperes, will cause insulation failure due to excessive voltage. The latter value is in substantial agreement with that calculated for a full-size cable when the earth resistivity is assumed to be 100 meter-ohms. It is evident from the above examples that in low-resistivity areas lightning troubles will not be a problem, and this is also borne out by experience on other routes installed in such territory during the last few years.

All cable installed in high-resistivity territory since 1942 has been provided with doubled core insulation and with shield wires, in spite of which considerable damage has been experienced on some routes, as between Atlanta and Macon. This appears to have been due partly to the circumstance that in many cases the insulation in splices and accessories has not been equal to that obtained in the cable through the use of extra core wrap, and that in some cases damage has been due to holes fused in the sheath due to arcing between the sheath and the shield wires. As an example, along the Atlanta-Macon route there are some 70 thunderstorm days per year and the average effective earth resistivity is about 1300 meter-ohms. The corresponding estimated rate of direct strokes to the cable is about 20 per 100 miles per year. It is estimated, in the manner outlined in section 3.0, that only stroke currents in excess of 80 kiloamperes are likely to damage the cable, so that on the basis of curve 2, Fig. 1, cable failures would be expected to occur about 2 times per 100 miles per year. The actual rate of trouble experienced on this route over two years has been about five times higher, so that some 50% of the strokes to the cable; i.e., currents in excess of 30 kiloamperes or so, appear to have caused cable failures, most of which occurred in splices and accessories.

It is evident from the above examples that careful examinations of trouble records are required before the observed rate of lightning failures can be adequately compared with that obtained from theoretical expectancy curves. If the cable as well as splices and accessories actually have a dielectric strength as assumed in the calculations, it is likely that the average rate of failures due to excessive voltages experienced over a long period will not be any greater than estimated from these curves.

Based on experience, an average of 4 sheath openings is required to repair damage caused by excessive voltage between the sheath and the cable conductors, as compared to about 2 sheath openings when the damage is due mainly to denting and fusing of the sheath, as in the case of full-size, tape-armored cable in low-resistivity territory. Although the damage may be confined to one point, it cannot usually be located by a single sheath opening.

III. Remedial Measures

3.1 *General*

From Fig. 12 it is evident that the rate of cable failures to be expected, and hence the need for remedial measures, depend greatly on the earth resistivity. Experience has indicated that lightning damage is likely to be encountered even when the surface resistivity is fairly low, provided the resistivity beyond depths of 10 or 20 ft. or so is very high. Considerably less trouble has been experienced where the resistivity below this depth is low, even where the surface resistivity has been high. The lightning stroke may then channel through the surface layer to the good conducting lower layer, so that direct strokes are not experienced as frequently in spite of the high surface resistivity. As a guide in applying protective measures, earth resistivity measurements are usually made along new cable routes.[23]

The curves given in Fig. 12 may also be used to find the lightning trouble expectancy when extra core insulation, shield wires or both are used. Thus when the insulation strength is doubled the effect is the same as if the sheath resistance is halved. If the shield wires reduce the voltage by a shield factor η, the effect is the same as if the sheath resistance is multiplied by η. Considering direct strokes only, curve 2 in Fig. 1 may be used to find the percentage reduction in lightning strokes that will damage the cable, when the stroke current which the cable is able to withstand is increased by extra insulation or shield wires.

3.2 *Extra Core Insulation*

One method of reducing failures caused by lightning strokes to buried cables is to increase the insulation between the cable conductors and the

sheath, no extra insulation being required between individual cable conductors. This has already been done for most new installations. The cable itself, cable stubs, loading cases, and gas alarm contactor terminals are all provided with sufficient extra insulation to double the dielectric strength between cable conductors and sheath. For a cable like that on which the measurements referred to before were made, such a measure would increase the stroke current which would damage the cable from 30,000 to 60,0000 amperes and would reduce the number of direct lightning strokes that could cause failure by direct arcing to the sheath to about 20

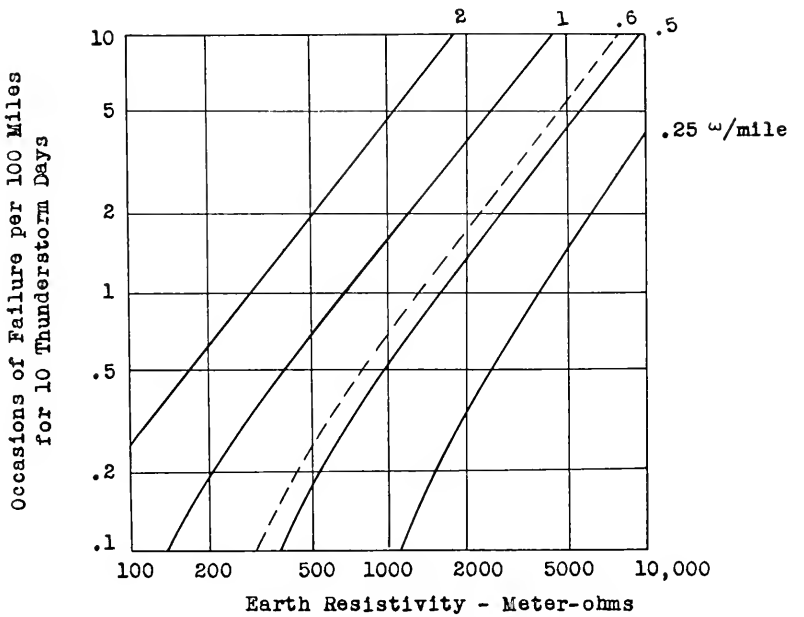


Figure 12—Theoretical lightning trouble expectancy curves showing number of times insulation failures due to excessive voltages would be expected per 100 miles for 10 thunderstorm days, for cables having sheath resistances as indicated on curves. Dashed line represents full-size cable.

per cent of the total instead of 50 per cent (see Fig. 1, curve 2). The number of failures due to direct strokes would thus have been reduced 2.5 times.

3.3 *Shield Wires*

Another method, employed in addition to the extra insulation where excessive lightning damage would otherwise be expected, is to bury shield wires over the cable. These conduct away part of the lightning current and thus reduce the amount that flows along the sheath. These wires may be plowed in with the cable, as has been done on several new routes, or may be installed afterward. When the wires and cables are plowed in

together, the arrangement shown in Fig. 13 has been used. The percentage of the current carried by the wires depends to a greater extent on their inductance relative to that of the sheath, than on their resistance. Two wires are employed, rather than a single wire of smaller resistance, in order to obtain a lower inductance than with a single wire.

On the route between Stevens Point and Minneapolis, where the shield wires were installed after the cable was in place, two 165-mil wires about twelve inches apart were plowed in some ten inches above the cable for a distance of eighty miles. Surge measurements made after these wires were installed indicated that the wires reduced the voltage between sheath and core conductors about 60 per cent, in substantial agreement with theoreti-

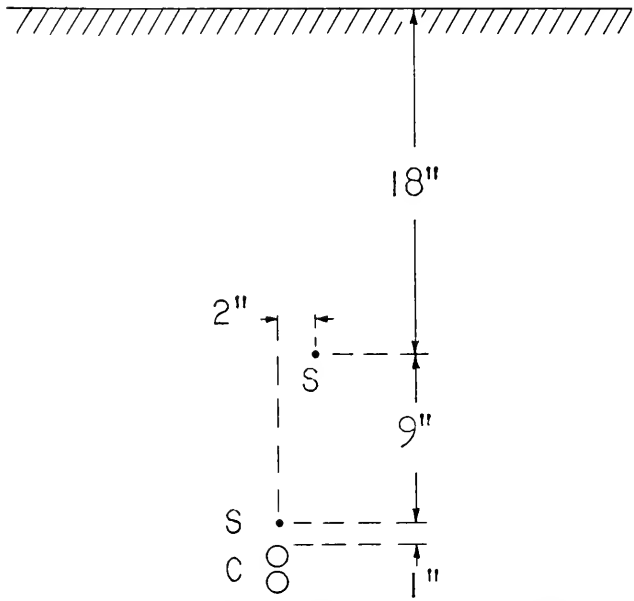


Figure 13—Position of shield wires, *S*, when they are plowed in with cable, *C*.

cal expectations. The cable would then withstand 75,000 amperes rather than 30,000 without shield wires. If a cable of normal construction were used on the above route, the shield wires would thus reduce the number of direct lightning strokes that would be expected to cause insulation failure about 3.3 times, from 50 per cent of the total without shield wires to about 15 per cent with shield wires (see Fig. 1, curve 2). If the insulation strength is only 1000 volts, however, about 80 per cent of all strokes would be expected to cause failure without shield wires and 40 per cent with shield wires, so that the reduction would be substantially smaller, as actually appears to be the case on the above route.

Aside from reducing core insulation failures, shield wires also minimize

damage to the sheath covering in the case of strokes to ground near the cable, particularly in the case of thermoplastic or rubber-covered cables. As mentioned in section 1.10, considerable current may enter the sheath of an insulated cable in the case of strokes to ground near the cable, due to numerous punctures in the insulation of the sheath. If conditions along the cable were uniform, the current through each puncture would be so small that the insulation would not be damaged. Due to variations in the resistivity of the soil and the presence of buried metallic structures, however, concentrated arcing may occur, and the insulation may then be damaged in spots, so that corrosion may be initiated. Shield wires reduce the voltage between the sheath and ground and thus the likelihood of damage to the sheath insulation.

It can be demonstrated theoretically, and has been proved by measurements, that when current enters shield wires next to a buried cable in good contact with the earth, the current in the sheath and the voltage between sheath and cable conductors is negligible. The reason for this is that current induced in the sheath by the shield wire current is equal and opposite to the current entering the sheath by virtue of leakage through the ground. Negligible voltages between sheath and core conductors would thus be obtained if shield wires were installed at such a distance that lightning strokes would be intercepted and direct strokes to the cable prevented. To prevent arcing to the cable of strokes to the shield wires, the separation between cable and shield wires would have to be at least 6 feet when the earth resistivity is 1000 meter-ohms, and greater for higher resistivities. Such wires cannot, therefore, be as easily installed in one plowing operation with the cable as shield wires at a smaller spacing. They are, therefore, not considered here, although they have been installed in one instance in a fairly short section where repeated lightning damage had been experienced.

When the sheath, as well as the shield wires, is in intimate contact with the earth, the propagation constant for the shield wires is the same as that for the sheath. In the case of a direct stroke to the cable or the shield wires, the voltage between sheath and shield wires will be so large that they will be in contact with each other at the stroke point by virtue of arcing. The current in the sheath is then:

$$I(x) = \frac{J}{2} \frac{Z_{22} - Z_{12}}{Z_{11} + Z_{22} - 2Z_{12}} e^{-\Gamma x} \tag{101}$$

where J is the total current at $x = 0$, and

$Z_{11} = R_1 + i\omega L_{11}$ = Unit length impedance of sheath
$Z_{22} = R_2 + i\omega L_{22}$ = Unit length impedance of shield wires
$Z_{12} = i\omega L_{12}$ = Unit length mutual impedance of sheath and shield wires

The voltage between sheath and core conductors at $x = 0$ then becomes:

$$\begin{aligned} V(0) &= \frac{J}{2} \frac{R}{\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}}} \frac{Z_{22} - Z_{12}}{Z_{11} + Z_{22} - 2Z_{12}} \left(\frac{1}{i\omega}\right)^{\frac{1}{2}} \\ &= \frac{J}{2} \frac{R}{\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}}} \eta \left(\frac{1}{i\omega}\right)^{\frac{1}{2}} \frac{p + \alpha_0}{p + \beta_0} \end{aligned} \tag{102}$$

where

$$\eta = (L_{22} - L_{12})/(L_{11} + L_{22} - 2L_{12}) \tag{103}$$

$$\alpha_0 = R_2/(L_{22} - L_{12}) \tag{104}$$

$$\beta_0 = (R_1 + R_2)/(L_{11} + L_{22} - 2L_{12}) \tag{105}$$

The function S' is in this case

$$S'(0, t) = \frac{J}{2} \frac{R}{\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}}} \eta \left[\left(\frac{1}{\pi t}\right)^{\frac{1}{2}} + \frac{\alpha_0 - \beta_0}{\beta_0^{\frac{1}{2}}} h(\beta_0^{\frac{1}{2}} t^{\frac{1}{2}})\right] \tag{106}$$

where the function h is defined as before.

When the shield wires are at a sufficient distance from the sheath, so that proximity effects may be neglected, the self and mutual inductances are as follows:

$$L_{11} - L_{12} = \frac{\nu}{2\pi} \log \frac{r_{12}}{r_{11}} \tag{107}$$

$$L_{22} - L_{12} = \frac{\nu}{2\pi} \log \frac{r_{12}}{r_{22}} \tag{108}$$

where

log = $\log_e$ and:

ν = $1.256 \cdot 10^{-6}$ henries per meter

r_{11} = Radius of sheath

r_{22} = Radius of shield wire

r_{12} = Distance between sheath and shield wire.

With more than one shield wire, r_{22} is the geometric mean radius and r_{12} their geometric mean separation from the sheath. When there are two cables, as is frequently the case, r_{11} is the geometric mean radius of the cables and R is their combined sheath resistance.

The surge voltage obtained by solution of (1), for a current as given by (19), is in this case:

$$\begin{aligned} V(0, t) = \frac{IR\eta}{2(\alpha^{\frac{1}{2}} + \beta^{\frac{1}{2}})} \Bigg[& \frac{a - \alpha_0}{a - \beta_0} a^{\frac{1}{2}} h(a^{\frac{1}{2}} t^{\frac{1}{2}}) \\ & - \frac{b - \alpha_0}{b - \beta_0} b^{-\frac{1}{2}} h(b^{\frac{1}{2}} t^{\frac{1}{2}}) + \frac{(\alpha_0 - \beta_0)(b - a)}{(a - \beta_0)(b - \beta_0)} \beta_0^{-\frac{1}{2}} h(\beta_0^{\frac{1}{2}} t^{\frac{1}{2}}) \Bigg] \end{aligned} \tag{109}$$

When $\alpha_0 = \beta_0$, the voltage with shield wires differs from that obtained without shield wires by the factor η. With two 165-mil wires 12 inches apart and 10 inches above the cable considered before, $\alpha_0 = 1.3 \cdot 10^3$, $\beta_0 = 1.5 \cdot 10^3$ and $\eta = .47$. In this case α_0 differs only slightly from β_0 so that the voltage is reduced by the factor η. Measurements made before and after the shield wires were installed indicated a reduction factor of .40 (i.e. the voltage with was .4 times the voltage without shield wires). The reasons for the smaller observed factor is partly that the shield wires are in more intimate contact with the earth than the sheath, and partly that the resistivity of the soil above the cable, where the shield wires are located, is somewhat smaller than the resistivity at the depth of the cable.

With 104-mil wires, $\alpha_0 = 3.2 \cdot 10^3$, $\beta_0 = 2.3 \cdot 10^3$ and $\eta = .49$. When the reduction factor is determined more accurately by calculating the crest voltage with shield wires from (109) and comparing it with the crest voltage without shield wires, a value of .52 is obtained as compared with .47 for 165-mil wires. It is thus seen that, within certain limits, the voltage reduction provided by shield wires depends to a comparatively small extent on the size of the wires, the resistance of 104-mil wires being about 2.5 times that of 165-mil wires.

3.4 *Lightning-Resistant Cable*

As mentioned before, buried cable may be covered by jute, thermoplastic, or rubber for protection against corrosion. The coating may be damaged by gophers or by lightning, and severe corrosion may be experienced at points where the coating is ruptured, particularly when thermoplastic or rubber coating is used. Even when the earth resistivity is low and protection against core insulation failures due to excessive voltage would not be required, the sheath coating may be damaged rather frequently.

Shield wires may effect a substantial reduction in core insulation failures and may also prevent damage to the coating in the case of strokes to ground at some distance from the cable. In the case of direct strokes, however, arcing between the shield wires and the sheath will damage the sheath coating and may also fuse a hole in the sheath, although there may be no insulation failures due to excessive voltage between the sheath and the cable conductors.

Reduction of damage to the coating and to the sheath occasioned by lightning, rodents, or corrosion and protection against core insulation failures occasioned by excessive voltage or crushing of the sheath may be secured by providing the sheath with a thermoplastic or rubber coating and an outside copper shield. If various auxiliary equipment connected to the sheath, such as load coils, gas pressure contactors and terminals are also properly

insulated from ground, currents will not then enter the sheath, except through the capacitance to the outside shield. The voltage across the core insulation will then be so small that core insulation failures will not occur, unless the voltage between the outside shield and the sheath is large enough to puncture the coating. Thus, if the resistance of the outside shield were 1.1 ohms per mile (10-mil copper shield around 1.2″ cable) and if the impulse breakdown voltage of the coating were 20,000 volts, breakdown of the insulation would not be expected except for currents in excess of 130 kiloamperes when the earth resistivity is as high as 4000 meter-ohms. For a cable of 2″ diameter with a 10-mil copper shield, breakdown of the thermo-

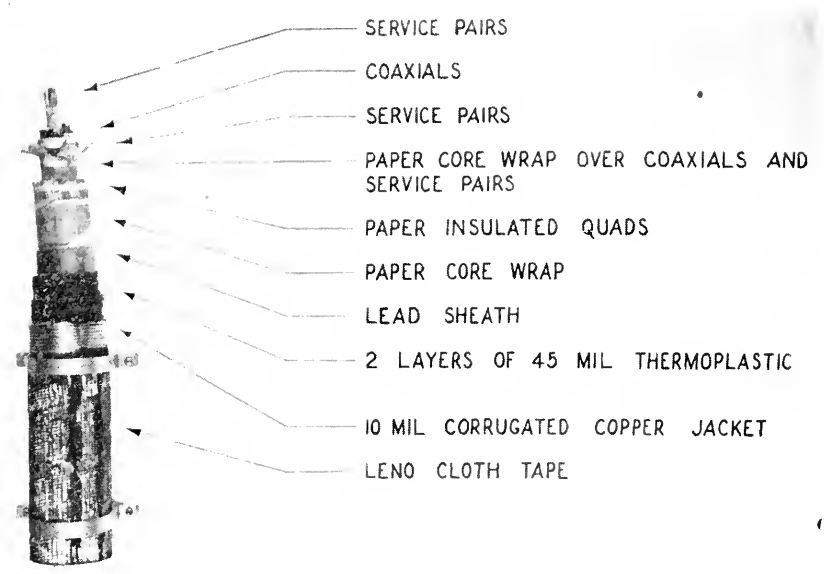


Figure 14—Thermoplastic covered, copper jacketed cable.

plastic insulation would not be expected except for currents in excess of 200 kiloamperes when the breakdown voltage of the coating is 20 kv and the earth resistivity is 4000 meter-ohms, or when the breakdown voltage is 10 kv and the earth resistivity 1000 meter-ohms. The above type of cable is also advantageous in that low-frequency induced voltages and noise due to static are reduced.

A cable of the above type is now being installed for a distance of about 180 miles along the Atlanta-Meridian route, where the effective earth resistivity varies between 1000 and 4000 meter-ohms and lightning storms occur frequently. The diameter of the lead sheath of this cable is about 2″, and the sheath is covered with two layers of thermoplastic each 45 mils

thick, with diametrically opposite seams having an overlap of about $\frac{1}{2}''$. The outside 10-mil copper shield is corrugated to facilitate bending and has a $\frac{1}{4}''$ overlap at the seam. The thermoplastic coating is flooded with thermoplastic cement to reduce moisture absorption. A photograph of this cable is shown in Fig. 14.

SUMMARY

Current in the sheath of buried cables, due to direct lightning strokes, strokes to ground near the cables or discharges between clouds, gives rise to voltages between the cable conductors and the sheath. The voltages are practically proportional to the square root of the earth resistivity and to the direct-current sheath resistance. For the latter reason they are substantially larger for carrier cables of the size now used than for the much larger voice-frequency cables. While direct strokes are usually most important, strokes to ground must be considered when the cables are of small size, even when the surface resistivity is low, provided the resistivity at greater depths is high. Under the latter conditions it is possible that for small cables, discharges between clouds over the cable may also cause failure.

For cables with thermoplastic or rubber coating the voltages between the sheath and the core conductors are much the same as for jute-covered cables. The coating of such cable is likely to be damaged by direct strokes and strokes to ground near the cable, in which case corrosion of the sheath may occur at such points.

Based on theoretical lightning expectancy curves, the incidence of lightning troubles increases faster than the sheath resistance or the earth resistivity. When the breakdown voltage of the core insulation is doubled by use of extra core wrap, or when shield wires are installed in situations where lightning damage is anticipated or has been experienced, a substantial reduction in lightning failures is to be expected. Shield wires will, however, not prevent damage to the sheath and the sheath coating.

Where the earth resistivity is very high and lightning storms occur frequently, doubled core insulation together with shield wires may not provide sufficient protection, even for cable of substantial size. Protection against various forms of lightning damage may then be secured by use of thermoplastic sheath coating of adequate dielectric strength together with an outside concentric copper shield.

REFERENCES

1. E. D. Sunde: "Lightning Protection of Buried Cable," *Bell Telephone Laboratories Record*, Vol. 21, No. 9, May 1943.
2. B. F. J. Schonland: "The Lightning Discharge," Clarendon Press, Oxford, England, 1938 or "Thunderstorms and Their Electrical Effects," *Proc. of the Phys. Soc.*, Vol. 55, Part 6, No. 312, Nov. 1, 1943.

3. K. B. McEachron: "Lightning to the Empire State Building," *Electrical Engineering*, Dec. 1938 and Sept. 1941.
4. W. W. Lewis and C. M. Foust: "Lightning Investigations on Transmission Lines" VIII—*A.I.E.E. Trans.*, Vol. 64, 1945, p. 107.
5. H. Grünewald: C.I.G.R.E. 1939, Report 323.
6. C. F. Wagner and G. D. McCann: "Lightning Phenomena," *Elec. Engg.*, Aug., Sept. and Oct. 1941.
7. C. E. R. Bruce and R. Golde: "The Mechanism of the Lightning Stroke and its Effect on Transmission Lines," *Jour. I.E.E.*, Part II, Dec. 1941.
8. E. D. Sunde: "Surge Characteristics of a Buried Bare Wire," *A.I.E.E. Transactions*, Vol. 59, 1940 or *Bell Telephone System Monograph* B-1279.
9. S. A. Schelkunoff: "Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shield," *Bell System Technical Journal*, October, 1934. Equation (5) is equivalent to Equation (82) of the above paper.
10. G. A. Campbell and R. M. Foster: "Fourier Integrals for Practical Applications," *Bell Telephone System Monograph* B-584, pair 807.
11. E. D. Sunde: "Currents and Potentials Along Leaky Ground Return Conductors," *Elec. Engg.*, December 1936 or *B.T.S. Monograph* B-970.
12. J. R. Carson: "Wave Propagation in Overhead Wires with Ground Return," *Bell Sys. Tech. Jour.*, October 1926. Z obtained by use of (34) and (35) and M by use of (36) and (37) of the paper referred to.
13. J. Riordan and E. D. Sunde: "Mutual Impedance of Grounded Wires for Stratified Two-Layer Earth," *Bell System Technical Journal*, April 1933 or *B.T.S. Monograph* B-726.
14. E. Hansson and S. K. Waldorf: "An Eight-Year Investigation of Lightning Currents and Preventive Lightning Protection on a Transmission System" presented at A.I.E.E. 1944 Winter Convention.
15. W. H. Alexander: "The Distribution of Thunderstorms in the United States, 1904–33," *Monthly Weather Review*, Vol. 63, 1935, page 157.
16. A. Matthias: "Modellversuche über Blitzeinschläge," *E.T.Z.* Aug. 12 and Sept. 9, 1937.
17. A. Matthias and W. Burkhardtsmaier: "Der Schutzraum von Blitzfang-Vorrichtungen und seine Ermittlung durch Modellversuche," *E.T.Z.*, June 8 and June 15, 1939.
18. C. F. Wagner, G. D. McCann and G. L. MacLane, Jr.: "Shielding of Transmission Lines", *A.I.E.E. Trans.*, Vol. 60, 1941, p. 313–328.
19. C. F. Wagner, G. D. McCann and C. M. Lear: "Shielding of Substations," *A.I.E.E. Trans.*, Vol. 61, 1942, p. 196–200.
20. P. L. Bellaschi, R. E. Armington and A. E. Snowden: "Impulse and 60-cycle Characteristics of Driven Grounds—II," *A.I.E.E. Trans.*, Vol. 61, 1942, p. 349.
21. R. Davis and T. E. M. Johnston: "Surge Characteristics of Tower Footing Impedance," *J.I.E.E.* Vol. 88, Part II, No. 5, Oct. 1941.
22. I. R. Eaton: "Impulse Characteristics of Electrical Connections to the Earth," *General Electric Review*, Oct. 1944, p. 41–50.
23. G. D. McCann: "The Effect of Corona on Coupling Factors between Ground Wires and Phase Conductors," *A.I.E.E. Trans.*, Vol. 62, 1943, p. 818–826.
24. G. Wascheck: "Earth Resistivity Measurements," *Bell Telephone Laboratories Record*, Vol. 19, No. 6, Feb. 1941.

Abstracts of Technical Articles by Bell System Authors

Ultra-Short-Wave Receiver for the Cape Charles-Norfolk Multiplex Radiotelephone Circuit.[1] D. M. BLACK, G. RODWIN and W. T. WINTRINGHAM. The requirements for an ultra-short-wave receiver for use in a multiplex radiotelephone link circuit are outlined. The technical details of a receiver designed to meet such requirements in the circuit between Cape Charles and Norfolk, Virginia, are described.

Ultra-Short-Wave Multiplex.[2] CHARLES R. BURROWS and ALFRED DECINO. The technical requirements of a twelve-channel ultra-short-wave multiplex system are discussed and the means of meeting them are described. The intermodulation between channels in equipment based on this design has been reduced to the point where it is possible to use twelve-channel radio systems in the toll plant. By employing a sufficient amount of envelope feedback, the transmitter can be operated with a high modulation factor without the use of spread sidebands.

Airplane Vibration Reproducer.[3] G. R. CRANE. This paper describes a reproducer set designed for use in the reproduction for analysis of multiple track film recordings. It is capable of reproducing simultaneously 13 variable-area tracks recorded side by side on standard 35-mm. film. Recorded signals between 5 and 3000 cps are accurately reproduced and may be analyzed for frequency components, amplitude, and phase relation.

Airplane Vibration Recorder.[4] J. C. DAVIDSON and G. R. CRANE. This paper describes a portable film recorder capable of simultaneously recording 13 variable-area tracks on 35-mm. film. It is intended for use in the analysis of airplane vibration or similar studies in which it is desirable to record disturbances (mechanical, acoustical, or electrical) from a number of sources in such a manner that the resultant record can be analyzed for frequency, amplitude, and phase relation. Film speeds of 12, 6, or 3 in. per sec. are available.

Application of Sound Recording Techniques to Airplane Vibration Analysis.[5] J. G. FRAYNE and J. C. DAVIDSON. This paper describes methods which have been developed for analysis of the various vibration components present in airplane structures. The complex wave forms are recorded on standard motion picture sound negatives during flight. These films later,

[1] *Proc. I. R. E.*, February 1945.
[2] *Proc. I. R. E.*, February 1945.
[3] *Jour. S. M. P. E.*, January 1945.
[4] *Jour. S. M. P. E.*, January 1945.
[5] *Jour. S. M. P. E.*, January 1945.

after proper development, are analyzed electrically, making possible a complete analysis on the ground and thereby reducing materially the time devoted to flight test, and also simplifying the process of analysis of complex wave forms.

Ultra-Short-Wave Transmitter for the Cape Charles-Norfolk Multiplex System.[6] R. J. KIRCHER and R. W. FRIIS. Design features of an unattended ultra-short-wave double-sideband multiplex transmitter are described. Forty decibels of envelope feedback is utilized over the 12- to 60-kilocycle band of the twelve type-K carrier-signal channels which modulate the last stage of the transmitter. Accessibility of apparatus and ease in maintenance contribute toward obtaining maximum reliability of the equipment in commercial service.

Paper Capacitors Containing Chlorinated Impregnants. Stabilization by Anthraquinone.[7] D. A. MCLEAN and L. EGERTON. This paper shows anthraquinone to be an effective stabilizer for capcitors having paper dielectrics containing chlorinated impregnants when aluminum electrodes are used and d-c. potentials are applied. One half per cent of anthraquinone prevents formation of the usual carbonized brown spots in the paper, and diminishes corrosion of electrodes and instability of leakage current. It increases the life under accelerated testing conditions by factors of four to one hundred fold, depending upon materials used and conditions of test. This development has added appreciably to the reliability of paper capacitors containing chlorinated impregnants, particularly for military equipment where high temperatures and high voltages are often encountered simultaneously. Solubility of anthraquinone in the usual chlorinated impregnants is limited. Where greater solubility is desired, the more soluble chloro and methyl derivatives can be used.

Reflex Oscillators.[8] J. R. PIERCE. This paper discusses qualitatively the behavior of reflex oscillators. Power production, electronic tuning, variation of frequency with resonator voltage, effect of modulation coefficient, and influence of load are considered. Two brief mathematical appendixes are included.

Cape Charles-Norfolk Ultra-Short-Wave Multiplex System.[9] N. F. SCHLAACK and A. C. DICKIESON. This paper describes the general features of a radio multiplex system which has been installed between Cape Charles and Norfolk, Virginia. The radio-frequency equipment operates in the vicinity of 160 megacycles. The system employs the 12 telephone channels of the type K cable carrier system which are in the frequency range 12 to 60 kilocycles.

[6] *Proc. I. R. E.*, February 1945.
[7] *Indus. & Engg. Chem.*, January 1945.
[8] *Proc. I. R. E.*, February 1945.
[9] *Proc. I. R. E.*, February 1945.

Contributors to this Issue

I. E. Fair, B.S. in Electrical Engineering, Iowa State College, 1929. Bell Telephone Laboratories, Radio Research Department, 1929–. Mr. Fair has been engaged in the study of piezoelectric crystals and crystal oscillators.

C. W. Harrison, B.S. in Electrical Engineering, Purdue University, 1938; M.S., Lehigh University, 1940. Bamberger Broadcasting Service, 1939–41; Bell Telephone Laboratories, 1941–. Engaged in the development of communication circuits.

E. D. Sunde, B.S., Haugesund, Norway, 1922; E.E., Darmstadt, Germany, 1926. American Telephone and Telegraph Company, 1927–33; Bell Telephone Laboratories, 1933–. Mr. Sunde has been engaged in studies of interference in telephone circuits from power lines and railway electrification and is now concerned with studies of protection of the telephone plant against lightning damage.

VOLUME XXIV JULY-OCTOBER, 1945 NOS. 3-4

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

$1.00 (Double Copy) *$1.50 per Year*

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are 50 cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

Vol. XXIV *July–October, 1945* *Nos. 3–4*

Physical Limitations in Electron Ballistics*

By J. R. PIERCE

Introduction

THE subject of this talk is "Physical Limitations in Electron Ballistics". It is pleasant to have a chance to talk about such physical limitations, because there is so little we can do about them. And, although these limitations are apt to be discouraging, a knowledge of them is very valuable, for it keeps us from spending time trying, like the inventors of perpetual motion machines, to do the impossible.

As electron ballistics is particularly subject to physical limitations, there are so many that it is impossible to discuss all of them thoroughly at this time. Also, many of the limitations are of a rather complicated nature, and to deduce them from basic principles in a quantitative way requires much thought and patience. I think the best I can do is to try to mention most of the chief limitations, as a warning to the uninitiated that rocks lie ahead in certain directions, but to concentrate attention on only a few of them. I have chosen this evening to devote particular attention to limitations that bear on the production and use of electron beams in which considerable current is required, such as those used in cathode ray tubes and high-frequency oscillators, and to mention only briefly as a sort of introduction problems pertaining more closely to low-current devices such as electron microscopes.

The Wave Nature of the Electron

One of the most important limitations in electron microscopy is the dual nature, wave and corpuscular, of the electron. Without making any attempt to justify or explain the combination of wave and particle concepts which is characteristic of modern physics, we may describe its consequence at once; very small objects don't cast distinct shadows. This cannot be explained merely in terms of the physical size of the electron and the object. When an electron beam is reflected from a surface of regularly

* A lecture given under the auspices of the Basic Science Group of the New York Section of the A.I.E.E., as a part of an Electron Ballistics Symposium, Columbia University, March 21, 1945.

spaced obstacles (the atoms in a crystal lattice, for instance) diffraction patterns are obtained, similar to those which may be obtained with waves of X-rays or light. It appears that electrons get around sufficiently small objects just as sound waves get around telephone poles, automobiles, and even houses, and if the objects are sufficiently small their effect on the electron flow will either be absent or will consist of a few ripples which are meaningless in disclosing the shape or size of the object.

The electron wave-length, which varies inversely as the momentum of the electron, may be simply expressed in terms of the energy V in electron volts. A simple non-relativistic expression which is only 5% in error at 100,000 volts (a high voltage for electron microscopes), is*

$$\lambda = \sqrt{150/V} \times 10^{-8} \text{ cm} \tag{1}$$

Thus for 30,000-volt electrons the wave-length is 7×10^{-10} cm or about 1.4×10^{-7} times the diameter of a hair and 1.2×10^{-5} times the length of a wave of yellow light.

In terms of this wave-length λ and the half angle of the cone of rays accepted by the objective, α, we can express the distance d between point objects which can just be distinguished in an electron microscope. This distance is

$$d = .61\lambda/\sin \alpha \tag{2}$$

For small values of α

$$2\alpha = 1/\text{f} \tag{3}$$

where f is the well known photographic f number, the ratio of the focal length to the lens diameter. We see that, just as with cameras, the smaller the f number the better. In electron microscopes a small f enables us to distinguish smaller objects.

Aberrations

Just as in cameras, the limitation to the f number is imposed by lens aberrations. But in electron lenses the aberrations are much more severe. Why is this so? Because with electron lenses we have less freedom of design than with optical lenses.

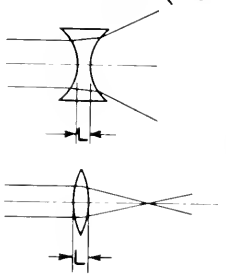
Consider an electric lens. The quantity analogous to the index of refraction for light is the square root of the potential with respect to the cathode. Now suppose that with a light lens we know the index of refraction at every point along the axis. Suppose, for instance, that the index of refraction is 1 everywhere along the axis except for a space L long

* The relativistic expression is

$$\lambda = (\sqrt{150/V}/\sqrt{1 + .98 \times 10^{-6} V}) \times 10^{-8} \text{ cm}$$

where it is 2, as in Fig. 1. Our lens may be converging or diverging; strong or weak. In the analogous electric case, however, the potential throughout the lens space must satisfy Laplace's equation, and this means that if it is specified along the axis it is known everywhere. We can easily see this by writing down Laplace's equation for an axially symmetrical field.

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial V}{\partial r}\right)+\frac{\partial^2 V}{\partial z^2}=0 \tag{4}$$



LIGHT-CONDITIONS OFF AXIS NOT FIXED BY CONDITIONS ON AXIS

ELECTRIC FIELD-FIELD OFF AXIS SPECIFIED BY POTENTIAL ON AXIS

$$V=\frac{1}{\pi}\int_0^{\pi} f(z+ir\cos\theta)\,d\theta$$

Fig. 1—Contrast between optical and electric focussing conditions.

The field near the axis may be expanded in powers of f

$$\frac{\partial V}{\partial r}=ar+\cdots \tag{5}$$

Substituting this into (4),

$$\frac{1}{r}\frac{\partial}{\partial r}(ar^2)=2a=\frac{-\partial^2 V}{\partial z^2}$$

$$\frac{\partial V}{\partial r}=\frac{-1}{2}\frac{\partial^2 V}{\partial z^2}r \tag{6}$$

As a matter of fact, the potential V (z,r) remote from the axis can be expressed in terms of the potential $V_0(z)$ on the axis as

$$V=\frac{1}{\pi}\int_0^{\pi} V_0(z+ir\cos\theta)\,d\theta \tag{7}$$

If we could introduce charges into our lens, Laplace's equation would no longer hold and we would have more freedom of design. The methods proposed for the introduction of charges comprise the use of free charges (space charge) which are largely uncontrollable, and the use of curved grids, which do more damage than good. In other words, the cures are worse than the disease.

Similar limitations apply to magnetic lenses, and in the end we find that because of the simplest form of aberration, spherical aberration, best definition is achieved in electron microscopes with f numbers of 100 or greater, while the f number of a light microscope objective corrected for spherical aberration and other defects as well may be around unity. Thus the electron microscope is severely handicapped, and this handicap is overcome only because electron waves are much less than 1/100 the length of light waves.

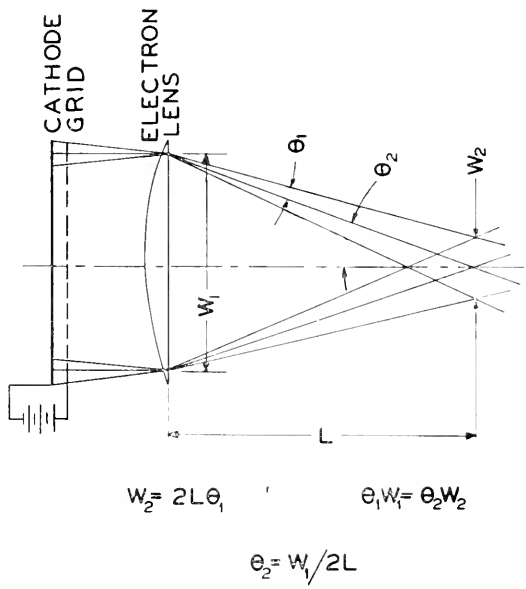


Fig. 2—Approximate relation between beam size and angular spread.

Thermal Velocities of Electrons

In many electron-optical systems, and particularly in such devices as cathode ray tubes, it is desirable to focus an electron beam into a small area, so as to produce a very small spot on a fluorescent screen, or to pass a considerable current through a small aperture. We might think at first that if our focusing system were good enough, that is, if it had very small aberrations, we could focus a current from a cathode of given area into as small a space as we desired. This, unfortunately, is not so. The obstacle is the thermal velocities of the electrons emitted by the cathode.

A simple example will show the sort of thing we should expect to take place. Figure 2 shows a plane cathode and near to it a positive grid so fine as to cause no appreciable deflections of the electrons which pass through it. Farther on we have an aberrationless electron lens designed to focus

the electron stream at a spot a distance L beyond it. The electrons will leave the cathode with some slight sidewise velocity components; so, electron paths will pass at several angles through a given point on the lens. The lens will bend these paths approximately equally, and hence we can see that at the point where the beam is narrowest it will still have some appreciable diameter W_2.

Now consider the beam at the lens. Suppose that through a given point all the paths lie within a cone of half angle θ. Then the width W_2 is approximately

$$W_2 = 2L\theta_1 \tag{8}$$

We can also see that the paths at W_2 will lie within an angle approximately

$$\theta_2 = W_1/2L \tag{9}$$

Hence we see that approximately

$$\theta_1 W_1 = \theta_2 W_2 \tag{10}$$

In other words, we can have a small spot through which electrons pass over a wide angular range, or we can have a broad beam in which all paths are nearly parallel, but we can't have a narrow spot and nearly parallel rays.

We see that the actual width of spot will depend on the thermal velocities, which are proportional to the square root of the cathode temperature, and on the forward velocity, which is proportional to the square root of the accelerating voltage. By using more involved arguments we discover that for any point in an electron stream, where the beam is wide, narrow, or intermediate, the current in an arbitrary direction chosen as the x direction can be expressed[4,*]

$$dj = \frac{4kT}{\pi m} j_0 v_x \epsilon^{(1/kT)(eV - mv^2/2)}\, dv_x\, dv_y\, dv_z \tag{11}$$

$$v = v_x^2 + v_y^2 + v_z^2$$

$$\text{when } v > \sqrt{2eV/m}; \tag{12}$$

$$\text{or } dj = 0 \tag{13}$$

$$\text{when } v < \sqrt{2eV/m} \tag{14}$$

Here j_o is the cathode current density, V is voltage with respect to the cathode, T is the absolute temperature of the cathode in degrees Kelvin, and v_x, v_y, and v_z are the three velocity components; dj is the element

* This expression neglects the effects of electron collisions, which may actually make the current density smaller.

of current density carried by electrons which have velocity components about v_x, v_y, v_z, lying in the little range of velocity $dv_x\ dv_y\ dv_z$.

The reason for restriction (12) is that if an electron starts with zero thermal velocity from the cathode, it will attain the velocity given by the right side of (12) by falling through the potential drop V. As electrons cannot have velocities smaller than this, we have (13) and (14).

By integrating (11) with appropriate limits we obtain a more specialized but very useful expression

$$j < j_m = j_o \left(1 + \frac{11600V}{T}\right) \sin^2\theta \tag{15}$$

For usual values of voltage, unity in the parentheses is negligible, and we can say that if all the electron paths approaching a given point in an electron beam lie within a cone of half angle θ, the current density j at that point cannot be greater than a limiting value j_m which is proportional to the

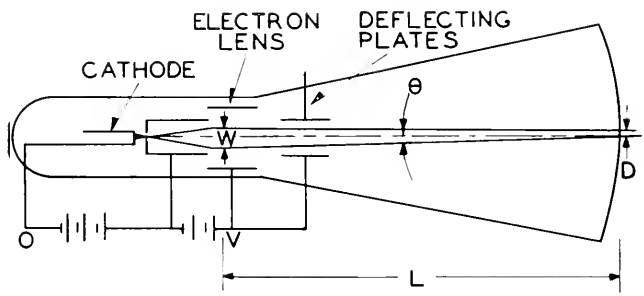


Fig. 3—Parameters important in determining spot size in a cathode ray tube.

cathode current density, to the voltage, to $\sin^2\theta$, and inversely proportional to the cathode temperature.

Let us see what this means in some practical cases. Figure 3 shows a cathode ray tube. The electron stream has a width W at the final electron lens, and is focused on a screen a distance L beyond the lens. The half angle of the cone of rays reaching the screen cannot be greater than

$$\sin\theta = \theta = W/2L \tag{16}$$

Suppose the spot must have a diameter not greater than d. Let the spot current be i. Then from (15),

$$j = \frac{4i}{\pi d^2} < j_0 \left(1 + \frac{11600V}{T}\right) (W/2L)^2,$$

$$i < \frac{\pi d^2}{4} j_0 \left(1 + \frac{11600V}{T}\right) (W/2L)^2. \tag{17}$$

Thus if for a given spot size we want to increase the spot current, and if we are limited to a given cathode current density because of cathode life, we must make V larger, W larger or L smaller.

Making W larger increases both lens and deflection aberrations. Making L smaller means that for a given linear deflection we must increase the angular deflection, and this too tends to defocus the spot. Because of these limitations, it is necessary to avail ourselves of the remaining variable and raise the operating voltage V.

Another illustration, perhaps a little more subtle, of the effect of thermal velocities, lies in the analysis of the properties of a type of vacuum tube amplifier known as the "deflection tube". In such a device, illustrated in Fig. 4, an electron stream from a cathode is accelerated and focused by a lens and deflected by a pair of deflecting electrodes so as to hit or miss an output electrode. Such a device may be used as an amplifier.

Now it is obvious that as the output electrode on which the beam is focused is moved farther away from the deflecting plates, a given deflecting voltage will produce a greater linear deflection of the beam at the output.

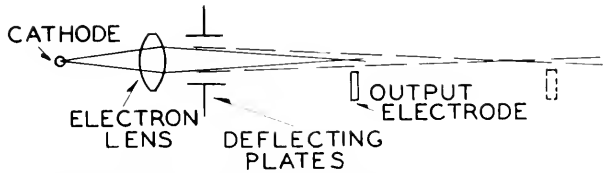


Fig. 4—Amplifying tube making use of electron deflection.

As this at first sight seems desirable; it has been seriously suggested not only that this be done, but that an elaborate electron optical system be interposed between the deflecting plates and the output electrode to amplify the deflection.

The merit of a deflection tube is roughly measured by the deflecting voltage required to move the beam from entirely missing the output electrode to entirely hitting the output electrode, and, of course, moving the output electrode farther away or putting lenses between the deflecting plates and the output electrode doesn't reduce this voltage at all. As we improve the deflection sensitivity by these means, we simply increase the spot size at the same time. Focusing our attention on the beam between the deflecting plates, we appreciate at once that the electron paths through each point will be spread over some cone of half angle θ, and that to change from a clean miss to a clean hit we must deflect the electrons through an angle of at least 2θ, regardless of what we do to the beam afterwards.

Returning for a moment to equation (15), we see that it says the current density can be less than a certain limiting value depending on θ. Yet

expression (15) was obtained by integrating a supposedly exact expression. What does this inequality mean?

The answer is that for the current to have the limiting value, electrons of *all allowable velocities* must approach *each part* of the spot from *all angles* lying within the cone of half angle θ. When the average current density in the spot is less than the limiting current density, the possibilities are

(a) Electrons are approaching each point in the beam from all angles, but along some angles only electrons which left the cathode with greater than zero velocity can reach the spot.

(b) Electrons leaving the cathode with all velocities can reach the spot, but at some portions of the spot electrons don't come in at all angles within the cone angle θ.

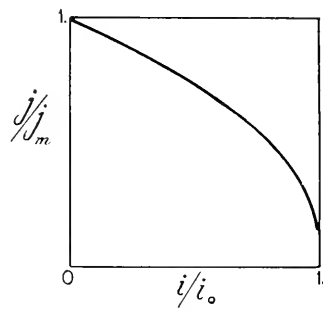


Fig. 5—Relation between nearness of approach to limiting current density and fraction of current utilized.

Thus, we can have less than the limiting current either because electrons do not reach the spot with all allowable velocities or from all allowable angles. Of course both factors may operate.

We can easily see how lens aberrations, which we know are present in all electron-optical systems, can prevent our attaining the limiting current density. There is a more fundamental limitation, however. It can be shown that even with perfect focusing, we must sort out and throw away part of the current in order to approach the limiting current density, and we can even derive a theoretical curve for the case of perfect focusing relating the fraction of the limiting current density which is attained to the fraction of the cathode current which can reach the spot. Figure 5 shows such a curve which applies for voltages higher than, say, 10 volts.

Usually, the failure to approach the limiting current density is chiefly caused by aberrations, and in ordinary cathode ray tubes the current density in the spot may be only a small fraction of the limiting value. A very close approach to the limiting current density has been achieved in a

special cathode ray tube designed by Dr. C. J. Davisson of the Bell Telephone Laboratories.

When we become thoroughly convinced that these equations expressing the effects of thermal velocities very much cramp our style in designing electron-optical devices, as good engineers we wonder if there isn't, after all, some way of getting around them. I don't think there is. The suggestion illustrated in Fig. 6 is a typical example of such an attempt. We know that in a strong magnetic field electrons tend to follow the lines of force. Why not use a very strong magnetic field with lines of force approaching the axis at a gentle angle to drag the electron stream toward the axis?

An electron off axis traveling parallel to the axis certainly will be dragged inward by such a field. The catch is that the field pulls the electron in because it makes the electron spiral around the axis. As the beam converges and the field becomes stronger, the pitch of each spiral decreases and the angular speed of each electron increases. Finally, if the field is strong enough, all the kinetic energy of the electron is converted from forward

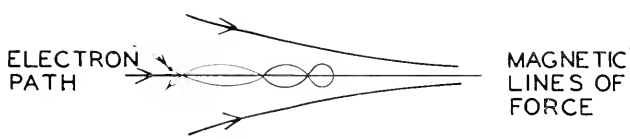


Fig. 6—Reflection of an electron by a magnetic field with strongly converging lines of force.

motion to revolution about the axis; the electron ceases to move into the field and bounces back out. It may be some small consolation to know that very high-current densities can be achieved by this means, but only because in their flat spiralling the electrons approach a spot at much wider angles with the axis than the small inclination of the lines of force.

Space Charge Limitations

In electron beam devices using reasonably large currents, the space charge of the electrons is a very serious source of trouble both in complicating design of the devices and in limiting their performance.

Let us begin our consideration right at the electron gun, the source of electron flow in many devices such as cathode ray tubes and certain high-frequency tubes. Electron guns are sometimes designed on the basis of radial space charge limited electron flow between a cathode in the form of a spherical cap of radius r_o and a concentric spherical anode a distance d from the cathode. It can be shown that by use of suitable electrodes external to the beam, radial motion can be maintained between cathode and anode along

straight lines normal to the cathode surface. A hole in the anode electrode will allow the beam to emerge from the gun. Because of the change in

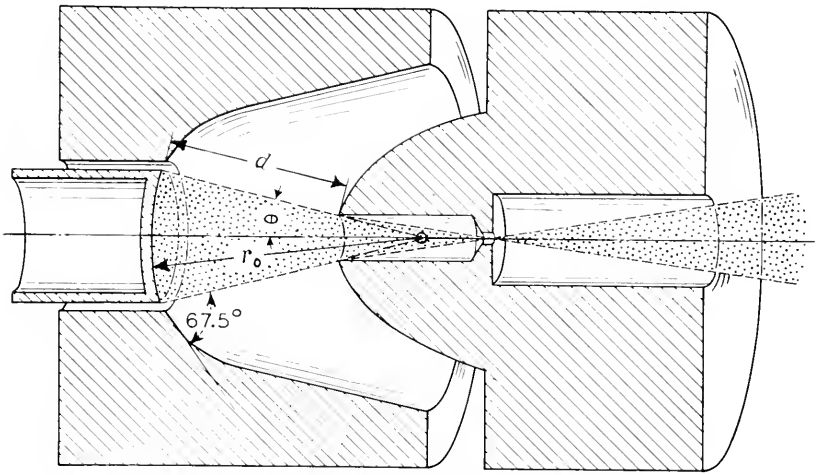


Fig. 7—Electron gun utilizing rectilinear flow.

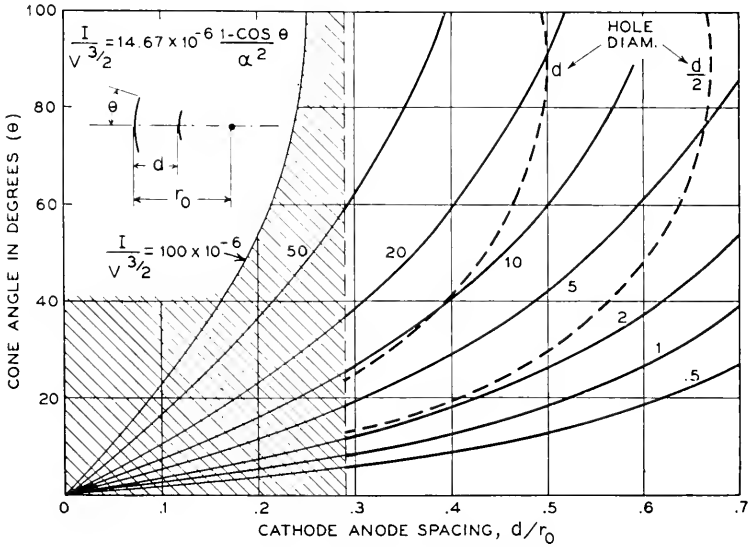


Fig. 8—Relation between perveance, angle of cone of flow, and cathode-anode spacing.

field near the hole, the hole acts as a diverging electron lens.[11] Figure 7 illustrates such a gun.[15] The curves shown in Fig. 8 relate to this sort of

electron gun. They are plots of a factor called the perveance, which is defined as

$$P = I/V^{3/2} \tag{18}$$

(that is, current divided by voltage to the 3/2 power) as a function of θ, the half angle of the cone of flow, and d/r_o, the ratio of cathode-anode spacing to cathode radius. In getting an idea of the meaning of the curves, we may note that a perveance of 10^{-6} means a current of 1 milliampere at 100 volts. It is obvious from the curves that to get very high values of perveance, that is, high current at a given voltage, θ must be large and the cathode-anode spacing must be small. Making θ large means that electrons approach the axis at steep angles; aberrations are bad and the beam tends to diverge rapidly beyond crossover. Moving the anode near to the cathode means that the hole which must be cut in the anode to allow the beam to pass through must be large, and cutting such a large hole in the anode defeats our aim of getting higher perveance; we can't pull electrons away from the cathode with an electrode which isn't there. Further, for ratios of spacing to cathode radius less than about .29, the lens action of the hole in the anode causes the emerging beam to diverge, which would make the gun unsuitable for many applications.

When we build guns for small currents at high voltages, such as cathode ray tube guns, space charge causes little trouble; when we try to obtain large currents at lower voltages, we find ourselves seriously embarrassed.

Suppose we now turn our attention to the effect of space charge in beams when the beam travels a distance many times its own width. Consider, for instance, the case of a circular disk forming a space charge limited cathode. Suppose we place opposite this a fine grid, and shoot an electron stream out into a conducting box, as illustrated in Fig. 9a. We immediately realize that there will be a potential gradient away from the charge forming the beam. In this case, the gradient will be toward the nearest conductor; that is outwards, and the electron beam will diverge.

How can we overcome such divergence? One way would be to arrange the boundary conditions in such a fashion that all the field would be directed along the beam instead of outwards; this might be done by surrounding the beam by a series of conducting rings and applying to them successively higher voltages as in 9b, the voltages which would occur in electron flow between infinite parallel planes with the same current density. Another way in which the same effect may be achieved is through use of specially shaped electrodes outside of the beam, as shown in Fig. 9c.[11] In maintaining parallel flow by these means, the electric field due to the electrons acts along the beam, and increases continually in magnitude with

distance from the cathode. We can in fact calculate the potential at any distance along the beam by the well known Child's law equation

$$I = 2.33 \times 10^{-6} AV^{3/2}/x^2$$
$$V = 5{,}690 x^{4/3} I^{2/3}/A^{2/3} \tag{19}$$

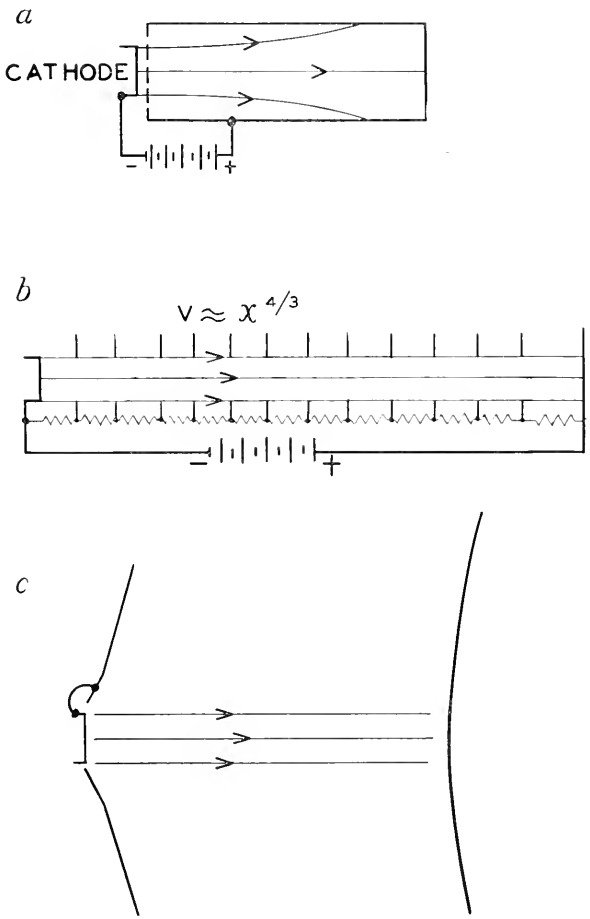


Fig. 9—Avoiding beam divergence by means of a longitudinal electric field.

Here V is the anode voltage, x the cathode-anode spacing, I the current in amperes and A the cathode area.

Suppose we take as an example

$$A = 1 \text{ cm}^2$$
$$I = .01 \text{ amp.}$$
$$x = 10 \text{ cm}$$

Then

$$V = 5{,}700 \text{ volts}$$

Thus to maintain parallel motion of the modest current of 10 milliamperes spread over an area of one square centimeter requires 5,700 volts. Moreover, the requirement of distributing this voltage smoothly along the beam would make it very difficult to put the beam to any use.

One means for mitigating the situation is to use an electron lens and direct the beam inward. Of course, the beam will eventually become parallel and then diverge again, but by this means a fairly large current can be made to travel a considerable distance. Some calculations made by Thompson and Headrick[12] cover this type of motion, with an especial emphasis on the problem in cathode ray tubes, in which the currents are moderate.

In order to confine large currents into beams, an axial magnetic field is sometimes used, as shown in Fig. 10 Here a cathode-grid combination shoots a beam of electrons into a long conducting tube. A long coil around the tube produces an axial magnetic field intended to confine the electron

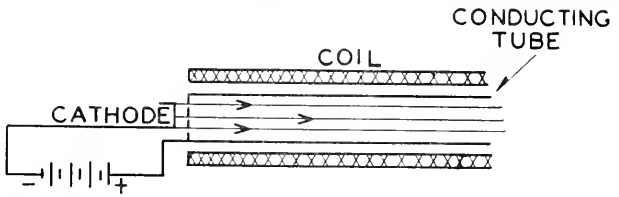


Fig. 10—Avoiding beam divergence by means of a longitudinal magnetic field.

paths in a roughly parallel beam. The radial electric field due to space charge will cause the beam to expand somewhat and to rotate about the axis. As the magnetic field is made stronger and stronger, the electrons will follow paths more and more nearly straight and parallel to the axis. For a given current and voltage, there is one sort of physical limitation in the strength of magnetic field we need to get a satisfactory beam. It is another effect that I wish to discuss.

Suppose we have a very strong magnetic field, in which the electrons travel almost in straight lines. We know, of course, that the radial electric field is still present, and this means that the potential toward the center of the beam is depressed; this in turn means that the center electrons are slowed down. This slowing down of course increases the density of electrons in the center of the beam. The result is that if for some critical voltage or speed of injection we increase current beyond a certain value, the process runs away, the potential at the center of the beam drops to zero, and another type of electron flow with a "virtual cathode" of zero electron velocity at the center of the beam is established. Thus, although the magnetic field

has overcome the diverging effect of the space charge, we still have a space charge limitation of the beam current. C. J. Calbick has calculated the value of this limiting current.[13] If the beam completely fills a conducting tube at a potential V with respect to the cathode, the limiting beam current is independent of the diameter of the beam and is

$$I = 29.3 \times 10^{-6} V^{3/2} \tag{20}$$

If the beam diameter is less than that of the conducting tube, the limiting current is lower.

But perhaps we can completely overcome the effects of space charge. Suppose we put a very little gas in the discharge space. Then positive ions will be formed. Any tendency of the electronic space charge to lower the potential and slow up the electrons will trap positive ions in the potential minimum and so raise the potential. Thus the gas enables us to get rid of the the slowing up effect of the space charge as well as its diverging effect.

Before we congratulate ourselves unduly, it might be well to make sure about the stability of an electron beam in which the electronic space charge is neutralized by heavy positive ions. Langmuir and Tonks, in their work on plasma oscillations, introduced a concept, extended later by Hahn and Ramo, which enables us to investigate this problem. The concept is that of space charge waves. It is found that in a cloud of electrons whose net space charge is neutralized by heavy, relatively immobile positive ions, small disturbances of the electron charge density produce a linear restoring force; and this, together with the mass of the electrons, makes possible a type of space charge wave which may be compared roughly with sound waves, although much of the detailed behavior of space charge waves is quite different from that of sound waves. We may express a disturbance in an electron beam in terms of these space charge waves and then examine the subsequent history of the disturbance as a function of time. This has been done[14] and the perhaps surprising result is that even when the electronic space charge is neutralized by heavy positive ions, the flow tends to collapse if the current is raised above a limiting value

$$I = 190 \times 10^{-6} V^{3/2} \tag{21}$$

It is true that this current is 6.5 times the limiting current in the absence of ions, but it is a limit nevertheless.

If this limit in the presence of ions seems unnatural, perhaps we should recall a mechanical analogy. Consider a vertical long column subjected to a load F. If we subject it to a sidewise force αF proportional to F, as shown in Fig. 11a, the behavior on increasing F will be a gradual deformation (analogous to the space charge lowering of potential in the absence of ions)

ending in collapse. However, even if, as in 11b, there is no sidewise loading and no bending during loading, we know from Euler's formula that beyond a certain loading the column will still collapse. This behavior is analogous to that of an electron beam in which the electronic space charge is neutralized by positive ions and there is no depression of potential in the beam.

This space charge limitation either in the presence or absence of ions allows the passage of quite a large current through a tube, as the table below will show:

Voltage	Current, amperes, no ions	Current, amperes, ions
1000	.927	6.01
100	.029	.190
10	.009	.060

We might therefore feel that the space charge is disposed of in a practical sense, and so it is in many cases.*

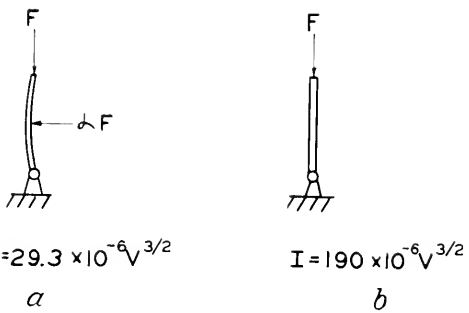


Fig. 11—Comparison of limiting stable beam currents with and without positive ions.

Power Dissipation Limitations

Having talked about various limitations imposed by wave effects, aberrations, thermal velocities and space charge on the electron flow in the beam itself, I want to close by discussing briefly a topic which seems hardly included in electron ballistics but yet is vital to any application in that field. I refer to the problems associated with power dissipation when electrons strike something and stop. This is a good deal like the problem imposed by suddenly coming down to earth while studying the sensations of a free fall. It is inevitable and may be fatal unless satisfactory provision is made for the dissipation of kinetic energy.

What I want chiefly to bring out are the consequences of scaling a given electronic device down in size. If we change the size of each part of an

* It appears that in many gas discharges, including those in which plasma oscillations are observed, the current is too high to allow persistence of the homogeneous flow upon which the plasma oscillation equations are based.

electron device in the ratio R, if we keep all voltages the same, and if we change all magnetic fields in the ratio $1/R$, electron current will remain the same (provided the cathode is still capable of giving space charge limited emission). Electron paths will remain exactly similar, though smaller; the power into the electron beam will remain the same, but what will happen to the power dissipation capabilities of the device and what will happen to the temperature?

In a device cooled by radiation alone and with cool surroundings, the radiating area varies as R^2, and since the radiation per unit area varies as T^4, the temperature will vary as $R^{-\frac{1}{2}}$.

In considering a case of cooling by conduction alone, think of a rod carrying a certain amount of power away. If all the dimensions of a rod are changed by a factor R, the length will be changed by a factor R, the cross sectional area will change by a factor R^2, and if the thermal conductivity

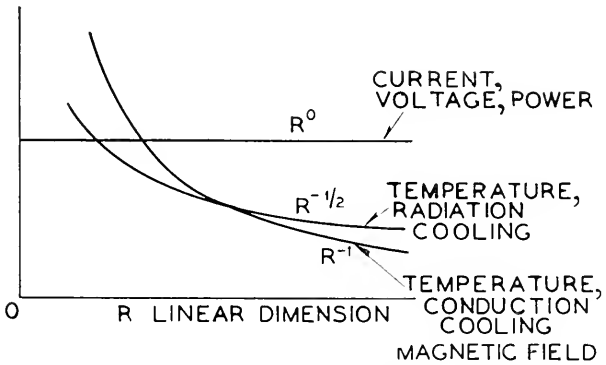


Fig. 12—Variation of magnetic field and temperature in scaling an electronic device.

remains constant the temperature will vary as R^{-1}. This is a faster rate of variation than in the case of cooling by radiation, and hence as the system is scaled to a smaller and smaller size, cooling by conduction will become negligible and radiation cooling only will remain effective and will determine the temperature.

Figure 12 gives an idea of the variation of various quantities discussed.

We want to make electronic devices smaller for a number of reasons; perhaps chiefly to reduce transit time and so to secure operation at higher frequencies. In doing this, we encounter the fundamental limitation of reduced power dissipation capabilities and increased temperature. What is the trouble? We have scaled everything. Or have we? The answer is, we have not. The electrons, atoms, and quanta are still the same size. Had we been able to scale these, we should have increased the heat conductivity and the radiating power of our device, and all would have been

well. As it is, if we make a tube for given power smaller and smaller, using the most refractory materials available we eventually reach a size of tube which will, despite our best efforts, melt, thaw, and resolve itself into a dew.

CONCLUSION

Perhaps after these somewhat gloomy words concerning physical limitations in electron ballistics, you may wonder how it is at all possible to surmount the difficulties mentioned. It certainly is not easy; all electronic devices represent compromises of one sort or another between fundamental physical limitations of electron flow on the one hand and structural complications on the other. In working with vacuum tubes one is perhaps troubled more by physical limitations, difficulties of construction, inadequacy of materials and the lack of quantitative agreement between complicated phenomena and relatively simple theories than in any other part of the electric art. It is for this reason that a friend of mine twisted an old aphorism into a new one and said, "Nature abhors a vacuum tube".

REFERENCES

Electron Microscopes

1. James Hillier and A. W. Vance: "Recent Developments in the Electron Microscope," *Proc. I.R.E.* 29, pp. 167–176, April, 1941.
2. L. Marton and R. G. E. Hutter: "The Transmission Type of Electron Microscope and Its Optics," *Proc. I.R.E.* 32, pp. 3–11, Jan. 1944.

Thermal Velocities

3. D. B. Langmuir: "Theoretical Limitations of Cathode-Ray Tubes," *Proc. I.R.E* 25, pp. 977–991, Aug., 1937.
4. J. R. Pierce: "Limiting Current Densities in Electron Beams," *Jour. App. Phys.*, 10, pp. 715–724, Oct., 1939.
5. J. R. Pierce: "After Acceleration and Deflection," *Proc. I.R.E.* 29, pp. 28–31, Jan., 1941.
6. R. R. Law: "Factors Governing the Performance of Electron Guns in Cathode-Ray Tubes," *Proc. I.R.E.* 30, pp. 103–105, Feb., 1942.
7. J. R. Pierce: "Theoretical Limitation to Transconductance in Certain Types of Vacuum Tubes," *Proc. I.R.E.* 31, pp. 657–663, Dec., 1943.

Space Charge

8. C. E. Fay, A. L. Samuel and W. Shockley: "On the Theory of Space Charge Between Parallel Plane Electrodes," *Bell Sys. Tech. Jour.* 17, pp. 49–79, 1938.
9. I. Langmuir and K. Blodgett: "Currents Limited by Space Chargeb etween Coaxial Cylinders," *Phys. Rev.* 22, pp. 347–356, 1923.
10. I. Langmuir and K. B. Blodgett: "Currents Limited by Space Charge between Concentric Spheres," *Phys. Rev.* 24, pp. 49–59, 1924.
11. J. R. Pierce: "Rectilinear Electron Flow in Beams," *Jour. of App. Phys.* 11, pp. 548–554, Aug., 1940.
12. B. J. Thompson and L. B. Headrick: "Space Charge Limitations on the Focus of Electron Beams," *Proc. I.R.E.*, 28, pp. 318–324, July, 1940.
13. C. J. Calbick: "Energy Distribution of Electrons within Dense Electron Beams," *Bull. Am. Phys. Soc.*, 19, No. 2, p. 14 (April 28, 1944).
14. J. R. Pierce: "Limiting Stable Current in Electron Beams in the Presence of Ions," *Jour. App. Phys.* 15, No. 10, pp. 721–726 (1944).
15. A. L. Samuel, "Some Notes on the Design of Electron Guns," Proc. IRE 33, pp. 233–240, April, 1945.

Electron Ballistics in High-Frequency Fields*

By A. L. SAMUEL

THIS, the final lecture of a series on Electron Ballistics, is not a summary of the material which has been previously presented but rather it is an attempt to show how the ballistic approach can be extended to the analysis of high-frequency devices. Much that might otherwise be said about ultra-high frequencies cannot be said because of secrecy requirements. However, there is considerable material which can be presented, within the limits of the necessary security regulations, which may be of interest to those who are not already well acquainted with the subject. I will, perforce, not be able to say anything specific about actual devices utilizing the principles to be discussed.

Many of the ultra-high-frequency devices which have come into use during the last few years have employed electron beams of one sort or another. These devices can be analysed in any one of a number of ways. For example, we can write the equation of space-charge flow. This approach considers the electric charge as a continuous fluid subject to Poisson's equation. The small-signal theory of Peterson and Llewellyn is an example of this type of analysis. Or if we wish we can consider the various types of wave motion which can exist in a space-charge region. The space-charge-wave analysis of Hahn and Ramo as applied to velocity-variation tubes is an example of this. In addition there is an electron-ballistic approach to the problem and it is with this method that we will be concerned in the present lecture.

Before we become involved in the details of the analysis, we should perhaps spend a few moments considering the relationship between these various methods. If we have an interaction taking place between electric fields and moving charges, we know at once from Newton's second law that the forces acting on the electrons must of necessity be equal and opposite to those acting on the fields. It is therefore a matter of small concern whether we consider the forces acting on the electrons and the effects of these forces on the electron motion or whether we consider the alteration in fields which the electron motion produces. We can, if we wish, compute the energy transfer to an electric field by the motion of an electric charge or we can compute the change in energy of the electron which accompanies this trans-

* Originally presented on April 11, 1945 as the concluding lecture of a symposium on Electron Ballistics sponsored by the Basic Science Group of the American Institute of Electrical Engineers.

fer. I was tempted to say "which results from this transfer" but this implies a cause and an effect, a notion which has no place in the present discussion. The dual aspect of any energy-transfer problem must always be kept in mind. Much needless discussion frequently arises between proponents of one point of view and those preferring the other when the only difference is one of language and both groups are really saying the same thing. The electron-ballistic approach yields a simple physical picture; it is capable of being applied to widely differing situations, but it is not well suited for a determination of the reactive contributions of an electron stream.

Basic Concepts

There are several concepts which we will find useful in our analysis. These concepts are extremely simple, so simple in fact that one is tempted to assume that they are well known. However, these concepts are so basic to the subject, and their results so far reaching that we must pause to consider them.

The first is the concept of total current, as distinguished from its components. One way of writing Kirchhoff's second law is

$$\text{Div. } J = 0 \tag{1}$$

This simply says that the total current entering or leaving any differential region in space is zero. This expression must of course be generalized by including displacement currents as proposed by Maxwell if applied to alternating currents. The current J is the total current density as here defined. An important consequence of equation (1), actually only an alternate way of stating it, is that the total current always exists in closed paths. Let us take a simple case of a two-element thermionic vacuum tube connected to a battery. Visualize the situation existing if but a single electron leaves the cathode and travels to the plate. The electron takes a finite time to cross from the cathode to the plate. During this time a current exists, the magnitude being given by the relationship

$$I = ev$$

and according to our premise this current is the same in every part of the circuit. The current begins at the instant that the electron leaves the cathode and it ceases when the electron arrives at the plate. In the apparently empty region ahead of the electron there must exist a displacement component, numerically equal to the conduction, or perhaps we should say convection component accounted for by the moving electron. An ammeter, were there one sufficiently sensitive and fast, connected in the external leads would read a current during this same interval of time.

I have chosen to talk about but a single electron to emphasize the electron-

ballistic aspect; however, the concept is much broader than this since it is not at all dependent upon a corpuscular concept of the electron. As a result of this property of the total current, the current to any electrode within a vacuum tube does not necessarily bear any relationship to the number of electrons which enter or leave it. Obviously then, currents can exist in the grid circuit of a three-element tube even though none of the electrons are actually intercepted by the grid. This current may have any phase relationship to an impressed voltage on the grid so that the grid may draw power from the external circuit, or it may deliver power to the external circuit, all without actually intercepting any electronic current. The grid-current component resulting from the electronic flow between cathode and plate may equally well bear a quadrature relationship to the impressed voltage, in which case it will either increase or decrease the apparent interelectrode capacitance. If these effects seem queer it is because one is still confusing the electronic component with the total current.

A second basic concept once stated becomes self-evident. This is to the effect that the only one thing which we can do to an electron is to change its velocity, that is, if we are to confine ourselves to the classical concept of an electron. We can change its longitudinal velocity, that is, alter its speed but not its direction other than possibly to reverse it, or we can introduce a transverse component to its velocity, that is, alter its direction as well as its speed. Thought of in this light all electronic devices in which a control is exercised over an electron stream are velocity-modulated devices. It might be argued that one could equally well say that *all we can do is to change the electron's acceleration (derivative of velocity) or its position (integral of velocity).* The singling out of velocity is in a sense arbitrary. It does, however, have some very interesting ramifications.

I might digress for a moment to elaborate on this idea. Since some of the newer devices have been labeled velocity-modulation tubes, there is a perfectly understandable tendency on the part of the uninitiated to assume that these tubes differ from earlier known devices, such as, for example, the space-charge-control tubes, the Barkhausen tube or the magnetron in the fact that they employ velocity modulation. The real difference lies elsewhere as we shall see in a few moments. At the same time that these newer devices were introduced, there was introduced a new way of looking at something which is very old in the art. This newer viewpoint, to my way of thinking, constitutes a far greater fundamental contribution than do the specific devices which have received so much attention. The pioneers in this new approach: Heil and Heil, Bruche and Recknagel, the Varian Brothers, Hahn and Metcalf, to mention a few, and the many other workers who lost in the race to publish their independent contributions in this field—all of these people deserve the greatest of praise for their stimulating contributions

to our thinking. My only point in all this discussion is to emphasize that the basic method of acting on the electron stream has not really been changed at all. The entire matter is summarized in the original statement that the only thing which we can do to an electron is to change its velocity.

Before going on to the next aspect of the problem there is a closely related concept which should be mentioned. This concept is that a change in the component of the velocity of an electron along one space coordinate does not introduce components of velocity in directions orthogonal to the first. For example, if an electron beam is deflected by a transverse electric field, there will be no accompanying change in the longitudinal velocity. The difficulty in the way of doing this in a practical case has nothing to do with the concept but only with the problem of producing unidirectional fields. Analyses of deflecting field problems which ignore the longitudinal components of the fringing fields are apt to be wrong. The problem of high-frequency deflecting fields has been treated in great detail in the literature and frequently with more acrimony than accuracy.

One further note should be added at this point. In an earlier lecture it was pointed out that the magnetic effects of an electromagnetic field are in general very much smaller than the electric effects. We will not stop to prove that this is still true at the frequencies which now interest us but will accept it without further discussion.

For our next concept we leave electron flow for a moment and consider the fields within a resonant cavity. You may very properly object that this has nothing to do with electron ballistics, and indeed it does not. However, we will find it necessary to discuss problems involving cavity resonators, and a failure to understand some of the properties of these circuit elements can cause a great deal of trouble. There are two conflicting approaches to this problem which I will attempt to reconcile.

The physicist when first presented with the problem of a resonant cavity is inclined to say: *This is a boundary value problem. The solution consists in writing Maxwell's equations subject to the conditions that the tangential component of E must be zero along the conducting walls. While a scalar and a magnetic vector potential can be defined, the field is not related to the former in the simple manner used in electrostatic problems.*

The engineer, on the other hand, is inclined to say: *This looks like an extension of the usual resonant circuit. A capacitance exists between the top and bottom walls of the cavity; charging currents will flow through the single turn toroidal inductance formed by the side walls. I would like to know what voltage difference exists between the top and bottom walls, and what currents exists in the side walls.*

Now, actually, I am maligning both the physicist and the engineer by my statements; nevertheless, there are these two approaches. Which is cor-

rect? Well, they both are. It is not correct to speak of an electrostatic potential within a resonant cavity; nevertheless, we may and do talk about the voltage between the top and bottom of a resonant cavity. What do we mean? Simply the maximum instantaneous line integral of the electric field taken along some specified path. In any practical device utilizing electron beams we are naturally interested in the path taken by the electrons. The fact that the line integral is different for different paths is of no great concern. We are interested in but one of these paths. We shall therefore have occasion to talk about voltages in cavities but we must always remember what is meant, and we must never for one instant forget that this voltage is not unique but that it depends upon some assumed path.

The second peculiarity of this voltage must also be emphasized. The line integral must be taken at a specified instant in time. In effect one takes a photograph of the field at some instant in time and then at one's leisure performs the integration.

Now, of course, an electron when projected through such a cavity will perform yet another type of integration. The change in squared velocity of the electron as expressed in volts will be given by the line integral of the field encountered by the electron; that is, integrated not instantaneously but with the electron velocity. This is not a simple process, because the electron velocity is continuously being changed by the field interaction and therefore the velocity with which the integration is performed depends upon the integrated value of the field up to the point in question. This has nothing to do with the concept of voltage in a resonant cavity. The cavity voltage can, however, be considered as the maximum change in squared velocity expressed in volts which an electron could receive if its entrance velocity was very large so that the transit time was small compared with the period of the cavity field.

The four basic concepts which I have chosen to recall to your mind are, by way of summary: (1) the total current is the same in all parts of a circuit, that is div. $J = 0$; (2) the only way we can act on an electron is to change its velocity; (3) the changes in the velocity component of an electron along any one rectangular coordinate have no effect on the velocity components along any other coordinate; and (4) for convenience, a voltage can be defined in a resonant circuit as the line integral of the electric field taken along some prescribed path.

Transit Angle

Since we are to deal with the interaction of electrons and high-frequency fields, we frequently find it convenient to measure electron velocity not directly but in terms of the equivalent potential difference through which an electron must fall to obtain the velocity in question, and the unit of measure

will be a volt. Instead of measuring the time required for an electron to traverse any given distance in seconds, it is also convenient to use, as a unit of time, one radian of angle at the operating frequency. We frequently refer to the transit angle of an electron rather than the transit time, although both terms are used. In fact, we may on occasion measure distances in terms of transit angle, and this usage is extended to measure dimensions transverse to the direction of travel of the electron beam. When used in this fashion, we mean that the dimension in question is such that were an electron to be projected in this direction with a velocity equal to that of the electrons in the main beam, the high-frequency field would change through the stated number of radians during the transit time.

The Five Functions in an Electronic Device

With this preliminary discussion out of the way we can now answer the question which has probably been troubling quite a few of you. If the only thing we can do to an electron is to change its velocity, then in what basic way does the velocity-modulation tube differ from the conventional negative grid tube or from the magnetron?

Well, this is an involved story. If we are to make any use at all of an electron beam we must in general perform five distinct operations or functions. First we must produce the beam. Then we must impress a signal of some sort onto the beam. From what I have just said this can be done only by varying the velocities of the electrons contained in the beam. The third operation consists in converting this variation into a usable form. It is in this way that the diverse forms of electronic devices differ to the greatest degree. We will go into this matter in more detail shortly. The fourth operation consists in abstracting energy from the beam, and the final operation consists in collecting the spent electrons. While these operations are distinct from an analytical point of view, in many actual devices they are performed more or less simultaneously and more than one operation may be performed by certain portions of the tube structure. In fact, in some devices, for example in the space-charge-control tube, the confusion is so great as to make the separation seem rather forced. This very confusion may partly explain why vacuum-tube engineers who were steeped in the art were so slow to realize the advantages of this new way of looking at things which I will call the velocity-modulation concept.

By way of mental exercise in this new way of thinking let us see how we can analyze a simple space-charge-control triode. Well, first of all we have to identify the electron gun which produces the beam. The electrons most certainly come from the cathode, but where is the first accelerating electrode? Actually there isn't any unless we think of the combined d-c field resulting from the d-c potentials on the grid and plate as assisted by

the initial emission velocities as performing this function. The next function, that of varying the electron velocities, is performed by the grid which varies the potential gradient in the vicinity of the cathode and hence the velocity of the electrons as they approach a potential minimum or virtual cathode which is formed a short distance in front of the cathode by the action of space charge. This virtual cathode performs the third function, that of conversion, by sorting out the electrons and allowing only those electrons with emission velocities greater than some specific value to pass. This, then, is one of the conversion mechanisms which we will call virtual-cathode sorting. In this example the virtual cathode occurs very close to the real cathode but this is not always the case. The fourth function, that of utilization, is performed by allowing the sorted electrons to traverse an electromagnetic field between the virtual cathode and the plate. This operation is completed by the time the electrons have reached the plate. Of course in the triode the plate then performs the final operation, that of collecting the spent electrons and dissipating the remaining energy as heat. It should be clearly realized, however, that this last function need not necessarily be performed by the same electrode which provides the output field. Indeed the so-called inductive-output tube proposed by Haeff is a space-charge-control tube in which these two operations are separated.

Conversion Mechanisms

But now to get back to a cataloguing of the different kinds of conversion mechanisms. The first general type involves sorting. The first kind which we have mentioned is by virtual-cathode sorting. A second kind of sorting might involve deflecting the electron beam in proportion to its longitudinal velocity instead of reflecting or transmitting it. Various deflection tubes have been proposed from time to time using this mechanism. We shall be forced to neglect this phase of the problem this evening because of time limitations but those of you who are interested will find the literature filled with detailed discussions. Still a third type of sorting, sometimes called anode sorting, is used in certain Barkhausen tubes when the plate is operated at or near the cathode potential so that fast electrons are collected while slow electrons are reflected and caused to retraverse a high-frequency field. There are still other types of sorting mechanisms but I will not burden you with these.

A second general type of conversion mechanism I will call bunching, to distinguish *sorting* in which electrons are separated according to their velocities from *bunching* in which electrons of differing velocities are brought together. Now it just happens that many of the older devices used sorting, while many of the newer devices use bunching but this is not universally the case. For example, the magnetron as used at high frequencies and the

cyclotron both employ a combination of sorting and bunching. A peculiar property of the motion of an electron in a magnetic field lies in the existence of the so called Larmor frequency. You will recall that the angular velocity of an electron in a magnetic field depends only upon the field-strength and not at all upon the electron's linear velocity. This time in seconds is given by

$$t = \frac{0.357 \times 10^{-6}}{H},$$

or in radians

$$\theta = 2\pi \frac{10600}{\lambda H}.$$

Electrons of widely differing velocity can thus revolve together in spoke-like bunches with the faster electrons going around larger circles than the slow ones, but just enough larger to keep them together. This, then, is one kind of bunching, which for simplicity we shall call magnetic bunching. It is used in the magnetron and in the cyclotron. We will have more to say on this subject a little later.

A second kind of bunching was used in some of the early Barkhausen tubes where the plate electrode was operated at a fairly high negative potential so that none of the electrons were able to reach it. Under such conditions a uniformly spaced stream of electrons with varying velocities is reflected as a bunched stream, the slower electrons being reflected almost at once and the faster electrons penetrating the retarding field for a greater distance and hence taking longer to return. This same type of bunching is used in a newer form of oscillator, commonly referred to as a reflex tube which was suggested by Hahn and Metcalf in 1939, and by others at about the same time. The reflex tube differs from the Barkhausen tube, not in the basic mechanisms so much as in the fact that the conversion mechanism occurs in a different region in the tube from the region devoted to velocity modulation and to energy abstraction. A second kind of bunching is then reflex bunching.

A third type of bunching was used in the diode oscillators of Muller and of Llewellyn. The mathematical research done by W. E. Benham may be mentioned as of interest in this connection. In these tubes a uniform stream of electrons becomes bunched simply through the fact that faster moving electrons overtake slower ones which precede them. In these earlier forms of tubes we again have the case where this conversion is performed simultaneously with one or more of the other processes so that it is very difficult to separate them. However, in 1935 Heil and Heil proposed a tube in which the conversion region was separated from the other regions of the

tube. This tube, the velocity-modulation tubes of Hahn and Metcalf, and the klystron tubes of the Varian Brothers, are alike in their use of transit-time bunching in a relatively-field-free drift tube. Since this separation of functions renders these devices much easier to analyze and since the structures are quite interesting in any case we will spend most of our time considering them and will, I fear, rather neglect some of the other types of tubes.

We will, of course, keep our analysis as general as possible so that the results may be applied to a variety of different devices.

Input Gap Analysis

Let us begin by a small-signal consideration of a uniform electron stream entering a region in which there is a longitudinal field defined as some function of the distance and of time. This can be the entire Llewellyn diode or it can be the input region of a klystron. We ask ourselves with what velocity will the electrons leave this region and what will be the net exchange of energy between the electrons and the field. At any point within the field a typical electron will experience an acceleration given by

$$\ddot{y} = \frac{1}{\xi} E + \eta f(y) f(t) \tag{1}$$

where η is proportional to the maximum amplitude of the h.f. field, but contains a numerical constant so that $\ddot{y}$ is expressed in centimeters per second per second. Now in the usual case $f(t)$ will be a simple sine function but $f(y)$ may assume a variety of forms. Again, by way of simplifying our work we will assume that it is also a sine function. Let us consider how we can go about solving this apparently simple equation. Unfortunately this expression can not be solved directly because the value of t at any plane (that is, the time of arrival of an electron at this plane) depends upon the interchange of energy between the electron and the field. Here we are forced back to the time-honored mathematical device of assuming a solution in the form of a series and then evaluating these coefficients. There is a large number of ways in which this can be done, and consequently a large number of different solutions which look very different but which all give comparable answers. Usually when such solutions are published, the arithmetical work is omitted leaving one with the feeling that there is something involved that is not within the ken of ordinary mortals. The fact is that the work is usually extremely tedious but actually very simple. It will be instructive to follow through one form of such an analysis in just enough detail to see the amount of work involved.

Since we are interested in the energy which is proportional to $\dot{y}^2$ we will write at once

$$(\dot{y}^2)_{y=a} = K = K_0 + \eta K_1 + \eta^2 K_2 + \eta^3 K_3 + \ldots$$

where the K's are a function of the transit time, of the field distribution and of the entrance phase, and we will proceed to evaluate these coefficients. The average energy per unit of change as expressed in volts is then simply

$\frac{\xi\bar{K}}{2}$ at the end of the field while the gain is:

$$V_{av} = (\xi/2)(\bar{K} - \bar{K}_0) = \frac{\eta\xi\bar{K}_1}{2} + \frac{\eta^2\xi\bar{K}_2}{2} + \cdots$$

where the bar means that we are averaging over all values of the entrance phase.

It is of interest to evaluate the value of velocity $\dot{y}^2$ which individual electrons receive as a function of the entrance phase. For small signals it is usually sufficient to evaluate $\dot{y}^2$ maximized with respect to the starting phase, then

$$V_{max} = (\xi/2)(K - K_0)_{max} = \left[\frac{\eta\xi K_1}{2} + \frac{\eta^2\xi K_2}{2} + \cdots\right]\max.$$

We can further define the ratio of V_{max} to the largest value it can have as a coefficient β, sometimes called the modulation coefficient.

But now to evaluate the K's. There are many ways of doing this as I have intimated. We will proceed by writing

$$y = y_0(t) + \eta y_1(t) + \eta^2 y_2(t) + \eta_3 y_3(t) + \ldots$$

where the y's are coefficients depending upon the transit time t which in itself is a function of the applied field thus

$$t = t_0 + \eta t_1 + \eta^2 t_2 + \eta^3 t_3 + \ldots.$$

We can then expand each function of time into a series remembering that

$$f(x + d) = f(x) + \frac{f'(x)\,d}{1!} + \frac{f''(x)\,d^2}{2!} \cdots$$

or for our particular case

$$y_0(t) = y_0(t_0) + \frac{\ddot{y}_0(t_0)[\eta t_1 + \eta^2 t_2 + \eta^3 t_3 + \cdots]}{1!} + \frac{\ddot{y}_0(t_0)[\eta t_1 + \eta^2 t_2 + \eta^3 t_3 + \cdots]^2}{2!} + \cdots$$

Now we can expand $y_1(t)$, $y_2(t)$ etc. in exactly the same way. Finally we get a collection of terms which can be grouped in like powers of η thus

$$y = y_0(t_0) + \eta\,[\text{terms in } \dot{y},\ \ddot{y},\ t_1,\ t_2,\ \text{etc.}] + \eta^2\,[\quad] \ldots$$

The coefficient of the η is in fact $\dot{y}_0(t_0)\,t_1 + y_1(t_0)$. We will not bother to write the rest. This expression can then be differentiated to get $\dot{y}$ and then

squared. However, we still have some undetermined coefficients the t_1, t_2 etc. terms. These we can evaluate by noting that we wish these values at $y = a$, where a is a fixed distance in the actual device. At this distance the t coefficients in the expression for y must have such values that the value of y does not change with the value of η. This can only be true if the individual expressions multiplying each power of η are each equal to zero. Equating these expressions to zero one can evaluate all of the t's. For example the first term yields

$$\dot{y}_0(t_0)t_1 + y_1(t_0) = 0$$

or

$$t_1 = -\frac{y_1(t_0)}{\dot{y}_0(t_0)}.$$

Introducing these values, differentiating and squaring, one finally gets an expression for $(\dot{y}^2)y = 0$ as a power series in y, the coefficients all being of a form easily evaluated for any specified field distribution. Since we have by definition called these coefficients K_0, K_1, etc. these values are then

$$K_0 = \dot{y}_0^2$$

$$K_1 = 2(\dot{y}_0\ddot{y} - y_0\ddot{y}_1)$$

$$K_2 = (\ddot{y}_1^2 - 2y_1\ddot{y}_1 + 2\ddot{y}_0\ddot{y}_2) - 2\ddot{y}_0 y_2 + \frac{\dddot{y}_0 y_1^2}{\dot{y}_0}$$

This then constitutes the formal solution of the problem. We must now particularize our problem to some specific field distribution and evaluate the y coefficients. Suppose, for example, that there is a uniform d.c. field (E of equation 1) and an alternating field which varies as some cosine function of distance. Then the latter is

$$f(y) = \cos\left(\frac{\pi y}{b} + c\right)$$

and

$$\ddot{y} = \frac{1}{\xi} E + \eta \cos(\omega t + \varphi) \cos\left(\frac{\pi y}{b} + c\right)$$

we must eliminate the y which appears in this expression and replace y by its equivalent

$$y = y_0 + \eta y_1 + \eta^2 y_2 + \ldots$$

and expanding

$$\cos\left(\frac{\pi y}{b} + c\right) = \cos\left(\frac{\pi y_0}{b} + c\right) + \frac{\frac{\pi}{b}\sin\left(\frac{\pi y_0}{b} + c\right)[\eta y_1 + \eta^2 y_2 + \cdots] + \cdots}{1!}$$

and as before equating like powers of η with $\ddot{y}$ defined as

$$\ddot{y} = \ddot{y}_0 + \eta\,\ddot{y}_1 + \eta^2\ddot{y}_2 + \cdots$$

we finally arrive at

$$\ddot{y}_0 = \frac{1}{\xi}E$$

$$\ddot{y}_1 = \cos(\omega t + \varphi)\cos\left(\frac{\pi y_0}{b} + c\right)$$

$$\ddot{y}_2 = -y_1\pi/b\cos(\omega t + \varphi)\sin\left(\frac{\pi y_0}{b} + c\right).$$

Now we need only integrate these expressions to obtain the values of the $\dot{y}$'s and the y's needed to evaluate the K's.

If we average $\dot{y}^2$ over all values of the starting phase we can write the energy contributed by the field to the electron's velocity. When this is done one finds that the odd powers of η are identically zero leaving only the even powers to be considered and for small signal analysis purposes we need only consider $\bar{K}_2$. The energy per electron expressed in volts is

$$V = 2.49 \times 10^{-8}E^2\lambda^2 f(\theta)$$

where $f(\theta) = \omega^2\bar{K}_2$, and the power is obtained by multiplying this expression by the beam current in amperes.

The end results can be expressed as curves of $f(\theta)$ against θ as shown in Fig. 1. Three examples are shown: the uniform field case and two different harmonic distributions as indicated by the smaller plot in the lower left-hand corner. You will note that there exist regions of positive $f(\theta)$ where the net transfer of energy is from the field to the electron and regions in which the transfer is in the other direction; the former portions are of considerable interest in connection with the input gaps in velocity modulation tubes, and for that matter in the cathode grid region of the negative grid tube although this is more complicated than is here indicated, as this transfer of energy constitutes a loss to the field which loads the input circuit. The latter portions may be utilized as was done in the Muller and Llewellyn diodes to obtain sustained oscillations.

If, as I have indicated, we maximize $\dot{y}^2$ as a function of the starting phase we can evaluate the modulation coefficient. The value for the uniform field

case, as shown in Fig. 2, is simply, $\beta = \dfrac{\sin \theta/2}{\theta/2}$. For future reference we will write the loss expression for this case as

$$f(\theta) = 2\ (1 - \cos \theta) - \theta \sin \theta.$$

Drift Space Analysis

Now let us consider the conversion region in a typical velocity-variation tube. Figure 3 is a drawing of several such devices with the conversion

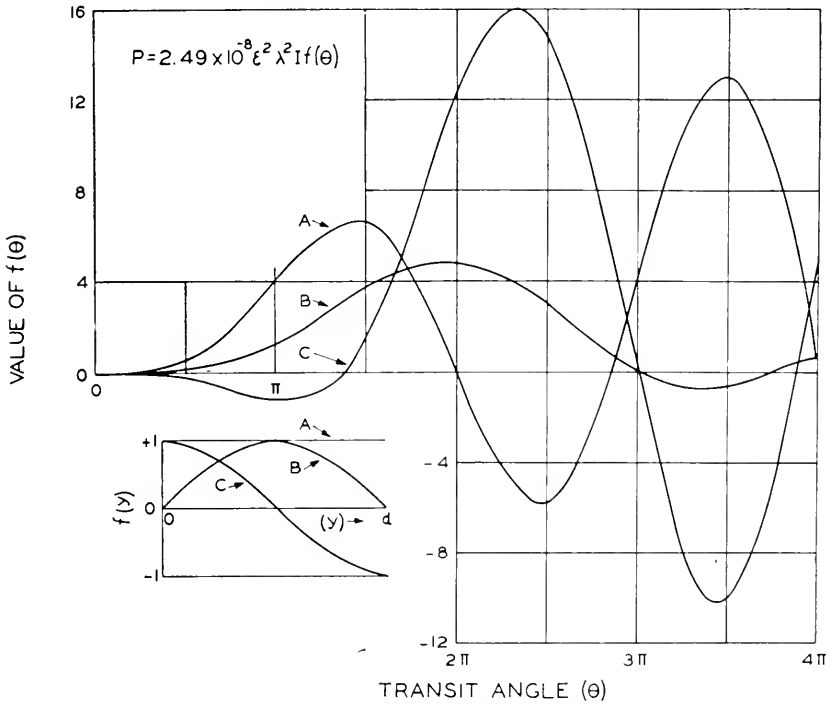


Fig. 1—The energy transfer between an initially uniform electron stream and a longitudinal electromagnetic field as a function of transit angle.

regions indicated. We will assume for the moment that the electrons enter this region with a small variation in velocity and at a perfectly uniform rate. Since the total number of electrons entering the region must be equal to the number of electrons leaving the region we may write

$$i_1\, dt_1 = i_0 dt_0$$

or

$$i_1 = i_0 \frac{dt_0}{dt_1}.$$

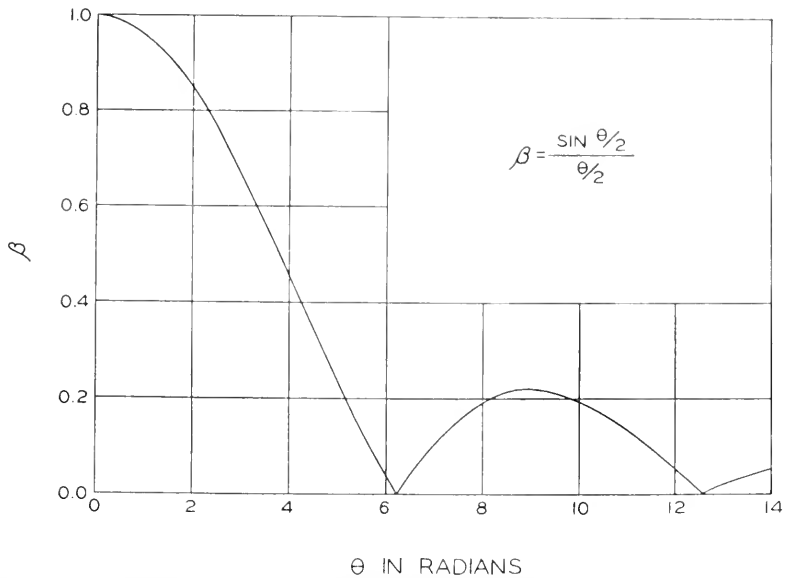
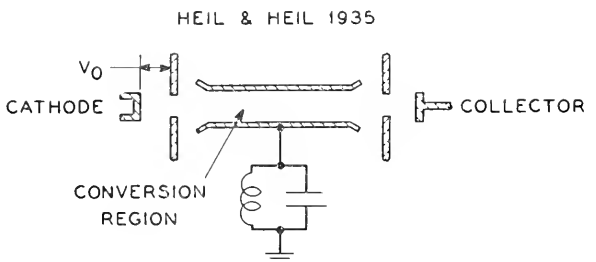


Fig. 2—The (velocity) modulation coefficient between an initially uniform electron stream and a uniform electromagnetic field as a function of transit angle.



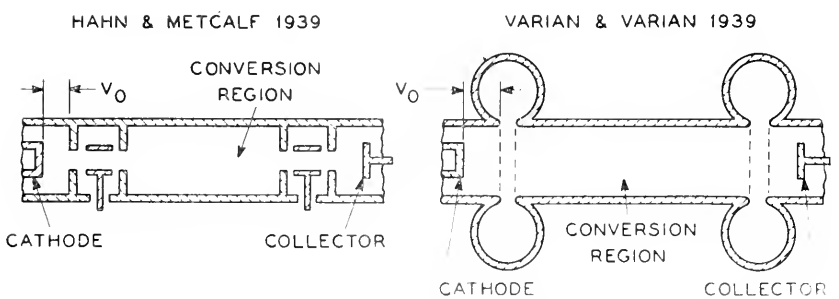


Fig. 3—Typical velocity variation devices employing transit-time bunching.

However, a relationship exists between t_1 and t_0,

$$t_1 = t_0 + \frac{\ell}{v}.$$

Where

$$v = v_0 \sqrt{1 + \alpha \sin \omega t_1},$$

$$t_1 = t_0 + \frac{\ell}{v_0 \sqrt{1 + \alpha \sin \omega t_1}}.$$

Now if $\alpha \ll 1$

$$t_1 = t_0 + \frac{\ell}{v_0}\left(1 - \frac{\alpha}{2} \sin \omega t_1 + \cdots\right)$$

and

$$\frac{dt_0}{dt_1} = 1 + \frac{\ell}{v_0}\frac{\alpha\omega}{2} \cos \omega t_1$$

and finally

$$i_1 = i_0\left(1 + \frac{\ell\alpha\omega}{v_0 \cdot 2} \cos \omega t_1\right)$$

but

$$\frac{\ell\omega}{v_0} = \theta$$

so that finally

$$i_1 = i_0\left(1 + \frac{\alpha\theta}{2} \cos \omega t_1\right).$$

This says that the velocity variation impressed on the beam at the entrance to the drift space or conversion region has resulted in a current variation at the output. For those of you who think in vacuum tube parameters it is of interest to differentiate this expression with respect to the a-c voltage and obtain the transconductance

$$G_m = \left|\frac{di_1}{dV_{a\cdot c\cdot}}\right|$$

rewriting

$$|i| = i_0\left(1 + \frac{V_{ac}\theta}{2V}\right)$$

$$\left|\frac{di_1}{dV_{ac}}\right| = \frac{\theta i_0}{2V}.$$

This result is obtained by neglecting all of the higher order terms and is therefore only a small signal theory of a very restricted sort.

Now let us consider what we have done. Well, we have followed a small interval of time through the drift tube. At the input this time dt_0 had a current i_0 associated with it; at the output the size of this unit of time is different—it is now dt_1 and the current associated with it is i_1. The physical picture corresponding to this phenomenon is that of a uniform distribution of electric charge becoming bunched with time as it traverses the drift space.

The next step in the analysis is to carry our approximation a step further and consider higher-order terms. Expanding the expression for i_1 and using our nomenclature the desired expression is

$$i_1 = i_0\left[1 + 2\left(J_1\left(\frac{\alpha\theta}{2}\right)\cos\omega t + J_2\left(2\frac{\alpha\theta}{2}\right)\cos 2\omega t + J_n\left(n\frac{\alpha\theta}{2}\right)\cos n\omega t\right)\right].$$

This equation is not exact since it neglects space charge effects but it does indicate the presence of harmonics in the beam current and it reveals certain non-linear effects which can also be illustrated by the so-called phase-focusing diagrams of Bruche and Rechnagel.

Phase-Focusing Diagrams

Bruche and Recknagel pointed out that an analogy exists between the focusing in space of a parallel light beam and the focusing in phase of the electrons in a uniform electron beam. In fact a small-signal theory can be developed entirely in terms of optical equations. We will not go into this aspect in detail but we will use their diagram (Fig. 4) to illustrate the bunching effect graphically. A uniform beam of electrons is represented by a series of parallel lines in distance and time coordinates, focus being indicated by a crossing of these lines after they have been deflected by the velocity modulation.

This general type of diagram has been popularized in this country by the Varians, and their associates under the name Applegate diagram, the only difference being an interchange of axis. Figure 5, taken from a recent paper by Dr. A. E. Harrison, illustrates this version of the Bruche and Recknagel diagram.

Now if instead of judging the current density by the density of the lines on the diagram, we make a plot of the current density as a function of time for different fixed distances from the input gap, the pictures are somewhat as shown on Fig. 6. Figure 7 represents a plot presented by Kompfner and combines in one illustration the type of presentation used by Tombs.

Phase Focusing in a Reflex Tube

It might be well to pause for a moment in our discussion of transit time bunching to consider how the phase focusing diagrams can be applied to a reflex tube. The elements of a modern reflex tube are shown in Fig. 8

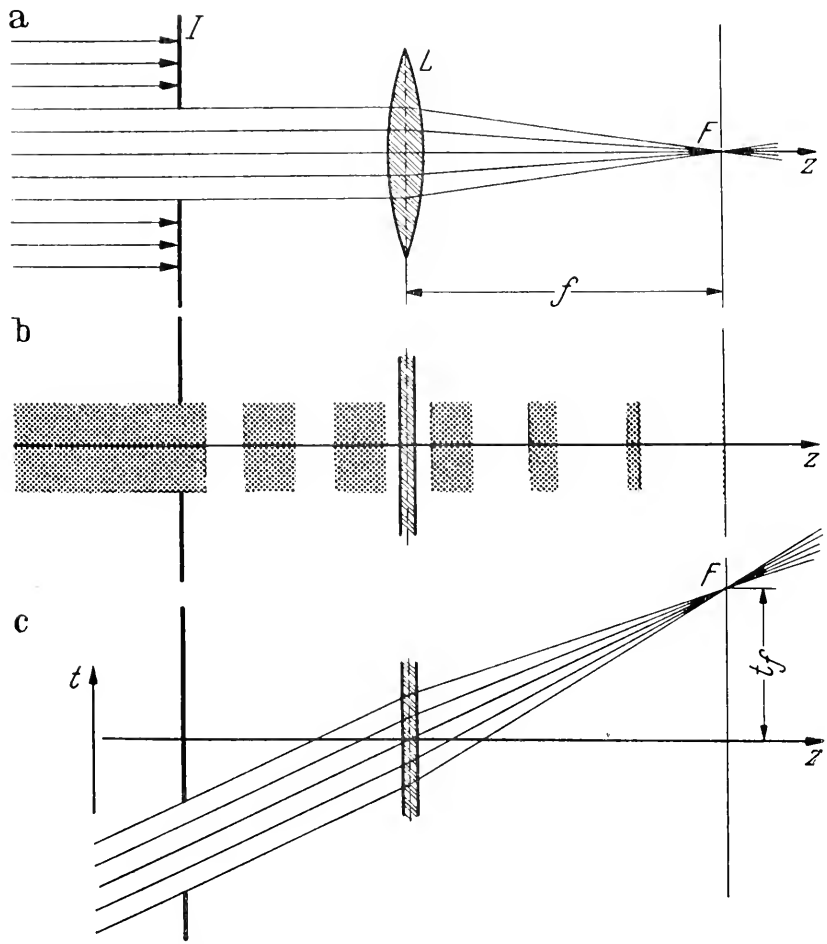


Fig. 4—The phase-focusing diagram of Bruche and Recknagel showing the analogy to optical focusing.

which was taken from a recent I.R.E. paper by Dr. J. R. Pierce. Electrons from the cathode pass through an input gap defined by two grids where they are modulated in velocity. In traveling in the retarding field produced by the repeller those electrons which passed the gap when the field was becoming progressively less accelerated, become bunched; the faster electrons

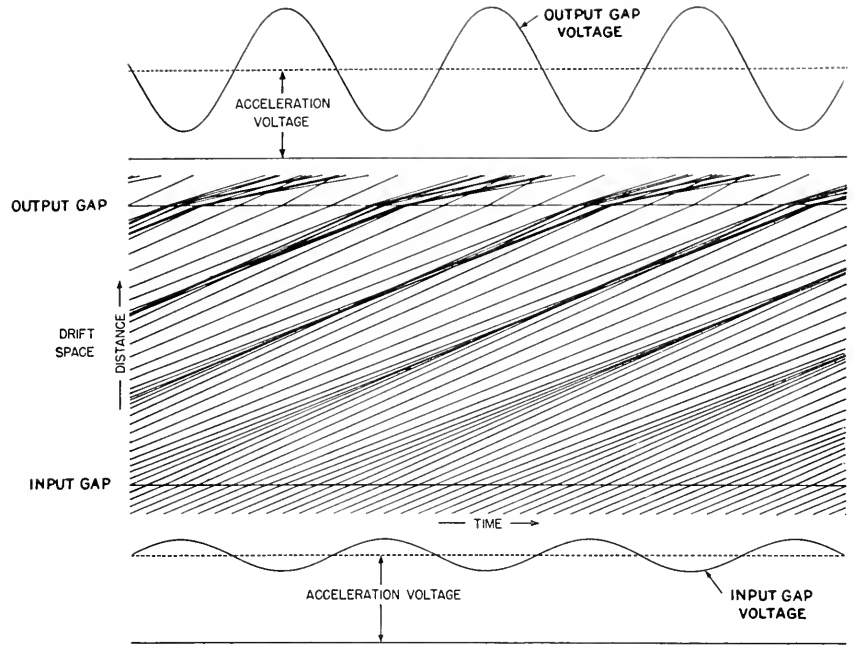


Fig. 5—Applegate's version of the phase-focusing diagram (Harrison).

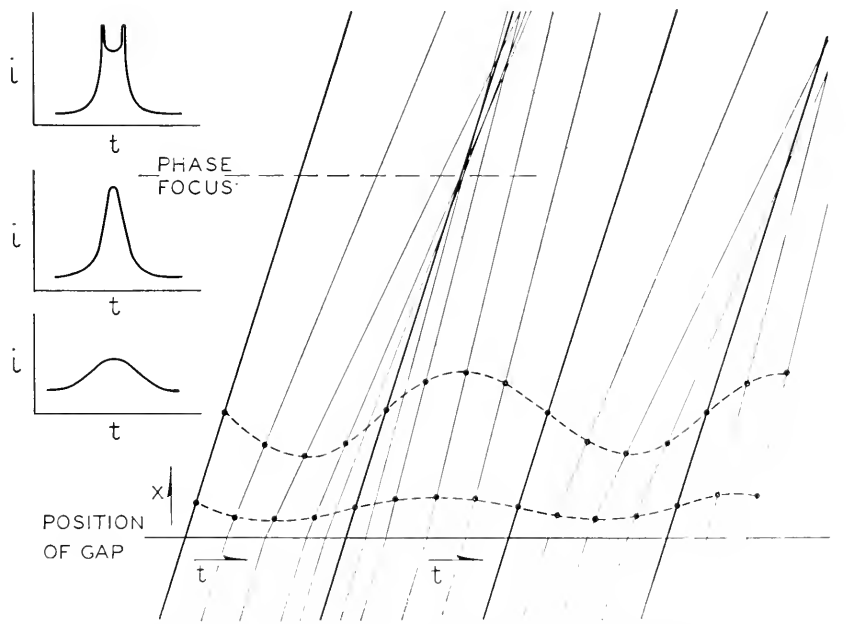


Fig. 6—The conduction-current distribution at different distances along the beam as predicted by the phase-focusing diagram.

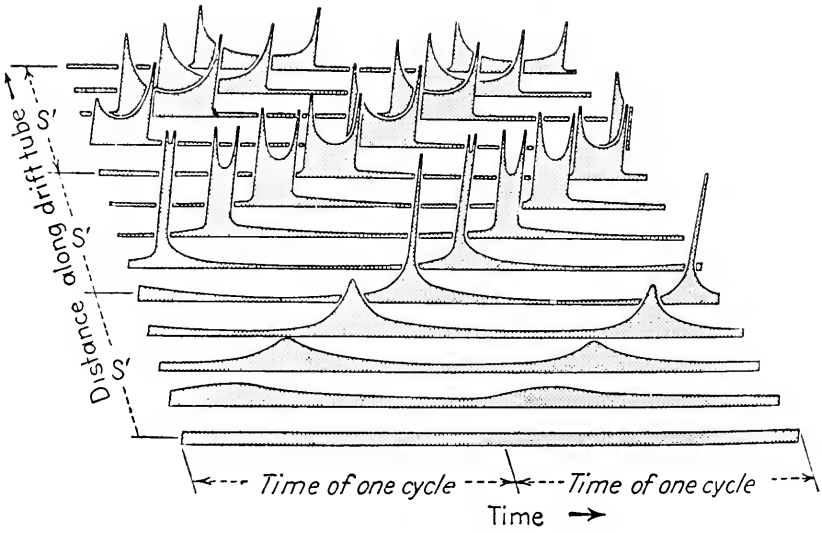


Fig. 7—Kompfner's presentation of the bunching effect.

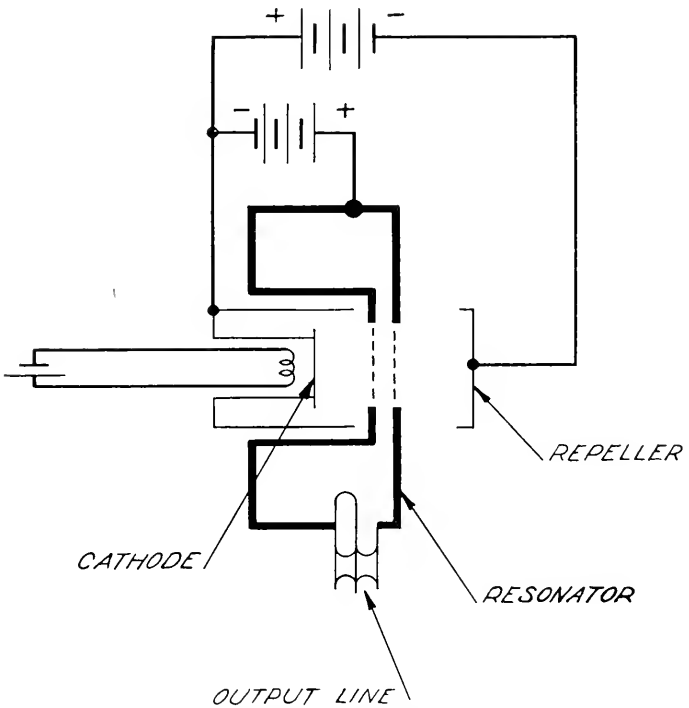


Fig. 8—The elements of a modern reflex tube (Pierce).

penetrating the field to a greater extent and waiting, as it were, for the slower electrons which follow to catch up. The electrons which pass across the gap while the field is becoming progressively more accelerating are spread out. If the retarding field is uniform it can be likened to the earth's gravitational field and the phase-focusing paths on our time-distance plot are parabolas. Figure 9, taken from Pierce's paper, illustrates this while Fig. 10 is such a plot taken from the paper by Harrison. One interesting and,

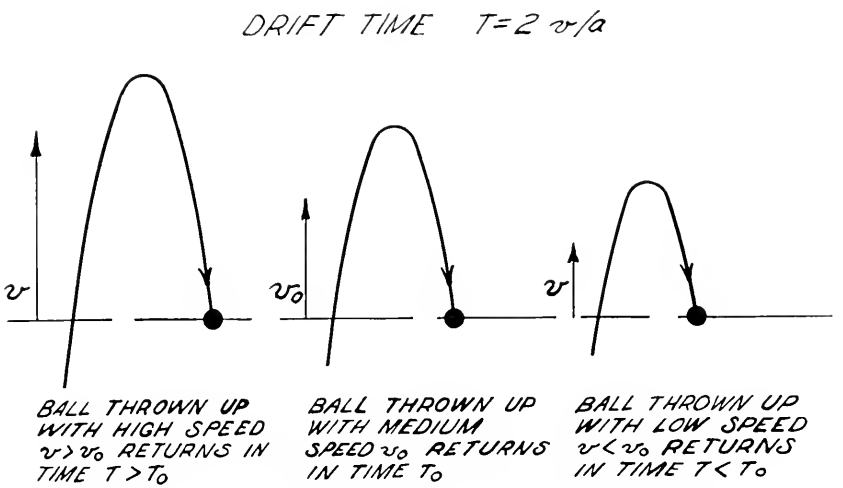


Fig. 9—The gravitational-field analogy to reflex bunching (Pierce).

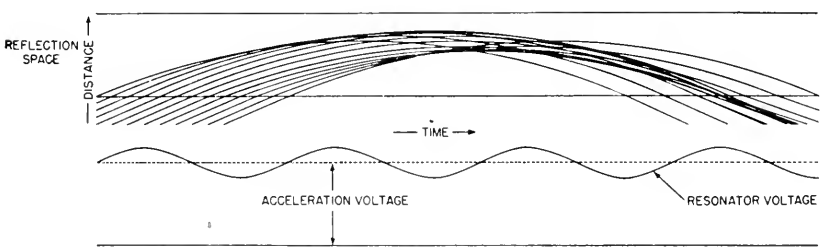


Fig. 10—The phase-focusing diagram for a reflex oscillator (Harrison).

in a way, unfortunate difference between reflection bunching and direct transit-time bunching is the fact that for reflection bunching the slow electrons catch up with the fast ones while the reverse is true for the other type. This means that if both types of bunching are present as shown in Fig. 11, (also taken from Harrison's paper) one will tend to undo the effect of the other.

Another way of combining effects of separate bunching actions is to build

a cascade transit-time-bunching amplifier in which a series of three gaps is used together with two drift spaces. The first gap velocity modulates the beam; this modulation is converted into a current modulation in the first drift space. The beam then excites the second cavity, which again velocity modulates the beam in quadrature with the original modulation. This action of course occurs in the output gap of a two-gap tube but it is not there used. Here this second and larger velocity modulation is converted to current modulation in the second drift space. The output is finally taken off the beam by the third gap. A phase-focusing diagram of this sort (again taken from Harrison's paper) is shown in Fig. 12.

Space-Charge-Wave Analysis

This phase-focusing approach is rather intriguing as one feels that one has a physical picture of what is going on. The picture is, however, very inexact except under certain highly specialized cases, as it completely ignores

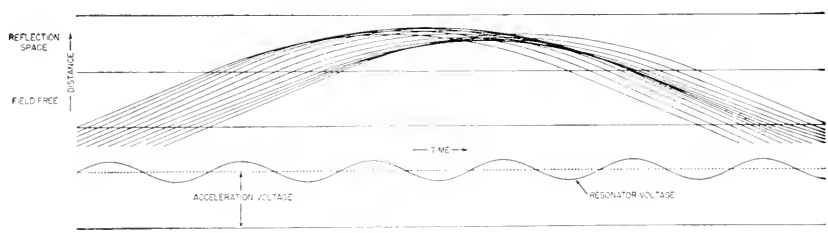


Fig. 11—Diagram showing reflex bunching combined with field-free transit-time bunching (Harrison).

space-charge effects. These space-charge effects are of two sorts: a d-c effect, if you will, and an r-f effect; that is, the presence of the electrons of the beam will alter the average velocity of the electrons at different parts of the beam, and will tend to undo the bunching action. Because of this second effect, the electrons are effectively prevented from passing each other as the graphical solution suggests. Instead, as the density of the electrons in the bunch becomes greater, the mutual repulsion forces tend to prevent a further concentration of charge. The electron bunch then tends to disperse. The action could be likened to the propagation of a sound wave in a moving column of air. While there are several approximate ways to handle this problem, Hahn was the first to propose a really satisfactory theory. Incidentally it should be noted that the Benham, Muller, Llewellyn and Peterson type of theory is capable of treating this aspect of the problem in a rigorous way and including all space-charge effects, but unfortunately these theories are limited in that they have been applied only to the parallel-plane case, and of course they are only small-signal theories.

Hahn's analysis starts by treating an infinitely long electron beam, using cylindrical co-ordinates and is limited to a small signal theory where the a-c motions are small compared to the d-c but it does not ignore the r-f effects of the space charge forces. The electron beam is thought of as a moving dielectric rod which is capable of propagating axial waves much as a dielec-

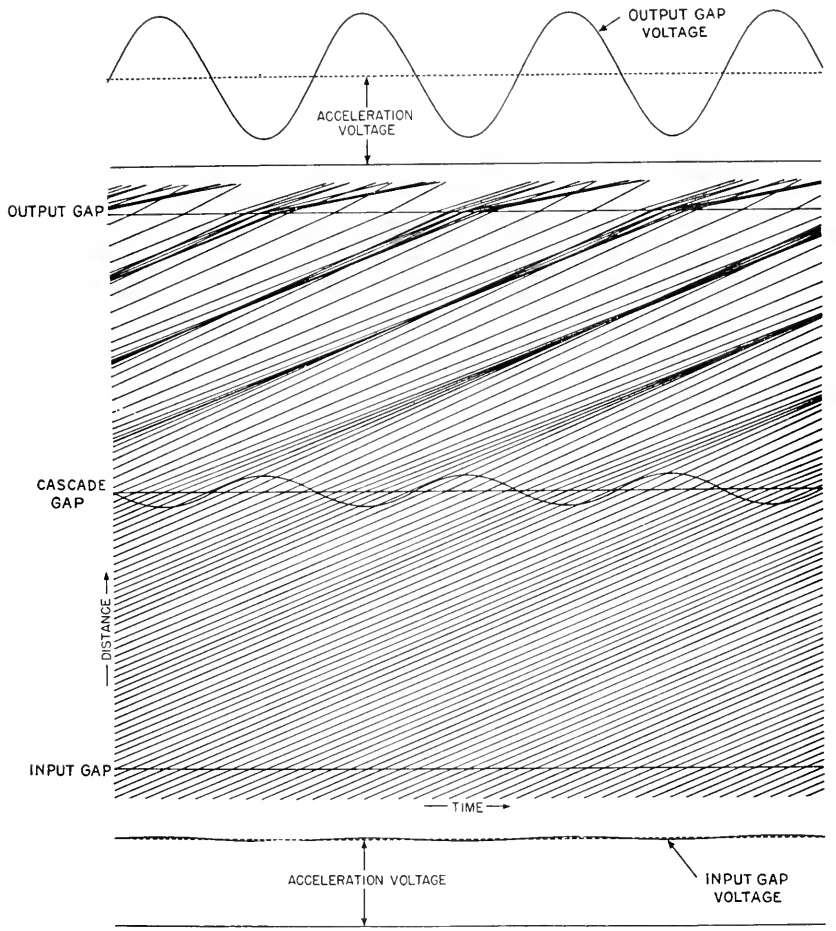


Fig. 12—Diagram for a cascade amplifier (Harrison).

tric wave guide will do. He assumes an axial magnetic field and a stream of positive ions having the same velocity axially and the same charge density. These ions are assumed to have infinite mass. The solution is much too complicated and involved to present here even in abstract. It involves the complete solution of Maxwell's equations subjected to the stated assump-

tions as restricted by the assumed boundary conditions at the edge of the beam.

It is found that two waves are possible, one traveling slightly faster than the electron beam and the second traveling slower. A point where the velocity components are in phase will correspond to the input to the beam, while points where the current components are in phase correspond to the desired positions for the output. The propagation constants for these two waves in a simplified special case where the magnetic field strength is infinite are given by Hahn, as well as expressions for the optimum drift tube length. He goes on to consider the case where the magnetic field is zero and finds that for this case the density of the charge does not vary much but instead the beam swells in and out so that instead of being lumps of charge with spaces between, the lumps appear in the outer boundary. Hahn has extended his general method of analysis to consider the modulation coefficient of gaps through which the beam must pass. His results are a great deal more general than those we have presented.

Ramo has reformulated Hahn's theory by means of retarded potentials for the most important case. This results in some simplification of the theory. He computes the more important design constants for a velocity modulated tube, such as the optimum drift tube length and the amount and phase of the transconductance. Those of you who are particularly interested are referred to the original paper. An interesting aspect brought out rather forcibly by Ramo's analysis is the existence of higher-order waves on the beam, always occurring in pairs, one faster and the other slower than the beam velocity.

The Magnetron

In what time remains I want to say just a very few words about the magnetron. This is a very complicated subject and one which cannot be adequately dealt with in an entire evening, and certainly not in the time remaining.

As you all know, the magnetron was invented and named by Dr. A. W. Hull. Habann, Zacek, Okabe and others pioneered in the use of the magnetron as an ultra-high-frequency oscillator. As envisioned today a magnetron is a two-element device, usually cylindrical with a centrally located cathode and a surrounding anode. The anode may be continuous or it may be split into a number of segments as suggested by Okabe, and these segments joined together either externally or internally by resonant circuits.

The basic ballistic problems of the magnetron, and hence the only problems which directly concern us at this time are (1) that of determining the

electron paths within the magnetron and having determined these paths (2) that of getting an understanding of the mechanism whereby electrons in traversing these paths are able to deliver energy to the connected high-frequency circuits. One might think that the first problem would be a relatively easy job. As a matter of fact the literature is surfeited with papers purporting to give the answer. Unfortunately almost all of the published work ignores the effect of space charge. A few moments' thought will suggest that space charge may be a controlling factor because of the long electron paths which are sure to result in crossed electric and magnetic fields, and indeed more detailed computations bear this out. Nevertheless the neglect of space charge greatly simplifies the problem. There are those who believe that the no-space-charge theories have no bearing on the way actual magnetrons work and that any correspondence between the predictions of such theories and the actual behavior of magnetrons is simply the result of an unfortunate coincidence. In fact Brillouin points out that the simplified form in which the Larmor theorem is applied by many, is in itself an approximation which was perfectly valid as originally applied by Larmor to the electronic orbits within the atom but which does not apply to conditions as they exist in the magnetron.

A number of recent workers have attempted to include the effects of space charge but have unfortunately largely restricted themselves to small signal theories while the magnetron is seldom operated under small signal conditions, at least not intentionally. Most theories are further restricted to a consideration either of the coaxial case where the cathode radius is small compared to the anode radius or of the plane case. Most practical structures are intermediate between these extremes.

As an example of the difficulties involved, Fig. 13, reproduced from a paper by Kilgore, shows the electron paths as computed neglecting space charge and also shows experimental proof that these paths actually exist. This illustration has been frequently reproduced and widely accepted. The experimental picture was obtained in the presence of gas, to make the electron beam path visible, and unfortunately the ionization which makes the beam visible also tends to neutralize space charge effects. The experimental arrangement departs still further from reality in that the electron emission from the cathode was restricted to a limited region so that the space charge forces were still further reduced. Now it is probably true that some magnetrons operate with electron paths as shown; still it is not true that all magnetrons operate in this way.

Contrasting with this picture which was until recently commonly accepted, Brillouin, Blewett and Ramo, and others have shown that stable distributions are possible in which a space charge of almost uniform density rotates with a uniform angular velocity about the axis. Brillouin goes so

far as to label the curves due to Kilgore as wrong, and pictures the possible electron trajectories as shown in Fig. 14.

One of the earliest papers to consider this newer picture of the electron paths in the magnetron was published by Posthumus in 1935. This was definitely a ballistic approach and hence suitable for discussing tonight.

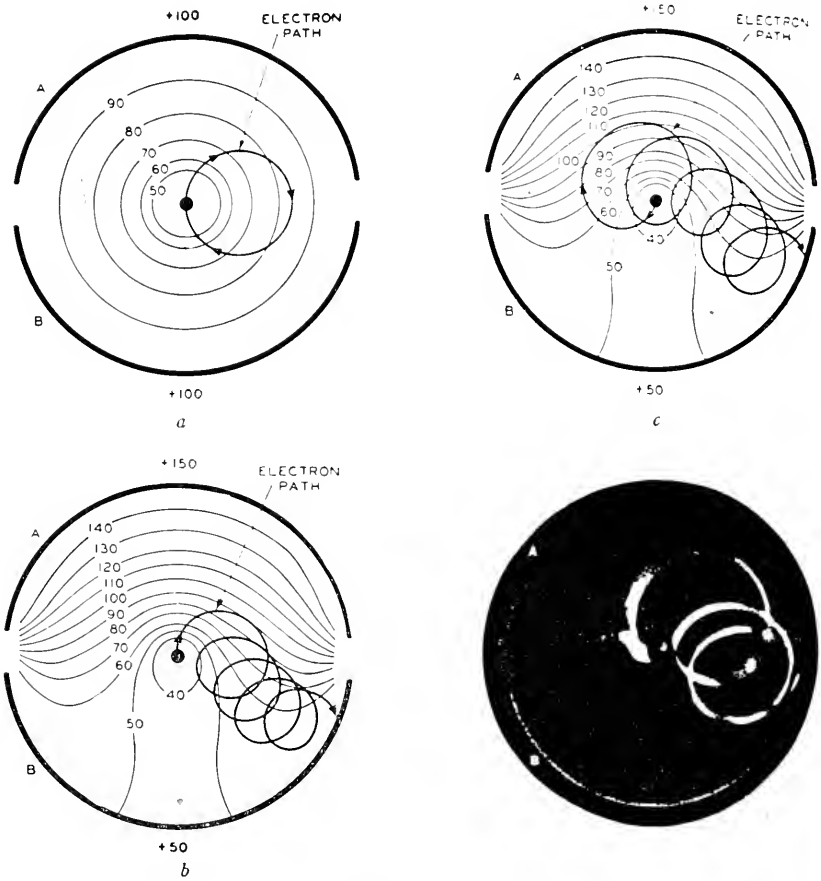


Fig. 13—Typical electron paths in a two-segment magnetron showing how electrons arrive at the plate-half of lower potential (Kilgore).

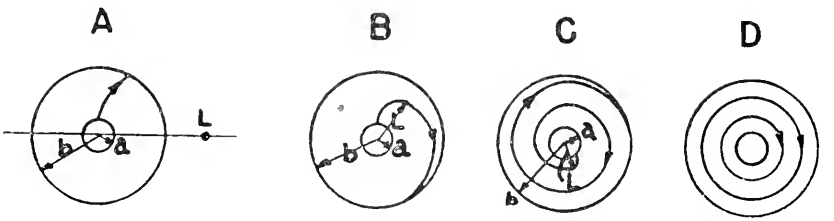
Posthumus limits his discussion to but one type of oscillation which can be obtained in the split-anode magnetron. Those of you who are familiar with the early literature on the magnetron will recall that two distinct types of oscillations were frequently described. One type usually called "electronic" was found to occur under conditions when the magnetic field was just

high enough to cut off the anode current under static conditions. This field has the value computed by Hull:

$$H = \frac{6.72\sqrt{V}}{R}$$

Hull's first computation, by the way, was made neglecting space charge, but, strangely enough, the result is not changed by space charge. These electronic oscillations were assumed to be related in frequency to the time of transit of an electron from the cathode to the anode, and at cutoff this is inversely proportional to the field strength, as expressed by the empirical relationship

$$\lambda H = 13{,}100.$$



Electronic trajectories for different magnetic fields
A—small magnetic field $L \gg b$
B—moderate magnetic field $L \approx b$
C—strong magnetic field $L \ll b$
D—critical magnetic field $L = 0$

Fig. 14—Electronic trajectories for different magnetic fields varying from weak fields to the critical field shown to the right (Brillouin).

In general, it was found that best operation occurred when the magnetic field was not quite perpendicular to the electric field. The efficiency and outputs as reported for this type of oscillator were always low, in spite of the large amount of effort devoted to it by an equally large number of workers. A second type of oscillation, usually referred to as negative resistance oscillations, has also been the subject of considerable study and some practical use has been made of it at relatively low frequencies.

Contrasting with this, Posthumus described a third kind of oscillation which he called rotating field oscillations. As in the electronic oscillations the preferred frequency is determined by the magnetic field-strength and the anode potential, the frequency being inversely proportional to the magnetic field-strength. Contrasting with the electronic oscillations, the rotating

field oscillations occur with the magnetic field-strength very much above the critical cutoff value and the efficiency on occasion reached as much as 70%. While a careful reading of the literature will reveal that some of the earlier experimenters were occasionally dealing with these oscillations, Posthumus' observations represent a new departure in magnetron theory and practice and one which we might do well to investigate.

Posthumus' approach consisted in studying the electron paths in a magnetron in detail in order to find the conditions under which electrons may reach the plate with considerably less energy than that corresponding to the plate potential. He assumed a magnetron having k pairs of plates and based his calculations on the supposition of a rotating electric field with k pairs of poles. In reality there exists a simple alternating field but this can be resolved into two rotating fields rotating in opposite directions. Power engineers will recognize this as identical with the procedure used in analyzing single-phase rotating machinery. Posthumus neglected the field opposite to the static angular velocity and considered only one component. This is an approximation but a fairly plausible one which can be partially justified.

In the absence of oscillations there is a radial electric field independent of the angular position and inversely proportional to radius (for the coaxial cylindrical case). When oscillations are present there is an additional radial field which varies as some periodic function of the angle and with a period 2π, and a tangential component of the same general type. For simplicity these functions are taken to be simple harmonic functions and can therefore be split into two circular rotating fields.

Posthumus writes the two simultaneous differential equations determining the path of an electron, neglecting space charge, and inquires if a solution is possible for an elecron path which travels at approximately the same angular velocity as the rotating field but lags it by an angle α. An equally satisfactory way of looking at this is to say that we transform our coordinates from a fixed system to one rotating with the field and inquire if a solution is possible where α the angular motion is always small. He finds that such a solution is indeed possible and that for the electron motion to be stable the value of α must be such that the electrons are somewhat behind the line for which the field has its maximum retarding value. The electrons are thus in a position to lose energy to the field and to spiral out toward the anode.

Posthumus defined the value of the electron's radial velocity squared at the anode as P and the total velocity squared at the anode as Q. Normalized plots of these two parameters are shown in Fig. 15 as a function of frequency. The upper plot shows the radial velocity. Obviously for electrons to reach the plate at all they must have a positive velocity at the plate. Electrons can therefore reach the plate with any given field value, say $Z = 2$,

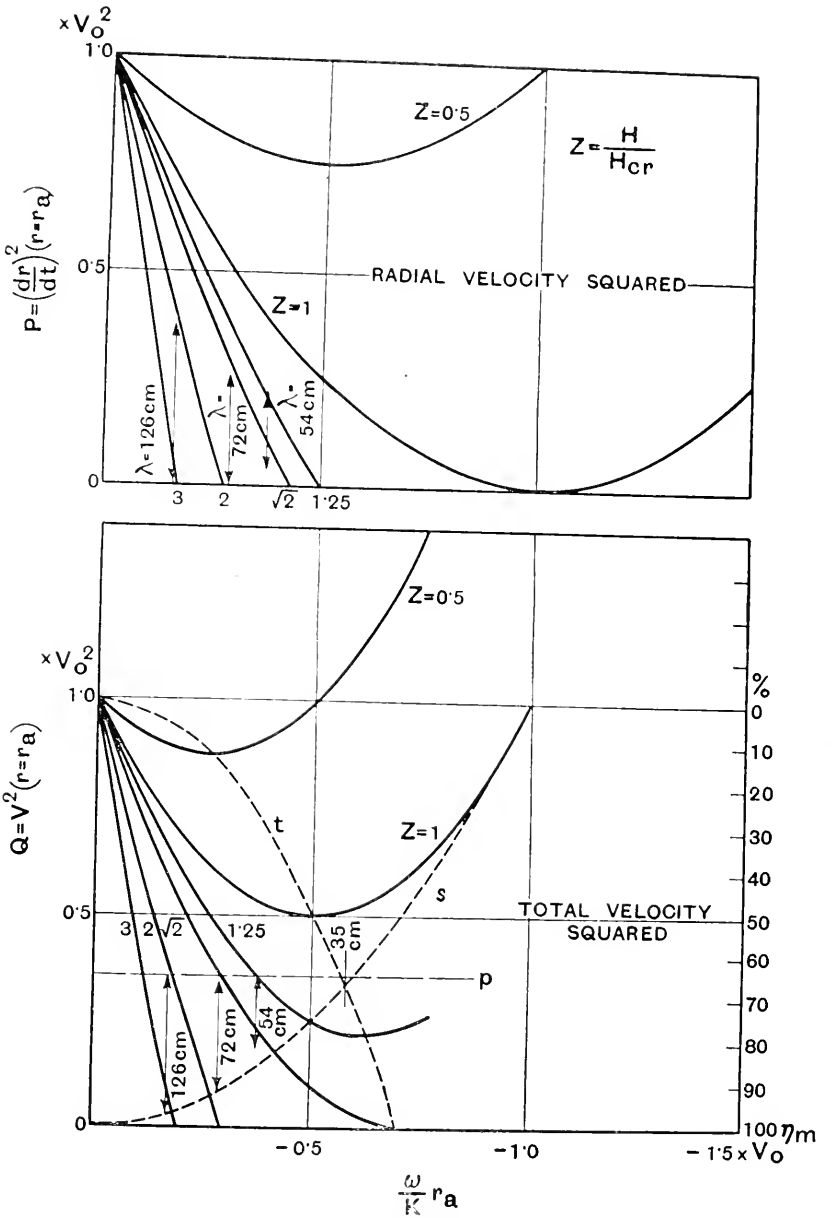


Fig. 15—Electron velocities in the magnetron according to Posthumus.

that is with a field equal to twice the cutoff value, for all frequencies less than the equivalent value defined by the intercept of the $Z = 2$ line with

the abscissa axis. The line for $P = 0$ appears on the lower curve as the dotted line s. Here the ordinate is the total velocity squared, normalized with respect to the value without oscillation. Efficiencies can therefore be put on the plot directly as shown by the right-hand scale in per cent. The line s is therefore a plot of the maximum possible efficiency. This refers to what we might call the electronic efficiency since no account is taken of circuit losses. Now in any physical device there are some circuit losses and hence a lower value of electronic efficiency for which sustained oscillations are not possible. The dotted line p is Posthumus' experimental value for this lower limit. Between the lines p and s, then, oscillations are possible at frequencies given by the abscissae and with field values shown on the solid lines. Actual data for an experimental tube are shown on the plot, oscillations occurring at the wavelengths indicated and over the ranges in field shown by the lines terminating in arrows.

One additional line t is shown on the plot connecting points on the different Z lines for which the efficiency is a maximum. The optimum design would be one based on the intersection of this line with the p line. Still other facts will appear from a detailed study of these results but we shall not be able to devote any more time to this interesting subject.

Conclusion

In concluding a talk of this sort and particularly in concluding a series of talks, it is usually appropriate to look ahead to the future and predict the trend of affairs, or perhaps to point out certain fruitful fields of research. I find this a singularly difficult thing to do. However, it is not revealing any military secrets to say that much of the progress of the last few years has been in the direction of making things work and not toward getting a clearer understanding of the underlying theory. If, for example, an illuminating approach could be devised which would make the problems associated with transverse fields, both electric and magnetic, appear as simple and straightforward as do longitudinal-electric-field problems, as a result of the velocity-modulation concept, then I believe even more striking advances could be made in the ultra-high-frequency field than those which the war years have brought forth.

Selected Bibliography

A. Papers of a Historical or Review Nature

H. Backhausen and K. Kurz, "The Shortest Waves Obtained with Vacuum Tubes", *Phys. Zeits*, vol. 21, p. 1 (1920).

E. W. B. Gill and G. H. Morrell, "Short Electric Waves Obtained by Valves", *Phil. Mag.*, vol. 44, p. 161 (1922).

A. W. Hull, "The Effect of a Uniform Magnetic Field on the Motion of Electrons between Coaxial Cylinders", *Phys. Rev.*, vol. 18, p. 31 (1921).

E. Habann, "A New Vacuum Tube Generator", *Zeits f. Hochfreq.*, vol. 24, p. 115 (1924).

A. Zacek, "A Method for the Production of Very Short Electromagnetic Waves", *Zeits f. Hockfreq.*, vol. 32, p. 172 (1928).
K. Okabe, "Production of Intense Extra-Short Electromagnetic Waves by Split-Anode Magnetron", *I.E.E. Jour.*, Japan, 1928, p. 284. See also I.R.E. Proc., vol. 17, p. 652 (1929); *I. R. E. Proc.*, vol. 18, p. 1748 (1930).
K. Kohl, "Continuous Ultra-Short Electric Waves", *Ergebnisse der exaktlen Naturwissenschaften*, vol. 9, p. 275 (1930). Bibliography of 135 titles.
M. J. Kelly and A. L. Samuel, "Vacuum Tubes as High-Frequency Oscillators", *Elec. Eng.*, vol. 53, p. 1504, Nov. 1934.
E. C. S. Megaw, "Electronic Oscillations", *Jour. I. E. E.* (London), vol. 22, p. 313, April (1933). (Bibliography of 47 titles.)
G. R. Kilgore, "Magnetron Oscillators for the Generation of Frequencies Between 300 and 600 Megacycles", *Proc. I. R. E.*, vol. 24, p. 1140, Aug. (1936).

B. Papers Dealing With Parallel Plane Electronics

W. E. Benham, "Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies".
Part I, *Phil. Mag.*, vol. 5, p. 641, March (1928).
Part II, *Phil. Mag.*, vol. 11, p. 457, February (1931).
Johannes Muller, "Electron Oscillations in High Vacuum", *Hockfreq. u Elek. Akus*, vol. 41, p. 156, May (1933).
F. B. Llewellyn, "Vacuum Tube Electronics at Ultra-High Frequencies", *I. R. E. Proc.*, vol. 21, p. 1532, Nov. (1933).
C. J. Bakker and G. de Vries, "Amplification of Small Alternating Tensions by an Inductive Action of the Electrons in a Radio Valve with Negative Anode", *Physica*, vol. 1, p. 1045, Nov. (1934).
C. J. Bakker and G. de Vries, "On Vacuum Tube Electronics", *Physica*, vol. 2, p. 683, July (1935).
G. Grunberg, "On the Theory of Operation of Electron Tubes", *Tech. Phys.* (U. S. S. R.), vol. 3, No. 2, p. 181, Feb. (1936).
D. O. North, "Analysis of the Effects of Space Charge on Grid Impedance", *I. R. E. Proc.*, vol. 24, p. 108, Jan. (1936).
W. E. Benham, "A Contribution to Tube and Amplifier Theory," *Proc. I. R. E.*, vol. 26, p. 1093, Sept. (1938).
F. B. Llewellyn and L. C. Peterson, "Vacuum Tube Networks", *Proc. I. R. E.*, vol. 32, p. 144, March (1944).

C. Velocity Modulation Papers

A. Arsenjewa-Heil and O. Heil, "Electromagnetic Oscillations of High Intensity", *Zeits f. phys.*, vol. 95, p. 752, and p. 62, Nov. & Dec. (1935).
(English Translation in *Electronics*, vol. 16, p. 7, July 1943).
E. Bruche and A. Recknagel, "On the Phase Focusing of Electrons in Rapidly Fluctuating Electric Fields", *Zeits f. phys.*, vol. 108, p. 459, March (1938).
W. C. Hahn and G. F. Metcalf, "Velocity-Modulated Tubes" *I. R. E. Proc.*, vol. 27, p. 106, Feb. (1939).
R. H. Varian and S. F. Varian, "A High-Frequency Oscillator and Amplifier", *Jour. Apl. Phys.*, vol. 10, p. 321, May (1939).
W. C. Hahn, "Small Signal Theory of Velocity Modulated Electron Beams", *G-E. Rev.*, vol. 42, p. 258, June (1939).
D. L. Webster, "Cathode Ray Bunching", *Jour. Apl. Phys.*, vol. 10, p. 501, July (1939).
W. C. Hahn, "Wave Energy and Transconductance of Velocity-Modulated Electron Beams", *G-E. Rev.*, vol. 42, p. 497, Nov. (1939).
S. Ramo, "The Electronic Wave Theory of Velocity Modulation Tubes", *Proc. I. R. E.*, vol. 27, p. 757, Dec. (1939).
J. R. Pierce, "Reflex Oscillators", *Proc. I. R. E.*, vol. 33, p. 112, Feb. (1945).

D. Graphical Solutions for Velocity Modulation Tubes

E. Bruche and A. Recknagel, "On the Phase Focusing of Electrons in Rapidly Fluctuating Electric Fields", *Zeits f. Phys.*, vol. 108, p. 459, March (1939).
D. M. Tombs, "Velocity-Modulation Beams", *Wireless Eng.*, vol. 17, p. 54, Feb. (1940).

R. Kompfner, "Velocity Modulation—Results of Further Considerations", *Wireless Eng.*, vol. 17, p. 478, Nov. (1940).
A. E. Harrison, "Graphical Methods for Analysis of Velocity Modulation Bunching", *Proc. I. R. E.*, vol. 33, p. 20, Jan. (1945).

E. Magnetrons

W. E. Benham, "Electronic Theory and the Magnetron", *Proc. Phys. Soc.*, vol. 47, p. 1, Jan. (1935).
K. Posthumus, "Oscillations in a Split Anode Magnetron", *Wireless Eng.*, vol. 12, p. 126, March (1935).
L. Tonks, "Motion of Electrons in Crossed Electric and Magnetic Fields with Space Charge", *Physik Z Sowjet*, vol. 8, p. 572 (1935).
H. Awendu, H. Thoma and M. Tombs, "Paths of Electrons in Magnetrons Taking Account of Space Charge", *Zeits. f. Phys.*, vol. 97, p. 202, Oct. (1935).
W. E. Benham, "Electronic Theory and the Magnetron Oscillator", *Proc. Phys. Soc.*, vol. 47, p. 1, Jan. (1935).
M. Grechowa, "Investigation of the Curve of the Electron Path in a Magnetron", *Tech. Phys. Jour.* (U. S. S. R.), vol. 3, p. 633 (1936).
S. V. Bellustin, "Theory of the Motion of Electrons in Crossed Electric and Magnetic Fields with Space Charge", *Physik Z. Towjet*, vol. 10, p. 251 (1936).
Grunberg and Wolkenstein, "The Influence of a Homogeneous Magnetic Field on the Motion of Electrons Between Coaxial Cylindrical Electrodes", *Tech. Phys. Jour.* (U. S. S. R.), vol. 8, p. 19 (1938).
F. Herringer and F. Hulster, "Oscillations in Tubes with Magnetic Fields", *Hock freq Tech u Elek Akus*, vol. 49, p. 123 (1937).
E. B. Moullin, "Consideration of the Effect of Space Charge in the Magnetron", *Phys. Soc. Proc.*, vol. 36, p. 94, Jan. (1940).
J. P. Blewett and S. Ramo, "High-Frequency Behavior of a Space Charge", *Phys. Rev.*, vol. 57, p. 635, April (1940).
J. P. Blewett and S. Ramo, "Propagation of Electromagnetic Waves in a Space Charge Rotating in a Magnetic Field", *Jour. Appl. Phys.*, vol. 12, p. 856, Dec. (1941).
L. Brillouin, "Practical Results from Theoretical Studies of Magnetrons", *Proc. I. R. E.*, vol. 32 p. 216, April (1944).

Dynamics of Package Cushioning

By RAYMOND D. MINDLIN

Introduction

MECHANICAL damage is a common occurrence in the transportation of packaged articles. The causes of failures are generally inadequate protective cushioning, lack of ruggedness of the outer packing container, or occasional abnormal weakness of the packaged article. The first of these difficulties is the subject of this paper.

One of the major influences in reducing the incidence of mechanical failures of packaged articles in recent years has been the use of the drop test. The drop test is performed simply by raising the package to a specified height and dropping it to the floor. The package and its contents are then examined for damage. This is a go-no-go test and requires a large number of samples before a reliable estimate of quality can be made. An adequate number of tests is prohibitive when the article packaged is costly. In such cases it is important, and in any case it is useful, to supplement the drop test data with measurements and calculations. It is also possible to evolve rational procedures for designing packages, as described in the present paper, so that a particular product will survive a drop test at any specified height, with a known factor of safety and with a minimum amount of space assigned for cushioning. The drop test then becomes only a check instead of playing an integral role in a cut and try design procedure.

Assuming that the outer container is adequate, the survival of a packaged article in a drop test still depends upon a large number of factors descriptive of the mechanical properties of both the cushioning medium and the packaged item. However, the more important properties can be grouped so that they may be replaced by knowledge of only the following factors:

(1) The magnitude of the maximum acceleration that the cushioning permits the packaged item to reach.
(2) The form of the acceleration-time relation.
(3) The strengths, natural frequencies of vibration and damping of the structural elements of the packaged article.

Part I of this paper is concerned primarily with methods for predicting maximum acceleration of the packaged article with emphasis on non-linear cushioning. Part II deals primarily with the prediction of the form of the acceleration-time relation. Part III deals with the effect of acceleration on

the packaged article and gives methods for determining whether or not the strength of the packaged article will be exceeded. The strength determinations themselves are not dealt with here; but the information in Part III is essential in interpreting and applying the data obtained in strength measurements. In Part IV some consideration is given to the influence of distributed mass and elasticity.

It cannot be emphasized too strongly that the determination of the mechanical properties of the packaged article, not dealt with in this paper, is an essential preliminary to a rational design procedure for packaging. The whole purpose in designing package cushioning is to limit the forces which may act on the packaged item. If one does not know to what values to limit the forces, a rational design procedure cannot be applied.

It is interesting to observe that the methods described here for analyzing and designing package cushioning are directly applicable to the design of shock mounts intended to protect equipment from the effects of a sudden change in velocity. All of the principles, formulas and design curves given here may be used in the shock mount problem with the simple substitution of $V^2/2g$ for h, where h is the height of drop in the packaging problem, g is the acceleration of gravity and V is the velocity change in the shock mount problem.

This paper is essentially a report on a study undertaken at the Bell Telephone Laboratories, Inc., in the Electronic Apparatus Development Department. The results have been applied to the packaging of large vacuum tubes and all of the examples used to illustrate the analysis and design procedures in the paper are taken from vacuum tube applications.

Miss H. A. Lefkowitz, Member of the Technical Staff, Bell Telephone Laboratories, assisted in the mathematical studies. The oscillograms, used as illustrations, were prepared under the supervision of Mr. F. W. Stubner, Member of the Technical Staff, Bell Telephone Laboratories. Figure 3.8.2 was taken from a thesis submitted by Mr. C. Ulucay in partial fulfillment of the requirements for the degree of Master of Science in the Department of Civil Engineering at Columbia University. The calculations for Figs. 3.5.1 to 3.5.6 and Fig. 3.2.2 for $\beta_1 > 0$ were performed on the Westinghouse Mechanical Transients Analyzer under the supervision of Dr. G. D. McCann, Transmission Engineer, Westinghouse Electric and Manufacturing Company.

Assumptions

The procedures to be described for the analysis and design of package cushioning are based on applications of a few simple laws of mechanics to an idealized mechanical system representing the package and its contents.

Essentially, a package consists of

1. Elements of the packaged article which are susceptible to mechanical damage.

2a. The packaged article as a whole.

2b. A cushioning medium (excelsior, cardboard spring pads, metal springs, etc.)

3. An outer container (cardboard carton, wood packing case, etc.)

The four major components are illustrated schematically in Fig. 0.2.1. The system is further idealized by "lumping the parameters"; for example, the outer container is considered as a single mass, the cushioning is considered as a massless spring with friction losses. The result of this idealization is to lose some of the fine detail of the real distributed system such as wave propagation through the cushioning and higher modes of vibration in

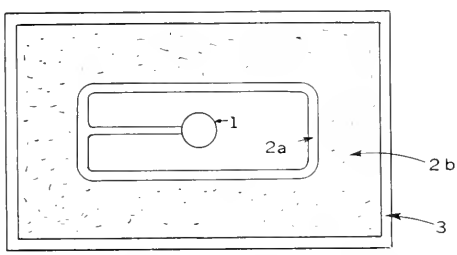


Fig. 0.2.1—Schematic representation of a package.
1. Element of packaged article
2a. Packaged article as a whole
2b. Cushioning
3. Outer container

the package structure and in the packaged article. Some consideration of these details is given in Part IV.

The idealized system is illustrated in Fig. 0.2.2. The major components of the system are as follows:

1. A structural element of the packaged item is represented by a mass (m_1) supported by a linear massless spring with or without velocity damping. The mass m_1 is assumed to be small in comparison with the mass of the whole packaged item.

2a. The whole packaged item is represented by a mass m_2.

2b. The cushioning is represented by a spring which may have a linear or non-linear load-displacement characteristic and which dissipates energy through velocity damping or dry friction. Permanent deformation of the cushioning is not considered, that is, in a repetition of the drop test it is assumed that the package has the same properties as before the first test. A properly designed package will have essen-

tially this characteristic. The mass of the cushioning is assumed to be small in comparison with m_2, except in Section 4.2.

3. The outer container is represented by the mass m_3. The impact of m_3 on the floor is assumed to be inelastic and during contact the relative displacement between m_3 and the initial position of the floor is assumed to be small in comparison with the relative displacement between m_2 and m_3. In other words, no spring action is assigned to the outer container and the floor is considered rigid.

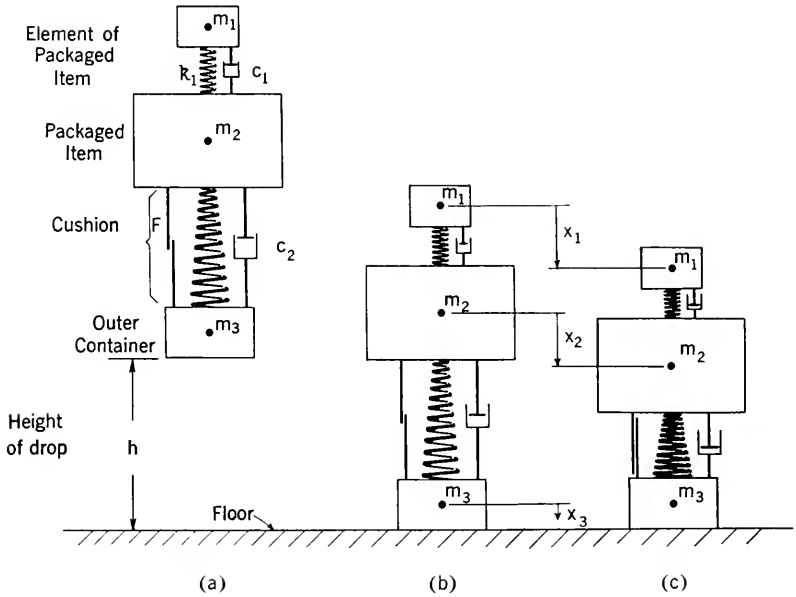


Fig. 0.2.2—Idealized mechanical system representing a package in a drop test.

PART I

MAXIMUM ACCELERATION AND DISPLACEMENT

1.1 Introduction

Most of Part I is concerned with the prediction of the maximum acceleration that the cushioning permits the packaged article (m_2) to attain. In many instances this will be all the information necessary for judging the suitability of a cushioning system. It will be all that is necessary if the shape and scale of the acceleration-time function satisfy certain criteria which are treated in detail in Parts III and IV. If these criteria are satisfied, the effect of the drop on the packaged article is found by multiplying the

dead load stresses (obtained in the usual manner) by the ratio of the maximum acceleration to the acceleration of gravity. If the criteria for the use of maximum acceleration alone are not satisfied, then Parts II and III will supply a numerical factor (the Amplification Factor) by which the maximum acceleration should be multiplied, and the remainder of the procedure is the same as before.

The determination of the maximum acceleration is founded on a knowledge of the load-displacement characteristics of the cushioning. When the cushioning system is simple enough, the load-displacement relation may be found or designed by purely analytical procedures. The tension spring package, discussed in Sections 1.7 and 1.8, is an example where such a treatment is possible. In many instances, as with distributed cushioning, the load-displacement relation is more easily found by test.

A load-displacement test is made by applying successively increasing forces, with weights or in a load testing machine, to the packaged item completely assembled in its package, and measuring the corresponding displacements. The force is applied usually by means of a rod inserted in a hole cut through the outer container and the cushioning to the packaged item. It is convenient to use a low loading rate in the test, and, in doing so, the effect of resisting forces that depend on velocity is lost. These forces are often of little importance but, in certain designs, it is necessary to consider them. This is done for velocity damping in Sections 2.5, 2.6, 3.2 and 3.5.

Most of Part I is concerned with cushioning having non-linear load-displacement characteristics. Linear cushioning is rarely encountered, but it will be treated first because of its simplicity and because it will be convenient later to express the maximum acceleration in non-linear cases in terms of the maximum acceleration in a hypothetical linear case.

1.2 Derivation of Equations of Motion

To introduce the method of analysis that will be used in Part I, the simplest possible system is considered first. The m_1 system is omitted entirely, the mass of the outer container (m_3) is neglected, and the cushioning is assumed to have no damping or friction. There remain only the mass m_2 (the mass of the packaged item alone) and the supporting spring, as shown in Fig. 1.2.1. If the spring is linear its displacement is proportional to the applied load throughout the range of use (see Fig. 1.4.1). The spring rate (k_2) of a linear spring is a constant usually expressed in terms of pounds per inch. The force (P) transmitted through a linear spring is therefore given by

$$P = k_2 x_2 , \tag{1.2.1}$$

where x_2 is the displacement of m_2 measured downward from its position at first contact of the spring with the floor (see Fig. 1.2.1). For a non-linear spring P will be some other function of x_2:

$$P = P(x_2). \tag{1.2.2}$$

To write the equation of motion for the mass m_2, we consider the forces acting on it at any instant. These are (see Fig. 1.2.2(b)) the spring force P and the weight $m_2 g$, where g is the acceleration of gravity. When x_2 is positive (i.e., a downward displacement of m_2 from its position at first contact of the spring with the floor) the spring exerts an upward force P

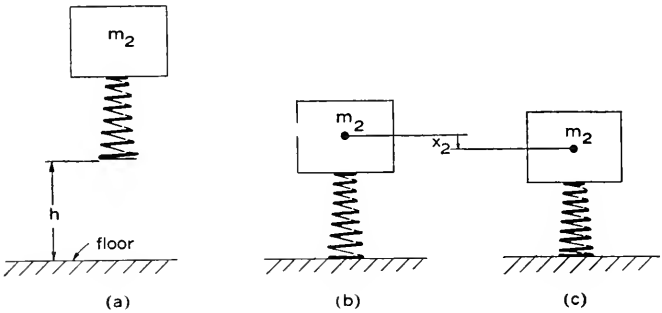


Fig. 1.2.1—Elementary system.

Fig. 1.2.2—Free body diagram for elementary system.
(a) Spring not in contact with floor.
(b) Spring in contact with floor.

on the mass, opposing the weight. The total downward force on m_2 is thus $m_2 g - P$. By the second law of motion, the product of the mass and its acceleration at any instant is equal to the applied force:

$$m_2 \ddot{x}_2 = m_2 g - P, \tag{1.2.3}$$

where the symbol $\ddot{x}_2$, representing the acceleration of m_2, stands for the second derivative of displacement with respect to time (d^2x_2/dt^2). Equation (1.2.3) is the law governing the motion of m_2 as long as the spring is in contact with the floor. When the spring is not in contact with the floor, it can exert no force on the mass so that, in writing the equation of motion that

governs before or after contact, the free-body diagram of Fig. 1.2.2(a) should be used. Then

$$\ddot{x}_2 = g. \tag{1.2.4}$$

Equation (1.2.4) holds (neglecting air resistance) from the instant the package starts to fall until the instant it strikes the floor and from it we can find the package velocity at the instant of first contact. Integrating (1.2.4) with respect to time, we find

$$\dot{x}_2 = gt + A, \tag{1.2.5}$$

where $\dot{x}_2$ is the velocity (dx_2/dt) and A is a constant of integration whose value is found from the initial condition that when $t = 0$ (the instant of release) $\dot{x}_2 = 0$. Thus $A = 0$ and

$$\dot{x}_2 = gt. \tag{1.2.6}$$

Integrating again,

$$x_2 = \tfrac{1}{2} gt^2 + B. \tag{1.2.7}$$

The value of the integration constant B is found from the initial condition that $x_2 = -h$ (the height of drop) when $t = 0$. Hence $B = -h$ and

$$x_2 = \tfrac{1}{2} gt^2 - h. \tag{1.2.8}$$

At the instant of contact, $x_2 = 0$ and, from (1.2.8), the time at first contact is given by $t_0^2 = 2h/g$. Substituting this value of t in (1.2.5) we find, for the velocity at first contact,

$$[\dot{x}_2]_{x_2=0} = \sqrt{2gh}. \tag{1.2.9}$$

We now have the initial conditions for finding the values of the integration constants in the solution of equation (1.2.3), which we proceed to obtain.

First multiply both sides of (1.2.3) by dx_2/dt and write $\ddot{x}_2 = \dfrac{d}{dt}\left(\dfrac{dx_2}{dt}\right)$:

$$m_2 \frac{dx_2}{dt} \frac{d}{dt}\left(\frac{dx_2}{dt}\right) + P \frac{dx_2}{dt} = m_2 g \frac{dx_2}{dt} \tag{1.2.10}$$

or

$$\tfrac{1}{2} m_2 \frac{d}{dt}\left(\frac{dx_2}{dt}\right)^2 + P \frac{dx_2}{dt} = m_2 g \frac{dx_2}{dt}.$$

Multiplying by dt and integrating once:

$$\tfrac{1}{2} m_2 \dot{x}_2^2 + \int^{x_2} P \, dx_2 = \int^{x_2} m_2 g \, dx_2 + C, \tag{1.2.11}$$

where C is a constant of integration whose value is determined by the initial conditions that $\dot{x}_2^2 = 2gh$ and $x_2 = 0$ at the instant of contact. Hence

$$C = m_2 gh + \int^0 P\, dx_2 .$$

Substituting the above value of C in (1.2.11), we have

$$\tfrac{1}{2} m_2 \dot{x}_2^2 + \int_0^{x_2} P\, dx_2 = m_2 g(h + x_2). \tag{1.2.12}$$

It may be observed that (1.2.12) is an energy equation in which the terms have the following meanings:

$\tfrac{1}{2} m_2 \dot{x}_2^2$ is the instantaneous kinetic energy of m_2,

$\int_0^{x_2} P\, dx_2$ is the energy stored in the spring at any instant. It is also equal to the area under the load-displacement curve up to the displacement x_2,

$m_2 g(h + x_2)$ is the potential energy of the mass at its initial height $h + x_2$ above the instantaneous position x_2.

Hence (1.2.12) expresses the law of conservation of energy.

Ordinarily h is very much larger than x_2 so that we may write, with good accuracy,

$$\tfrac{1}{2} m_2 \dot{x}_2^2 + \int_0^{x_2} P\, dx_2 = m_2 gh. \tag{1.2.13}$$

To the same approximation, equation (1.2.3) becomes

$$m_2 \ddot{x}_2 + P = 0. \tag{1.2.14}$$

Equation (1.2.14) and its first integral, equation (1.2.13), are convenient forms for calculating events at any time during contact. Their use will be illustrated in Part II. For calculating only maximum displacement and acceleration, the equations become simpler. Let

W_2 = weight of the packaged article ($= m_2 g$),

d_m = maximum displacement of the packaged article,

G_m = absolute value of maximum acceleration of the packaged article in terms of number of times gravity ($G_m = |\ \ddot{x}_2/g\ |_{\max}$),

P_m = maximum force exerted on packaged article by cushioning.

We shall limit our study to the practical regions where $P > 0$ when $x_2 > 0$. Then it may be seen from (1.2.13) that x_2 is a maximum when $\dot{x}_2$ is zero, hence

$$\int_0^{d_m} P\, dx_2 = W_2 h, \tag{1.2.15}$$

and, from (1.2.14),

$$G_m = \frac{P_m}{W_2}, \tag{1.2.16}$$

where P_m is the maximum value of P. If $P(x_2)$ is a monotonic function, P_m may be obtained from (1.2.2) by substituting d_m for x_2:

$$P_m = P(d_m). \tag{1.2.17}$$

In the unusual case where $P(x_2)$ is not monotonic, the maximum value of P in the interval $0 \leq x_2 \leq d_m$ must be chosen instead of equation (1.2.17).

The general procedure is to calculate d_m from (1.2.15), P_m from (1.2.17) and then G_m from (1.2.16). If P can be expressed analytically in terms of x_2 and if the integral in (1.2.15) can be evaluated in terms of elementary functions, simple formulas can be found for d_m and G_m. If this is not possible, then the integration can be performed graphically or numerically. Both of these procedures will be illustrated. In either case the maximum acceleration and displacement are obtained in terms of the weight of the packaged item, the height of drop and parameters descriptive of the load-displacement characteristics of the cushioning.

1.3 Linear Elasticity

For cushioning with a linear load-displacement relation, equation (1.2.1) applies. Substituting this value of P in (1.2.15), and performing the integration, we find

$$d_m = \sqrt{\frac{2hW_2}{k_2}}. \tag{1.3.1}$$

From (1.3.1) and (1.2.17),

$$P_m = \sqrt{2hW_2k_2}, \tag{1.3.2}$$

and, from (1.3.2) and (1.2.16),

$$G_m = \sqrt{\frac{2hk_2}{W_2}}. \tag{1.3.3}$$

Notice that equation (1.3.3) holds only if there is space available for a displacement d_m and if the cushioning is linear and capable of transmitting a force P_m. Also, from (1.3.3) and (1.3.1),

$$d_m = \frac{2h}{G_m} \tag{1.3.4}$$

and

$$k_2 = \frac{W_2G_m^2}{2h} = \frac{2hW_2}{d_m^2}. \tag{1.3.5}$$

Example: Find the properties of the linear cushioning required so that the maximum acceleration will be $50g$ in a 3 ft. drop of a 20 lb. article.

From (1.3.4),

$$\text{necessary travel, } d_m = \frac{2 \times 36}{50} = 1.44 \text{ inches.}$$

From (1.3.5),

$$\text{spring rate, } k_2 = \frac{20 \times (50)^2}{2 \times 36} = 694 \text{ lbs/in.}$$

From (1.2.16)

$$\text{Maximum force } P_m = 20 \times 50 = 1000 \text{ lbs.}$$

1.4 Cushioning with Non-Linear Elasticity

In practice it is rarely that a packaging system has linear spring characteristics. Departure from linearity may be due to

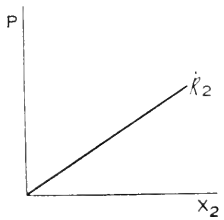


Fig. 1.4.1—Linear elasticity. Class A.

1. Non-linear geometry, such as in the tension spring package described in Section 1.7.

2. Non-linear characteristics of distributed cushioning materials such as excelsior and rubber.

3. Abrupt change of stiffness such as occurs if the packaged item can strike the wall of the container.

For the purpose of developing design formulas it is desirable to have analytical functions to represent load-displacement characteristics. It is not feasible to have only one family of functions with adjustable parameters to fit all possible shapes of load-displacement curves. Therefore, all the practical shapes have been divided into six general classes, most of which are associated with simple functions having one or two adjustable parameters. The six classes are as follows:

Class A—Linear Elasticity. This has already been treated. Its load-displacement function is

$$P = k_2 x_2. \tag{1.4.1}$$

Class B—Cubic Elasticity. This includes cushioning which does not bottom in the anticipated range of use, but the slope of the load-displacement function generally increases with increasing displacement as in the curved full line of Fig. 1.4.2. A suitable load-displacement function is

$$P = k_0 x_2 + r x_2^3 . \tag{1.4.2}$$

k_0 is the initial spring rate of the cushioning, as shown by the slope of the dashed straight line in Fig. 1.4.2, and r determines the rate of increase of the spring rate. The same function can be used if the slope of the curve decreases gradually with increasing load as shown by the curved dashed line in Fig. 1.4.2. In this case the parameter r is negative.

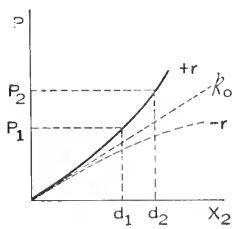


Fig. 1.4.2

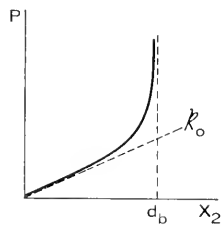


Fig. 1.4.3

Fig. 1.4.2—Cubic elasticity. Class B.
Fig. 1.4.3—Tangent elasticity. Class C.

Class C—Tangent Elasticity. Cushioning that bottoms, but not very abruptly, can be represented by the load-displacement function

$$P = \frac{2k_0 d_b}{\pi} \tan \frac{\pi x_2}{2d_b} . \tag{1.4.3}$$

Referring to Fig. 1.4.3, k_0 is the initial spring rate and d_b is the maximum available displacement. The figure shows how the stiffness of the cushioning (i.e., the slope of the curve) increases as the displacement approaches the maximum available (d_b) at hard bottoming. The shape of the curve is typical of load-displacement curves for a great variety of packages with distributed cushioning.

Figure 1.4.7 illustrates the wide variety of shapes of non-linear cushioning characteristics that can be obtained with the single function given by equation (1.4.3) simply by varying the parameter k_0; and a similar set is given by each value of d_b. Although these families of curves do not include all possible shapes, one of them can usually be found to fit a practical shape for cushioning of this class over the anticipated range of use.

Class D—Bi-linear Elasticity. This is characterized by a load-displace-

ment curve consisting of two straight line segments. The load displacement function is (see Fig. 1.4.4)

$$\left.\begin{aligned} P &= k_0 x_2 \qquad 0 \leqq x_2 \leqq d_s \\ P &= k_b x_2 - (k_b - k_0) d_s \quad x_2 \geqq d_s \end{aligned}\right\}. \tag{1.4.4}$$

It is useful especially in situations where very abrupt bottoming is possible.

Class E—Hyperbolic Tangent Elasticity. When the mechanism of the cushioning is such as to limit the maximum force that can be transmitted over a considerable displacement range, the load-displacement function

$$P = P_0 \tanh \frac{k_0 x_2}{P_0} \tag{1.4.5}$$

is useful. P_0 is the asymptotic value of the force and k_0 is the initial spring rate (see Fig. 1.4.5).

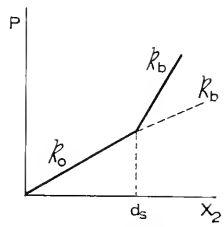


Fig. 1.4.4

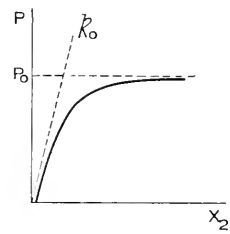


Fig. 1.4.5

Fig. 1.4.4—Bi-linear elasticity. Class D.
Fig. 1.4.5—Hyperbolic tangent elasticity. Class E.

Class F—Anomalous Elasticity. In occasional instances the load-displacement curve of the cushioning cannot be matched accurately enough by any of the five preceding functions. In such cases a numerical integration procedure can be used, as described in Section 1.15.

1.5 Cushioning with Cubic Elasticity (Class B)

Substituting (1.4.2) in (1.2.15) and performing the integration, we have:

$$\frac{k_0 d_m^2}{2} + \frac{r d_m^4}{4} = W_2 h. \tag{1.5.1}$$

Now, let

$$d_0 = \sqrt{\frac{2W_2 h}{k_0}}, \tag{1.5.2}$$

that is, d_0 is the displacement that would take place if the elasticity were linear (see equation (1.3.1)) with a constant spring rate k_0 equal to the initial spring rate of the cubic elasticity. Also let

$$B = \frac{4W_2 hr}{k_0^2}. \tag{1.5.3}$$

Then, from (1.5.1), (1.5.2) and (1.5.3)

$$\frac{d_m}{d_0} = \sqrt{\frac{2}{B}(-1 + \sqrt{1 + B})}. \tag{1.5.4}$$

Equation (1.5.4) is plotted in Fig. 1.5.1 which shows graphically how the maximum displacement d_m compares with the "equivalent linear displacement d_0" as the parameter B is varied. Note that B depends on the weight of the packaged item, the height of drop and the shape of the load displacement curve (as determined by k_0 and r).

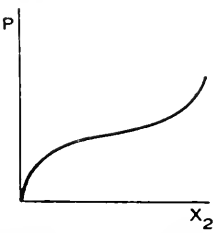


Fig. 1.4.6—Anomalous elasticity. Class F.

Similarly we can compare the maximum acceleration G_m with the maximum (G_0) that would obtain if the load displacement curve were linear with spring rate k_0. The latter acceleration is given by

$$G_0 = \sqrt{\frac{2hk_0}{W_2}} \tag{1.5.5}$$

and the former is obtained by finding P_m from (1.2.17) and then, from (1.2.16),

$$\frac{G_m}{G_0} = \sqrt{\frac{2}{B}(1 + B)(-1 + \sqrt{1 + B})}. \tag{1.5.6}$$

Equation (1.5.6) is plotted in Fig. 1.5.2.

1.6 Procedure for Finding Maximum Acceleration and Displacement for Cushioning with Cubic Elasticity

If the load-displacement curve of a cushioning system has the general appearance of Fig. 1.4.2 (where the slope increases or decreases gradually

with displacement) the following procedure may be used for estimating the effectiveness of the cushioning.

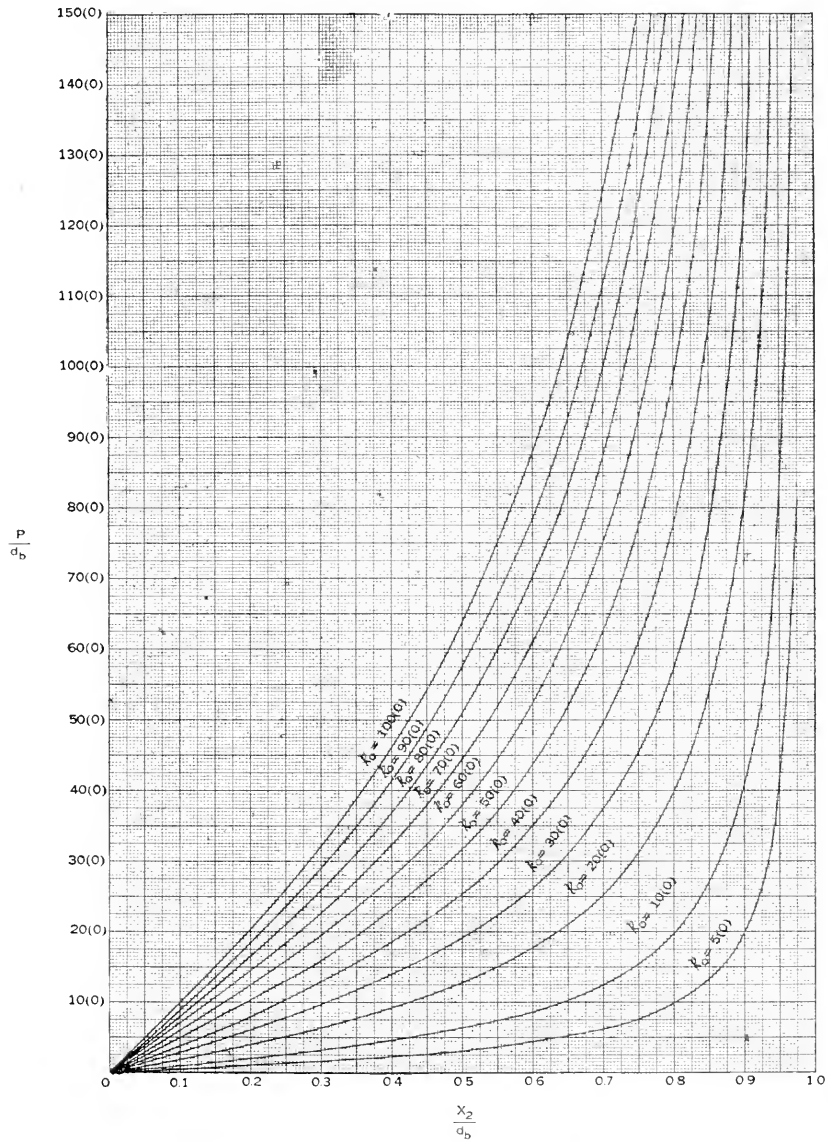


Fig. 1.4.7—Family of load displacement curves for cushioning with tangent elasticity.

a. Select the point on the load-displacement curve for which the load is equal to the weight of the packaged item multiplied by the allowable G_m. Call this load P_2 and the corresponding displacement d_2.

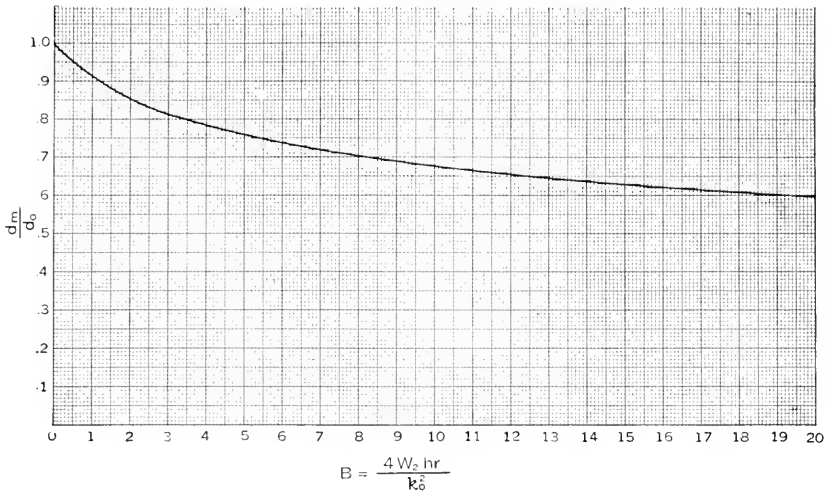


Fig. 1.5.1—Maximum displacement for cushioning with cubic elasticity. See equation (1.5.4).

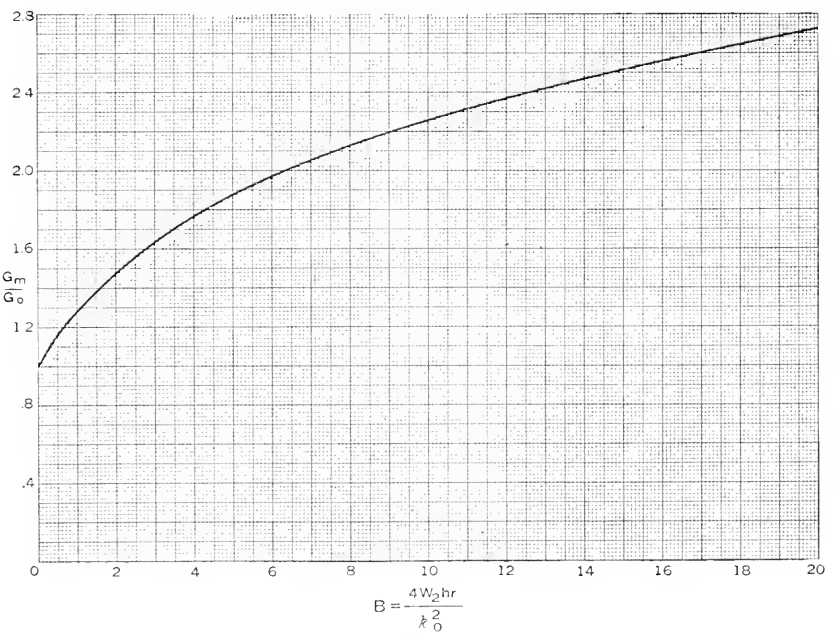


Fig. 1.5.2—Maximum acceleration for cushioning with cubic elasticity. See equation (1.5.6).

b. Select another point (d_1, P_1) about half way toward the origin from (d_2, P_2). See Fig. 1.4.2.

c. Calculate

$$k_0 = \frac{\frac{P_1}{d_1} d_2^2 - \frac{P_2}{d_2} d_1^2}{d_2^2 - d_1^2} \tag{1.6.1}$$

and

$$r = \frac{\frac{P_2}{d_2} - \frac{P_1}{d_1}}{d_2^2 - d_1^2}. \tag{1.6.2}$$

d. Using the known weight, W_2, of the packaged item, the specified height of drop h, and k_0 and r from (1.6.1) and (1.6.2), calculate B, d_0 and G_0 from (1.5.2), (1.5.3) and (1.5.5). Then calculate the maximum acceleration G_m and maximum displacement d_m from (1.5.6) and (1.5.4) or find their values from Figs. 1.5.1 and 1.5.2.

Example: A large vacuum tube, weighing 22.5 lbs, was packed in a 7″ x $7\frac{3}{4}$″ x 15″ carton which was supported on corrugated cardboard spring pads in a $10\frac{1}{2}$″ x $11\frac{1}{2}$″ x $18\frac{1}{2}$″ carton. The latter was, in turn, packed in 28 pounds of excelsior in a 25″ x 25″ x 30″ carton. The tube is rated at 50g and the package is intended for a drop of three feet.

A rod was inserted through a hole cut through the three cartons to the tube. Load was applied to the rod and the displacement of the tube was measured. The data obtained were

P (*load in lbs*)	x_2 (*displacement in inches*)
0	0
100	$\frac{5}{16}$
200	$\frac{5}{8}$
300	$\frac{7}{8}$
400	$1\frac{1}{16}$
500	$1\frac{3}{16}$
600	$1\frac{3}{8}$
700	$1\frac{1}{2}$
800	$1\frac{9}{16}$
900	$1\frac{5}{8}$
1000	$1\frac{3}{4}$
1100	$1\frac{13}{16}$
1200	$1\frac{7}{8}$
1300	$1\frac{15}{16}$
1400	2

The data are plotted in Fig. 1.6.1. The resulting curve is suitable for classification as either Class B or Class C cushioning. Considering it, for the present, as Class B, we take $P_2 = 22.5 \times 50 = 1225$, and from the curve, $d_2 = 1.9$ inches. Also, from the curve, take $d_1 = 1$ inch and $P_1 = 365$

lbs. Substituting these values in (1.6.1) and (1.6.2), we find $k_0 = 255$ and $r = 108$. Then

$$B = \frac{4W_2hr}{k_0^2} = \frac{4 \times 22.5 \times 36 \times 108}{(255)^2} = 5.4,$$

$$G_0 = \sqrt{\frac{2hk_0}{W_2}} = \sqrt{\frac{2 \times 36 \times 255}{22.5}} = 28.6.$$

Entering Fig. 1.5.2 with $B = 5.4$ we find $G_m/G_0 = 1.9$. Hence

$$G_m = 28.6 \times 1.9 = 55$$

This is close enough to the $50g$ rating of the tube to call the cushioning safe insofar as maximum acceleration is concerned.

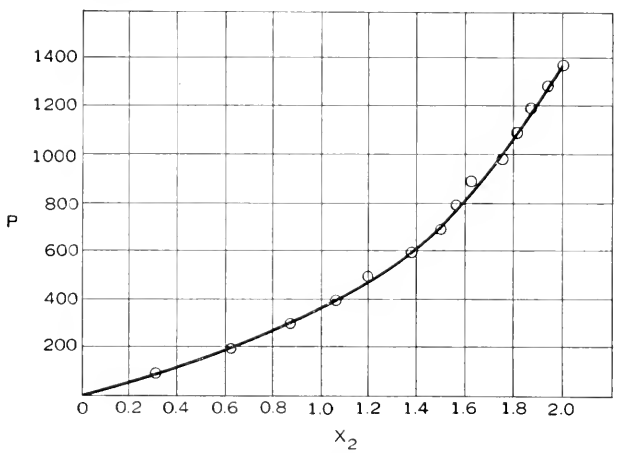


Fig. 1.6.1—Experimental load-displacement curve for a corrugated cardboard spring pad and excelsior cushion.

The maximum displacement, obtained by entering Fig. 1.5.1 with $B = 5.4$ and finding $d_m/d_0 = 0.75$. Then $d_m = 0.75 \times 2h/G_0 = 1.95$ inches. Hence, the package is much larger than necessary since approximately 8 inches of cushioning thickness is supplied to accommodate 2 inches of displacement.

1.7 The Tension-Spring Package (Class B)

The tension spring package is useful when the allowable G_m is so small and height of drop so great that a large displacement (say $d_m >$ several inches) is required. The decision as to whether or not a tension spring package is indicated may be made on the basis of a preliminary estimate of displacement based on the linear case. Suppose the height of drop is to be 60 inches and the allowable acceleration for the packaged item is

$20g$. Then, from Equation 1.3.4, the approximate displacement that will be required is

$$d_m = \frac{2 \times 60}{20} = 6 \text{ inches.} \tag{1.7.1}$$

The actual maximum displacement in a tension spring package will prove to be somewhat more than 6 inches, but the preliminary calculation shows the displacement to be large enough to warrant the use of this type of cushioning.

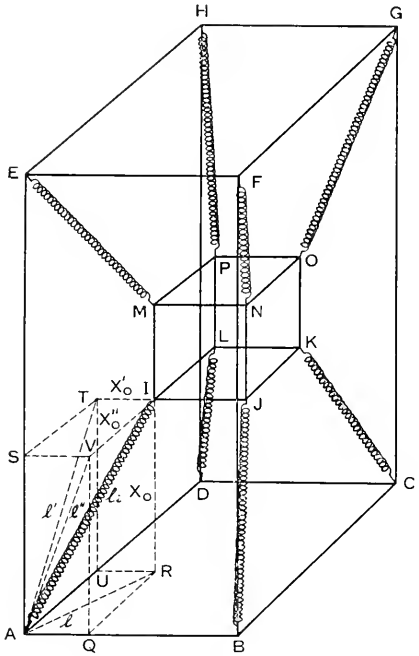


Fig. 1.7.1—Schematic diagram of a tension spring package.

A schematic diagram of a typical tension spring package is shown in Fig. 1.7.1 and a photograph of one design is given in Fig. 1.7.2. The packaged item is suspended on eight identical helical tension springs which diverge to the outer frame. The analysis and design procedures described in this and the following section apply equally well if the springs *converge* from the packaged item to the outer frame. With a slight modification, indicated in the next section, the procedure also applies if four of the springs (say, BJ, DL, EM, OG in Fig. 1.7.1) are omitted.

In all cases, however, we shall consider only systems having reflected symmetry about each of three mutually perpendicular planes through the center of gravity of the packaged article.

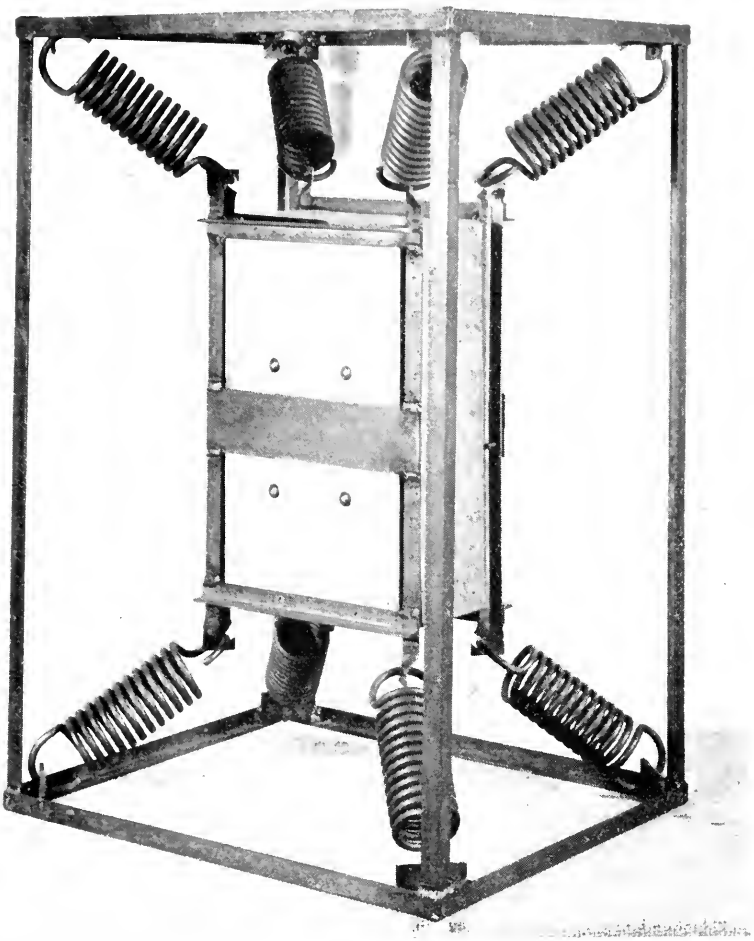


Fig. 1.7.2—A tension spring package.

The load-displacement characteristics of the spring system may be found by statical considerations. We shall examine, first, the displacement in the vertical direction in Fig. 1.7.1, using the following notations:

P = force applied to the suspended object.

x_2 = displacement of suspended object,

x_0 = perpendicular distance (IR, Fig. 1.7.1) from inner spring support point (I, Fig. 1.7.1) to nearest plane, perpendicular to displacement direction and containing four outer spring support points (A, B, C, D, Fig. 1.7.1);

ℓ_i = distance (IA) between spring support points when suspended article is in equilibrium position,
ℓ = projection of ℓ_i on plane $ABCD$,
f = ℓ minus length (between hooks) of unstretched spring,
k = spring rate of each spring.

Consider, first, the action of one pair of springs, say EM and GO of Fig. 1.7.1, independent of the remainder of the suspension. Since EM and GO lie in parallel vertical planes and the points M and O remain in the initial planes of their respective springs during a vertical displacement, the two springs may be considered to lie in the same plane, and to be translated horizontally in this plane so that their outer ends are separated by a distance 2ℓ. Hence Fig. 1.7.3 may be used to represent the independent action of this pair of springs and it is required to find the force Q' needed to transform Fig. 1.7.3(a) to Fig. 1.7.3(b). Initially there are two springs, each of length $\ell - f$

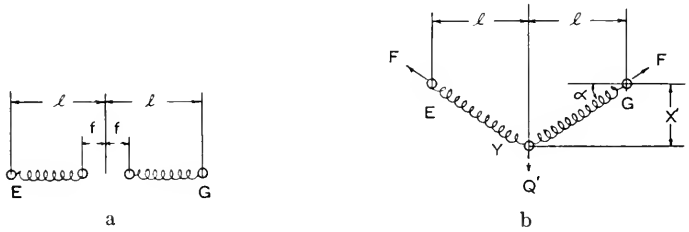


Fig. 1.7.3—Diagram used in discussion of tension spring package.

and spring constant k, with no initial tension in them. One end of one spring is fixed at point E and one end of the other spring is fixed at a point G distant 2ℓ from E. The springs are then stretched so that the two initially free ends are located at a point Y equidistant from E and G and distant x_2' from line EG. The axis of each spring makes an angle α with EG, where

$$\sin \alpha = \frac{x_2'}{\sqrt{\ell^2 + x_2'^2}}. \tag{1.7.2}$$

In this state the axial force F in each spring is

$$F = k[\sqrt{\ell^2 + x_2'^2} - \ell + f] \tag{1.7.3}$$

and the force Q', required to equilibrate the two forces F is $2F \sin \alpha$. Considering the force Q' as a function of the displacement x_2', we write

$$Q'(x_2') = \frac{2x_2' k}{\sqrt{\ell^2 + x_2'^2}} [\sqrt{\ell^2 + x_2'^2} - \ell + f] \tag{1.7.4}$$

or
$$Q'(z') = 2k\ell\left[z' + \frac{(1 - b)z'}{\sqrt{1 + z'^2}}\right] \tag{1.7.5}$$

where
$$z' = \frac{x_2'}{\ell}, \qquad b = \frac{f}{\ell}.$$

Consider, next, the configuration shown in Fig. 1.7.4(a), where one end of each of four springs is fixed at a corner of a rectangle of length 2ℓ and width $2x_0$. Each spring is again of length $\ell - f$. The four free ends of the springs are drawn together at a common point X at the center of the rectangle (see Fig. 1.7.4(b)). The system is in equilibrium in this position. A force Q is then applied at X in the plane of the rectangle and normal to the side 2ℓ.

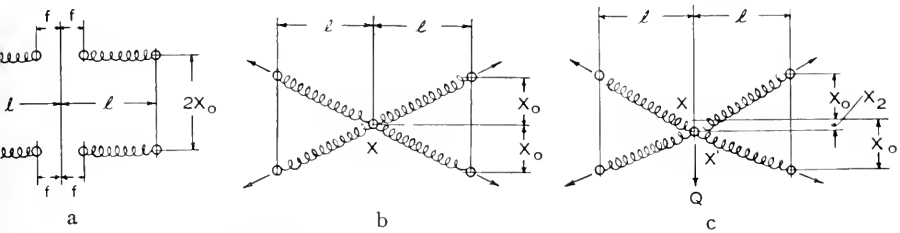


Fig. 1.7.4—Action of springs in a tension spring package.

The common point X is displaced a distance x_2 to X' (see Fig. 1.7.4(c)). Writing $z = x_2/\ell$, $a = x_0/\ell$, we observe that

$$Q(z) = Q'(z + a) + Q'(z - a), \tag{1.7.6}$$

or, from equation (1.7.5),

$$Q(z) = 2k\ell\left\{2z - (1 - b)\left[\frac{z + a}{\sqrt{1 + (z + a)^2}} + \frac{z - a}{\sqrt{1 + (z - a)^2}}\right]\right\}. \tag{1.7.7}$$

The standard tension spring package has two sets of four springs so that the force P required to displace the common point X a distance x_2 is

$$P(z) = 2Q(z). \tag{1.7.8}$$

If x_2 is small in comparison with ℓ (i.e., z is small in comparison with unity), equation (1.7.8) may be written approximately as

$$P(z) = 4k\ell\left[2z - \frac{(1 - b)z(2 - z^2)}{(1 + a^2)^{3/2}}\right]. \tag{1.7.9}$$

Even when x_2 becomes almost as large as ℓ, equation (1.7.9) has been found, experimentally, to be remarkably accurate.
Writing

$$K = 8k\left[1 - \frac{1 - b}{(1 + a^2)^{3/2}}\right] \tag{1.7.10}$$

and

$$c = \frac{(1 + a^2)^{3/2}}{1 - b} - 1, \tag{1.7.11}$$

equation (1.7.9) becomes

$$P = K\ell\left(z + \frac{z^3}{2c}\right). \tag{1.7.12}$$

It is seen, by comparison with (1.4.2) that this is Class B cushioning (cubic elasticity). K is the initial spring rate and c determines the rate of increase of stiffness with displacement. With the notation k_0, r of Section 1.5, we see that

$$k_0 = K \tag{1.7.13}$$

$$r = \frac{K}{2c\ell^2}. \tag{1.7.14}$$

Hence equations (1.5.6) and (1.5.4) may again be used to calculate maximum acceleration and displacement. B has the same meaning as before (Eq. 1.5.3).

To predict the performance, in the vertical direction (Fig. 1.7.1), of an existing tension spring package the same procedure as outlined in Section 1.6 may be used, except that it is not necessary to have a load-displacement curve for calculating k_0 and r. Instead, these parameters may be calculated directly from equations (1.7.10), (1.7.11), (1.7.13) and (1.7.14). The remainder of the procedure is the same as in Section 1.6(d).

To predict the performance perpendicular to another face, say $AEHD$ of Fig. 1.7.1, it is only necessary, in the calculation of k_0 and r, to substitute x_0' for x_0, ℓ' for ℓ (see Fig. 1.7.1) and, in place of b:

$$b' = 1 - \frac{\ell}{\ell'}(1 - b). \tag{1.7.15}$$

The initial spring rate K for any direction of acceleration may be calculated from the initial spring rates K_1, K_2, K_3 in the three directions normal to the faces of the frame by using the relation

$$\frac{1}{K^2} = \frac{s^2}{K_1^2} + \frac{t^2}{K_2^2} + \frac{u^2}{K_3^2}, \tag{1.7.16}$$

where s, t, u are the direction cosines of the acceleration direction with respect to the normals to the faces of the frame. It is seen, from (1.7.16), that the spring rate is given by the radius to the surface of an ellipsoid whose principal semi-axes are K_1, K_2, K_3.

The displacement direction does not necessarily coincide with the acceleration direction. The angle θ between them is given by

$$\cos\theta = K\left[\frac{s^2}{K_1} + \frac{t^2}{K_2} + \frac{u^2}{K_3}\right] \tag{1.7.17}$$

where K is defined by equation (1.7.16).

The spring characteristics may be made the same in all directions and the displacement direction may be made to coincide with the acceleration direction by setting

$$x_0 = x_0' = x_0'' = \frac{\ell}{\sqrt{2}}$$

and

$$\ell = \ell' = \ell''$$

(see Fig. 1.7.1). This makes $b = b' = b''$, $c = 0.828$ and $k/K = 0.274$ in the calculations of the next section.

1.8 Procedure for Designing Tension Spring Packages

The design of a tension spring package, as contrasted with the analysis of one, must proceed without initial knowledge of values for the parameters k_0 and r, since these cannot be known until the springs are designed. Therefore equations (1.5.4) and (1.5.6) cannot be used directly. For design purposes they are transformed to the following set of formulas:

$$\sqrt{c} = \sqrt{\frac{(1+a^2)^{3/2}}{1-b}} - 1 \tag{1.8.1}$$

$$L = \frac{h}{\sqrt{c}\,\ell G_m} = \frac{1}{8}\sqrt{(\sqrt{N} + \sqrt{N+8})^2 + \frac{16}{N}} \tag{1.8.2}$$

$$\frac{d_m}{\sqrt{c}\,\ell} = \sqrt{2(-1+\sqrt{1+B})} \tag{1.8.3}$$

$$M^2 = N = \frac{W_2 G_m^2}{2hK} = \frac{2}{B}(1+B)(-1+\sqrt{1+B}) \tag{1.8.4}$$

$$\frac{e}{\ell} - b = -1 + \sqrt{1 + \frac{d_m + x_0}{\ell}}. \tag{1.8.5}$$

These formulas have been converted to design curves which are given in Figs. 1.8.1 to 1.8.5. The curves are for use in connection with the following routine procedure which has been found useful in designing the springs for tension spring packages. Reference should be made to Table I.

1. Enter, on Line 1, Table I, the weight (W_2) in pounds, of the suspended item. This includes the weight of the cradle or other holding arrangement and one-third the estimated weight of the springs.

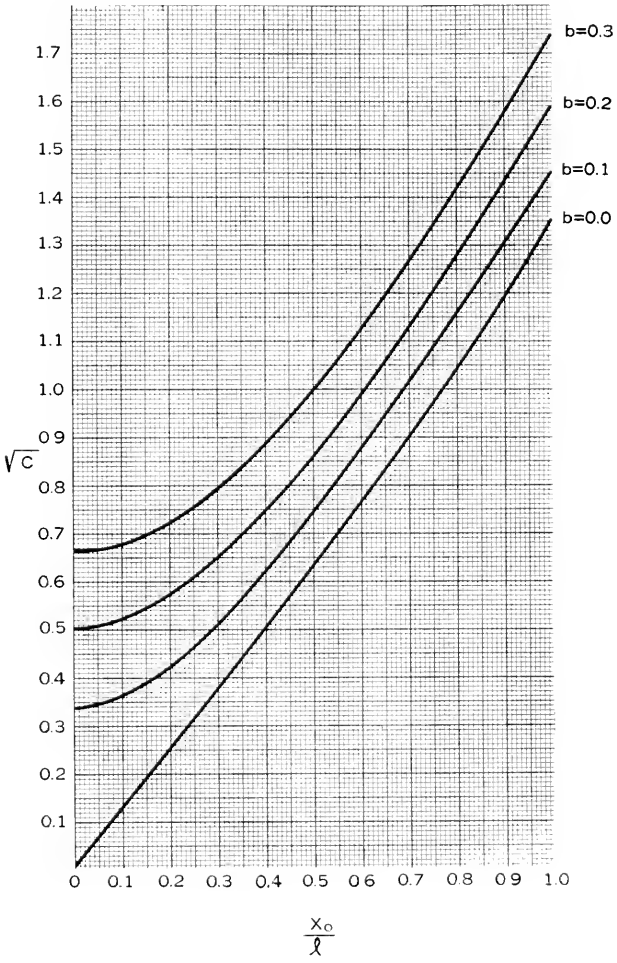


Fig. 1.8.1—Tension spring package design curve. Equation (1.8.1).

2. Enter, on Line 2, the height of drop (h) in inches.
3. Enter, on Line 3, the maximum allowable acceleration (G_m) in units of "number of times gravity." This should be determined beforehand from tests on the item to be packaged.
4. Enter, on Line 4, the dimension x_0 (inches).

5. Enter, on Line 5, the dimension ℓ (inches). For a package to have the same spring rate in all directions, $\ell = x_0\sqrt{2}$ is a necessary condition.
6. Enter, on Line 6, the value chosen for b. As b becomes greater than zero, the stiffness of the whole suspension increases for a given stiffness of individual springs. The reverse happens for b less than zero.

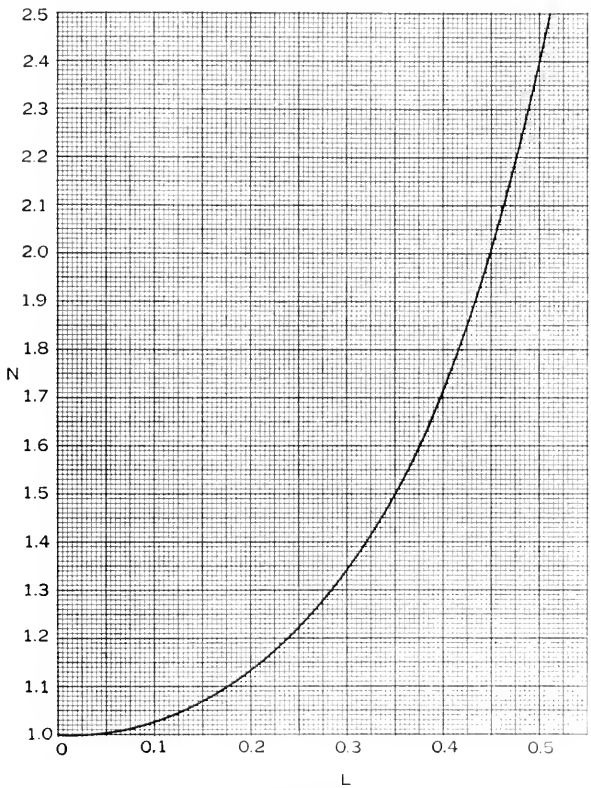


Fig. 1.8.2—Tension spring package design curve. Equation (1.8.2).

7. Calculate x_0/ℓ.
8. Enter Fig. 1.8.1 with x_0/ℓ and find $\sqrt{c}$.
9. Calculate $L = h/(\sqrt{c}\,\ell G_m)$.
10. Enter Fig. 1.8.2 with L and find N.
11. Calculate $K = (W_2 G_m^2)/(2hN)$. This is the initial spring rate of the suspension in the direction of x_0.
12. Calculate $f = 3.13\,(K/W_2)^{\frac{1}{2}}$. This is the natural frequency of vibration (cycles per second) of the suspension for small amplitudes in the x_0 direction. This should not be close to the natural frequency of

vibration of any element in the packaged item, which should be determined by test beforehand (see, also, Section 4.2). *In any case it is advisable to provide damping for the suspension.*

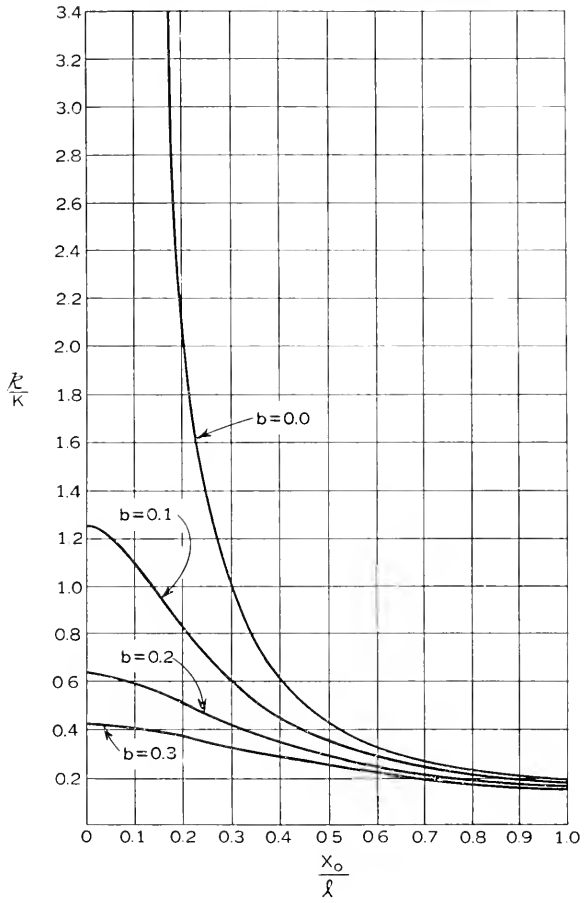


Fig. 1.8.3—Tension spring package design curve. Equation (1.7.10).

13. Enter Fig. 1.8.3 with x_0/ℓ and find k/K. If a four-spring package is desired, instead of an eight-spring package, (see Section 1.7) the value of k/K found on Fig. 1.8.3 should be multiplied by two before entering it on Line 13 in Table I. This is the only change required in the procedure.
14. Calculate $k = K\left(\frac{k}{K}\right)$. This is the spring rate of each of the springs in pounds per inch.

15. Calculate $B = (2W_2h)/(Kc\ell^2)$.
16. Enter Fig. 1.8.4(a) or (b) with B and find $d_m/(\sqrt{c}\,\ell)$.
17. Calculate $d_m/\ell = \sqrt{c}\cdot d_m/\sqrt{c}\,\ell$.
18. Calculate $d_m = \ell\cdot d_m/\ell$.
19. Calculate $(d_m/\ell) + (x_0/\ell)$.

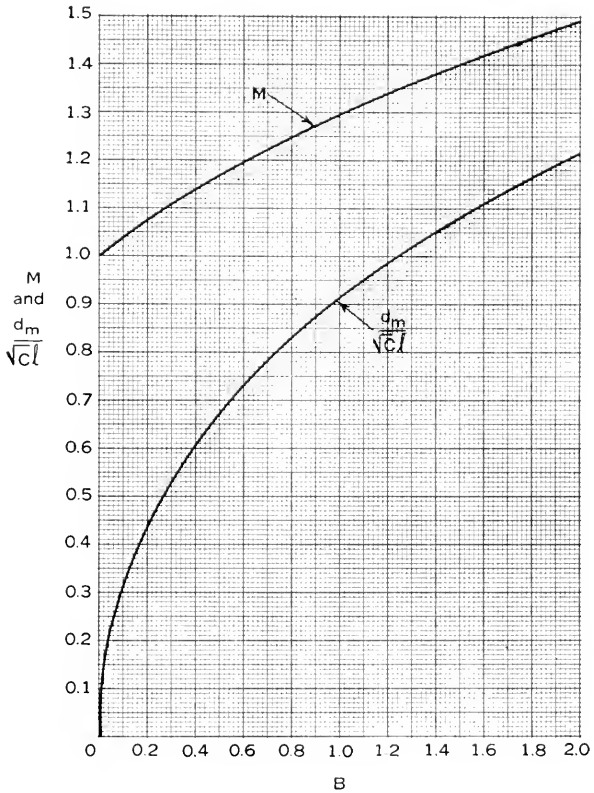


Fig. 1.8.4(a)—Tension spring package design curve. Equations (1.8.3) and (1.8.4).

20. Enter Fig. 1.8.5 with $(d_m/\ell) + (x_0/\ell)$ and find $(e/\ell) - b$. e is the stretch of each spring (in inches) when the displacement is d_m inches.
21. Calculate $F_m = k\cdot(e/\ell)\cdot\ell$. This is the maximum load (in pounds) on each spring.

22, 23, 24, 25, 26. These are the coil diameter, wire diameter, number of turns, fiber stress and length of coils. These quantities are calculated from the ordinary formulas, charts or slide rules for helical springs, using the values of k and F_m from Lines 14 and 21.

27. The length inside hooks is entered on Line 27 to group all of the spring specifications.

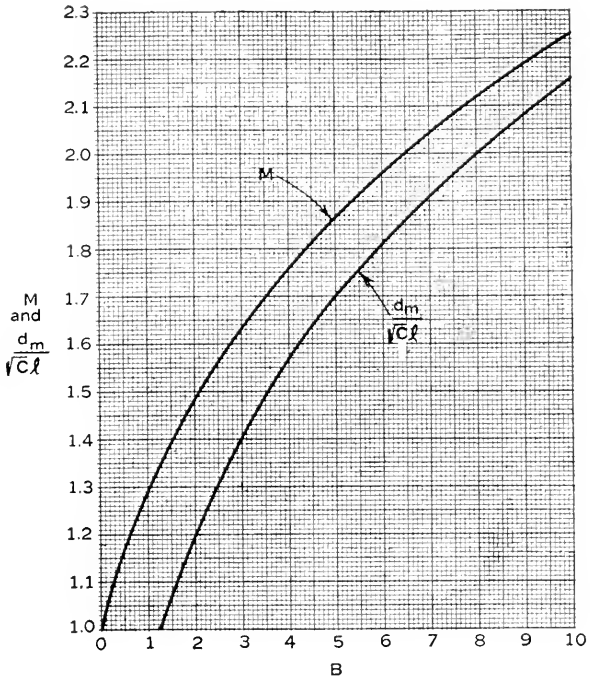


Fig. 1.8.4(b)—Tension spring package design curve. Equations (1.8.3) and (1.8.4).

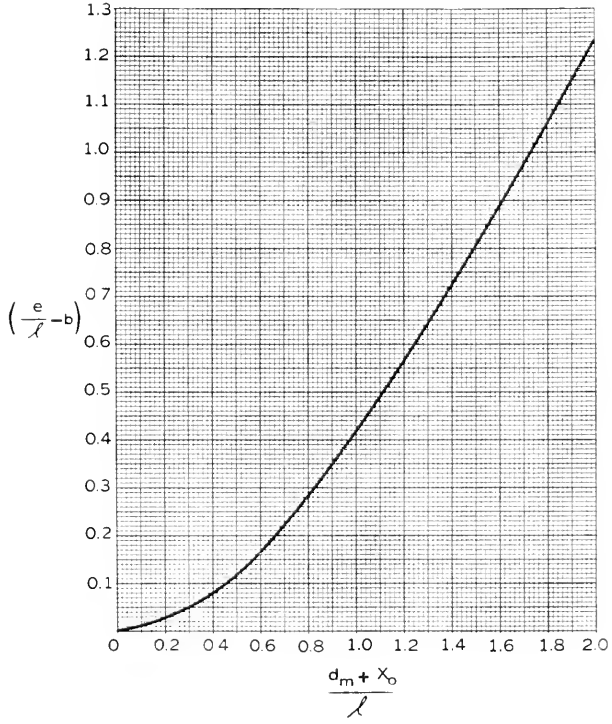


Fig. 1.8.5—Tension spring package design curve. Equation (1.8.5).

As an example of the calculations, Table I contains the entries for the design of springs for a 21 pound article (including $\frac{1}{3}$ the estimated spring weight) which is to be packaged so as not to exceed $35g$ in a five-foot drop.

TABLE I

Computation Form for Tension Spring Packages

1. W_2 (lbs.)	21
2. h (ins.)	60
3. G_m	35
4. x_0 (ins.)	5
5. ℓ (ins.)	7.07
6. b	0
7. Calc. x_0/ℓ	0.707
8. Find $\sqrt{c}$ from Fig. 1.8.1	0.91
9. Calc. $h/\sqrt{c}\,\ell G_m = L$	0.269
10. Find N from Fig. 1.8.2	1.265
11. Calc. $W_2G_m^2/2hN = K$ (lbs/in.)	169.0
12. Calc. $f = 3.13\ (K/W_2)^{\frac{1}{2}}$ (cyc./sec.)	8.9
13. Find k/K from Fig. 1.8.3	0.274
14. Calc. $k = K \cdot k/K$ (lbs/in.)	46.5
15. Calc. $B = 2W_2h/Kc\ell^2$	0.368
16. Find $d_m/\sqrt{c}\,\ell$ from Fig. 1.8.4	0.575
17. Calc. $d_m/\ell = \sqrt{c} \cdot d_m/\sqrt{c}\ell$	0.518
18. Calc. $d_m = \ell \cdot d_m/\ell$ (ins.)	3.68
19. Calc. $d_m/\ell + x_0/\ell$	1.220
20. Find e/ℓ from Fig. 1.8.5 and line 6	0.580
21. Calc. $F_m = k \cdot e/\ell \cdot \ell$ (lbs)	191.0
22. Coil diameter (ins.)	1.40
23. Wire diameter (ins.)	0.207
24. Number of turns	19
25. Fiber Stress (lbs./sq. in.) $\cdot\ 10^{-3}$	80
26. Length of Coils (ins.)	3.93
27. Length inside hooks (ins.)	7.07

1.9 Cushioning with Tangent Elasticity (Class C)

This is one of the most frequently encountered classes of cushioning since it includes a very common type of bottoming (Figs. 1.4.3 and 1.4.7). The load-displacement function (equation (1.4.3)) takes into account the fact that the cushioning can be compressed only to a definite amount d_b.

To find formulas for maximum acceleration and displacement, we proceed as follows. Substitute equation (1.4.3) in (1.2.15) and perform the integration, obtaining

$$\frac{4k_0d_b^2}{\pi^2}\log\cos\frac{\pi d_m}{2d_b} = -W_2h, \tag{1.9.1}$$

which may be written as

$$\tan\frac{\pi d_m}{2d_b} = \sqrt{\exp\left(\frac{\pi^2 W_2 h}{2k_0 d_b^2}\right) - 1}\,. \tag{1.9.2}$$

Equation (1.9.2) can then be substituted into (1.4.3) to obtain the maximum force P_m in accordance with (1.2.17):

$$P_m = \frac{2k_0 d_b}{\pi} \sqrt{\exp\left(\frac{\pi^2 W_2 h}{2k_0 d_b^2}\right) - 1}. \tag{1.9.3}$$

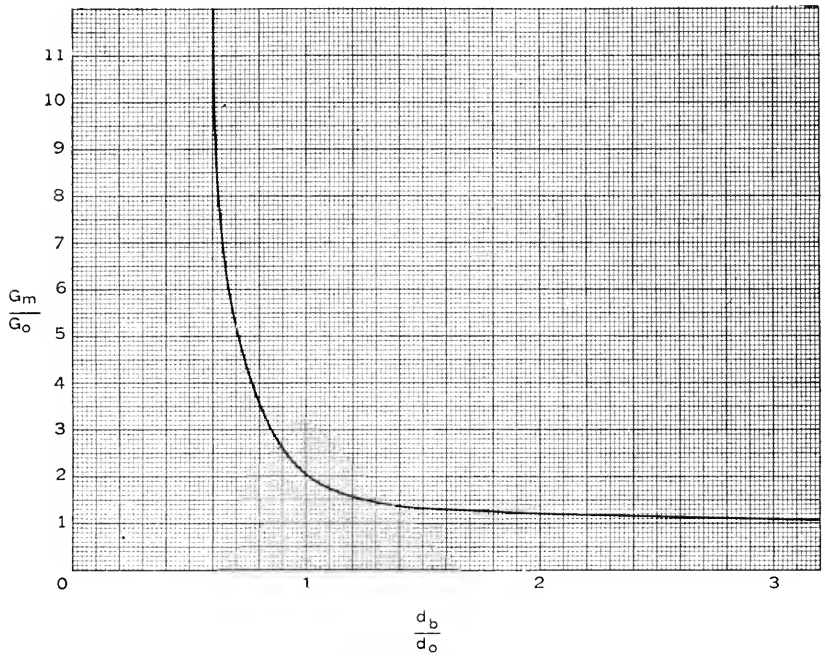


Fig. 1.9.1—Curve for finding maximum acceleration for cushioning with tangent elasticity. See equation (1.9.4).

The maximum acceleration is then obtained from (1.2.16) and may be written in the form

$$\frac{G_m}{G_0} = \frac{2d_b}{\pi d_0} \sqrt{\exp\left(\frac{\pi d_0}{2d_b}\right)^2 - 1}, \tag{1.9.4}$$

where d_0 and G_0 are defined just as in (1.5.2) and (1.5.5). G_0 is the maximum acceleration that would obtain if the cushioning did not bottom, that is, if the spring rate remained constant at its initial value k_0. d_0 is the maximum displacement that would be reached under the same linear conditions. Hence G_m/G_0 is a multiplying factor to be applied to a hypothetical linear cushioning to take into account the effect of bottoming. The multiplying factor depends only on the ratio (d_b/d_0) of the amount of space actually available to the amount of space that would be used under linear conditions.

The ratio G_m/G_0 is plotted against the ratio d_b/d_0 in Fig. 1.9.1. It may be seen that the multiplying factor increases very rapidly as the displacement ratio (d_b/d_0) falls below unity. For example, if the cushioning, with tangent elasticity, reaches hard bottoming (d_b) when only 80% of the required displacement (d_0) is attained, the acceleration is multiplied by 3.5; if only 60% of the required displacement is available, the acceleration is multiplied by 11.5.

Example: To illustrate with a numerical example, consider the case already discussed in Section 1.3, where we found that a spring rate of 694 lbs/in and a displacement of 1.44 inches were required to limit a 20-pound article

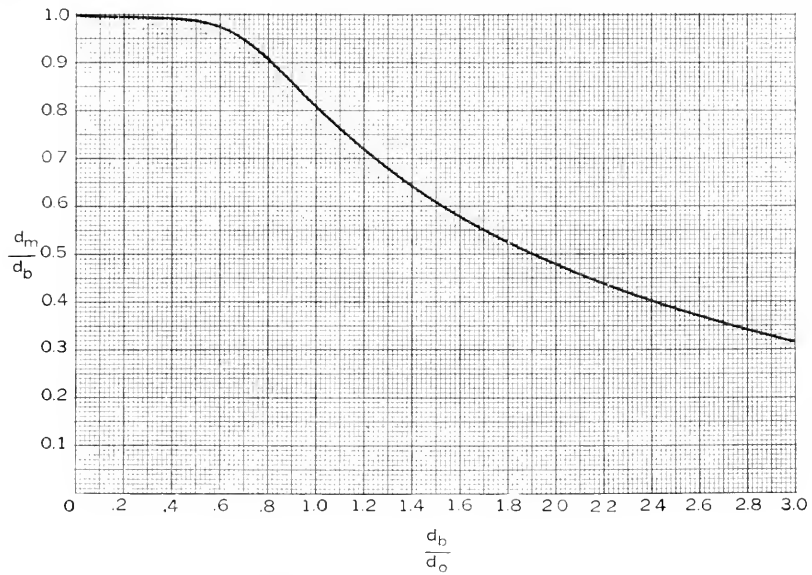


Fig. 1.9.2—Curve for finding maximum displacement for cushioning with tangent elasticity. See equation (1.9.5).

to an acceleration of 50g in a 3-foot drop with linear cushioning. Let us suppose that only 1.15 inches are available, instead of 1.44 inches, and that the cushioning has tangent elasticity starting with a spring rate of 694 lbs./in. Entering the curve of Fig. 1.9.1 at $d_b/d_0 = 1.15/1.44$ we find $G_m/G_0 = 3.5$. Hence the maximum acceleration will be 175g instead of the required 50g. This illustrates the wide variations in acceleration that may occur as a result of minor variations of dimensions in high G packages.

It is not necessarily true that the 175g test is 3.5 times as severe as the 50g test for all elements of the supported structure, since the severity depends also on the shape and scale of the acceleration-time relation. The factor may be more or less than 3.5 but it will be very close to this value for

all high-frequency elements of the structure. This subject is treated in detail in Parts II and III.

The maximum displacement d_m, in the case of tangent elasticity, may be calculated from equation (1.9.2) or, in terms of d_b/d_0, from

$$\frac{d_m}{d_b} = \frac{2}{\pi} \cos^{-1} \exp\left[-\frac{\pi^2}{8}\left(\frac{d_0}{d_b}\right)^2\right]. \tag{1.9.5}$$

The ratio d_m/d_b is plotted against d_b/d_0 in Fig. 1.9.2.

The use of Fig. 1.9.2 can be illustrated with the example already calculated, in which $d_b/d_0 = 1.15/1.44 = 0.8$. Entering the abscissa of Fig. 1.9.2 with $d_b/d_0 = 0.8$ we find $d_m/d_b = 0.915$. Hence the maximum displacement will be $0.915 \times 1.15 = 1.05$ inches.

1.10 Optimum Shape of Load-Displacement Curve for Tangent Elasticity

It is possible to choose the best shape for the load-displacement curve of the cushioning from those represented in Fig. 1.4.7. This will be, of course, not the best of all possible curves, but only the best among "tangent elasticity" curves. The best shape is defined as the one that yields the smallest maximum acceleration (G_m) for a given weight (W_2), height of drop (h) and available space (d_b). This leaves the initial spring rate (k_0) as the only remaining variable. To find its optimum value (say k_0'), set equal to zero the derivative of G_m (equation (1.9.4)) with respect to k_0, remembering that G_0 and d_0 are functions of k_0. The result is

$$\left(1 - \frac{\pi^2 W_2 h}{4 d_b^2 k_0'}\right) \exp\left(\frac{\pi^2 W_2 h}{2 d_b^2 k_0'}\right) - 1 = 0, \tag{1.10.1}$$

from which

$$k_0' = \frac{3.1 W_2 h}{d_b^2}. \tag{1.10.2}$$

Substituting (1.10.2) in (1.9.4) we find the minimum value (G_m') of maximum acceleration to be

$$G_m' = \frac{3.9h}{d_b}. \tag{1.10.3}$$

To illustrate the application of equations (1.10.2) and (1.10.3), consider again the case of the 20-pound article dropped from a height of three feet. We found that a linear spring, with a spring constant of 694 lbs/in, would limit the maximum acceleration to $50g$ if 1.44 inches of displacement were available. If only 1.15 inches of displacement are available, and the initial spring rate is kept at 694 lbs/in, we found the maximum acceleration to be

175g if the cushioning bottoms with tangent elasticity. Now, according to equation (1.10.2), the best initial spring rate for cushioning with tangent elasticity would be

$$k_0' = \frac{3.1 \times 20 \times 36}{(1.15)^2} = 1690 \text{ lbs/in.}$$

In this case, equation (1.10.3) gives, for the maximum acceleration,

$$G_m' = \frac{3.9 \times 36}{1.15} = 122g.$$

Hence, confronted with a space limitation less than that required for a 50g linear spring, it is better to use an initial spring rate higher than that for the 50g linear spring in order to strike an economical balance between displacement and bottoming. The best balance, among cushionings having tangent elasticity, is obtained by using equation (1.10.2).

If no factor of safety is considered, it would be still better not to use a bottoming type of cushion at all. From equations (1.3.5) and (1.3.3) it can be seen that a linear spring with a constant of 1090 lbs/in will give only 63g with a displacement of 1.15 inches. Such a spring, though, would bottom very sharply at a drop slightly higher than 3 ft. and would give an acceleration much greater than cushioning with tangent elasticity which bottoms more gradually. This may be important if there are high-frequency, brittle elements in the packaged article (see Part III).

1.11 Procedure for Finding Maximum Acceleration and Displacement for Cushioning with Tangent Elasticity (Class C)

To illustrate the use of the equations and curves for Class C cushioning, the same example used for Class B will be used, as it was observed that the experimental load-displacement curve in that example (Fig. 1.6.1) is a border line one which can be treated as either B or C.

By laying a straight edge along the first part of the curve (Fig. 1.6.1), the average initial spring rate is found to be 305 lbs/in. This value is taken as k_0 in the present case.

The next step is to find a value of d_b such that a graph of P/d_b vs x_2/d_b will fall slightly above the curve $k_0 = 30(0)$ in Fig. 1.4.7; d_b must be greater than 2 inches, since that displacement was obtained in the experiment. As a trial take $d_b = 2.25$ inches and test it at one point, say the experimental point $P = 300$ lbs., $x_2 = \frac{7}{8}$ in. Then $P/d_b = 133$ and $x_2/d_b = 0.39$. The point (0.39, 133) falls below the curve $k_0 = 30(0)$ in Fig. 1.4.7. Next try $d_b = 2.5$ inches. In this case, for the experimental point $P =$

300, $x_2 = \frac{7}{8}$, we find $P/d_b = 120$, $x_2/d_b = .35$. This falls slightly above the $k_0 = 30(0)$ curve as required. The whole experimental curve is then plotted to the coordinates $P/2.5$ vs $x_2/2.5$ and is found to fit as closely as necessary. Hence the parameters are adopted as $k_0 = 305$, $d_b = 2.5$.

We can now calculate the maximum acceleration that the tube will receive in, say, a three-foot drop test. First calculate, from equations (1) and (2),

$$d_0 = \sqrt{\frac{2hW_2}{k_0}} = \sqrt{\frac{2 \times 36 \times 22.5}{305}} = 2.31,$$

$$G_0 = \sqrt{\frac{2hk_0}{W_2}} = \sqrt{\frac{2 \times 36 \times 305}{22.5}} = 31.3.$$

Then $d_b/d_0 = 1.08$. Entering Fig. 4 with this value we find $G_m/G_0 = 1.82$. Hence the maximum acceleration is:

$$G_m = 31.3 \times 1.82 = 57g.$$

Finally, entering Fig. 5 with $d_b/d_0 = 1.08$ we find $d_m/d_b = 0.8$. Hence the maximum displacement is $d_m = 0.8 \times 2.5 = 2.0$ inches. This indicates that the load-displacement test was carried far enough to cover the range up to a three-foot drop.

It may be observed that the results obtained, by treating the same data as Class B or Class C cushioning, agree within a few per cent. This is because, in the example chosen, both B and C curves can be made to fit the experimental load-displacement curve.

1.12 Consequences of Abrupt Bottoming (Class D)

It is useful to examine cushioning systems that can bottom more abruptly than Class C cushioning. Abrupt bottoming is possible, for example, in a tension spring package lacking a snubbing device. An estimate of the increase in acceleration can be made by studying the case of bilinear elasticity (Fig. 1.4.4). Here we have a spring rate k_0 up to a displacement d_s, following which the cushioning has a different spring rate k_b. k_0 represents the average spring rate before bottoming and k_b can represent the much greater stiffness of the wall of the container.

If $d_0 > d_s$, that is, if

$$\sqrt{\frac{2W_2h}{k_0}} > d_s, \tag{1.12.1}$$

the suspended article will bottom and the maximum displacement and acceleration are obtained by using both of the equations (1.4.4) in evaluating the integral in (1.2.15). Thus,

$$\int_0^{d_s} k_0 x_2 \, dx_2 + \int_{d_s}^{d_m} [k_b x_2 - (k_b - k_0)d_s] \, dx_2 = W_2 h. \tag{1.12.2}$$

The remainder of the procedure for finding G_m is the same as before. The value of d_m found from (1.12.2) is substituted for x_2 in the second of

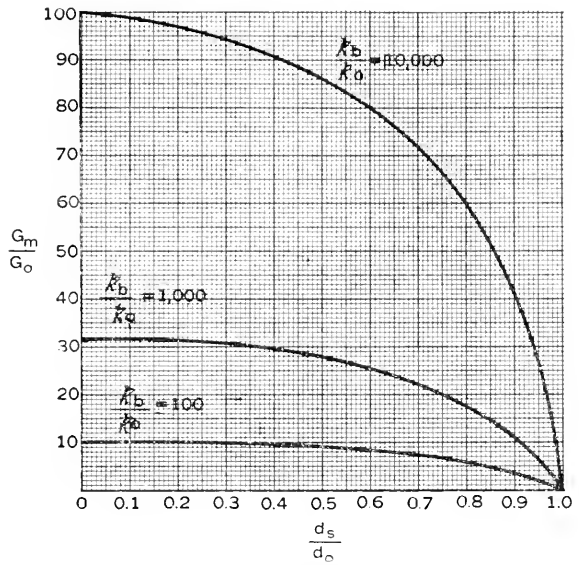


Fig. 1.12.1—Curves for finding maximum acceleration as a result of abrupt bottoming. See equation (1.12.3).

(1.4.4) and the value of P_m, thus obtained, is used, in (1.2.16), with the result:

$$G_m = G_0 \sqrt{\frac{k_b}{k_0} + \frac{d_s}{d_0^2}\left(1 - \frac{k_b}{k_0}\right)}, \tag{1.12.3}$$

where G_0 is the acceleration that would be reached if a displacement d_0 were available:

$$G_0 = \sqrt{\frac{2hk_0}{W_2}}. \tag{1.12.4}$$

The ratio G_m/G_0 is plotted against d_s/d_0 in Fig. 1.12.1 for several values of k_b/k_0. Since, in practice, k_b might be thousands of times as great as k_0, it may be seen that the increase in maximum acceleration can be very large even when d_s is only slightly less than d_0. It is apparent that a snubbing device is desirable in a tension spring suspension. This is especially true when considering high-frequency elements of the packaged article. It will be shown, in Part III, that low-frequency elements are not affected as much as might be expected from consideration of maximum acceleration alone.

1.13 Cushioning with Hyperbolic Tangent Elasticity (Class E)

In the preceding sections, there have been considered four types of elasticity (linear, cubic, tangent and bilinear) that fit the load-displacement characteristics of the more common cushioning materials and devices. There now remains the problem of finding more nearly ideal shapes of elasticity. By "more nearly ideal" is meant a shape which will result in a smaller maximum displacement for a given maximum acceleration. This is important in the packaging of very delicate articles if shipping space is limited.

It may be observed (equation (1.2.15)) that the total area under the load-displacement curve is equal to the maximum energy of the system. The maximum ordinate of the enclosed area is proportional to the maximum acceleration. Hence, if we wish to (1) limit the maximum acceleration (2) accomodate a given kinetic energy and (3) have as small a displacement as possible, the best shape for the load displacement function is $P =$ constant, where the constant is the product of the supported mass and the maximum allowable acceleration.

It is not practical to obtain this ideal shape exactly, for there will always be a finite initial spring rate and a rounding off of the load-displacement curve to the limiting maximum load. A function which represents this practical condition (and also includes the ideal case) is the hyperbolic tangent function mentioned in Section 1.4:

$$P = P_0 \tanh \frac{k_0 x_2}{P_0}. \tag{1.13.1}$$

The formulas for maximum acceleration and displacement are found in the same way as for the other classes of cushioning with the results:

$$d_m = \frac{P_0}{k_0} \cosh^{-1} \exp\left(\frac{W_2 h k_0}{P_0^2}\right) \tag{1.13.2}$$

or

$$d_m = \frac{d_0 P_0}{W_2 G_0} \cosh^{-1} \exp\left(\frac{W_2^2 G_0^2}{2P_0^2}\right) \tag{1.13.3}$$

and

$$G_m = \frac{P_0}{W_2} \tanh \frac{k_0 d_m}{P_0} \tag{1.13.4}$$

or

$$G_m = \frac{P_0}{W_2} \sqrt{1 - \exp\left(-\frac{W_2^2 G_0^2}{P_0^2}\right)} \tag{1.13.5}$$

where, as before

$$d_0 = \sqrt{\frac{2hW_2}{k_0}}, \qquad G_0 = \sqrt{\frac{2hk_0}{W_2}}.$$

Equations (1.13.3) and (1.13.5) are plotted, in Figs. 1.13.1 and 1.13.2, against the dimensionless parameter P_0/W_2G_0. The latter is the ratio of the maximum force, that the hyperbolic tangent cushioning will transmit,

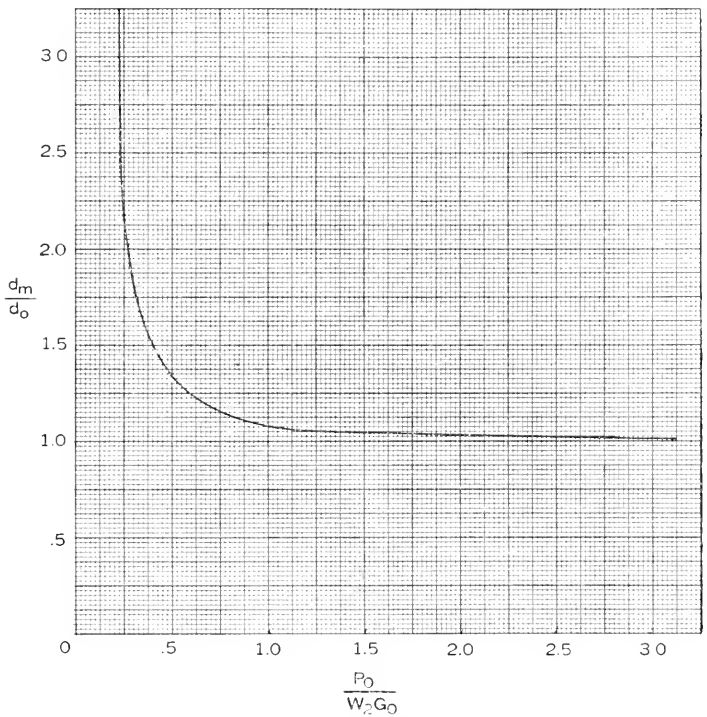


Fig. 1.13.1—Maximum displacement for cushioning with hyperbolic tangent elasticity. See equation (1.13.3).

to the force that linear cushioning would transmit under the conditions specified.

To find the value of k_0 which yields the minimum value of acceleration for a given maximum displacement, differentiate (1.13.4) with respect to k_0 and set the result equal to zero:

$$\operatorname{sech}^2 \frac{k_0 d_m}{P_0} = 0. \tag{1.13.6}$$

This is satisfied by $k_0 \to \infty$, which represents the rectangular load displacement curve and confirms the conclusion reached from energy considerations.

Taking the limit of (1.13.4) as $k_0 \to \infty$, we find the optimum acceleration to be

$$G'_m = \frac{P_0}{W_2}. \tag{1.13.7}$$

The corresponding maximum displacement is found, from (1.13.2) to be

$$d'_m = \frac{W_2 h}{P_0} = \frac{h}{G'_m}. \tag{1.13.8}$$

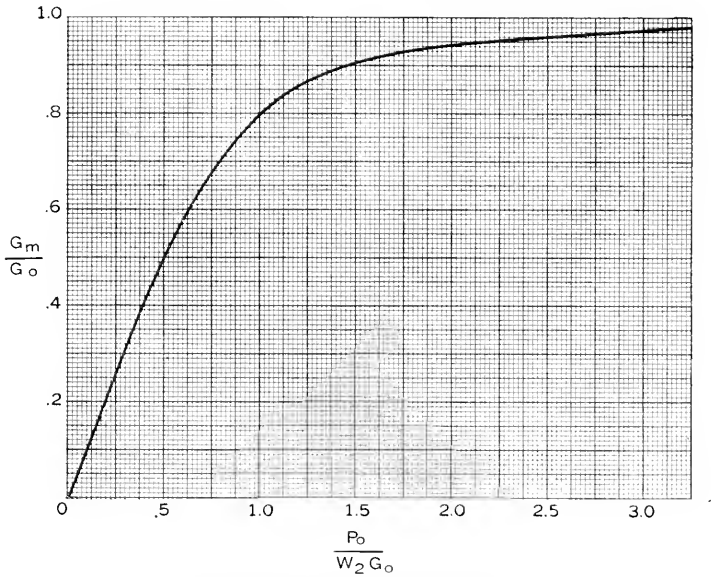


Fig. 1.13.2—Maximum acceleration for cushioning with hyperbolic tangent elasticity. See equation (1.13.5).

1.14 Minimum Space Requirements for Various Classes of Cushioning

It is interesting to compare the minimum amount of space for displacement that can be attained with the various kinds of cushioning that have been discussed.

Hyperbolic Tangent Elasticity	$d'_m = \frac{h}{G'_m}$
Linear Elasticity	$d_m = \frac{2h}{G_m}$
Tangent Elasticity	$d'_m = \frac{3.9h}{G'_m}$

Cubic elasticity will give a d_m somewhat more or less than $2h/G_m$ depending upon whether the parameter r is positive or negative.

It is seen that a factor of almost four can be gained, in the linear dimensions of the cushioning space required, by replacing the tangent type of cushioning with the hyperbolic tangent type.

There are several ways of obtaining a load-displacement curve with a shape similar to the hyperbolic tangent curve. One of the most interesting is suggested by the fact that the load-displacement curve of a strut has approximately this shape. Hence a bristle brush has the proper characteristics.

TABLE II

1	2	3	4	5	6	7	8
n	$\Delta(x_2)_n$	$(x_2)_n$	P_n	$\Delta A_n = \frac{1}{2}\Delta(x_2)_n \times (P_n + P_{n-1})$	$A_n = A_{n-1} + \Delta A_n$	$h_n = \frac{A_n}{W_2}$	$G_n = \frac{P_n}{W_2}$
0	—	0	0	0	0	0	0
1	0.500	0.500	120	30	30	1.6	6.5
2	0.100	0.600	150	13.5	43.5	2.4	8.1
3	0.100	0.700	205	17.0	61.3	3.3	11.1
4	0.100	0.800	290	24.8	86.1	4.7	15.7
5	0.100	0.900	410	35.0	121.1	6.6	22.2
6	0.100	1.000	585	49.8	170.9	9.2	31.6
7	0.050	1.050	730	32.9	203.8	11.1	39.5
8	0.050	1.100	950	42.0	245.8	13.3	51.4
9	0.050	1.150	1370	58.0	303.8	16.4	74.0
10	0.025	1.175	1680	38.1	341.9	18.5	91.0
11	0.025	1.200	2240	49.0	390.9	21.1	121.0
12	0.0125	1.2125	2620	30.4	421.3	22.8	141.5
13	0.0125	1.225	3200	36.4	457.7	24.7	173.0

1.15 Numerical Method for Analyzing Class F Cushioning

The numerical method to be described is one that has been adapted from a graphical one used by the Committee on Packing and Handling of Radio Valves of the British Radio Board. The method has advantages of simplicity in concept and ease of application, especially when the load-displacement curve of the cushioning does not resemble closely one of the Classes A to E described above. It has the disadvantage that it does not yield, directly, numerical factors by which the spring rate or depth of cushioning should be changed in the event that the analysis reveals inadequate or more than adequate protection.

The method is based on the fact that the area under the load-displacement curve of the cushioning represents the energy stored in, or absorbed by, the cushion. The total amount of energy that must be transferred is equal to the product of the weight (W_2) of the suspended item and the height (h) of drop. By finding the abscissa (x_2) and its ordinate (P) which

include an area W_2h, the maximum displacement is immediately x_2 and the maximum acceleration is the quotient P/W_2, in accordance with equations (1.2.15) and (1.2.16).

As actually applied in the present instance, the British method was modified slightly to make the procedure a routine numerical one. The

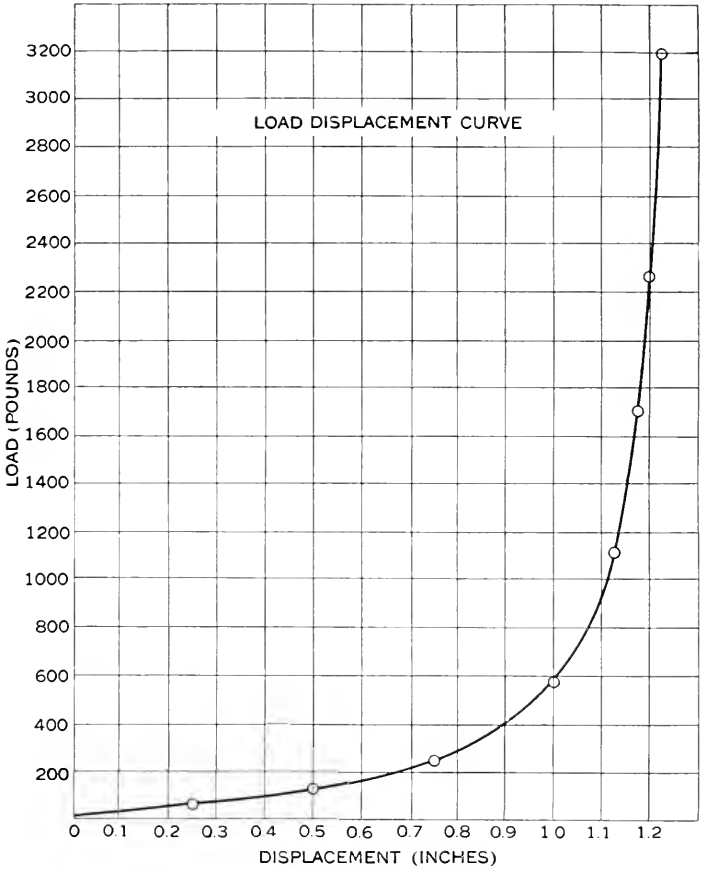


Fig. 1.15.1—Experimental load-displacement curve for Table II.

computing form is given in detail in Table II, in which the data are taken from an experimental load-displacement curve (Fig. 1.15.1) for the end spring pads of a vacuum tube package. The load-displacement data are listed in Columns 3 and 4 of Table II. The meaning of each column in the table is as follows.

Column 1. $n(= 1, 2, 3\cdots)$ is the number that identifies the displacement (and corresponding load) up to which the area under the load-displacement curve is to be calculated.

Column 2. $\Delta(x_2)_n$ is the increment of displacement between $(x_2)_{n-1}$ and $(x_2)_n$. $\Delta(x_2)_n = (x_2)_n - (x_2)_{n-1}$, see Fig. 1.15.2. Note that, as the curve becomes steeper, $\Delta(x_2)_n$ is taken smaller for better accuracy.

Column 3. $(x_2)_n$ is the displacement associated with the n^{th} point (see Fig. 1.15.2).

Column 4. P_n is the load that produces displacement $(x_2)_n$.

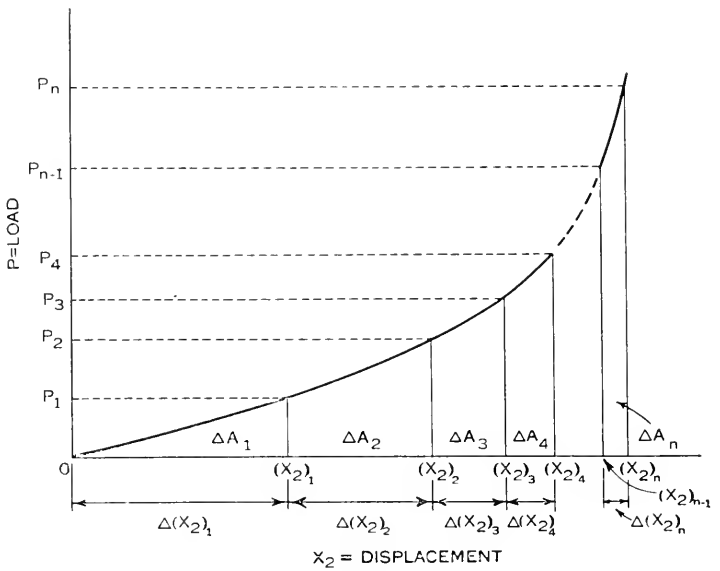


Fig. 1.15.2—Graphical illustration of numerical method of calculating area under load-displacement curve. See Table II.

Column 5. $\Delta A_n = \frac{1}{2}\Delta(x_2)_n(P_{n-1} + P_n)$ is the area of the trapezoid with altitude $\Delta(x_2)_n$ and bases P_{n-1} and P_n. It is approximately the energy absorbed by the cushioning in displacing from $(x_2)_{n-1}$ to $(x_2)_n$.

Column 6. A_n is the sum of all the trapezoidal areas from $x_2 = 0$ to $x_2 = (x_2)_n$. It is approximately the total energy the cushioning can absorb in displacing an amount $(x_2)_n$ beginning at zero displacement. Note that A_0 is always equal to zero.

Column 7. $h_n = A_n/W_2$ is the height of fall that will cause the cushioning to displace an amount $(x_2)_n$. In Table II, $W_2 = 18.5$ pounds.

Column 8. $G_n = P_n/W_2$ is the maximum acceleration experienced by the suspended mass when dropped from a height h_n.

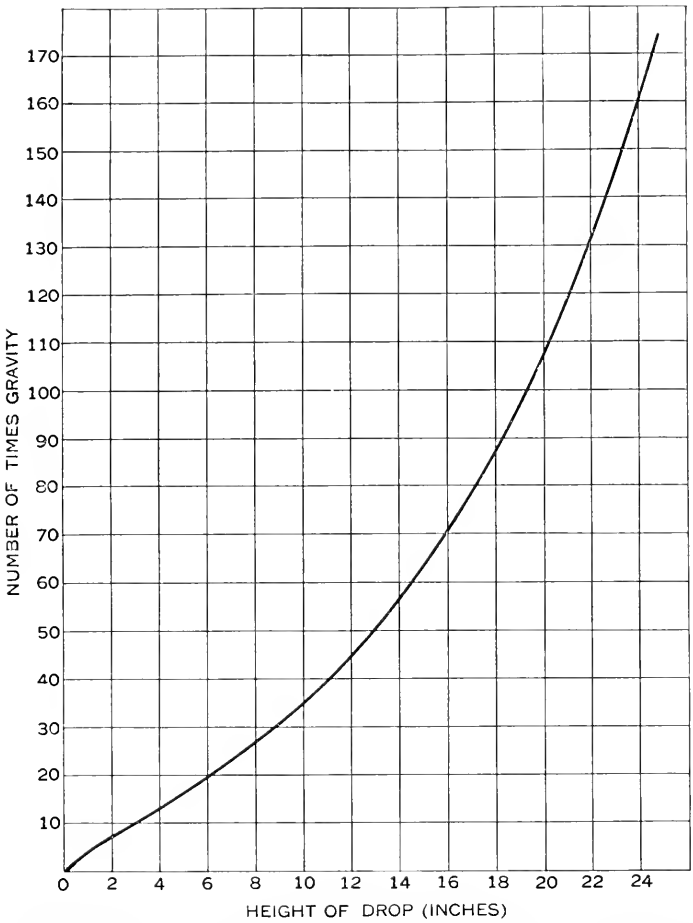


Fig. 1.15.3—Maximum acceleration vs. height of drop for an 18.5 pound article supported on cushioning with the load-displacement curve of Fig. 1.15.1. See Table II.

From the last two columns of the table a curve of height of drop vs. the corresponding acceleration may be plotted as in Fig. 1.15.3.

PART II

ACCELERATION-TIME RELATIONS

2.1 Introduction

In Part I we were concerned primarily with the *maximum* acceleration of the packaged item. In this part we shall study the details of the variation of acceleration with time in order to have this information available for our

study, in Part III, of its influence on the response of elements of the packaged item.

The first case to be considered will be the simple single mass and linear spring example described in Sections 1.2 and 1.3. Following this the phenomenon of rebound of the package will be considered. The influence of velocity damping and dry friction will be studied; and, finally, the effects of non-linearity of the cushion elasticity on the acceleration-time relation will be investigated.

2.2 Acceleration-time Relation for Linear Elasticity

Returning to the elementary example studied in Sections 1.2 and 1.3, we first write the equation of motion for the mass m_2, on its linear spring of spring rate k_2 (see Fig. 1.2.1.). Equation (1.2.3) becomes

$$m_2\ddot{x}_2 + k_2x_2 = m_2g. \tag{2.2.1}$$

Using the initial conditions

$$[x_2]_{t=0} = 0, \tag{2.2.2}$$

$$[\dot{x}_2]_{t=0} = \sqrt{2gh}, \tag{2.2.3}$$

the solution of (2.2.1) is

$$x_2 = \frac{\sqrt{g^2 + 2\omega_2^2 gh}}{\omega_2^2} \sin(\omega_2 t - \alpha) + \frac{g}{\omega_2^2} \tag{2.2.4}$$

or

$$x_2 = \sqrt{\frac{W_2^2}{k_2^2} + d_m^2} \sin(\omega_2 t - \alpha) + \frac{W_2}{k_2}, \tag{2.2.5}$$

where

$$\omega_2 = \sqrt{\frac{k_2}{m_2}} = 2\pi f_2 = \frac{2\pi}{T_2} \tag{2.2.6}$$

and

$$\alpha = \tan^{-1} \frac{g}{\omega_2 \sqrt{2gh}} = \tan^{-1} \frac{W_2}{k_2 d_m}. \tag{2.2.7}$$

ω_2 is the circular frequency, f_2 is the frequency and T_2 is the period of vibration of the mass m_2 on its spring; d_m has the same definition as in Section 1.3 (equation (1.3.1)).

Now, W_2/k_2 is the static displacement of the mass m_2 on its spring. This is usually very small in comparison with the maximum displacement (d_m) during impact. Hence W_2/k_2 will be neglected, and (2.2.5) becomes

$$x_2 = d_m \sin \omega_2 t. \tag{2.2.8}$$

Differentiating (2.2.8) twice with respect to t, we find, for the acceleration

$$\ddot{x}_2 = -\omega_2^2 d_m \sin \omega_2 t = -\omega_2 \sqrt{2gh} \sin \omega_2 t. \tag{2.2.9}$$

Hence the absolute magnitude of the maximum acceleration is

$$G_m = \frac{|\ddot{x}_2|_{\max}}{g} = \frac{\omega_2^2 d_m}{g} = \sqrt{\frac{2hk_2}{W_2}} \tag{2.2.10}$$

as before.

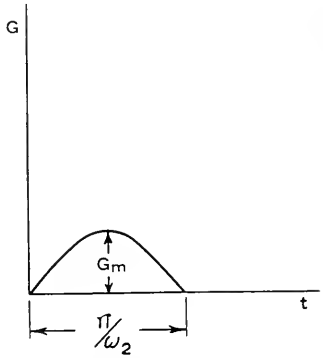


Fig. 2.2.1

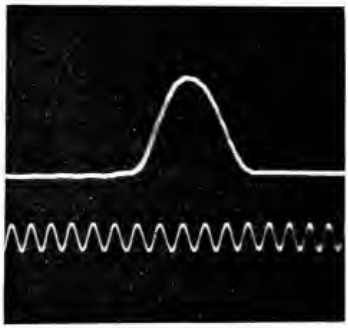


Fig. 2.2.2

Fig. 2.2.1—Half-sine-wave pulse acceleration. See equation (2.2.9).
Fig. 2.2.2—Oscillogram of a half-sine-wave pulse obtained with a piezo-crystal accelerometer.

Equation (2.2.9) shows that the acceleration varies sinusoidally with time. It rises from its initial zero value to its maximum in a time $\pi/2\omega_2$, at which time the displacement also reaches its maximum value. The acceleration returns to zero again at time π/ω_2. At this time the displacement is also zero. This is the end of the range of applicability of equation (2.2.9); for when t becomes slightly greater than π/ω_2, a tension in the spring is required. Since no mechanism, such as a large mass m_3 (Fig. 0.2.2), has been supplied, to allow a tension in the spring to develop, the system will rebound from the floor at the end of the half period π/ω_2. The acceleration is thus a half-sinusoidal pulse of duration $\pi/\omega_2 = T_2/2$ and amplitude $G_m g$ as illustrated in Fig. 2.2.1. An oscillogram of such a pulse obtained with a piezo-crystal accelerometer is shown in Fig. 2.2.2.

2.3 Package Rebound.

The presence of the mass of an outer container will affect the acceleration after the first half cycle of displacement. The outer container is represented by the mass m_3 in the general idealized system illustrated in Fig. 0.2.2 and in the simpler system (Fig. 2.3.1) that we shall consider now.

Two pairs of equations are necessary to describe the action of the system; one pair applies during the time of contact of m_3 with the floor and the second pair applies if rebound occurs.

The mass m_3 is assumed to be inelastic (see Section 0.2) so that, during the interval of its contact with the floor, the equation of motion for m_2 will be the same as before (2.2.1). In addition, there will be an equation of equilibrium for the mass m_3 :

$$R = k_2 x_2 + m_3 g \,, \tag{2.3.1}$$

where R is the upward force exerted by the floor on m_3 .

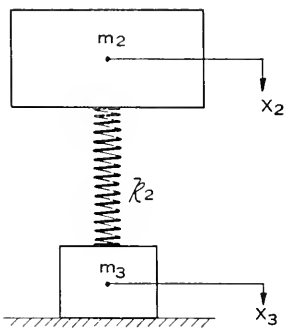


Fig. 2.3.1—Two-mass system representing packaged article, linear cushioning and outer container.

Equations (2.3.1) and (2.2.1) will hold as long as R is positive. To find out when $R > 0$, solve (2.2.1) for $k_2 x_2$ and substitute in (2.3.1):

$$R = W_2 + W_3 - m_2 \ddot{x}_2 \,. \tag{2.3.2}$$

That is, a necessary condition for rebound is that the mass of the cushioned article, multiplied by its maximum acceleration, exceeds the total weight of the package. The condition for rebound may be written

$$G_m > \frac{W_2 + W_3}{W_2} \,. \tag{2.3.3}$$

This is a necessary, but not a sufficient, condition for rebound because there will be energy losses as a result of damping and permanent deformation. G_m will generally have to be considerably greater than the right hand side of (2.3.3) for rebound to occur.

If rebound does not occur, equation (2.2.9) continues to apply, except for damping which will be considered in Section 2.5.

2.4 Motion After Rebound

If rebound occurs, the equations of motion for the two masses, m_2 and m_3, are

$$m_2\ddot{x}_2 + k_2(x_2 - x_3) = m_2 g, \tag{2.4.1}$$

$$m_3\ddot{x}_3 - k_2(x_2 - x_3) = m_3 g. \tag{2.4.2}$$

Multiplying (2.4.1) by m_3 and (2.4.2) by m_2 and subtracting, we find

$$m\ddot{y} + k_2 y = 0, \tag{2.4.3}$$

where

$$y = x_2 - x_3, \tag{2.4.4}$$

$$m = \frac{m_2 m_3}{m_2 + m_3}. \tag{2.4.5}$$

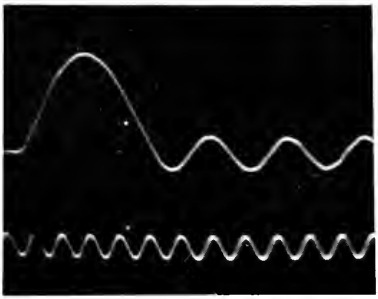


Fig. 2.4.1—Oscillogram illustrating the half-sine pulse followed by the higher frequency, lower amplitude vibration of the packaged article in a rebounding package.

Equation (2.4.3) is the equation governing the vibration of the two-mass system as a simple oscillator. The circular frequency of the vibration is

$$\omega = \sqrt{\frac{k_2}{m}} \tag{2.4.6}$$

and it may be noticed that this frequency is always greater than ω_2 (equation (2.2.6)). This fact is important in estimating the effect of vibrations on elements of the packaged item (Section 3.5).

ω is also the frequency of vibration of the packaged article during the interval of free fall. This vibration (usually of small amplitude) is initiated by the sudden release of the dead load displacement of the packaged article.

As an intermediate step in obtaining the acceleration after rebound we shall find the magnitude of the relative displacement (y) of the two masses. To do this it is necessary to solve equation (2.4.3) with the appropriate boundary conditions. Calling t_r the time at which m_3 leaves the floor, we

must find y and $\dot{y}$ at $t = t_r$. Since m_3 is motionless at $t = t_r$, the relative displacement and velocity at that time are identical with x_2 and $\dot{x}_2$ respectively. The former is simply the stretch of the spring necessary to just pull the mass m_3 off the floor, i.e.,

$$[y]_{t=t_r} = [x_2]_{t=t_r} = -\frac{W_3}{k_2}. \tag{2.4.7}$$

To find the velocity at $t = t_r$, substitute (2.4.7) in (2.2.4) and also substitute t_r for t in the latter. This gives an equation for determining t_r. Then, returning to (2.2.4), differentiate it once to obtain $\dot{x}_2$ and substitute for t the value t_r just found. The result is

$$[\dot{x}_2]_{t=t_r} = [\dot{y}]_{t=t_r} = -\sqrt{2gh - \frac{gW_3(2W_2 + W_3)}{W_2 k_2}}. \tag{2.4.8}$$

The solution of (2.4.3) with initial conditions (2.4.7) and (2.4.8) is

$$y = -\frac{1}{\omega}\sqrt{2gh - \frac{W_3 g}{k_2}}\sin(\omega t - \zeta), \tag{2.4.9}$$

where

$$\zeta = \omega t_r - \tan^{-1}\frac{\omega[y]_{t=t_r}}{[\dot{y}]_{t=t_r}}.$$

We are now in a position to find the acceleration of the packaged item after rebound. Substitute y of (2.4.9) for $x_2 - x_3$ in (2.4.1) to obtain

$$\ddot{x}_2 = g + \frac{\omega_2^2}{\omega}\sqrt{2gh - \frac{W_3 g}{k_2}}\sin(\omega t - \zeta). \tag{2.4.10}$$

To obtain a simple formula for the ratio of the maximum accelerations after and before rebound, let us assume that both are much greater than gravitational acceleration. Then if

G_r = maximum number of g's after rebound, (maximum of (2.4.10))

$G_m = \sqrt{\frac{2hk_2}{W_2}}$ = maximum number of g's before rebound,

we find, from (2.4.10), neglecting the term g outside the radical,

$$\frac{G_r}{G_m} = \frac{W_3}{W_2 + W_3}\sqrt{1 - \frac{W_3}{2hk_2}}. \tag{2.4.11}$$

Hence, the maximum acceleration after rebound is always less than the maximum acceleration before rebound. Therefore, conditions after rebound need only be examined when the frequency after rebound (see equation (2.4.6)) is near the natural frequency of vibration of a critical element of the packaged item (see Section 3.5).

The complete acceleration history of a rebounding package with undamped linear cushioning is thus a half sine wave pulse of amplitude $G_m = \sqrt{2hk_2/W_2}$ and duration π/ω_2 followed by an oscillating acceleration of amplitude given by (2.4.11) and frequency given by (2.4.6). Such a wave shape is shown in Fig. 2.4.1.

2.5 Influence of Damping on Acceleration

The presence of damping in cushioning is always desirable to prevent the building up of large amplitudes as a result of periodic disturbances. However, damping also has an effect on the maximum acceleration that is attained in a drop test. From the latter point of view there is an optimum amount of damping and an amount that should not be exceeded if the maximum undamped acceleration is not to be exceeded.

We shall consider the case of a linear cushion with damping proportional to velocity. The system is represented in Fig. 2.5.1. With the addition

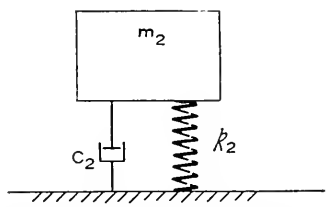


Fig. 2.5.1—Idealization of linear cushioning with velocity damping.

of the damping term the equation of motion of m_2, during contact of the package with the floor, is

$$m_2\ddot{x}_2 + c_2\dot{x}_2 + k_2x_2 = 0, \tag{2.5.1}$$

in which c_2 is the damping coefficient of the cushioning. Equation (2.5.1) is more conveniently expressed as

$$\ddot{x}_2 + 2\beta_2\omega_2\dot{x}_2 + \omega_2^2x_2 = 0, \tag{2.5.2}$$

where

$$\omega_2 = \sqrt{\frac{k_2}{m_2}}, \tag{2.5.3}$$

$$\beta_2 = \frac{c_2}{2m_2\omega_2}. \tag{2.5.4}$$

ω_2 is the undamped circular frequency of vibration of m_2 on its spring and β_2 is the fraction of critical damping. $\beta_2 = 0$ means no damping and $\beta_2 = 1$ means just enough damping so that there will be no oscillation if the packaged article is displaced and released.

The acceleration solution of (2.5.2), with the initial conditions of the drop test (see (2.2.2) and (2.2.3)) is

$$\ddot{x}_2 = -\frac{\omega_2\sqrt{2gh}}{\sqrt{1-\beta_2^2}}\, e^{-\beta_2\omega_2 t}\cos\left(\omega_2 t\sqrt{1-\beta_2^2}+\gamma\right) \tag{2.5.5}$$

where

$$\tan\gamma = \frac{2\beta_2^2-1}{2\beta_2\sqrt{1-\beta_2^2}}. \tag{2.5.6}$$

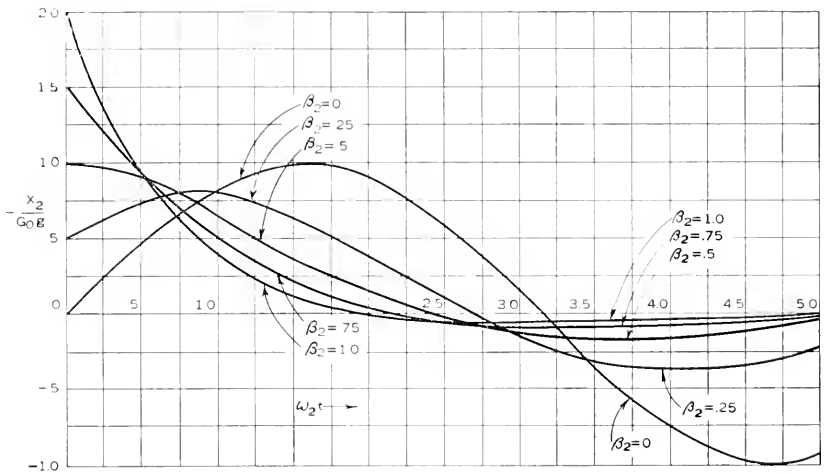


Fig. 2.5.2—Acceleration-time curves for linear cushioning with various amounts of damping (no rebound). See equation (2.5.5).

The acceleration is thus a damped sinusoid with an abruptly reached initial value whose magnitude depends upon the amount of damping. For small damping, the initial acceleration is small and then the acceleration increases, but never reaches the value that would be reached without any damping. For high damping ($\beta_2 > 0.5$) the initial value is greater than without any damping and falls off thereafter. Figure 2.5.2 shows the shapes of the acceleration time curves for several values of β_2. All of the curves are for no rebound. It may be seen, from equation (2.5.5) and Fig. 2.5.2 that the addition of damping changes the shape of the acceleration-time relation in three ways. First, a damped sinusoid replaces the pure sinusoid; second, the frequency is reduced; and, third, the initial phase is changed.

It is useful to consider in detail the effect of damping on *maximum* acceleration. Let

G_m = maximum number of g's with damping

$G_0 = \sqrt{\frac{2hk_2}{W_2}}$ = maximum number of g's without damping.

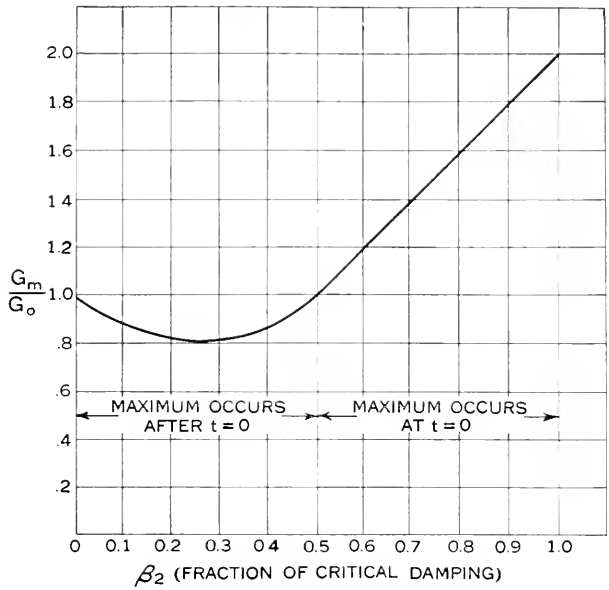


Fig. 2.5.3—Influence of velocity damping on maximum acceleration. See equation (2.5.5).

Then, from (2.5.5), at $t = 0$

$$\frac{G_m}{G_0} = 2\beta_2 \tag{2.5.7}$$

and, after $t = 0$,

$$\frac{G_m}{G_0} = e^{-\beta_2\omega_2 t_m}, \tag{2.5.8}$$

where t_m, the time at which the maximum occurs, is given by

$$\tan \omega_2 t_m \sqrt{1 - \beta_2^2} = \frac{(1 - 4\beta_2^2)\sqrt{1 - \beta_2^2}}{\beta_2(3 - 4\beta_2^2)}. \tag{2.5.9}$$

The largest value of G_m/G_0 from (2.5.7) and (2.5.8) is plotted against β_2 in Fig. 2.5.3. It is shown there that, as the damping is increased from zero, the maximum acceleration first decreases to a minimum of 80% of G_0 and then increases to G_0 at 50% of critical damping. In this interval the maximum acceleration occurs after $t = 0$. For damping greater than $\beta_2 = 0.5$ the maximum acceleration occurs at the instant of contact and increases in direct proportion to β_2.

2.6 Influence of Damping on Rebound

In considering rebound without damping, it was found that rebound does not occur unless the product of the maximum acceleration and the sus-

pended mass exceeds the total weight of the package. It was not necessary to distinguish between maximum acceleration on the first downstroke and first upstroke, since these are the same when there is no damping. With damping, however, the maximum acceleration on the first downstroke is

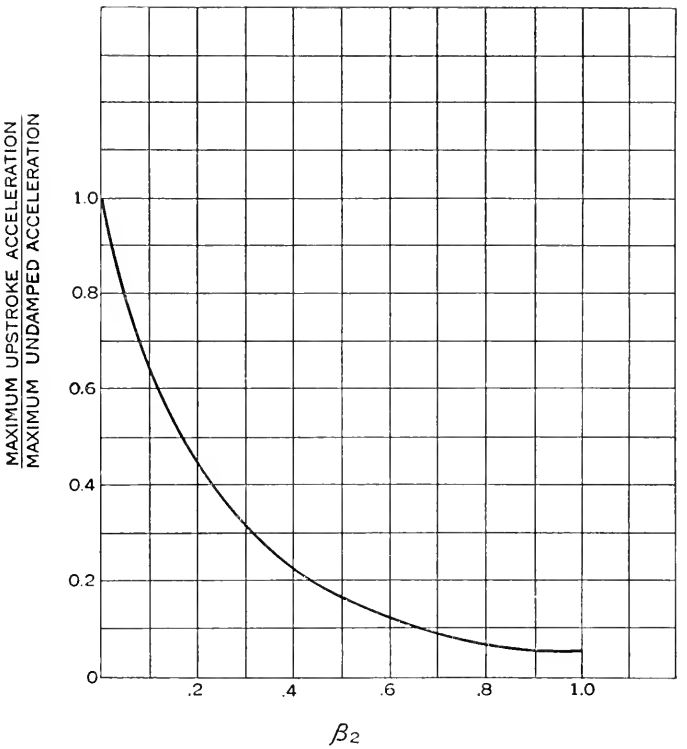


Fig. 2.6.1—Influence of velocity damping on maximum upstroke acceleration. See equation (2.5.5).

greater than that on the first upstroke (Fig. 2.5.2) and it is the latter that controls rebound. Hence damping inhibits rebound.

For example, with 50% of critical damping ($\beta_2 = 0.5$), equations (2.5.8) and (2.5.9) and Fig. 2.5.2 show that for the first downstroke $G_m/G_0 = 1$ while for the first upstroke $G_m/G_0 = 0.164$. Hence the tendency to rebound is reduced by a factor of six when damping to the extent of 50% of critical is added to an undamped package.

The ratio of the maximum acceleration on the first upstroke to the maximum undamped acceleration is plotted in Fig. 2.6.1 for various values of β_2 .

2.7 Influence of Dry Friction on Acceleration and Displacement

By "dry friction" is meant friction that is independent of velocity except for sign. During contact of the package with the floor the motion of m_2 might be opposed by a constant friction force F. Such a force is developed, for example, in a package with corrugated spring pad cushioning by rubbing against the side and end pads in a top or bottom drop. A typical idealized

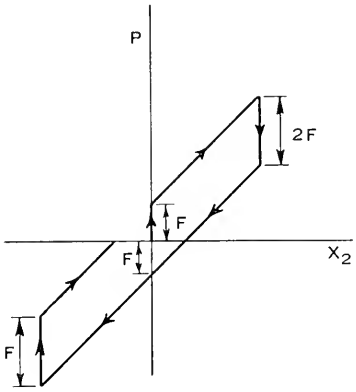


Fig. 2.7.1—Load vs. displacement for cushioning with dry friction.

load-displacement curve is shown in Fig. 2.7.1. For the first downstroke of m_2, the equation of motion of m_2 is

$$m_2\ddot{x}_2 + k_2x_2 = -F. \tag{2.7.1}$$

With initial conditions

$$[x_2]_{t=0} = 0, \qquad [\dot{x}_2]_{t=0} = \sqrt{2gh} \tag{2.7.2}$$

the solution of (2.7.1) is

$$x_2 = \sqrt{d_0^2 + \left(\frac{F}{k_2}\right)^2}\sin(\omega_2 t + \alpha) - \frac{F}{k_2} \tag{2.7.3}$$

where

$$d_0 = \sqrt{\frac{2W_2h}{k_2}},$$

$$\tan\alpha = \frac{F}{k_2 d_0} = \frac{F}{W_2 G_0},$$

$$G_0 = \sqrt{\frac{2hk_2}{W_2}}.$$

d_0 and G_0 are the maximum displacement and acceleration that would obtain if no friction were present. From (2.7.2) the maximum displacement with friction is

$$d_m = \sqrt{d_0^2 + \left(\frac{F}{k_2}\right)^2} - \frac{F}{k_2}. \tag{2.7.4}$$

Hence, the presence of friction decreases the maximum displacement since $d_m < d_0$.

From (2.7.3) the acceleration is

$$\frac{\ddot{x}_2}{g} = -\sqrt{G_0^2 + \left(\frac{F}{W_2}\right)^2} \sin(\omega_2 t + \alpha), \tag{2.7.5}$$

so that the maximum acceleration is

$$G_m = \sqrt{G_0^2 + \left(\frac{F}{W_2}\right)^2}, \tag{2.7.6}$$

which is greater than the maximum acceleration without friction.

It would appear, at first glance, that cushioning with friction always gives a greater acceleration than the corresponding cushioning without friction. However, the reverse is actually true provided we allow the same displacement in both cases. This may be done, as may be seen from (2.7.4), by decreasing the spring rate in the cushioning with friction to

$$k_F = k_2 - \frac{2F}{d_0}. \tag{2.7.7}$$

The maximum acceleration in the cushioning with friction is then, from (2.7.6),

$$G_F = G_0 - \frac{F}{W_2}, \qquad 0 \leqq \frac{F}{W_2} \leqq \frac{G_0}{2}. \tag{2.7.8}$$

That is, for the same maximum displacement, the maximum acceleration is reduced by the addition of dry friction.

2.8 Acceleration-Time Relation for Cubic Elasticity

As an example of the effect of nonlinearity of the cushioning on the shape of the acceleration-time function, the case of cubic elasticity (Class B) will be analyzed. The system to be considered is illustrated in Fig. 1.2.1, and the load-displacement relation for the cushioning is given by

$$P = k_0 x_2 + r x_2^3. \tag{2.8.1}$$

Substituting (2.8.1) in (1.2.13) and performing the indicated integration, we find

$$\dot{x}_2^2 = 2gh - \frac{k_0}{m_2} x_2^2 - \frac{r}{2m_2} x_2^4. \tag{2.8.2}$$

Remembering that $\dot{x}_2 = dx_2/dt$, we solve (2.8.2) for dt:

$$dt = \frac{dx_2}{\sqrt{2gh - \frac{k_0}{m_2} x_2^2 - \frac{r}{2m_2} x_2^4}}. \tag{2.8.3}$$

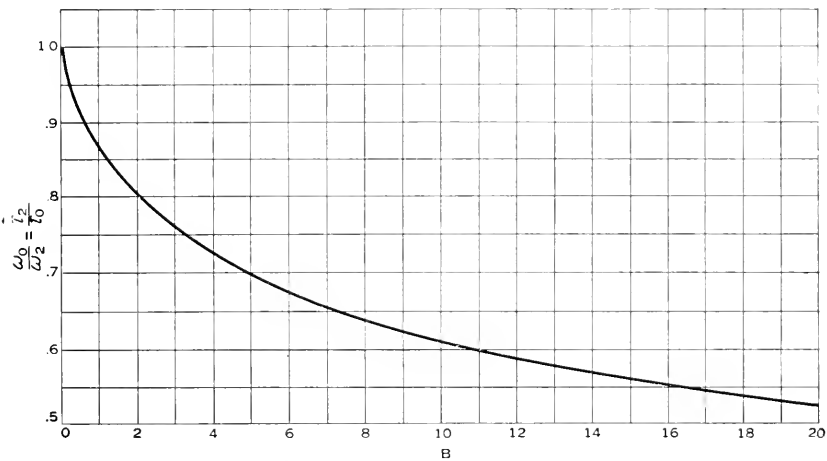


Fig. 2.8.1—Duration of acceleration pulse for cushioning with cubic elasticity. See equation (2.8.10).

Then, with the initial condition $x_2 = 0$ when $t = 0$, the integral of (2.8.3) is

$$t = \int_0^{x_2} \frac{dx_2}{\dot{x}_2} = \int_0^{x_2} \frac{dx_2}{\sqrt{2gh - \frac{k_0}{m_2} x_2^2 - \frac{r}{2m_2} x_2^4}}. \tag{2.8.4}$$

To integrate (2.8.4), let

$$Z = \frac{x_2}{\sqrt{k^2(x_2^2 - d_m)^2 + d_m^2}}, \tag{2.8.5}$$

where

$$k^2 = \tfrac{1}{2}\left\{1 - \frac{1}{\sqrt{1 + B}}\right\} \tag{2.8.6}$$

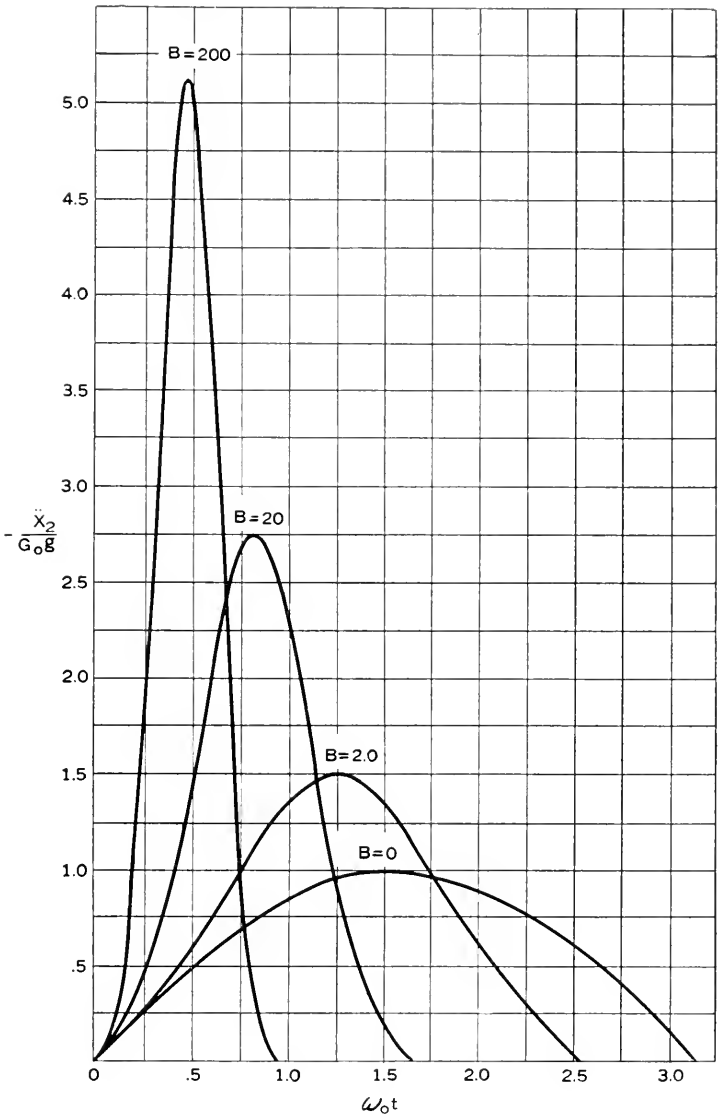


Fig. 2.8.2—Acceleration-time curves for cushioning with cubic elasticity. See equation (2.8.14).

and B and d_m are as given in Part I:

$$B = \frac{4W_2 hr}{k_0^2} \tag{1.5.3}$$

$$d_m = d_0 \sqrt{\frac{2}{B}(-1 + \sqrt{1 + B})}. \tag{1.5.4}$$

Then (2.8.4) becomes

$$t = \frac{1}{\omega_c} \int_0^Z \frac{dZ}{\sqrt{(1 - Z^2)(1 - k^2 Z^2)}}, \tag{2.8.7}$$

in which the integral is the elliptic integral of the first kind (see Hancock "Elliptic Integrals," John Wiley and Sons, New York, 1917). In (2.8.7),

$$\omega_c = \omega_0 (1 + B)^{1/4}, \tag{2.8.8}$$

where $\omega_0 = \sqrt{k_0/m_2}$ is the radian frequency that would obtain if the cushioning were linear with spring rate k_0. The motion for the linear case has a half period, or pulse duration $\tau_0 = \pi/\omega_0$. The half-period (τ_2) of the motion with cubic elasticity is twice the time required for x_2 to increase from 0 to Z_m.

From 2.8.5

$$\begin{aligned} [Z]_{x_2=0} &= 0, \\ [Z]_{x_2=d_m} &= 1. \end{aligned} \tag{2.8.9}$$

Hence, from (2.8.7), the half-period is

$$\tau_2 = \frac{2}{\omega_c} \int_0^1 \frac{dZ}{\sqrt{(1 - Z^2)(1 - k^2 Z^2)}} = \frac{2K}{\omega_c} \tag{2.8.10}$$

where K is the complete elliptic integral of the first kind. The duration of the acceleration pulse is therefore $2K/\omega_c$. We can define a radian frequency of the acceleration by

$$\omega_2 = \frac{\pi}{\tau_2} = \frac{\pi\omega_c}{2K} = \frac{\pi\omega_0(1 + B)^{1/4}}{2K}. \tag{2.8.11}$$

The ratio ω_0/ω_2 (i.e., τ_2/τ_0) is plotted in Fig. 2.8.1 which illustrates how the pulse duration decreases as the parameter B increases. Hence, for a given cushioning with cubic elasticity, the pulse duration decreases as the height of drop increases. This is in contrast to the linear case in which the duration is independent of the height of drop.

To find the acceleration $\ddot{x}_2$, we return to (2.8.7) and write it in the form of an elliptic function:

$$\operatorname{sn} \omega_c t = Z. \tag{2.8.12}$$

Substituting the expression for Z given in (2.8.5) and solving for x_2, we find

$$x_2 = d_m cn(\omega_c t - K). \tag{2.8.13}$$

Finally, differentiating (2.8.13) twice with respect to t, we find the acceleration to be

$$\ddot{x}_2 = \omega_c^2 d_m [2k^2 sn^2(\omega_c t - K) - 1] cn(\omega_c t - K). \tag{2.8.14}$$

The ratio $-\ddot{x}_2/G_0 g$ is plotted in Fig. 2.8.2 against a radian coordinate $(\omega_0 t)$ for several values of B. It may be seen that, as B increases, the maximum acceleration increases, the duration of the pulse decreases (see Fig. 2.8.1) and the acceleration-time curve becomes bell shaped. For reference, the sinusoid for the linear case ($B = 0$) is plotted in the figure.

Figure 2.8.2 is plotted for perfect rebound. If rebound does not occur, the curves continue, mirrored in the time axis, so as to form a periodic vibration of period $2\tau_2$.

2.9 Acceleration-Time Relation for Tangent Elasticity

In this section the effect of tangent elasticity on the shape of the acceleration-time relation will be studied. The shape of the load displacement curve is given by

$$P = \frac{2k_0 d_b}{\pi} \tan \frac{\pi x_2}{2d_b}. \tag{1.4.3}$$

The system considered is again that shown in Fig. 1.2.1. Referring to the energy equation (1.2.13):

$$\frac{m_2 \dot{x}_2^2}{2} + \int_0^{x_2} P dx_2 = m_2 gh, \tag{1.2.13}$$

we substitute the above value of P and perform the indicated integration to obtain, for the velocity,

$$\dot{x}_2 = \sqrt{2gh + \frac{8k_0 d_b^2}{m_2 \pi^2} \log \cos \frac{\pi x_2}{2d_b}}. \tag{2.9.1}$$

Then, as in Section 2.8,

$$t = \int_0^{x_2} \frac{dx_2}{\dot{x}_2} = \int_0^{x_2} \frac{dx_2}{\sqrt{2gh + \frac{8k_0 d_b^2}{m_2 \pi^2} \log \cos \frac{\pi x_2}{2d_b}}} \tag{2.9.2}$$

and the half-period (τ_2) of the motion is again twice the time required for x_2 to increase from 0 to d_m. Hence

$$\tau_2 = 2 \int_0^{d_m} \frac{dx_2}{\sqrt{2gh + \frac{8k_0 d_b^2}{m_2 \pi^2} \log \cos \frac{\pi x_2}{2d_b}}} \tag{2.9.3}$$

where, from Section 1.9,

$$d_m = \frac{2d_b}{\pi} \cos^{-1} \exp \left[-\frac{\pi^2}{8} \left(\frac{d_0}{d_b} \right)^2 \right]. \tag{1.9.5}$$

The radian frequency of the acceleration is

$$\omega_2 = \frac{\pi}{\tau_2}$$

and this is to be compared with the frequency

$$\omega_0 = \frac{\pi}{\tau_0} = \sqrt{\frac{k_0}{m_2}}$$

that would obtain if the cushioning were linear with spring rate k_0. The ratio ω_0/ω_2 (i.e., τ_2/τ_0) was obtained by numerical integration of (2.9.3)

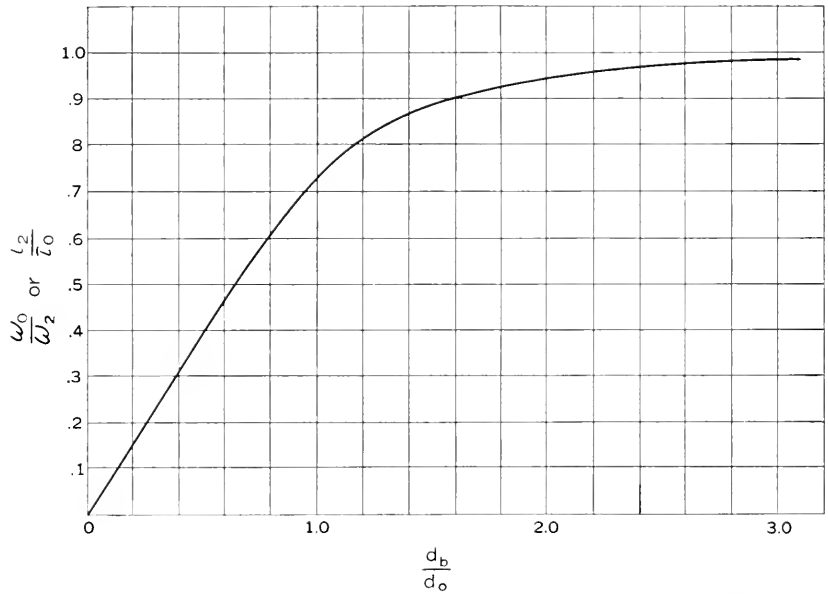


Fig. 2.9.1—Duration of acceleration pulse for cushioning with tangent elasticity. See equation (2.9.3).

and is plotted in Fig. 2.9.1 against the ratio d_b/d_0. The figure shows that for $d_b/d_0 < 1$, the pulse duration varies almost linearly with d_b/d_0. As the bottoming distance becomes larger than that required for linear cushioning with spring rate k_0, the pulse duration approaches asymptotically the duration π/ω_0 for the linear case.

As d_b/d_0 decreases, the pulse duration becomes shorter, but the maximum acceleration increases, in accordance with equation (1.9.4) and Fig. 1.9.1. The shapes of the acceleration-time curves for several values of d_b/d_0 are illustrated in Fig. 2.9.2. They are more sharply peaked than the corresponding curves for cubic elasticity (Fig. 2.8.2) as might be expected from

the fact that the load-displacement curve for tangent elasticity rises more rapidly than that for cubic elasticity; that is, the bottoming is harder.

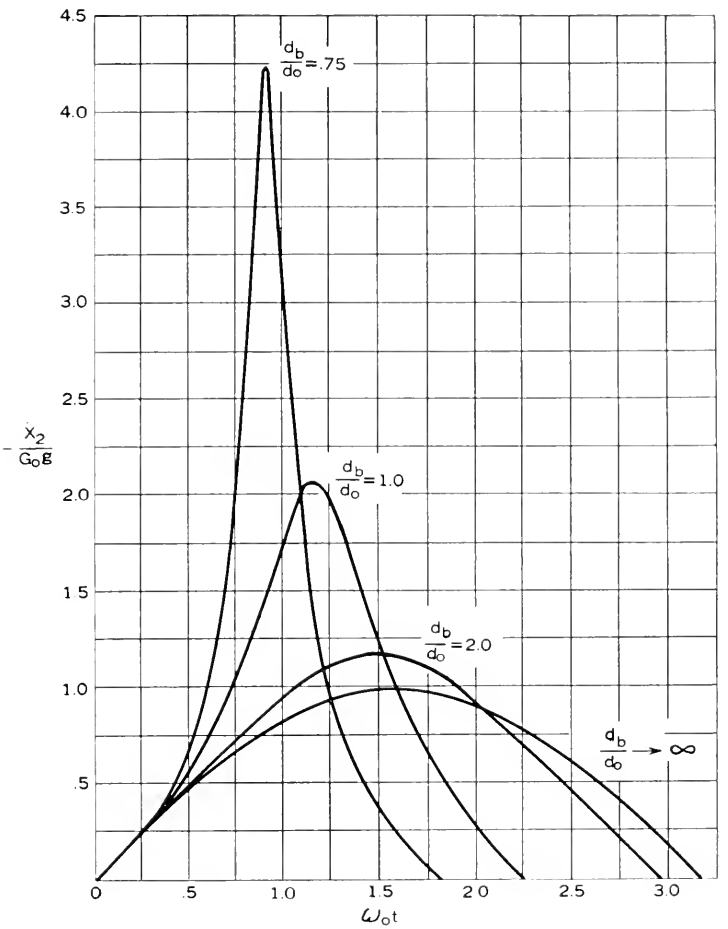


Fig. 2.9.2—Acceleration-time curves for cushioning with tangent elasticity.

The curves of Fig. 2.9.2 were obtained by numerical integration of equation (2.9.2), to obtain x_2 as a function of t, following which these values were substituted in the equation

$$m_2 \ddot{x}_2 + \frac{2k_0 d_b}{\pi} \tan \frac{\pi x_2}{2d_0} = 0$$

to obtain $\ddot{x}_2$. It may be observed that the maximum values of the curves are the values dictated by equation (1.9.4).

In performing the numerical integrations of equations (2.9.2) and (2.9.3), it is found that the integrand becomes infinite when $x_2 = d_m$ since at this point the velocity is zero. In order to avoid this difficulty, it was assumed[1] that, for a small distance in the neighborhood of d_m, the acceleration is constant with magnitude $G_m g$ as obtained from equation (1.9.4). The procedure is described in further detail in Section 2.12.

Figure 2.9.2 gives the acceleration-time curve for perfect rebound. If rebound does not occur, the acceleration is a periodic vibration, each successive half period having the shape shown, with alternating sign.

2.10 Acceleration-Time Relation for Abrupt Bottoming

By abrupt bottoming, we mean bilinear cushioning (Class D) as treated in Section 1.12. The load-displacement relation is (see equation (1.4.4) and Fig. 1.4.4)

$$\left.\begin{aligned} P &= k_0 x_2 \qquad 0 \leqq x_2 \leqq d_s \\ P &= k_b x_2 - (k_b - k_0) d_b \qquad x_2 \gtrless d_s \end{aligned}\right\}. \tag{1.4.4}$$

Considering, again, the system illustrated in Fig. 1.2.1, the equation of motion of m_2, before bottoming, is

$$m_2 \ddot{x}_2 + k_0 x_2 = 0 \qquad 0 \leqq x_2 \leqq d_s \tag{2.10.1}$$

with initial conditions

$$[x_2]_{t=0} = 0, \qquad [\dot{x}_2]_{t=0} = \sqrt{2gh}. \tag{2.10.2}$$

The solution of (2.10.1) is then

$$x_2 = \frac{\sqrt{2gh}}{\omega_0} \sin \omega_0 t, \qquad 0 \leqq x_2 \leqq d_s, \tag{2.10.3}$$

where

$$\omega_0 = \sqrt{\frac{k_0}{m_2}}. \tag{2.10.4}$$

The time (t_s) at which x_2 reaches d_s is found from (2.10.3):

$$t_s = \frac{1}{\omega_0} \sin^{-1} \frac{\omega_0 d_s}{\sqrt{2gh}} = \frac{1}{\omega_0} \sin^{-1} \frac{d_s}{d_0}, \tag{2.10.5}$$

where

$$d_0 = \sqrt{\frac{2W_2 h}{k_0}};$$

[1] See Timoshenko, "Vibration problems in Engineering," D. Van Nostrand Co., New York, Second Edition (1937) page 123.

i.e., d_0 is the displacement that would have been reached if the spring rate remained constant.

The velocity of m_2 at time t_s is

$$[\dot{x}_2]_{t=t_s} = \sqrt{2gh}\cos\omega_0 t_s = \sqrt{2gh\left(1 - \frac{d_s^2}{d_0^2}\right)}. \tag{2.10.6}$$

If $\sqrt{2gh} > \omega_0 d_s$, the displacement will exceed d_s and the equation of motion becomes

$$m_2\ddot{x}_2 + k_b x_2 - (k_b - k_0)d_s = 0, \qquad x_2 \geqq d_s. \tag{2.10.7}$$

The solution of (2.10.7), with initial conditions

$$\left.\begin{aligned}[x_2]_{t=t_s} &= d_s \\ [\dot{x}_2]_{t=t_s} &= \sqrt{2gh\left(1 - \frac{d_s^2}{d_0^2}\right)},\end{aligned}\right\} \tag{2.10.8}$$

is

$$\left.\begin{aligned}x_2 = \frac{k_0 d_0}{k_b}\sqrt{\frac{k_b}{k_0} + \frac{d_s^2}{d_0^2}\left(1 - \frac{k_b}{k_0}\right)}\sin(\omega_b t + \alpha - \omega_b t_s) \\ + \left(1 - \frac{k_0}{k_b}\right)d_s \qquad x_2 \geqq d_s\end{aligned}\right\} \tag{2.10.9}$$

where

$$\left.\begin{aligned}\tan^2\alpha &= \frac{k_0}{k_b\left(\frac{d_0^2}{d_s^2} - 1\right)} \\ \omega_b &= \sqrt{\frac{k_b}{m_2}}.\end{aligned}\right\} \tag{2.10.10}$$

By differentiating (2.10.3) and (2.10.9) twice with respect to t, the accelerations for the two regions are found to be

$$\ddot{x}_2 = -G_0 g\sin\omega_0 t, \qquad 0 \leqq x_2 \leqq d_s, \tag{2.10.11}$$

$$\ddot{x}_2 = -G_0 g\sqrt{\frac{k_b}{k_0} + \frac{d_s^2}{d_0^2}\left(1 - \frac{k_b}{k_0}\right)}\sin(\omega_b t + \alpha - \omega_b t_s), \tag{2.10.12}$$

$$x_2 \gtrless d_s,$$

where

$$G_0 = \sqrt{\frac{2hk_0}{W_2}}. \tag{2.10.13}$$

Typical shapes of the acceleration pulse represented by equations (2.10.11) and (2.10.12) are shown in Fig. 2.10.1. The curves are drawn for $d_s/d_0 =$

0.5 and for several values of k_b/k_0. The peak values of the curves are the same as given by equation (1.12.3). The curve marked $k_b/k_0 = 1$ is the sinusoid of the linear case with duration

$$\tau_0 = \frac{\pi}{\omega_0}. \tag{2.10.14}$$

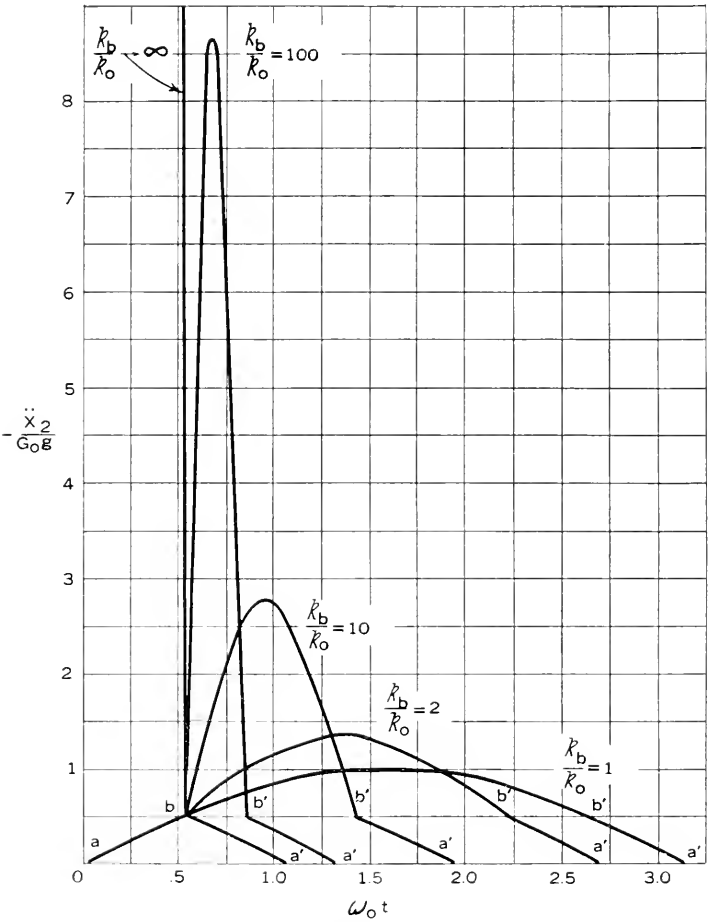
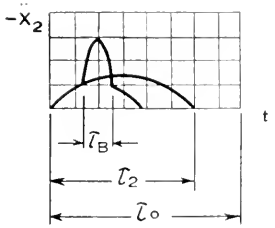


Fig. 2.10.1—Acceleration-time curves for cushioning with bi-linear elasticity. $d_s/d_0 = 0.5$. See equations (2.10.11) and (2.10.12).

As before, if the package does not rebound, the acceleration shown is mirrored in the time axis after each half cycle, to form a vibration of period $2\tau_2$.

It is useful to know the duration of the complete pulse (aa' in Fig. 2.10.1) and also the duration of bottoming (bb' in Fig. 2.10.1). Calling the former τ_2 and the latter τ_B, we have, from equations (2.10.11) and (2.10.12)



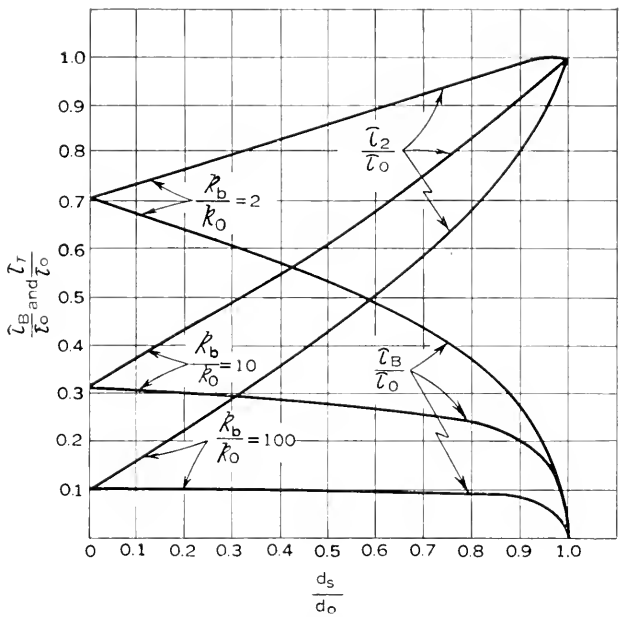


Fig. 2.10.2—Pulse durations for cushioning with bi-linear elasticity. See equations (2.10.15) and (2.10.16).

$$\frac{\tau_2}{\tau_0} = \frac{2}{\pi}\sin^{-1}\frac{d_s}{d_0} + \sqrt{\frac{k_0}{k_b}}\left[1 - \frac{2}{\pi}\tan^{-1}\sqrt{\frac{k_0}{k_b\left(\frac{d_0^2}{d_s^2} - 1\right)}}\right] \tag{2.10.15}$$

$$\frac{\tau_B}{\tau_0} = \sqrt{\frac{k_0}{k_b}}\left[1 - \frac{2}{\pi}\tan^{-1}\sqrt{\frac{k_0}{k_b\left(\frac{d_0^2}{d_s^2} - 1\right)}}\right]. \tag{2.10.16}$$

These two equations are plotted in Fig. 2.10.2 for several values of k_b/k_0.

2.11 Acceleration-Time Relation for Hyperbolic Tangent Elasticity

The relation between acceleration and time for hyperbolic tangent elasticity is found by the same procedure that was used for tangent elasticity

in Section 2.9. The system considered is that shown in Fig. 1.2.1 and the load displacement curve of the cushioning is given by

$$P = P_0 \tanh \frac{k_0 x_2}{P_0}.$$

Substituting the above expression for P in the energy equation (1.2.13), we find the velocity to be

$$\dot{x}_2 = \sqrt{2gh - \frac{2P_0^2}{m_2 k_0} \log \cosh \frac{k_0 x_2}{P_0}}. \tag{2.11.1}$$

Then, as before,

$$t = \int_0^{x_2} \frac{dx_2}{\dot{x}_2} \tag{2.11.2}$$

and the half period (τ_2) of the motion is twice the time required for x_2 to increase from 0 to d_m, or

$$\tau_2 = 2 \int_0^{d_m} \frac{dx_2}{\dot{x}_2}, \tag{2.11.3}$$

where, from Section 1.13,

$$d_m = \frac{d_0 P_0}{W_2 G_0} \cosh^{-1} \exp \left(\frac{W_2^2 G_0^2}{2P_0^2} \right). \tag{1.13.3}$$

The radian frequency of the acceleration is defined as

$$\omega_2 = \frac{\pi}{\tau_2}$$

and this is to be compared with the frequency

$$\omega_0 = \frac{\pi}{\tau_0} = \sqrt{\frac{k_0}{m_2}}$$

that would obtain if the cushioning had a constant spring rate equal to the initial spring rate (k_0) of the hyperbolic tangent cushioning. The ratio ω_0/ω_2 (or τ_2/τ_0) is plotted, in Fig. 2.11.1, against the dimensionless parameter P_0/W_2G_0 (see Section 1.13). It may be observed that the pulse duration becomes very long when P_0/W_2G_0 is small, i.e., when the horizontal portion of the load displacement curve (Fig. 1.4.5) comes into play. The influence on the shape of the acceleration-time curve is illustrated in Fig. 2.11.2. The curve marked $P_0/W_2G_0 \rightarrow \infty$ is the sinusoid for the linear case. For small values of P_0/W_2G_0 the curve approaches a square wave.

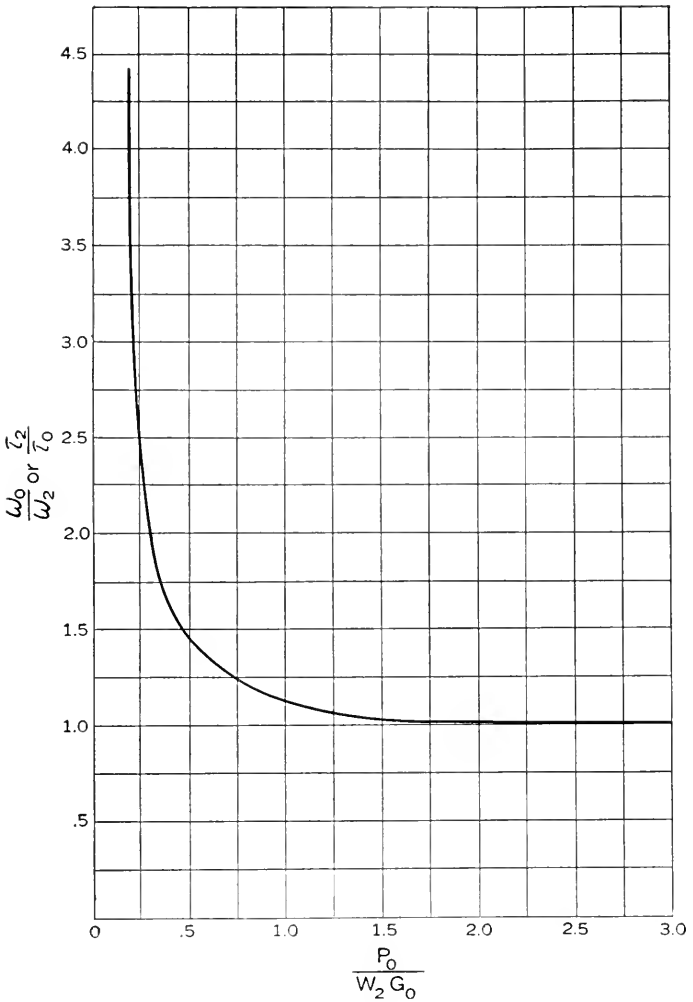


Fig. 2.11.1—Duration of acceleration pulse for cushioning with hyperbolic tangent elasticity.

2.12 Numerical Procedure for Finding Acceleration-Time Relation for Class F Cushioning

When the load-displacement curve does not resemble one of Classes A to E, the acceleration-time relation may be found by numerical integration. Combining the energy equation,

$$\frac{m_2 \dot{x}_2^2}{2} + \int_0^{x_2} P \, dx_2 = m_2 g h, \tag{1.2.13}$$

with the equation relating time and velocity,

$$t = \int_0^{x_2} \frac{dx_2}{\dot{x}_2}, \tag{2.12.1}$$

we find

$$t = \sqrt{\frac{W_2}{2g}} \int_0^{x_2} \frac{dx_2}{\sqrt{W_2 h - \int_0^{x_2} P\, dx_2}}. \tag{2.12.2}$$

As an example, consider the problem of a 15-pound article supported on cushioning with the load-displacement curve shown in Fig. 2.12.1. The package is to be dropped from a height of 3 feet. The computations are

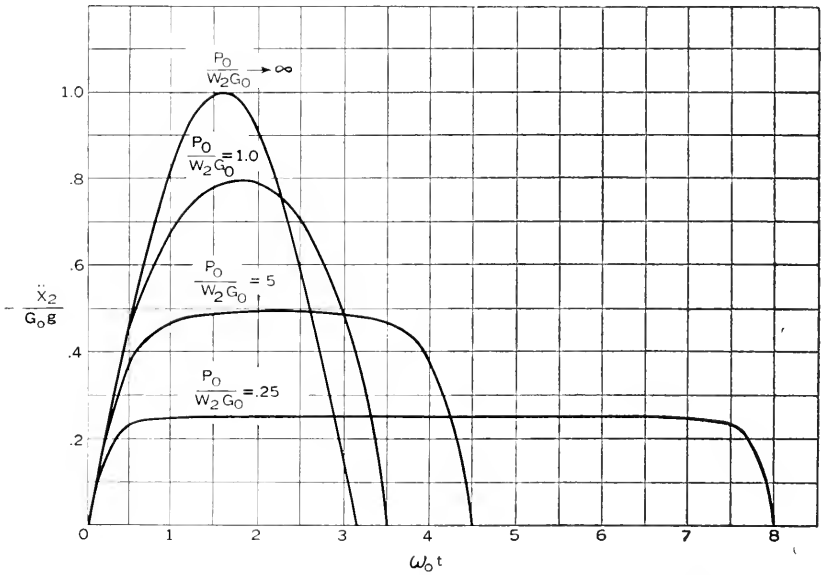


Fig. 2.11.2—Acceleration-time curves for cushioning with hyperbolic tangent elasticity.

given in detail in Tables III and IV. The headings of Columns (1) to (8) of Table III are the same as in Table II, Section 1.15. A_n is the integral under the radical of equation (2.12.2). Column (10) of Table III is the integrand of Equation (2.12.2), i.e., it is proportional to the reciprocal of the velocity expressed as a function of displacement. The function is plotted in Fig. 2.12.2 and its integration is performed in Table IV. In columns (11), (12) and (13), intervals of x_2 are chosen to suit the shape of the curve. The values for column (14) are taken from column (10). Columns (15) and (16) perform the same operations on the integrand $(W_2 h - A_n)^{-\frac{1}{2}}$ that are performed in Columns (5) and (6) of Table III on the integrand P.

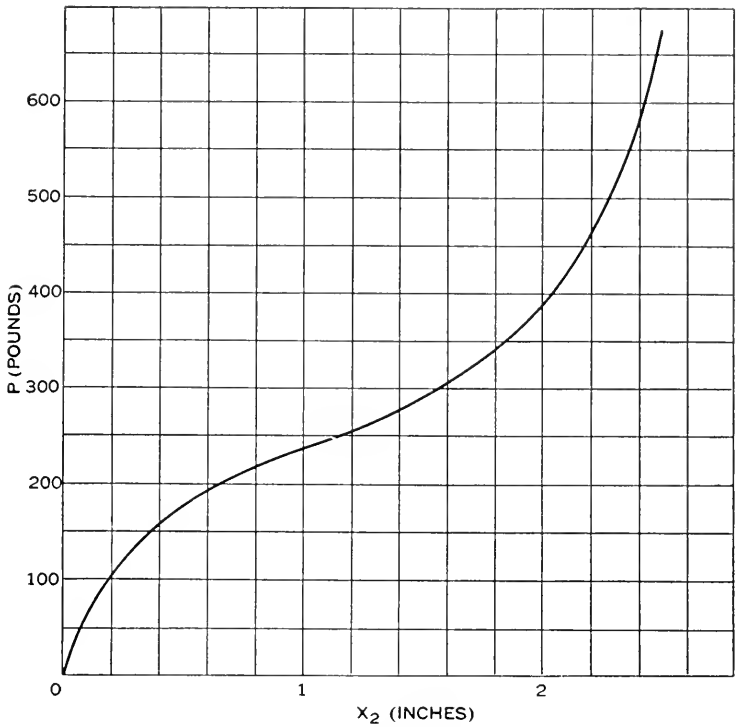


Fig. 2.12.1—A load displacement curve for Class F cushioning.

TABLE III

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
n	$\Delta(x_2)_n$	$(x_2)_n$	P_n	$\frac{\Delta(x_2)_n}{2}(P_n + P_{n-1})$	$A_n = \int_0^{(x_2)_n} P dx_2$	$h_n = \frac{A_n}{W_2}$	$G_n = \frac{P_n}{W_2}$	$W_2 h - A_n$	$\frac{1}{\sqrt{W_2 h - A_n}}$
0	0	0.0	0	0	0	0	0	540	0.0431
1	.20	0.2	105	10.5	10.5	0.7	7.0	529.5	0.0435
2	.20	0.4	155	26.0	36.5	2.4	10.3	503.5	0.0446
3	.20	0.6	192	34.7	71.2	4.8	12.8	468.8	0.0462
4	.20	0.8	217	40.9	112.1	7.5	14.5	427.9	0.0483
5	.20	1.0	237	45.4	157.5	10.5	15.8	382.5	0.0511
6	.20	1.2	257	49.4	206.9	13.9	17.1	333.1	0.0547
7	.20	1.4	277	52.9	259.8	17.3	18.5	280.2	0.0597
8	.20	1.6	305	58.2	318.0	21.2	20.3	222.0	0.0671
9	.20	1.8	342	64.7	382.7	25.5	22.8	157.3	0.0798
10	.20	2.0	392	73.4	456.1	30.4	26.1	83.9	0.109
11	.05	2.05	405	19.9	476.0	31.8	27.0	64.0	0.125
12	.05	2.10	422	20.7	496.7	33.2	28.1	43.3	0.152
13	.05	2.15	440	21.6	518.3	34.6	29.4	21.7	0.215
14	.01	2.16	445	4.42	522.7	34.8	29.7	17.3	0.240
15	.01	2.17	450	4.48	527.2	35.2	30.0	12.8	0.279
16	.01	2.18	455	4.52	531.7	35.5	30.3	8.3	0.347
17	.01	2.19	457	4.56	536.3	35.8	30.5	3.7	0.521
18	.01	2.20	462	4.60	540.9	36.1	30.8	0	∞

TABLE IV

(11)	(12)	(13)	(14)	(15)	(16)	(17)
n	$\Delta(x_2)_n$	$(x_2)_n$	$f_n = \frac{1}{\sqrt{W_2 h - A_n}}$	$\frac{\Delta(x_2)_n}{2}(f_n + f_{n-1})$	$D_n = \int_0^{(x_2)_n} \frac{dx_2}{\sqrt{W_2 h - A_n}}$	$t = D_n \times \sqrt{\frac{W_2}{2g}}$
0	0	0	0.0431	0	0	0
1	0.4	0.4	0.0446	0.0175	0.0175	0.0024
2	0.4	0.8	0.0483	0.0185	0.0360	0.0050
3	0.4	1.2	0.0547	0.0206	0.0566	0.0079
4	0.4	1.6	0.0671	0.0243	0.0809	0.0112
5	0.2	1.8	0.0798	0.0149	0.0958	0.0133
6	0.2	2.0	0.109	0.0189	0.1147	0.0160
7	0.1	2.1	0.152	0.0131	0.1278	0.0178
8	0.05	2.15	0.215	0.0092	0.1370	0.0190
9	0.03	2.18	0.347	0.0083	0.1453	0.0202
10	0.01	2.19	0.521	0.0043	0.1496	0.0208
11	0.01	2.20	∞			0.0221

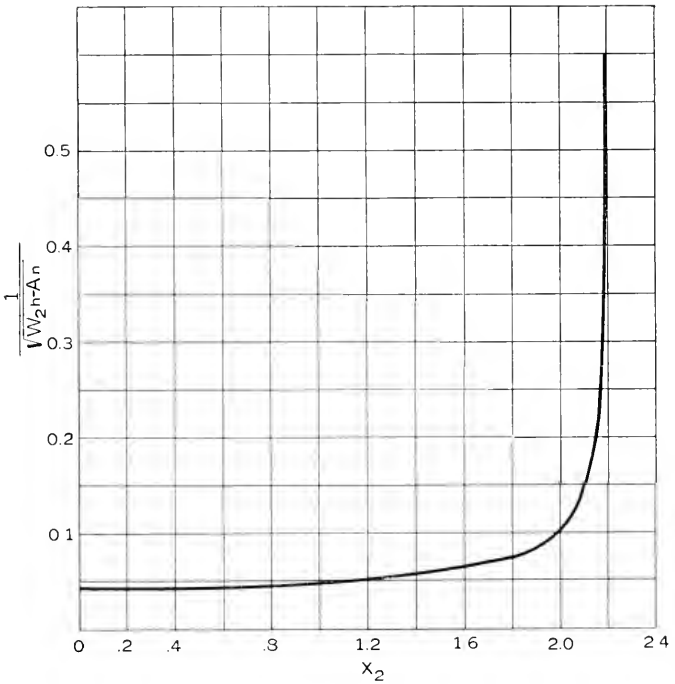


Fig. 2.12.2—Plot of Column (3) vs. Column (10) of Table III.

A difficulty arises because the integrand $(W_2 h - A_n)^{-\frac{1}{2}}$ becomes infinite for the maximum displacement (see Column (14)). This is avoided by assuming that the acceleration is constant in the last interval[2] and has the

[2] Timoshenko, "Vibration Problems in Engineering," D. Van Nostrand Co., New York, Second Edition (1937) page 123.

value given in Column (8), Table III, for the maximum height of drop. Then,

$$\Delta(x_2)_n = \tfrac{1}{2}G_m g \Delta t^2 \tag{2.12.3}$$

or

$$\Delta t = \sqrt{\frac{2\Delta(x_2)_n}{G_m g}}. \tag{2.12.4}$$

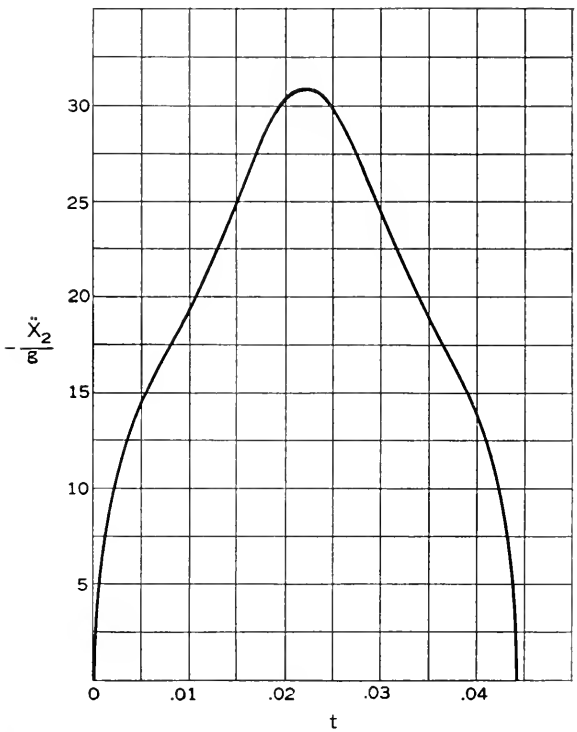


Fig. 2.12.3—Acceleration-time curve (for the cushioning shown in Fig. 2.12.1) obtained by numerical integration.

In the present instance,

$$(\Delta x_2)_n = 0.01 \text{ inches}$$

$$G_m g = 30.8 \times 386 = 11900 \text{ in/sec.}^2$$

Hence, from (2.12.4), $\Delta t = 0.0013$ sec. and the last entry in Column (17) is obtained by adding this value of Δt to the preceding entry.

The final curve of acceleration vs. time is obtained by plotting the entries of Column (17) against the entries of Column (8), Table III, for corresponding values of x_2. The result is shown in Fig. 2.12.3.

PART III

AMPLIFICATION FACTOR

3.1 Introduction

If the maximum acceleration, of the packaged article as a whole, is reached very slowly, the severity of the disturbance experienced by a structural element of the packaged article is very nearly proportional to the maximum acceleration. Roughly speaking, "very slowly" means that the time, during which the acceleration undergoes a major change in magnitude, is long in comparison with the natural period of vibration of the element under consideration. When this is so, no transient vibration is excited in the element. The displacement response of an element under very slowly varying conditions is called the "static response". Under more rapidly varying conditions the dynamic response to the same maximum acceleration may be greater or less than the static response. The ratio (A) of the maximum dynamic response to the static response is called the amplification factor. In general, for a given acceleration disturbance, very low-frequency elements have amplification factors less than unity, while the amplification factors are greater than unity for elements whose natural frequencies are near or above the disturbing frequencies. The numerical value of the amplification factor depends not only on the manner in which the disturbing acceleration varies with time, but also on the "reference acceleration", i.e., the value of acceleration for which the static response is calculated. Usually the reference acceleration chosen for calculating the static response is the maximum value (G_m) of the disturbing acceleration. However, when special circumstances are being investigated, such as the effect of damping or abrupt bottoming, the reference acceleration is taken to be G_0, which is the acceleration that would be reached if the damping or bottoming were absent. In such cases the amplification factor includes both the effect of rate of change of acceleration and the effect of the special conditions.

When the reference acceleration is G_m the amplification factor will be denoted by A_m and when the reference acceleration is G_0 the amplification factor will be denoted by A_0. The symbol G_e will be used to designate the slowly applied acceleration that would produce the same maximum displacement as the transient acceleration, i.e., $G_e = A_m G_m$ or $G_e = A_0 G_0$. The symbol G_s will be used to denote the safe value of G_e, for an element of the packaged article, as determined by a strength test or by calculation. In specifying G_s some judgement is required to take into account the effects of plastic deformation in comparing tests made on greatly different time scales. Good judgement is also necessary in deciding whether or not the

assumptions listed in Section 0.2 are valid in each application. The general procedure for using amplification factors is as follows. We first find the value of the reference acceleration (in units of *number of times gravity*) from Part I. From Part II we find the properties of the acceleration-time relation which give us the information required for entering one of the curves of Part III and finding the amplification factor. Then, the product of the reference acceleration and the amplification factor ($A_m G_m$ or $A_0 G_0$) is a number (G_e) by which the weight of the structure is to be multiplied when calculating its deflection or stress by the usual static methods of elementary strength of materials. Alternatively, G_e must be found not to exceed G_s.

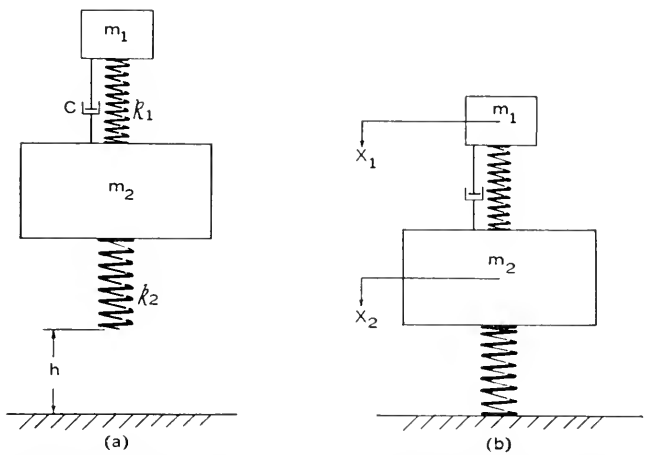


Fig. 3.2.1—Idealized system used in calculating amplification factors for linear undamped cushioning with perfect rebound. (a) initial position, (b) first contact with floor.

In the following sections the amplification factors for typical transient accelerations encountered in package drop tests are calculated. The amplification factor curves that are plotted are entirely analogous to the familiar "resonance curves" for steady sinusoidal vibration, except that in this case the disturbing forces are transients of various shapes. It will be seen from the curves that the maximum acceleration, as calculated by the methods of Part I or as measured by an accelerometer, is not always a true measure of the severity of the disturbance.

3.2 Amplification Factors for a Half-Sine-Wave Pulse Acceleration

The first case to be treated is the response of an element of the packaged item to the transient acceleration that would occur in a package with linear

undamped cushioning and perfect rebound. Figure 3.2.1 illustrates the idealized system, and it may be noted that the mass m_3 is omitted, as is required for perfect rebound (Section 2.3). At first we shall consider that the mass m_1 is undamped and later we shall consider the effect of damping in this element.

The mass m_1 is taken to be small in comparison with m_2, so that the motion of the latter is the same as we found it to be in Section 2.2 where m_1 was not considered. Hence the acceleration of m_2 is a half-sine wave pulse:

$$\ddot{x}_2 = -\omega_2 \sqrt{2gh} \sin \omega_2 t, \qquad (0 \leqq t \leqq \pi/\omega_2). \tag{3.2.1}$$

The equation of motion of m_1 is

$$m_1\ddot{x}_1 + k_1(x_1 - x_2) = 0. \tag{3.2.2}$$

Let x be the relative displacement of m_1 with respect to m_2, i.e.,

$$x = x_1 - x_2 . \tag{3.2.3}$$

x is proportional to the force in the spring (k_1x) and to the acceleration of m_1 and hence is proportional to the deflection, strain and stress in the element which the system m_1, k_1 represents.

Substituting (3.2.3) in (3.2.2), we find:

$$m_1\ddot{x} + k_1x = -m_1\ddot{x}_2 . \tag{3.2.4}$$

This equation holds for the duration π/ω_2 of the pulse $\ddot{x}_2$. The initial conditions for x are

$$[x]_{t=0} = [\dot{x}]_{t=0} = 0 \tag{3.2.5}$$

so that the solution of (3.2.4) is

$$x = \frac{\omega_2 \sqrt{2gh}}{\omega_2^2 - \omega_1^2}\left[\frac{\omega_2}{\omega_1} \sin \omega_1 t - \sin \omega_2 t\right], \qquad \left(0 \leqq t \leqq \frac{\pi}{\omega_2}\right). \tag{3.2.6}$$

It may be seen that x is composed of a forced displacement at the acceleration frequency ω_2, on which is superposed a free vibration at the natural frequency, ω_1, of the element. The maximum value of the relative displacement is

$$x_{\max} = \frac{\sqrt{2gh}}{\omega_1\left(\frac{\omega_1}{\omega_2} - 1\right)} \sin \frac{2n\pi}{\frac{\omega_1}{\omega_2} + 1}, \qquad \left(0 \leqq t \leqq \frac{\pi}{\omega_2}\right). \tag{3.2.7}$$

in which n is a positive integer chosen so as to make the sine term as large as possible while the argument remains less than π.

(3.2.7) gives the maximum dynamic response of the element m_1 during the interval of impact. To find the amplification factor we must compare

$x_{\max}$ with the "static response" i.e. with the value (x_{st}) that x would have if the acceleration $\ddot{x}_2$ reached the same maximum value ($\omega_2\sqrt{2gh}$) in a very long time. The resulting value may be found from (3.2.4) by omitting the transient term $m_1\ddot{x}$. Then

$$x_{st} = \omega_2 \sqrt{2gh}\,\frac{m_1}{k_1}$$

or

$$x_{st} = \frac{\omega_2}{\omega_1^2}\sqrt{2gh}\,. \tag{3.2.8}$$

The amplification factor for the interval $0 \leqq t \leqq \pi/\omega_2$ is then

$$A_m = \frac{x_{max}}{x_{st}} = \frac{\dfrac{\omega_1}{\omega_2}}{\dfrac{\omega_1}{\omega_2} - 1} \sin \frac{2n\pi}{\dfrac{\omega_1}{\omega_2} + 1}, \qquad (0 \leqq t \leqq \pi/\omega_2). \tag{3.2.9}$$

It should be observed that A_m depends only on the frequency ratio ω_1/ω_2. That is, since $\omega_1/\omega_2 = \tau_2/\tau_1$, the amplification factor depends only on the ratio of the pulse duration to the half period of vibration of the element.

Thus far we have studied only the motion in the interval $0 \leqq t \leqq \pi/\omega_2$. We must not, however, overlook the possibility of larger displacements of m_1 with respect to m_2 occurring after rebound. In fact, examination of (3.2.6) reveals that x has no maximum in the interval $0 \leqq t \leqq \pi/\omega_2$ when $\omega_1 < \omega_2$. It is very likely, then, that larger values will occur at later times.

After rebound, m_1 executes free vibrations with respect to m_2. We have to compare the magnitude of $x_{\max}$, in the interval $0 \leqq t \leqq \pi/\omega_2$, with the amplitude of the free vibration. Calling the relative displacement during free vibration x' and measuring a time coordinate t' from the instant the package leaves the floor, we have

$$m_1\ddot{x}' + k_1x' = 0, \tag{3.2.10}$$

with initial conditions

$$\begin{aligned} [x']_{t'=0} &= [x]_{t=\pi/\omega_2}, \\ [\dot{x}']_{t'=0} &= [\dot{x}]_{t=\pi/\omega_2}. \end{aligned} \tag{3.2.11}$$

The solution of (3.2.10) with initial conditions (3.2.11) is

$$x' = \frac{\omega_2^2\sqrt{4gh\left(1 + \cos\dfrac{\omega_1}{\omega_2}\pi\right)}}{\omega_1(\omega_2^2 - \omega_1^2)} \sin\left(\omega_1 t' + \frac{\omega_1\pi}{2\omega_2}\right), \tag{3.2.12}$$

$$t \leqq \frac{\pi}{\omega_2}.$$

Then

$$A_m = \frac{x'_{max}}{x_{st}} = \frac{2\frac{\omega_1}{\omega_2}\cos\frac{\pi\omega_1}{2\omega_2}}{1 - \frac{\omega_1^2}{\omega_2^2}}, \qquad \left(t \geqq \frac{\pi}{\omega_2}\right). \tag{3.2.13}$$

We find, on comparing (3.2.13) with (3.2.9) that for $\omega_1 < \omega_2$ equation (3.2.13) gives the larger value of A_m, while for $\omega_1 > \omega_2$ equation (3.2.9)

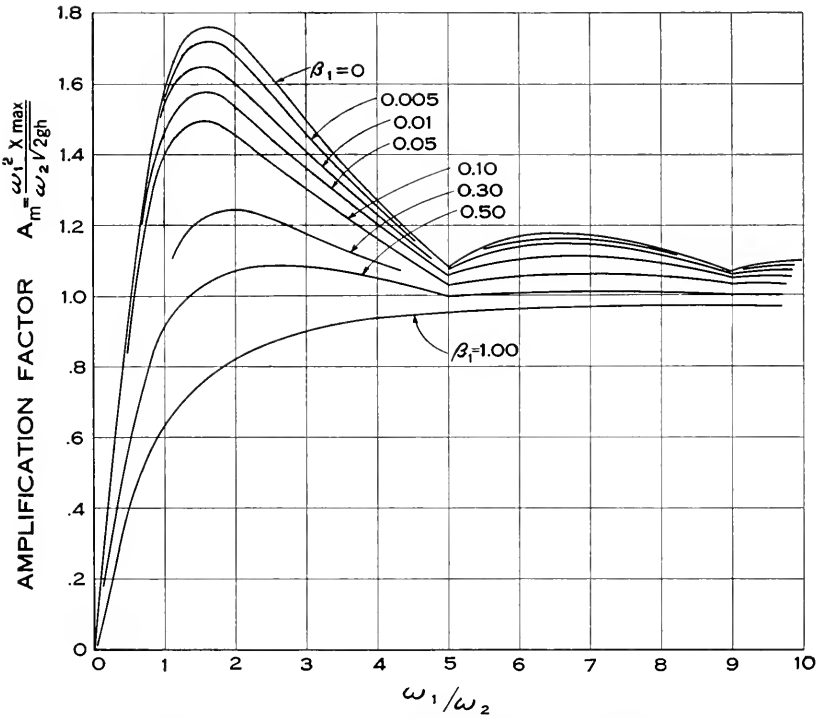


Fig. 3.2.2—Amplification factors for linear undamped cushioning with perfect rebound. See Fig. 3.2.1 and equations (3.2.9) and (3.2.13).

gives the larger value of A_m. That is, when the duration of impact is shorter than the half-period of vibration of the element, the maximum displacement (and stress) in the element occurs after the impact is over.

The curve marked $\beta_1 = 0$ in Fig. 3.2.2 is a plot of the largest value of A_m from (3.2.9) and (3.2.13) with the frequency ratio ω_1/ω_2 as abscissa. (3.2.13) was used for $\omega_1/\omega_2 \leqq 1$ and (3.2.9) for $\omega_1/\omega_2 \geqq 1$. The maximum value of A_m is 1.76 and occurs at $\omega_1/\omega_2 = 1.6$. Hence, at this frequency ratio, the deformation of the element is 1.76 times as great as would be expected from a calculation using the maximum value of acceleration alone as in Part I.

On the other hand, for frequency ratios $\omega_1/\omega_2 < 0.5$ the severity of the shock can be very much less than might be expected from the calculations of Part I. For very small values of ω_1/ω_2 the amplification factor may be seen from (3.2.13) to be equal to $2\omega_1/\omega_2$. For large values of ω_1/ω_2 (stiff elements) Fig. 3.2.2 shows that the amplification factor is very nearly unity and the methods of Part I can be used without additional calculation.

When damping of the element of the packaged article is considered, the amplification factors are less than without damping. The applicable equations of motion during and after impact are obtained by inserting velocity damping terms in (3.2.4) and (3.2.10):

$$m_1\ddot{x} + c_1\dot{x} + k_1x = -m_1\ddot{x}_2\,, \qquad 0 \leqq t \leqq \frac{\pi}{\omega_2} \tag{3.2.14}$$

$$m_1\ddot{x}' + c_1\dot{x}' + k_1x' = 0, \qquad t \geqq \frac{\pi}{\omega_2}. \tag{3.2.15}$$

If we express the damping of the element m_1 as the fraction of critical damping

$$\beta_1 = \frac{c_1}{2\sqrt{m_1k_1}}, \tag{3.2.16}$$

(as in Section 2.5) equations (3.2.14) and (3.2.15) become

$$\ddot{x} + 2\beta_1\omega_1\dot{x} + \omega_1^2x = -\ddot{x}_2\,, \qquad 0 \leqq t \leqq \frac{\pi}{\omega_2}, \tag{3.2.17}$$

$$\ddot{x}' + 2\beta_1\omega_1\dot{x}' + \omega_1^2x' = 0, \qquad t \geqq \frac{\pi}{\omega_2}. \tag{3.2.18}$$

The amplification factors for equations (3.2.17) and (3.2.18), with boundary conditions (3.2.5) and (3.2.11), respectively, were obtained on the Westinghouse Mechanical Transients Analyzer[3] for $\beta_1 = 0.005, 0.01, 0.05, 0.10, 0.30, 0.50$ and 1.00. The curves are shown in Fig. 3.2.2.

3.3 Application of Half-Sine-Wave Amplification Factors

As an example of the use of the amplification factor curves of Fig. 3.2.2, let us consider the following problem:

[3] Arrangements for performing these calculations were made through the courtesy of Mr. A. C. Monteith, Manager of Industry Engineering, and Mr. C. F. Wagner, Manager of Central Station Engineering, Westinghouse Electric and Manufacturing Co. Dr. G. D. McCann, Transmission Engineer, was in immediate charge of the project. For a description of the analyzer see "A New Device for the Solution of Transient-Vibration Problems by the Method of Electrical-Mechanical Analogy" by H. E. Criner, G. D. McCann and C. E. Warren, *Journal of Applied Mechanics*, Vol. 12, No. 3 (1945) pp. A-135 to A-141.

It is required to judge the suitability of a proposed package for a large vacuum tube weighing 10 pounds. Strength tests have been made on the tube in a shock testing machine which produces a half-sine-wave acceleration pulse of 25 milliseconds duration. The weakest element of the tube is found to be the cathode structure, for which the safe maximum acceleration in the drop testing machine is 200g. The cathode structure has a natural vibration frequency of 120 cycles per second and has 1% of critical damping. The proposed package has essentially linear, undamped cushioning with a spring rate of 3300 pounds per inch and an available displacement of $\frac{3}{4}$ inch. The outer container weighs much less than the tube so that the package may be expected to rebound. Is the cushioning suitable for protecting the cathode in a drop of 5 feet?

First find the maximum G that the tube will experience in a 5 ft. drop of the package (equation 1.3.3):

$$G_m = \sqrt{\frac{2hk_2}{W_2}} = \sqrt{\frac{2 \times 60 \times 3300}{10}} = 199.$$

The accompanying maximum displacement is, from equation (1.3.4),

$$d_m = \frac{2h}{G_m} = \frac{2 \times 60}{199} = 0.6 \text{ in.}$$

The available displacement ($\frac{3}{4}$ inches) is therefore sufficient and the maximum acceleration (199g) is slightly less than the safe maximum acceleration (200g) found with the shock testing machine. However, before the cushioning is approved it is necessary to investigate the frequency effects. The duration of acceleration in both the shock machine and in the package must be considered.

The amplification factor for the element tested in the shock machine is found as follows. First find the frequency corresponding to the 25 millisecond pulse:

$$f_2 = \frac{1}{2 \times .025} = 20 \text{ c.p.s.}$$

The ratio of the element frequency to the shock machine frequency is

$$\frac{f_1}{f_2} = \frac{\omega_1}{\omega_2} = \frac{120}{20} = 6.$$

Entering Fig. 3.2.2 with $\omega_1/\omega_2 = 6$, we read, from the curve $\beta_1 = 0.01$, $A_m = 1.14$. The 200g test in the shock machine is, therefore, equivalent to a slowly applied acceleration of $G_s = 200 \times 1.14 = 228g$.

To find the corresponding quantity for the package drop, first find the cushion frequency:

$$f_2 = \frac{1}{2\pi}\sqrt{\frac{k_2}{m_2}} = \frac{1}{2\pi}\sqrt{\frac{3300 \times 386}{10}} = 57 \text{ c.p.s.}$$

The ratio of the element frequency to the package frequency is therefore

$$\frac{f_1}{f_2} = \frac{\omega_1}{\omega_2} = \frac{120}{57} = 2.1.$$

Entering Fig. 3.2.2 with $\omega_1/\omega_2 = 2.1$ we read, from the curve $\beta_1 = 0.01$, $A_m = 1.59$. The $199g$ acceleration pulse in the package drop is therefore equivalent to a slowly applied acceleration of $G_e = 199 \times 1.59 = 316g$. This is almost 40% in excess of the value ($228g$) found to be safe from the shock machine data. The cushioning is therefore judged to be inadequate. The procedure for finding the correct spring rate for the cushioning is as follows. It is known that we must have

$$G_e \leqq G_s.$$

Therefore, take

$$G_e = A_m G_m = 228.$$

Now

$$G_m = \sqrt{\frac{2hk_2}{W_2}} = \omega_2\sqrt{\frac{2h}{g}}.$$

Therefore

$$A_m\omega_2 = 409 \text{ rad/sec.}$$

Also

$$\omega_1 = 2\pi \times 120 = 754 \text{ rad/sec.}$$

Then, with successive trial values of ω_2, we calculate ω_1/ω_2, enter Fig. 3.2.2, read the corresponding value of A_m from curve $\beta_1 = 0.01$ and test to see if the product $A_m\omega_2 = 409$. The combination which satisfies the test is found to be

$$\omega_2 = 280 \text{ rad/sec.}$$

$$\omega_1/\omega_2 = 2.69$$

$$A_m = 1.47.$$

Then

$$k_2 = \omega_2^2 m_2 = \frac{(280)^2 \times 10}{386} = 2030 \text{ lbs./in.}$$

$$G_m = \sqrt{\frac{2hk_2}{W_2}} = 155$$

$$d_m = \frac{2h}{G_m} = .77 \text{ in.}$$

Hence the spring rate of the cushioning should be reduced from 3300 lbs./in. to 2030 lbs./in. and the available space should be increased to accomodate the 0.77 inch maximum displacement before bottoming.

3.4 Special Treatment of Strong, Low Frequency Elements

The product of the amplification factor (A_m) and the maximum acceleration (G_m) must be equal to or less than the maximum allowable slowly applied acceleration (G_s):

$$G_e = A_m G_m \leqq G_s . \tag{3.4.1}$$

For frequency ratios

$$\frac{\omega_1}{\omega_2} < \frac{1}{2}, \tag{3.4.2}$$

Figure 3.2.2 shows that, approximately,

$$A_m = 2\frac{\omega_1}{\omega_2} \tag{3.4.3}$$

for $\beta_1 \leqq 0.10$. When this is so, we may combine (3.4.1) and (3.4.3) to obtain the criterion

$$2\frac{\omega_1}{\omega_2} G_m \leqq G_s . \tag{3.4.4}$$

Now,

$$G_m = \sqrt{\frac{2hk_2}{W_2}} = \omega_2 \sqrt{\frac{2h}{g}} . \tag{3.4.5}$$

Hence the criterion (3.4.4) may be written as

$$2\omega_1 \sqrt{\frac{2h}{g}} \leqq G_s \tag{3.4.6}$$

or

$$f_1 \leqq \frac{1.1\, G_s}{\sqrt{h}} \tag{3.4.7}$$

where h is in inches and f_1 is in cycles per second.

It may be observed that (3.4.7) *is independent of the properties of the cushioning.* Hence, as long as (3.4.2) is satisfied, any cushioning at all may be used for an element that satisfies (3.4.7) regardless of the magnitude of the maximum acceleration G_m. In particular, rigid mounting is suitable for such an element. The only precaution to be observed is that the maximum acceleration and duration must not be unfavorable for other elements of the packaged article.

Example: A 9-pound vacuum tube has an anode structure for which the safe maximum acceleration is $200g$ as determined in a centrifuge test. The natural vibration frequency of the anode is 35 cycles per second and the damping is 1% of critical. What cushioning around the tube is required to protect the anode from damage in a package drop of 3 feet?

Calculate

$$\frac{1.1G_s}{\sqrt{h}} = \frac{1.1 \times 200}{\sqrt{36}} = 36.7 \text{ c.p.s.}$$

This is greater than $f_1 = 35$ c.p.s and hence any cushioning is safe for the anode. The results of calculations for cushioning with spring rates of 50, 500, 5000, 5×10^5 and 5×10^7 pounds per inch are given in the following table:

k_2(lbs./in.)	G_m	$\frac{\omega_1}{\omega_2}$	A_m	$A_m G_m$
50	20	4.74	1.12	22
500	63	1.47	1.65	106
5×10^3	200	.474	0.9	180
5×10^5	2,000	.0474	0.09	180
5×10^7	20,000	.0047	0.009	180

In each case the product of $A_m G_m$ is less than the allowable 200 and, as long as the combination of G_m and the amplification factors for other elements does not exceed the allowable $A_m G_m$ for those elements, the cushioning is suitable. The precaution to observe is that higher-frequency elements shall not have amplification factors such that $A_m G_m$ may be excessive for them.

3.5 Amplification Factors for Damped Sinusoidal Acceleration

If the outer container of the package is heavy enough (see Section 2.3) there will be no rebound and the packaged item will vibrate in the cushioning after impact. For linear cushioning with velocity damping, the acceleration produced by the vibration will be a damped sinusoid (equation (2.5.5) and Fig. 2.5.2). The system to be considered is shown in Fig. 3.5.1. To determine the effect of the damped vibration of m_2 on the mass m_1,

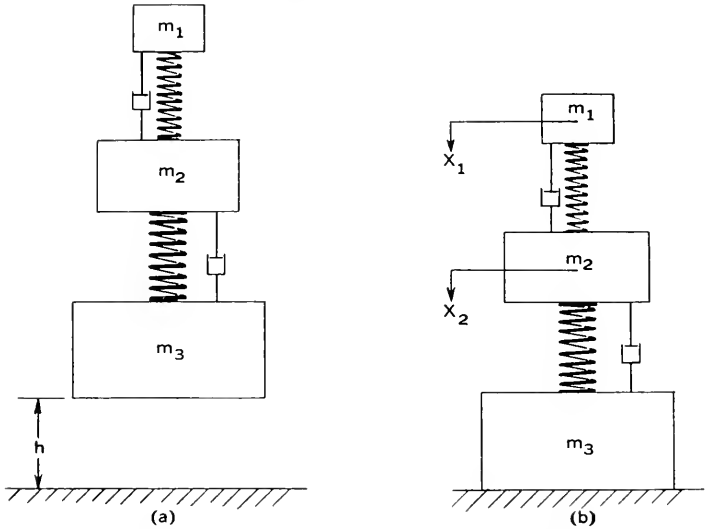


Fig. 3.5.1—Idealized system for linear damped cushioning with no rebound.
(a) initial position, (b) first contact with floor.

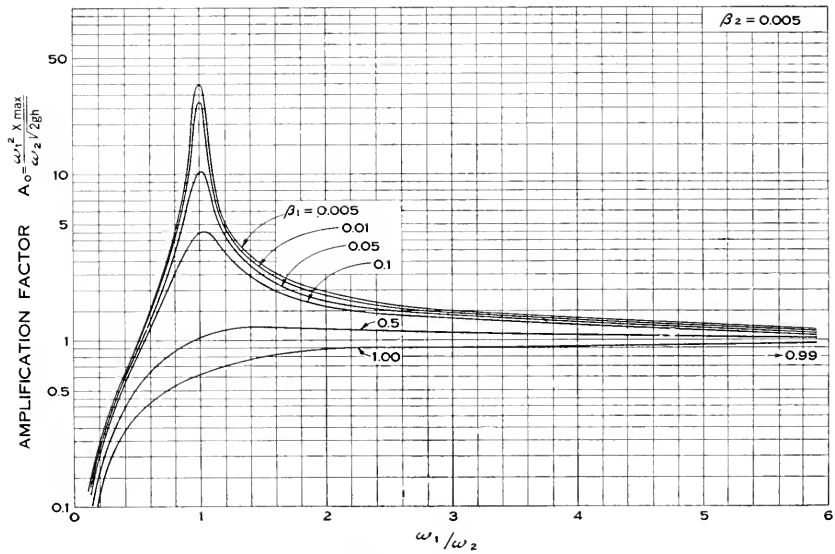


Fig. 3.5.2—Amplification factors for linear damped cushioning with no rebound.
$\beta_2 = 0.005$. See equations (3.5.1) and (3.5.2).

we note that the equation of motion and initial conditions are identical with (3.2.17) except that for the acceleration $\ddot{x}_2$ we use the damped sinusoid, equation (2.5.5), instead of the half-sine pulse (3.2.1). The solution of (3.2.17), i.e. the relative displacement $(x_1 - x_2)$ of m_1 with respect to m_2, is

$$x = \frac{\omega_2^2 \sqrt{2gh}}{2\omega_2' \omega_1'} \{e^{-\beta_1\omega_1 t}[A \sin(\omega_1' t + \gamma - \delta) + B \sin(\omega_1' t - \gamma - \zeta)] - e^{-\beta_2\omega_2 t}[A \sin(\omega_2' t + \gamma - \delta) - B \sin(\omega_2' t + \gamma + \zeta)]\} \qquad (3.5.1.)$$

where

$$\omega_1 = \sqrt{\frac{k_1}{m_1}}$$

$$\omega_2 = \sqrt{\frac{k_2}{m_2}}$$

$$\omega_1' = \omega_1\sqrt{1 - \beta_1^2}$$

$$\omega_2' = \omega_2\sqrt{1 - \beta_2^2}$$

$$1/A = \sqrt{(\beta_2\omega_2 - \beta_1\omega_1)^2 + (\omega_1' - \omega_2')^2}$$

$$1/B = \sqrt{(\beta_2\omega_2 - \beta_1\omega_1)^2 + (\omega_1' + \omega_2')^2}$$

$$\tan\gamma = \frac{2\beta_2^2 - 1}{2\beta_2\sqrt{1 - \beta_2^2}}$$

$$\tan\delta = \frac{\omega_1' - \omega_2'}{\beta_2\omega_2 - \beta_1\omega_1}$$

$$\tan\zeta = \frac{\omega_1' + \omega_2'}{\beta_2\omega_2 - \beta_1\omega_1}.$$

The relative displacement of m_1 with respect to m_2 is seen to consist of a forced, damped vibration (ω_2, β_2) on which is superposed the free damped oscillations (ω_1, β_1) of m_1.

The amplification factor

$$A_0 = \frac{x_{\max}}{x_{st}} = \frac{\omega_1^2 x_{\max}}{\omega_2\sqrt{2gh}} \qquad (3.5.2)$$

is plotted in Figs. 3.5.2 to 3.5.7 for six values of β_1 and six values of β_2. These curves were obtained by direct solution of the differential equation on the Westinghouse Mechanical Transients Analyzer.[4] The amplification factor in this case includes the effect of damping; i.e., the reference acceleration is not the maximum acceleration of m_2, but is the maximum acceleration that m_2 would reach if the damping β_2 were zero. Consequently, the amplification factors for large values of ω_1/ω_2 do not approach

[4] See footnote, Section 3.2. Only enough data were obtained with the analyzer to find the general shapes of the curves, so that the fine structure is not revealed. Checks on the analyzer results were made by computing A_0 from equations (3.5.1) and (3.5.2) for $\omega_1/\omega_2 = 1$, $\beta_1 = \beta_2$; $\omega_1/\omega_2 = 0$; $\omega_1/\omega_2 \to \infty$.

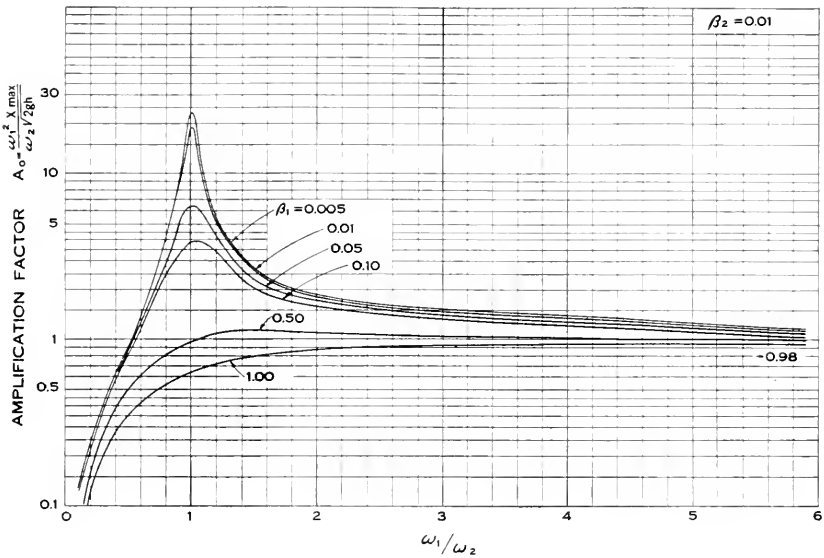


Fig. 3.5.3—Amplification factors for linear damped cushioning with no rebound. $\beta_2 = 0.01$. See equations (3.5.1) and (3.5.2).

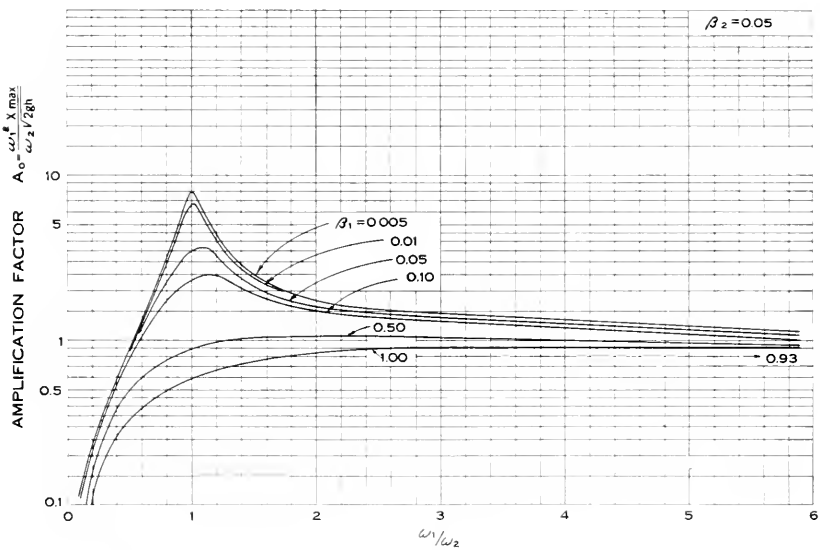


Fig. 3.5.4—Amplification factors for linear damped cushioning with no rebound. $\beta_2 = 0.05$. See equations (3.5.1) and (3.5.2).

unity. For example, the curve for $\beta_1 = 0.005, \beta_2 = 1$ (Fig. 3.5.7) approaches a value of nearly four as $\omega_1/\omega_2 \to \infty$. The factor four is composed of two

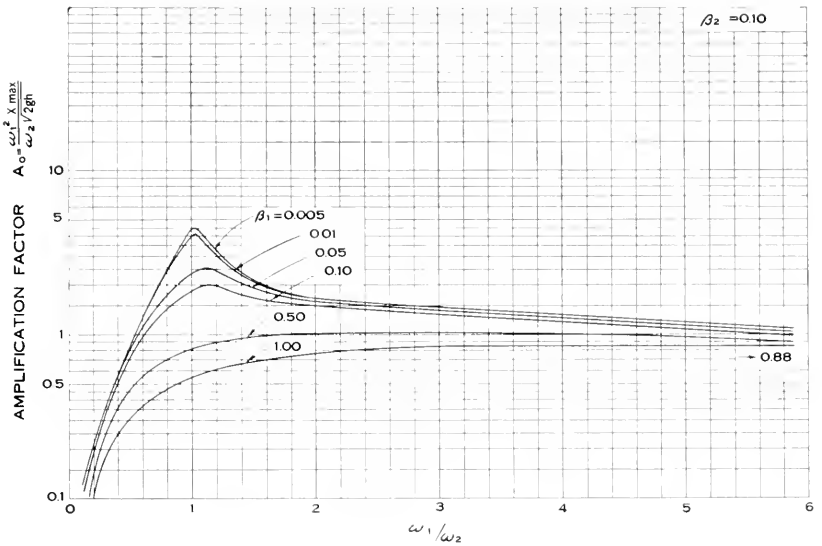


Fig. 3.5.5—Amplification factors for linear damped cushioning with no rebound. $\beta_2 = 0.1$. See equations (3.5.1) and (3.5.2).

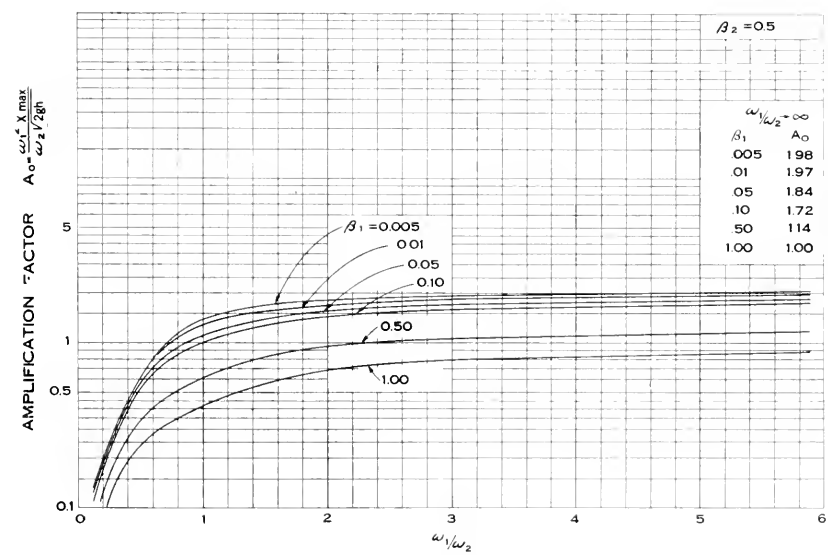


Fig. 3.5.6—Amplification factors for linear damped cushioning with no rebound. $\beta_2 = 0.5$. See equations (3.5.1) and (3.5.2).

factors of two. The first arises from the fact that the maximum value of acceleration, for $\beta_2 = 1$, is twice the value that would be reached if β_2

were equal to zero (see Fig. 2.5.3). The second factor (of nearly two) is due to the fact that the maximum acceleration is reached at time $t = 0$ when $\beta_2 = 1$ (see Fig. 2.5.2) and the response of an almost undamped system ($\beta_1 = 0.005$) to a suddenly applied and subsequently maintained acceleration is double the response to a slowly applied acceleration (see curve (a) Fig. 3.8.1). For $\beta_1 > 0$ and $\beta_2 < 1$ the amplification factor is less than four, as $\omega_1/\omega_2 \to \infty$, in accordance with the curves plotted in Fig. 3.5.8.

Example: A 1.5-pound vacuum tube is to be packed in a container whose estimated weight will be at least 50 pounds. The cathode structure of the tube has a natural frequency of 25 c.p.s. with damping 0.5% of critical

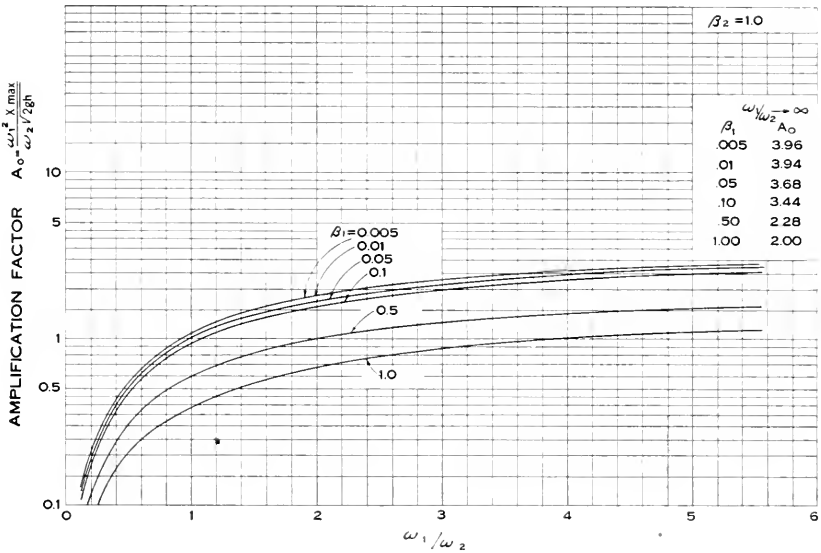


Fig. 3.5.7—Amplification factors for linear damped cushioning with no rebound. $\beta_2 = 1.0$. See equations (3.5.1) and (3.5.2).

and its safe acceleration, as determined in a centrifuge, is 90g. What spring rate of cushioning is suitable for protecting the cathode in a drop of five feet? It is specified that the cushioning shall have damping 50% of critical.

Assuming linear cushioning, the spring rate that would be prescribed, by considering maximum acceleration alone, is

$$k_2 = \frac{W_2 G_m^2}{2h} = \frac{1.5 \times (90)^2}{2 \times 60} = 101 \text{ lbs./in.}$$

Considering damping, Fig. 2.5.3 shows that 50% of critical damping does

not change G_m. To find the amplification factor we must first decide if the package will rebound. With 50% of critical damping, the maximum ac-

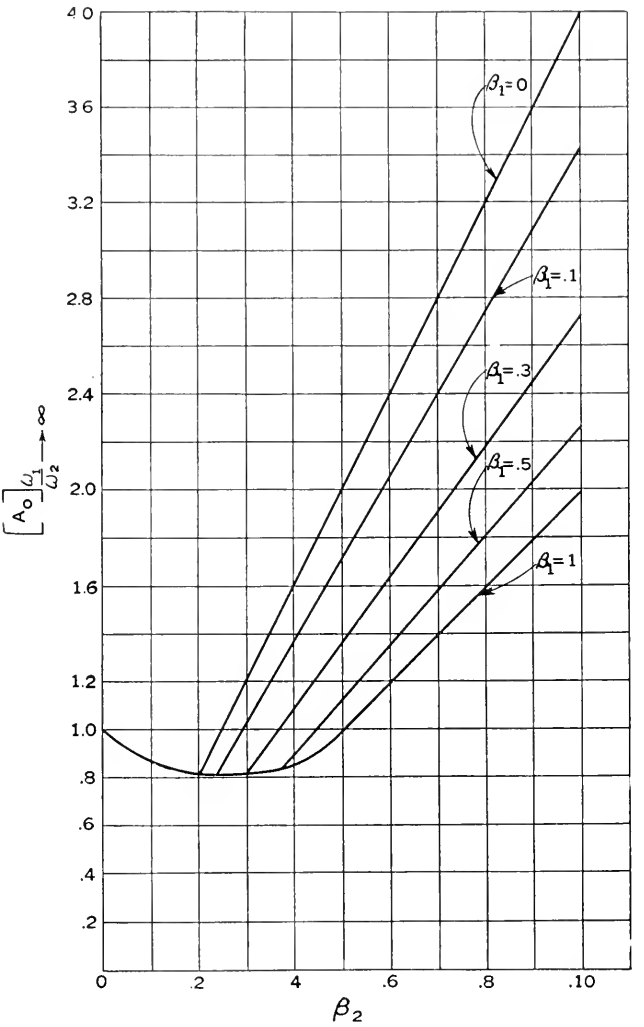


Fig. 3.5.8—Limiting values of amplification factors for linear damped cushioning with no rebound. $\omega_1/\omega_2 \to \infty$. See equations (3.5.1) and (3.5.2).

celeration on the first upstroke is 0.164 G_m (see Section 2.6 and Fig. 2.6.1). Then, $0.164 \times 90 \times 1.5 = 22$ lbs. which is less than the estimated weight of the outer container. The package will not rebound and Fig. 3.5.6 should be used for the amplification factor.

The frequency of vibration of the tube in its cushion will be

$$\omega_2 = \sqrt{\frac{k_2}{m_2}} = \sqrt{\frac{101 \times 386}{1.5}} = 159 \text{ rad./sec.}$$

Hence $\omega_1/\omega_2 = 2\pi \times 25/159 = 0.99$ and, from Fig. 3.5.6, $A_0 = 1.4$. Hence $G_e = 90 \times 1.4 = 126$, which is greater than the allowable $G_s = 90$, so that the 101 lb./in. cushion is unsatisfactory.

To obtain satisfactory cushioning, set

$$A_0 G_0 = 90,$$

that is

$$A_0 \omega_2 = \frac{90}{\sqrt{\frac{2h}{g}}} = 159 \text{ rad./sec.}$$

Noting that $\omega_1 = 2\pi \times 25 = 157$, we find from Fig. 3.5.6 that there are two values of ω_2 (90 and 600 rad/sec.) that satisfy the criterion $A_0\omega_2 = 159$ rad/sec. The first gives

$$\omega_2 = 90 \text{ rad/sec.}$$
$$A_0 = 1.8$$
$$k_2 = 31.5 \text{ lbs./in.}$$
$$G_0 = 50$$
$$G_e = 90$$
$$d_m = 2.4 \text{ in.}$$

The second gives

$$\omega_2 = 600 \text{ rad/sec.}$$
$$A_0 = 0.27$$
$$k_2 = 1400 \text{ lbs./in.}$$
$$G_0 = 335$$
$$G_e = 90$$
$$d_m = 0.36 \text{ in.}$$

The second solution requires less space for cushioning than the first but should be used only if the remainder of the tube can endure the high acceleration of 335g. Otherwise the 50g package should be used.

3.6 Amplification Factors for the Pulse Acceleration of Cubic Cushioning

In a rebounding package with undamped Class B cushioning, the packaged article (m_2) will undergo a pulse acceleration of duration π/ω_2 as given by equation (2.8.11). The shape of the pulse is illustrated in Fig. 2.8.2 and its functional form is

$$\ddot{x}_2 = \frac{4K^2\omega_2^2 d_m}{\pi^2}\left[2k^2 sn^2\left(\frac{2K\omega_2 t}{\pi} - K\right) - 1\right] cn\left(\frac{2K\omega_2 t}{\pi} - K\right). \quad (2.8.14)$$

To determine the influence of the shape and duration of this pulse on the amplification factor, we proceed as before by substituting (2.8.14) in the

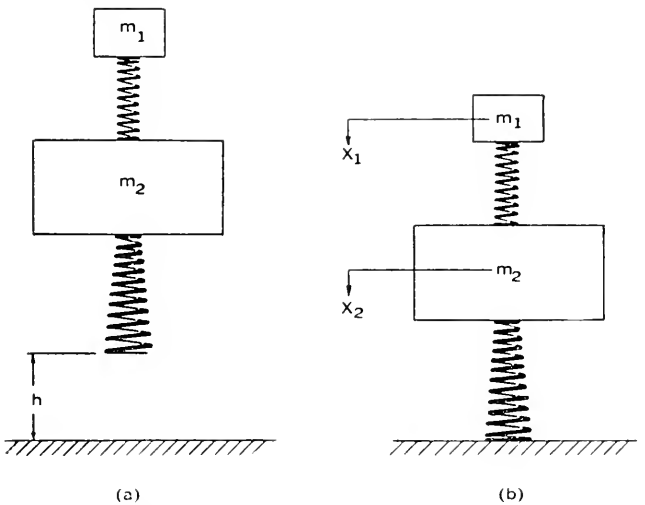


Fig. 3.6.1—Idealized system used in calculating amplification factors for non-linear, undamped cushioning with perfect rebound.

differential equation governing the relative displacement ($x = x_1 - x_2$) between m_1 and m_2 (see Fig. 3.6.1):

$$\ddot{x} + \omega_1^2 x = -\ddot{x}_2 . \quad (3.6.1)$$

With boundary conditions $x(0) = \dot{x}(0) = 0$, the solution of (3.6.1) may be written as

$$x = \frac{1}{\omega_1}\int_0^t \ddot{x}_2(\lambda) \sin \omega_1 (\lambda - t)\, d\lambda \quad (3.6.2)$$

and the maximum value of x may be expressed by

$$x_{\max} = \frac{1}{\omega_1}\int_0^{t_m} \ddot{x}_2(\lambda) \sin \omega_1(\lambda - t_m)\, d\lambda, \quad (3.6.3)$$

where t_m is the time at which the largest value of x occurs.

The amplification factor, in this case, will be taken as the ratio of $x_{\max}$ to the relative displacement (x_{st}) resulting from a slow application of the maximum value of $\ddot{x}_2$. From (3.6.1),

$$x_{st} = \frac{|\ddot{x}_2|_{\max}}{\omega_1^2} = \frac{G_m g}{\omega_1^2}, \tag{3.6.4}$$

where G_m is given by equation (1.5.6). Then

$$A_m = \frac{x_{\max}}{x_{st}} = \frac{x_{\max}\omega_1^2}{G_m g} \tag{3.6.5}$$

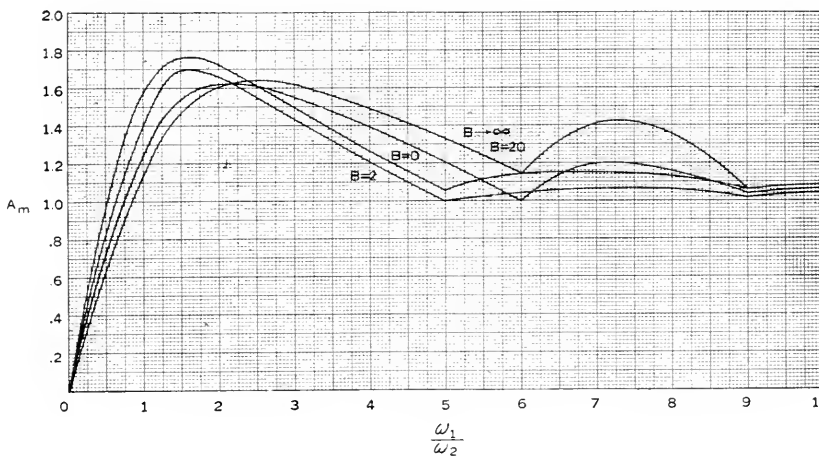


Fig. 3.6.2—Amplification factors for undamped cushioning with cubic elasticity. Perfect rebound. See equation (3.6.6).

or

$$A_m = \frac{\omega_1}{G_m g}\int_0^{t_m} \ddot{x}_2(\lambda)\sin\omega_1(\lambda - t_m)\,d\lambda. \tag{3.6.6}$$

A_m was evaluated, mostly by graphical methods, for four values of B (0, 2, 20 and ∞) and the results are plotted in Fig. 3.6.2. Observing that $B = 0$ corresponds to linear cushioning, it may be noted that cubic non-linearity in the cushioning does not change the amplification factor by more than 35% even in the most extreme case ($B \to \infty$). The severity of the shock, however, may be much greater for the cubic cushioning than for linear cushioning with a spring rate equal to the initial spring rate (k_0) of the cubic cushioning. This is because A_m is multiplied by G_m to obtain G_e and, for large values of B, G_m may be much larger than the maximum acceleration for the linear case. In other words, in comparing Class B

with Class A cushioning the difference in maximum acceleration, rather than the difference in amplification factors, is usually more important.

Example: Consider the example given in Section 1.6 and let it be required to determine the effect of pulse duration on a cathode structure with a 200 c.p.s. natural frequency of vibration. In Section 1.6 we found that

$$B = 5.4 \qquad k_0 = 255$$

$$G_0 = 28.6 \qquad r = 108.$$

$$G_m = 55$$

With $B = 5.4$, enter Fig. 2.8.2 and find

$$\frac{\omega_0}{\omega_2} = 0.88.$$

Now

$$\omega_0 = \sqrt{\frac{k_0 g}{W_2}} = \sqrt{\frac{255 \times 386}{22.5}} = 66.1 \text{ rad./sec.}$$

Hence

$$\omega_2 = \frac{66.1}{0.88} = 75 \text{ rad./sec.}$$

Then, with $\omega_1/\omega_2 = 2\pi \times 200/75 = 16.7$, enter Fig. 3.6.2 and find $A_m =$ approximately 1.0. Hence G_e is about the same as G_m and the conclusions reached for this problem in Section 1.6 are not altered.

3.7 Amplification Factors for Abrupt Bottoming

The amplification factors for bilinear elasticity have not been computed in complete detail. They can be obtained approximately by using the duration curves (Fig. 2.10.2) and the amplification curves for the linear case (Figs. 3.2.2 and 3.5.2 to 3.5.7). It is useful, however, to calculate the amplification factors for extremely abrupt bottoming ($k_b \rightarrow \infty$) to obtain a general understanding of the accompanying phenomena.

The system to be considered is illustrated in Fig. 3.7.1. It is assumed that the impact between m_2 and the base (occurring at $t = t_s$, $x_2 = d_s$) has a coefficient of restitution of unity. Hence m_2 will strike the base with velocity

$$[\dot{x}_2]_{t=t_s} = \sqrt{2gh\left(1 - \frac{d_s^2}{d_0^2}\right)}$$

(see equation (2.10.8)) and leave it at a velocity of the same magnitude but opposite sign. Perfect rebound of the whole package is also assumed.

The acceleration pulse will then look like the curve marked $k_b/k_0 \rightarrow \infty$ in Fig. 2.10.1.

There will be three regions in which to consider the relative displacement x:

$$\text{Region 1} \qquad 0 < t < t_s$$

$$\text{Region 2} \qquad t_s < t < 2t_s$$

$$\text{Region 3} \qquad t > 2t_s$$

The relative displacement ($x = x_1 - x_2$) of m_1 with respect to m_2 will have

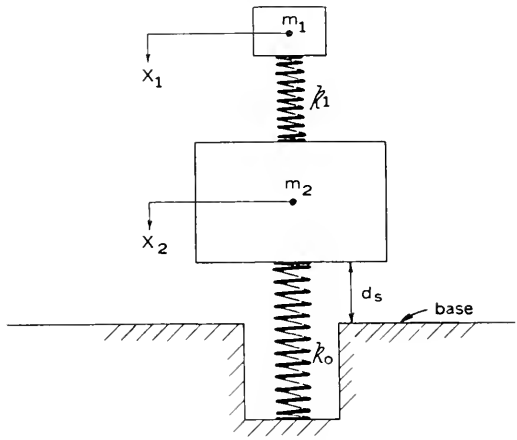


Fig. 3.7.1—Idealized system representing abrupt bottoming.

the same functional form for Region 1 as in the linear case (see equation (3.2.6)), and the amplification factor is, by analogy with (3.2.9),

$$A_0 = \frac{\dfrac{\omega_1}{\omega_0}}{\dfrac{\omega_1}{\omega_0} - 1} \sin \frac{2n\pi}{\dfrac{\omega_1}{\omega_0} + 1}, \qquad 0 < t < t_s, \tag{3.7.1}$$

where

$$\omega_0^2 = k_0/m_2 .$$

For Region 2, we use the differential equation

$$\ddot{x} + \omega_1^2 x = \omega_0 \sqrt{2gh} \sin \omega_0(t - 2t_s) \tag{3.7.2}$$

and, as initial conditions at $t = t_s$, we use the terminal conditions for Region 1 with the sign of $[\dot{x}_2]_{t=t_s}$ reversed. The amplification factor for this region is found to be (by the same method as in Section 3.2):

$$A_0 = \frac{\omega_1^2}{\omega_0\sqrt{2gh}}\sqrt{A^2 + B^2}\sin(\omega_1 t_m + \eta - \omega_1 t_s) - \frac{1}{1 - \left(\frac{\omega_0}{\omega_1}\right)^2}\sin\left(\omega_0 t_m - \omega_0 t_s - \sin^{-1}\frac{d_s}{d_0}\right), \quad (3.7.3)$$

$$t_s < t < 2t_s$$

where

$$\frac{\omega_1^2}{\omega_0\sqrt{2gh}}A = -\frac{\frac{\omega_0}{\omega_1}}{1 - \left(\frac{\omega_0}{\omega_1}\right)^2}\sin\left(\frac{\omega_1}{\omega_0}\sin^{-1}\frac{d_s}{d_0}\right)$$

$$\frac{\omega_1^2}{\omega_0\sqrt{2gh}}B = \frac{\frac{\omega_0}{\omega_1}}{1 - \left(\frac{\omega_0}{\omega_1}\right)^2}\left[2\frac{\omega_1^2}{\omega_0^2}\sqrt{1 - \frac{d_s^2}{d_0^2}} - \cos\left(\frac{\omega_1}{\omega_0}\sin^{-1}\frac{d_s}{d_0}\right)\right]$$

$$\eta = \tan^{-1}\frac{A}{B}$$

and t_m is the root of

$$\frac{\omega_1^3}{\omega_0^2\sqrt{2gh}}\sqrt{A^2 + B^2}\cos(\omega_1 t_m + \eta - \omega_1 t_s) - \frac{1}{1 - \left(\frac{\omega_0}{\omega_1}\right)^2}\cos\left(\omega_0 t_m - \omega_0 t_s - \sin^{-1}\frac{d_s}{d_0}\right) = 0$$

that yields the largest value of A_0 in equation (3.7.3). Region 3 is governed by

$$\ddot{x} + \omega_1^2 x = 0 \quad (3.7.4)$$

and the initial conditions are the terminal conditions of Region 2. By the same method as was used in Section 3.2, we find

$$A_0 = \frac{2\frac{\omega_1}{\omega_0}}{\left(\frac{\omega_1}{\omega_0}\right)^2 - 1}\left[\left(\frac{\omega_1}{\omega_0}\right)^2\sqrt{1 - \frac{d_s^2}{d_0^2}} - \cos\left(\frac{\omega_1}{\omega_0}\sin^{-1}\frac{d_s}{d_0}\right)\right], \quad (3.7.5)$$

$$t > 2t_s.$$

The largest value of A_0 from equations (3.7.1), (3.7.3) and (3.7.5) is plotted against ω_1/ω_0 in Fig. 3.7.2 for several values of d_s/d_0.

Notice that the amplification factor is A_0 rather than A_m. That is, the reference acceleration is G_0 rather than G_m. This is necessary because G_m is infinite in the present instance. Hence Fig. 3.7.2 cannot be compared directly with Figs. 3.2.2 and 3.6.2. However, it is interesting to observe that, for $\omega_1/\omega_2 < 0.5$, (low frequency elements) abrupt bottoming has no harmful effect. For high-frequency elements, the severity of bottoming is very great even when very nearly all of the required space (d_0) is available. For example, if 90% of the required space is available ($d_s/d_0 = 0.9$) and the frequency of the element is ten times the package frequency,

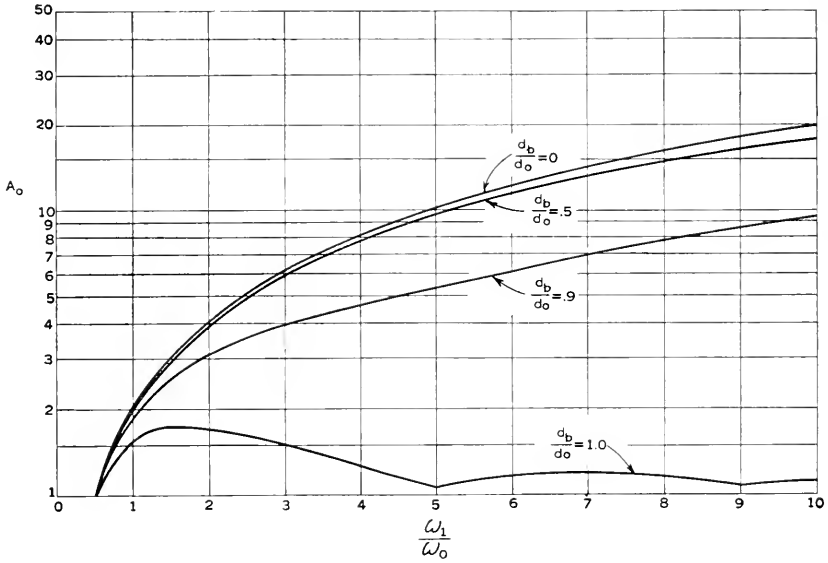


Fig. 3.7.2—Amplification factors for abrupt bottoming. See equations (3.7.1), (3.7.3) and (3.7.5).

the severity of the shock is almost ten times as great as it would be if the additional 10% of space were available.

3.8 General Influence of Shape of Acceleration-Time Curve on Amplification Factor

When amplification factor curves are not available for a special shape of acceleration-time curve, an approximate value of A_m may be obtained by interpolation between or extrapolation from the curves of the preceding sections. The shape of the acceleration-time curve and its duration (τ_2) or frequency (ω_2) should be found, first, by the methods described in Part II. The shape found should then be compared with the standard shapes shown in Part II, for which amplification factors are given in Part III.

The amplification factor found in this way will generally be within 25%

of the true value because amplification curves for pulse accelerations do not differ greatly even for very different acceleration-time curves as long as the

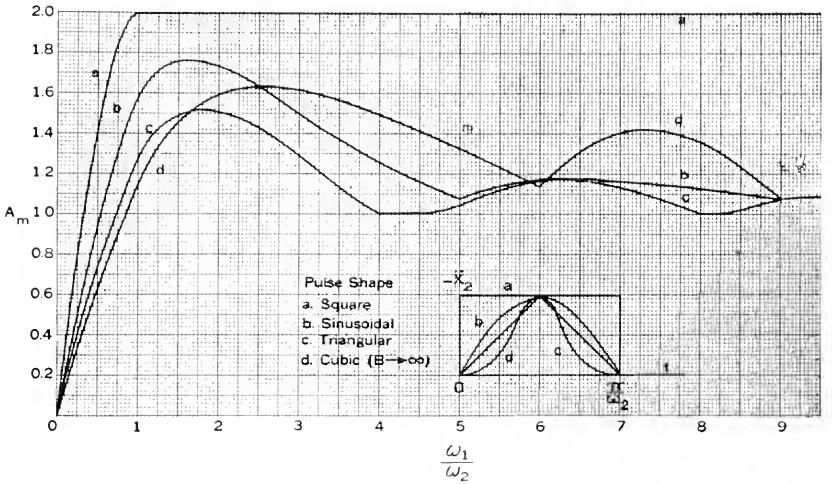


Fig. 3.8.1—Dependence of amplification factor on shape of symmetrical acceleration pulse.

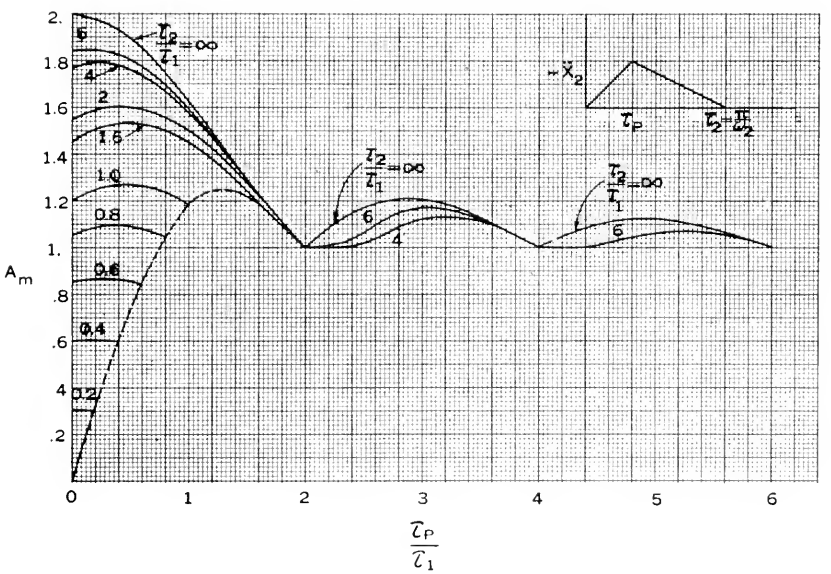


Fig. 3.8.2—Effect of asymmetry of an acceleration pulse on amplification factor.

amplitudes and frequencies are adjusted to the same scales. This is illustrated in Fig. 3.8.1 where the amplification factor curves are drawn for square wave, half-sine wave, triangular and cubic pulses.

Amplification factors for small values of ω_1/ω_2 may be calculated very accurately if it is observed that the initial slope of the amplification factor curve for a pulse acceleration is proportional to the area under the acceleration-time curve. For example, noting that the initial slope of the amplification factor curve for the half-sine wave pulse is 2, we assign the value 2 to the area under the half-sine wave. On the same scale, the area under a square wave pulse is π and under a triangular pulse is $\pi/2$. Accordingly, the initial slopes of the amplification factor curves for the latter two pulses are π and $\pi/2$ respectively.

As an additional aid in finding amplification factors for unusual cases, Fig. 3.8.2 is given to show the effect of asymmetry of an acceleration pulse. The pulse is triangular in shape but the time (τ_P) taken to reach the peak value of acceleration may have any value from zero to the total duration (τ_2) of the pulse.

PART IV

DISTRIBUTED MASS AND ELASTICITY

4.1 Introduction

It is important to be aware of the conditions under which the assumption of lumped parameters is permissible. In Parts I and II the cushioning medium was assumed to be massless, so that wave propagation (or surges) through it was ignored. Such surges will contribute to the acceleration imposed on the packaged article and we should be able to predict both the magnitudes and frequencies of the additional disturbances. If this is done, the information in Part III may be used to obtain at least a rough estimate of the resulting effects. In Part III itself the effects of accelerations were determined by studying the response of a system having only one degree of freedom; that is, an element of the packaged article was assumed to be a single mass supported by a massless spring. Every real element, of course, has an infinite number of degrees of freedom, so that it is important to discover the contribution, of the higher modes of vibration of an element, to the overall response.

Both of these problems (distributed parameters of mass and elasticity in the cushioning medium and in an element of the packaged article) are studied in this part. One example of each type is considered, and the choice of the example in each case was influenced by considerations of expediency, namely that the mathematical derivations be relatively simple and lead to solutions for which not too lengthy computations are necessary to yield results that can be applied practically. At the same time, the examples chosen are believed to give some insight into several of the physical phe-

nomena involved. The treatment is by no means complete, but a more detailed investigation is beyond the scope of this paper.

4.2 Effect of Distributed Mass and Elasticity of Cushioning on Acceleration of Packaged Article

Referring to Fig. 4.2.1, we consider the packaged article, of mass m_2, to be supported by distributed cushioning of mass m_c and depth ℓ. The cushioning may be a pad, say of rubber, in which case ℓ is the pad thickness, or it may be a helical metal spring, in which case ℓ is the coil length. The package is dropped vertically from a height h and has attained a velocity v at the instant of contact ($t = 0$) of the outer container and the floor. The outer container is assumed to be heavy enough so that there is no rebound. A horizontal plane in the cushioning is located by a coordinate x measured from the end of the cushioning attached to the outer container. The vertical

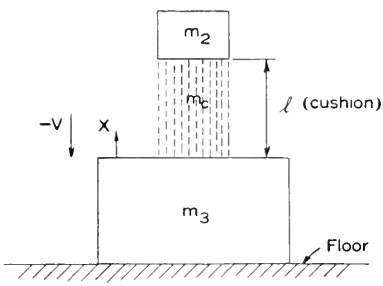


Fig. 4.2.1—Packaged article of mass m_2, supported on distributed cushioning of depth ℓ and mass m_c, depicted at the instant of first contact of the outer container (m_3) and the floor.

displacement of the plane x is designated by u. The undamped motion of the cushioning after contact is governed by the one-dimensional wave equation:

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \tag{4.2.1}$$

in which a is the velocity of propagation of longitudinal waves in the cushioning. If the cushioning is continuous,

$$a^2 = \frac{E}{\rho}, \tag{4.2.2}$$

where E is the modulus of elasticity and ρ is the density of the cushioning. If the cushioning is a helical spring,

$$a^2 = \frac{k\ell^2}{m_c}, \tag{4.2.3}$$

where k is the spring rate.

The initial conditions of the system are

$$[u]_{t=0} = 0, \tag{4.2.4}$$

$$\left[\frac{\partial u}{\partial t}\right]_{t=0} = -v. \tag{4.2.5}$$

The boundary conditions are

$$[u]_{x=0} = 0, \tag{4.2.6}$$

$$k\ell \left[\frac{\partial u}{\partial x}\right]_{x=\ell} = -m_2 \left[\frac{\partial^2 u}{\partial t^2}\right]_{x=\ell}. \tag{4.2.7}$$

Equation (4.2.7) expresses the requirement that the force on the upper end of the cushioning must balance the inertia force of the packaged article. For continuous cushioning $k\ell$ should be replaced by EA, where A is the cross-sectional area of the cushioning.

A solution of (4.2.1) satisfying conditions (4.2.4) and (4.2.6) is

$$u = \sum_{n=1}^{\infty} A_n \sin \frac{\omega_n x}{a} \sin \omega_n t, \tag{4.2.8}$$

where ω_n is the n^{th} root of a transcendental equation to be obtained from (4.2.7) and A_n is a constant to be determined by (4.2.5). Substituting (4.2.8) in (4.2.7) and equating coefficients of like terms of the series, we obtain the transcendental equation

$$\frac{\omega_n \ell}{a} \tan \frac{\omega_n \ell}{a} = \frac{m_c}{m_2}. \tag{4.2.9}$$

Substituting (4.2.8) in (4.2.5) we obtain, by the usual methods of expansion into trigonometric series,

$$A_n = -\frac{2v}{\omega_n \left(\frac{\omega_n \ell}{a} + \frac{1}{2} \sin \frac{2\omega_n \ell}{a}\right)}. \tag{4.2.10}$$

Hence the complete solution of the problem is

$$u = -\sum_{n=1}^{\infty} \frac{2v \sin \frac{\omega_n x}{a} \sin \omega_n t}{\omega_n \left(\frac{\omega_n \ell}{a} + \frac{1}{2} \sin \frac{2\omega_n \ell}{a}\right)}. \tag{4.2.11}$$

Our chief interest is in the acceleration of m_2. Making use of (4.2.7) and (4.2.9) we find, from (4.2.11), that this acceleration is

$$\left[\frac{\partial^2 u}{\partial t^2}\right]_{x=\ell} = v\omega_0 \sum_{n=1}^{\infty} B_n \sin \omega_n t, \tag{4.2.12}$$

where

$$\omega_0^2 = \frac{k}{m_2} \tag{4.2.13}$$

and

$$B_n = \frac{2\sqrt{\frac{m_c}{m_2}\left(\frac{m_c^2}{m_2^2} + \frac{\omega_n^2 l^2}{a^2}\right)}}{\frac{m_c}{m_2} + \frac{m_c^2}{m_2^2} + \frac{\omega_n^2 l^2}{a^2}}. \tag{4.2.14}$$

The acceleration of m_2 is, therefore, a sum of sinusoids of frequency ω_n and amplitude $v\omega_0 B_n$. Now, $v\omega_0$ is the maximum acceleration that m_2 would attain if the mass of the cushioning were negligible. Calling G_n the maximum acceleration in the n^{th} mode and G_0 the maximum acceleration neglecting the mass of the cushioning, as in Part I, we have

$$\frac{G_n}{G_0} = B_n. \tag{4.2.15}$$

But B_n depends only on the ratio m_c/m_2, as may be seen from equations (4.2.9) and (4.2.14). Similarly the ratio of the frequency (ω_n) of any mode to the frequency (ω_0) with massless cushioning depends only on m_c/m_2, as may be seen from equations (4.2.3), (4.2.9) and (4.2.13). *Hence, both the amplitude and frequency ratios for the acceleration in any mode depend only on the ratio of the mass of the cushioning to the mass of the packaged article.* The ratios G_n/G_0 and ω_n/ω_0 are plotted against m_c/m_2 in Figs. 4.2.2 and 4.2.3 for the first five modes. *It may be seen from these figures that the accelerations in the higher modes can be very important.* For example, if the cushioning weighs half as much as the packaged article the maximum acceleration in the second mode is about 40% of the acceleration in the first mode and the latter is about the same as found by the elementary method of Part I. This could have a disastrous effect on an element of the packaged article if the latter had a fundamental frequency near that of the second mode of the cushioning, the latter being found, from Fig. 4.2.3, to be about five times the fundamental frequency of the package.

It must be remembered that damping has been neglected in the above investigation and damping in the cushioning will serve to mitigate the severity of the higher mode accelerations to a great extent. However, the danger is always present at the start of a design and the possibilities of unfavorable combinations should be studied in every case.

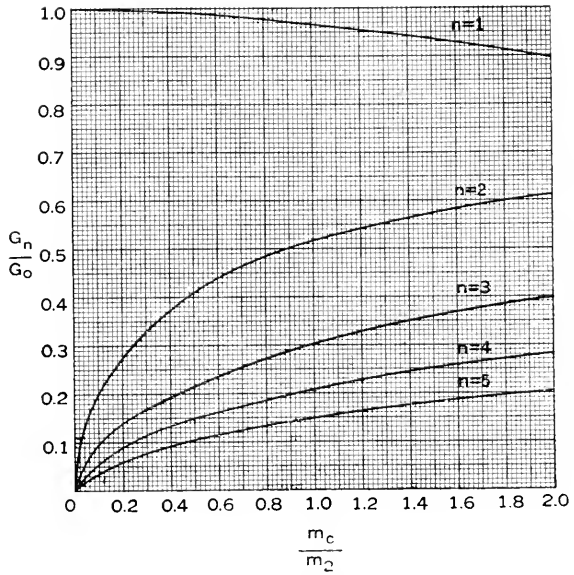


Fig. 4.2.2—Influence of ratio of mass of cushioning (m_c) to mass of packaged article (m_2) on acceleration ratio. The numerator of the acceleration ratio is the maximum acceleration (G_n) in the nth mode of vibration transmitted through the cushioning. The denominator of the acceleration ratio is the maximum acceleration $\left(G_0 = \sqrt{2hk_2/m_2g}\right)$ that the mass m_2 would experience if the mass of the cushioning were negligible. See equations (4.2.15), (4.2.14), (4.2.9).

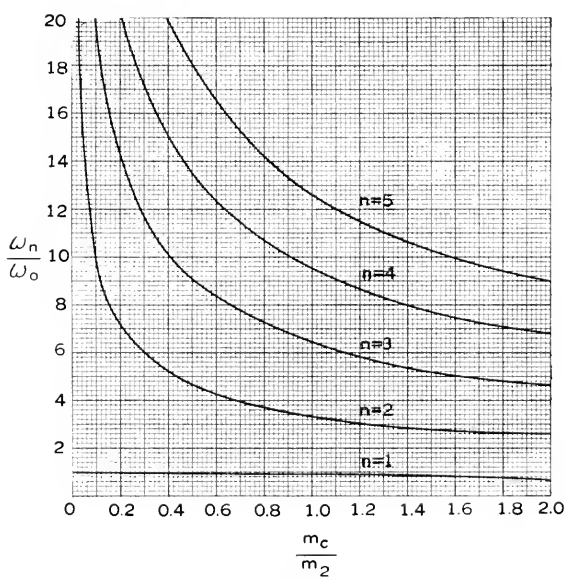


Fig. 4.2.3—Influence of ratio of mass of cushioning (m_c) to mass of packaged article on frequency ratio. The numerator of the frequency ratio is the frequency (ω_n) of the nth mode of vibration transmitted through the cushioning. The denominator of the frequency ratio is the frequency $\left(\omega_o = \sqrt{k_2/m_2}\right)$ of vibration of the mass m_2 neglecting the effect of the mass of the cushioning. See equations (4.2.9), (4.2.13), and (4.2.3).

4.3 Effect of Distributed Mass and Elasticity, of an Element of the Packaged Article, on the Amplification Factor for a Half-Sine-Wave Pulse Acceleration

In this section we shall determine the contribution of the higher modes of vibration of a structural element to its total response to a half-sine-wave pulse acceleration. For the shape of the element, we choose a prismatic bar because this leads to the simplest mathematical formulation of the problem and such a bar is also a common structural element. Other considerations influence the choice of direction of acceleration with respect to the axis of the bar. The transverse direction (cantilever) is the most practical from a physical standpoint, but, for purposes of comparison with the one-degree-of-freedom system, the parallel (axial) direction of acceleration is the more logical. Both problems lead to solutions in the form of infinite series, but, in the latter case, the expression for the strain at a fixed

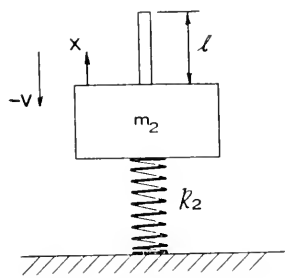


Fig. 4.3.1—The system studied in Section 4.3 depicted at the instant of contact with the floor.

end can be summed in terms of elementary functions without difficulty. Since it is necessary to determine maximum values of strain over a wide range of frequency ratios for the plotting of an amplification factor curve, an enormous reduction in the time required for accurate computations is obtained by choosing the axial case. Furthermore, the axial case appears to contain the essential features which might result in differences between the response of a one-degree-of-freedom system and a continuous one.

The complete system to be studied is illustrated in Fig. 4.3.1. To the mass m_2, supported on massless cushioning of constant spring rate k_2, is attached one end of an elastic prismatic bar, of length ℓ, cross sectional area A, modulus of elasticity E, and density ρ, with its axis oriented vertically. The system is dropped from a height h so that its velocity is v at the instant of contact of the cushioning with the floor. The mass of the bar is supposed to be small in comparison with m_2 and perfect rebound is assumed, so that the motion of m_2 during contact is a half-sine wave of frequency

$$\omega_2 = \sqrt{\frac{k_2}{m_2}}. \tag{4.3.1}$$

The maximum acceleration of m_2 is thus $v\omega_2$. If this magnitude of acceleration were reached very slowly, so as not to excite transient longitudinal waves in the bar, the maximum force between the bar and m_2 would be the product of the acceleration and the mass of the bar:

$$F = v\omega_2 \rho A \ell. \tag{4.3.2}$$

Hence the strain at the end of the bar attached to m_2 would be

$$\epsilon_0 = \frac{v\omega_2 \rho \ell}{E}. \tag{4.3.3}$$

Our problem is to find the ratio of the maximum transient strain to ϵ_0.

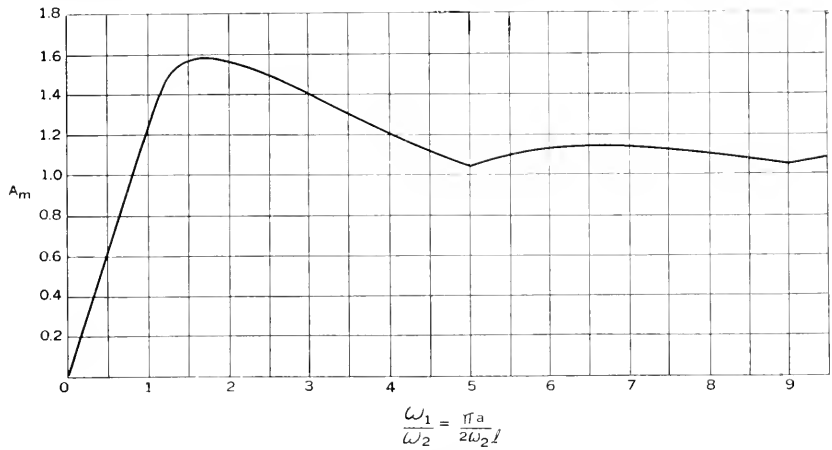


Fig. 4.3.2—Amplification factors for an element of the packaged article having distributed mass and elasticity. The package has linear undamped cushioning and perfect rebound. See equations (4.3.15), and (4.3.14).

Let u be the displacement of a transverse plane section of the bar distant x from the end attached to m_2. Then, the equation of motion of the bar is

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \tag{4.3.4}$$

where a is the velocity of propagation of longitudinal waves in the bar:

$$a^2 = \frac{E}{\rho}. \tag{4.3.5}$$

Taking the instant of first contact of the cushioning with the floor to be $t = 0$, we know, from Part II, that the system will leave the floor when $t =$

π/ω_2. We shall therefore treat separately, as in Part III, the motion during contact

$$0 \leqq t \leqq \frac{\pi}{\omega_2}$$

and after rebound

$$t \geqq \frac{\pi}{\omega_2}.$$

During the first interval, the initial and boundary conditions are

$$[u]_{t=0} = 0, \tag{4.3.6}$$

$$\left[\frac{\partial u}{\partial t}\right]_{t=0} = -v, \tag{4.3.7}$$

$$[u]_{x=0} = \frac{-v}{\omega_2} \sin \omega_2 t, \tag{4.3.8}$$

$$\left[\frac{\partial u}{\partial x}\right]_{x=l} = 0. \tag{4.3.9}$$

The first and second conditions state that, at the instant of contact, all points in the bar are moving with the approach velocity v, without relative displacement. The third condition prescribes the half-sine wave motion of the end of the bar that is attached to m_2. The fourth condition states that the strain at the free end of the bar is always zero.

By the usual methods, a solution of (4.3.4) satisfying conditions (4.3.6) to (4.3.9) is found to be

$$u = -\frac{v \cos \frac{\omega_2}{a}(\ell - x) \sin \omega_2 t}{\omega_2 \cos \frac{\omega_2 \ell}{a}} + \frac{8v\ell}{\pi^2 a} \sum_{n=1,3,5\cdots}^{\infty} \frac{\sin \frac{n\pi x}{2\ell} \sin \frac{n\pi a t}{2\ell}}{n^2\left[\left(\frac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} \tag{4.3.10}$$

$$\left(0 \leqq t \leqq \frac{\pi}{\omega_2}, \frac{n\pi a}{2\omega_2 \ell} \neq 1\right).$$

The displacement is seen to be a forced vibration at the frequency (ω_2) of the applied acceleration, on which are superposed the free vibrations of the bar given by the series expression. The frequency of the fundamental mode of vibration of the bar is $\pi a/2\ell$ and the frequencies of the higher modes are the odd integral multiples of the fundamental.

The strain at the attached end of the bar is

$$\epsilon = \left[\frac{\partial u}{\partial x}\right]_{x=0}$$

$$= -\frac{v}{a}\left\{\tan\frac{\omega_2 \ell}{a}\sin\omega_2 t - \frac{4}{\pi}\sum_{n=1,3,5\cdots}^{\infty}\frac{\sin\dfrac{n\pi a t}{2\ell}}{n\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]}\right\} \tag{4.3.11}$$

$$\left(0 \leqq t \leqq \frac{\pi}{\omega_2}, \frac{n\pi a}{2\omega_2 \ell} \neq 1\right).$$

It may be verified that the sum of the series in (4.3.11) is given by

$$\frac{4}{\pi}\sum_{n=1,3,5\cdots}^{\infty}\frac{\sin\dfrac{n\pi a t}{2\ell}}{n\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} = \tan\frac{\omega_2 \ell}{a}\sin\omega_2 t + \cos\omega_2 t - 1, \tag{4.3.12}$$

$$\left(0 < t < \frac{2\ell}{a}\right).$$

It should be observed that the summation is valid only in the interval $0 < t < 2\ell/a$. However, the series is periodic with half period $2\ell/a$ and includes only the odd terms, so that the function repeats itself with reversed sign after each interval $2\ell/a$. Hence the summation, valid for all t, can be written

$$\frac{4}{\pi}\sum_{n=1,3,5\cdots}^{\infty}\frac{\sin\dfrac{n\pi a t}{2\ell}}{n\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]}$$

$$= (-1)^k\left[\tan\frac{\omega_2 \ell}{a}\sin\omega_2\left(t - \frac{2k\ell}{a}\right) + \cos\omega_2\left(t - \frac{2k\ell}{a}\right) - 1\right] \tag{4.3.13}$$

$$k = m \quad \text{when} \quad \frac{2m\ell}{a} < t < \frac{2(m+1)\ell}{a}$$

$$m = 0, 1, 2, 3, \cdots .$$

We may, therefore, rewrite (4.3.11) in the form

$$\frac{\epsilon a}{v} = -\tan\frac{\omega_2 \ell}{a}\sin\omega_2 t + (-1)^k\left[\tan\frac{\omega_2 \ell}{a}\sin\omega_2\left(t - \frac{2k\ell}{a}\right)\right.$$

$$\left. + \cos\omega_2\left(t - \frac{2k\ell}{a}\right) - 1\right] \tag{4.3.14}$$

$$0 \leqq t \leqq \frac{\pi}{\omega_2}$$

$$k = m \text{ when } \frac{2m\ell}{a} < t < \frac{2(m+1)\ell}{a}$$

$$m = 0, 1, 2, 3 \cdots .$$

The expression (4.3.14) is simple enough so that the maximum value (ϵ_m) of the strain at the attached end can be obtained without difficulty for any ratio of the fundamental frequency ($\omega_1 = \pi a/2\ell$) of the bar to the frequency (ω_2) of the disturbing acceleration. The amplification factor

$$A_m = \left| \frac{\epsilon_m}{\epsilon_0} \right| = \frac{|\epsilon_m| E}{v\omega_2\rho\ell} = \frac{2\omega_1 a}{\pi\omega_2 v} |\epsilon_m| , \tag{4.3.15}$$

may then be calculated. The results of these calculations are plotted in Fig. 4.3.2. The important feature of this curve is that the amplification factor is everywhere less than the corresponding amplification factor for the one-degree-of-freedom system (Fig. 3.2.2, $\beta_1 = 0$). *Hence the assumption of lumped parameters is on the side of safety as regards amplification factor.*

It is interesting to observe that the curve of A_m vs. ω_1/ω_2, for this case, is a straight line between $\omega_1/\omega_2 = 0$ and $\omega_1/\omega_2 = 1$. This arises from the fact that, for $\omega_1/\omega_2 \leqq 1$, equation (4.3.14) reduces to

$$\frac{\epsilon a}{v} = \cos \omega_2 t - 1, \qquad \left(\frac{\omega_1}{\omega_2} \leqq 1\right). \tag{4.3.16}$$

Hence, when the duration of shock is less than the half period of the fundamental mode of vibration, the maximum value of strain occurs at the end of impact and is equal to twice the ratio of the approach velocity to the velocity of wave propagation in the bar.

The whole solution of the problem is not yet completed; for, although it is fairly evident from the fact that there is at least one maximum in the interval $0 \leqq t \leqq \pi/\omega_2$ for all values of ω_1/ω_2, it must be verified that the maximum strain (and, therefore, the amplification factor) is never greater after $t = \pi/\omega_2$ than before. Defining a new time coordinate

$$t' = t - \frac{\pi}{\omega_2}, \tag{4.3.17}$$

we have, for the initial and boundary conditions of equation (4.3.4) for $t \geqq \pi/\omega_2$,

$$[u]_{t'=0} = \frac{8v\ell}{\pi^2 a} \sum_{n=1,3,5\cdots}^{\infty} \frac{\sin \dfrac{na\pi^2}{2\omega_2 \ell} \sin \dfrac{n\pi x}{2\ell}}{n^2\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} \tag{4.3.18}$$

$$\left[\frac{\partial u}{\partial t}\right]_{t'=0} = \frac{v \cos \dfrac{\omega_2}{a}(\ell - x)}{\cos \dfrac{\omega_2 \ell}{a}} + \frac{4v}{\pi} \sum_{n=1,3,5\cdots}^{\infty} \frac{\cos \dfrac{na\pi^2}{2\omega_2 \ell} \sin \dfrac{n\pi x}{2\ell}}{n\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} \tag{4.3.19}$$

$$[u]_{x=0} = vt', \tag{4.3.20}$$

$$\left[\frac{\partial u}{\partial x}\right]_{x=l} = 0. \tag{4.3.21}$$

The first and second conditions state that the displacement and velocity of every point in the bar must be the same at the beginning of the second interval as at the end of the first interval; the expressions in (4.3.18) and (4.3.19) are obtained from (4.3.10). The third condition prescribes the constant velocity of departure from the floor of the mass m_2 and, therefore, of the end of the bar attached to it. The last condition states, again, that the strain at the free end of the bar is zero.

It may be verified that a solution of (4.3.4) satisfying conditions (4.3.18) to (4.3.21) is

$$u = vt' + \sum_{n=1,3,5\cdots}^{\infty} C_n \sin \frac{n\pi x}{\ell} \sin\left(\frac{n\pi a t'}{2\ell} + \gamma_n\right) \tag{4.3.22}$$

$$(t' \gtrless 0,\ t \gtrless \pi/\omega_2),$$

where

$$C_n \sin \gamma_n = \frac{8v\ell \sin \dfrac{na\pi^2}{2\omega_2 \ell}}{\pi^2 a n^2\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} \tag{4.3.23}$$

$$C_n \cos \gamma_n = \frac{8v\ell\left(1 + \cos \dfrac{na\pi^2}{2\omega_2 \ell}\right)}{\pi^2 a n^2\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} \tag{4.3.24}$$

$$\frac{n\pi a}{2\omega_2 \ell} \neq 1.$$

Hence, the strain at the attached end of the bar is

$$\epsilon = \left[\frac{\partial u}{\partial x}\right]_{x=0} = \frac{4v}{\pi a} \sum_{n=1,3,5\cdots}^{\infty} \frac{\sin \dfrac{n\pi a t}{2\ell}}{n\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]} + \frac{4v}{\pi a} \sum_{n=1,3,5\cdots}^{\infty} \frac{\sin \dfrac{n\pi a t'}{2\ell}}{n\left[\left(\dfrac{n\pi a}{2\omega_2 \ell}\right)^2 - 1\right]}. \tag{4.3.25}$$

The two series may be summed, as before, with the result

$$\frac{\epsilon a}{v} = (-1)^k \left[\tan \frac{\omega_2 \ell}{a} \sin \omega_2\left(t - \frac{2k\ell}{a}\right) + \cos \omega_2\left(t - \frac{2k\ell}{a}\right) - 1\right]$$
$$+ (-1)^{k'} \left[\tan \frac{\omega_2 \ell}{a} \sin \omega_2\left(t' - \frac{2k'\ell}{a}\right) + \cos \omega_2\left(t' - \frac{2k'\ell}{a}\right) - 1\right] \tag{4.3.26}$$

$$t \gtreqless \pi/\omega_2\,, \qquad t' = t - \pi/\omega_2$$

$$k = m \quad \text{when} \quad \frac{2m\ell}{a} < t < \frac{2(m+1)\ell}{a}\,,$$

$$k' = m' \quad \text{when} \quad \frac{2m'\ell}{a} < t' < \frac{2(m'+1)\ell}{a}$$

$$m = 0, 1, 2, 3 \cdots \qquad m' = 0, 1, 2, 3 \cdots .$$

Once more, the expression for the strain at the attached end of the bar is in a form suitable for rapid calculation and it can be shown the ϵ in equation (4.3.26) for $t \gtreqless \pi/\omega_2$ is never greater than the ϵ in equation (4.3.14) for $0 \leqq t \leqq \pi/\omega_2$ for the same ω_1/ω_2. Hence, Fig. 4.3.2 and the conclusions following equations (4.3.15) and (4.3.16) need not be modified.

NOTATIONS

A	Cross sectional area of a bar element of the packaged article. Also, a constant of integration.
A_0	Amplification factor when the reference acceleration is G_0. Ratio of maximum dynamic response to the response to a slowly applied acceleration of magnitude $G_0 g$.
A_m	Amplification factor when the reference acceleration is G_m. Ratio of maximum dynamic response to the response to a slowly applied acceleration of magnitude $G_m g$.
A_n	In Section 1.15, the sum of all the trapezoidal areas from $x_2 = 0$ to $x_2 = (x_2)_n$. Also, in Section 4.2, the coefficient of the nth term of a series.
ΔA_n	The area of a trapezoid with altitude $\Delta(x_2)_n$ and sides P_{n-1} and P_n.
a	x_0/l in the tension spring package. Also, in Part IV, the velocity of propagation of longitudinal waves.

B	A parameter of cushioning with cubic elasticity defined in equation (1.5.3). Also, a constant of integration.
B_n	Coefficient in the n^{th} term of a series.
b	f/l in the tension spring package.
C	A constant of integration.
C_n	Coefficient of the n^{th} term of a series.
c	A constant defined in equation (1.7.11)
c_1	Damping coefficient of an element of a packaged article.
c_2	Damping coefficient of linear cushioning.
cn	The elliptic cosine function.
d_0	Hypothetical displacement that would result if initial spring rate were maintained.
d_b	Maximum possible displacement of packaged article in cushioning with tangent elasticity.
d_m	Maximum displacement of packaged article.
d_m'	Value of d_m when $k_0 = k_0'$.
d_s	Displacement of bi-linear cushioning at which the spring rate changes from k_0 to k_b.
E	Modulus of elasticity.
e	In the tension spring package the stretch of a spring when the displacement is d_m.
exp ( )	$e^{(\,)}$, where e is the Naperian base 2.718$\cdots$
F	In section 2.7, a frictional force.
F_m	In the tension spring package, the maximum force on a spring.
f	In the tension spring package, the difference between l and the distance between hooks of an unstretched spring.
f_1	Frequency of vibration of an element of the packaged article.
f_2	Frequency of vibration of the packaged article on its cushioning.
G_0	Hypothetical maximum acceleration (in number of times g) that would result if initial spring rate were maintained.
G_e	A_mG_m or A_0G_0, i.e. the slowly applied acceleration (in number of times g) that will produce the same maximum response as a transient acceleration of maximum value G_m or G_0.
G_F	Maximum acceleration (in number of times g) in cushioning with friction and spring rate k_F.
G_m	Absolute value of maximum acceleration of packaged article in units of "number of times gravitational acceleration."
G_m'	Value of G_m when $k_0 = k_o'$.

G_n In section 1.15, the maximum acceleration (in number of times g) experienced by the suspended mass when dropped from a height h_n. In Part IV, the maximum acceleration (in number of times g) of the n^{th} mode of vibration.

G_r Maximum acceleration (in number of times g) after rebound.

G_s Safe value of G_c.

g Gravitational acceleration.

h Height of drop.

h_n In Section 1.15, the height of fall that will cause the cushioning to displace an amount $(x_2)_n$.

K In the tension spring package, the initial spring rate of the suspension. In Section 2.8, the complete elliptic integral of the first kind.

K_1, K_2, K_3 The initial spring rates in the three mutually perpendicular directions normal to the faces of the package frame.

k In the tension spring package, the spring rate of a spring. In Section 2.8, the modulus of an elliptic integral.

k, k' In Section 4.3, 0, 1, 2, 3, $\cdots$.

k_0 Initial spring rate of non-linear cushioning.

k_0' Optimum value of initial spring rate k_0.

k_1 Spring rate of lumped elasticity of element of packaged article.

k_2 Spring rate of linear cushioning.

k_b Spring rate of bilinear cushioning after bottoming.

k_F Spring rate defined in equation (2.7.7).

L Constant defined in equation (1.8.2).

l In the tension spring, the projection of l_i on a horizontal plane. In Section 4.2, length of cushioning. In Section 4.3, length of element of packaged article.

l_i In the tension spring package, the distance between the two support points of a spring when the suspended article is in the equilibrium position.

M Constant defined in equation (1.8.4), equal to G_m/G_0.

m Reduced mass defined in equation (2.4.5).

m, m' In Section 4.3, 0, 1, 2, 3, $\cdots$.

m_1 Lumped mass of element of packaged article.

m_2 Lumped mass of packaged article.

m_3 Lumped mass of outer container.

m_c Mass of cushioning.

N M^2.

n	0, 1, 2, 3, $\cdots$.
P	Force transmitted through cushioning.
P_0	Asymptotic value of force transmissible through cushioning with hyperbolic tangent elasticity.
P_m	Maximum force exerted on packaged article by cushioning.
P_n	In Section 1.15, the load that produces displacement $(x_2)_n$.
R	Force between package and floor.
r	Coefficient of cubic term in load-displacement function for cushioning with cubic elasticity.
s, t, u	The direction cosines of the acceleration direction with respect to the normals to the faces of the package frame.
sn	The elliptic sine function.
T_2	The period of vibration of the packaged article on its cushioning.
t	Time coordinate.
t'	$t - \frac{\pi}{\omega_2}$
t_0	Time of first contact of package with floor.
t_m	Time at which maximum displacement or acceleration occurs.
t_r	Time at which package leaves floor on rebound.
t_s	Time at which the displacement reaches d_s.
u	Displacement in x direction.
v	Approach velocity.
W_2	Weight of packaged article.
W_3	Weight of outer container.
x	$x_1 - x_2$; relative displacement of m_1 with respect to m_2.
$\dot{x}$	$\dot{x}_1 - \dot{x}_2$.
$\ddot{x}$	$\ddot{x}_1 - \ddot{x}_2$.
x'	Relative displacement of m_1 with respect to m_2 at time t'.
x_0	In the tension spring package, the perpendicular distance from an inner spring support point to the nearest plane, perpendicular to the displacement direction and containing four outer spring support points.
x_1	Displacement of m_1.
$\dot{x}_1$	Velocity of m_1.
$\ddot{x}_1$	Acceleration of m_1.
x_2	Displacement of m_2.
$\dot{x}_2$	Velocity of m_2.

$\ddot{x}_2$	Acceleration of m_2.
$x_{\max}$	Maximum value of x.
$(x_2)_n$	In Section 1.15, the displacement associated with the n^{th} point.
x_{st}	The value x would have if the acceleration reached its maximum value in a very long time.
$(\Delta x_2)_n$	In Section 1.15, equals $(x_2)_n - (x_2)_{n-1}$.
y	$x_2 - x_3$.
z	x_2/l (tension spring package).
$\alpha, \gamma, \delta, \zeta, \eta$	Phase angles.
β_1	Fraction of critical damping of an element of the packaged article.
β_2	Fraction of critical damping of package cushioning.
γ_n	Phase angle of n^{th} term of series (equation (4.3.22)).
ϵ	Strain at attached end of element under transient conditions.
ϵ_0	Strain at attached end of element under non-transient conditions.
ϵ_m	Maximum strain at attached end of element under transient conditions.
θ	Angle between the displacement direction and the acceleration direction.
π	3.14159$\cdots$.
ρ	Density (mass per unit of volume)
τ_0	Pulse duration of a half-sine-wave acceleration.
τ_2	Pulse duration associated with non-linear cushioning.
τ_B	Duration of bottoming of cushioning with bi-linear elasticity.
τ_P	Time required to reach peak value of a triangular acceleration pulse.
ω	Radian frequency defined in equation (2.4.6).
ω_1	Radian frequency of vibration of an element of the packaged article.
ω_1'	Radian frequency of vibration of damped element of packaged article.
ω_2	Radian frequency of vibration of the packaged article on its cushioning.
ω_2'	Radian frequency of vibration of the packaged article on damped cushioning.
ω_b	A frequency defined in equation (2.10.10).
ω_c	A frequency defined in equation (2.8.8).
ω_n	Radian frequency of nth mode.

Abstracts of Technical Articles by Bell System Authors

Dimensional Stability of Plastics.[1] ROBERT BURNS. Because of inherent insulating properties, rigid plastics play an important part in the design and manufacture of precision electrical apparatus. Almost invariably, practical design considerations require that the plastics have reasonable structural possibilities since it is rarely practicable to disassociate completely electrical and structural functions.

This paper discusses one of the important factors in the successful use of plastics in precision devices, namely, dimensional stability. Since plastics are organic compounds, one must be prepared to accept a degree of instability not usually encountered in metals. The measurement of this property is therefore of prime importance to the user of plastics since the data provide a basis for design adjustment which frequently is the difference between failure and success.

The various types of dimensional change are reviewed. Data illustrating the separate effects of humidity, drying, and cycling procedures are submitted. The influence of fabricating processes such as compression or injection molding, and sheeting, is included.

Some Numerical Methods for Locating Roots of Polynomials.[2] THORNTON C. FRY. It is the purpose of this paper to discuss the location of the roots of polynomials of high degree, with particular reference to the case of complex roots. This is a problem with which the Laboratories has been much concerned in recent years because of the fact that the problem arises rather frequently in the design of electrical networks. Attention is not given to strictly theoretical methods, such as the exact solution by elliptic or automorphic functions: nor to the development of roots in series or in continued fractions, though such methods exist and one at least—development of the coefficients of a quadratic factor—is of great value in improving the accuracy of roots once they are known with reasonable approximation.

Instead, the paper deals with just two categories of solutions: one, the solution of the equations by a succession of rational operations, having for their purpose the dispersion of the roots; the other, a method depending on Cauchy's theorem regarding the number of roots within a closed contour.

Thermistor Technics.[3] J. C. JOHNSON. This paper is confined to a study of how the three basic types of thermistors, namely, externally-heated or ambient temperature type, the directly-heated type, and also the indirectly-

[1] *A.S.T.M. Bulletin*, May 1945.
[2] *Quarterly Applied Mathematics*, July 1945.
[3] *Electronic Industries*, August 1945.

heated type, are used in simple feedback amplifiers as regulation and control devices to effect the economies inherent in an entirely electrical system by eliminating such mechanical devices as motor-driven condensers, sliding contacts and rotary switches.

Dynamic Measurements on Electromagnetic Devices.[4] E. L. NORTON. A method is presented by which measurements of flux may be made at any desired time during the operate cycle of an electromagnet. Apparatus is described which operates the magnet cyclically at an accurately held rate, and provides a means for measuring flux either by the use of a search coil or by the operating winding of the magnet itself. When using a search coil, it is connected to a direct-current milliammeter at the time in the cycle at which the value of the flux is desired and disconnected at the end of the cycle or just before the magnet is energized for the next pulse. If proper precautions are taken, the steady reading of the instrument is an accurate measure of the difference in the flux in the coil between the time it is connected to the meter and the time it is removed, or, since the latter is zero except for residual flux, the reading is a direct measure of flux.

The same apparatus may be used for the measurement of instantaneous current by the addition of an air core mutual inductance, and its use is extended to the measurement of armature position and velocity by the addition of a photoelectric cell and the proper amplifiers.

A form of vacuum tube filter is described which effectively filters the pulses from the indicating instrument without affecting the accuracy of the measurements.

Coaxial Cables and Television Transmission.[5] HAROLD S. OSBORNE. Communication techniques and facilities useful to the entertainment industry have evolved naturally from the Telephone Companies' main objective—the transmission of speech. The development of carrier systems for long-distance transmission and technical features involved in the latest carrier medium—the coaxial cable—are reviewed. The television transmission capabilities of this medium, both now and what may be expected shortly after the war, are mentioned. The extensive system of such cables planned for the next five years, supplemented by radio relay systems to the extent that these prove themselves as a part of a communications network, will provide an excellent beginning for a nation-wide television transmission network. Planned primarily to meet telephone requirements, this network of cables will be suitable to meet the transmission needs of the television industry.

The Performance and Measurement of Mixers in Terms of Linear-Network Theory.[6] L. C. PETERSON AND F. B. LLEWELLYN. This paper discusses

[4] *Elec. Engg., Transactions Section*, April 1945.
[5] *Jour. S.M.P.E.*, June 1945.
[6] *Proc. I.R.E.*, July 1945.

the properties of mixers in terms of linear-network theory. In Part I the network equations are derived from the fundamental properties of nonlinear resistive elements. Part II contains a résumé of the appropriate formulas of linear-network theory. In Part III the network theory is applied, first to the case of simple nonlinear resistances, and next to the more general case where the nonlinear resistance is embedded in a network of parasitic resistive and reactive passive-impedance elements. In Part IV application of the previous results is made to the measurement of performance properties. The "impedance" and the "incremental" methods of measuring loss are contrasted, and it is shown that the actual loss is given by the incremental method when certain special precautions are taken, while the impedance method is in itself incomplete.

A Figure of Merit for Electron-Concentrating Systems.[7] J. R. Pierce. Electron-concentrating systems are subject to certain limitations because of the thermal velocities of electrons leaving the cathode. A figure of merit is proposed for measuring the goodness of a device in this respect. This figure of merit is the ratio of the area of the aperture which, in an ideal system with the same important parameters as the actual system, would pass a given fraction of the cathode current to the area of the aperture which in the actual system does pass this fraction of the cathode current. Expressions are given for evaluating this figure of merit.

A 60-Kilowatt High-Frequency Transoceanic-Radiotelephone Amplifier.[8] C. F. P. Rose. Here is described a high-frequency radio amplifier recently developed for the transoceanic-telephone facilities of the Bell System at Lawrenceville, New Jersey. In general, the amplifier is capable of delivering 60 kilowatts of peak envelope power when excited from a 2-kilowatt radio-frequency source. It is designed to operate as a "class B" amplifier for transmitting either single-channel double-sideband or twin-channel single-sideband types of transmission. Features are described which permit rapid frequency-changing technique from any preassigned frequency to another lying anywhere within the spectrum of 4.5 to 22 megacycles.

Some Notes on the Design of Electron Guns.[9] A. L. Samuel. A method is outlined for the design of electron guns based on the simple theory first published by J. R. Pierce. This method assumes that the electrons are moving in a beam according to a known solution of the space-charge equation, and requires that electrodes exterior to the region of space charge be shaped so as to match the boundary conditions at the edge of the beam. An electrolytic tank method is used to obtain solutions for cases which are not amenable to direct calculation. Attention is given to some of the

[7] *Proc. I.R.E.*, July 1945.
[8] *Proc. I.R.E.*, October 1945.
[9] *Proc. I.R.E.*, April 1945.

complications ignored by the simple theory and to some of the practical difficulties which are encountered in constructing guns according to these principles. An experimental check on the theory is described, together with some information as to the actual current distribution in a beam produced by a gun based on this design procedure.

Microwave Radiation from the Sun.[10] G. C. SOUTHWORTH. During the summer months of 1942 and 1943, a small but measurable amount of microwave radiation was observed coming from the sun. This appeared as random noise in the outputs of sensitive receivers designed to work at wavelengths between one and ten centimeters. Over a considerable portion of the range, the energy was of the same order of magnitude as that predicted by black-body radiation theory.

Attempts were made to determine the effect of the earth's atmosphere on this radiation. Measurements made near sunrise or sunset, when the path through the earth's atmosphere was relatively long, differed only slightly from those made at noon. This suggested that any absorption that may have been present was small. In this connection it is of interest that small temperature differences could be noted between points below the horizon and the sky immediately above. This also suggested that the earth's atmosphere was relatively transparent.

In another kind of measurement the parabolic receiver was centered on the sun and its output was observed as the sun's disc moved out of the aperture of the receiver. The directional pattern so obtained indicated that at the shorter wave-lengths the sun's apparent diameter was considerably larger than that measured by ordinary optical means. This suggested that there may have been some refraction or perhaps scattering by the earth's atmosphere.

Resistive Attenuators, Pads and Networks—An Analysis of their Applications in Mixer and Fader Systems (Part Eight of a Series).[11] PAUL B. WRIGHT. In last month's discussion, the series-connected fader and the parallel-connected fader systems were considered, together with an analysis of their performance expressed both algebraically and in terms of the hyperbolic functions of a real variable. In this installment, the series-parallel-connected fader system discussion is continued and equations describing the complete behavior of this type network system are developed. This is followed by further analytical work dealing with the parallel-series-connected fader and mixer system and several lesser known systems which are quite useful to use. These are the *multiple bridge* and the *lattice network systems* which may be utilized to advantage for some applications. All of

[10] *Jour. Franklin Institute*, April 1945.

[11] Communications, September 1945. (*Preceding parts of this Series appeared in earlier issues of Communications.*)

the equations which are derived are shown in the algebraical, hyperbolical and symbolical forms. The key chart which was presented earlier in this series may be used to great advantage when checking the definitions of the symbols used which are not specifically defined in the text. This procedure also may be directly applied to the hyperbolic equations shown. It is of course necessary to take into account that, in general, subscripts are used in most of the equations in the text while the key chart does not have any subscripts. This does not, however, alter the fundamental forms nor their definitions in terms of the propagation function, theta. To avoid the necessity for extensive interpolation of the hyperbolic function tables to find the correct numerical values for the various functions used throughout the text, a series of tables providing all of the functions required is presented.

Contributors to this Issue

Raymond D. Mindlin, B.A., Columbia University, 1928; B.S., 1931; C.E., 1932; Ph.D., 1936. Assistant 1932–38, Instructor 1938–40, Assistant Professor 1940–45, Associate Professor 1945–, Department of Civil Engineering, Columbia University. Consultant, Section T, National Defense Research Committee (later Office of Scientific Research and Development), 1940–45, on the development of the rugged radio proximity fuze and on mathematical problems in fire control. Consultant, Bell Telephone Laboratories, 1943–44, Member of the Technical Staff 1944–45, Consultant 1945–. Dr. Mindlin has been concerned with the fields of mathematical and experimental mechanics.

J. R. Pierce, B.S. in Electrical Engineering, California Institute of Technology, 1933; Ph.D., 1936. Bell Telephone Laboratories, 1936–. Engaged in study of vacuum tubes.

A. L. Samuel, A.B., College of Emporia (Kansas), 1923; S.B. and S.M. in Electrical Engineering, Massachusetts Institute of Technology, 1926. Additional graduate work at M.I.T. and at Columbia University. Instructor in Electrical Engineering, M.I.T., 1926–28. Mr. Samuel joined the Technical Staff of the Bell Telephone Laboratories in 1928, where he has been engaged in electronic research and development. Since 1931, his principal interest has been in the development of vacuum tubes for use at ultra-high frequencies.