The Bell System Technical Journal

January, 1929

Decibel—The Name for the Transmission Unit

By W. H. MARTIN

IN 1923 the "mile of standard cable" was replaced in the Bell System by a new unit for expressing telephone transmission efficiencies and levels. At that time the generic term "transmission unit" was taken to designate this new unit, since it was considered desirable to defer the adoption of a more distinctive name until this unit had been given further consideration by others who would have use for a unit of this type. This new unit is defined by the statement that two amounts of power differ by one transmission unit when they are in the ratio of $10^{.1}$, and any two amounts of power differ by N transmission units when they are in the ratio of $10^{N(.1)}$. In accordance with this, the number of transmission units corresponding to the ratio of any two powers is ten times the common logarithm of that ratio.

For a unit of this kind, it is evidently desirable to have universal use. Accordingly, the Bell System, prior to its adoption of the transmission unit, discussed this matter with various foreign telephone administrations, and suggested their consideration of the use of the proposed "transmission unit." A number of these administrations expressed a favorable attitude towards this unit.

In 1924 there was organized the International Advisory Committee on Long Distance Telephony in Europe. The purpose of this committee, which is composed of representatives of the various telephone administrations of Europe, is to recommend standards and practices for the development of telephone service between the European countries. One of the early considerations of this committee was this proposal of the universal standardization of a unit for telephone transmission work. This brought forth a difference of view, since some of the countries represented on this committee wished to continue their use of a unit based on naperian or natural logarithms, for which the basic power ratio is e^2. The characteristics of the unit based on decimal logarithms and that based on natural logarithms and their relative merits were discussed in a number of papers which were published at that time * and so need not be rehearsed here.

At the request of the International Advisory Committee, representatives of the Bell System attended some of their meetings at which this matter was discussed. In this joint consideration there arose the suggestion that the fundamental unit on the decimal basis be defined to be equal in magnitude to that of ten transmission units, so that the basic power ratio would be 10^1. The units proposed thus came to one based on the power ratio of 10^1 and the other on the power ratio of e^2, with the provision that decimal submultiples of either unit could be employed, using the customary prefixes to give the proper indication. On this basis, the numbers of the two kinds of units corresponding to a given power ratio, differ by about 15 per cent. It was further suggested that the naperian unit be called the "neper" and that the fundamental decimal unit be called the "bel," these names being derived respectively from the names of Napier, the inventor of natural logarithms, and Alexander Graham Bell.

These joint considerations have had the following results. The European International Advisory Committee has recommended to the various European telephone administrations that they adopt either the decimal or naperian unit and designate them the "bel" and "neper" respectively. The Bell System has adopted the name "decibel" for the "transmission unit," based on a power ratio of $10^{\cdot 1}$. This is in accordance with the terminology for the decimal unit, the prefix "deci" being the usual one for indicating a one-tenth relation. For convenience, the symbol "db" will be employed to indicate the name "decibel."

* *References:*

"The Transmission Unit and Telephone Transmission Reference Systems," W. H. Martin, *A.I.E.E. Trans.*, Vol. 43, pp. 797–801, 1924, and *B.S.T.J.*, Vol. 3, pp. 400–408, July, 1924.

"The Transmission Unit," R. V. L. Hartley, *Electrical Communications*, Vol. 3, pp. 34–42, June, 1924.

"Telephone Transmission Standards," Dr. F. Breisig, *The Electrician*, Vol. 94, pp. 114–115, Jan. 30, 1925.

"Telephone Transmission Units," Colonel T. F. Purves, *The Electrician*, Vol. 94, pp. 144–145, Feb. 6, 1925.

"Telephone Transmission," G. D. Hood, *The Electrician*, Vol. 94, p. 369, March 27, 1925.

"Telephone Transmission Standards," Dr. F. Breisig, *The Electrician*, Vol. 94, pp. 454–455, April 17, 1925.

"The Telephone Transmission Unit," Colonel T. F. Purves, *The Electrician*, Vol. 94, pp. 535–542, May 8, 1925.

"Telephone Transmission Unit," F. B. Jewett, *The Electrician*, Vol. 94, pp. 562–563, May 15, 1925.

"Telephone Transmission Units," Dr. F. Breisig, *The Electrician*, Vol. 94, p. 630, May 29, 1925.

"The Telephone Transmission Unit," Dr. F. Luschen, *The Electrician*, Vol. 95, pp. 94–95, July 24, 1925.

The Principles of Electric Circuits Applied to Communication[1]

By H. S. OSBORNE

SYNOPSIS: This paper discusses the method of presenting in the curricula of engineering schools the fundamental electrical principles, emphasizing the desirability of presenting them as far as practical in a general way and of making clear the relations of specific applications, such as the relation between circuit theory equations as applied to power systems and to telephone systems, and the relation between ordinary circuit theory and the generalized electromagnetic equations. An outline is given of some interesting problems arising and results obtained in the application of electric principles to telephone systems.

ENGINEERING education shares with other forms of education the general movement toward greater emphasis on the unity of subject matter which plays such an important part in the development of modern educational methods. By the unity of subject matter I refer to the aim to reduce as far as practicable the number of separate compartments in which the educational subject matter is kept, and to present this subject matter under a smaller number of broader headings. Whereas the curriculum must be divided into a certain number of different courses for administrative and practical reasons, I take it the modern educational method is opposed to the presentation of these courses as individual entities, separated from other subjects, but insists rather that the curricular partitions be kept as low as possible so that the student may appreciate as fully as possible the continuity of each subject with its neighbors, and may obtain a good perspective of the close mutual relations of the different parts of the educational material and realize their mutual dependence and the large areas in which they are jointly applicable.

The wisdom of this move is evident to men in the industries as well as to educators. The tremendously rapid growth of fundamental electrical science and of the electrical industries have both worked rapidly in this direction. The time has long passed when it was at all possible to cover both fundamental electrical science and its applications in a four year engineering course. I judge that the old question "Is it more important to teach the fundamentals of electrical science

[1] Presented at Pittsburgh July 18, 1928, at the Summer School for Electrical Engineering Teachers under the auspices of the Society for the Promotion of Engineering Education.

The author expresses his appreciation of the assistance given him by Mr. C. O. Gibbon, American Telephone and Telegraph Company, in the preparation of this paper.

or their application?" is no longer even discussible, as far as the engineering school is concerned. Applications must indeed be learned by the student, but this can best be done after graduation, with the help of the industry employing him, and under the stimulus of the necessity of learning his job and preparing himself for greater usefulness in the organization. The question before the colleges is not "Shall we teach fundamentals?" but "What fundamentals shall we teach, and how can these most effectively be presented? How much of the fundamentals of our present far flung electrical science can we convey to the students in a four year course?"

The same question naturally presents itself in this discussion. The communication field, from the nature of the problems which it presents, has a great wealth of material relating to the principles of electric circuits, both as regards the practical application of these principles and research extending our fundamental knowledge of these principles. What phases of this subject matter should be presented here? In this Mr. Hammond's circular to the members of the summer school staffs is a guide. He points out that among the various purposes of the summer schools the principal aim is to produce tangible results for improving methods of teaching. I do not understand that I am expected, in talking with a group of educators, to discuss teaching methods directly, and will not presume to do so. However, in response to your invitation to discuss "The Principles of Electric Circuits Applied to Communication" I shall refer to the use in the communication field of those principles within the scope of student work which appear to have the broadest general application, including all fields of electrical engineering. Also, it will no doubt be of interest for us to give some consideration to the relative conditions, including similarities and differences, of application of these principles to communication problems and to other branches of the electrical industries.

Limiting the discussion in this way necessarily results in leaving untouched many phases of the application of electrical principles to communication which are of great interest but of less application in other fields, and so I have left out such important matters as, for example, modulation and demodulation, the balancing of line impedance characteristics by artificial lines, inductive effects between different circuits, and the performance characteristics of various types of apparatus.

Also, of course, this will in no sense be a general discussion of the work of the transmission engineers of the telephone companies. While their work is based on the application of electrical principles, and requires that they understand those principles, the theoretical work which they do on electric circuits is, with the exception of a few men, relatively small, and their work is also very largely based on the use of economic principles, knowledge of the telephone system, general business principles and common sense.

General Principles

Communication circuits in general are very complicated networks and the application of electric principles to these networks involves the solution of numerous new problems and the development of a great many practical approximations. To be effective in this work it is of prime importance that the young engineer have a good grounding in the general fundamental principles of direct and alternating current theory. By a good grounding we mean an appreciation of the generality of these principles so that they can be applied by the student to the problems of his particular work, no matter in what branch of electrical engineering that work may be. He needs also to have a facility in their application to new problems. The relations of resistance, reactance, conductance and susceptance, the use of Kirchhoff's laws, the relations of resonance and conditions for maximum transfer of energy between two branches of a network—all these we would list as a matter of course. We should include also in the list of basic fundamentals some simple but extremely useful practical theorems, including the reciprocity theorem and Pollard's or Thevenin's theorem.

Along with these fundamental principles we believe it helpful to a man in any branch of electrical work to have absorbed the idea of the equivalence of networks of different types; for example, the expression, in terms of equivalent T or Pi circuits or other convenient form of the characteristics of any 4 pole network (that is, a network with two input and two output terminals) from the measurements which can be made at its terminals. Fundamental training of this sort gives the man a mastery in the solution of electrical problems, not only through ability to place a given problem in its most convenient form, but by assisting the engineer in the formation of correct and simple physical ideas regarding the processes which are taking place.

A very good illustration of this point is given by the transformer. Most young men graduating from engineering schools, I believe, think of the operation of the transformer in terms of its vector diagram. This is a valuable way of getting a physical picture of the effect of a transformer which is useful in certain types of problems but less convenient in others. If in addition to the vector diagram the student is taught the equivalent network of a transformer, as is done in some texts, he can more easily apply it to other types of problems, and the equivalent T network of the transformer is so simple that it is very helpful in showing the variation in the performance of the circuit with changes in the constants of any part of the circuit.

Familiarity with the fundamental principles of networks and equivalent circuits is particularly helpful in case of a man who comes in contact with problems associated with networks made up of a number of similar sections, as for example, electric filters, which already play such an important part in certain branches of the communication art.

In the study of fundamental principles it would appear to be very helpful in enabling the student to get an appreciation of their generality if the specific problems and illustrations used in his work are drawn from the various fields of electric work rather than from a single field. The communication field is replete with specific problems illustrating these principles which are very suitable for the use of the student.

Transmission Line Theory

In discussing the application of the principles of electric circuits to communication it is natural to give particular attention to transmission line theory because of its importance in connection with the transmission of electrical energy for any purpose whatever, including both power and communication services, and because of the interest of the problems it involves. Transmission line theory in one sense dates back to Lord Kelvin who in 1855 applied laws of diffusion of heat to the determination of the flow of electricity through long submarine cables. This solution ignored the effect of line inductance which was unimportant in the particular problem to which Lord Kelvin applied this solution, but which is very important in any general transmission line theory. Through the work of Heaviside and others the general transmission line theory was at an early date applied to telephony. It is, of course, in relatively recent years that the great development of long distance power transmission lines has made the general theory of value in power transmission work, the performance of early alternating current systems being adequately represented by approximate formulas, entirely neglecting the effect of the capacity of the line. It is indeed not long ago that the effect of line capacity, assumed lumped at one or two points, became important and still more recently that it became necessary in power problems to take more accurate account of the distribution of the capacity.

It is no doubt partly as a result of the historical development of the application of transmission line theory to telephony and to power transmission problems, and partly the result of differences in conditions which the solution must meet in the two fields of application that it seems in the past not always to have been made clear to students that the power line equations and the telephone line equations are simply special solutions of the same general line formula.

To illustrate this point it is desirable to refer to a few well-known equations. The differential equations for what may be called the classical transmission line theory are given in equation 1. Equation 2 is a solution of these differential equations for the steady state, for the circuit indicated in Fig. 1.

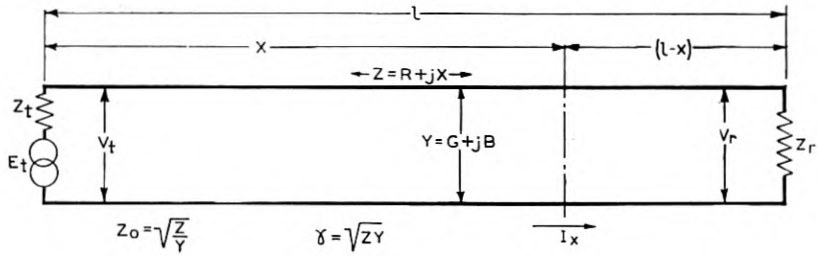


Fig. 1

$$\left.\begin{aligned}\left(L\frac{d}{dt}+R\right) I &= -\frac{\partial}{\partial x} V,\\ \left(C\frac{d}{dt}+G\right) V &= -\frac{\partial}{\partial x} I,\end{aligned}\right\} \tag{1}$$

$$\begin{aligned}I_x = \frac{E_t}{Z_0+Z_t}\Bigg\{ &\epsilon^{-\gamma x}\left[1+\left(\frac{Z_0-Z_t}{Z_0+Z_t}\cdot\frac{Z_0-Z_r}{Z_0+Z_r}\right)\epsilon^{-2\gamma l}\right.\\ &\left.+\left(\frac{Z_0-Z_t}{Z_0+Z_t}\cdot\frac{Z_0-Z_r}{Z_0+Z_r}\right)^2\epsilon^{-4\gamma l}+\cdots\right]\\ &+\epsilon^{-\gamma(l-x)}\left[\frac{Z_0-Z_r}{Z_0+Z_r}\epsilon^{-\gamma l}+\frac{Z_0-Z_t}{Z_0+Z_t}\left(\frac{Z_0-Z_r}{Z_0+Z_r}\right)^2\epsilon^{-3\gamma l}\right.\\ &\left.+\left(\frac{Z_0-Z_t}{Z_0+Z_t}\right)^2\left(\frac{Z_0-Z_r}{Z_0+Z_r}\right)^3\epsilon^{-5\gamma l}+\cdots\right]\Bigg\}.\end{aligned} \tag{2}$$

This equation indicates the current flowing at any point in the circuit for a given impressed voltage. The solution in this form seems to have special educational value because it gives a very clear physical picture of what is taking place in the transmission line. As was early pointed out by Heaviside, the current flowing in any simple circuit such as indicated can be considered to be built up from a directly transmitted wave which at the transmitting end has the magnitude of $\frac{E_t}{Z_t+Z_0}$. This direct wave is attenuated as it is propagated along the line, and at the receiving end is reflected in the ratio of the difference between the receiving impedance and characteristic line impedance to the sum of these impedances; propagated back toward the transmitting end with continued attenuation; reflected there if the transmitting end impedance is not equal to the characteristic line impedance, the doubly reflected wave propagated toward the receiving end, and so on in an infinite series of propagations and reflections.

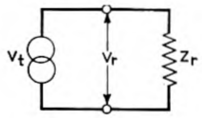
In both power and communication work one quantity which is of great importance in considering the characteristics of the transmission line is the ratio of the voltage at the transmitting end to that at the receiving end. This is, of course, readily derived from equation 2, and is presented in equation 3 in the beautifully compact form offered by the use of hyperbolic functions of the propagation constant of the line.

$$\frac{V_t}{V_r} = \cosh \gamma l + \frac{Z_0}{Z_r} \sinh \gamma l. \tag{3}$$

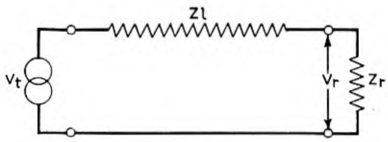
Considering first the application of this general transmission line theory to power transmission circuits, it would appear to be of great value for the student to appreciate the relationship between the general formula and the approximations used for short lines. This is brought out clearly by equation 4 and the diagrams and equations presented under 5.

$$\frac{V_t}{V_r} = 1 + \frac{Zl}{Z_r} + \frac{YZl^2}{2!} + \frac{YZ^2l^3}{3!Z_r} + \frac{Y^2Z^2l^4}{4!} + \frac{Y^2Z^3l^5}{5!Z_r} + \dots \tag{4}$$

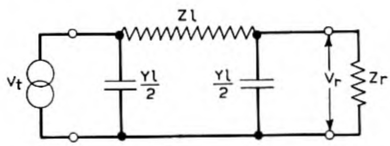
Equation 4 is simply the development of the general formula of equation 3 into a series of terms of ascending powers of Zl, the total resistance and reactance of the line, and of Yl the total shunt admittance of the line. The comparison of this series expansion with the results indicated by various approximate methods is of considerable interest. The first term, unity, is naturally the ratio of transmitting and receiving voltages with no transmission line, as indicated in 5*a*. With the addition of the second term $\left(\frac{Zl}{Z_r}\right)$ one has the result obtained by entirely ignoring the capacity of the line as indicated in 5*b*. The first three terms of the series give the result obtained by assuming that one half of the capacity of the line is concentrated at each end of the line as indicated in 5*c*, namely, a simple Pi network. The simple T network, assuming the capacity all concentrated at the middle as indicated in 5*d*, gives 4 terms, but you will note that the fourth term



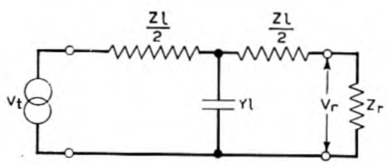
$$\frac{V_t}{V_r} = 1 \tag{5a}$$



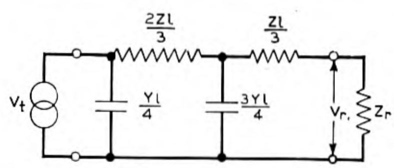
$$\frac{V_t}{V_r} = 1 + \frac{Zl}{Z_r} \tag{5b}$$



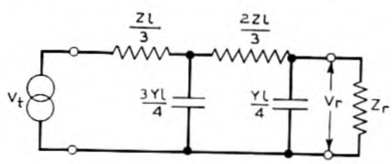
$$\frac{V_t}{V_r} = 1 + \frac{Zl}{Z_r} + \frac{YZl^2}{2!} \tag{5c}$$



$$\frac{V_t}{V_r} = 1 + \frac{Zl}{Z_r} + \frac{YZl^2}{2!} + \frac{YZ^2l^3}{4Z_r} \tag{5d}$$



$$\frac{V_t}{V_r} = 1 + \frac{Zl}{Z_r} + \frac{YZl^2}{2!} + \frac{YZ^2l^3}{3!Z_r} \tag{5e}$$



$$\frac{V_t}{V_r} = 1 + \frac{Zl}{Z_r} + \frac{YZl^2}{2!} + \frac{YZ^2l^3}{3!Z_r} + \frac{Y^2Z^2l^4}{4!} \tag{5f}$$

is 50 per cent. greater in magnitude than the fourth term of the series, and this approximation, therefore, does not commend itself for determining the voltage ratio, since if one wishes a precision requiring four terms of the series it is naturally better to use the correct fourth term. In order to correctly represent four terms of the series by an equivalent network it is necessary, as indicated in 5*e*, to assume one fourth of the capacity concentrated at the sending end of the line and the other three fourths concentrated at a point two thirds distant from the sending end. Finally, if this unsymmetrical network be reversed in direction, as indicated in 5*f*, the first five terms of the series are accurately reproduced.

The degree of approximation represented by dropping off various terms of this series is indicated for three typical cases in Figures 2, 3, and 4. Fig. 2 represents a typical 11,000 volt distribution line. It is to be noted that even neglecting the third term, which is the first in which the capacity of the line appears, results in an error of only 0.4 per cent. as indicated by consulting the values of ratios between successive terms of the series which are given with the diagram. The first ratio shown is that between the second and first term, and the

Fig. 2

20 Miles, 11 Kv., Three Phase Distribution Line
200 Kw. at 90% Power Factor, 60 ~
Ratio of Terms:

$$\frac{Zl}{Z_r} = .0540\underline{/2°}[1 \ \tfrac{1}{3} \cdots]$$

$$YlZ_r = .0668\underline{/116°}[\tfrac{1}{2} \ \tfrac{1}{4} \cdots]$$

second ratio when divided by two is that between the third and second term. The ratios between successive terms in the series are equal alternately to these two values divided by coefficients increasing in simple arithmetical progression as indicated. This fact would seem to make the series very convenient for computations.

To take the opposite extreme of power transmission, Fig. 3 has been prepared showing the degree of approximation resulting from the use of different numbers of terms of the series for a 220 kv. line, 250 miles

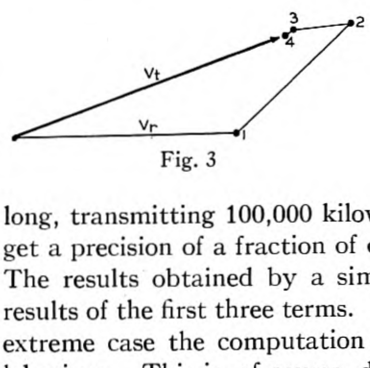


Fig. 3

250 Mile, 220 Kv., Three Phase Power Line
100,000 Kw. at 90% Power Factor, 60 ~
Ratio of Terms:

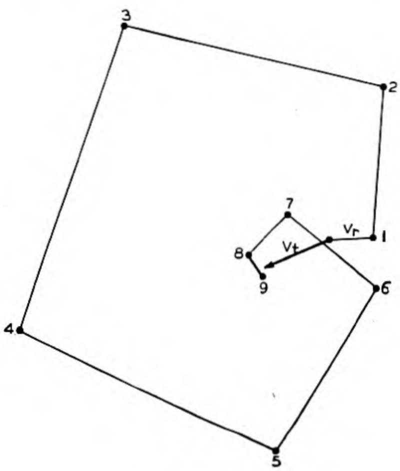
$$\frac{Zl}{Z_r} = .464\underline{/60°}[1 \ \tfrac{1}{3} \cdots]$$

$$YlZ_r = .588\underline{/116°}[\tfrac{1}{2} \ \tfrac{1}{4} \cdots]$$

long, transmitting 100,000 kilowatts. Here it is seen that in order to get a precision of a fraction of one per cent. four terms are necessary. The results obtained by a simple Pi network are indicated by the results of the first three terms. It is seen that even for this somewhat extreme case the computation by the series expansion is not at all laborious. This is, of course, due to the fact that even for 250 miles at 60 cycles the power line represents only about one-fifteenth of a wave-length.

Series expansions of the type discussed are, of course, not novel. I have dwelt on these somewhat, however, because of the value which they appear to have in clarifying the students' ideas about electric transmission, and because few students appear to be familiar with this method of treatment.

It is no doubt true that in many cases the student can best start transmission line theory with a simple approximation. It would seem, however, that before he gets through his study of the principles of electric circuits he should have a clear picture of the physical processes involved in the propagation of electric power over transmission lines, such as is given by equation 2, and of the assumptions involved in the various approximations which may be presented to him. If the scope of the course is not such as to permit the derivation of the general equations from the differential equations, it is possible to get equation



100 Mile, 8BWG Copper Telephone Circuit, 1000 ~
Terminating Impedance Equals Characteristic Impedance

Ratio of Terms:

$$\frac{Zl}{Z_r} = YlZ_r$$
$$= 3.5\underline{/83^\circ}[1 \ \tfrac{1}{2} \ \tfrac{1}{3} \ \tfrac{1}{4} \ \tfrac{1}{5} \ \tfrac{1}{6} \ \tfrac{1}{7} \ \tfrac{1}{8} \cdots]$$

Fig. 4

4 by a method of successive approximations as is shown in at least one textbook, and from 4 to derive the general equation 2.

In contrast with the power line cases, Fig. 4 indicates the results obtained by the series computation of a relatively short telephone toll line, an open wire circuit 100 miles long, and using 1,000 cycles as one typical telephone frequency. Although this line is only a little more than half a wave-length, the solution by the series for this case is quite laborious and indeed impracticable, and of course, would be even more so with the longer lengths or the higher attenuations ordinarily involved in telephone circuits.

The form of the general equation under discussion which is found most convenient for telephone use is given in equation 6.

$$\frac{E_t}{V_r} = \frac{Z_t + Z_r}{Z_r} \cdot \epsilon^{\gamma l} \cdot \frac{\dfrac{Z_0 + Z_t}{2\sqrt{Z_0 Z_t}} \cdot \dfrac{Z_0 + Z_r}{2\sqrt{Z_0 Z_r}}}{\dfrac{Z_t + Z_r}{2\sqrt{Z_t Z_r}}} \cdot \left[1 - \frac{Z_0 - Z_t}{Z_0 + Z_t} \cdot \frac{Z_0 - Z_r}{Z_0 + Z_r} \cdot \epsilon^{-2\gamma l} \right]. \quad (6)$$

This applies to the same circuit as equations 3 and 4, the only difference being that for telephone purposes the ratio E_t/V_r is used rather than V_t/V_r since in the telephone case the internal voltage of the generating apparatus rather than the terminal voltage is the more convenient reference voltage. Otherwise the two equations are identical consisting simply of a rearrangement of terms. The outstanding feature of the arrangement shown in equation 6 is that it consists of the product of a number of terms rather than the sum of a series of terms as in equation 4. The first term represents the ratio of voltages which would be obtained if there were no line present, namely $\frac{Z_t + Z_r}{Z_r}$. The second term illustrates the effect of propagation over the line itself without allowance for reflection at the terminals. The next three terms, two in the numerator and one in the denominator are the factors which make allowance for this reflection. Each is dependent upon simply the magnitude of two impedances, and their inclusion in the equation represents the fact that inserting the line between the two impedances inserts the reflection factors between the line and the transmitting impedance at one end, and the line and receiving impedance at the other end, and takes out the reflection factor directly between the transmitting and receiving impedances. The reflection factor becomes unity in any case in which the two impedances are equal. The last term of the equation is called the interaction factor because it represents the effect of multiple reflection back and forth between the two terminals of the line. This factor necessarily is complicated as it depends upon the characteristics of the line and on both transmitting and receiving impedances. It will be noted that this factor becomes substantially equal to unity in case either the transmitting impedance and the line impedance are approximately equal, the receiving impedance and the line impedance are approximately equal, or the attenuation of the circuit is considerable. In most practical telephone circuits these conditions are approximated sufficiently closely so that the departure of this factor from unity can usually be neglected.

In general, practical telephone circuits are, of course, a good deal more complicated in form than the simple circuit indicated above. It is, however, necessary to have for practical telephone purposes, means for readily determining to a good degree of approximation the overall efficiency of these complicated circuits as a part of the everyday work of certain departments of the telephone companies. The process of expressing the efficiency in terms of the product of a number of factors provides a convenient means for doing this. Under these conditions the large factors such as line attenuation and certain other factors are determined for the individual circuits, whereas factors which are close to unity can be treated approximately by tables representing various types of cases rather than individual circuits.

The convenience of treatment of circuit equations in this way for telephone use has led to the use of a logarithmic measure for expressing the efficiencies of telephone circuits. Although the computations can often be made in terms of currents or voltages, where changes of impedance are involved, such as inequality ratio transformers, we are, of course, concerned with variations in power rather than variations in either the current or voltage. The losses in a telephone circuit are therefore expressed in terms of the logarithm of the ratio of input power to output power. The unit is so defined that 10 transmission units correspond to a ratio of 10, 20 transmission units correspond to a ratio of 100, etc. The overall efficiency of practical telephone circuits from transmitter to receiver is in many cases in the order of 20 transmission units, that is, the circuit delivers at the receiving end one per cent. of the power delivered to it at the transmitting end.

Line Efficiency

A problem of great importance in electric transmission in all fields is that of obtaining the maximum transmission line efficiency practicable within economic limits. Certain comparisons between power transmission and telephone transmission would seem to be of interest.

The losses in a unit length of transmission circuit include both resistance losses and leakage losses, and may be represented by the formula $I^2r + V^2g$. If both r and g are constant with variations in voltage and current, it is easy to show that the maximum efficiency of transmission takes place when the voltage and current are so adjusted that these two parts of the line losses are equal in magnitude.

Actually in a well designed power transmission circuit without Corona losses V^2g is very small. Hence the solution of the problem of increasing line efficiency is to raise the transmission voltage, thus decreasing the transmission current. This is accomplished in the shorter power transmission lines by using step-up and step-down transformers at the ends of the line, the line itself having a relatively small effect on the impedances thus obtained. For the longer lines, however, the line characteristics play a very important part in this process, and this is best illustrated by consideration of the telephone case.

In the ordinary telephone line the leakage losses are also small compared with the resistance losses, and the voltage can be raised by a factor of 2 or 3 before these losses become equal. The solution of the problem here as in the case of the power line is therefore to raise the transmission voltage and decrease the current, that is, to increase the impedance of the line. In a telephone case, however, this cannot be done by changing the impedance of the terminal apparatus since, as has been pointed out, practically all of the power is absorbed in line losses, and therefore the impedance at the transmitting end of the line is not appreciably affected by the impedance of the receiving apparatus. In order to raise the ratio of voltage to current on the telephone line, it is therefore necessary to operate on the line itself.

The impedance of an electrically long transmission circuit is very approximately equal to $\sqrt{L/C}$. In telephone lines the most practical way to increase this impedance is to increase the inductance. This may be done, as you know, either by uniformly distributed inductance or by lumped inductance, providing certain essential conditions are met, and the result is what is called a "loaded" telephone circuit. It is particularly to be noted that this is not a resonance phenomenon. On the contrary, loading in this way tends to decrease the variations with frequency of the efficiency of transmission of the circuit, and when so proportioned as to give maximum power efficiency, results in distortionless transmission.

In the very long power lines on the other hand, it is desired to transmit efficiently only one frequency, the fundamental. The use of methods depending upon resonance is therefore permissible, and in fact the method which will undoubtedly have increasing use in the future as with increasing length of power transmission lines the effects of line capacity become more important, will be the partial neutralization of the effects of this capacity by shunt inductances distributed along the line, that is, induction machines, or synchronous machines underexcited. This reduces the equivalent capacity of the lines by supplying at least a part of the charging current at the intermediate points. Thus is a similar end obtained in the two cases by different means.

Long Line Phenomena

The notable success of the vacuum tube amplifier has made a great change in the character of the problems encountered in the design of very long telephone circuits, and our discussion would not be complete without a brief consideration of the nature of these problems. A detailed discussion of these problems is given in numerous papers in recent technical literature.

With the amplifier it is possible in a very large measure to overcome at a relatively small cost the effects of power losses in the telephone circuit. Whereas before the general use of the amplifier it was necessary to make every possible effort to further improve the efficiency of the long circuits in order to increase the distance over which satisfactory telephone transmission could be given, with the amplifier of today there is a limit to the amount of money which can properly be spent merely for the improvement of the volume efficiency of the line as the line losses can always be made up if desirable by the use of amplifiers. As a result, in general, very long telephone circuits have become electrically so long that factors other than power efficiency determine the limits of their effectiveness. While a quantitative theoretical discussion of these problems is necessarily in large measure beyond the scope of undergraduate work at the present time, this may not long be the case, and in any event a general appreciation of these phenomena is of a good deal of interest.

Although these effects are common to all long circuits in principle, they are most prominent in very long telephone cable circuits as these are electrically the longest circuits in use. For example, the propagation constant of a toll circuit in cable between New York and Chicago is at 1,000 cycles approximately $50 + j300$. This is approximately the same as the propagation constant of a high voltage power line transmitting power at 60 cycles of 25,000 to 50,000 miles in length. If there were no intermediate amplification in the circuit the ratio of input to output power would be ten to the 45th power so that with our usual telephone input of about one milli-watt the circuit would deliver only one electron in each two months, and even if all the power available in New York City or Chicago could be used at the input without burning up the circuit, the received current would be utterly inappreciable. However, there are far more practical reasons than this for frequent intermediate amplification. The lower limit to which the power level can be permitted to fall in the circuit is limited by the disturbances picked up from other telephone circuits in the same cable or from other electric circuits outside the cable, and the maximum power level is, of course, limited by considerations of economy in the design of the amplifiers. These limits result in the use of amplification in these circuits at approximately 50 mile intervals, there being 17 intermediate points of amplification between New York and Chicago on the shortest route.

With such a circuit the variation of efficiency with temperature is very rapid and the emergence of the sun from under the clouds could make as much as 1,000 fold-difference in the amount of power delivered. It is therefore necessary for practical operation to control the power gain introduced by the amplifiers by means of pilot wires in the cable subject to the same temperature variations as the talking circuit, and in this way to compensate automatically for the effect on the propagation constant of varying temperature.

In order to obtain for these cable circuits as far as possible the benefits of high voltage transmission, the circuits are loaded. This loading also in large measure equalizes the efficiency of transmission over the frequency range required for the transmission of speech.

The velocity of wave propagation over conductors loaded with inductance is, of course, relatively slow compared with the velocity of light. In the case of loaded telephone circuits in cable the velocity for the two types of circuit in general use is respectively 10,000 miles a second and 20,000 miles a second. The low velocity circuits are loaded with more inductance and are of higher efficiency and therefore preferable from the standpoint of volume. It is necessary, however, for the long circuits to use facilities of higher velocity and lower efficiency because of several very interesting phenomena.

For one thing, with circuits of high efficiency conforming to present day standards, the currents reflected from the distant end because of irregularity of impedance between the line and the terminal apparatus are by no means inappreciable. When the time for the propagation of these currents over the line and back to the transmitting end is very short, the reflected currents can be large without interference with service as they are indistinguishable from the sound directly heard by the speaker. If a circuit from New York to Chicago were used on the lower speed of the two types of circuit mentioned above, however, the reflected currents would arrive about one fifth of a second late. An interval of this magnitude would result in serious confusion to the speaker due to hearing his words twice, by direct transmission and after reflection from the distant end of the circuit. With the high speed facilities the time interval is reduced to one tenth of a second, and the interfering effect is very much smaller. Even with the high speed facilities, however, the effect is sufficient so that on circuits of over a few hundred miles in length special devices known as echo suppressors are used to intercept the echo currents and prevent their interference with the speaker.

Another important effect is the imperfect equalization of the transmission of different frequencies within the range of important telephone frequencies. This imperfection can be offset by the use at intervals of correcting networks which introduce a distortion opposite to that produced by the line, and the design of such networks is one of the interesting problems which has been worked out in connection with these very long circuits. The distortion is not only one of magnitude but also one of phase due to the difference in the velocity of wave propagation of component currents of different frequencies. This may be extremely important on long circuits of the heavier type of loading in which any minor disturbance at the transmitting end of the line is transmitted in such a way that the low frequency components appear first at the receiving end, followed by progressively higher frequency components and causing disturbing transient noises somewhat similar to the chirrupping of a bird. Phase distortion is less on the more lightly loaded circuits but still remains of enough importance to require the use in some cases of networks to equalize the distortion of phase.

In this discussion of very long circuits I have talked of telephone circuits. In the case of both telephone and telegraph circuits the fundamental requirements are the same, namely, the propagation of currents within a certain range of frequencies without excessive distortion and without interference from other electric circuits. The principal difference in the problems is in the range of frequencies which is important in the two cases, that for telegraph being much lower and more limited in extent than that for telephone. Another difference is that in the case of telephony phase distortion is important only in producing different time of arrival of different components, whereas in telegraphy the effect of phase distortion in distorting wave shape is also of importance. The telegraph problem can, like the problem of transients in telephone lines, be approached theoretically from the performance of a circuit when a potential is applied suddenly at one end. It has been shown, however, that the treatment of the circuit in terms of its steady state characteristics for the propagation of alternating currents over a range of frequencies leads to results identical with those reached by the transient treatment, and for most cases the steady state method of treatment rather than the transient method of treatment is found to be more convenient to handle for purposes of circuit and apparatus design.

Electromagnetic Theory

There is a further gap which it is to be hoped can in the future be bridged for those students who are sufficiently advanced to become familiar with the general electromagnetic theory. To what extent it is practical to teach general electromagnetic theory to undergraduates is a question which is perhaps beyond the scope of this paper. However, those five differential equations as formulated by Maxwell and Lorentz which form the general mathematical statement of the fundamental discoveries of Ampere, Faraday and the other pioneers of electric science are the Magna Charta of electric science of today, and with the rapid development of the electrical arts in various directions constant recourse must be had to these fundamental equations for the establishment of correct electrical principles.

The simplified electric circuit theory which we have just discussed, which may be called the classic theory, serves very well for the great bulk of problems of electrical transmission of today. However, already there are situations both in the power transmission art and in the communication art in which the approximations which these equations involve are not valid, and for the solution of practical cases recourse must be had to the more general equations. These practical cases include the distribution of current in the earth when one side of a circuit is grounded, the inductive effects produced in other electrical circuits from a grounded circuit, and the transmission characteristics of submarine cables in which the sea water forms a part of the return path.

The classic circuit theory expresses the electrical quantities in terms of the total currents flowing in conductors and the voltages between these conductors, and expresses the aggregate energies in terms of inductances and capacities, and the dissipation in terms of resistances. The general equations of the electromagnetic theory express the electric quantities in terms of elementary current and charge densities, and the electric and magnetic fields are expressed in terms of field strengths. These, then, are the rigorous equations in differential form. In the classic theory the current and charge densities and field strengths are integrated into more easily manipulated totals. In other words, the electric circuit theory deals with macroscopic or large scale phenomena and the electromagnetic theory deals with microscopic or small scale phenomena.

What are the approximations involved in the classic theory and what conditions must be met for these to be good approximations? This matter has been treated in a very interesting way in some recent papers by Mr. John R. Carson. In brief, Mr. Carson's papers point out that the classic circuit theory applied to transmission lines involves the following assumptions:

1. That the solution is concerned only with conditions at some distance from the terminals of the circuit and can therefore ignore the changes in distribution of electric and magnetic fields near the terminals known as end effects. That means that the electric and magnetic fields are propagated along the line in the same way as the currents and voltages.
2. That the propagation constant (per centimeter) is very small compared with unity and that the real part of the propagation constant is not large compared with the imaginary part, that is, the attenuation is not large compared with the phase change.
3. That in the conductors the loss due to the transmission current (that is, the axially flowing current) is large compared with the loss due to the charging currents.
4. That in the dielectric the propagation of energy is nearly parallel to the axis of the conductors and the dissipation in the dielectric is negligible.
5. That the fields of the currents and charges are propagated at an infinite velocity, that is, that radiation is neglected.

These assumptions, it can be shown, are very good for the ordinary case of an efficient transmission system. The effect of modification of the field at the terminals influences only a few feet of the line and is negligible in amount. Assumptions 2, 3 and 4 can very readily be shown to be true from the characteristics of the conductors and dielectrics involved in transmission, and as regards neglecting of radiation Mr. Carson has shown that for the ordinary transmission system the losses by radiation are in the order of one ten-thousandth of those in the conductors.

It is to be noted, however, that these assumptions place certain limitations on the application of the classic theory which are important in certain cases. The limitations are as follows:

1. The electric and magnetic fields are accurately expressed only for points relatively near the conductors. At great distances from the conductors the radiation field becomes important in comparison with the inductive field because of the much more rapid rate of decrease in intensity of the inductive field with distance.
2. The electric and magnetic fields are not accurately expressed near the terminals of the circuits.

3. The approximations do not apply if the conductors are quite imperfect or if the dielectric is highly dissipative. This affects the application in certain practical cases as indicated above.
4. The classic theory does not apply to circuits of the usual dimensions for extremely high frequencies in the order of millions of cycles.

From the standpoint of the best appreciation of the fundamentals of the electrical arts it might be said that the ideal course for students in electrical theory would start with the fundamental discoveries of Ohm, Oersted, Ampere, Faraday and Henry, pass through the generalized mathematical statement of the laws which they discovered to the general electromagnetic equations of Maxwell and Lorentz, and then from this focal point derive the various approximations of electric theory as applied to the various electrical arts; electric light and power, electric transportation, telephone, telegraph and radio. Unfortunately, teachers and students as well as the rest of us are hampered by questions of time and this approach is not proposed as a practical undergraduate course at the present time.

Conclusion

Electrical science and its practical applications are undergoing a very rapid development and expansion. The science of today is the engineering of tomorrow. These facts result in increasing importance in the mastery by the engineering student during his time at college of the electrical principles of broadest general usefulness, rather than learning specific applications of these principles. By a mastery is meant such an appreciation of the scope and limitations of the principles that he is able to apply them correctly to new conditions as they come up.

It is not intended to express a judgment on the extent to which the curricula of engineering schools should go in presenting electrical theory, and it is, of course, recognized that different schools have different conditions to meet which will naturally result in somewhat different courses. It is not proposed that the specific forms of electrical theory applying directly to telephone problems should be taught.

It is proposed for consideration, however, that whatever the scope of general electrical principles which is taught, these be so presented that the student have a clear picture of what they mean and of how and where they apply, and that he also should appreciate the relation to the general principles of any specific cases presented and the approximations which they involve. As far as practicable, all principles should be related to the general electromagnetic theory, the fundamental basis of all our electrical science.

Magnetic Properties of Perminvar[1]

By G. W. ELMEN

SYNOPSIS: This paper describes the magnetic properties of a group of iron-nickel-cobalt alloys, named "perminvar." With certain heat treatments these alloys have unusual constancy of permeability and extremely small hysteresis losses at low flux densities, and peculiarly shaped hysteresis loops constricted in the middle as the maximum flux densities of the loops are increased. Methods of preparing and heat treating the alloys are described, limits of composition, and changes in the magnetic properties with composition and with different heat treatments are illustrated. A theory of constitutional changes effected by heat treatment and responsible for the unusual magnetic properties is suggested.

IN 1921 the writer was investigating the magnetic properties of a series of permalloys[2] to which a few per cent of a third metal was added to the nickel and iron. One of these alloys contained cobalt. Magnetic measurements indicated that up to moderate field strengths the permeability of this nickel-cobalt-iron alloy was remarkably constant. The constancy was materially better than for soft iron, notwithstanding the fact that the initial permeability was several times higher. This was unusual, as small permeability variation ordinarily is found only in materials with low permeability. Measurements of other magnetic properties were equally surprising. When the hysteresis loop was traced for a cycle which carried the flux density up to a few thousand gauss, it was found to have an extraordinary form in that it was sharply constricted in the middle. These and other differences which were observed indicated that this alloy was a new type of magnetic material in which the magnetic properties were fundamentally different from those of previously known materials.

This discovery aroused a great deal of interest for it was recognized that magnetic materials possessing these properties were of great scientific and technical importance. In order to develop the possibilities which this alloy suggested, an exploration of the whole field of the iron-nickel-cobalt series was undertaken. For it was, of course, apparent that the alloy which had aroused our interest must be one of a group of compositions which possessed similar properties in a greater or less degree. In this survey, alloys varying in 10 per cent steps in composition and including the whole range of the ternary

[1] Reprinted from *The Journal of the Franklin Institute*, Vol. 206, No. 3, September, 1928.

[2] Arnold & Elmen, *Jour. of Frank. Inst.*, May, 1923, p. 621.

series of these metals were made up and their magnetic properties measured.

These measurements showed the range of compositions which shared in such unusual magnetic properties, and indicated that heat treatment was an important factor in the development of these properties. A large number of alloys have been made up in this range, for which the variations in composition were evenly distributed but much smaller than for the initial survey. From these alloys a few were selected which appeared to be specially suited for magnetic uses in electrical communication circuits. Our experience with these alloys has been that when good grades of commercial materials are used, the castings are readily reduced mechanically to the desired dimensions, and the magnetic properties from different castings of the same composition are quite uniform.

We felt that these alloys were so unique as regards magnetic quality that they should be grouped in a class under a common name which should readily distinguish them from other materials. We have chosen "perminvar" as the name for alloys in the iron-cobalt-nickel series, which are characterized, when properly heat treated, by constancy of permeability for a considerable range of the lower part of the magnetization curve, by small hysteresis loss throughout the same range of flux densities, and by a hysteresis loop constricted at the origin for medium flux densities.

This paper describes the magnetic properties of the perminvar group of alloys. Results are given for several alloys selected to show the variation in magnetic properties when the proportions of the constituent metals are varied over a wide range. Detailed measurements under a variety of magnetic conditions and heat treatments are recorded for the composition 45 per cent nickel, 25 per cent cobalt and 30 per cent iron. This composition is a typical one and was chosen early in our experimental work as specially suitable for commercial uses, for it had, in addition to the unusual properties in which we were most interested, a fairly high initial permeability.

Preparation of Alloys

The alloys were cast from the best available commercial materials. Armco iron, electrolytic nickel and commercial cobalt were melted together in the desired proportions in a silica crucible in a high frequency induction furnace. Before pouring, one half of one per cent of metallic manganese was added to the molten metal. Part of this manganese deoxidized the metal and went into the slag, and the remainder, usually about one half of the added amount, remained in the alloy. The alloys also contained small amounts of carbon (less than .03 per cent), silicon (less than .1 per cent) and traces of sulphur and phosphorus. The alloys were cast into bars 18 in. long and 3/4 in. in diameter. The bars were rolled or swedged into 1/4 in. rods and drawn from that size to .062 in. diameter wire. This wire was flattened and trimmed into tape 1/8 in. × .006 in. The material was annealed several times in the reduction process, for the cold working hardened the alloys rapidly and made them difficult to work.

To prepare the tape for heat treatment and subsequent magnetic measurements, about 30 ft. of it was wound spirally into a ring of 3 in. inside diameter, the ends being spot welded to the adjacent turns. Care was taken to wind the rings loosely to prevent the turns of tape from sticking during annealing.

A number of such rings were packed in a nichrome pot. Some iron dust was usually placed in the pot to take up the oxygen and thus prevent the oxidation of the rings. Further protection was secured by luting the joint between the pot and its cover. The pot was placed in an electrical resistance furnace, the temperature of the furnace raised to 1000° C. and held at that temperature for one hour. The current was then turned off and the pot cooled with the furnace. Ten hours were required for the furnace to cool to the temperature of the room. Between 700° C. and 400° C. the rate of cooling was approximately 1.5° per minute.

Three rings of each composition were always annealed together. One of these rings received no further heat treatment. The second ring was placed for 15 minutes in a furnace held at 600° C., then removed and cooled rapidly on a copper plate. In some cases, the third ring was heated 24 hours at 425° C.

In the discussions and in the figures and tables, the rings which received the first heat treatment only are referred to as "annealed," those reheated to 600° C. and rapidly cooled as "air quenched," and those held for a long time at 425° C. as "baked."

Magnetic Measurements

Permeabilities at low magnetizing forces were measured on unwound rings with an inductance bridge, and an a.c. permeameter.[3] From these measurements initial permeabilities were computed. For elevated temperature, measurements were made with a similar permeameter provided with a furnace compartment.[4] The bridge was also used for measuring permeabilities to small a.c. magnetizing forces when d.c. forces are superposed on the magnetic circuit. For these

[3] G. A. Kelsall, *J. O. S. A.* and *R. S. I.*, **8**, pp. 329–338, 1924.
[4] G. A. Kelsall, *J. O. S. A.* and *R. S. I.*, **8**, pp. 669–674, 1924.

measurements the rings were wound with insulated copper wire after being placed in thin annular wooden boxes to protect them from strain. Hysteresis losses were computed for a few alloys from effective resistance measurements at low flux densities, made on wound rings.

Magnetization and permeability curves and hysteresis loops were plotted from ballistic galvanometer measurements. Galvanometer measurements also were made on a few alloy rods 11 in. long and 1/8 in. diameter, at a magnetizing force of 1500 gauss. For these measurements the rods were placed in a long solenoid and the induction measured by means of an exploring coil at the center of the rod.

Properties of the 45 Per Cent Ni, 25 Per Cent Co, 30 Per Cent Fe, Composition

Measurements for the composition 45 per cent nickel, 25 per cent cobalt and 30 per cent iron in the annealed condition are plotted in Figs. 1–7 and tabulated in Table I. The curves in Fig. 1 illustrate the permeability characteristics for this composition (No. 858–1) and for a sample of annealed Armco iron. For magnetizing forces below 1.7 gauss, the permeability is substantially constant, the variation being less than 1 per cent. This constancy is remarkable for a magnetic material having an initial permeability nearly double that of iron. Within the same range of field strengths the permeability of the iron sample rises from an initial value of 250 through a maximum of 7,000 at a magnetizing force of 1.3 gauss and decreases to 6,300.

TABLE I.

Hysteresis Loss with Different Heat Treatments for [45% Ni—25% Co—30% Fe] Composition Perminvar

Heat Treatment	B	Ergs per Cm.^{3} per Cycle
Air Quenched	568	18.7
	722	32
	993	57
	1,503	119
	5,010	850
	14,810	2,500
Baked at 425° C. for 24 Hours	600	0
	795	0
	1,003	15.27
	1,604	163
	4,950	1,736
	13,810	4,430
Annealed	570	0
	820	9.54
	974	15.65
	1,508	93.20
	5,050	1,185
	8,480	2,500
	14,900	3,375

Another property of the perminvar alloy closely related to the constancy of permeability is the extremely small hysteresis loss in the range of magnetizing forces and flux densities in which the permeability is constant. This is illustrated in Fig. 2 where line *a* represents the plot of the upper half of the hysteresis loop for a maximum flux

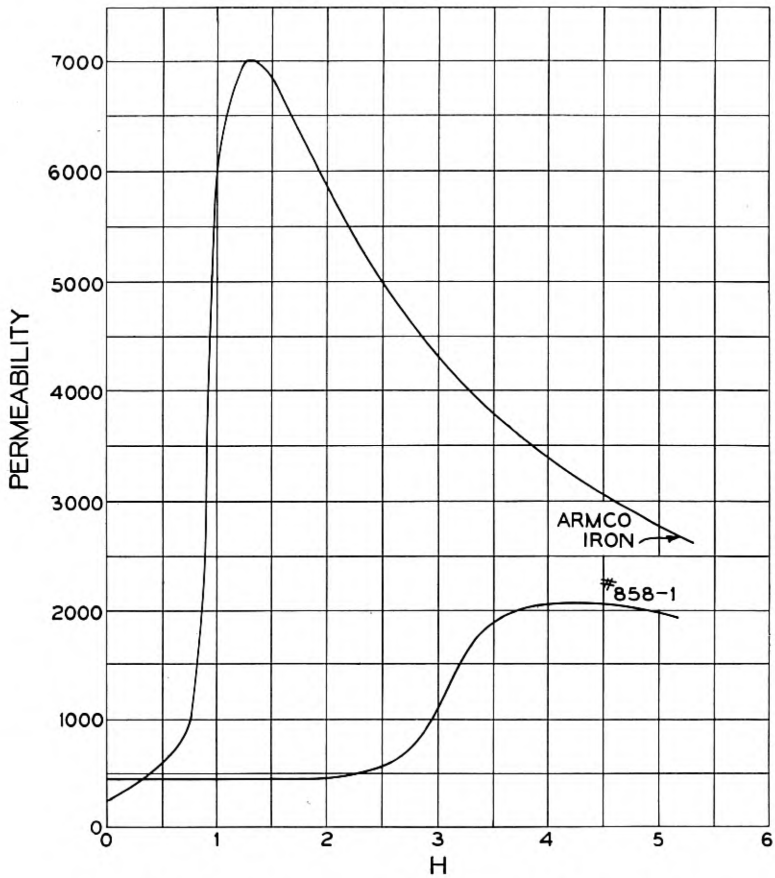


FIG. 1—Permeability curves for Armco iron and Perminvar (45% Ni—25% Co—30% Fe)

density of approximately 600 gauss. Curves *b* and *c* are similar plots for silicon steel ($3\frac{1}{2}$ per cent silicon) and Armco iron respectively. While the hysteresis loops for Armco iron and silicon steel have considerable areas amounting to 33 and 14 ergs per cubic centimeter per cycle respectively, there is no measurable area for the perminvar alloy. Although the ballistic method of measurements which was used in obtaining these curves, does not indicate very small losses readily it is evident that the losses in the perminvar alloy are of a different order of magnitude from those of the other two materials. In order to obtain additional information in regard to the hysteresis loss of this alloy at low flux densities, the sample was measured by the inductance bridge method. It was found that the hysteresis loss at a flux density of 100 gauss was $.024 \times 10^{-3}$ ergs per cubic centimeter per cycle. The best material in this regard previously known was permalloy, for which a sample containing approximately $78\frac{1}{2}$ per cent nickel, measured under similar conditions, had a hysteresis loss of 33×10^{-3} ergs per cubic centimeter per cycle.

The growth of the hysteresis loss and the appearance of measurable areas, and the peculiar shapes of the loops for this composition as the flux densities increase are illustrated in Fig. 3. The curve for a

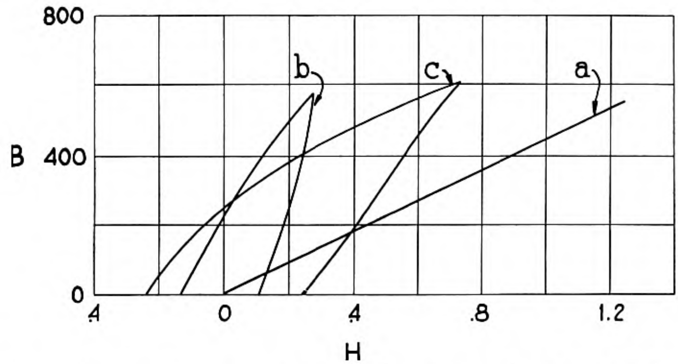


FIG. 2—Hysteresis characteristics: *a*—Perminvar (45% Ni—25% Co—30% Fe); *b*—silicon steel; *c*—Armco iron

maximum flux density of 580 gauss in Fig. 3, is from the same data as Curve *a* in Fig. 2. The circles in this plot indicate points on the ascending branch, and the dots, points on the descending branch. The hysteresis loop broadens out so that it has a measurable area when the maximum flux density is increased to 800 gauss. The existence of a close relation between the hysteresis losses and the constancy of permeability is quite apparent from the permeability curve in Fig. 1 and the curves in Fig. 3. While the permeability remains constant there is practically no hysteresis loss but as it begins to change this loss appears and increases quite rapidly with increase in permeability. The increase in the energy loss and the changes in the shapes of the loops as the flux density increases also are illustrated by these curves. The most striking hysteresis characteristic of these

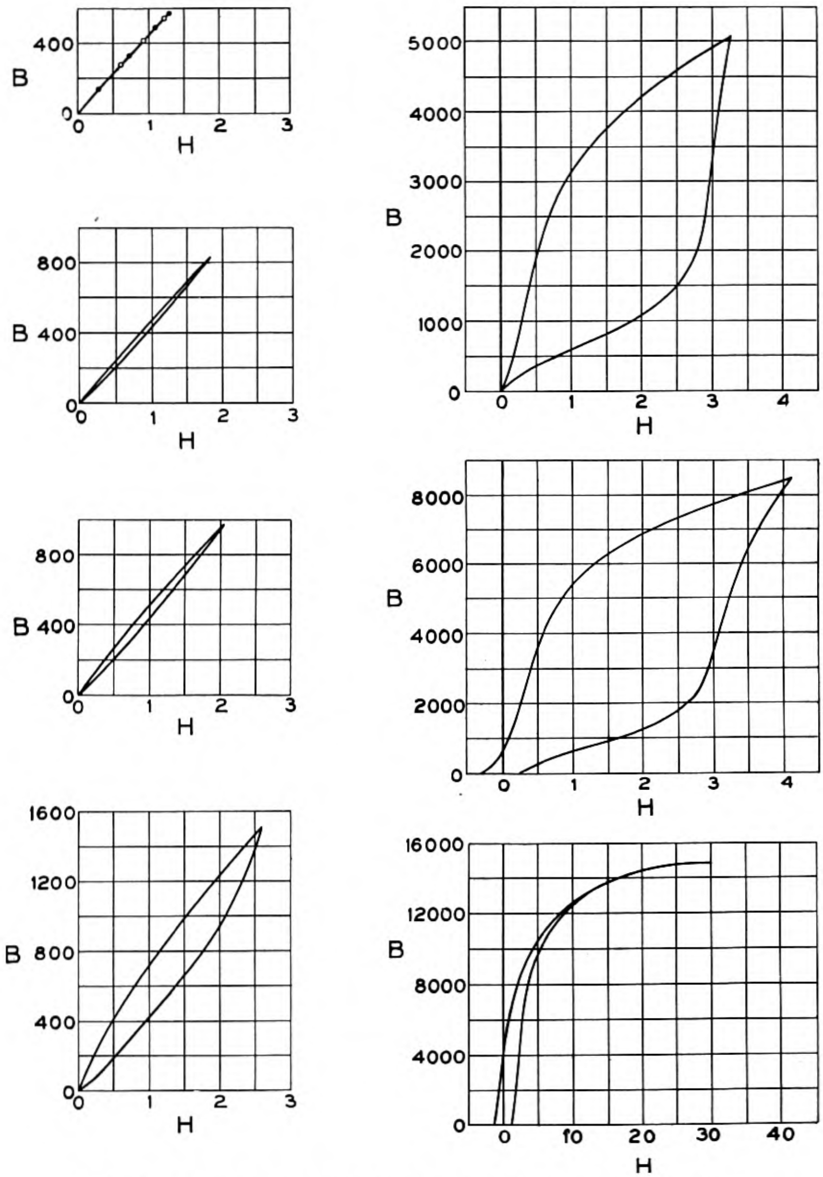


FIG. 3—Upper halves of hysteresis loops of Perminvar (45% Ni—25% Co—30% Fe) annealed

loops is the absence of coercivity. For loops having maximum flux densities of 5,000 or less the ascending and descending branches pass through the origin. For greater flux densities the coercivity begins to be measurable, but there is still a considerable constriction of the loop for a maximum flux density of 8,000 gauss. It is only in the loop for 15,000 gauss, that the constriction at the origin has disappeared and the loop resembles those for ordinary magnetic materials. In Table I the hysteresis losses for the complete loops are tabulated.

Fig. 4 illustrates graphically how the permeability measured with a constant alternating current magnetizing force of about .0021 gauss and 200 cycles per second is affected when a steady magnetizing force

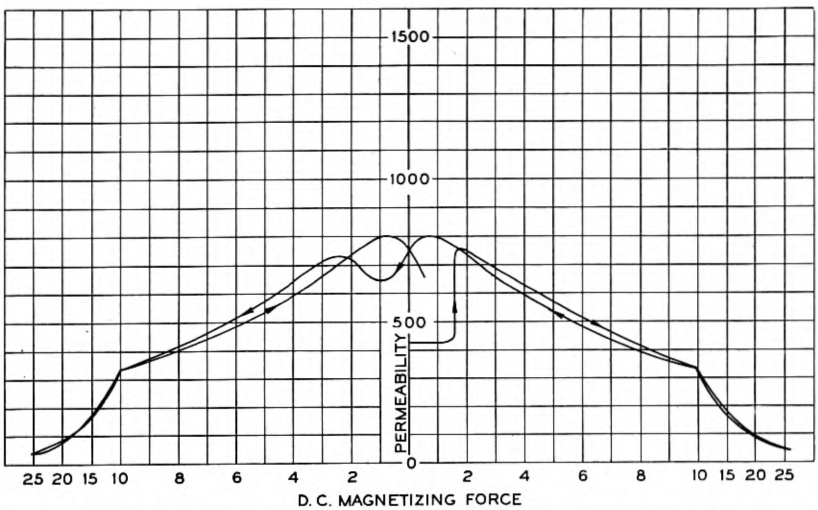


Fig. 4—The effect of superposed d.c. fields on the a.c. permeability of Perminvars (45% Ni—25% Co—30% Fe)

is superposed on the magnetic circuit, the steady force being produced by a direct current. The arrows in the figure indicate the direction in the progress of the permeability as the direct current magnetizing force is varied. The permeability is substantially constant as the direct current magnetizing force increases up to approximately 1.7 gauss and it then suddenly rises as the force is increased beyond that value. This is the same field strength at which the permeability begins to increase as shown in Fig. 1.

Another characteristic of this material not found in ordinary magnetic substances also is shown in Fig. 4. After an applied d.c. magnetizing force of 25 gauss is removed, the permeability has risen from 460 to 750. With ordinary materials, after such magnetization, the permeability is reduced, in some cases from 40 to 60 per cent. With the increase in permeability its constancy disappears both for the superposed condition as shown in Fig. 4, and for ordinary magnetizations at low field strengths. The hysteresis losses are also increased for corresponding flux densities. These changes in the

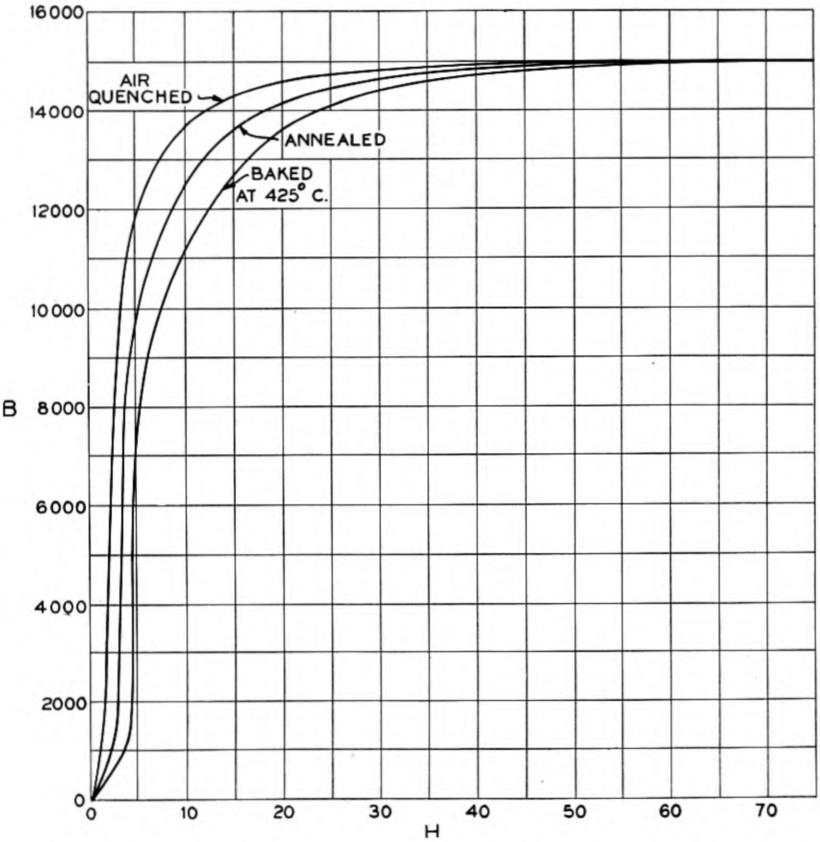


FIG. 5—Magnetization curves for Perminvar (45% Ni—25% Co—30% Fe)

magnetic properties are largely removed by demagnetization. The ordinary method of demagnetization by reversals of a slowly decreasing magnetizing force has been less successful than it is with iron in returning the material to its initial state. Addition of an a.c. force superposed on the d.c. helps materially in restoring the original magnetic properties.

Effects of Heat Treatment

The manner in which the magnetic properties of this composition are affected by the rate of cooling is illustrated in Figs. 5–7. The measurements plotted in these figures are from three rings, air quenched, annealed and baked, respectively. For weak fields there are large differences in the magnetic properties for these rings. The initial permeability for the quenched ring is more than twice that of the baked one. With increased field strength this difference decreases and disappears for fields over 50 gauss. The permeability variations as the strength of the field increases also show the remarkable change

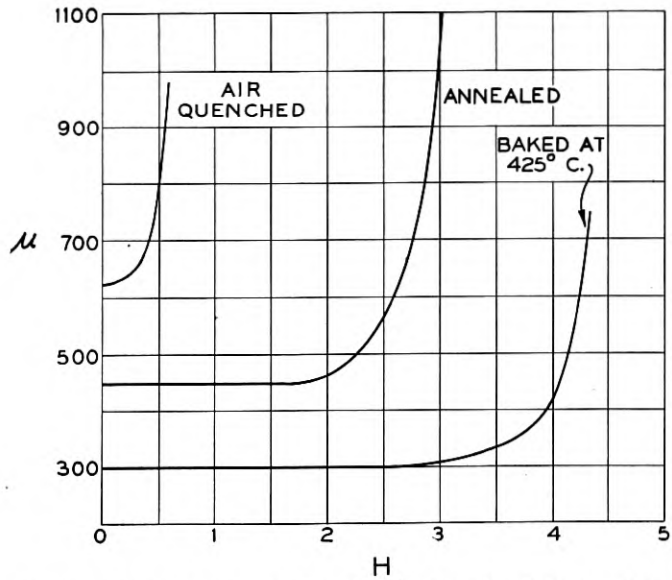


Fig. 6—Permeability curves for Perminvar (45% Ni—25% Co—30% Fe)

which heating for a long time in the critical temperature range produced in these alloys. For the quenched ring the permeability increased from 620 to 800 for an increase of field strength from 0 to .5 gauss. For the baked ring the permeability remains constant for fields up to 2.5 gauss.

The hysteresis loss and the shapes of the loops also are affected greatly by the heat treatment. This is illustrated in Fig. 7, where loops for a number of flux densities are plotted for two sample rings, one baked at 425° C. and the other air quenched, and in Fig. 3 where loops for the same maximum flux densities are plotted for an annealed ring. The energy losses integrated for complete loops are tabulated in Table I.

These curves show that the rate of cooling determines the magni-

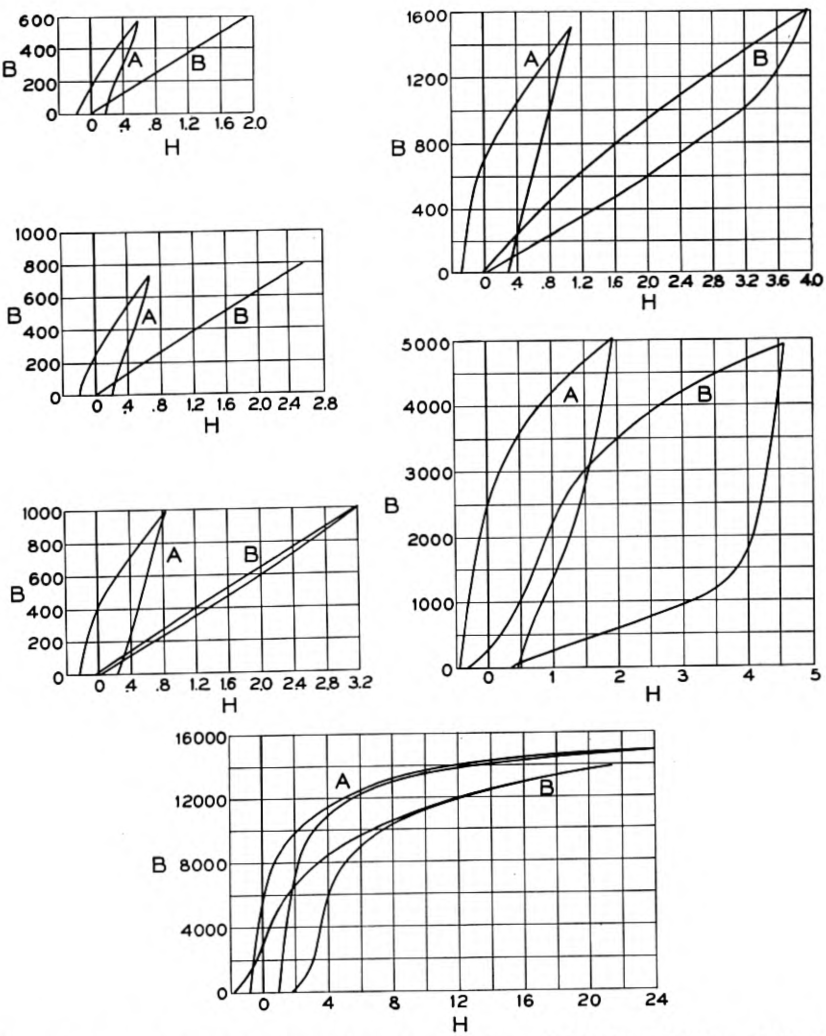


FIG. 7—Upper halves of hysteresis loops for Perminvar (45% Ni—25% Co—30% Fe) *A*—air quenched, *B*—baked at 425° C.

tudes of the hysteresis losses and the shapes of the hysteresis loops. For the air quenched rings the shapes of the loops for the different flux densities resemble those for ordinary magnetic materials although a few of them show traces of the perminvar characteristics. If a

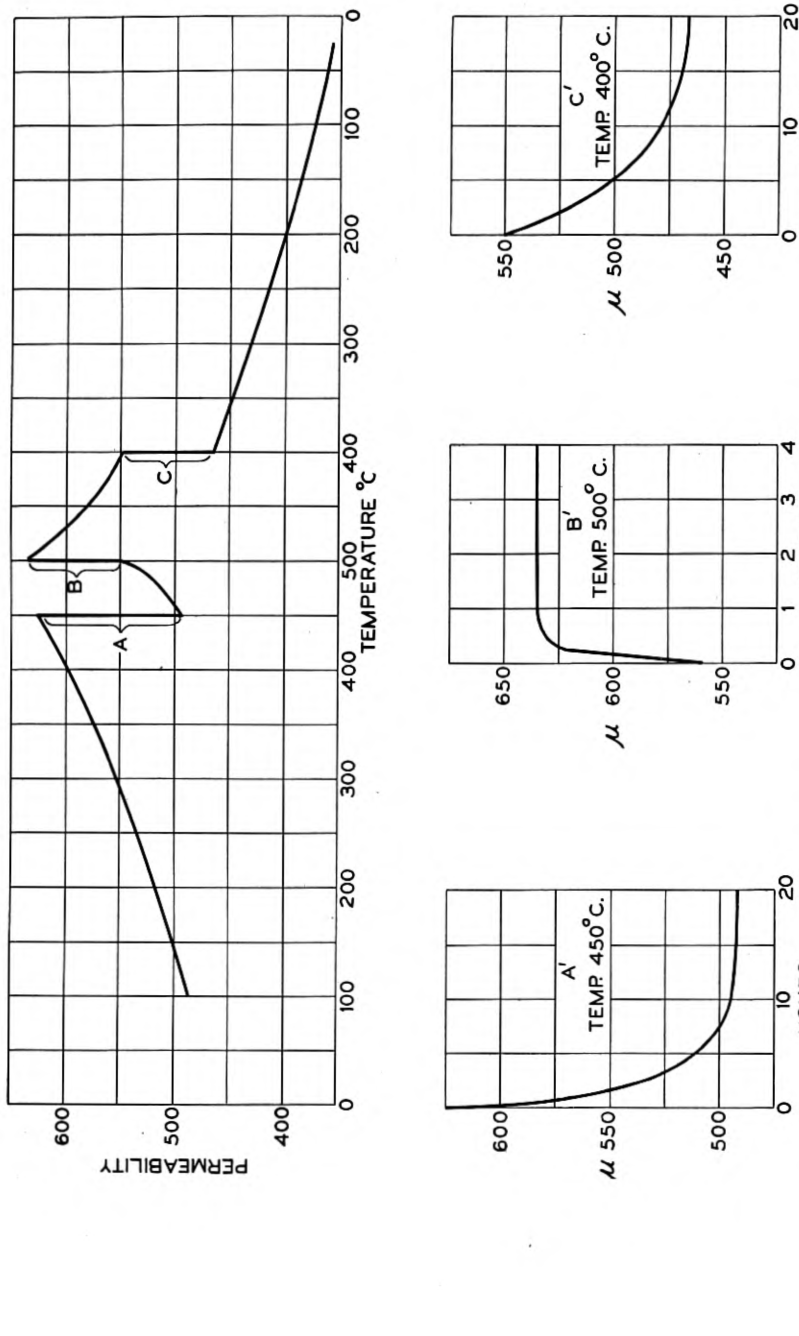


Fig. 8—Permeability—Temperature curve for Perminvar (45% Ni—25% Co—30% Fe)

more rapid cooling rate had been used these characteristics no doubt would have disappeared completely. The hysteresis loops all have considerable areas. The one for 568 gauss, the lowest flux density measured, represents an energy loss of 18.7 ergs per centimeter cube. For the same flux density, the hysteresis loops for the annealed and the baked rings have no measurable areas, the ascending and descending branches of the loops falling on the straight lines shown in

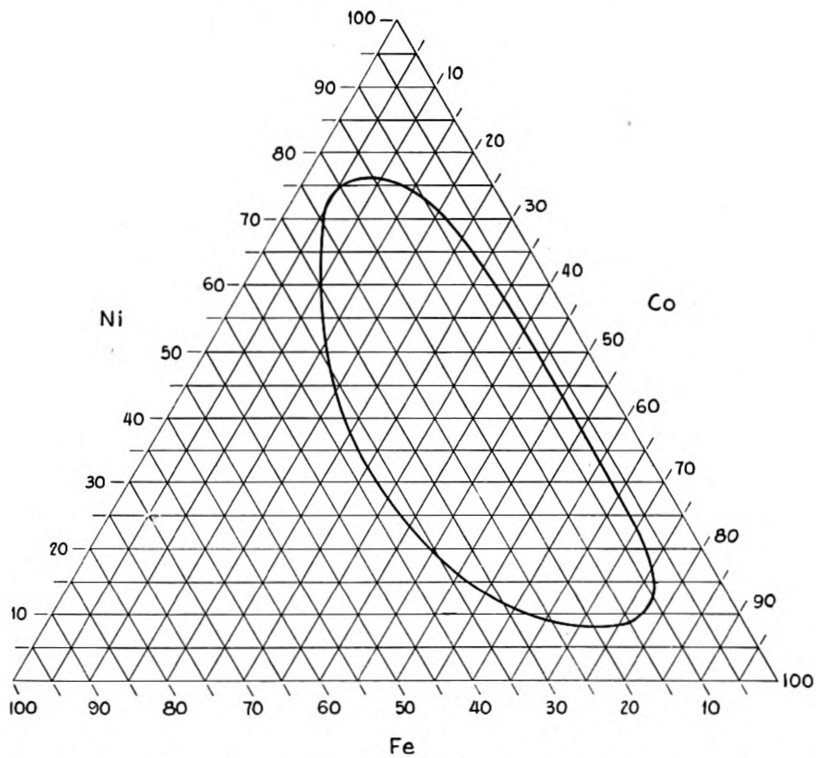


FIG. 9—Composition diagram Ni—Fe—Co Series. Area enclosed by the curve shows compositions with marked Perminvar characteristics

the figure. This absence of measurable area extends up to a flux density of nearly 1,000 gauss for the baked ring. The increase of energy loss as the flux density increases above this value, however, is considerably more rapid than for the quenched ring. At 1,500 gauss, the hysteresis loss is a little greater than for the quenched ring and when the flux density is increased to 5,000 gauss, the loss is more than double.

It was shown above that the degree to which perminvar characteristics are developed depends on the rate of cooling through the critical temperature range, and that baking at 425° C. gave the most characteristic results. The manner in which the permeability of the 45 per cent nickel, 25 per cent cobalt and 30 per cent iron composition changes in this temperature range is illustrated in Fig. 8. The temperature of an annealed ring was increased from that of the room to 450° C. where it was held constant for twenty hours. It was then raised to 500° C. and held for four hours, then lowered to 400° C. where it was held for twenty hours, and finally cooled to room temperature. Permeability measurements were made at these temperatures with an a.c. magnetizing force of .02 gauss.

Inspection of these curves shows that in the range 400°–500° C., the permeability lags behind the temperature, and that the time required for the permeability to reach a constant value increases very rapidly below 450° C. The changes in final permeabilities with temperature decrease also rapidly below 450° C. In fact, when the difference in permeability caused by the temperature coefficient is corrected for, the permeability of the alloy after heating at 400° C. is not very different from what it is after heating at 450° C. Other experiments show that the critical temperature range extends below 400° C., but, as would be expected, the decrease in permeability is very small. The range also extends above 500° C. for this alloy and some experiments indicate that the upper limit is the magnetic transformation temperature which for this alloy is 725° C.

Effects of Variation of Compositions

The composition range within which the magnetic properties characteristic of perminvar are developed pronouncedly by annealing, is represented by the area enclosed by the curve in the triangular composition diagram Fig. 9. Magnetic properties for a few of the compositions in this area are plotted in Figs. 10 and 11. Table 2 gives their chemical analyses, initial permeability (μ_0), the maximum permeability ($\mu_{max.}$), the magnetizing forces (H) and the flux densities (B) in gauss to which the alloys may be brought with a permeability variation not over 1 per cent, also the $(B - H)$ values for a magnetizing force of 1,500 gauss for some of the alloys, and the resistivity in microhms-cm. The hysteresis losses for a number of flux densities are given in Table 3.

The area enclosed in Fig. 9 shows that approximately one third of the alloys in the Ni-Fe-Co series show some of the characteristic perminvar properties in the annealed condition. The proportions of nickel and cobalt may be varied through a wide range. A great deal

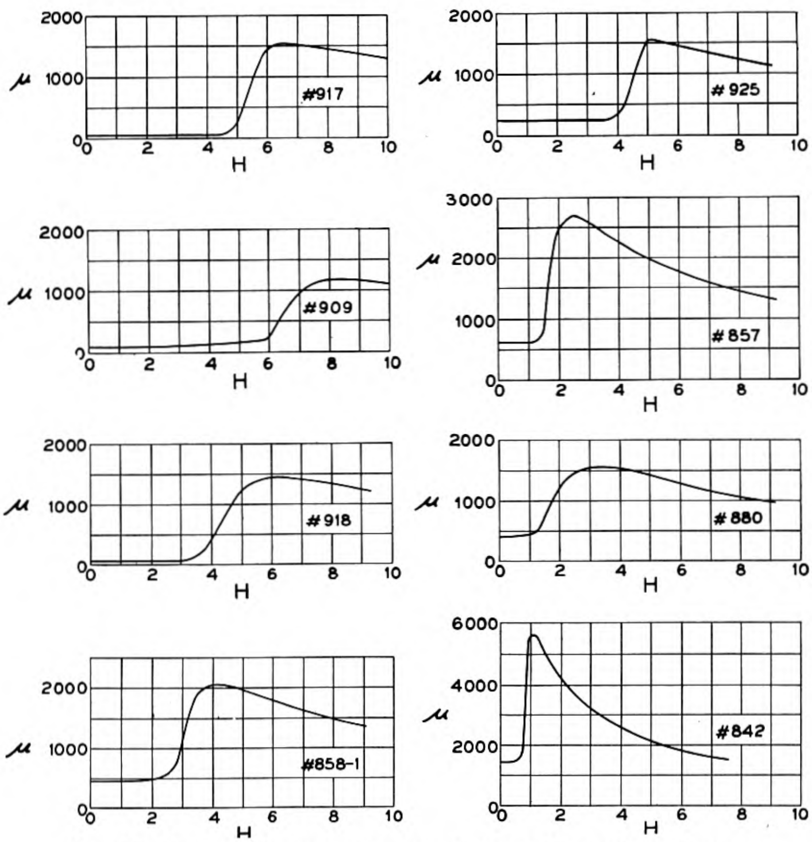


FIG. 10—Permeability curves for several compositions of Perminvar. Chemical analysis given in Table II

TABLE II

CHEMICAL COMPOSITION AND MAGNETIC PROPERTIES OF PERMINVARS

Casting Number	Chem. Analysis				Magnetic Properties					Resistivity
	Ni	Co	Fe	Mn	μ_0	μ_m	H for 1% Ch. in μ	B for 1% Ch. in μ	$B - H$ for H = 1,500	Microhm-Cm.
917	11.35	68.10	20.36	.35	57	1,545	4.2	242	18,400	15.38
909	20.85	49.18	29.74	.31	98	1,180	4.0	396	18,200	16.59
918	20.73	68.35	10.58	.39	51	1,447	2.5	129	17,400	12.35
858–1	45.12	23.83	30.69	.46	449	2,075	1.75	793	15,600	18.63
925	50.47	29.28	20.15	.33	231	1,555	3.2	746	14,600	14.55
857	59.66	14.76	24.97	.60	631	2,680	1.15	733		17.5
880	70.29	15.23	14.57	Tr.	390	1,570	.55	216		14.13
842	73.29	5.97	20.7	.20	1,430	5,600	.55	795		15.56

less variation in the iron content is permissible, being less than one half of the amounts of the other two constituents. The manner in which each of the metals affects the magnetic properties is not very clearly indicated by the numerical values in the table. Iron and

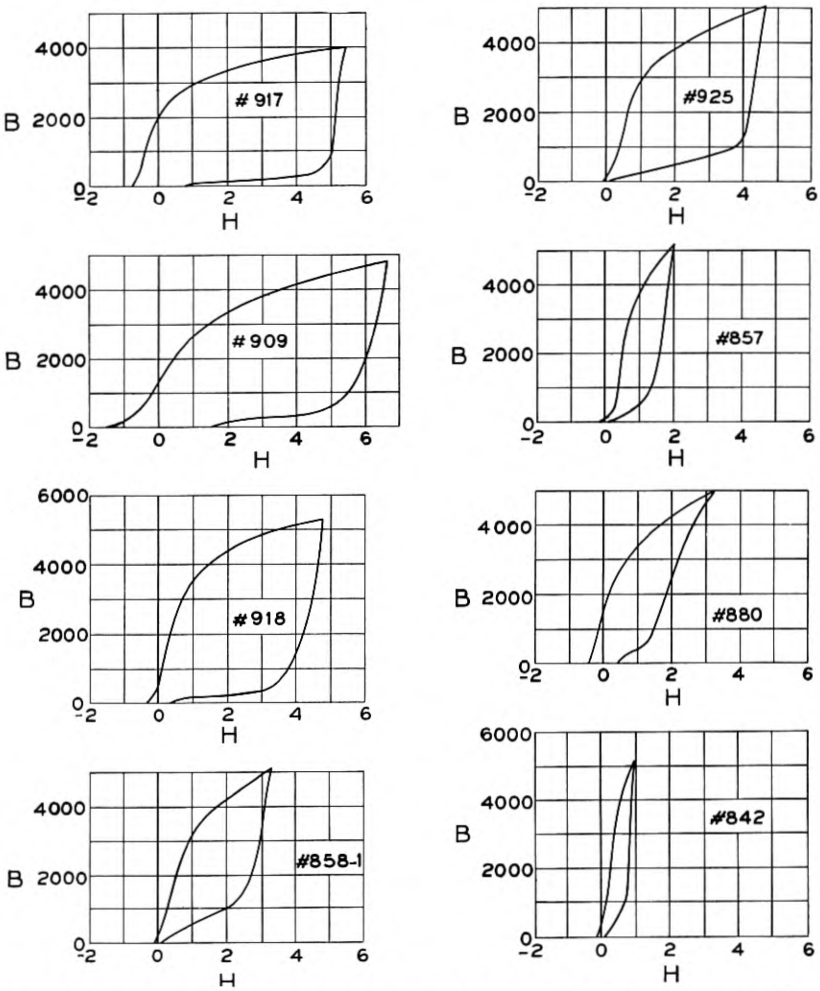


FIG. 11—Upper halves of hysteresis loops for several Perminvar compositions. Chemical analysis is given in Table II

cobalt appear to increase the constancy of the permeability but decrease the initial permeability values. Nickel increases the initial permeability but large percentages decrease the constancy. On the whole, in the alloys with high nickel content, the combination of high

TABLE III

HYSTERESIS LOSS

Casting Number	B	Ergs per Cm.^{3} per Cycle	B	Ergs per Cm.^{3} per Cycle	B	Ergs per Cm.^{3} per Cycle	B	Ergs per Cm.^{3} per Cycle	B	Ergs per Cm.^{3} per Cycle
917	280	8	625	150	1,700	1,040	3,950	2,740	15,500	14,160
909	—	—	620	36	1,030	299	4,800	3,330	15,200	12,460
918	—	—	750	247	—	—	5,270	2,605	—	—
858–1	566	0	827	9.5	1,508	93	5,050 8,480	1,185 2,500	14,900	3,375
925	560	0	—	—	2,000	468	5,000	2,020	13,200	4,965
857	145	0	840	8.4	1,520	88	5,200 8,250	632 1,240	13,250	1,508
880	320	0	—	—	1,470	116	5,000	1,012	10,300	2,435
842	500	0	—	—	1,530	23	5,100	348	11,500	783

permeability and fair constancy makes for a larger range of flux densities in which the permeability is constant and consequently also increases the range of flux densities with low hysteresis loss.

Experiments on several alloys of this series indicated that by baking the alloys at 425° C. the area enclosed by the curve in Fig. 8 would be increased considerably, and possibly would include some of the binaries of these metals.

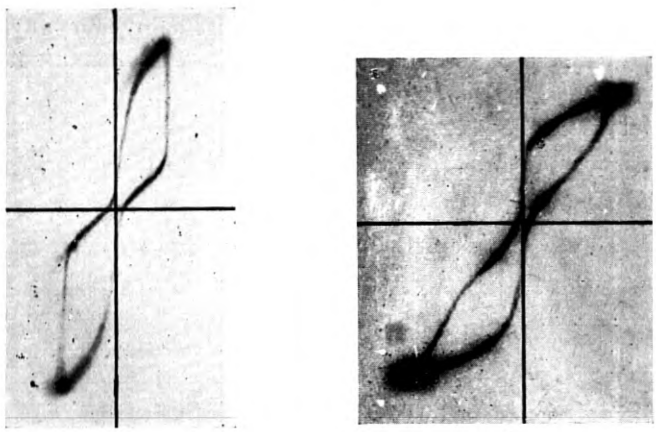
Discussion

While this paper is concerned primarily with the study of the magnetic properties of these alloys and the dependence of these properties on composition and on heat treatment, some of the results are of considerable theoretical interest, as they suggest the manner in which the unusual magnetic properties are acquired by the alloys. It was shown in the heat treating experiments that the unusual magnetic properties resulted from suitable heat treatment of certain compositions. Slow cooling through a rather narrow temperature range, or continuous heating for a long time at the lower end of this range resulted in alloys which had marked perminvar characteristics. Rapid cooling through this temperature range usually did not develop these characteristics. From the measurements at elevated temperatures, Fig. 8, it was shown that in the temperature range from 400° C. to 500° C., the change in the alloys is quite rapid at the higher temperature, but that the rate of stabilization slows up as the temperature decreases. When the alloy is heated and cooled through a temperature cycle in this manner, the permeability changes progressively and at each temperature in the cycle the alloy reaches a stable condition if the rate of cooling or heating is slow enough. There is a striking similarity in the manner in which these changes in permeability are developed, and in the progress of the constitutional changes in an alloy which at high temperatures is a homogeneous solid solution, but as the temperature falls becomes saturated and segregates into a mixture of two solid solutions of different concentration.

That such a segregation takes place in the slowly cooled alloys is also supported by a study of the differences in the shapes of the hysteresis loops of the quenched and the slowly cooled alloys. Ordinarily, the widest part of a hysteresis loop of a homogeneous material is the intercept on the H axis. All the loops of the air-quenched alloys have these characteristics. Gumlich [5] has shown that if a magnetic circuit is made of two materials of different magnetic properties, the loops may assume a variety of shapes, ranging from

[5] E. Gumlich, *Arch. f. Elektrotechnik*, Vol. 9, p. 153, 1920.

that of the homogeneous material to one in which two branches converge at the origin into a single line. This constriction of the hysteresis loop is also illustrated for a parallel bimetallic magnetic circuit in Fig. 12 where loops *a* and *b* are traced for a perminvar core and a bimetallic rod, respectively. The rod was 15 in. long and consisted of a core of .04 in. diameter unannealed piano wire and a .006 in. wall permalloy tube, heat treated to give high permeability, and fitting closely to the wire. Though the magnetic circuit condition for the perminvar core is not the same as for the bi-metallic rod, the similarity of the two loops is marked and supports the theory that the constricted loop of the perminvar core is caused by segregation in the alloy.



a *b*

FIG. 12—Hysteresis loops: *a*, Perminvar; *b*, Bi-metallic rod. Loops traced with a cathode ray oscillograph

The electrical resistance also is affected by the slow cooling. An air-quenched alloy of the 45 per cent nickel, 25 per cent cobalt and 30 per cent iron composition was 10 per cent lower in resistivity after it had been baked at 425° C. This change is also in line with the idea that segregation takes place when the perminvar alloys are cooled slowly.

While these considerations point to a satisfactory explanation for the constriction of the hysteresis loops they do not explain the extremely low hysteresis losses of the alloys at low flux densities. This characteristic of perminvar suggests that one of the constituents which is segregated by the heat treatment is itself a material of much lower hysteresis loss than any previously known material and that the other constituent suffers relatively little change of magnetization at low magnetizing forces.

To the engineer these alloys are of unusual interest. They may be used to advantage for magnetic structures where the magnetizing forces do not exceed the limits of constancy of permeability for the various compositions. Interesting results have been obtained with the 45 per cent nickel, 25 per cent cobalt and 30 per cent iron composition for continuous loading of telephone conductors and for cores of loading and filter coils used in high quality transmission and in carrier current circuits. For such purposes high resistivity is also desired, and it has been found that the addition of a few per cent of other metals such as molybdenum serves for this purpose. For circuits requiring greater constancy or higher permeability other compositions are more suitable. The best alloy for any specific circumstance may be selected from a study of the magnetic properties of the various compositions.

The Aluminum Electrolytic Condenser [1]

By H. O. SIEGMUND

SYNOPSIS: In this paper the anodic film-forming properties of aluminum are discussed and the unique electrical qualities of film-coated aluminum anodes are described. Special reference is made to an aluminum electrolytic condenser of the type used in low pass electric wave-filters of direct-current telephone power plant equipment. Electrical characteristics of condensers are given and the manner is described in which the operation and life of the units are influenced by variations in composition of the electrodes and the electrolyte.

INTRODUCTION

SINCE the discovery about 75 years ago [2] of the unusual polarizing effect of aluminum it has become well known that certain metals, notably aluminum and tantalum, as anodes in a suitable electrolyte become coated with a film having remarkable electrical properties. Films formed in this manner are characterized by the influence of impressed potential on their electrical resistance.

A representative relationship between applied voltage and resistance per 1,000 sq. cm. of film on an aluminum anode in an ammonium borate electrolyte [3] is shown in Fig. 1. This resistance characteristic imparts to the film the capability of conducting current more freely in one direction than in the other; of breaking down as an insulation between the metallic electrode and the solution when voltages above a critical value are applied; and in combination with the thinness of the film of holding a substantial charge of electricity at potentials below the breakdown voltage.

Each of these characteristics provides the principle around which a distinctive class of electrical apparatus has been developed. The electrolytic rectifier, widely used in small direct-current supply sets for battery charging and radio purposes, employs the unidirectional conducting characteristic. The aluminum electrolytic lightning arrester, used extensively for protection of direct-current railway equipment, depends for its operation upon the breakdown characteristic of the film. And finally the aluminum electrolytic condenser, now being

[1] Presented before the American Electrochemical Society, at Bridgeport, Conn., April 26, 1928.
[2] Wheatstone, *Phil. Mag.*, **10**, 143 (1855).
[3] The exact values of resistance are somewhat unstable, and depend on the time between readings and whether successively increasing or decreasing values of potential are applied. However, the general shape of the curve and the magnitude of the values are representative.

used in direct-current telephone power plant equipment [4] utilizes the dielectric property of the film to provide electrostatic capacity.

Electrical Qualities of Films on Aluminum

There are a number of electrolytes, including various concentrations of phosphates, borates, tartrates, carbonates and others in which films can be formed on aluminum to withstand maximum potentials

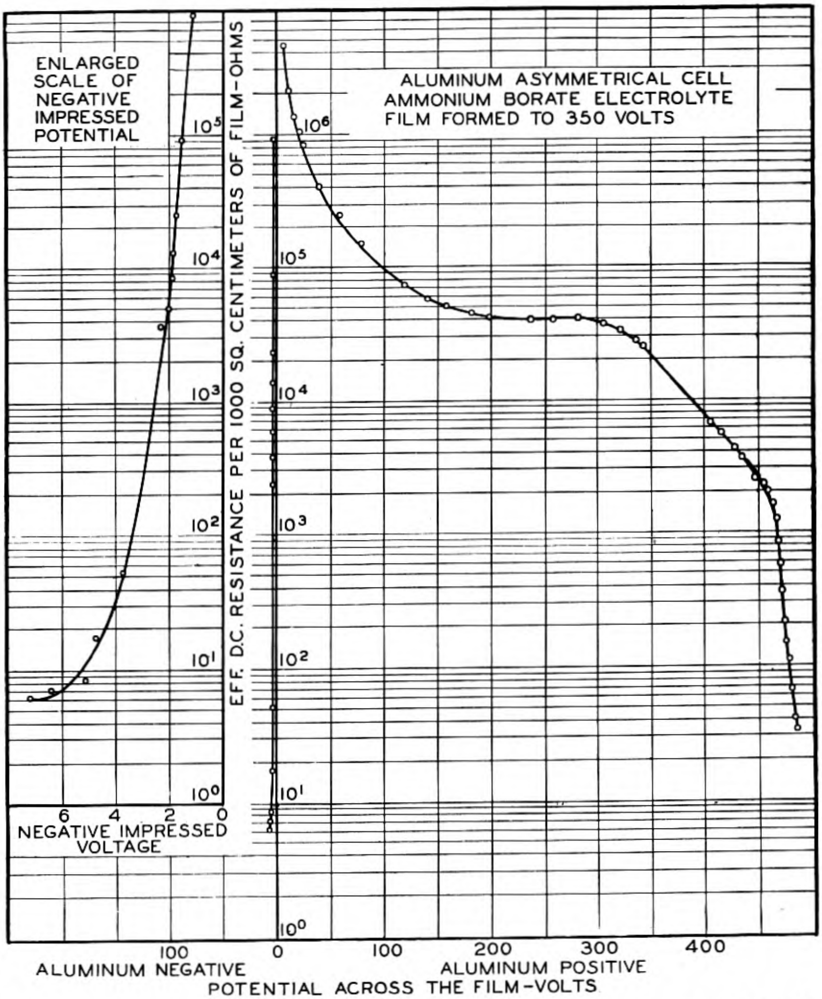


Fig. 1—Influence of impressed potential upon the electrical resistance of the film on an aluminum electrode

[4] Young, R. L., *Bell System Techn. J.*, 6, 708 (1927).

upwards of 300 volts, at least for limited periods. If a film is formed on a piece of aluminum to this maximum voltage and the metal is then made the anode in an electrolytic cell across which variable potential can be applied, a current corresponding to a density of less than a microamp. per sq. cm. of filmed surface will flow when a potential of one tenth the maximum voltage is impressed.

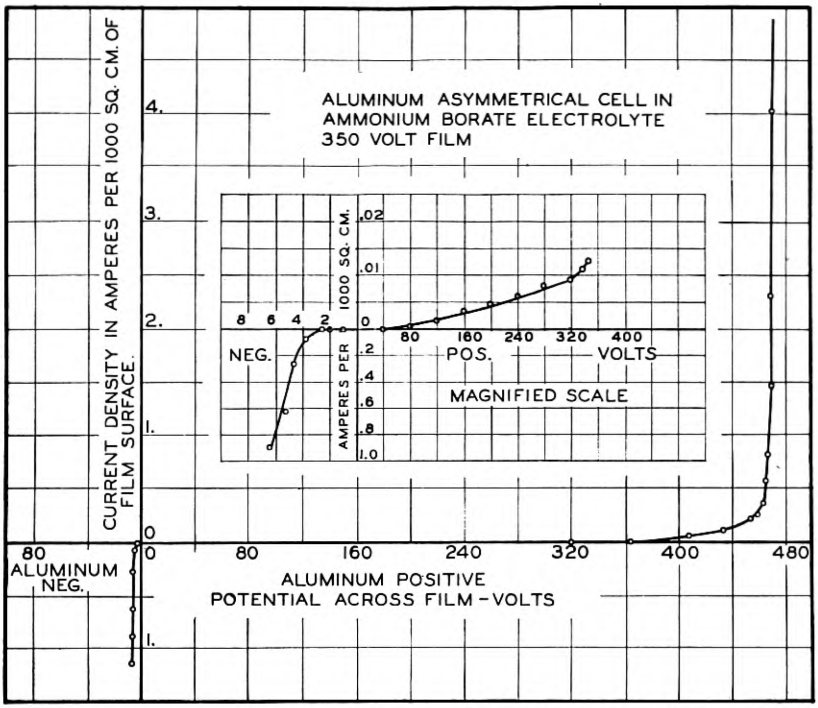


FIG. 2—Influence of potential on the current through an electrolytic cell with a film-coated aluminum electrode

As the potential is increased this "leakage" current will increase at a rate somewhat greater than proportionate to the voltage. As the maximum or breakdown potential is approached it will be noticed, if the room is darkened, that the anode begins to glow uniformly over the surface with a pale light and with further increases in voltage sparks begin to scintillate over the entire electrode, being noticed first at the surface of the electrolyte. The current through the cell becomes appreciable under this condition and increases more rapidly until at voltages slightly above the sparking potential the cell acts virtually as a short circuit.

Upon reduction of the voltage, however, the insulating properties of the film are restored and the current decreases with decreasing potential in substantially the same relation to voltage as before. The sparking over the surface will be observed to cease at about the same potential at which it began, the glow will disappear and the low leakage-current values will be obtained when the voltage is reduced sufficiently.

Upon reversal of potential on the aluminum electrode, however, there is a much larger flow of current, the value of which is limited by a counter voltage of several volts, and by the low internal resistance of the cell with negative potential applied. Typical current-voltage relations for an aluminum cell are shown graphically in Fig. 2. These relations correspond to the curve in Fig. 1, showing the variation of resistance with potential of a "filmed" aluminum electrode in ammonium borate electrolyte.

Capacity of Aluminum Films

Like the ordinary paper or mica static condenser, the electrolytic condenser consists of two conducting surfaces separated by an insulator. The high-resistance film constitutes the insulator in the electrolytic cell, and the electrolyte on one side of the film and the metal of the film-bearing electrode on the other provide the two conducting surfaces. The cathode in this type of cell merely provides a means for making electrical contact with the electrolyte.

When a film is formed upon a smooth polished aluminum surface the coating is transparent. If observed under favorable illuminating conditions the "filmed" surface is seen to be colored and may be either green, yellow, red or blue, depending upon the thickness of the film. This is attributed to light interference and indicates that the thickness of the film is in the order of the length of light waves. The actual thicknesses of films on aluminum have been determined to be from 0.001 to 0.00001 mm.,[5] depending upon conditions of formation.

Because of this extreme thinness of the dielectric and its high insulation resistance when positive potential is applied, unusually large capacities per unit area of surface can be obtained. The capacity of a film formed to 30 volts on aluminum is about 0.18 microfarad per sq. cm. of dielectric surface, or about 1,000 times that of paper condensers. The capacity per unit area is approximately inversely

[5] Zimmerman, *Trans. Am. Electrochem. Soc.*, **7**, 309 (1905); Sutton and Willstrop, *Engineering*, **124**, 442 (1927); Slepian, *Trans. Am. Electrochem. Soc.*, September, 1927, to be printed in Vol. 54 of the *Transactions*.

proportional to the potential at which the film is formed, indicating that the thickness of the dielectric is directly proportional to the voltage of formation.

Effect of Impressed Voltage on Capacity

When an electrolytic cell with a "formed" anode has impressed on its terminals a voltage greater than the formation voltage, the film must build up to the new potential before the electrical characteristics of the cell become stable. At this higher voltage the capacity of the cell will be reduced to correspond to the increased potential. Where large plate areas are involved the direct application of a potential above the formation voltage results in a heavy flow of current, which may overheat and damage the cell if not properly limited.

If a voltage is impressed on a condenser lower than the potential applied during the formation of the film, the cell will operate satisfactorily, but the capacity will not be immediately affected and will correspond to the potential at which the film was originally formed. This is illustrated in Table I, which shows the capacities of an electrolytic cell measured with applied voltages of different values below the voltage of formation.

However, if a condenser operates for a long time at a reduced voltage the excess film will be removed slowly by the chemical action of the electrolyte, and the capacity will increase gradually to a value depending upon the operating voltage. The rate of change of capacity under these conditions is affected by the temperature at which the cell operates, and by the conductivity of the electrolyte. As is illustrated by the curves shown in Figs. 3 and 4 the change becomes more rapid when these factors are increased.

TABLE I

Applied Potential Volts—D.C.	Capacity Readings Microfarads
49.2	1008
43.0	1002
36.7	1008
30.4	1013
24.3	1013
18.3	1013

Series Connected Condensers for Direct and Alternating Current Service

In the discussion of the current-voltage relations in Fig. 1, it was noted that an aluminum cell with one electrode of non-film-forming metal conducts current freely when the aluminum is cathode. Accordingly this type of cell is capable of holding a charge of electricity and serving as a condenser only while the aluminum is at a higher positive potential than the electrolyte.

A cell with a non-film-forming cathode makes a suitable condenser to operate on direct-current or pulsating-current circuits, in

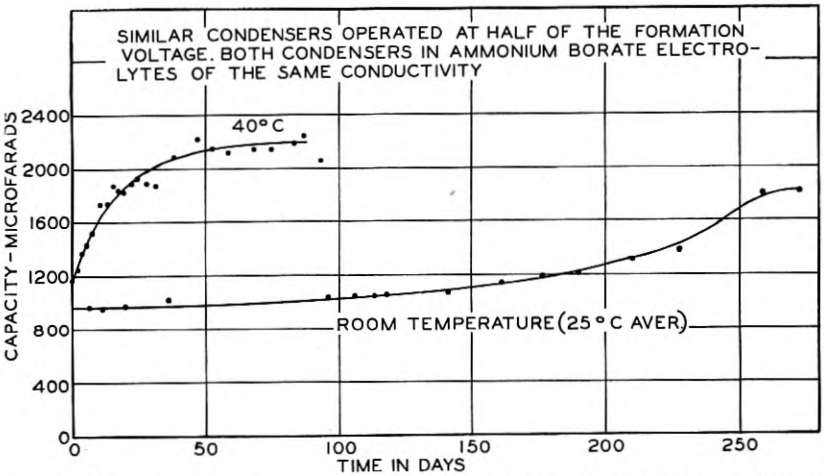


Fig. 3—Effect of electrolyte temperature on the rate of capacity change in condensers operating at voltages below the formation voltage

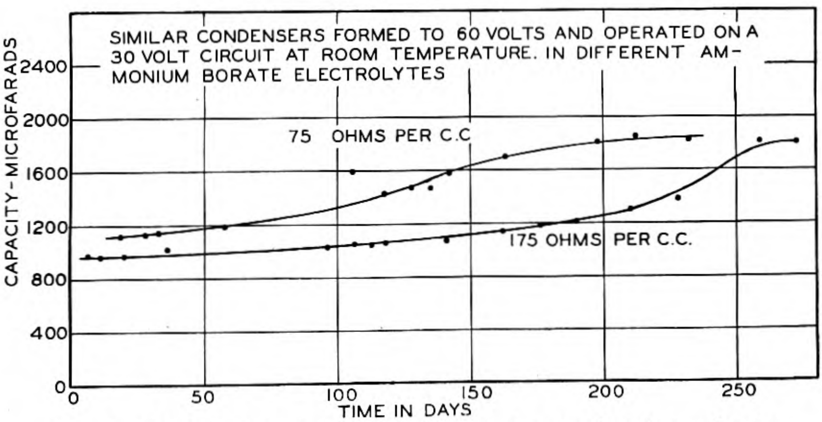


Fig. 4—Effect of electrolyte conductivity on the capacity change due to reduced operating voltage

which the aluminum always remains positively charged. On alternating-current circuits, however, such a cell will operate as a rectifier rather than as a condenser, unless two similar units are connected in a series-opposed relationship.

A suitable condenser for operation on alternating current can also be made by having two electrodes of film-forming metal in the same solution, the electrical relations between the "filmed" electrodes being the same in this case as in the series-opposed arrangement of two asymmetrical cells. In either case one or the other of the film-forming electrodes opposes the flow of current during each half cycle.

If we consider the conditions that exist at the instant of maximum potential in one direction, the electrode that is then anode acting as a condenser, receives its maximum charge. As soon as the potential begins to decrease from this maximum the accumulated electricity begins to flow from the charged plate through the circuit, but in so doing the opposing film-forming electrode becomes an anode, enabling it to hold the charge given up by the discharging electrode.

In this way, as the alternating potential varies between maximum values in each direction, the charge is transferred from the capacity provided by one film-forming electrode to the other, the sum of the charges on these two electrodes at every instant remaining constant. It can be shown that two series-opposed asymmetrical cells of capacities C_1 and C_2, or two "formed" electrodes of these capacities in the same electrolyte, have a resulting capacity equal to

$$\frac{C_1C_2}{C_1 + C_2}.$$ [6]

As illustrated in Table II, which gives the results of measurements on two asymmetrical cells formed to different voltages and connected in various series combinations, it will be noted that this relationship

TABLE II

MEASUREMENTS AT 60 CYCLES

	Biasing Potential Volts D.C.	Capacity Microfarads	Equivalent Series Res. Ohms
Condenser "A"	24.6	1529	0.165
Condenser "B"	44	962	0.23
"A" and "B" series—aiding	50	587	0.385
	43.8	582	0.38
	25	588	0.38
"A" and "B" series—opposing			
"B" positive	37.3	588	0.395
	18.3	588	0.38
No bias	0	594	0.385
"A" positive	12	588	0.39
	24	588	0.39
Calculated from measurements of "A" and "B"		589	0.395

[6] Zimmerman, *Trans. Am. Electrochem. Soc.*, **7**, 323 (1905).

is true both on alternating and direct-current circuits. If unidirectional potential is applied this relation holds without regard to the polarity or the magnitude of the impressed voltage so long as this voltage does not make the potential across the film opposing the flow of direct current greater than the voltage to which this film was formed.

In the case of asymmetrical cells connected in series-opposed relation the expression is correct either when the positive plates

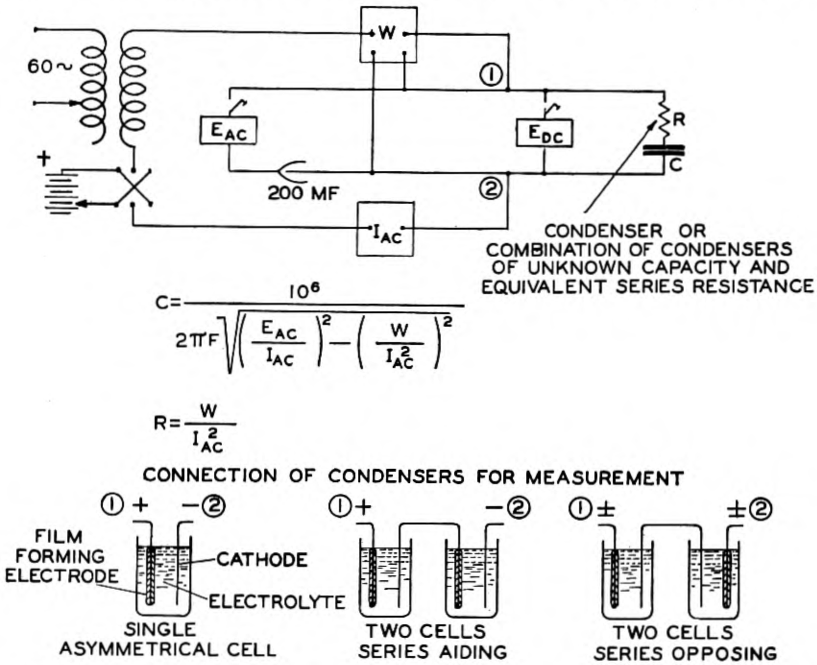


FIG. 5—Determination of capacity and equivalent series resistance of electrolytic condensers with 60-cycle alternating current superimposed upon a variable unidirectional potential.

of the two cells are connected together at the mid-point or when the negatives are so connected. On direct current it applies for series-aiding as well as for series-opposed connections.

It will be recognized that the resultant capacity $\dfrac{C_1C_2}{C_1 + C_2}$ is the same as is obtained when two ordinary static condensers of values C_1 and C_2 are connected in series. However, the internal distribution of electrostatic charges in opposed electrolytic condensers is quite different from that in series-connected static condensers, due to the rectifying characteristic of the films.

These determinations were made with a voltmeter, ammeter and wattmeter at 60 cycles, and were in conformity with measurements over a range of frequencies made on an impedance bridge. In the cases of measurements made with unidirectional potential applied, suitable bias was provided by superimposing the alternating current by means of a transformer in a battery circuit, the d.c. potential of which could be varied by changing the number of cells in series.

The apparatus for these measurements is shown in Fig. 5. It will be noted that a high-capacity blocking condenser is required in the a.c. voltmeter circuit, but is omitted from the potential circuit of the wattmeter. This condenser is inserted to block the unidirectional potential, which otherwise would be read by the voltmeter, but the power that the wattmeter indicates due to this potential is of no importance, and except for direct-current leakage would actually be zero.

Losses in Aluminum Cells and their Effect on Electrical Impedance

In the matter of electrical impedance characteristics, the electrolytic condenser does not approach a perfect capacitance as nearly as the more familiar forms of static condensers. Three sources of energy loss in the electrolytic condenser impart to it an equivalent series resistance, as a result of which the condenser current leads the impressed voltage by a phase angle somewhat less than 90°.

The first of these losses is the dielectric hysteresis loss, which, as in the case of the paper condenser, is approximately proportional to the frequency. The second loss is due to the resistance of the electrolyte and, in the case of aluminum condensers, may be of appreciable magnitude because of the low electrical conductivity of suitable electrolytes. This electrolyte resistance remains practically constant over a wide range of frequencies.

The third possible loss is due to the leakage-resistance of the film, which in its effect is similar to a high resistance in parallel with the condenser. Ordinarily this loss is negligible because the leakage current is less than a microampere per sq. cm. of film surface.

Conditions Affecting the Life of Condensers

To be successful from a commercial point of view an electrolytic condenser must have long life and must not require frequent attention. Otherwise the advantage in the matter of mounting space and the cost per unit capacity is offset by the depreciation and maintenance costs involved. There are two common conditions affecting the life of aluminum condensers that must be controlled if the cells are to operate satisfactorily.

The first concerns the chemical action of the electrolyte on the electrodes and the film. This action, which is merely a matter of the film dissolving and forming aluminum hydroxide in the solution, takes place when the cell is off circuit as well as when potential is impressed. With impressed potential, new film forms under the influence of the leakage current to replace that which is dissolved, but in time the fluid becomes saturated with aluminum hydroxide, which may precipitate as a white jelly and adversely affect the life of the condenser.

The second consideration involves corrosion of the positive electrodes. The susceptibility of aluminum to corrosion is well known, and in the use of electrolytic condensers anodic corrosion is the most damaging irregularity that can occur.

Obviously then an electrolyte must be chosen that does not rapidly dissolve the film, and the material for the electrodes as well as for the electrolyte must be selected and prepared to prevent serious corrosion of the "formed" aluminum plates.

Commercial Applications and Designs of Electrolytic Condensers

Reference has already been made to the use in telephone systems of electrolytic condensers. The principal applications of this device involve its use in low-pass electric wave-filters. These filters are placed in the supply circuits associated with central office storage batteries to eliminate noise-producing ripples and pulsations, introduced by battery charging-apparatus and signaling equipment, from the direct current furnished to telephone instruments. That is, the filters are used to exclude hum and other disturbing noises from the subscribers' circuits.

In Fig. 6 is shown an electrolytic condenser of the type designed for direct-current filter service. When prepared for operation on 24-volt d.c. circuits, the capacity of this cell is nominally 1,000 mf. at 1,000 cycles, and for 48 volts is about 600 mf. at the same frequency. The cell is 8 inches wide, 10.25 inches long and 14.25 inches high (20 x 26 x 36 cm.). Completely assembled it weighs about 42 pounds (19 kg.), including 22 pounds (10 kg.) of electrolyte.

The container for the condenser is made of heat-resisting glass which reduces possible breakage due to temperature variations. The electrodes, both of aluminum, are rigid and are bolted to a porcelain cover to keep them in proper space relation. Two supporting bolts, one from each electrode properly marked with respect to polarity, extend through the cover to provide the terminals for the condenser.

A thin layer of high grade paraffine oil is used on top of the con-

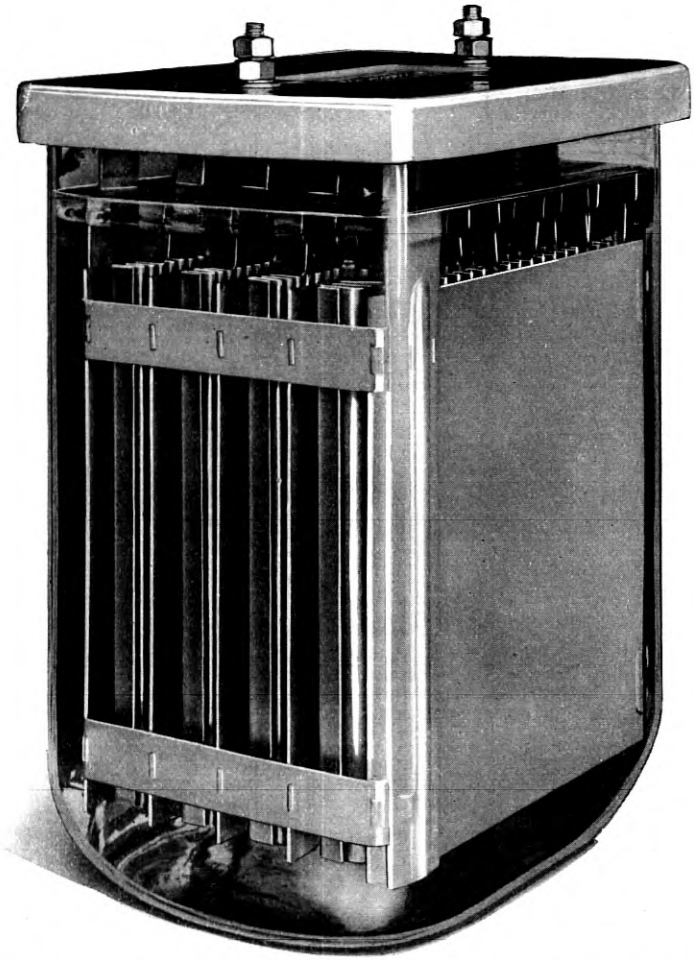


FIG. 6—Aluminum electrolytic condenser, designed for direct-current filter service

denser fluid to prevent evaporation and to keep the inside of the cell from sweating under varying room temperature conditions. The cover is sealed to the glass jar with paraffine to provide additional protection against evaporation and to prevent dirt from getting into the cell.

The Anode Construction and Material

The construction of the electrodes is shown in Fig. 7. The positive electrode on which the dielectric film is formed is made of four corrugated aluminum plates, each supported by four integral ears. In an assembled condenser the positive plate surfaces are entirely immersed in electrolyte, the ears extending up through the oil and providing contact with the positive terminal.

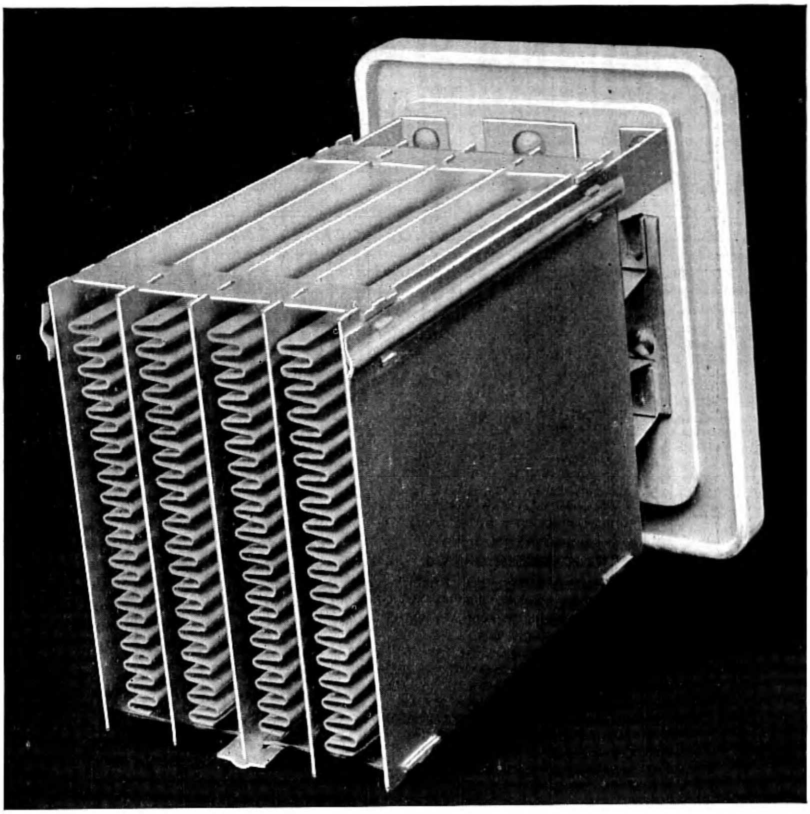


Fig. 7—Aluminum electrolytic condenser, showing construction of the electrodes

The material for the positive plates is aluminum of special composition, selected on a basis of properties which influence the formation of the film, the leakage current and the life of the metal. In general, the higher the purity of the aluminum the more rapid is the formation of the film and the lower is the resultant leakage current.

It has been noticed that the unit-area capacity for high-purity metals is slightly lower than for metals containing small quantities of alloying materials, but this difference is of inconsequential importance. The difference in the rate of film formation under similar conditions between 99.1 per cent aluminum and 99.6 per cent aluminum is shown in Fig. 8. After 24 hours in these particular cases the

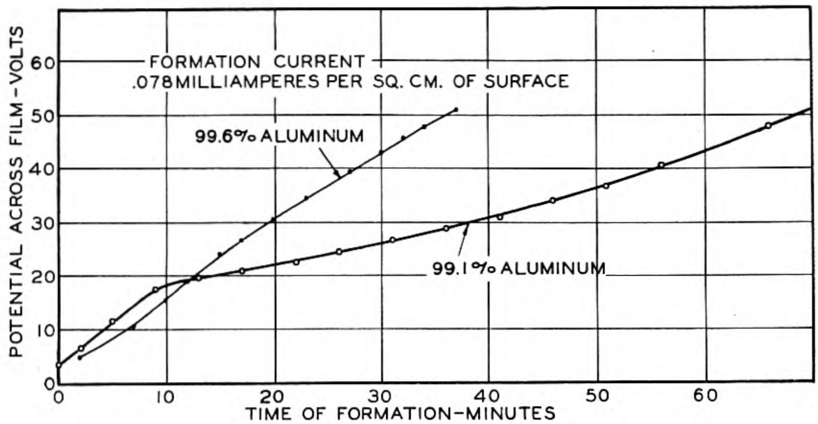


FIG. 8—Rate of film formation as influenced by composition of metal

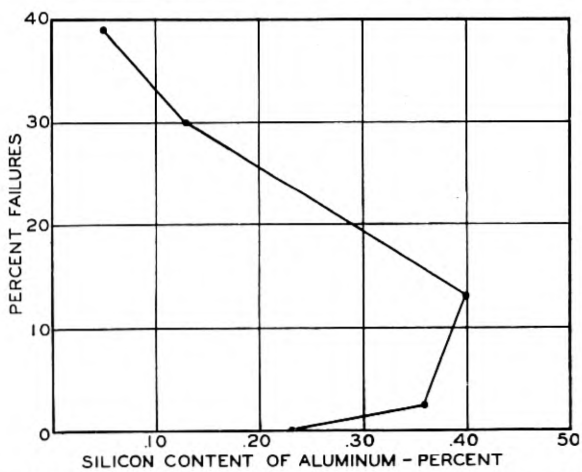


FIG. 9—The relation between silicon contents of anodes and failures due to corrosion in groups of aluminum condensers classified with respect to anode composition

leakage current of the 99.1 per cent aluminum was about 3 microamperes per sq. cm., or nearly six times that of the 99.6 per cent material at the formation potential of 60 volts.

In the matter of life, however, the purer metals seem to be more readily attacked by agencies capable of causing electrolytic corrosion.

This is illustrated by the performance of groups of condensers having anodes of different compositions as represented in Fig. 9. This curve shows the percentage of the cells in each group affected by corrosion in relation to the amount of silicon in the anode aluminum. It will be noted that the group of condensers with anodes of the purest metal (that is, least silicon) gave the least satisfactory results.

The Negative Electrode

The negative electrode consists of five rectangular flat plates, having a combined useful surface area about 35 per cent of the total positive surface. The negative plates or cathodes are of aluminum, but they do not have a film formed on them because their sole function is to provide contact with the condenser fluid. In an ammonium borate electrolyte, such as is used in these condensers, there are a number of other materials, including tin and carbon which can be used for the negative electrodes.

However, aluminum was chosen because it is light, relatively strong, easily worked and mounted, and can be cleaned by the same process used to clean the positive plate metal. A question might properly be raised as to the use of an aluminum negative electrode, particularly one having less area than the positive, because the formation of a film on the negative electrode will result in a reduction of the electrostatic capacity of the cell.

In normal operation with aluminum negatives there is a tendency for a film to form, even though the condenser is operated on direct-current circuits, because the negative electrode is an anode during the interval that the condenser discharges. However, this disadvantage with respect to the use of aluminum cathodes is overcome by making these plates of metal, which is really a rich aluminum alloy, containing enough other substances such as silicon to impede the formation of a film on its surface.

With material containing less than 99 per cent aluminum it is possible to have as much as 3.5 amperes a.c. in the condenser circuit or an alternating-current density of about 1 milliamp. per sq. cm. of negative plate surface without forming sufficient film on the negative plates to affect the capacity of the cell.

Prevention of Contamination in Manufacture and Installation

The initial formation of the film on the positive electrodes is carried out by chemical cleaning and electrochemical formation processes, in which the purity of the materials used as well as the composition of the electrodes is of importance. Because of the delicate nature of the film and the desirability of keeping it clean, the electrodes are designed so that in the manufacture, packing and installation of the condenser it is unnecessary to touch either the positive or the negative plate surfaces after the film is formed.

The fluid in which the condensers operate is shipped in sealed glass containers to prevent contamination, and the routine to be followed in the installation of the condensers emphasizes the need for cleanliness on the part of the installer. However, the design

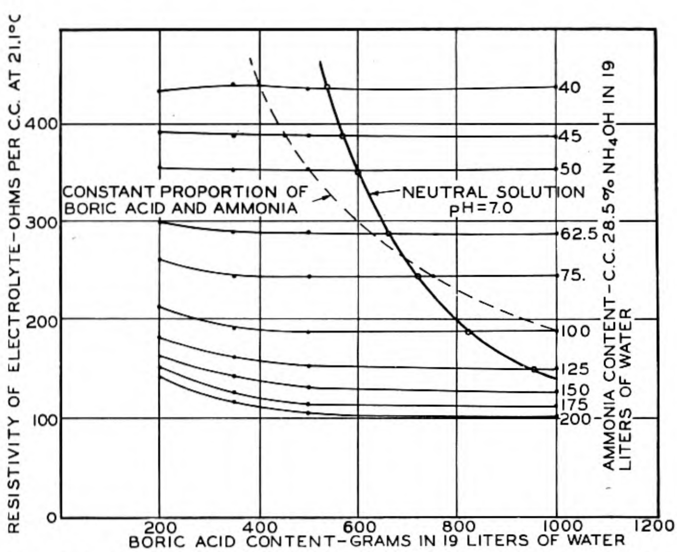


FIG. 10—Specific resistance chart of ammonium borate electrolyte

of the condenser and the method of supplying the condenser fluid have been arranged so that satisfactory installations can be made with ordinary skill and simple precautions.

THE COMPOSITION OF THE ELECTROLYTE

The electrolyte or condenser fluid as the solution is called is a mixture of ammonia, boric acid and water. A chart showing the specific resistance of different ammonium borate electrolytes is shown in Fig. 10. Two curves on this chart illustrate an interesting characteristic of these solutions. The heavy solid curve shows the neutral ammonium borate solutions, that is, those having a hydrogen ion concentration of 10^{-7} mols per liter.

This curve was determined colorimetrically with suitable indicators against standard solutions, and was checked by measurements with the hydrogen electrode. The dotted curve shows different dilutions for the same proportion of ammonia and boric acid. Since these curves intersect it is evident that the acidity of an ammonium borate solution decreases with dilution, and acid solutions may become alkaline by the addition of sufficient water.

Within the ranges of compositions and concentrations shown, it will also be noted from the set of approximately horizontal full line curves that the specific resistance of the electrolyte is practically independent of the boric acid content. That is, within these limits, the conductivity of the solution is substantially determined by the ammonia content, the amount of boric acid in the electrolyte affecting principally the degree of acidity or alkalinity of the solution.

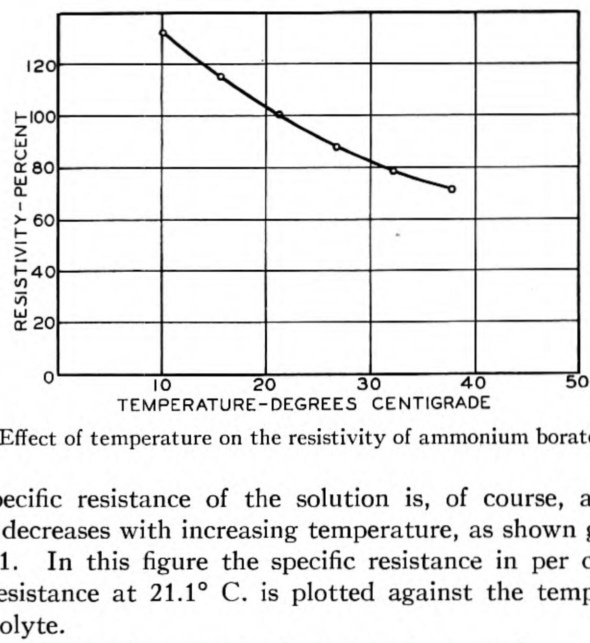


FIG. 11—Effect of temperature on the resistivity of ammonium borate electrolyte

The specific resistance of the solution is, of course, affected by heat and decreases with increasing temperature, as shown graphically in Fig. 11. In this figure the specific resistance in per cent of the specific resistance at 21.1° C. is plotted against the temperature of the electrolyte.

In the selection of a suitable electrolyte for a condenser the choice is influenced by the life of the solution, the effect of the specific resistance of the electrolyte on the electrical characteristics of the cell, and the susceptibility of the electrodes to corrosion in the solution.

The life of the solution, as has been explained, is determined by the rate at which it becomes saturated with aluminum hydroxide. In general, the lower the specific resistance of the solution, the more quickly does aluminum hydroxide form. On this basis the advantage of a high-resistance electrolyte is obvious.

But a limit is reached beyond which increases in specific resistance produce objectionable additions to the electrical impedance of the condensers, because of the high internal resistance set up in the cells.

In a condenser, as shown in Fig. 6 at room temperatures of about 25° C., a fluid having a specific resistance of 75 ohms per cc. will last from six months to a year without need for renewal because of the precipitation of aluminum hydroxide.

With fluid of 150 ohms per cc. the period of useful life is from one to three years, and with a 300-ohm solution is upward of five years, possibly never requiring renewal within the useful life of the cell. The rate at which the precipitate forms in a given solution is greatly accelerated at elevated temperatures and at 40° C., for example, the solution remains free from a white precipitate only about one third as long as at 25° C.

With respect to the effect of the acidity or alkalinity of the solution on the operation of aluminum condensers the difference is not readily distinguishable. Films can be formed and cells can be operated both in acid and alkaline electrolytes, and the electrical characteristics, except for resistance effects due to different solution conductivities, are essentially the same in both kinds of electrolyte. Somewhat better results, with respect to corrosion of electrodes have been obtained, however, with alkaline solutions, particularly under unfavorable operating conditions.

The Relation between Voltage of Film Formation and Operating Voltage

Electrolytic condensers, used on circuits associated with storage batteries, must be capable of operating at potentials throughout the range of voltage variations due to charging and discharging the batteries. Both the 24-volt and the 48-volt type condensers described can be used in circuits up to 140 per cent of their normal voltage, provision for this variation being made by the initial formation of a film to a potential somewhat above the maximum operating value.

After a condenser is connected in service the excess thickness of film is removed slowly by the chemical action of the electrolyte, because the film on the anode is maintained only at a thickness corresponding to the operating voltage. Ordinarily, therefore, the capacities of these condensers increase from their initial values, and stabilize at new values depending upon the maximum potential in the cycle of operating-voltage variations.

Capacity and Resistance Characteristics

The capacity and resistance characteristics plotted against frequency for a condenser of the type illustrated in Fig. 6, with a film formed to 46 volts d.c. are shown in Fig. 12. Curves No. 1 show the values

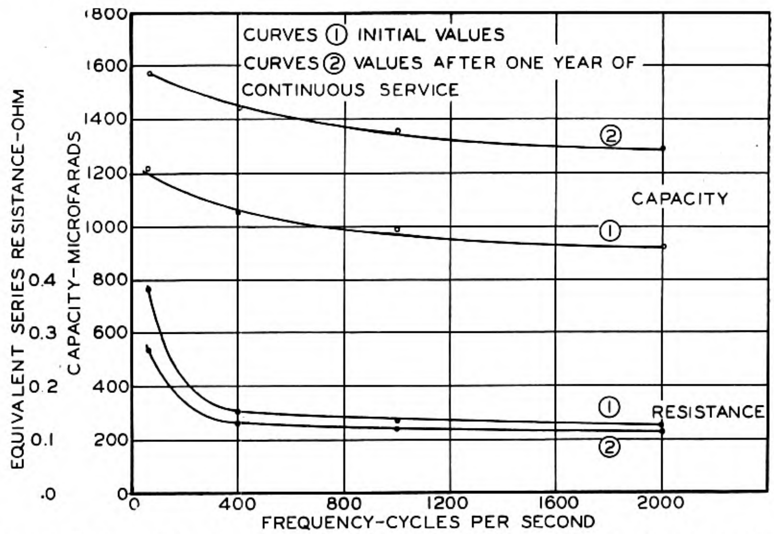


Fig. 12—Effect of frequency on capacity and resistance of an aluminum condenser formed to 46 volts and operated at 28 volts

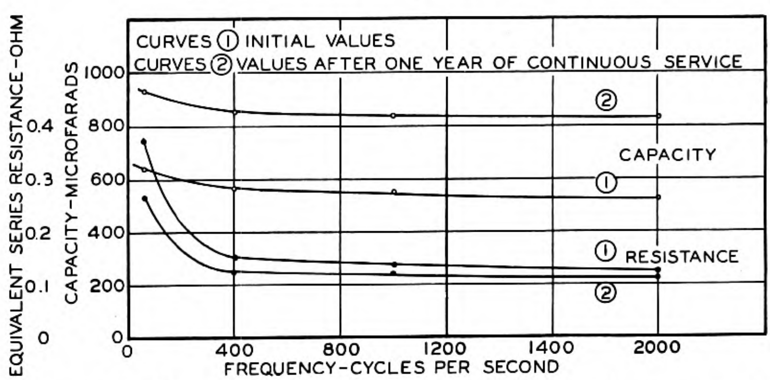


Fig. 13—Effect of frequency on capacity and resistance of an aluminum condenser formed to 100 volts and operated at 66 volts

measured at the time the condenser was put in service, and curves No. 2 show the values after one year of continuous service on a 28-volt battery with maximum voltage of 32. A second set of curves for a similar condenser formed to 100 volts and operated on a 66-volt battery, maximum 75, are shown in Fig. 13.

It will be noted that the capacities of the condensers decrease with increasing frequency. While there is a slight decrease in the unit-area capacity of films that accompanies a rise of frequency [7] this drooping characteristic is due principally to the corrugated shape of the plates on which the film is formed. The resistance through the electrolyte from the negative electrode to the film at the mouth of a "U" shaped corrugation, is less than that to the portion of the film at the bottom of the "U."

Thus, as the frequency increases, the alternating-current density at the mouth of the corrugations increases, while that in the trough decreases, resulting in a decrease of the effective capacity of the unit. The change in capacity due to frequency is greater in con-

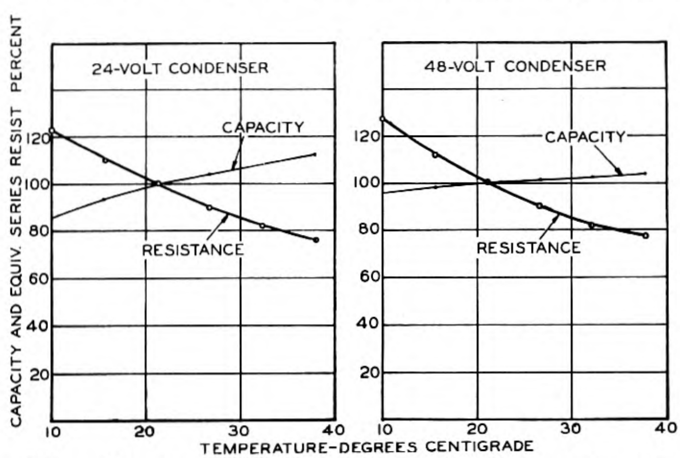


FIG. 14—Effect of temperature upon capacity and equivalent series resistance of 24-volt and 48-volt condensers at 1,000 cycles

densers with high-resistance electrolytes, and is more pronounced in the case of low-voltage films. The series resistance of the condensers is also observed to fall quite rapidly as the frequency is increased. This is due to the inverse proportionality between frequency and the component of the total condenser resistance, representing dielectric loss.

Since the change in capacity with frequency is caused by the difference at the crest and in the trough of the corrugations of the ratios of unit-area capacity to associated electrolyte resistance, changes in electrolyte resistance, caused by variations in temperatures, also influence the effective capacity of the cells.

The temperature effect is particularly noticeable at high frequencies,

[7] De Bruyne and Sanderson, *Trans. Faraday Soc.*, **23**, 42 (1927).

and, as in the case of the change of capacity with frequency, is more pronounced in condensers with low-voltage films. In Fig. 14 the effect of temperature on the electrical characteristics, at 1,000 cycles for 24-volt and 48-volt condensers, is shown. Here the change in resistance is due to the negative temperature coefficient of the electrolyte which, with increasing temperature, causes a reduction in that component of the total condenser resistance representing the resistance of the electrolyte.

In the normal adaptations of these condensers in low-pass electric wave-filters, the inherent changes in capacity and resistance with frequency and temperature cause no serious engineering difficulties. The limitations imposed by these variations are more than offset by the advantages of the corrugated structure in the matter of compactness and simplicity of design.

Corrosion of the Condenser Anodes

Notwithstanding the care taken in the manufacture and installation of condensers, there remains some possibility that the positive electrode will be attacked by corrosion. This corrosion may make its appearance as a gray growth or pitting on the surface of the anodes, or on the anode supports where they extend through the electrolyte. It is usually accompanied by the deposit of a granular or finely divided gray substance, probably aluminum oxide, which collects in the bottom of the condenser jar. Corrosion may occur shortly after a condenser is put in service or months can elapse before it appears.

In cases where corrosion has occurred, provided proper materials were used in the manufacture of the condenser, it has been noticed that the electrical properties of the cells were not seriously impaired. Such units have often continued to perform satisfactorily from a circuit standpoint for a number of years after the electrodes were attacked, even though the mechanical structure of the electrodes was damaged and weakened.

The long life that can be obtained from a corroded condenser under these circumstances is due in part to the tendency for areas affected to heal and restore the condenser to normal conditions. This indicates that the influence responsible for corrosion dissipates itself or it may be carried away from the corroded area by some of the products of the action to lie inert in the bottom of the cell. A number of spots on aluminum anodes have been observed, where the attack on the metal has ceased and a film has formed over the affected surface.

The leakage current of a condenser increases substantially when corrosion occurs, going up in some cases 20 or 30 times, but this current again decreases and approaches the original value. The capacity likewise increases when corrosion occurs, but this merely lowers the impedance of the condenser which, in most cases, is not objectionable. The performance of a condenser that has corroded and continued to operate satisfactorily is shown in Fig. 15. The curves on the chart show how the leakage current, capacity and resistance varied during continuous operation on a 65-volt d.c. circuit over a period of two years, both before and after the anodes were attacked by corrosion.

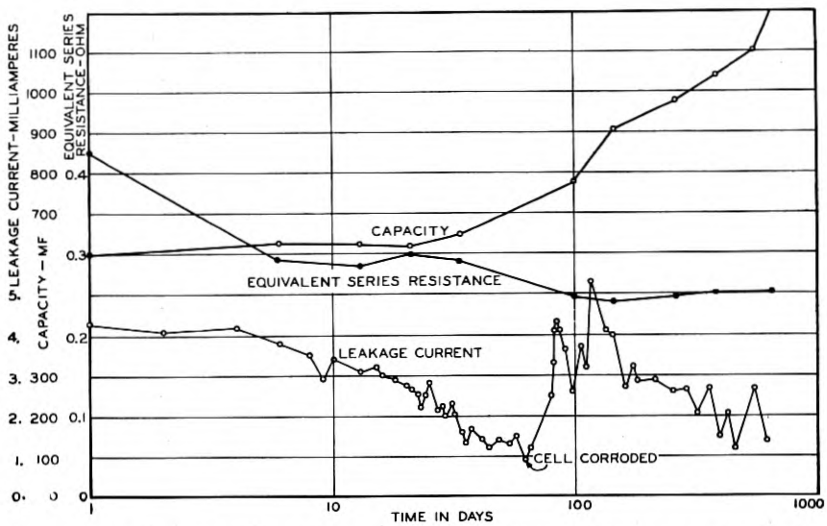


FIG. 15—The performance of a condenser, the anodes of which were attacked by corrosion after 62 days of operation at 65 volts direct current

Under certain conditions it is possible for the products of corrosive action to cause short circuits within a condenser, not by bridging from the positive to the negative plates because the corrosion product, aluminum oxide, is a non-conductor, but by accumulating between the plates and forcing the positive plates out of position into contact with the negative.

Also if corrosion occurs on the positive terminals eating through the supports one of the plates may drop and cause a short circuit. These possibilities of trouble are minimized by suitable design, and can be cleared up when they occur by removing the electrodes from the solution and repositioning or removing the deranged anode plates.

Because of the tendency for the vigor of the corrosive attack on aluminum anodes to decrease as the action continues and because the cell will continue to operate satisfactorily from an electrical standpoint, the best practice for condensers affected by corrosion is to leave them alone unless it is necessary to clear a short circuit due to a buckled plate or a severed anode support. Several cases have been reviewed where condensers are continuing to operate without maintenance, though corrosion was experienced nearly four years ago.

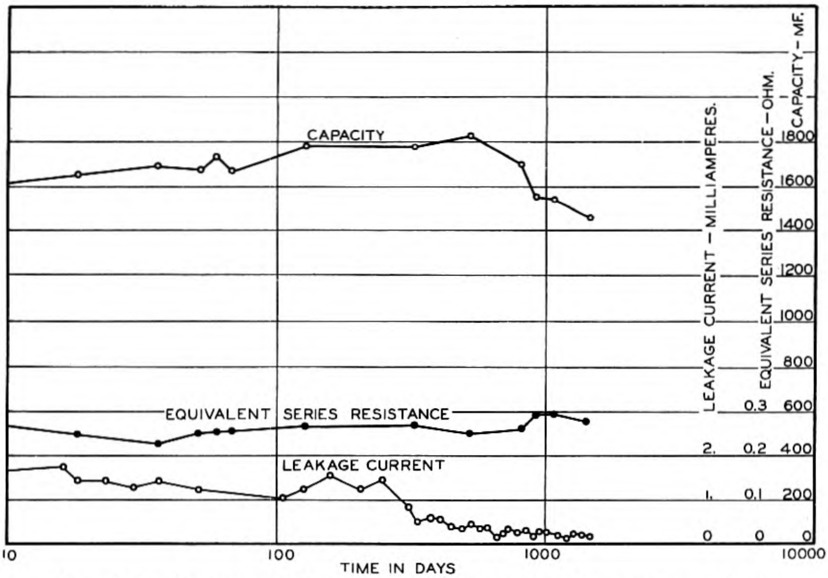


FIG. 16—Performance of a normal condenser operating continuously at 28 volts, direct current

The Life of Aluminum Condensers

With respect to the life of normal aluminum condensers, a number of cells have been on trial in actual service installations for more than five years, and several have been operating under test conditions for about eight years. In general, after the capacity of a condenser has stabilized at the value corresponding to the operating voltage, the cell continues to operate at a comparatively constant but slowly decreasing capacity due, perhaps, to a gradual thickening of the film with age.

The leakage current also decreases with length of service. The capacity, resistance and leakage current of a typical condenser operated for several years is shown in Fig. 16. In this case the fluid had never been renewed since the condenser was installed and after the period of service shown the electrolyte was still free from the white precipitate of aluminum hydroxide.

What the ultimate life of a condenser of this type will be remains for future determination. Judging from the appearance of cells that have operated for six and seven years there would seem to be years more of useful service to be rendered by these cells.

Contemporary Advances in Physics, XVII

The Scattering of Light with Change of Frequency

By KARL K. DARROW

SCATTERING of light is one of the commonest of all phenomena, which does not in the least imply that it is one of the most commonplace. Even its practical importance entitles it to high respect. We are often told that were it not for scattering, the sky would not be blue; the sun and the stars would stand out amazingly brilliant against a background black as coal. It is probable, however, that if scattering were suddenly to be suspended, the disappearance of the sky would be one of the least of our worries. Everything else would disappear, except what was self-luminous. The visible world would consist of the sun, the other stars, and flames, some electrical discharges, the filaments of incandescent lamps, and some substances glowing feebly with fluorescence or phosphorescence. Nothing else could be seen except as a silhouette, apart from objects so translucent that they could be viewed as a stereopticon slide against a flame. Happily no such calamity impends; and we may unconcernedly consider the theoretical importance of the process, which is great. As some might say, the scattering of light is one of the battlegrounds between the undulatory and the corpuscular theories. Metaphors of combat are however not appropriate; it is necessary to reconcile the theories, not to smash one or the other. Now it happens that some of the phenomena of scattering may be interpreted by the one theory, and some by the other; and some can be explained by either, which is most auspicious; for if this can some day be said of all the phenomena of light, the goal of our desires will have been attained.

Also scattering of light has just sprung into prominence as the most inviting and the most ardently invaded field of physics, because of a discovery such as, it had been supposed, could never happen again. It seemed that experimental physics had been so thoroughly developed —or, to change over to an ancient metaphor, that the field had been so thoroughly harvested and then so exhaustively gleaned—that nothing important could possibly remain to be discovered unless by measurements of great precision, or by radically new apparatus, or by applying voltages or other agencies on a scale as yet untried. Yet in the spring of 1928 a mode of scattering of visible light was discovered, by a physicist working with a quite ordinary spectroscope, a quite ordinary source of light and some very familiar chemicals, all of which had been available to everyone for at least fifty years. The physicist who thus saw what for half a century the whole world had overlooked was C. V. Raman of Calcutta.

Within a few months it was observed that what Raman had discovered was one of the special cases of a very general principle, which had already been stated almost as clearly as one can state it even today, but had failed of due recognition, evidently for want of some conspicuous example. Other special cases had indeed already been discovered among the phenomena of X-rays, but for some reason or other all except one had failed to make the impression which they should have made. Raman's discovery brought the principle sharply into relief. I shall not state it fully at this point; but this is its practical consequence: *when light falls upon matter, the scattered rays need not all have the same frequency as the infalling.* Most of the scattered light may preserve the initial frequency unchanged as formerly it was supposed that all did; but some part of it may be modified or shifted.

The present paper is devoted to this principle almost entirely; but to give it the proper background, I must at least mention some of the other aspects of scattering. Light may be scattered by particles of any size, from atoms up to grains of dust or droplets of mist; even the reflection from larger bodies is to be considered as the resultant of light-scattering from the particles of which these are made. Scattering by granules or droplets may be analyzed by the electromagnetic theory, the substance of the particle being considered as a medium with a different index of refraction and a different coefficient of absorption from the surrounding air. This however is distinct from the case with which we have now to deal, the scattering of light by molecules or atoms—although the transition from one type to the other should be very interesting and instructive.

The effect produced by molecules or atoms—it is known as the Tyndall effect, from its earliest thorough student—is fairly conspicuous in liquids or crystalline solids, though a very small trace of dust or finely-dispersed precipitate produces a brilliant scattering which completely overwhelms it. In gases it is difficult to see, but not impossible. The blue of the sky is a specific instance which is easy to observe, because the air is so deep.

The features of the scattered light which are commonly studied—I am speaking now of the time when it was assumed that there is no change in frequency—are its intensity, its polarization, and a property which is rather vaguely known as "coherence." The intensity of the light scattered from an incident beam, of which the wave-length is varied while the intensity is kept constant, increases very rapidly indeed with decreasing wave-length—in some cases, as the inverse fourth power of the wave-length. (This is the reason why the sky is blue; the molecules of oxygen and nitrogen are not especially tuned to blue light, but the waves from the sun are more powerfully scattered the nearer they lie to the violet end of the visible spectrum.) The scattered light is more or less polarized, even when the primary light is not polarized at all. In some cases the light deflected through 90° is perfectly plane-polarized; by the undulatory theory this signifies a very simple sort of vibrator in the atom, a vibrator which is attracted by an equal restoring-force whichever way it is displaced from its centre of vibration. In other cases the polarization is different, and other inferences about the atom-model may be drawn from it. As for the "coherence," it is a very important property—important for the theorist. If the scattering atoms contain vibrators which the infalling waves maintain in forced oscillations, and which themselves send out the scattered light, then these scattered or "secondary" wave-trains should interfere with one another. If in particular the vibrators, or let us say the atoms, form a regular lattice in space—a cubic lattice, for example—there should be destructive interference; the secondary wavetrains should completely destroy one another in all directions save that of the ongoing primary beam,[1] and there should be no perceptible scattering at all. The perceptible scattering is then a measure, to speak rather vaguely, of the *irregularity* in the arrangement of the atoms. This theory seems to be confirmed, for the light scattered without change of wave-length. Whether it is true also for the shifted light will probably soon be known.

Scattering of light with unchanged frequency is easy to explain by either wave-theory or corpuscle-theory. To those who think of light as waves and of atoms as systems of vibrators, it is a consequence of forced vibrations. The fluctuating or alternating electric field which—coupled with an alternating magnetic field—constitutes the beam of light, seizes upon an electrified portion of the atom and swings it to and fro in synchronous vibration. From this swinging electric charge, the scattered waves originate. It is evident that the forced vibrations and the scattering should be especially intense, when the frequency of the light coincides with a natural frequency of vibration of the atom, for then there is resonance. Now it is a fact that scattering is especially intense, when the infalling light agrees in frequency with any of

[1] More precisely, in all directions save those of the Laue diffraction-beams, which however for crystals and visible light do not occur.

certain spectrum lines which the atom may emit spontaneously. To those who think of light as a hail of corpuscles—"quanta"—scattering is rebounding of the quanta from atoms which they strike "elastically"; that is to say, as one elastic sphere striking another. This is easy to picture; but then we are left without any obvious explanation of the fact which was just mentioned—the fact that this sort of rebounding takes place especially often, when the quanta agree in frequency with those which the atom can naturally emit. Finally, for the single case in which the incident frequency agrees with that of a spectrum line and the scattering is very abundant, one can employ a compromise-theory; the atom is struck as by a bullet which sets it to vibrating freely with one of its own natural frequencies, as a bell which is struck by its clapper.

Scattering of light with change of frequency is certainly more complicated. The advocates of waves and oscillators must conceive that in the atom there goes on a process similar to what, in the art of electrical communication, is known as *modulation*. The frequency of the infalling light is modulated with some frequency characteristic of the atom. If the compromise-theory is valid, there are several cases in which one easily sees how this happens. Thus if a straight spring is alternately contracting and expanding with a frequency n_0, and at the same time is revolving around an axis perpendicular to its length with a constant angular velocity $2\pi n_1$, its ends will seem to a stationary observer to be moving with a motion compounded of two frequencies—$(n_0 + n_1)$ and $(n_0 - n_1)$; and if waves are sent out, they will have these frequencies jointly.[2] If the "spring" is an electrical doublet lying perpendicular to a magnetic field, it revolves automatically as it vibrates, and sends forth electromagnetic waves which are discriminated by the spectroscope into two lines of these two frequencies. Such is the explanation by wave-theory of the "normal Zeeman effect"; and while the actual effect of magnetic fields upon the light emitted by the atoms which they influence is not often exactly thus, it is sufficiently nearly so to prove that this interpretation is a step on the right path. If however one wishes to maintain the uncompromising wave-theory, and suppose that the vibrators in the atoms are kept going in forced vibration by the continually-acting waves of light, then modulation does not necessarily occur—not at least with the con-

[2] This idea was introduced by the elder Lord Rayleigh in the course of some speculations on the emission of light by rotating atoms, and was later turned to account in explaining the fine-structure of the bands which constitute the spectra of molecules.

ventional atom-models.[3] Models however can be devised which account for modulation, and perhaps they will become more popular.[4] In one simple case related to the Compton effect, forced vibrations result in waves which do not coincide in frequency with the primary waves; this is the case of a free electron, which the magnetic force in the light-stream pushes more and more rapidly forward as the electric force makes it swing more and more rapidly crosswise; and as the electron gains speed, the frequency of the waves which it sends to a stationary observer steadily sinks.

To the thoroughgoing advocate of the corpuscle-theory, however, the problem of the scattered light of shifted frequency seems simple; or, at all events, the first step in explaining it seems obvious. Frequency of light, when multiplied by the universal constant h, is the measure of the energy of the corpuscles of the light. Change of frequency therefore means transfer of energy. If a quantum of frequency n_0 flies onto an atom and a quantum of frequency n_1 flies away, energy in the amount $h(n_0 - n_1)$ stays behind with the atom. If n_1 is greater than n_0, as sometimes happens, the departing quantum takes with it some energy which belonged to the atom as well as all that was brought by the oncoming quantum. As yet there is no picture of the process by which the energy is passed between the matter and the light. But we are not supposed to ask the quantum-theory for such pictures. Perhaps one reason why it seems so much stronger than the wave-theory is, that of the latter we have expected so much more.

However, visualizable or not, the corpuscle-theory implies that if the scattered light differs in frequency from the infalling light, individual molecules or atoms are receiving or giving energy in quantities equal to the frequency-difference multiplied by h. Not any and every amount of energy may be annexed or ceded by a molecule or an atom—only certain sharply definite, distinct and separate amounts, equal to the energy-differences between the state in which the particle happens initially to be, and one or another of its various other "permitted" stationary states. There are exceptions to the rule, as I will state immediately; but in experiments performed with visible light they are not apparent. Correspondingly the frequency-shifts of the scattered light are limited to certain distinct and separate values;

[3] Thus an electron which is subject at once to a quasi-elastic restoring-force, a sinusoidal electric field, and a constant magnetic field perpendicular to the electric field, describes a fixed orbit with a single frequency equal to that of the electric force, and there is no modulation. Mistakes in this respect have been made by various people who theorized about the Wood-Ellett effect.

[4] Such models have been devised by Hartley and by Kennard.

the spectrum of the scattered light due to a single primary spectrum line consists of that line accompanied by a number of others, separate from it and from one another. Intermediate frequencies do not occur, for they would correspond to transfers of energy in quantities which the atoms are not able to offer or accept. The shifted lines which do occur, the "Raman lines," reveal the energy-values of the stationary states of the scattering particles.

We consider next the exceptions to the rule—the cases in which a scattering particle may accept or surrender any quantity of energy whatever within an appreciable continuous range, instead of merely certain separate discrete amounts. This may be possible if the energy conceded by the quantum is employed in altering the speed of the particle, or in breaking the particle into pieces and imparting speed to these—if it becomes kinetic energy of translatory motion of the molecule or atom, or of the fragments thereof. Translatory motion is non-quantized, which is a way of saying that it is not under the dominion of quantum-conditions which allow to it some values and deny it others. Any amount of kinetic energy of translation is permitted to a molecule or an atom, so far as we know. This suggests that any amount of energy may be transferred when such a particle meets a corpuscle of light, provided that so long as the energy is held by the molecule or the atom it is held in this form. But there is another limitation to be remembered—that imposed by Newton's principle of the conservation of momentum. If a swiftly-moving corpuscle of relatively small mass m strikes a slowly-moving body of much larger mass M, the latter cannot gain much speed in the encounter; for it cannot acquire speed without acquiring momentum, and if it were to accept for that purpose more than a very small fraction of the energy of m, it would have to take more momentum than all that m possesses.

Now relatively to an atom, a corpuscle of light is a body of very small mass and very swift flight indeed; and a quantum of frequency n cannot transfer to an atom of mass M, for use as kinetic energy of translation, more than the fraction $2hn/MC^2$ of its own initial energy—more than the quantity $2h^2n^2/MC^2$ altogether.[5] For a quantum belonging to the visible spectrum the fraction $2hn/MC^2$ is of the order of 10^{-8} even for an impact with the lightest of all atoms. The utmost possible shift in frequency of the scattered light would bear only this proportion to the primary frequency, and would be indistinguishable. But the higher the frequency of the quantum, and the lower the mass of the

[5] The formula is approximate, but the approximation is very close in all practical cases. For the derivation of this and the accurate formula, see A. H. Compton, *Bull. Nat. Res. Council*, 20 (1922) or my *Introduction to Contemporary Physics*, pp. 148–149.

scattering particle, the greater this maximum possible transfer of energy and this maximum possible frequency-shift become; and for a collision between an X-ray quantum and a free electron, it attains the order 10^{-2} of the primary frequency, and is very appreciable. In fact, the frequency-shift occurring when X-ray quanta transfer energy to free electrons and these employ it as kinetic energy was the first of all to be observed. It is simply the Compton effect. It was noticed first towards 1904 and was described as "softening of the scattered X-rays," and in 1922 was for the first time properly measured and properly interpreted by Arthur Compton. The scattered rays include every frequency from that of the primary rays, n_0 let us call it, downward to the lower limit $(1 - 2hn/MC^2)n_0$, as they should.[6]

If the primary quantum breaks the particle which it strikes into two or more fragments—as for instance when an atom is ionized or a molecule dissociated—the requirement of conservation of momentum no longer limits the amount of energy which it may pass to these. It must give at least enough energy to ionize or to dissociate the particle; beyond this, so far as we know *a priori*, any extent of transfer is permitted. Hence we should expect to observe in the spectrum of the scattered light a continuous band, commencing at the frequency which is less than that of the primary light by the quotient of h into the ionizing-potential or the dissociation-potential, and extending towards lower frequencies indefinitely far. More precisely, we should expect to observe as many of these bands as there are modes of dissociation or modes of ionization feasible by light.

No one, so far as I know, has yet observed any bands corresponding to dissociation of molecules or to the detachment of loosely-bound electrons by visible or ultra-violet light. In the X-ray region, however, it is different. In the spectra of scattered X-ray bands answering to this description, and suggesting that X-ray quanta have extracted deep-lying tightly-bound electrons from atoms and have conferred kinetic energy upon them, have in fact been reported. Several such spectra were depicted in 1923 and 1924 by G. L. Clark and W. Duane.

If it should turn out in any special case that quanta could extract electrons from atoms, but could not confer extra kinetic energy of translatory motion on them—a restriction which there is no evident

[6] This is disguised by the fact that the rays scattered in any one direction (relatively to the primary beam) are of a single frequency. If we observed simultaneously rays scattered in all directions, we should see a continuous band of light extending between n_0 and the stated lower limit. This condition was approached, though not purposely, in some of the earlier researches on the Compton effect.

The shift which should occur if quanta of the visible spectrum are scattered by free electrons is very small, but sufficiently large to be appreciable; however, this type of scattering does not seem to occur to a perceptible extent, for it has been sought in vain (P. A. Ross).

reason to foresee—then the aforesaid bands would be reduced to lines, shifted from the primary line through frequency-intervals equal to the ionizing-potentials divided by h. Such lines were observed in the early spring of 1928 by B. Davis and D. P. Mitchell, in the spectrum of X-rays scattered by graphite.

There is one more way in which corpuscles of light can dispose of part of their energy—in setting into vibration atoms which are built into the structure of crystal lattices. Many crystals behave as if they contained oscillators having natural frequencies of the order 10^{12}–10^{14}, and able to emit light of the corresponding wavelengths, which are in the infra-red region of the spectrum. Some of the quanta which strike such a crystal lose energy in being scattered, and the energy which they lose is equal to h times one or another of these oscillation-frequencies. This effect was discovered independently in the late winter or early spring of 1928 by G. Landsberg and L. Mandelstam, and by C. V. Raman and K. S. Krishnan.

Presently I will quote in more detail the data which establish all these facts; but first it is urgent to point out that there is another phenomenon in nature, a phenomenon long known and well known, with which the scattering of light with altered frequency can readily be confused; indeed it is often difficult, and I suspect that it may sometimes be impossible, to tell whether in an actual case we have the one or the other before us. I refer, of course, to fluorescence. The description of fluorescence, indeed, reads exactly like the description of scattering of light with change of frequency. Light of one frequency falls upon a substance, and light of another frequency emerges from it. How then shall we discriminate between the two?

According to the ordinary conception of fluorescence—a conception which has attained to the rank of a definition—the molecule or the atom absorbs a quantum of the incident light, and is put thereby into an excited state; and after a longer or a shorter time, it passes spontaneously into a state different both from the excited and from its original state, and emits a quantum which is not of the same frequency as the one which it absorbed. (I am considering fluorescence of gases or of dilute solutions, where one can suppose that the quanta are absorbed or emitted by individual molecules; more complex cases are too complex, for the time being.) Let the original state be symbolized by N, and the final state by A, and the temporary state by B; denote by $E_{AN} = hn_{AN}$ the energy-difference between A and N, positive if the energy in state A is the greater; use the letters n_{BN} and n_{BA} correspondingly, and denote by n_0 the primary frequency. Then there will be no fluorescence at all unless $n_0 = n_{BN}$; unless, that is to say, the primary quanta have exactly the right energy to transfer a molecule from its original state to some other (excited) state. Suppose however that this condition is fulfilled; then the quantum of fluorescence-light is emitted when the molecule passes from state B to state A, and therefore has the frequency n_{BA}. But because of the energy-relations, we have

$$n_{BA} = n_{BN} - n_{AN} = n_0 - n_{AN},$$

which means that the quantum of fluorescence-light has exactly the same energy as a quantum of the primary light would retain, if it had been scattered from the molecule after communicating to this latter the energy requisite to transfer it from its original state N to the state A. The fluorescence-light is shifted in frequency from the primary line by exactly the same interval as the Raman line corresponding to the transfer of the molecule from N to A.

It follows then that one can never decide by measurement of wave-length whether a line in the scattered light is due to fluorescence or to scattering with change of frequency.[7] This is inevitable; for in either case the molecule involved in the process starts from the same initial and ends in the same final state, and the frequency of the departing quantum depends on nothing but the difference between these two. The only question at issue is, whether the molecule has gone from the initial to the final state directly, or *via* the temporary state B.

One way of solving the question seems obvious: to vary the frequency of the light with which the molecules are irradiated, and notice whether the shifted line which is under observation—a line is identified by the amount of its shift, not by its actual frequency, so that a given line travels along the spectrum *pari passu* with the primary light—makes its appearance when and only when the quantum-energy hn_0 of the primary beam coincides with the energy-difference between the initial state of the molecule and any one of its other states. In other words: does the shifted line appear only when the primary quanta can be absorbed by the molecules—when the infalling light coincides with a line of the absorption-spectrum of the substance? or does it appear always? If the latter is the case, it is the Raman effect which we have before us, at least when the wave-length of the infalling light is such that its quanta are not absorbed.

If however the shifted line is most intense when the primary fre-

[7] This may be too strongly stated; one might observe a fluorescence-line emitted—to use the foregoing symbols—by reason of the transition from A to N (not that from B to A) which could not be a Raman line correlated with a transfer of the molecule out of the state N. On the other hand it could be a Raman line associated with a transfer of the molecule out of state A, so that one would have to assess the relative likelihoods of the states A and N among the molecules.

quency coincides with a frequency in the spectrum of the scattering substance, it does not necessarily follow that we are dealing with a case of fluorescence. Scattering without change of frequency is very much intensified, when such coincidence is brought about. Experience teaches this, and the wave-theory also; for a vibrator scatters waves most powerfully, when they and it are in resonance together. Scattering with change of frequency may follow the same rule. The proof of fluorescence, then, turns finally on this: can it be shown that there is an interval of time between the moment when the primary quantum impinges on the molecule, and the moment when the secondary quantum leaves it?

Often with solid substances one can actually see that the secondary rays continue to emerge for an appreciable time after the primary rays are discontinued; but with gases no such great delay has so far been observed.[8] However, there are sometimes indications that between the arrival of the primary quantum and the departure of the secondary quantum, there is an interval of time during which something can happen to the molecule or atom—something which changes the nature of the departing quantum, and may even prevent it from ever being born. Pure rarefied gaseous mercury and sodium and iodine, to take three instances, emit light vividly when they are illuminated; but if they are made very dense, the intensity of the emitted light is much reduced, or its spectrum is entirely changed, or both of these things happen; so also when they are mixed with gases such as hydrogen or argon.

Such results, it is clear, are difficult or impossible to explain if the emitted light consists of primary quanta which have rebounded *instantaneously* from collisions with (say) mercury atoms; for such collisions would be more numerous when the gas became denser, and not much less numerous when the gas was diluted with argon; and the rebounding quanta would vary proportionately in number, while the frequency-shifts which they display would not be changed (unless one were to hit two or more atoms in succession). However, if the mercury atoms absorb the primary quanta, and hold on to their energy for a while, and subsequently by some independent process release it, all these effects are quite easy to interpret. The atom which has accepted the energy of a quantum, and has not yet decided to disgorge it in the form of a secondary quantum, may meet another atom and unload the energy in part or altogether. It is certain that this can happen; for when mercury vapor is mixed for instance with thallium

[8] Methods whereby it might be possible to measure the time-interval have been suggested by Ruark.

vapor and the mixture is bombarded with quanta which mercury atoms can absorb and thallium atoms cannot, we nevertheless presently find the thallium atoms possessed of some of the energy which was sent in. Collisions of atoms occur more frequently, the denser and the warmer the gas or the mixture of gases becomes; the opportunities for diversion of energy, which otherwise would be re-radiated as fluorescent light, become correspondingly more numerous. When therefore the light emitted by an illuminated gas changes its spectrum or fades away as the gas is densified or contaminated, the probabilities are that it is true fluorescence-light.[9] The influence of a magnetic field upon the character of the emitted light may also furnish evidence.

One sees therefore that the distinction between scattering and fluorescence is by no means immediate. Even the case which seems most explicit of all—where the primary light agrees with one of the spectrum-frequencies of the atom and the secondary light is unshifted, as when rarefied sodium vapor is illuminated by one of the *D*-lines and re-radiates it—is not exempt from doubt. Very likely part of the re-radiated rays is fluorescence-light and part consists of scattered quanta. It is obvious why Raman thought at first that he was observing fluorescence. Others very likely had already noticed the Raman effect, and classified it merely as another instance of the already well-known phenomenon.

I will now relate some of the details of the recent experiments which have suggested that light may actually be scattered with change of frequency.

The Raman Effect

The scattering of visible[10] light with change of frequency was first discovered by a man who was working with molecular liquids. It is interesting and instructive to consider why the effect, so obvious under these circumstances, had eluded the numerous and notable physicists who had studied—exhaustively, it was thought,—the influences of light on gases and of gases on light.

In the first place, a liquid contains many more molecules per unit volume than a gas, and therefore offers many more opportunities for collisions of quanta with molecules. This is essential, for collisions which result in excitation of the molecule and in scattering of the quan-

[9] I suspect that Saha, in concluding that the resonance-spectra of vapors discovered by Wood are actually examples of scattering with change of frequency, did not take sufficient account of some of these phenomena. Consider for instance those observed by Wood and Loomis (*Jour. Franklin Inst., 205*, pp. 489–495).

[10] I will reserve the name "Raman Effect" for the scattering with shift-of-frequency of light of the visible and adjacent ranges of the spectrum, as in the X-ray region the effect was earlier discovered.

tum with altered frequency are evidently relatively rare; otherwise they could not have escaped the notice of those who have studied gases.[11] Even scattering without change of frequency is unusual, unless the primary light coincides exactly with a spectrum-line of the molecule; the blue of the sky is conspicuous only because the air is so thick; in the laboratory, light scattered with unshifted wave-length by a gas can be seen only if the gas is dense, the primary light blindingly brilliant, and the eye thoroughly rested.

But if it had occurred to any physicist to seek for the effect with (say) mercury atoms, by crowding the atoms together into the liquid form, he would certainly have rejected the idea the moment after it flashed across his mind; indeed it would probably never have flashed; for as

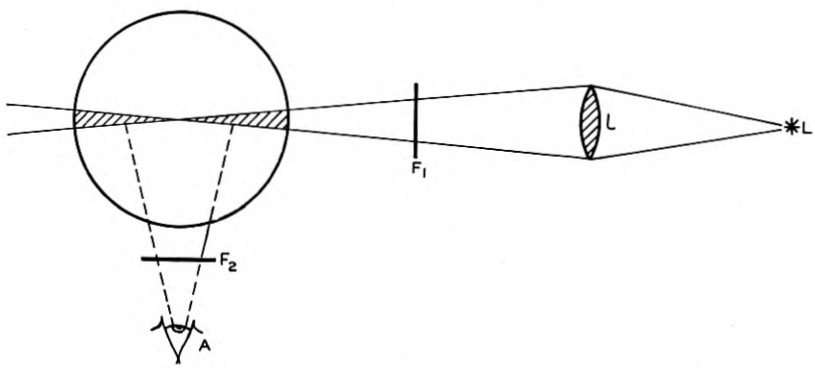


Fig. 1—Sketch of scheme for observing fluorescence and scattering. (After Pringsheim.)

soon as atoms are forced into such close proximity, their excited states, or those at least with which we are now concerned, simply disappear. The *free* mercury atom, for instance, has a "4.9-volt" excited state—that is to say, a stationary state into which it will pass over, if being initially in its normal state it receives an acceptable offer of 4.9 equivalent volts of energy. Were we to bombard mercury vapor with 6-volt quanta—i.e. with corpuscles of light, each possessing 6 equivalent volts of energy—we might expect some of these to transfer 4.9 equivalent volts to atoms which they strike, and rebound as 1.1-volt quanta. But there is no reason to expect anything of the sort with *liquid* mercury; there is no reason to suppose that the atoms are avid to grasp this particular amount of energy, and plenty of reason to suppose that they are not. The same holds for every other excited state of a free atom, of which the energy-excess over the normal state is smaller

[11] Also one would expect to find, in the light scattered from liquids, quanta which have suffered two or more collisions; such have not yet been reported, so far as I know.

than the energy at the disposal of corpuscles of visible or ultra-violet light. Modern atomic theory makes this vanishing of the excited states seem very plausible; for the said excited states correspond to particular arrangements of the electrons at the surface of the atom, which are completely disorganized when atoms are crowded close together. But, apart from theory, it is an experimental fact.

One can therefore scarcely hope to find scattered light with a spectrum composed of discrete separate lines, unless one can find atoms or complexes of atoms of which the low-energy excited states remain discrete, separate and accessible when the substance is liquefied or solidified. With atoms, as I have said, this appears to be impossible. The high-energy excited states, in which one or another of the deep-lying electrons is absent from the atomic system, are indeed the same whether the atoms are free or are crowded together into a solid or a liquid; but corpuscles of light of the visible or the ultra-violet spectrum have not energy enough to excite them, and therefore for the time being they fall out of our purview. Molecules, however, do possess excited states, into which they may be transferred from the normal state by offering them one or two or three, or even a fraction of one equivalent volt of energy; and these they possess, even when jammed together in the liquid state.

These excited states of molecules correspond, according to modern theory, to various amplitudes of vibration of the atom-nuclei relatively to one another within the molecule. The simplest case, of course, is that of the diatomic molecule, of which there are so many examples—oxygen and hydrogen and nitrogen, for instance. The nuclei of the two atoms, being positively charged, repel each other; but the electrons, aided possibly by additional magnetic fields, exert upon each nucleus a force which tends to push them together; and there is a certain internuclear distance of equilibrium, for which the two opposing forces balance. If the nuclei are displaced slightly from their points of equilibrium, they vibrate. Vibration is a quantized form of motion; only certain amplitudes and certain energy-values are permitted. The low-energy excited states of the molecules correspond to the permitted amplitudes and the permitted vibration-energies; when a quantum excites one of these and rebounds with the remainder of its energy, the energy which it gives up is spent in augmenting the vibrations of the nuclei.

These low-energy excited states are responsible for Raman's discovery; partly because, as I have stated, they survive when the molecules are jammed together into a liquid—a fact which evidently means that the electrons, which produce the force upon the nuclei countervailing their reciprocal repulsion, are shielded from the outer world, presumably by other electrons lying still farther outward from the nuclei; and partly because their energy-values are so conveniently low. This latter point can best be illustrated with an example. To perceive a relatively feeble optical effect, or one which is expected to be feeble, it is best to produce it in the visible spectrum—not merely in order to observe it with the eye; the major reasons are rather, that in

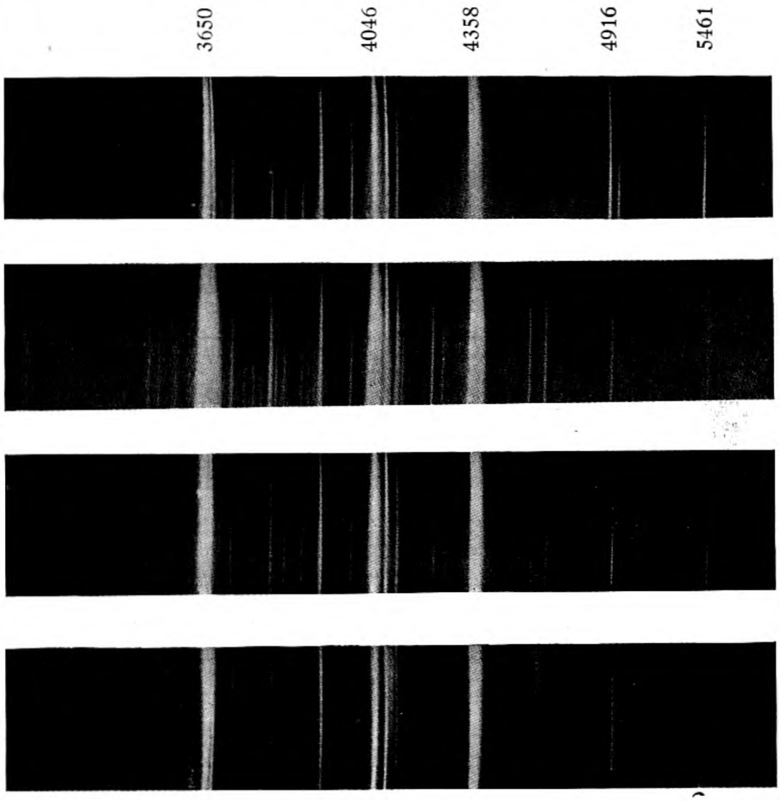


Fig. 2—At top, spectrum of primary light (mercury arc); below, spectra of light scattered by benzene, toluene, pentane respectively. (C. V. Raman; *Indian Journal of Physics.*)

going away from the visible spectrum-range in one direction we find the photographic plates becoming rapidly less sensitive, while in the other direction the transmission of the rays through matter grows steadily worse. Suppose then that one tries to produce the Raman effect by light near the high-frequency limit of the visible—say about 4000 Å., where the quantum-energy is about 3 equivalent volts. If the scattered quantum in its turn is to be in the visible spectrum, its wave-length must be less than some 8000 Å., its energy more than roughly 1.5 equivalent volts. The primary corpuscle of light must therefore not cede to the molecule or atom more than $(3-1.5)$ or 1.5 equivalent volts of energy—the material particle must therefore be able to receive energy in quantities less than this, quantities preferably which are small fractions of an equivalent volt; its excited states should differ from the normal state by energy-differences of this order; and molecules satisfy this condition.[12]

En somme, then, the scattering of light with change of frequency was never discovered in all the abundant work on the common monatomic

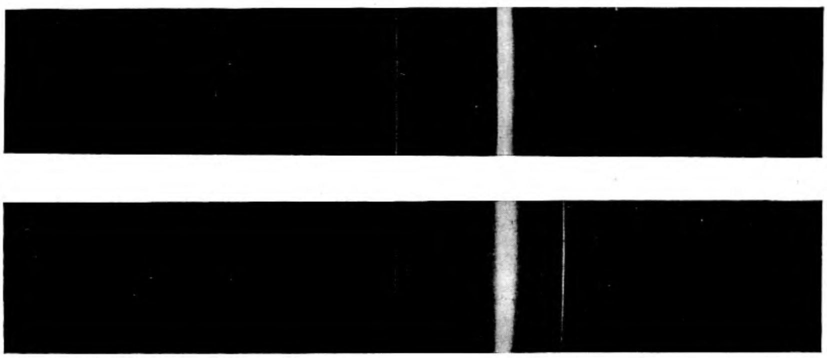


Fig. 3—At top, spectrum of primary light, reduced by a filter almost to a single strong line; below, that of light scattered by benzene, showing shifted lines due to that single line (4358). (C. V. Raman; *Indian Journal of Physics*.)

gases, because: first, the atoms were too few in unit volume to cause much scattering; second, they could not be squeezed together without destroying their power of entering distinct excited states; third, the excited states had energy-values so high, that a quantum of the visible or the near ultra-violet spectrum striking a normal atom would simply not have had energy enough to transfer it into one of them. It was discovered by Raman in the way in which he discovered it, because he was working with molecular liquids where: first, the molecules were numerous in unit volume and scattering was frequent; second, in spite of being squeezed together the molecules retained their power of enter-

[12] Atoms however do not always satisfy it; in particular, those of the noble gases and of mercury, the substances most often used in optical experiments on monatomic gases, possess no excited states differing from the normal state by less than four equivalent volts—another reason why the Raman effect was not sooner discovered. The more massive of the alkali metals have excited states superior to the normal by about 1.5 equivalent volts, while metals of the third column of the periodic table would be very favorable.

ing distinct excited states; third, the excited states had energy-values so low that a quantum of the violet, blue or green regions of the spectrum striking a molecule had plenty of energy to excite it and yet have some left over. Subsequently one of Raman's associates (Ramdas), photographing with very long exposure, detected the effect with ether vapor. Carrelli and his colleagues obtained it with molecular salts in aqueous solution; [13] and perhaps in the course of time somebody may overcome the obstacles, and demonstrate the scattering of light with change of frequency by free atoms of a monatomic gas.

We will now consider some of the photographs which were made by Raman, and later with improvements of technique by Wood, Langer and Meggers, Brickwedde and Peters.

Some of Raman's earliest published pictures are reproduced in Fig. 2. At the top we see the spectrum of the primary light—that of the mercury arc, the source of light employed, I think, in all the researches thus far published. The strong lines near 4046 and 4358 are responsible for most of the Raman lines thus far observed by anyone; other strong lines are those near 5461, and the pair at 5770 and 5790. (These last look much fainter in Fig. 2 than in some of the others, but such variations are due to the photographic plates employed, and should not be heeded.) The three spectra below are, in order, those of the light scattered by benzene, toluene and pentane. The new lines are extremely numerous—so much so, that some care is required to determine for each new line which is the primary line whence it is shifted. This may be done by filtering out from the primary light all but one of its strong lines. In making the photographs in Fig. 3, Raman and Krishnan used a filter which removed from the infalling light almost all the quanta but those of the wave-length 4358 (though 3650 and 4046 are still seen dimly in the spectrum, the topmost one in the figure). The spectrum of the scattered light, below, now shows additional lines which are certainly made of quanta which originally had the wave-length 4358.

Figs. 4 and 5 show the spectra scattered by benzene and carbon tetrachloride, as photographed by Wood.[14] The "fat" lines from left to right are the unshifted lines 4046, 4358, 5461 and the aforesaid doublet 5770–5790. (The rich adjacent spectrum is a "comparison" spectrum of iron.) Most of the lines companioning 4358 and 5461 on both sides are shifted lines. Notice, in the spectrum scattered by

[13] With salts which are completely dissociated in solution they failed, as they expected, to obtain it. Possibly the effect may some day be used as a measure of percentage of dissociation!

[14] I am much indebted to Professor Wood for furnishing me with prints of these, and to Dr. Langer for plates from which the next two figures were made.

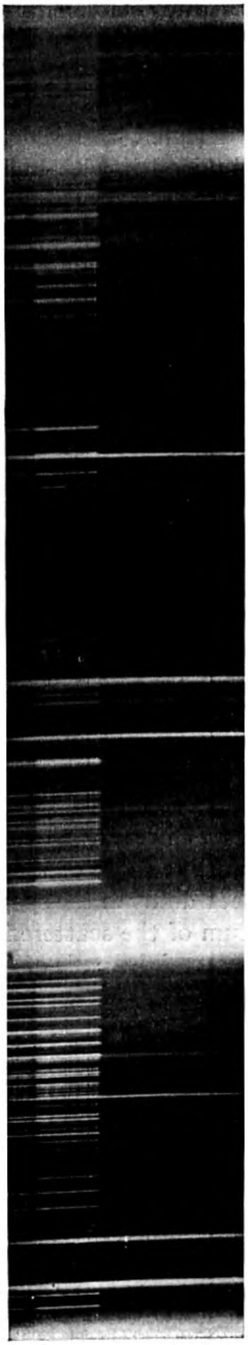


Fig. 4—Light scattered by benzene (with comparison spectrum). The four fat lines from left to right are 4046, 4358, 5461 and the doublet 5770–5790 of the unshifted scattered light. (R. W. Wood.)

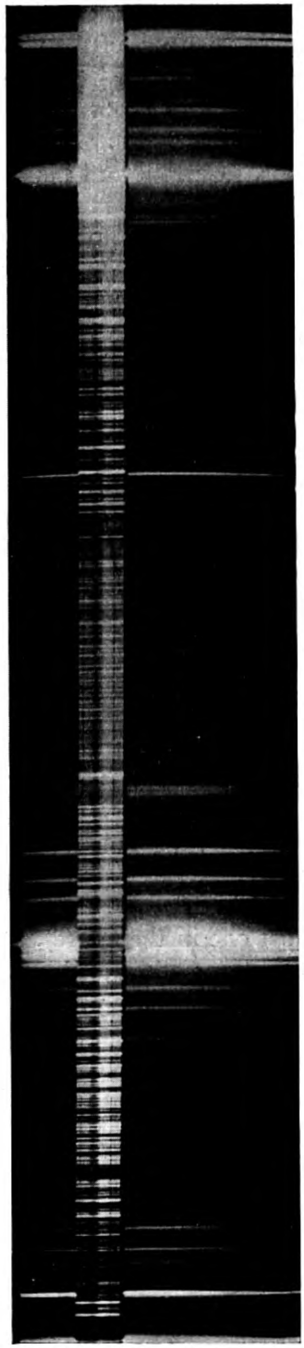


Fig. 5—Light scattered by carbon tetrachloride; note the array of lines shifted each way from 4358. (R. W. Wood.)

carbon tetrachloride, the triad of lines to the right of 4358, and the equally-spaced triad to its left. These latter are "anti-Stokesian" lines—I will presently explain the name—and consist of quanta which have received as much energy from molecules as the quanta of lowered frequencies have given up. With carbon tetrachloride they are extraordinarily bright. Notice again how these lines and the hazy doublet still further out are repeated to the right of 4046, where they are interspersed with other lines which are primary lines unshifted; and again on both sides of 5461.

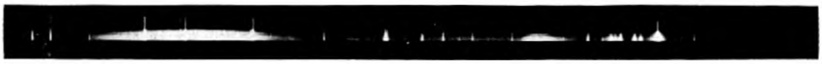


Fig. 6—Ultraviolet light scattered by sulphuric acid; at the right, 2536 accompanied by numerous shifted lines. (R. M. Langer.)

In Fig. 6 we pass to another region of the spectrum, the ultra-violet. On the extreme right is the strong line 2536 of the mercury spectrum, scattered unshifted by sulphuric acid; the numerous lines beside it are Raman lines.

In Fig. 7 the scattering substance is water; the novel feature of the scattered light is a set of diffuse bands, each shifted from a certain line of the primary spectrum. It is frequently observed that the shifted lines, and the unshifted lines as well, are broader and hazier than those of the primary light, an effect attributed to conversion of the energy of the quanta into energy of rotation of the molecules. Water however shows a quite remarkable broadening, if this be the

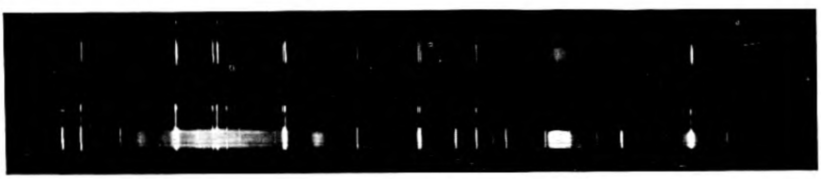


Fig. 7—Light scattered by water; note the diffuse bands shifted from the sharp lines. (R. M. Langer.)

way to describe it; it has been attributed to association of molecules.

A few special words must be said about the shifted lines which are of higher frequency than the primary light responsible for them—not because they are really more remarkable than the others, but because somehow they seem less to be expected. They are called by the monstrous name "anti-Stokesian" because in fluorescence such lines run counter to a principle laid down by Stokes; it seems cruel to perpetuate a mistake in this way. They consist of quanta which have received energy from the molecules which they have struck. They must therefore have struck molecules which were not originally in the state of lowest energy, or "lowest state" for short. Such a molecule might have been in the lowest state, until some antecedent quantum came along and gave it energy, lifting it into a higher state; the second quantum then undid the work of the first, taking away the energy which the first had given and being shifted as far towards higher as the former towards lower frequencies. It does not seem likely that this double process occurs often, although pairs of lines with equal and opposite shifts are reported in several cases. Much more commonly, in all probability, the molecules which at any moment are in other states than the lowest are there because of the interchanges of energy which are always taking place between the particles of substances in thermal equilibrium. The laws of thermal equilibrium are such, that if in a substance at room-temperature the molecules have one or more excited states differing from the lowest state only by fractions of a volt, quite an appreciable fraction among them are at any moment in one or another of those states. The higher the temperature, the greater this fraction; in consonance with which fact it is observed, that the warmer the scattering liquid the more prominent are these "anti-Stokesian" lines.

The shifted lines, in the light scattered at right angles to the primary beam (the only direction which has been utilized to any extent), are partially polarized. The electric vector is stronger in the direction perpendicular to the primary beam than in the direction parallel to it, as one would expect. The degree of polarization varies enormously from one shifted line to another, and may be either greater or less than that of the unshifted lines. Cabannes thought it to be constant for lines shifted by the same amount from different primary lines, but the ampler data of Carrelli do not seem to bear him out.

The test for "coherence" of the shifted light, which is made by bringing the scattering substance near to the state of "critical opalescence" where the irregularity of the arrangement of the atoms is greatest—the shifted lines should brighten *pari passu* with the unshifted, if their light is coherent—has been made by at least four people (Raman, Bogros and Rocard, Martin); but the results are oddly discordant.

One more most valuable service of the shifted lines remains to be mentioned. In the foregoing pages I have stressed the fact that some of them are known to agree, that is to say their frequency shifts are known to coincide, with the frequencies of lines of the infra-red spectrum of the scattering substance. But there are cases in which the infra-red spectrum of the scatterer has not yet been explored; and there we may deduce its lines from the frequency-shifts of the Raman lines in the visible spectrum. Moreover, there are regions of the infra-red spectrum which are very difficult to explore, because of such technical reasons as the insensitiveness of photographic plates; and the Raman lines make it possible to discover some at least of the features of these, by observations made in the most convenient region of the spectrum. Perhaps this will turn out to be the most fruitful of the consequences of Raman's discovery.

Scattering of Light with Transfer of Energy to Vibrations in Solids

The scattering of light with shift of frequency from solids was discovered, independently and almost simultaneously, by C. V. Raman and K. S. Krishnan in India and by G. Landsberg and L. Mandelstam in Russia. As seen on the photographs of the spectra of the scattered light, the effect is altogether like the Raman effect of liquids and vapors. The lines of the primary spectrum, scattered without change of frequency, are accompanied by companions shifted mostly towards lower, but in occasional cases towards higher frequencies. The infalling quanta therefore sometimes cede energy to quantized motions within the solid substance, and sometimes—but much less frequently—receive energy from these.

The first and obvious question is: do the frequency-shifts agree with lines of the infra-red spectra of the solid substance? For studying the infra-red spectra of solids there are, be it remembered, two classical methods. One is the familiar way of dispersing a beam of light which has traversed the solid, and looking for absorption lines or bands in its spectrum. To find a good dispersing-agent in the far infra-red is however not easy; and there is an alternative method, in which the beam of light is reflected several times over from samples of the solid. At each incidence of the beam upon the crystal, the waves of frequencies which do not coincide with natural frequencies of the substance go on through, while the waves of frequencies which do coincide are mostly reflected. Thus, after several reflections, the "residual" beam is composed of one or a few wave-lengths, those of the principal absorption-lines of the crystal; and these are measured by operating on the beam with a special interferometer, or in some other way. This is the method of "residual rays," or "Reststrahlen," which was developed and much exploited during the nineties of the last century and the opening years of this. The spectrum-lines of the crystalline substance are its "Reststrahlen"; and these are to be compared with the frequency shifts observed in and near the visible spectrum. Here is an instance: Landsberg and Mandelstam working with quartz observed frequency-shifts corresponding to infra-red lines of wave-lengths 9μ, 13.5μ, 21.5μ, 48μ, and 81μ respectively; there are Reststrahlen of wave-lengths 8.7, 12.8 and 20.7, while the other two wave-lengths cited lie in gaps of the infra-red spectrum unexplored as yet.

A photograph by F. G. Brickwedde [15] which I reproduce as Figure 8 illustrates this effect. The very broad black band is due to primary light of wave-length about 2536, scattered without change of fre-

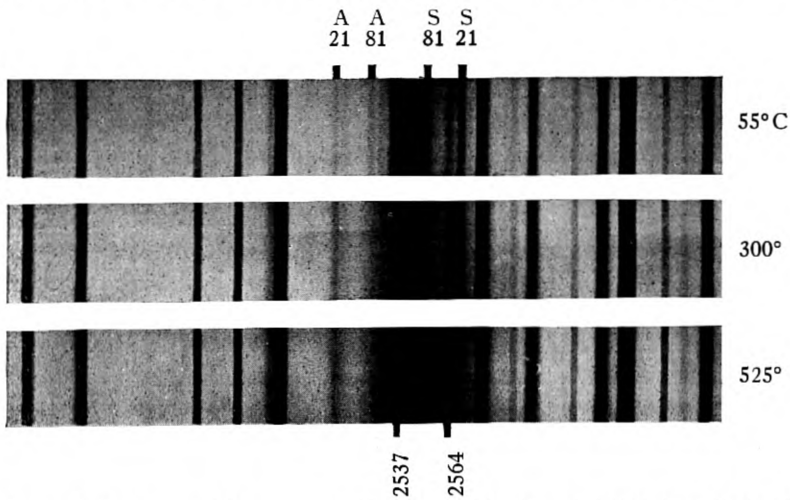


Fig. 8—Light scattered by a quartz crystal; unshifted line is 2536, lines denoted by S and A are shifted towards lower and higher frequencies respectively. (F. G. Brickwedde.)

quency; its excessive width is due to the great intensity of this light. The lines marked S21 and S81 consist of quanta which have spent part of their energy in exciting vibrations, those corresponding to the Reststrahlen of wave-lengths 21μ and 81μ respectively. The lines marked A21 and A81 consist of quanta which have received energy from these vibrations. These "Anti-Stokesian" lines evidently grow more intense as the temperature of the crystal is raised, as they should, for the reason which I have already stated: as the substance grows warmer, the percentage of the molecules spontaneously in vibration is increased. In addition, the shifted lines all move inward toward the position of the unshifted light as the temperature rises—one may see this by comparing those on the low-frequency side with

[15] I am much obliged to Dr. Brickwedde for furnishing me with a print of this photograph.

the line marked 2564, which is a faint line of the primary light scattered without change in frequency. This diminution in the shift signifies that the natural frequencies of the lattice-vibrations are declining, which is to be expected from the expansion and relaxation of the crystal which the rise of temperature brings about.

There arises now an interesting question. X-ray analysis reveals that there are certain chemical compounds, organic chiefly, of which the molecules retain their identity when crystallization occurs; they set themselves side by side in a regular lattice, but the arrangement of the atoms in each of them is not greatly altered. At the other extreme, there are compounds of which the molecules disintegrate completely when solidification takes place, and the atoms arrange themselves without any reminiscence of their earlier relations; a familiar instance is sodium chloride, in the crystal of which every atom of either kind, Na or Cl, is surrounded by six of the other kind all equally distant from it. Intermediate cases occur, as for instance that of $CaCO_3$, where each atom-group or "radical" CO_3 retains its identity but not its coupling to one single Ca atom. Now when molecules or radicals survive within the crystal, oscillations of atoms inside these atom-groups are probably not different in character from the oscillations which occur within the same molecules when they are wandering freely in a liquid or a gas. But in the case of a crystal like sodium chloride, the oscillations must be controlled by the forces which hold the atoms of the crystal together; they are truly lattice-vibrations. Perhaps the difference between the two is not really profound; but it will be interesting to find out whether quanta may or may not transfer energy with equal ease to vibrations of either type, *i.e.* whether in the two cases the shifted lines are comparably bright. According to Carrelli, Pringsheim and Rosen, all of the shifted lines thus far observed with solids, except probably those obtained with quartz, correspond to vibrations within molecules or radicals which remain intact in the crystal.

The Compton Effect

The Compton effect, in the restricted sense—the sense in which I shall use the term—is simply the scattering ensuing on collisions of corpuscles of light with free electrons. So much has been written about the effect [16] that it is scarcely necessary for me to do more in this place than mention the laws of these collisions. The energy which the quantum loses is converted into kinetic energy of translatory

[16] Cf. for instance this Journal, April, 1925; "Introduction to Contemporary Physics," pp. 146–160; H. Kallmann, H. Mark, "Ergebnisse der exakten Naturwissenschaften," *5* (1926).

motion of the electron. Momentum also is conserved in the encounter; the momentum of the quantum is equal in magnitude to hn/c before and to hn'/c after the collision, n and n' standing for the frequency before and after; these momenta are of course vectors parallel to the directions along which the corpuscle of light approaches and recedes, respectively; and their difference is the momentum which the electron acquires. These two conditions limit very severely the transfer of energy from quantum to electron. All of the quanta deflected through a given angle from their original line of flight suffer the same loss of energy and the same shift in frequency. The relation between shift of frequency and angle of deflection θ takes its simplest form when we write it as a relation between shift of wave-length, $\Delta\lambda$, and angle θ:

$$\Delta\lambda = \frac{h}{mc}(1 - \cos\theta).$$

The maximum shift of wave-length occurs when the quantum is reflected straight backward along its original line of approach; it is evidently $2h/mc$, and the corresponding maximum frequency shift is $2hn^2/mc^2$, as I stated earlier.

The predicted relation of wave-length shift and angle of scattering has been verified to the most thoroughgoing extent;[17] and the recoiling electrons have been observed as they dash off with the energy which the corpuscle of light has lost. Research on the Compton effect is now confined almost entirely to the problem of its likelihood of occurrence—*i.e.*, given a substance with P atoms per unit volume irradiated by a stream of X-rays composed of N quanta per unit area per second, what are the relations between the number of these quanta which are scattered in the fashion just described, and the frequency of the X-rays and the nature of the atoms? The corresponding problem for the Raman effect will undoubtedly soon come into the foreground. It is, of course, slightly annoying that we do not know *a priori* how many of the electrons of (say) the carbon atom, or how many of the electrons in a piece of graphite, are to be regarded as "free" electrons. This seems to be one of the facts which we shall have to deduce as best we may from these researches on the likelihood of the Compton effect. The data thus far acquired may be summarized in this way: the higher the frequency of the quanta, and the lower the mass of the atoms, the more abundant these collisions are. High-frequency X-rays poured

[17] Even to the point where with general assent it is taken for granted and the measurements are used as data for evaluating the constant m, the mass of the electron!

upon lithium or paraffin give the best "yield"; visible light, none perceptible.

The shifted X-rays scattered at 90° are polarized, very nearly completely (Lukirsky, Kallmann and Mark).

The Scattering of Quanta Attended by Extraction of Bound Electrons

Almost immediately after Arthur Compton had measured and interpreted the scattering which is due to collisions of quanta with free electrons, he and others realized that corpuscles of light might possibly encounter atoms in such a way that they extracted bound electrons, and thereupon were scattered with a corresponding abatement of their energy. In fact it seemed most probable that the electrons responsible for the Compton effect were themselves not quite free, but very lightly bound; and that a careful study of the scattered quanta, the "shifted" or "modified" line, would reveal that they had spent energy in dissolving the bonds as well as in imparting kinetic energy to the electrons. Researches on this topic were numerous, and are still continuing. At that time, the Raman effect had not yet been discovered; and perhaps it did not seem natural to accept the transfer of energy from corpuscles of light to atoms as a *general* phenomenon, apart from special cases so easy and so beautiful to visualize as the elastic impacts of quanta against free electrons. At all events, when in 1923 and 1924 data were published by G. L. Clark and W. Duane which to the present-day onlooker seem to declare the effect in the most forthright fashion, they made no such impression.

The experiments of Clark and Duane were involved in a long controversy, in which the reality even of the Compton effect was called into question. Data were obtained by some experimenters, which others could not or at least did not reproduce. The questions at issue speedily reached the point, where no outsider could risk a judgment unless he was himself a great expert in the study of X-ray scattering. Unfortunately the experiments were terminated, when the reality of the Compton effect was established. I say "unfortunately," for it now seems as if in the nature of things both sides must have been right. Compton had discovered the transfer of energy from quanta to free or nearly free electrons; Clark and Duane must have discovered the transfer of energy from quanta to bound electrons, or the process in which a corpuscle of light uses part of its energy in ionizing an atom, part in giving speed to the liberated electron, and retains the remainder. It would be a very desirable result of the present-day revival of interest in scattering, if somebody should reinvestigate this entire field.

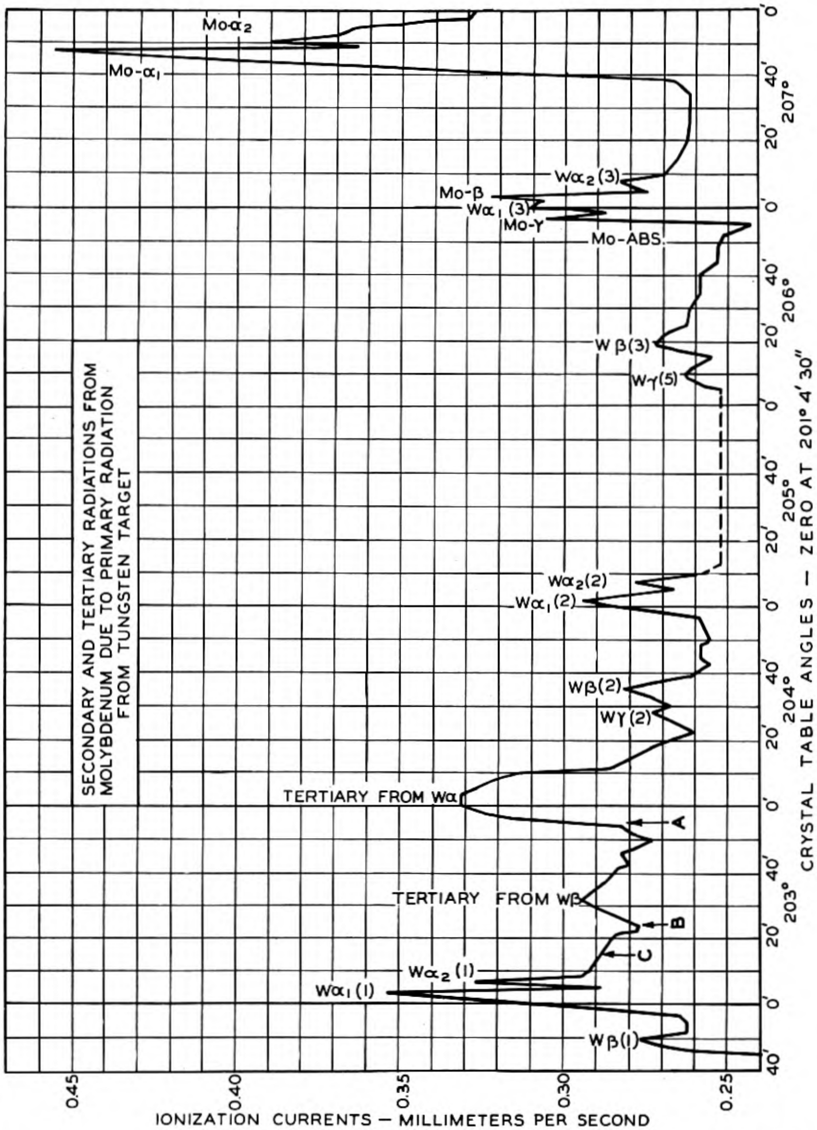


Fig. 9—X-rays scattered by molybdenum, the sharply-pointed peaks being unshifted lines. (G. L. Clark, W. Duane; *Proc. Nat. Acad. Sciences.*)

I will pass over the theory first propounded for these data, which I have quoted elsewhere, and over the paper in which Compton advanced practically the present theory; and reproduce one of the spectrum-curves of scattered radiation from the work of Duane and Clark. This curve represents the scattering of high-frequency X-rays by molybdenum. The primary light was composed of several lines, the various lines of the K-spectrum of tungsten. In the scattered spectrum, these lines appear unshifted as narrow sharp peaks. We consider the two at the extreme left, and then the two humps which rise from the points marked B and A. These points correspond to the wave-lengths of quanta which originally belonged to the spectrum-lines at the extreme left, and have lost exactly the energy necessary to extract a K-electron from the molybdenum atom. The humps in all probability are composed of quanta, which have extracted such electrons and in addition have given them greater or smaller amounts of extra kinetic energy.

Years after the work of Clark and Duane had been discontinued, Bergen Davis and D. P. Mitchell undertook to study what they designated as the "fine structure" of the lines in the spectra of scattered X-rays. They had many improvements of technique at their disposition, improvements many of which were due to Davis himself; for instance they had a spectrometer of Davis' design, by which it was possible to appreciate the true narrowness of a very narrow X-ray line, instead of having it spread out by defects of the apparatus into a simulation of a wide band. They irradiated graphite with the spectrum-line known as $K\alpha_1$ of molybdenum, of which the wave-length is 0.721 Å.; and in the spectrum of the rays scattered at 90° they found not only this line, but four others of slightly greater wave-lengths of which three are shown in Fig. 10. The outermost, beyond the right-hand limit of the picture, comprises quanta which collided with free electrons —it is the Compton shifted line. The outermost of the other three consists of incident quanta which have given up just the amount of energy required to extract one of the K electrons—one of the pair which are by far the most tightly bound of all the six which belong to the carbon atom. More precisely, the extraction energy of these electrons is evaluated from the X-ray spectrum of carbon at 287 equivalent volts, while the loss of energy suffered by the quanta is estimated by Davis and Mitchell at 279; the difference of less than 4 per cent is within the uncertainty of experiment. The two innermost of the shifted lines, composed as they are of quanta which have yielded up 29 and 50 equivalent volts respectively, are presumably tokens of collisions in which superficial electrons were torn away from carbon atoms. That their shifts do not agree very well with the values of extraction-energies suggested by certain spectroscopic data is not in the least surprising. None of the spectroscopic data was obtained with solid carbon; and the superficial electrons of the atom are most sensitive of all to such changes of environment as occur when it is incorporated into a lattice. In all probability, the frequency-shifts of these lines, when divided by h, are the best values yet available for the amounts of energy required to extract superficial electrons from carbon atoms in the graphite lattice.[18]

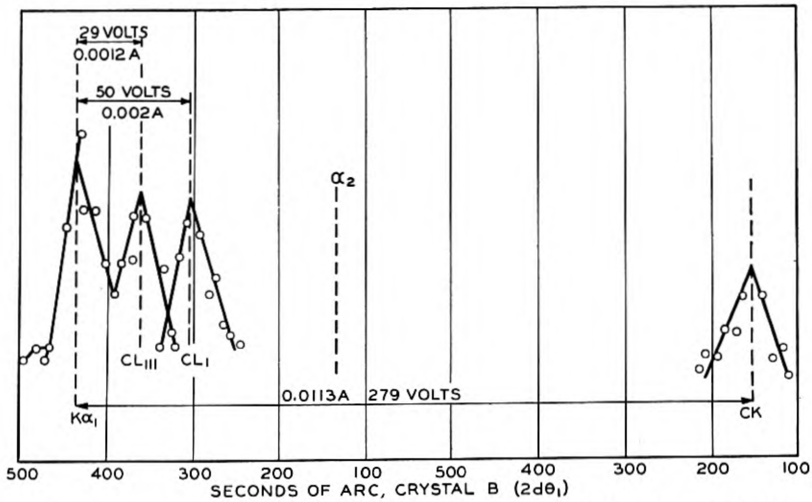


Fig. 10—X-rays scattered by graphite, the peak on the left being the unshifted line. (B. Davis, D. P. Mitchell; *Physical Review.*)

Davis and Mitchell point out a feature of these scattered rays which should be stressed: the shifted lines are sharp, implying that the quanta simply extract electrons, without endowing them with divers quantities of kinetic energy to boot. They later observed such lines in the rays scattered by aluminium.

The General Principle

The general principle of which all these scattered phenomena of scattering are special illustrations now stands forth very clearly, and

[18] It seems paradoxical that the quanta which collide with free electrons should confer on them several hundred of equivalent volts of energy, while those which strike these loosely-bound superficial electrons should lose but half-a-dozen. The contrast is due to the requirement of conservation of momentum; where only two particles are involved the relatively large energy-transfer is entailed, but where there are three among which the momentum may be divided, the limitation ceases.

may be expressed in any of several ways, laying greater emphasis now on one aspect and now on another.

From the viewpoint of the atom, and using the notions of the undulatory theory, one may say that *the atom (or the molecule) modulates the incident light with frequencies of its own.*

Again from the viewpoint of the atom, but using now the notions of the corpuscle-theory of light, one may say that *the atom or molecule may take part but not necessarily all of the energy of an incident quantum, converting this energy in any of numerous ways.*

From the viewpoint of the quantum, however, the essential feature of the principle is this: *a quantum may lose part of its energy or receive energy in an encounter with a molecule or atom, retaining its identity even though its frequency is changed.*

The first who stated the principle with anything like its proper generality was probably Smekal; in the following year (1924) it was developed by Kramers and Heisenberg. They knew of no examples but the Compton effect, and curiously enough no one was tempted to search for other instances, though Foote and Ruark considered whether any of the phenomena already known in optical spectra could be related to it.[19] Partial anticipations crop out here and there, especially in the work of Compton, Jauncey and their associates; for it was early suspected that the electrons responsible for the Compton shift are not altogether free, but very loosely bound to atoms; it was assumed that the incident quanta must spend energy enough to break the bonds as well to set the electrons into motion, and efforts were made to disclose this breaking of the bonds.[20]

It is the third of the foregoing formulations of the principle which I wish to stress in closing—the principle from the viewpoint of the quantum, the authorization of the quantum to give up part of its energy and retain the rest. To the unprejudiced mind this must seem very natural. We have accepted for years the principle that an electron may give up part of its energy and keep the rest—that the life-history of an electron is an endless sequence of gains and losses of kinetic energy, of speedings-up and slowings-down, during which the identity of the electron is never lost. Why should we not have thought likewise about the quantum? Yet it has been almost an article

[19] I am told that Kramers tried vainly to persuade a number of experimental physicists to look for the effect. At present they must be feeling like the astronomers whom Adams vainly pressed to make haste in looking for the planet Neptune, until finally someone else discovered it.

[20] Jauncey and Compton anticipated in 1927 the idea that atoms in a lattice may acquire energy of vibration from incident quanta, and discovered an important restriction which should be noted; apparently the lattice or some third particle must be involved in the impact.

of faith that a quantum must give all of its energy, or none—either vanish altogether, or retain its frequency unchanged.

Of course, till 1922 there was no compelling evidence that a corpuscle of light may suffer a change of frequency in rebounding from a particle of electricity or matter. However, it does not seem to have occurred to anyone that the want of evidence was in any way surprising, or that it should be possible to find quanta scattered with change of energy. The reason for this satisfaction, I suspect, was perfectly simple. It did not seem possible that a quantum should give up part of its energy, for its energy was inseparably linked with its frequency, and its frequency seemed to be its one indissoluble and characteristic feature. As well say that an electron might lose part of its charge and still be the same electron, or that an atom might lose part of its mass and yet remain the same atom, as that a quantum might give up part of its frequency without ceasing to be itself!

Now this contention—if one may call it a contention—lost its force through the discovery that electrons also are endowed with frequency and wave-length, or in other words that negative electricity like light possesses both qualities of corpuscles and qualities of waves. Whenever a corpuscle of electricity parts with kinetic energy, whenever a corpuscle of light parts with energy, the associated wave-length is augmented. If we suppose that an electron retains its identity when its wave-length changes, how can we deny like continuity of existence to a quantum? If we admit that an electron may suffer change of wave-length in rebounding from an atom, how may we be surprised when a quantum does the like? It is true that the corpuscle of electricity has other features than wave-length: a charge which apparently never changes, a mass which apparently never falls below a certain minimum. The quantum does not have an immutable quality corresponding to charge, and we do not know of any lower limit to its mass short of complete disappearance. But for either sort of corpuscle the wave-length is in principle variable. We say that all electrons are of one kind, but may have any speed. Should we not also say that all quanta are of a single kind, though they may have any frequency?

BIBLIOGRAPHY

(The year is 1928, unless otherwise stated)

S. K. Allison, W. Duane: Proc. Nat. Acad. Sci., *10*, pp. 196–199 (1924).
J. A. Becker: Proc. Nat. Acad. Sci., *10*, pp. 342–345 (1924).
C. E. Bleecker: Zs. f. Phys., *50*, pp. 781–786.
A. Bogros, Y. Rocard: C. R., *186*, pp. 1712–1713.
M. Born: Naturwiss., *16*, p. 673.
F. G. Brickwedde, M. F. Peters: Bull. Amer. Phys. Soc., Nov. 15.
M. de Broglie: C. R., *187*, p. 697.
J. Cabannes: C. R., *186*, pp. 1201–1202, 1714–1715; *187*, pp. 654–656.

J. Cabannes, P. Daure: C. R., *186*, pp. 1533–1534.
A. Carrelli, P. Pringsheim, B. Rosen: Zs. f. Phys., *51*, pp. 511–519.
G. L. Clark, W. Duane: Proc. Nat. Acad. Sci., *9*, pp. 413–424 (1923); *10*, pp. 41–47, 92–96, 148–152, 191–196 (1924).
A. H. Compton: Phys. Rev., (2) *24*, pp. 168–177 (1924).
A. Cotton: C. R., *186*, pp. 1475–1476.
B. Davis, D. P. Mitchell: Phys. Rev., (2) *31*, p. 1119; *32*, pp. 331–335; Bull. Am. Phys. Soc., Dec. 12.
P. Daure: C. R., *186*, pp. 1833–1835; *187*, pp. 940–941.
K. K. Darrow: Science, *68*, pp. 488–490.
P. D. Foote, A. E. Ruark: Science, *61*, pp. 263–264 (1925).
R. V. L. Hartley: Bull. Am. Phys. Soc., Dec. 12.
G. E. M. Jauncey, A. H. Compton: Nature, *120*, p. 549 (1927).
H. Kallmann, H. Mark: Zs. f. Phys., *36*, pp. 120–142 (1926).
E. H. Kennard: Bull. Amer. Phys. Soc., Dec. 12.
H. Kornfeld: Naturwiss., *16*, p. 653.
H. A. Kramers, W. Heisenberg: Zs. f. Phys., *31*, pp. 681–706 (1925).
H. S. Krishnan: Nature, *122*, pp. 477–478, p. 650.
G. Landsberg, L. Mandelstam: Naturwiss., *16*, pp. 557–558; C. R., *187*, pp. 109–111.
R. M. Langer, W. F. Meggers: Bull. Amer. Phys. Soc., Nov. 15.
F. A. Lindemann, T. C. Keeley, N. R. Hall: Nature, *122*, p. 921.
P. Lukirsky: Nature, *122*, pp. 275–276.
W. H. Martin: Nature, *122*, pp. 506–507.
P. Pringsheim: Naturwiss., *16*, pp. 597–605; Zs. f. Phys., *50*, pp. 741–755.
I. Ramakrishna Rao: Indian Jour. of Phys., *3*, pp. 123–130.
C. V. Raman: Indian Jour. of Phys., *2*, pp. 387–398.
C. V. Raman, K. S. Krishnan: Nature, *121*, pp. 501–502, 619, 711; *122*, pp. 12–13, 169, 278; Indian Jour. of Phys., *2*, pp. 398–420.
L. A. Ramdas: Nature, *122*, p. 57; Indian Jour. of Phys., *3*, pp. 131–136.
Y. Rocard: C. R., *186*, pp. 1107–1109; Annales de Physique, *10*, pp. 116–232.
A. E. Ruark: Nature, *122*, pp. 312–313.
M. N. Saha, D. S. Kothari, G. R. Toshniwai: Nature, *122*, p. 398.
A. Smekal: Naturwiss., *11*, pp. 873–875 (1923); *16*, pp. 612–613 (1928).
S. Venkateswaran: Indian Jour. of Phys., *3*, pp. 105–122.
D. L. Webster: Proc. Nat. Acad. Sci., *10*, pp. 191–196 (1924).
R. W. Wood: Nature, *122*, p. 349; Phil. Mag., (7) *6*, pp. 729–742.

Ground Return Impedance: Underground Wire with Earth Return

By JOHN R. CARSON

SYNOPSIS: In certain transmission problems principally those relating to induction and interference phenomena, it is necessary to know the transmission characteristics of a circuit composed of an underground wire with earth return. These can be evaluated by well known engineering formulas provided the ground return impedance is known. The present paper gives the mathematical solution of this problem and shows that the ground return impedance is substantially independent of the depth of the wire below the surface.

THE object of this note is to give the solution of a problem of considerable interest and practical importance which does not appear to have been solved heretofore; this is the "ground return" impedance, per unit length, of a circuit composed of an underground wire or cable with earth return.

The physical system and the problem may be more explicitly described and explained as follows: An underground wire or cable parallel to and at depth h below the surface of the ground is surrounded by a concentric dielectric cylinder of external radius a. The earth then forms the return path for currents flowing in the wire. The ground return impedance Z_g is then defined as the ratio of the mean axial electric intensity at the external surface of the dielectric sheath to the current flowing in the wire.

When the earth extends indefinitely in all directions about the wire so that circular symmetry obtains, the problem is quite simple, and the formula for the ground return impedance, denoted in this case by Z_g^0, is well known. In practice, however, we are interested principally in the case where the wire is close to the surface of the earth, so that the distribution of return current in the ground is anything but symmetrical. For this case the formula for the ground return impedance, which it is the object of this note to state and discuss, is

$$Z_g = (1 + c)Z_g^0. \tag{1}$$

Here the correction term c, which takes care of the departure from circular symmetry, is given by

$$c = \frac{1}{K_0(a'i\sqrt{i})}\int_0^\infty \frac{\sqrt{\mu^2 + i} - \mu}{\sqrt{\mu^2 + i} + \mu} \cdot \frac{e^{-2h'\sqrt{\mu^2+i}}}{\sqrt{\mu^2 + i}} d\mu. \tag{2}$$

In formula (2),

$a' = a\sqrt{\alpha}$,

$h' = h\sqrt{\alpha}$,

$\alpha = 4\pi\lambda\omega$ where λ is the conductivity of the ground in elm. c.g.s. units, ω is 2π times the frequency,

$i = \sqrt{-1}$.

$K_0(a'i\sqrt{i})$ is the Bessel Function of the second kind; it is equal to $\frac{i\pi}{2} H_0^{(1)}(a'i\sqrt{i})$ where $H_0^{(1)}$ is the Hankel function as defined by Jahnke u. Emde in their Funktionentafeln. Denoted by ker a' + i kei a' the function $K_0(a'i\sqrt{i})$ has been computed and tabulated by the British Association. The only restriction on formula (2) is that the radius a is supposed small compared with the depth h.

Now the ground conductivity λ lies between 10^{-14} and 10^{-12}, while the depth h will not in practice exceed a few meters ($h \leq 10^3$). Under such circumstances, at ordinary frequencies, h' will be exceedingly small compared with unity, and a' still smaller. Consequently in evaluating the infinite integral in (2), it is permissible to take $e^{-2h'\sqrt{\mu^2+i}}$ as unity, since for $\mu > 2$, the rest of the integrand converges as $i/4\mu^3$.

Now we have

$$\int_0^\infty \frac{\sqrt{\mu^2+i}-\mu}{\sqrt{\mu^2+i}+\mu}\cdot\frac{d\mu}{\sqrt{\mu^2+i}} = \frac{1}{i}\int_0^\infty \left\{ \sqrt{\mu^2+i} - 2\mu + \frac{\mu^2}{\sqrt{\mu^2+i}} \right\} d\mu = \frac{1}{2}$$

and hence c of formula (2) becomes

$$c = \frac{1}{2}\,\frac{1}{K_0(a'i\sqrt{i})}.$$

Furthermore since a' by hypothesis is very small compared with unity, we can replace K_0 by its limiting form for vanishingly small arguments which is approximately

$$\log\,(1/a').$$

We thus get, finally, the approximate formula, valid for most practical applications,

$$Z_g = \left\{ 1 + \frac{1}{2\log\,(1/a')} \right\} Z_g^0. \tag{3}$$

The interesting and somewhat surprising feature of this formula is that the value of the correction term $1/2 \log\,(1/a')$ likely to occur in practical applications amounts at most to 0.05 to 0.10. On the other hand, with the wire close to the surface of the ground, the conducting area of the ground return path is just one half the area available when the ground extends indefinitely in all directions and the return impedance is Z_g^0. In other words, the departure from circular symmetry means only a very small increase in the ground return impedance. In fact this increase is so small and the ground conductivity actually so variable, the correction is hardly justified by the precision of the data, so that, in most engineering applications, we may take Z_g as equal to Z_g^0 with an error probably less than that involved in other factors, and lack of precision in data.

Derivation of Preceding Formulas

The derivation of the preceding formulas is not without interest. Since, however, this derivation is, in general, an adaptation of the methods employed in my paper 'Wave Propagation in Overhead Wires with Ground Return' (*B. S. T. J.*, Oct., 1926) it will be outlined rather than given in detail.

Take the axis of the wire as the origin and Y as the vertical axis; then the surface of the ground is the plane $y = h$. Let a unit current flow in the wire and take the axis of the wire as the Z axis. In the ground ($\rho = \sqrt{x^2 + y^2} \geqq a$) the axial electric intensity will be written

$$E = \frac{Z_g^0}{K_0(a'i\sqrt{i})} K_0(\rho' i\sqrt{i}) + E' = E^0 + E', \tag{4}$$

where $\rho' = \sqrt{\alpha}\sqrt{x^2 + y^2}$ and K_0 is the Bessel function of the second kind, related to the Hankel function by the equation

$$K_0(a'i\sqrt{i}) = \frac{\pi i}{2} H_0^{(1)}(a'i\sqrt{i}).$$

The first term on the right hand side of (4) represents the circularly symmetrical distribution which would alone exist if the surface of the ground were removed to an infinite distance, while E' is a secondary distribution due to reflection at the surface of the earth ($y = h$). Inspection of equation (4) shows that when $\rho = a$, E is the required return impedance Z.

Strictly speaking E^0 should be written as

$$E^0 = \frac{Z_g^0}{K_0(a'i\sqrt{i})} \{K_0(\rho' i\sqrt{i}) + h_1 K_1(\rho' i\sqrt{i})\cos\theta + h_2 K_2(\rho' i\sqrt{i}) \cos 2\theta + \cdots\}, \tag{5}$$

the harmonic terms representing secondary reflection at the surface of the dielectric cylinder ($\rho = a$). If a is made sufficiently small, however, the harmonic terms become negligible. In view of this fact, the large amount of tedious additional analysis required, if the harmonic terms are retained, is not believed to be justified by the practical applications contemplated. E^0 will therefore be taken as in formula (4).

The secondary electric intensity E' can always be written as the Fourier integral

$$E' = \int_0^{\infty} F(\mu) e^{y'\sqrt{\mu^2+i}} \cos x'\mu \, d\mu, \quad 0 \leqq y \leqq h, \quad (6)$$

where $x' = x\sqrt{\alpha}$, $y' = y\sqrt{\alpha}$, and the Fourier function F is to be determined. For the formulation of the boundary conditions at $y = h$ we also require the expansion of $K_0(\rho' i\sqrt{i})$ as a Fourier integral; the required expansion is *

$$K_0(\rho' i\sqrt{i}) = \int_0^{\infty} \frac{1}{\sqrt{\mu^2+i}} e^{-y'\sqrt{\mu^2+i}} \cos x'\mu \, d\mu, \quad \rho > 0. \quad (7)$$

In the dielectric, the magnetic forces H_x, H_y are taken as

$$\left.\begin{aligned} H_x &= \int_0^{\infty} \phi(\mu) e^{-y\mu} \cos x\mu \, d\mu \\ H_y &= -\int_0^{\infty} \phi(\mu) e^{-y\mu} \sin x\mu \, d\mu \end{aligned}\right\} \quad y \geqq h. \quad (8)$$

In the ground, on the other hand, we have

$$\begin{aligned} i\omega H_x &= -\frac{\partial}{\partial y} E, \\ i\omega H_y &= \frac{\partial}{\partial x} E. \end{aligned} \quad (9)$$

In order to satisfy the boundary conditions at the plane $y = h$, we equate H_x as given by (8) and (9), and H_y as given by (8) and (9). The explicit formulas for H_x and H_y are derived from (9) by substituting the Fourier integral for K_0, as given by (7), in (4) and differentiating as indicated.

The two equations resulting from equating the two expressions for H_x and the two expressions for H_y can be solved for the Fourier function

* See note at end of this paper.

$F(\mu)$. With this determined the required impedance Z_g is simply by (4)

$$Z_g = Z_g^0 + E_{\rho'}' = 0 \tag{10}$$

on the assumption that $a' = a\sqrt{\alpha}$ is quite small compared to unity. This gives formula (1).

Note: The expansion (7), which is believed to be novel, was derived by a limiting process rather too long and unsuitable for inclusion in this paper. It and the following additional expansions are quite useful in certain problems on wave propagation.

$$\cos\theta \cdot K_1(\rho' i\sqrt{i}) = -\sqrt{i}\int_0^\infty e^{-y'\sqrt{\mu^2+i}}\cos x'\mu\, d\mu,$$

$$\sin\theta \cdot K_1(\rho' i\sqrt{i}) = -\sqrt{i}\int_0^\infty \frac{\mu}{\sqrt{\mu^2+i}} e^{-y'\sqrt{\mu^2+i}}\sin x'\mu\, d\mu,$$

where

$$\cos\theta = y/\rho = y/\sqrt{x^2+y^2},$$

$$\sin\theta = x/\rho = x/\sqrt{x^2+y^2}.$$

Application to the Binomial Summation of a Laplacian[1] Method for the Evaluation of Definite Integrals

By E. C. Molina

Introduction

THE numerical evaluation of the incomplete Binomial Summation, a problem of major importance for many statistical and engineering applications of the Theory of Probability, is a question for which a satisfactory solution has not as yet been obtained. Several approximation formulas have been presented,[2] each of which gives good results for some limited range of values of the variables involved; but a formula of wide applicability is still a desideratum.

The purpose of this paper is to submit for consideration an approximation formula which seems to meet the situation to a measurable extent. The writer derived it by applying to the equation

$$(1)\qquad \sum_{x=c}^{x=n}\binom{n}{x}p^x(1-p)^{n-x}=\frac{\displaystyle\int_0^p x^{c-1}(1-x)^{n-c}dx}{\displaystyle\int_0^1 x^{c-1}(1-x)^{n-c}dx},$$

a method which is peculiarly efficacious for approximately evaluating definite integrals when the integrands contain factors raised to high powers.

The method used constitutes the subject matter of Chapter I, Part II, Book I of Laplace's "Théorie Analytique des Probabilités." Poisson applied the method to the integrals in the equation

$$(2)\qquad \sum_{x=c}^{x=n}\binom{n}{x}p^x(1-p)^{n-x}=\frac{\displaystyle\int_{(1-p)/p}^{\infty} x^{n-c}/(1+x)^{n+1}dx}{\displaystyle\int_0^{\infty} x^{n-c}/(1+x)^{n+1}dx}$$

and published a first approximation, together with its derivation, in his "Recherches sur la Probabilité des Jugements." Poisson's approximation seems never to have been used and was less fortunate than his famous limit to the binomial expansion which also was lost sight of until it reappeared under the caption "law of small numbers."

[1] Presented before International Congress of Mathematicians at Bologna, Italy in September, 1928.

[2] For an excellent resumé of some well-known formulas, together with a discussion of their limitations, reference may be had to C. Jordan, "Statistique Mathématique," articles 37 and 38.

While the integrals in equations (1) and (2) are well known equivalent forms for the complete and incomplete Beta functions, the equations themselves are not so familiar although one or the other will be found in Laplace, Poisson, Boole (Differential Equations) and at least two other places.

Approximate Formula

The approximate formula derived from equation (1) and submitted herewith for consideration is

$$(3) \qquad \sum_{x=c}^{x=n} \binom{n}{x} p^x(1-p)^{n-x} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{T} e^{-t^2} dt - \frac{S_i e^{-T^2}}{2\sqrt{\pi}},$$

where S_i is the ith approximation to the infinite series

$$(4) \quad S = \frac{\sum_{s=1} R_s T^{s-1}[1 + (s-1)T_1^{-2} + (s-1)(s-3)T_1^{-4} \cdots]}{1 + \sum_{s=1} R_{2s}[1\cdot 3\cdot 5 \cdots (2s-1)]2^{-s}},$$

$T_1 = T\sqrt{2}$,

$$(5) \quad T^2 = (n-1)\log\frac{n}{n-1} + (c-1)\log\frac{c-1}{a} + (n-c)\log\frac{n-c}{n-a},$$

and $a = np$; T to be taken negative when $a < (c-1)n/(n-1)$.

The first, second and third approximations to the infinite series S are

$$S_1 = R_1, \qquad S_2 = \frac{R_1 + R_2 T}{1 + R_2/2}, \qquad S_3 = \frac{R_1 + R_2 T + R_3(1 + T^2)}{1 + R_2/2},$$

where

$$R_1 = 4[(n-c) - (c-1)]/3\sqrt{2(n-1)(n-c)(c-1)},$$
$$R_2 = (1/6)[1/(n-c) + 1/(c-1) - 13/(n-1)],$$
$$R_3 = -(4/15)R_1[R_2 + 6/(n-1)].$$

It will be noted that R_2, $|R_1|$ and $|R_3|$ are symmetric functions of $(n-c)$ and $(c-1)$.

For the limiting case (Poisson's Exponential Binomial Limit) where $n = \infty$, $p = 0$ but $np = a$, we have

$$T^2 = 1 + (c-1)\log(c-1)/a + (a-c),$$
$$R_1 = 4/3\sqrt{2(c-1)},$$
$$R_2 = 1/6(c-1),$$
$$R_3 = -(4/15)R_1R_2.$$

Numerical Results

Since it is easy to compute the binomial summation directly when either c or $n - c$ is small, the practical value of an approximate formula depends on its efficiency for large values of these quantities.

The analysis given below under the heading "Derivation of the Approximate Formula" indicates that the successive R_s's in the series for S decrease when $\sqrt{c-1}$ and $\sqrt{n-c}$ increase. Therefore, when these two quantities are large, a few terms of the approximate formula (3) may be expected to give satisfactory results. As a matter of fact, the formula gives good results when $\sqrt{c-1}$ and $\sqrt{n-c}$ are not large. To confirm this statement the Tables [3] given at the end of this paper are submitted. In the 4th column of each table are given 10^6 times the true values of

$$P = \sum_{x=c}^{x=n} \binom{n}{x} p^x (1-p)^{n-x}.$$

In the columns headed Δ_1, Δ_2 and Δ_3 are given 10^6 times the differences between the true values and those obtained by applying formula (3) with the first, second and third approximations to S respectively. Table I in Czuber's "Wahrscheinlichkeitsrechnung" was used for evaluating the probability integral in equation (3).

The range of values of P covered by the tables is such that at the lower end of each section $P \ngtr .0005$ while at the upper end $P \nless .9995$, except where this latter condition would call for a value of $c < 2$. Of course, a larger or smaller range might have been given. The decision as to this question was based on the fact that several writers on the theory of statistics, when dealing with the normal law of errors, speak of an error exceeding 3 or 4 times the standard deviation as being a very improbable event. In order to keep the number of pages required for the tables within reasonable bounds computations were made only for even values of c.

The values of $a = np$ used are such that each of the values $p = 1/2$, $p = 1/10$ and $p = 1/20$ occurs twice; likewise each of the values $n = 100$, $n = 50$ and $n = 30$ occurs twice.

A greater degree of accuracy than that indicated by the tables can, of course, be obtained by working out and using R_4, $R_5 \cdots$; for this purpose, recourse should be had to equation (12) below and the details immediately following it. The only practical limitation to the use of formula (3) would appear to be the number of places given

[3] I am greatly indebted to Miss Nelliemae Z. Pearson of the Department of Development and Research both for supervising the work of my computers and contributing personally several sections of the tables.

by the existing tables for the probability integral. However, this difficulty is encountered only when P, or $(1 - P)$, is small, in which case T is large and the integral

$$\int_{-\infty}^{T} e^{-t^2} dt$$

may be readily evaluated by computing the first few terms of the series

$$[e^{-T^2}/2T\sqrt{\pi}][1 - T_1^{-2} + (1\cdot 3)T_1^{-4} - (1\cdot 3\cdot 5)T_1^{-6} \cdots],$$

where, as above, $T_1 = T\sqrt{2}$.

When P is very small, the difference $c - a = c - np$ is relatively large compared to a, and for this latter case recourse may be had to the approximate formula published by the writer in the *American Mathematical Monthly* for June, 1913.

Derivation of the Approximate Formula

Following Laplace closely, let us set

$$y(x) = Ye^{-t^2}, \tag{6}$$

where $Y = y(x_0)$ is the maximum value of $y(x)$. Then

$$\int_0^p y dx = Y \int_{-\infty}^{T} e^{-t^2} \left(\frac{dx}{dt}\right) dt, \tag{7}$$

the upper limit T being given by the equation

$$y(p) = y(x_0) e^{-T^2}. \tag{8}$$

Assuming dx/dt expanded in powers of t so that

$$dx/dt = \sum_{s=0} D_{s+1} t^s \tag{9}$$

and setting $R_s = D_{s+1}/D_1$, equation (7) reduces to

$$\int_0^p y dx = YD_1 \sum_{s=0} R_s \int_{-\infty}^{T} t^s e^{-t^2} dt.$$

Our fundamental equation (1) may now be written

$$\sum_{x=c}^{x=n} \binom{n}{x} p^x (1 - p)^{n-x} = \frac{\sum_{s=0} R_s \int_{-\infty}^{T} t^s e^{-t^2} dt}{\sum_{s=0} R_s \int_{-\infty}^{\infty} t^s e^{-t^2} dt}. \tag{10}$$

Integrating by parts and separating the terms involving $\int e^{-t^2}dt$ from the terms containing e^{-t^2}, we obtain equations (3) and (4).

To determine $R_s = D_{s+1}/D_1$, note that equation (6) gives $t = (\log Y - \log y)^{1/2}$ and set $v(x) = (x - x_0)/(\log Y - \log y)^{1/2}$ so that x may be written in the form

$$x = x_0 + v(x)t.$$

This form for x gives the expansion (Lagrange's Theorem for the simple case where $f(x) = x$; see "Modern Analysis" by Whittaker and Watson)

$$x = \sum_{s=0} \frac{t^s}{s!}\left(\frac{d^{s-1}v^s}{dx^{s-1}}\right)_{x=x_0}.$$

Comparing this expansion for x with the previous expansion (9) for dx/dt, we obtain

$$D_1 = v(x_0)$$

and

$$\frac{D_{s+1}}{D_1} = R_s = \left(\frac{1}{s!v(x)}\cdot\frac{d^s v^{s+1}}{dx^s}\right)_{x=x_0}.$$

Up to this point no particular form has been attributed to the function $y(x)$. From now on we deal with the function which constitutes the integrand of the integrals in equation (1).

The function $y(x) = x^{c-1}(1 - x)^{n-c}$ gives the expansion $(\log Y - \log y) = (x - x_0)^2[A_0 + A_1(x - x_0) + A_2(x - x_0)^2 \cdots]$, where $x_0 = (c - 1)/(n - 1)$ is the value of x for which $y(x)$ is a maximum and

$$A_s = \frac{1}{(s+2)!}\left[\frac{d^{s+2}(\log Y - \log y)}{dx^{s+2}}\right]_{x=x_0}$$

or

$$(11) \qquad A_s = \frac{(n-1)^{s+2}}{s+2}\left[\left(\frac{1}{n-c}\right)^{s+1} + (-1)^s\left(\frac{1}{c-1}\right)^{s+1}\right].$$

We are now prepared to evaluate R_s. Set

$$g = A_0 + A_1(x - x_0) + A_2(x - x_0)^2 \cdots$$

and

$$g_s = d^s g/dx^s.$$

Then

$$v = g^{-1/2},$$

$$\frac{dv^2}{dx} = -g^{-2}g_1,$$

$$\frac{d^2v^3}{dx^2} = (3/2)g^{-7/2}[(5/2)g_1^2 - g_2 g],$$

$$\frac{d^3v^4}{dx^3} = -2g^{-5}[g_3g^2 - 9g_2g_1g + 12g_1^3].$$

Therefore, since $g_s = s!A_s$ when $x = x_0$,

$$R_1 = -A_0^{-3/2}A_1,$$
$$R_2 = (3/2)A_0^{-3}[(5/4)A_1^2 - A_0A_2],$$
$$R_3 = -2A_0^{-9/2}[A_3A_0^2 - 3A_2A_1A_0 + 2A_1^3].$$

Substituting for A_0, A_1, A_2 and A_3 the expressions derived by giving s the values 0, 1, 2 and 3 respectively in equation (11), we obtain for R_1, R_2 and R_3 the functions of n and c given on page 2.

For values of s greater than 3 the direct evaluation of $d^s v^{s+1}/dx^s$ by successive differentiation becomes very tedious. It will be found much more practical to use the following procedure,[4] where D is a symbol of operation, $A = A_0$ and $b = A_1$.

$$A_0^{-1/2}R_s = (1/s!)\left(\frac{d^s g^{-(s+1)/2}}{dx^s}\right)$$
$$= \left[\frac{dA^{-(s+1)/2}}{1!dA}\right]D^{s-1}b + \left[\frac{d^2A^{-(s+1)/2}}{2!dA^2}\right]D^{s-2}b^2 + \cdots$$
$$+ \left[\frac{d^{s-1}A^{-(s+1)/2}}{(s-1)!dA^{s-1}}\right]Db^{s-1} + \left[\frac{d^sA^{-(s+1)/2}}{s!dA^s}\right]b^s$$

or

$$R_s = A_0^{1/2}\sum_{m=1}^{m=s}\left[\frac{d^mA^{-(s+1)/2}}{m!dA^m}\right](D^{s-m}b^m). \tag{12}$$

The following equations give the details requisite for the formation of R_s to R_8 inclusive; A_s can be computed from equation (11).

$Db = A_2$, $D^2b = A_3$, $D^3B = A_4$, $D^4b = A_5$,

$D^5b = A_6$, $D^6b = A_7$, $D^7b = A_8$,

$Db^2 = 2A_1A_2$,

$D^2b^2 = 2A_1A_3 + A_2^2$,

$D^3b^2 = 2A_1A_4 + 2A_2A_3$,

$D^4b^2 = 2A_1A_5 + 2A_2A_4 + A_3^2$,

$D^5b^2 = 2A_1A_6 + 2A_2A_5 + 2A_3A_4$,

$D^6b^2 = 2A_1A_7 + 2A_2A_6 + 2A_3A_5 + A_4^2$,

$Db^3 = 3A_1^2A_2$,

$D^2b^3 = 3A_1^2A_3 + 3A_1A_2^2$,

$D^3b^3 = 3A_1^2A_4 + 6A_1A_2A_3 + A_2^3$,

[4] See DeMorgan's "Differential and Integral Calculus," 1842, page 328, art. **214.**

$$D^4b^3 = 3A_1^2A_5 + 6A_1A_2A_4 + 3A_1A_3^2 + 3A_2^2A_3,$$
$$D^5b^3 = 3A_1^2A_6 + 6A_1A_2A_5 + 6A_1A_3A_4 + 3A_2^2A_4 + 3A_2A_3^2,$$
$$Db^4 = 4A_1^3A_2,$$
$$D^2b^4 = 4A_1^3A_3 + 6A_1^2A_2^2,$$
$$D^3b^4 = 4A_1^3A_4 + 12A_1^2A_2A_3 + 4A_1A_2^3,$$
$$D^4b^4 = 4A_1^3A_5 + 12A_1^2A_2A_4 + 6A_1^2A_3^2 + 12A_1A_2^2A_3 + A_2^4,$$
$$Db^5 = 5A_1^4A_2, \qquad Db^6 = 6A_1^5A_2,$$
$$D^2b^5 = 5A_1^4A_3 + 10A_1^3A_2^2, \qquad D^2b^6 = 6A_1^5A_3 + 15A_1^4A_2^2,$$
$$D^3b^5 = 5A_1^4A_4 + 20A_1^3A_2A_3 + 10A_1^2A_2^3, \; Db^7 = 7A_1^6A_2.$$

To illustrate the use of the procedure given above, let us evaluate R_4. We have

$$\begin{aligned} A_0^{-1/2}R_4 &= \left(\frac{dA^{-5/2}}{1\,!dA}\right)D^3b + \left(\frac{d^2A^{-5/2}}{2\,!dA^2}\right)D^2b^2 + \left(\frac{d^3A^{-5/2}}{3\,!dA^3}\right)Db^3 + \left(\frac{d^4A^{-5/2}}{4\,!dA^4}\right)b^4 \\ &= -(5/2)A_0^{-7/2}(A_4) + (1/2)(5/2)(7/2)A_0^{-9/2}(2A_1A_3 + A_2^2) \\ &\quad -(1/6)(5/2)(7/2)(9/2)A_0^{-11/2}(3A_1^2A_2) \\ &\quad +(1/24)(5/2)(7/2)(9/2)(11/2)A_0^{-13/2}A_1^4 \end{aligned}$$

or

$$\begin{aligned} R_4 = (5/2)A_0^{-6}[-A_0^3A_4 + (7/2)A_0^2(A_1A_3 + A_2^2/2) \\ - (1/2)(7/2)(9/2)A_0A_1^2A_2 \\ + (1/24)(7/2)(9/2)(11/2)A_1^4]. \end{aligned}$$

Tables Indicating Degree of Accuracy Obtainable by Use of Formula (3) for Evaluating

$$P = \sum_{x=c}^{x=n} \binom{n}{x} p^x(1 - p)^{n-x}.$$

P_1 = 1st approximation, $\Delta_1 = (P - P_1)10^6$,

P_2 = 2d approximation, $\Delta_2 = (P - P_2)10^6$,

P_3 = 3d approximation, $\Delta_3 = (P - P_3)10^6$,

$a = np$,

$$T^2 = (n-1)\log\frac{n}{n-1} + (c-1)\log\frac{c-1}{a} + (n-c)\log\frac{n-c}{n-a},$$

$$I = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{T} e^{-t^2}dt.$$

TABLE I

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
			$a = 1.5,\ n = \infty,\ p = 0$			
2	+ .3074653	668154	442174	15991	9518	−1348
4	− .7612106	140849	65643	10816	1989	30
6	−1.5874105	12386	4456	1641	303	8
8	−2.2985028	577	170	103	20	0
			$a = 1.5,\ n = 30,\ p = .05$			
2	+ .2865166	657333	446458	21266	17382	−2083
4	− .8219430	122536	60772	3684	5617	− 97
6	−1.6966449	8211	3282	− 113	914	40
8	−2.4640017	246	85	− 24	50	7

TABLE II

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
			$a = 5,\ n = \infty,\ p = 0$			
2	1.5461442	985613	959576	−1681	2590	−798
4	.6837566	833222	734978	−2036	1897	−138
6	.0000000	500000	384044	2986	1036	− 4
8	− .5960752	199621	133376	4219	616	17
10	−1.1358169	54105	31832	2128	286	13
12	−1.6349406	10385	5452	604	123	2
14	−2.1027717	1471	692	107	11	− 5
16	−2.5454242	159	68	14	1	− 1
			$a = 5,\ n = 100,\ p = .05$			
2	1.5596227	986295	962920	− 373	3331	−732*
4	.6839234	833281	742162	639	3248	−141
6	− .0162780	490817	384001	2889	2108	− 38
8	− .6310024	186097	127961	1989	1431	− 8
10	−1.1912234	46029	28188	566	703	1
12	−1.7124507	7723	4274	75	208	2
14	−2.2138799	914	463	2	38	0
16	−2.6715388	79	37	− 1	5	0
			$a = 5,\ n = 50,\ p = .1$			
2	1.5742756	987005	966214	830	3962	−724
4	.6844600	833471	749706	3289	4437	−212
6	− .0335708	481067	383877	2599	3022	− 62
8	− .6689826	172053	122145	− 334	2076	31
10	−1.2524740	38258	24538	− 789	954	51
12	−1.7994619	5467	3220	− 278	244	24
14	−2.3191412	520	285	− 48	36	6
16	−2.8175965	34	17	− 5	3	0

* $|P - P_3| > |P - P_1|$.

TABLE III

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
			$a = 10, n = \infty, p = 0$			
2	2.5879363	999874	999499	− 47	67	−37
4	1.8406742	995381	989662	− 533	275	−53
6	1.2386541	960089	932912	−1532	518	−50
8	.7094191	842135	779778	−1585	547	−27
10	.2274981	626172	542069	79	425	− 8
12	− .2200272	377838	303223	1786	322	1
14	− .6408864	182375	135535	2077	239	4
16	−1.0401811	70640	48740	1374	146	4
18	−1.4215063	22199	14277	628	68	0
20	−1.7875189	5737	3454	216	25	1
22	−2.1402533	1236	699	58	7	0
24	−2.4813121	225	119	12	1	− 1
			$a = 10, n = 100, p = .1$			
2	2.6528972	999912	999679	− 3	71	−24*
4	1.8917619	996268	992164	− 15	432	−39*
6	1.2715533	963931	942424	278	1070	−45
8	.7213308	846163	793949	997	1349	−36
10	.2161911	620099	548710	1213	1179	−19
12	− .2564838	358406	296967	503	982	− 2
14	− .7042404	159638	123877	− 222	771	10
16	−1.1320595	54691	39891	− 376	470	13
18	−1.5434535	14526	10007	− 222	206	9
20	−1.9410214	3025	1979	− 80	66	4
22	−2.3267578	500	312	− 20	15	1
24	−2.7022383	66	40	− 4	2	0

* $|P - P_3| > |P - P_1|$.

TABLE IV

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
			$a = 15, n = \infty, p = 0$			
4	2.6780004	999924	999788	− 18	10	− 4
6	2.1229551	998660	997207	− 141	54	− 9
8	1.6324888	989520	981998	− 525	146	− 14
10	1.1843012	953019	930147	−1066	242	− 15
12	.7670044	860975	815249	−1198	274	− 10
14	.3737500	701445	636783	− 515	245	− 4
16	.0000000	500000	431911	582	203	0
18	− .3574541	306598	251141	1311	168	2
20	− .7009899	160758	124781	1351	132	2
22	−1.0324325	72134	53106	961	89	2
24	−1.3532229	27826	19464	523	49	1
26	−1.6645241	9287	6184	228	22	0
28	−1.9672925	2700	1715	82	8	− 1
30	−2.2623270	689	418	25	2	0

TABLE IV—*Continued*

$a = 15, n = 30, p = .5$

6	2.6019552	999883	999837	52	24	− 12
8	2.0184138	997845	997388	559	152	− 77
10	1.4626537	980704	978613	2676	411	−239
12	.9237039	904277	899756	5946	488	−343
14	.3942720	711436	707667	5025	227	−205
16	− .1313195	426335	427768	−1916	− 73	72
18	− .6581761	175978	180797	−6392	−382	301
20	−1.1915875	45979	49369	−4405	−497	316
22	−1.7378702	6991	8062	−1346	−278	149
24	−2.3057782	555	715	− 191	− 69	32
26	−2.9097701	19	30	− 11	− 7	3

TABLE V

c	T	$I(10^6)$	$P(10^6)$	Δ_1	Δ_2	Δ_3
	$a = 25, n = \infty, p = 0$					
10	2.6086661	999888	999778	− 11	3	− 1
12	2.2291734	999191	998583	− 51	11	− 2
14	1.8705496	995920	993531	− 159	30	− 4
16	1.5289263	984700	977705	− 364	59	− 6
18	1.2015564	955365	939522	− 617	90	− 7
20	.8863972	894998	866422	− 765	109	− 7
22	.5818753	794717	752697	− 651	110	− 7
24	.2867455	657452	606120	− 253	101	− 6
26	.0000000	500000	447076	268	92	− 3
28	− .2791919	346482	299814	678	84	1
30	− .5515253	217703	182105	837	75	2
32	− .8175896	123790	100070	761	62	3
34	−1.0778902	63708	49782	561	45	3
36	−1.3328647	29718	22460	350	29	2
38	−1.5828952	12593	9212	189	17	2
40	−1.8283181	4860	3445	90	9	2
42	−2.0694313	1713	1178	38	4	1
44	−2.3065005	553	370	15	2	1
	$a = 25, n = 50, p = .5$					
14	2.3698187	999598	999531	80	16	− 9
16	1.9447371	997023	996699	404	57	− 32
18	1.5274793	984621	983580	1329	126	− 74
20	1.1159208	942735	940539	2844	170	−113
22	.7083182	841759	838881	3763	140	−109
24	.3031406	665931	664094	2415	60	− 56
26	− .1010188	443200	443862	− 873	− 20	19
28	− .5055162	237333	239944	−3426	−102	87
30	− .9117246	98634	101319	−3499	−166	117
32	−1.3211006	30859	32454	−2056	−157	95
34	−1.7352770	7063	7673	− 774	− 91	50
36	−2.1561545	1147	1301	− 191	− 33	17
38	−2.5860897	128	153	− 31	− 8	4

A New Method for Obtaining Transient Solutions of Electrical Networks

By W. P. MASON

SYNOPSIS: A new method for obtaining transient solutions of electrical networks is developed in this paper which depends upon the fact that a distortionless line can be made to approach as a limit all three of the circuit elements, resistance, inductance and capacity. The process of solution consists in solving for the current in a distortionless line—which is ordinarily a simple process—and then proceeding to the limiting case of the distortionless line which approaches the element or elements of interest. Some examples are worked out and a derivation of the Laplacian integral solution is given. It is interesting to note that this method gives a formal solution of the Laplacian integral equation.

THE following paper sets forth a new method for obtaining the transient solutions of electrical networks, which it is believed has some advantages over other methods of solution, in that the operations required for solution are quite simple, and also because this method presents a more definite physical picture of the processes involved. By means of this method, the current at any time t can be obtained, due to an applied voltage which is zero when t is less than zero, and is $E_0 \cos(\omega t + \theta)$ when t is greater than zero. This type of voltage includes as a special case the applied voltage, which is zero when t is less than zero, and is unity when t is greater than zero, and hence the solutions obtained by this method reduce to the cases discussed by Heaviside,[1] when ω and θ are taken equal to zero.

This method gives directly the more compact Laplacian integral equation solution, first obtained by Carson, and in addition gives a method for solving this integral equation, if its solution is not already known.

I. Method of Solution

All practical schemes for solving the transient type of circuit problem, including the Laplacian integral equation, and the generalized Fourier integral solution, are made to depend on the known and easily determined steady state solution. This implies that all circuits which have the same steady state solution, have also the same transient or time solution. The method described in the present paper rests on the same basis.

The method of solution used here depends upon the fact that the

[1] Heaviside, "Electromagnetic Theory," Volume II.

distortionless line can be made to approach as a limit, all three of the electrical elements, resistance, inductance, and capacity, and that the complete solution for the current in a distortionless line can be obtained by adding the incident current and the sum of the reflected currents which can occur up to the time of interest. That is the distortionless line has a true velocity of propagation, and hence the current at any time will be the initial current and the sum of the reflections which can occur up to the time of interest. All of the three electrical elements, resistance, inductance, and capacity, can be considered as limiting cases of the distortionless line. Hence the process of solution consists in solving for the current in the distortionless line, and then proceeding to the limiting value of the line which coincides with the element of interest.

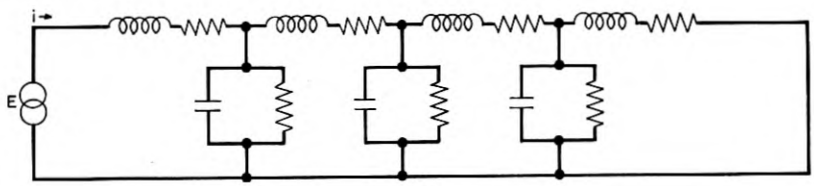


Fig. 1-A.

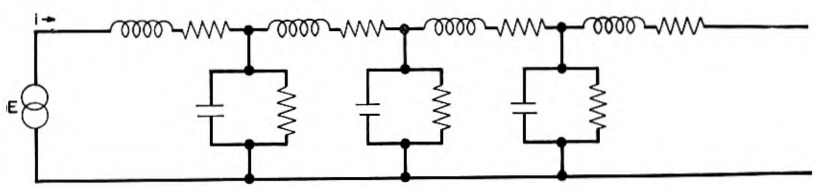


Fig. 1-B.
Diagrammatic representations of lines.

A. The Distortionless Line

Since the distortionless line is the tool by means of which problems are solved by this method, a brief discussion of lines [2] is given here. If a voltage is suddenly applied to a transmission line, the current at any point in the line is zero for a certain time and then begins to build up to its final or steady state value. If there is no distributed inductance in the line, the current begins to build up immediately.

For a distortionless line, however, the current is zero for the time required to propagate the wave to the point of interest and then instantaneously reaches its steady state value. To show this let us consider the equations for a transmission line. A line has distributed

[2] For a more complete discussion of lines see "Transmission Circuits for Telephone Communication," K. S. Johnson, page 144.

series resistance R_u and inductance L_u, and distributed shunt capacity C_u and leakage conductance G_u, as shown on Fig. 1-A, where the letter u indicates the values per unit length. If i denotes the current and v the voltage at a distance x from one end of the line, the well known differential equations are

$$\left.\begin{aligned} \left(L_u \frac{d}{dt} + R_u\right) i &= -\frac{\partial}{\partial x} v, \\ \left(C_u \frac{d}{dt} + G_u\right) v &= -\frac{\partial}{\partial x} i. \end{aligned}\right\} \tag{1}$$

If we eliminate v from these two equations we have

$$L_u C_u \frac{d^2 i}{dt^2} + (R_u C_u + G_u L_u) \frac{di}{dt} + R_u G_u i = \frac{\partial^2 i}{\partial x^2}. \tag{2}$$

Similarly, if i is eliminated, the resulting equation is the same as (2) with i replaced by v. Since we are dealing with simple harmonic forces, the current i varies as $\cos(\omega t + \theta)$ where ω is 2π times the frequency of vibration and θ is an arbitrary phase angle. It is usually more convenient to consider i as varying according to the time factor

$$i = \dot{i} e^{j(\omega t + \theta)} = \dot{i}[\cos(\omega t + \theta) + j \sin(\omega t + \theta)],$$

where $\dot{i}$ is the maximum amplitude of i. The solution obtained on this assumption is called the symbolic solution, and the real solution is obtained from the symbolic solution by taking the real part. Substituting the symbolic form of i in equation (2), this equation reduces to

$$[(j\omega)^2 L_u C_u + j\omega[R_u C_u + G_u L_u] + R_u G_u] i = \frac{\partial^2 i}{dx^2}. \tag{3}$$

The solution for a line can be specified in terms of two parameters, the characteristic impedance and the propagation constant of the line. To show this we note that the solution of (3) is

$$i = A e^{-\Gamma x} + B e^{\Gamma x}, \tag{4}$$

where $\Gamma^2 = [R_u + j\omega L_u][G_u + j\omega C_u]$ and A and B are constants. From the last of equations (1) we have

$$v = -\frac{\frac{\partial}{\partial x} i}{G_u + j\omega C_u} = \frac{\Gamma}{G_u + j\omega C_u}[A e^{-\Gamma x} - B e^{\Gamma x}]. \tag{5}$$

When $x = 0$, $i = i_0$, and $v = v_0$. From (4) and (5) solving for A and B we have

$$A = \frac{i_0}{2} + \frac{v_0/2}{\sqrt{\frac{R_u + j\omega L_u}{G_u + j\omega C_u}}}; \qquad B = \frac{i_0}{2} - \frac{v_0/2}{\sqrt{\frac{R_u + j\omega L_u}{G_u + j\omega C_u}}}.$$

Substituting these values in (4) and (5), we have the equations

$$\begin{aligned} i &= i_0 \cosh \Gamma x - \frac{v_0 \sinh \Gamma x}{\sqrt{\frac{R_u + j\omega L_u}{G_u + j\omega C_u}}}, \\ v &= v_0 \cosh \Gamma x - i_0 \sqrt{\frac{R_u + j\omega L_u}{G_u + j\omega C_u}} \sinh \Gamma x. \end{aligned} \tag{6}$$

In this equation $\Gamma x = P$, the propagation constant of the line, and

$$\sqrt{\frac{R_u + j\omega L_u}{G_u + j\omega C_u}} = Z_0,$$

the characteristic impedance of the line. If we are interested in a given length of line l, the parameters are

$$P = \sqrt{(R + j\omega L)(G + j\omega C)}; \qquad Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \tag{7}$$

where R, L, G, and C are the total distributed constants for the length of line considered. The characteristic impedance is the impedance looking into a line of infinite length as can readily be seen from either of equations (6) by letting x, the length, approach infinity. In this case $\cosh \Gamma x = \cosh P$ approaches $\sinh P$ and both approach infinity. Then from (6)

$$\frac{v_0}{i_0} = \left(\frac{\cosh P - i/i_0}{\sinh P}\right) Z_0 \to Z_0 \quad \text{when} \quad x \to \infty,$$

since i/i_0 can never be larger than 1.

The physical significance of the propagation constant is that e^{-P} represents the ratio of currents or voltages at the two ends of the line when the line is connected to an infinite line of the same characteristic impedance. To show this, suppose we terminate the section considered in an infinite line, which as we have seen above has an impedance Z_0. Then in equations (6), $v = v_1$ the output voltage and

$i = i_1$ the output current. We let $v_1 = i_1 Z_0$. Eliminating either v_0 or i_0, we have

$$v_1 = v_0 e^{-P} \qquad \text{or} \qquad i_1 = i_0 e^{-P}. \tag{8}$$

In the following work it is necessary to know the impedance of a short circuited line and that of an open circuited line. For the short circuited line, the voltage at the far end is zero, so putting $v = 0$ in the last of equation (6), we have

$$v_0/i_0 = Z_0 \tanh P. \tag{9}$$

For the open circuited line we put $i = 0$, obtaining

$$v_0/i_0 = Z_0 \coth P. \tag{10}$$

So far we have discussed the general transmission line. For the distortionless line there exists the relation

$$\frac{R}{L} = \frac{G}{C}. \tag{11}$$

Substituting this relation in equation (7), these parameters reduce to

$$Z_0 = R_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{R}{G}} \quad \text{and} \quad P = \sqrt{RG} + j\omega\sqrt{LC} = A + j\omega D. \tag{12}$$

This equation shows that the characteristic impedance becomes a resistance R_0, while the propagation constant equals a real constant A plus $j\omega$ times the constant D. To show how wave transmission takes place in an infinite line, these values are substituted in equation (8), giving

$$v_1 = v_0 e^{-(A+j\omega D)}; \qquad i_1 = i_0 e^{-(A+j\omega D)}.$$

To find the real solution, we take the real part of this symbolic solution, obtaining

$$v_1 = \bar{v}_0 e^{j(\omega t+\theta)} e^{-(A+j\omega D)} = \bar{v}_0 e^{-A} e^{j[\omega(t-D)+\theta]},$$

or, taking the real part,

$$\begin{aligned} & v_1 = \bar{v}_0 e^{-A} \cos [\omega(t - D) + \theta] \\ \text{and} \quad & \\ & i_1 = \bar{i}_0 e^{-A} \cos [\omega(t - D) + \theta], \end{aligned} \tag{13}$$

where the dash over v_0 and i_0 indicate the maximum amplitude of these quantities. Since $v_0 = \bar{v}_0 \cos (\omega t + \theta)$, these equations show that either v_1 or i_1 has the same form as v_0 or i_0 respectively, attenuated by a factor e^{-A} and delayed in time by an amount D.

B. Condition for Obtaining Lumped Constants from a Distortionless Line

In a distortionless line there is only one necessary relation between the constants, equation (11). Hence, we are at liberty to vary the constants subject only to this relation. In the following work we wish to make the distortionless line degenerate into resistances, inductances, and capacities, or combinations of these.

For example, suppose that we wish to obtain a resistance from a distortionless line. To obtain this we take a short circuited line as shown on Fig. 1-A and let R be finite, $L \to 0$, $G \to 0$, and $C \to 0^2$ in order to satisfy equation (11). The shunt elements will all vanish and the series inductance disappears, leaving only the series resistance in the line. Since the line is short circuited, the line degenerates into a resistance. The line parameters for this case become

$$R_0 = \sqrt{\frac{R}{G}} = \sqrt{\frac{L}{C}} \to \infty^{1/2}; \; P = \sqrt{RG} + j\omega\sqrt{LC} \to (0^{1/2} + j\omega 0^{3/2})$$

and

$$R_0 P = \sqrt{\frac{R}{G}}(\sqrt{RG}) + \sqrt{\frac{L}{C}}(j\omega\sqrt{LC}) \to R. \tag{14}$$

There are three combinations of lumped elements which can be obtained from the short circuited line and three combinations which can be obtained from the open circuited line. These are listed in the following table, together with the resulting line parameters.

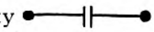
Limiting Cases of the Distortionless Line

A. Limiting Values with Short Circuited Line

Equation	Assumed Line Constants				Resulting Line Parameters			Resulting Lumped Element
	R	L	G	C	R_0	P	R_0P	
(14)	Finite	0	0	0^2	$\infty^{1/2}$	$0^{1/2} + j\omega 0^{3/2}$	R	Resistance
(15)	0	Finite	0^2	0	$\infty^{1/2}$	$0^{3/2} + j\omega 0^{1/2}$	$j\omega L$	Inductance
(16)	Finite	Finite	0	0	$\infty^{1/2}$	$0^{1/2} + j\omega 0^{1/2}$	$R+j\omega L$	Resistance and Inductance

The first three cases result from the suppression of the shunt elements in the short circuited line, and the line parameters are characterized by $R_0 \to \infty$; $P \to 0$. The last three cases result from the suppression of the series elements in the open circuited line, and are characterized by $R_0 \to 0$; $P \to 0$.

B. Limiting Values with Open Circuited Line

Equation	Assumed Line Constants				Resulting Line Parameters			Resulting Lumped Element
	R	L	G	C	R_0	P	R_0/P	
(17)	0	0^2	Finite	0	$0^{1/2}$	$0^{1/2} + j\omega 0^{3/2}$	$1/G$	Resistance 
(18)	0^2	0	0	Finite	$0^{1/2}$	$0^{3/2} + j\omega 0^{1/2}$	$1/j\omega C$	Capacity 
(19)	0	0	Finite	Finite	$0^{1/2}$	$0^{1/2} + j\omega 0^{1/2}$	$1/(G+j\omega C)$	Resistance and Capacity 

C. Solution for a Resistance and Inductance in Series

As a first example of a transient solution obtained by this method let us consider the case of an inductance and resistance in series with a source of alternating voltage. To solve this problem, consider the case of a voltage in series with a distortionless line, short circuited, as shown on Fig. 1-*A*. The current immediately flowing on application of the voltage will be i_0 where

$$i_0 = \frac{E}{R_0}. \tag{20}$$

This current is transmitted down the line and completely reflected at the far end, returning to the near end. The first reflected current entering the generator is then

$$i_1 = i_0 e^{-2P}, \tag{21}$$

where P is the propagation constant of the line. Upon reaching the near end, the current is completely reflected in the same phase and again enters the line. At the end of the first reflection, the current entering the line is

$$i = i_0(1 + 2e^{-2P}). \tag{22}$$

After $(n - 1)$ reflections and passages through the line, the current is

$$\begin{aligned} i &= i_0[1 + 2e^{-2P} + 2e^{-4P} + \cdots + 2e^{-2(n-1)P}] \\ &= i_0\left[2\left(\frac{1 - e^{-2nP}}{1 - e^{-2P}}\right) - 1\right]. \end{aligned} \tag{23}$$

Now the time at which the nth reflection occurs will be

$$t = n(2D),$$

where D is the time of delay in passing the network once. For a distortionless line

$$D = \sqrt{LC}. \tag{24}$$

Hence, we can replace n by

$$n = \frac{t}{2D} = \frac{t}{2\sqrt{LC}}. \tag{25}$$

Since $D \to 0$ and $n \to \infty$, the time scale becomes continuous. In equation (23) we insert the values given in (16) and (25), appropriate to the limiting case considered here, namely

$$P = \frac{R + j\omega L}{R_0}; \qquad R_0 \to \infty; \qquad n = \frac{t}{2\sqrt{LC}} = \frac{tR_0}{2L}$$

and note that $2P \to 0$ so that $e^{-2P} \to 1 - 2P$; then

$$i \to \frac{E}{R_0}\left[\frac{2(1 - e^{-tR_0/L[(R+j\omega L)/R_0]})}{1 - 1 + 2(R + j\omega L)/R_0} - 1\right]$$

$$= E\left[\frac{1 - e^{-t(R/L+j\omega)}}{R + j\omega L} - \frac{1}{R_0}\right].$$

But $R_0 \to \infty$ and hence the solution is

$$i = E\left[\frac{1 - e^{-t(R/L+j\omega)}}{R + j\omega L}\right]. \tag{26}$$

This is the symbolic or complex algebra solution of the equation

$$L\frac{di}{dt} + Ri = E. \tag{27}$$

In general it is desirable to obtain the current due to an applied voltage of the form

$$E = E_0 \cos(\omega t + \theta).$$

This solution can be obtained directly from the symbolic solution given in (26) by taking the real part. The result is

$$i = E_0\left[\frac{\cos(\omega t + \theta - \varphi) - \cos(\theta - \varphi)e^{-tR/L}}{\sqrt{R^2 + \omega^2 L^2}}\right]. \tag{28}$$

It will be found that (28) is a solution of (27) for an applied voltage $E_0 \cos(\omega t + \theta)$.

D. General Method for Determining Reflections

The method for obtaining the successive current reflections of the line given in the preceding section, is laborious to carry out in complicated cases and hence it is desirable to obtain a simple method for determining the reflections. Such a method is the expansion of the expression for the impedance by the binomial theorem in order to get the successive reflections. In the above example the current i is

$$i = E/(R + j\omega L).$$

We note that the expression for the impedance is approached by that of a short circuited line when the R and L of the line are finite and the capacity and leakance approach zero. Hence

$$i = \frac{E}{R + j\omega L} \rightarrow \frac{E}{R_0 \tanh P} = \frac{E(1 + e^{-2P})}{R_0(1 - e^{-2P})}.$$

Now the expansion of

$$\frac{1}{1 - e^{-2P}} = 1 + e^{-2P} + e^{-4P} + \cdots.$$

Hence

$$i \rightarrow \frac{E}{R_0}(1 + e^{-2P})(1 + e^{-2P} + e^{-4P} + \cdots)$$

$$= \frac{E}{R_0}[1 + 2e^{-2P} + 2e^{-4P} + \cdots + 2e^{-2(n-1)P} + \cdots],$$

which is the expression for the reflections given by equation (23).

In all the following problems it will be found that a similar process for obtaining the reflections can be followed. It is evident that any method which gives an expression for the current in the form

$$i = E[a_0 + a_1 e^{-2j\omega D} + a_2 e^{-4j\omega D} + \cdots + a_n e^{-2nj\omega D} + \cdots] \quad (29)$$

will give the reflections, for if we take the real part of this expression we have

$$i = E_0[a_0 \cos(\omega t + \theta) + a_1 \cos[\omega(t - 2D) + \theta] + \cdots \\ + a_n \cos[\omega(t - 2nD) + \theta] + \cdots].$$

Each term represents a current which adds to the original current after a time of delay $2D$, $4D$, $\cdots$ $2nD$, and hence the nth terms represents the nth reflection. Therefore any method, such as the above, which gives the current in the form of equation (29), will give the reflections.

E. Simpler Form for Replacing an Impedance

In the preceding section, the transient solution of an inductance and resistance in series was obtained by replacing the impedance $R + j\omega L$ by the expression

$$R_0 \tanh P, \text{ where } R_0 P = R + j\omega L \text{ and } R_0 \to \infty;\ P \to 0.$$

Tanh P has a numerator and a denominator both of which must be expanded in order to obtain the reflections. If a single term can be used, the expansion becomes simpler. In order to effect such a simplification, it is necessary to find a physical structure, which has only one term in its impedance expression and which approaches a resistance and inductance as a limit.

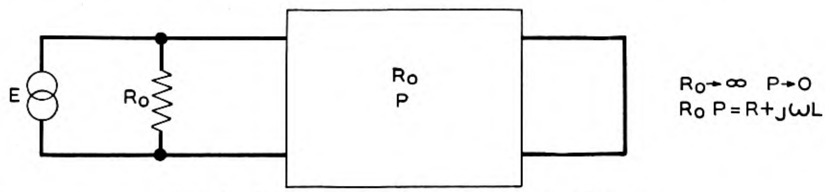


Fig. 2—Short circuited line and shunt resistance.

Such a structure is shown on Fig. 2. It consists of a short circuited line shunted by a resistance R_0. The current into the combination is

$$i = \frac{E}{\dfrac{R_0 \times R_0 \tanh P}{R_0 + R_0 \tanh P}} = \frac{E}{\dfrac{R_0(1 - e^{-2P})}{2}}. \tag{30}$$

If now in the short circuited line, we let R and L be finite and $G \to 0$, $C \to 0$, the combination obviously reduces to a resistance and inductance in series, since the infinite shunt will not affect the result. Hence the replacement of a resistance and inductance in the equation

$$i = \frac{E}{R + j\omega L}$$

by the expression in (30), is justified.

The solution of (30) is gotten by expanding the expression and is

$$i = \frac{2E}{R_0}[1 + e^{-2P} + e^{-4P} + \cdots + e^{-2(n-1)P} + \cdots] = \frac{2E[1 - e^{-2nP}]}{R_0[1 - e^{-2P}]}.$$

Upon substituting in the values $R_0 P = R + j\omega L$, $n = t/2D$, and letting $R_0 \to \infty$; $P \to 0$, we have

$$i = E\left[\frac{1 - e^{-t[R/L + j\omega]}}{R + j\omega L}\right]$$

in agreement with equation (26).

Similarly, when we have the expression

$$\frac{1}{G + j\omega C}$$

we can replace it by

$$\frac{2R_0}{1 - e^{-2P}} \quad \text{where} \quad \frac{R_0}{P} = \frac{1}{G + j\omega C} \quad \text{and} \quad R_0 \to 0,\ P \to 0. \tag{31}$$

The structure which gives the impedance in (31) is an open circuited line in series with a resistance R_0. The impedance of the combination is

$$R_0 + R_0 \coth P = R_0\left[1 + \frac{1 + e^{-2P}}{1 - e^{-2P}}\right] = \frac{2R_0}{1 - e^{-2P}} \cdot$$

If then the series impedances of the line approach zero, $R_0 \to 0$ and the impedance of the combination approaches

$$\frac{1}{G + j\omega C} \cdot$$

F. Solution for a Resistance and Capacity in Series

As a second example let us obtain the solution of a resistance and capacity in series. To obtain the solution we solve the case of a

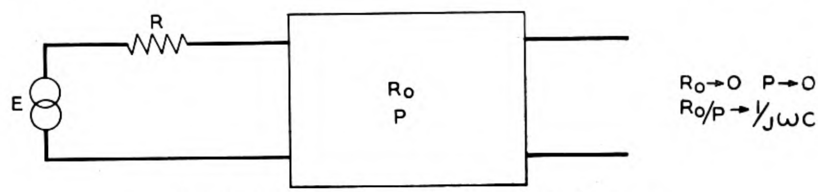


Fig. 3—Open circuited line and series resistance.

resistance in series with an open circuited line as shown on Fig. 3. The steady state solution for the current in this circuit is

$$i = \frac{E}{R + \dfrac{1}{j\omega C}} \cdot$$

Replacing $\dfrac{1}{j\omega C}$ by $\dfrac{2R_0}{1 - e^{-2P}}$, and substituting in the above equation there results

$$i = \frac{E}{R + \dfrac{2R_0}{1 - e^{-2P}}} \quad \text{when} \quad R_0 \to 0 \quad \text{and} \quad P \to 0, \quad \text{and} \quad \frac{R_0}{P} \to \frac{1}{j\omega C}$$

in accordance with equation (18).

After some rearrangements this can be put into the form

$$i = \frac{E}{R + R_0}\left[1 - \frac{R_0}{R}\,\frac{e^{-2F}(1 + e^{-2P})}{1 - e^{-2(F+P)}}\right], \tag{32}$$

where

$$e^{-2F} = \frac{R}{R + 2R_0}\cdot$$

Expanding equation (32) in the form of a series, there results

$$i = \frac{E}{R+R_0}\left[\frac{R+3R_0}{R+2R_0} - \frac{R_0}{R}\left(\frac{2R+2R_0}{R+2R_0}\right)(1+e^{-2(F+P)}+e^{-4(F+P)} + \cdots)\right]\cdot$$

Summing up n terms of this series, we have

$$i = \frac{E}{R + R_0}\left[\frac{R + 3R_0}{R + 2R_0} - \frac{R_0}{R}\left(\frac{2R + 2R_0}{R + 2R_0}\right)\left(\frac{1 - e^{-2n(F+P)}}{1 - e^{-2(F+P)}}\right)\right]\cdot \tag{33}$$

Since in the above expression $R_0 \to 0$, we can obtain the value of F by writing the first terms of the expansion for the exponential

$$e^{-2F} = 1 - 2F + \frac{(2F)^2}{2!} + \cdots = \frac{R}{R + 2R_0} = 1 - \frac{2R_0}{R} + \cdots.$$

Hence

$$F \to \frac{R_0}{R}\cdot$$

If now in equation (33) we proceed to the limit, letting

$$R_0 \to 0; \quad P \to 0; \quad \frac{R_0}{P} = \frac{1}{j\omega C}; \quad n = \frac{t}{2D}$$

there results the equation

$$i = \frac{E}{R}\left[1 - \frac{1 - e^{-t(1/RC+j\omega)}}{1 + j\omega CR}\right]\cdot \tag{34}$$

This equation is the symbolic solution of the integral equation

$$Ri + \frac{1}{C}\int idt = E_0 e^{j(\omega t+\theta)}.$$

If we wish the solution corresponding to the impressed voltage, $E_0 \cos(\omega t + \theta)$, we take the real part of (34), obtaining

$$i = E_0\frac{\cos(\omega t + \theta + \delta) - [\sin(\theta - \delta)\tan\delta]e^{-t/RC}}{\sqrt{R^2 + 1/\omega^2C^2}},$$

where

$$\tan\delta = \frac{1}{\omega RC}\cdot$$

II. Solution for M Sections of All-Pass Lattice Network

The process for solving any type of problem is to replace any resistance and inductance in series by $R_{0_1}(1 - e^{-2P_1})/2$, and any capacity and leakance in parallel by $(1 - e^{-2P_2})/2R_{0_2}$, where $R_{0_1} \to \infty$; $R_{0_2} \to 0$; $R_{0_1}P_1 = R + j\omega L$; $R_{0_2}/P_2 = 1/(G + j\omega C)$.

The next problem considered here is the solution for any number of sections of all-pass lattice type network [3] as shown on Figure 4.

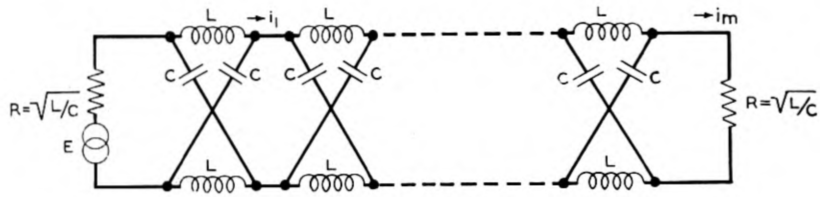


Fig. 4—Sections of all pass lattice network.

These networks have the property of passing all frequencies without attenuation, and they are much used as phase equalizers.

The steady state equation for the current at the end of the first section, when this section is terminated at each end by a resistance $R = \sqrt{Z_1 Z_2}$, is

$$i_1 = \frac{E}{2R}\left[\frac{\sqrt{Z_2} - \sqrt{Z_1}}{\sqrt{Z_2} + \sqrt{Z_1}}\right].$$

The current flowing out of the mth section of such a structure takes the form

$$i_m = \frac{E}{2R}\left[\frac{\sqrt{Z_2} - \sqrt{Z_1}}{\sqrt{Z_2} + \sqrt{Z_1}}\right]^m. \tag{35}$$

In the structure considered $Z_1 = j\omega L$; $Z_2 = 1/j\omega C$, and $\sqrt{L/C} = R$. In accordance with section I-B, we replace the inductance by a short circuited line, and the capacity by an open circuited line. For the first line, in the limiting case, we have by equation (15),

$$R_0 \to \infty; \quad R_0 P = j\omega L.$$

For the second line, by equation (18), we have

$$\overline{R}_0 \to 0; \quad \frac{\overline{R}_0}{\overline{P}} = \frac{1}{j\omega C}.$$

There is no loss of generality if we take the propagation constants for

[3] See for example *B. S. T. J.*, July 1928, page 510.

9

the two lines equal so that

$$\frac{\overline{R}_0}{\overline{P}} \times R_0 P = j\omega L \times \frac{1}{j\omega C} = \frac{L}{C} = R^2.$$

Hence $\overline{R}_0 = R^2/R_0$. Substituting these values in equation (35), we have

$$i_m = \frac{E}{2R}\left[\frac{1 - [R_0/2R](1 - e^{-2P})}{1 + [R_0/2R](1 - e^{-2P})}\right]^m. \tag{36}$$

After some simple rearrangements, equation (36) takes the form

$$i_m = \frac{E}{2R}\left[\frac{4R}{R_0}\left(\frac{1}{1 - e^{-2[(R/R_0)+P]}}\right) - 1\right]^m. \tag{37}$$

If m equals 1, the solution can be obtained exactly as discussed in the first example in section (1), and it is

$$i_1 = \frac{E}{2R}\left[\frac{1 - j\omega\sqrt{LC} - 2e^{-t[(1/\sqrt{LC})+j\omega]}}{1 + j\omega\sqrt{LC}}\right].$$

The solution for m sections of lattice network is discussed in the Appendix, and it is there shown that the solution can be written in the form

$$\begin{aligned} i_m = \frac{E}{2R}\Bigg[\left(\frac{1 - j\omega\sqrt{LC}}{1 + j\omega\sqrt{LC}}\right)^m &- e^{-t[(1/\sqrt{LC})+j\omega]}\Bigg[\frac{\left(\frac{t}{\sqrt{LC}}\right)^{m-1}}{(m-1)!} \times \frac{2^m}{1 + j\omega\sqrt{LC}} \\ &+ \frac{\left(\frac{t}{\sqrt{LC}}\right)^{m-2}}{(m-2)!}\left(\frac{2^m}{(1 + j\omega\sqrt{LC})^2} - \frac{m2^{m-1}}{1 + j\omega\sqrt{LC}}\right) \\ &+ \cdots + \frac{t}{\sqrt{LC}}\left[\frac{(1 - j\omega\sqrt{LC})^m}{(1 + j\omega\sqrt{LC})^{m-1}} - (-1)^m\left(m + 1 - \frac{m}{1 + j\omega\sqrt{LC}}\right)\right] \\ &+ \left(\frac{1 - j\omega\sqrt{LC}}{1 + j\omega\sqrt{LC}}\right)^m - (-1)^m\Bigg]\Bigg]. \end{aligned} \tag{38}$$

Equation (38) represents the symbolic or complex algebra solution for the current at the end of the mth section of a lattice network as shown on Fig. 4. It is usually desirable to obtain the current due to an applied voltage $E_0 \cos(\omega t + \theta)$. This can be obtained from equation (38) by taking the real part of the equation and rejecting the imaginary part. The process of doing this is simple and the result obtained is

$$i_m = \frac{E}{2R}\left[\cos(\omega t + \theta - 2m\varphi) - 2e^{-(t/\sqrt{LC})}\left[\frac{\left(\frac{2t}{\sqrt{LC}}\right)^{m-1}}{(m-1)!}\cos(\theta-\varphi)\cos\varphi\right.\right.$$

$$+ \frac{\left(\frac{2t}{\sqrt{LC}}\right)^{m-2}}{(m-2)!}[2\cos^2\varphi\cos(\theta - 2\varphi) - m\cos\varphi\cos(\theta-\varphi)]$$

$$+ \frac{\left(\frac{2t}{\sqrt{LC}}\right)^{m-3}}{(m-3)!}\left[4\cos^3\varphi\cos(\theta-3\varphi) - 2m\cos^2\varphi\cos(\theta-2\varphi)\right.$$

$$\left.\left.\left. + \frac{m(m-1)}{2!}\cos\varphi\cos(\theta-\varphi)\right] + \cdots\right]\right], \tag{39}$$

where $\tan\varphi = \omega\sqrt{LC}$.

For example, the solutions for one and two section networks take the form

$$i_1 = \frac{E}{2R}[\cos(\omega t + \theta - 2\varphi) - 2e^{-(t/\sqrt{LC})}\cos\varphi\cos(\theta-\varphi)]$$

$$i_2 = \frac{E}{2R}\left[\cos(\omega t + \theta - 4\varphi) - 2e^{-(t/\sqrt{LC})}\left[2\cos^2\varphi\cos(\theta-2\varphi)\right.\right. \tag{40}$$

$$\left.\left. - 2\cos\varphi\cos(\theta-\varphi) + \frac{2t}{\sqrt{LC}}\cos(\theta-\varphi)\cos\varphi\right]\right].$$

It appeared desirable to obtain some numerical calculations on the building up of current in this type of network. This calculation has been carried through for two section, four section, and six section networks. The current has been calculated for a constant voltage, for an alternating voltage whose frequency is the resonant frequency of the network, and for one whose frequency is twice the resonant frequency of the network. The current building up for a constant applied voltage is shown for the three networks on Fig. 5. The current building up in a two section network, in which an alternating voltage whose frequency equals the resonant frequency of the network, is shown on Fig. 6. The steady state and the transient terms are shown separately in the dotted lines, and the complete solution in the full line. The applied voltage is of the form $E_0 \cos \omega t$ and hence θ is taken as zero in equation (39). Similarly, curves for two, four,

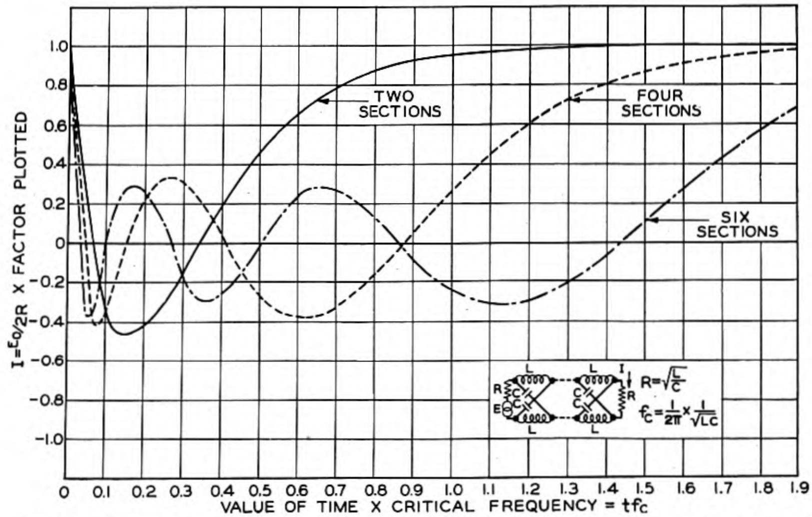


Fig. 5—Current resulting from the application of a constant voltage, E, on several sections of lattice network. The current plotted is the current in the termination of the network.

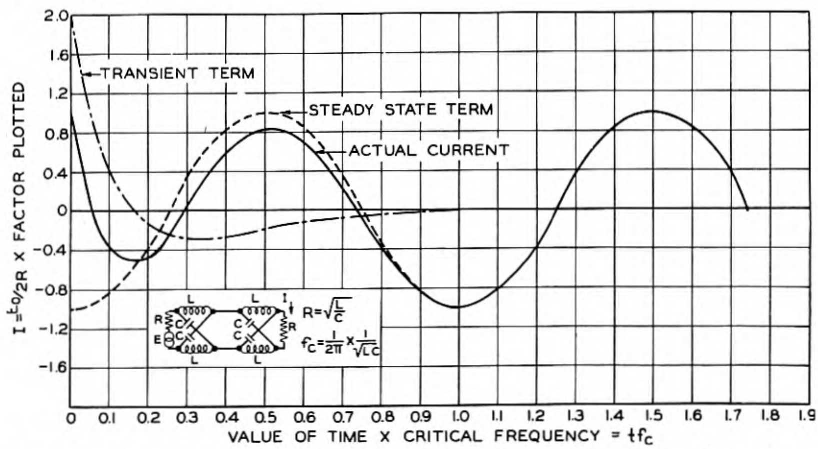


Fig. 6—Current resulting from the application of an alternating voltage, $E = E_0 \cos \omega_c t$, on a two section lattice network. The current plotted is the current in the termination of the network. The frequency of the applied voltage is the resonant frequency, f_c, of the network.

and six sections are shown on Fig. 7 and Fig. 8. Fig. 7 shows the transient terms and Fig. 8 the complete solution. Fig. 9 and Fig. 10 show similar curves for a frequency twice the resonant frequency of the network. In addition, the solution for infinite frequency is readily obtained from (39) since for this frequency $\varphi = 90°$. Then

$$i = \frac{E}{2R}[\cos(\omega t + \theta - m\pi)].$$

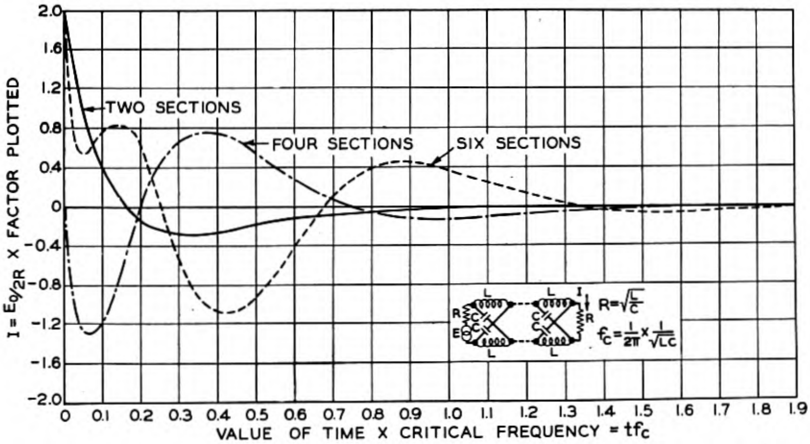


Fig. 7—Transient current resulting from the application of an alternating voltage, $E = E_0 \cos \omega_c t$, for several sections of lattice network. The current plotted is the current in the termination of the network. The frequency of the applied voltage is the resonant frequency, f_c, of the network.

III. Laplacian Infinite Integral Equation and Its Formal Solution

The solution of circuit problems by means of the Laplacian integral equation has been used by Carson [4] to a large extent. It is interesting to note that this integral form can be derived in a simple manner by means of this expansion method, and that this method provides a means for solving the Laplacian integral equation.

Any impedance Z is made up of resistances, inductances and capacities, and hence the current i can be represented by a series

$$i = \frac{E}{Z} = E[a_0 + a_1 e^{-j2\omega D} + a_2 e^{-j4\omega D} + \cdots + a_n e^{-j2n\omega D} + \cdots]. \quad (41)$$

The interpretation of this expansion from a physical standpoint is that the current is Ea_0, for the first interval of time $2D$, $E[a_0 + a_1 e^{-j2\omega D}]$ for the next interval of time $2D$, etc. Hence at the time $t = n(2D)$, the current i will be given by the sum of n terms

[4] See "Electric Circuit Theory and the Operational Calculus," *B. S. T. J.*, October 1925, and following.

of this series. We can, therefore, express the current i at the time t by the integral

$$i = E\left[\int_0^t a(t)e^{-j\omega t}dt + a_0\right], \tag{42}$$

where the value of $a(t)$ for any interval of time $(n-1)2D$ to $n(2D)$ is the constant of the above series a_{n-1} divided by $2D$. The value of this integral for an infinite time must reduce to the steady state value of $i = E/Z$, hence

$$\frac{E}{Z} = E\left[\int_0^\infty a(t)e^{-j\omega t}dt + a_0\right].$$

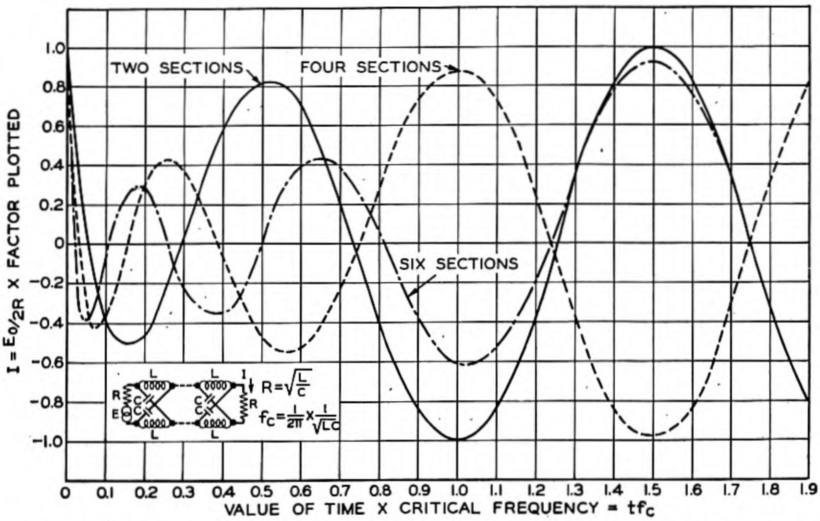


Fig. 8—Current resulting from the application of an alternating voltage, $E = E_0 \cos \omega_c t$, on several sections of lattice network. The current plotted is the current in the termination of the network. The frequency of the applied voltage is the resonant frequency, f_c, of the network.

Cancelling out the common factor E, we have the infinite integral equation

$$\frac{1}{Z(j\omega)} = \left[\int_0^\infty a(t)e^{-j\omega t}dt + a_0\right]. \tag{43}$$

The physical interpretation of the quantity $a(t)$ is readily obtained by reference to equation (42). If we set $\omega = 0$ and $E = 1$ in this equation we have

$$i = \int_0^t a(t)dt + a_0 = \int_0^t a(t)dt + h(0),$$

where $h(0)$ is a constant denoting the current when t is zero. Now i at any time t is the indicial or direct current admittance, designated by $h(t)$, hence $a(t)$ is

$$a(t) = \frac{d}{dt}(h(t)).$$

The infinite integral equation (43) takes the form

$$\frac{1}{Z(j\omega)} = \left[\int_0^\infty \frac{d}{dt}(h(t))e^{-j\omega t}dt + h(0) \right]. \tag{44}$$

This integral equation does not have quite the same form as Carson's integral equation but can be readily put into that form by means of

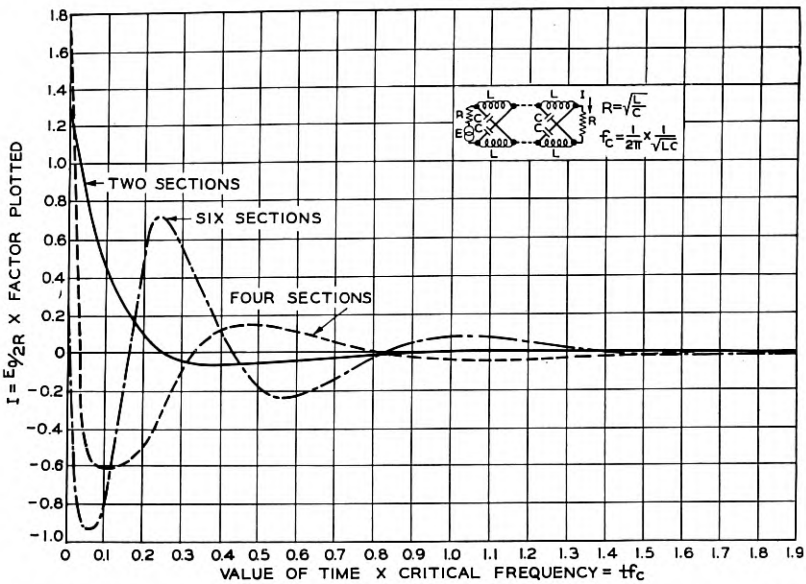


Fig. 9—Transient current resulting from the application of an alternating voltage, $E = E_0 \cos 2\omega_c t$, on several sections of lattice network. The current plotted is the current in the termination of the network. The frequency of the applied voltage is twice the resonant frequency, f_c, of the network.

Borel's theorem [5] which is given below. Suppose that $1/Z'$ and $1/Z''$ are two admittances, which when multiplied together give the admittance $1/H$. The admittances $1/Z'$ and $1/Z''$ have the expansions

$$\frac{1}{Z'} = [a_0' + a_1'e^{-2j\omega D} + a_2'e^{-4j\omega D} + \cdots];$$

$$\frac{1}{Z''} = [a_0'' + a_1''e^{-2j\omega D} + a_2''e^{-4j\omega D} + \cdots].$$

[5] See *B. S. T. J.*, October 1925, page 722.

The expansion $1/H$, then takes the form

$$\frac{1}{H} = \frac{1}{Z'Z''} = [a_0 + a_1 e^{-2j\omega D} + \cdots + a_n e^{-2nj\omega D} + \cdots]$$
$$= [a_0'a_0'' + [a_0'a_1'' + a_0''a_1']e^{-2j\omega D} + \cdots$$
$$+ [a_0'a_n'' + \cdots + a_k'a_{n-k}'' + \cdots + a_n'a_0'']e^{-2nj\omega D} + \cdots].$$

We have then the relation

$$a_n = [a_0'a_n'' + \cdots + a_k'a_{n-k}'' + \cdots + a_n'a_0''].$$

Now a_n'' is the value of a'' when $t = n(2D)$, hence the above relation can be put into the form of an integral

$$a(t) = \int_0^t a''(\tau)a'(t - \tau)d\tau,$$

where $\tau = k/2D$, and the complete relation is

$$\frac{1}{H} = \frac{1}{Z'Z''} = \int_0^\infty a(t)e^{-j\omega t}dt + a_0$$
$$= \int_0^\infty \left[\int_0^t a''(\tau)a'(t - \tau)d\tau\right] e^{-j\omega t}dt + a_0'a_0''.$$

Suppose now that we let $Z' = j\omega$; $Z'' = Z(j\omega)$. We know from Heaviside's rule and from Section I that $1/j\omega$ has the direct current or indicial admittance solution

$$\frac{1}{j\omega} = t = h(t).$$

Hence

$$a'(t) = \frac{d}{dt}h'(t) = \frac{d}{dt}t = 1 \qquad \text{and} \qquad a_0' = 0,$$

and the infinite integral equation

$$\frac{1}{H} = \frac{1}{j\omega Z''(j\omega)} = \int_0^\infty \left[\int_0^t a''(\tau)a'(t - \tau)d\tau\right] e^{-j\omega t}dt + a_0'a_0''$$

takes the form

$$\frac{1}{j\omega Z''(j\omega)} = \int_0^\infty \left[\int_0^t a''(\tau)d\tau\right] e^{-j\omega t}dt = \int_0^\infty h''(t)e^{-j\omega t}dt.$$

Dropping the primes, we have the Laplacian integral equation

$$\frac{1}{j\omega Z(j\omega)} = \int_0^\infty h(t)e^{-j\omega t}dt. \tag{45}$$

Hence (43) is equivalent to Carson's integral equation, if $(j\omega)$ is replaced by p.

It will be noted that in deriving this equation use is made only of the general form of the expansion of admittances. For particular admittances, the values of the a's in equation (43) or the h's in equation (45) can be derived directly from an expansion of the admittance function, as shown in the foregoing work. Hence, if the solution of

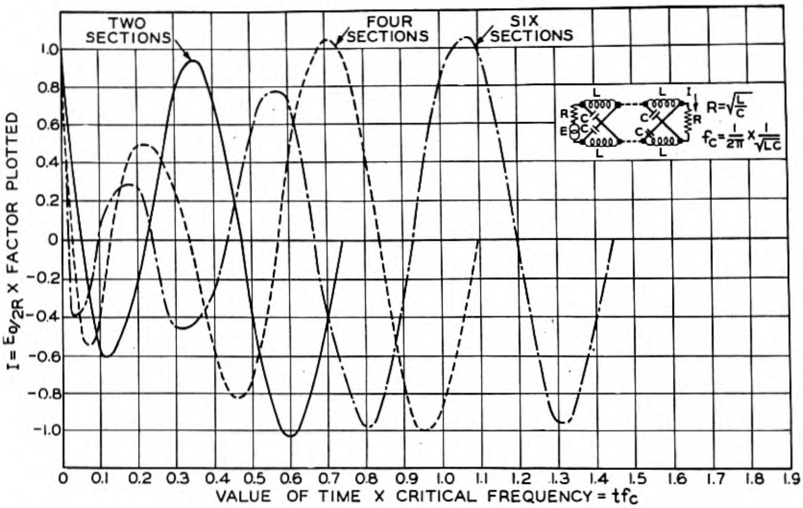


Fig. 10—Current resulting from the application of an alternating voltage, $E = E_0 \cos 2\omega_c t$, on several sections of lattice network. The current plotted is the current in the termination of the network. The frequency of the applied voltage is twice the resonant frequency, f_c, of the network.

the integral equation is not known from a table of integrals, one method for obtaining its solution is the expansion method developed above. This method may then have some application as a method for solving integral equations.

A. Illustrative Example

As an illustration of the use of this method in solving integral equations we will consider the equation

$$\frac{1}{\sqrt{(j\omega)^2 + 2\lambda j\omega}} = \int_0^\infty a(t)e^{-j\omega t}dt. \tag{46}$$

The expression on the left can be written

$$\frac{1}{\sqrt{j\omega}\sqrt{2\lambda + j\omega}}. \tag{47}$$

Noting that the square of the first factor has the form of an inductance and the second the form of a resistance and inductance in series we replace

$$j\omega \to R_{0_1}\left(\frac{1 - e^{-2P_1}}{2}\right) \qquad \text{where} \qquad R_{0_1} \to \infty;\ P_1 \to 0$$

$$R_{0_1}P_1 = j\omega$$

and $2\lambda + j\omega \to R_{0_2}\left(\frac{1 - e^{-2P_2}}{2}\right)$ where $R_{0_2} \to \infty$; $P_2 \to 0$; and $R_{0_2}P_2 = 2\lambda + j\omega$. We note that P_1 has the form $j\omega D$ where $R_{0_1}D = 1$, while P_2 has the form $A + j\omega D$ where $R_{0_2}A = 2\lambda$ and $R_{0_2}D = 1$. Substituting these values in (47), we have

$$\frac{1}{\sqrt{\frac{R_{0_1}R_{0_2}}{4}(1 - e^{-2j\omega D})(1 - e^{-2(A+j\omega D)})}}.$$

Expanding these two factors by the binomial theorem, we have

$$\begin{aligned}\frac{2}{\sqrt{R_{0_1}R_{0_2}}}\bigg[\,1 + \frac{1}{2}e^{-j(2\omega D)} &+ \frac{1/2 \times 3/2}{2!}e^{-j(4\omega D)} \\ &+ \cdots + \frac{(2n)!e^{-j(2n\omega D)}}{2^{2n}(n!)^2} + \cdots\bigg] \\ &\times\left[\,1 + \frac{1}{2}e^{-2(A+j\omega D)} + \cdots + \frac{(2n)!e^{-2n(A+j\omega D)}}{2^{2n}(n!)^2} + \cdots\right].\end{aligned}$$

At this point we make use of Stirling's theorems on factorials which states that when K is large

$$K! = \left(\frac{K}{e}\right)^K \sqrt{2\pi K}.$$

The typical terms of the above expression become

$$\frac{(2n)!e^{-j(2n\omega D)}}{2^{2n}(n!)^2} = \frac{\left(\frac{2n}{e}\right)^{2n}\sqrt{4\pi n}\; e^{-j(2n\omega D)}}{2^{2n}\left(\frac{n}{e}\right)^{2n} 2\pi n} = \frac{e^{-j(2n\omega D)}}{\sqrt{\pi n}}.$$

Inserting this value in the above expression and multiplying, there results

$$\sqrt{\frac{4}{R_{0_1}R_{0_2}}}\Bigg[1 + e^{-j(2\omega D)}[1/2e^{-2A} + 1/2] + \cdots + e^{-j(2n\omega D)} \left(\frac{e^{-2nA}}{\sqrt{\pi n}} + \frac{e^{-2(n-1)A}}{2\sqrt{\pi(n-1)}} + \cdots + \frac{e^{-2(n-k)A}}{\sqrt{\pi(n-k)}\sqrt{\pi k}} + \cdots + \frac{1}{\sqrt{\pi n}}\right) + \cdots\Bigg]. \tag{48}$$

Since the value of $a(t)$ is given by the factor multiplying the term $e^{-j2n\omega D}$ divided by $2D$ we can write

$$a(t) = \frac{2}{\sqrt{R_{0_1}R_{0_2}}2D}\left[\frac{e^{-2nA}}{\sqrt{\pi n}} + \frac{e^{-2(n-1)A}}{2\sqrt{\pi(n-1)}} + \cdots + \frac{e^{-2(n-k)A}}{\sqrt{\pi(n-k)}\sqrt{\pi k}} + \cdots + \frac{1}{\sqrt{\pi n}}\right]. \tag{49}$$

This expression can be written as the sum

$$a(t) = \sum_{k=0}^{k=n} \frac{e^{-2(n-k)A}}{\pi\sqrt{(n-k)k}}.$$

We introduce now the value $n = t/2D$ and define a new variable τ by $k = \tau/2D$. Inserting these values in the above summation and noting that $A/D = 2\lambda/R_{0_2}/1/R_{0_2} = 2\lambda$, we have

$$a(t) = \sum_{k=0}^{k=n} \frac{e^{(t-\tau)2\lambda}2D}{\pi\sqrt{(t-\tau)\tau}}. \tag{50}$$

But $2D = d\tau$, the element of time, so that the summation can be written as the integral

$$a(t) = \frac{e^{-2\lambda t}}{\pi}\int_0^t \frac{e^{2\lambda\tau}d\tau}{\sqrt{(t-\tau)\tau}}. \tag{51}$$

The value of the integral is $\pi e^{\lambda t}I_0(\lambda t)$, where $I_0(\lambda t)$ is the Bessel's functions $J_0(i\lambda t)$, when $i = \sqrt{-1}$. To show this we can expand the exponential and integrate the series giving

$$\pi\left[1 + \lambda t + \frac{1.3(\lambda t)^2}{(2!)^2} + \frac{1.3\cdot 5(\lambda t)^3}{(3!)^2} + \cdots\right].$$

This can be recognized as the series expansion of $e^{\lambda t}I_0(\lambda t)$. Hence the value of (51) is

$$a(t) = e^{-\lambda t} I_0(\lambda t), \tag{52}$$

which is the solution of the integral equation.

IV. Other Types of Boundary Conditions

The solutions obtained before are all on the assumption that no energy exists in the network before the voltage is applied. Other types of boundary conditions are sometimes desirable, but these can all be derived from the above solutions.

The next most important case is the case where the network has come to its equilibrium value and the voltage is suddenly taken off. This condition is the same as would result if a negative voltage E were suddenly applied to the circuit, and hence the solution is the steady state value of the current minus the current which flows on application of the voltage E.

Another type of boundary condition sometimes occurring is the one where energy exists in the network when t is zero. This may be taken account of by assuming that the voltage is applied before t equals zero. To take account of this condition analytically, examine the expansion

$$i = \frac{E}{Z} = E[a_0 + a_1 e^{-2j\omega D} + a_2 e^{-4j\omega D} + \cdots + a_n e^{-2nj\omega D} + \cdots].$$

According to the above assumption, the voltage is applied when $t = -t_0$, hence for n we substitute

$$n = \frac{t}{2D} + \frac{t_0}{2D} \cdot$$

The above series is then replaced by the integral

$$i = E\int_{-t_0}^{t} a(t + t_0) e^{-j\omega(t+t_0)} dt. \tag{53}$$

Another boundary condition of interest occurs when the voltage is taken off before an equilibrium value has been reached. If we count time as starting when the voltage is taken off, or what amounts to the same thing, when a negative voltage is applied, the symbolic solution takes the form

$$i = E[\int_{-t_0}^{t} a(t + t_0) e^{-j\omega(t+t_0)} dt - \int_0^t a(t) e^{-j\omega t} dt]. \tag{54}$$

Appendix

The expression

$$\left[\frac{4R}{R_0}\left(\frac{1}{1 - e^{-2(R/R_0+P)}}\right) - 1\right]^m$$

can be expanded into the form

$$\left[\frac{4R}{R_0}\left(\frac{1}{1 - e^{-2(R/R_0+P)}}\right)\right]^m - m\left[\frac{4R}{R_0}\left(\frac{1}{1 - e^{-2(R/R_0+P)}}\right)\right]^{m-1} + \cdots + (-1)^m. \quad (55)$$

and hence the general solution depends only on the solution of the general form

$$\left[\frac{4R}{R_0}\left(\frac{1}{1 - e^{-2(R/R_0+P)}}\right)\right]^m. \quad (56)$$

If equation (56) is expanded by the binomial theorem, there results the expression

$$\left(\frac{4R}{R_0}\right)^m\left[1 + me^{-2(R/R_0+P)} + \frac{m(m+1)}{2!}e^{-4(R/R_0+P)} + \cdots + \frac{(m+n-1)!e^{-2n(R/R_0+P)}}{n!(m-1)!} + \cdots\right]. \quad (57)$$

For any value of m, we can write the typical term of (57) as

$$\frac{(m+n-1)!}{n!(m-1)!} = \frac{\left(\frac{m+n-1}{e}\right)^{m+n-1}\sqrt{2\pi(m+n-1)}}{\left(\frac{n}{e}\right)^n\sqrt{2\pi n}(m-1)!}$$

$$= \frac{e^{-(m-1)}\left(\frac{m+n-1}{n}\right)^n(m+n-1)^{m-1/2}}{(m-1)!\sqrt{n}}. \quad (58)$$

by Stirling's theorem on factorials. Now n for any finite value of time approaches infinity, while m for any finite term in the series is finite. Hence (58) can be written as

$$\frac{e^{-(m-1)}\left(1 + \frac{m-1}{n}\right)^n n^{m-1}}{(m-1)!}.$$

The limit of $\left(1 + \frac{m-1}{n}\right)^n$ as $n \to \infty$ is $e^{(m-1)}$.[6] Hence

$$\frac{(m+n-1)!}{n!(m-1)!} = \frac{n^{m-1}}{(m-1)!}.$$

[6] See "Probability and Its Engineering Uses," T. C. Fry, page 107.

The value of (57), then reduces to

$$\left(\frac{4R}{R_0}\right)^m \sum_{t=0}^{t=n} \frac{n^{m-1}}{(m-1)!} e^{-2n(R/R_0+P)}. \tag{59}$$

If now we substitute $n = t/2D$, $P = j\omega D$; $R_0 P = j\omega L$ (59) reduces to the integral

$$\left(\frac{4R}{R_0}\right)^m \times \frac{1}{(2D)^m} \int_0^t \frac{t^{m-1} e^{-[(1/\sqrt{LC})+j\omega]t}}{(m-1)!} dt$$
$$= \left(\frac{2}{\sqrt{LC}}\right)^m \times \frac{1}{(m-1)!} \int_0^t t^{m-1} e^{-t[(1/\sqrt{LC})+j\omega]} dt. \tag{60}$$

If we integrate (60) by parts, successively, there results the series

$$2^m \left\{ \left(\frac{1}{1+j\omega\sqrt{LC}}\right)^m - e^{-t(1/\sqrt{LC}+j\omega)} \left[\frac{1}{(1+j\omega\sqrt{LC})^m} + \frac{\left(\frac{t}{\sqrt{LC}}\right)}{(1+j\omega\sqrt{LC})^{m-1}} + \cdots + \frac{\left(\frac{t}{\sqrt{LC}}\right)^{m-1}}{(m-1)!\,(1+j\omega\sqrt{LC})} \right] \right\}. \tag{61}$$

The complete solution of (55) is then obtained by adding terms of the kind given in (61). For example the steady state term is given by the series

$$\frac{2^m}{(1+j\omega\sqrt{LC})^m} - \frac{m2^{m-1}}{(1+j\omega\sqrt{LC})^{m-1}} + \frac{\frac{m(m-1)}{2!}2^{m-2}}{(1+\omega\sqrt{LC})^{m-2}} + \cdots + (-1)^m$$
$$= \frac{2^m - m2^{m-1}(1+j\omega\sqrt{LC}) + \frac{m(m-1)}{2!}2^{m-2}(1+j\omega\sqrt{LC})^2 + \cdots + (-1)^m(1+j\omega\sqrt{LC})^m}{(1+j\omega\sqrt{LC})^m}.$$

This is readily seen to be the binomial expansion of

$$\left[\frac{2-(1+j\omega\sqrt{LC})}{1+j\omega\sqrt{LC}}\right]^m = \left[\frac{1-j\omega\sqrt{LC}}{1+j\omega\sqrt{LC}}\right]^m. \tag{62}$$

The other terms given in equation (38) follow directly by addition and reference to equations (55) and (61).

Acoustic Considerations Involved in Steady State Loud Speaker Measurements

By L. G. BOSTWICK

SYNOPSIS: Certain difficulties encountered in acoustic measurements of the performance of loud speakers are described. Because of the nature of these difficulties it has not yet been possible to specify a complete and simple set of measurements or conditions which will completely express the performance of a loud speaker. Data are given showing the performance of two representative types of loud speakers both when measured in outdoor space free from reflections and when measured under varying conditions in a specially treated acoustic laboratory. The differences serve to emphasize the importance of certain precautions in the making of indoor acoustic measurements.

SCOPE

IN view of the general misconception of the meaning of many claims which are made regarding the operation of loud speakers, it appears desirable to discuss in some detail the requirements which should be taken into account in making measurements for setting up ratings of loud speaker performance. For example, claims to "uniform response at all frequencies in the audible range" or "flat characteristic" can not be accurately applied to loud speakers which have thus far been made available. In many cases the claims for a loud speaker are based upon carefully made electrical measurements but these are often obtained in such a manner and under such conditions that they do not represent the performance of the loud speaker as it would be normally observed, and therefore are misleading. The main consideration in making loud speaker measurements is not the electrical circuit arrangement or apparatus of the measuring system but rather the acoustic conditions under which the magnitude of the sound output of the loud speaker is determined. It is the purpose of this paper to discuss steady state loud speaker measurements particularly from the standpoint of the more important acoustic factors which are involved and which must be properly considered in order to be able to interpret the significance of any measurements obtained.

LOUD SPEAKER PERFORMANCE INDICES AND FACTORS INVOLVED IN THEIR DETERMINATION

Efficiency.—In power engineering and other branches of engineering, efficiency (a measure of the degree to which a device performs the functions for which it is designed) is defined as the ratio of the power delivered to a load to the power absorbed from a source of supply. Since in power transmission systems the purpose of a machine is to draw a limited amount of power from a relatively unlimited source and to deliver this power to a load with a minimum loss in the machine itself, this ratio constitutes a useful measure of the performance. If it were of interest a similar quantity could likewise be used as a measure of the performance of a loud speaker. In this latter case however, the function is not in general to draw a limited amount but as much power as possible from a supply source and to radiate maximum power to the air or load. A measure of the efficiency would therefore have to involve the ability of the loud speaker to take maximum power from the supply and might be defined as the ratio of acoustic power P_A radiated to the maximum electrical power P_E which the supply circuit is capable of delivering under optimum impedance conditions. Thus the efficiency η at a specified frequency would be defined by the ratio

$$\eta = \frac{P_A}{P_E}. \tag{1}$$

Assuming the impedance of the electrical supply source for a loud speaker essentially non-reactive (as is almost invariably the case) and of a constant magnitude r suitable to the requirements of the loud speaker, then the maximum power which the supply circuit is capable of delivering under optimum impedance conditions with an open circuit supply voltage e would be

$$P_E = \frac{e^2}{4r}. \tag{2}$$

These quantities are all readily measureable. The determination of the quantity P_A however is more difficult.

For measuring the acoustic energy or power stored in or transmitted through the medium adjacent to a loud speaker, the condenser transmitter is probably the most suitable free space acoustic measuring device. The ruggedness of this transmitter for an instrument of this type and the straightforward manner in which it can be used recommend it for practical loud speaker measurements. The condenser transmitter is not, however, an acoustic power indicating device but is a device having a high impedance compared to the impedance of the acoustic system in which it is used. It is therefore an acoustic measuring device which is analogous to an electrical voltmeter and can be calibrated by the thermophone [1] or other means to measure the excess pressure in a medium resulting from a sound wave. Other acoustic measuring devices such as the Rayleigh disc, thermal devices

[1] "The Thermophone," E. C. Wente, *Physical Review*, Vol. XIX, No. 4, April, 1922.

of different kinds, etc. can of course be used but these in general are considerably more difficult to use than the condenser transmitter, especially for free space acoustic measurements and the measured quantities bear no more simple relation to the acoustic power or energy.

Assuming the condenser transmitter then to be the acoustic measuring or indicating device, the problem becomes one of how and where in the medium to measure the pressure so that the measurement will bear some readily deducible relation to the acoustic power delivered by the loud speaker. The answer depends upon the nature of the acoustic medium in which the measurements are made. The simplest relations between excess r.m.s. pressure in the medium and the acoustic power exist when the pressure measurements are made in an infinite medium or in a room in which the incident energy at the walls is completely absorbed. Under such conditions the acoustic power from a loud speaker could be obtained by measuring the pressure at all points on the surface of a sphere having a radius several times that of the largest dimension of the sound radiating surface and with the sound radiating surface at the center of the sphere. The acoustic power would then be

$$P_A = \frac{1}{\rho c}\int\int p^2 ds,[2] \tag{3}$$

where ρ is the density of the air; c is the velocity of sound propagation; p is the excess r.m.s. pressure; and ds is the surface of the sphere. This relation, however, is generally true only when the radius of the spherical surface is sufficiently large so that the sound radiating surface appears as a point source. From equations (1), (2) and (3) the efficiency of a loud speaker could then be expressed in terms of excess pressure measurements over the surface of the sphere in an infinite medium as follows:

$$\begin{aligned} \eta = \frac{P_A}{P_E} = \frac{1}{\rho c}\frac{\int\int p^2 ds}{\frac{e^2}{4r}} \\ = \frac{K\int\int p^2 ds}{\frac{e^2}{r}}. \end{aligned} \tag{4}$$

Within an enclosure where there are sound reflections from the bounding surfaces, the determination of the acoustic power delivered by a loud speaker would involve the measurement of the pressure at

[2] "Theory of Vibrating Systems and Sound," Crandall, pp. 92 and 120.

all points within the enclosure in order to obtain the average energy density and then making use of known relations between energy density and the rate of energy flow into the room. Under steady state conditions, the total potential energy stored within the room would be

$$E_p = \frac{1}{2\rho c^2}\int\int\int p^2 dv.^{3} \tag{5}$$

Assuming the room to be large so that the region within say a wave-length of the loud speaker is a small portion of the total volume of the room, it can be said with a reasonable approximation that the potential and kinetic energies stored within the room as the sound is transmitted are equal. The total energy would therefore be twice the potential energy or

$$E = 2E_p = \frac{1}{\rho c^2}\int\int\int p^2 dv. \tag{6}$$

This latter statement may be roughly justified in a simple manner by considering the sound radiated by the loud speaker as consisting of two components, one of which is completely absorbed at the walls and the other completely reflected. In considering separately that component which is absorbed, the loud speaker can be thought of as in an infinite medium and under these conditions (excluding the region within say a wave-length of the loud speaker) the acoustic impedance of the medium is essentially non-reactive. The potential and kinetic energies of the sound transmitted would therefore be equal. That component which is transmitted to the medium and completely reflected at the walls produces an ideal standing wave system. In such a system along the direction of the standing wave the total energy is alternately all kinetic and all potential and since this transition takes place the potential and kinetic energies must be equal.

Considering the room volume V, the average energy density would be

$$\overline{E} = \frac{1}{\rho c^2 V}\int\int\int p^2 dv. \tag{7}$$

If the loud speaker emits power into the room at a rate P_A, the average energy density in the room after a steady state has been reached is

$$\overline{E} = \frac{4P_A}{ac},^{4} \tag{8}$$

[3] "The Dynamical Theory of Sound," Lamb, p. 208, Second Edition.
[4] "Theory of Vibrating Systems and Sound," Crandall, p. 210.

where a is the absorbing power of the room obtained from the sum of the products of the areas of the absorbing surfaces in the room and their respective absorption coefficients. From (7) and (8) the acoustic power delivered by the loud speaker could then be expressed in terms of the excess r.m.s. pressure throughout a large room as follows:

$$P_A = \frac{a}{4V\rho c}\int\int\int p^2 dv \tag{9}$$

and the efficiency of the loud speaker would therefore be

$$\begin{aligned}\eta = \frac{P_A}{P_E} &= \frac{\frac{a}{4V\rho c}\int\int\int p^2 dv}{\frac{e^2}{4r}} \\ &= \frac{K'\int\int\int p^2 dv}{\frac{e^2}{r}}.\end{aligned} \tag{10}$$

Response.—By determining the mean square pressure at all points in the measuring room or at all points on the surface of a large sphere in an infinite medium as discussed above, it is therefore possible to measure the "efficiency" of a loud speaker with a pressure indicating device. Such a method for determining the merits of a loud speaker at all frequencies of usual interest, however, would obviously be quite impracticable. Furthermore, unless the radiation from the loud speaker is uniform over a spherical surface it is not of particular interest to know the magnitude of the total acoustic power or the value of the quantity η since the configuration of the sound field about a loud speaker may change decidedly with frequency, with the result that variations in sound loudness at different frequencies in a particular region may be large even though the total power output from the loud speaker may be constant. In order then that the measured characteristic shall convey a true idea of the performance as it might be observed by the ear, the square of the pressure at one representative listening position or the average of the squares of the pressures in a small region wherein an observer might normally be located may be considered instead of the average throughout the room. In this manner a sort of specific efficiency measure would be obtained in that it is a measure of the efficiency with respect to the acoustic power transmitted through the specified position or region. Throughout the remainder of this paper, this specific measure of the efficiency is called the "response" of the loud speaker as measured at a specified position or positions and is expressed in transmission units (TU) relative to a fixed arbitrary reference condition of 1 volt, 1 ohm, and 1 bar. In other words the acoustic power density at a specified position or the average acoustic power density at specified positions in the medium produced by the loud speaker under test per unit of available electrical supply power, is expressed relative to the acoustic power density produced by a fictitious standard loud speaker which when placed in the location of the loud speaker under test will produce a mean square pressure of one unit at the specified position or positions in the room when the ratio

$$\frac{e^2}{r} = 1.$$

The response in TU is thus expressed by the relation:

$$\mathrm{TU} = 10 \log_{10} \frac{\dfrac{K'\overline{p^2}}{\dfrac{e^2}{r}}}{K'\dfrac{\dfrac{1}{1^2}}{1}} = 10 \log_{10} \frac{\overline{p^2}}{\dfrac{e^2}{r}}$$

$$= 20 \log_{10} \frac{\bar{p}}{\dfrac{e}{\sqrt{r}}},$$

where $\bar{p}$ is the r.m.s. pressure at a specified position or the square root of the mean square r.m.s. pressures at specified positions in the medium; and e and r are as defined above.

Measuring Considerations in a Reflectionless Medium.—In a medium where there are no sound reflections from the bounding surfaces, two factors are most likely to cause the measured response of a loud speaker to vary with frequency. These factors are independent and their effects of about equal importance. The first and most apparent is the inherent dynamical characteristics of the loud speaker which involves its ability to transfer maximum power from the electrical supply to the acoustic medium. Any variation with frequency in the acoustic power output of a loud speaker when supplied by constant, available electrical power will, of course, cause corresponding variations in the response provided the square of the pressure at the measuring position is indicative of the power transmitted through this position. In order that this latter condition be strictly true, the measuring position in general should be, at a distance from the loud speaker, large compared to the dimensions of the radiating surface. Otherwise, the wave front at the measuring position would not be spherical and the indicated pressure might result largely from the cyclic storage and absorption of energy by the loud speaker in the immediate vicinity of the radiator. With a proper location of the condenser transmitter, however, a response-frequency characteristic provides a useful measure of the dynamical perfection of a loud speaker.

The second factor which may cause large variations in the response of a loud speaker in a reflectionless medium is the change in the space distribution of the radiated sound with frequency. Although the total acoustic power delivered by a loud speaker may be constant, the power density at certain positions in the medium may change greatly with frequency due to the interference of sound originating at different parts of the radiating surface. Unless the radiation from the loud speaker is spherical, this interference phenomenon will result in a concentration of sound power in certain regions in the medium and a diminution in others. The locations of these regions change with frequency, radiator dimensions and the mode of vibration of the radiating surface.

For the case of a piston diaphragm radiator in a large rigid wall, it is possible to calculate the variations with frequency in the excess pressure at points in the sound field. Such calculations [5] and confirming experimental data show that in the sound field along the center perpendicular (a line normal to the surface of the piston at the center) to a piston radiator there is a succession of sound pressure maxima and minima out to a distance equal to approximately $\frac{D^2f_1}{4500}$ feet (where D is the piston diameter and f_1 is the energizing frequency). Beyond this distance these maxima and minima points disappear and the pressure varies inversely as the distance. If then, the response of a loud speaker with a piston diaphragm is measured with the condenser transmitter at a distance less than $\frac{D^2f}{4500}$ feet (where f is the highest measuring frequency), the response-frequency characteristic will have a succession of peaks and depressions which are independent of but which may be difficult to distinguish from those caused by poor dynamical characteristics of the loud speaker itself. On the other hand, if the condenser transmitter is located at any distance greater than $\frac{D^2f}{4500}$ feet, the response-frequency characteristic obtained will not

[5] "The Directional Effect of Piston Diaphragms," Backhaus and Trendelenburg, *Zeitschrift f. Techn. Physik.*, Vol. 7, pp. 630–635, 1926. Also "Theory of Vibrating Systems and Sound," Crandall, pp. 137–149.

have abrupt irregularities due to interference and any two curves so obtained will only differ in magnitude. A curve obtained under this latter condition would therefore show the response-frequency variations as these would be observed at any distance greater than $\frac{D^2 f}{4500}$ feet in this same direction.

While the above facts relate to a piston diaphragm radiator because more definite statements can be made regarding its sound field, similar effects due to the irregular distribution of the sound field are involved in the case of any other loud speaker which does not radiate as a pulsating sphere. Using such piston diaphragm considerations as a basis it has been found possible to predict suitable measuring conditions for any particular loud speaker. The fundamental requisite is that the pressure indicator be located at a distance from the loud speaker commensurate with the typical listening distance in order that the response-frequency characteristic shows variations which would normally be observed and which are therefore of interest. If, however, the typical listening distance is greater than $\frac{D^2 f}{4500}$ feet (where D is roughly the diameter of the radiating surface and f is the highest measuring frequency) response-frequency measurements at this distance will show the response-frequency variations at any greater distance in the same direction so that measurements at the greater distances would not be necessary. If the most likely position of a listener is at a distance less than $\frac{D^2 f}{4500}$ feet, the response-frequency characteristic obtained with the condenser transmitter at such a position will have irregularities due to interference but since these irregularities would be heard they should be charged against the loud speaker and such a curve would be indicative of the performance. In this latter case response measurements at a distance greater than $\frac{D^2 f}{4500}$ feet are sometimes valuable for loud speaker design work in order to distinguish those variations due to poor dynamical characteristics of the loud speaker itself and those due to poor sound field distribution characteristics.

Measuring Considerations in a Medium with Reflections.—If the sound energy reflected to the condenser transmitter position from the bounding surfaces of the medium is comparable in magnitude with the energy reaching this position directly from the loud speaker, standing waves will exist and the sound pressure may vary greatly with frequency at any fixed transmitter position although the acoustic power transmitted through this position may be constant. Response measurements with the condenser transmitter at any one position can therefore mean very little under such conditions.

To our knowledge it is practically possible only by working outdoors under very particular conditions to obtain a medium sufficiently free from reflections to make suitable response measurements at all frequencies with the condenser transmitter located at any one position. By using a room with all dimensions very large compared to the distance between the condenser transmitter and the loud speaker (which distance is determined by the size of the loud speaker and the highest measuring frequency as discussed above) and covering the walls with sound absorbing material, it is possible to reduce the reflected energy at the transmitter position to a small value over a considerable frequency range but any practical method of reducing the reflected sound to a negligible value at all frequencies of interest in loud speaker measurements is as yet not available.

In a plane standing wave system the energy density at points of maximum pressure or minimum velocity is equal to $\frac{p^2}{\rho c^2}$, where p is the r.m.s. pressure at these points. The locations of these maximum pressure points change with frequency but if the position of the condenser transmitter in loud speaker response measurements is changed at each measuring frequency to a maximum pressure point within a suitable region the indicated pressure will be a measure of the energy transmitted through this region. The measured response would then be approximately the same as would be measured in a reflectionless medium except for a magnitude difference due to the addition of the reflected energy. Such a procedure for loud speaker response measurements indoors would thus be suitable if it were not for the fact that the standing wave system in the room is usually of a very complicated configuration in three dimensions instead of being simple. The probability of being able to locate a position in any desired region of the sound field of a loud speaker where the pressures of each of the standing waves which may traverse this position are a maximum, is obviously remote.

A method of response measurement making use of the mean square pressure instead of the maximum value is more practicable. With a single frequency sinusoidal sound source, the pressure-space distribution of each of the standing waves in a room is likewise sinusoidal. The maximum r.m.s. pressure squared of each standing wave would then be twice the mean r.m.s. pressure squared over a half wave-length or any multiple of a half wave-length; also approximately twice the mean over any distance large compared to a wave-length since this latter average approaches the half wave-length mean. The mean square pressure is therefore just as suitable as the maximum value as a measure of the energy density and further, it lends itself more readily to the determination of the energy density in the case where there are standing waves in several directions. For this latter case the energy density within a specified region is proportional to the mean square pressure in all directions or at all points within the volume of a sphere having a diameter large compared to the wave-length. The response of the loud speaker at any frequency can accordingly be measured in a room with reflections by averaging the squares of the pressures throughout a suitable volume.

The above method of response measuring indoors however, is not entirely independent of the measuring room. If the reflected energy is large the response measurements will be affected by the variation in the absorption power of the room with frequency so that a large room with absorbing material having as uniform and large absorption characteristics as possible over the measuring frequency range is still desirable. Some sound absorbing materials have uniform absorption characteristics but when this is the case the absorbing power is apt to be very low. The use of such materials results in extremely large pressure variations within the room so that a measuring device having a sufficient amplitude range to average the squares of the pressures is difficult to obtain. For this reason and because of the fact that the region through which it is necessary to average the squares of the pressures at low frequencies becomes prohibitive, a large room is most desirable so that the difference between the direct and reflected energies at the transmitter position will be as large as possible. Indoor measurements under these conditions approach infinite medium measurements.

The use of a large room also results in less reaction of the room inclosure on the loud speaker itself. While under most conditions such reactions have little effect on the acoustic output power of the loud speaker, in small measuring rooms at very low frequencies where the absorption is low and the radiator of the loud speaker (perhaps designed for a large auditorium) is large, the phase and magnitude of the reflected energy at the radiating surface of the loud speaker may be such as to cause large variations in the acoustic impedance of the medium on the area adjacent to this surface. This variation in the acoustic impedance of the loud speaker load will cause variations in the acoustic power density at the transmitter position. Response measurements on loud speakers of large dimensions and particularly measurements on such devices at very low frequencies are consequently more indicative of the capabilities of the loud speaker when obtained outdoors or in a very large room.

Experimental Data

General.—To determine the extent to which the above discussed acoustic effects may influence the result of a loud speaker performance measurement and to show the measuring conditions under which such acoustic factors are encountered, response-frequency characteristics of loud speakers were measured under various acoustic conditions. These are described in the following paragraphs.

In all these measurements a calibrated condenser transmitter approximately $2\frac{7}{8}''$ in diameter was used as the acoustic detector. The thermophone calibration on this transmitter showed that with constant pressure on the diaphragm the voltage produced between the grid and filament of the associated vacuum tube was very nearly independent of frequency. Corrections for such small variations as did exist however have been made in all the following curves. In addition a tapered correction of .6 TU per 100 cycles increased in frequency between 1,100 and 2,100 cycles and a constant correction of 6 TU for frequencies above 2,100 cycles have been subtracted from the response measurements to correct for the fact that the indicated pressure approaches twice the free space pressure at frequencies at which there is reflection from the diaphragm. This latter correction was obtained by making response-frequency measurements on a loud speaker under a fixed set of conditions; first, with the condenser transmitter freely suspended in the sound field as in all the following curves and then with the transmitter located at the center of a round baffle 12″ in diameter. When in the baffle complete sound reflection from the transmitter occurred at a frequency lower than that at which reflection began to take place from the transmitter alone. From the difference between the two response-frequency curves so obtained, it was therefore possible to definitely locate the transition frequency range between 1,100 and 2,100 cycles and to evaluate the transmitter reflection correction.

The number of measurements made in order to define any response-frequency characteristic depended upon the nature of the curve. If no abrupt changes in the response were observed in making the measurements, approximately 10 measurements per octave were obtained. Otherwise the frequency of the oscillator was changed by small steps and a sufficient number of measurements made to clearly define the curve.

Measuring System.—The circuit arrangement of the measuring system used in making the measurements is shown in schematic form on Fig. 1. An oscillator having a suitable frequency range and power output was alternately connected by means of a two-position switch through a transformer to the loud speaker under test and to the input terminals of an attenuator calibrated in TU. The transformer had a ratio such that the loud speaker being measured was always connected to an impedance equal to that for which it was designed to be connected. The condenser transmitter in series with the output terminals of the attenuator and located in the medium as will be discussed later, was connected to a voltage indicating system consisting of an amplifier, a thermocouple and a meter. A low-pass filter was included in the

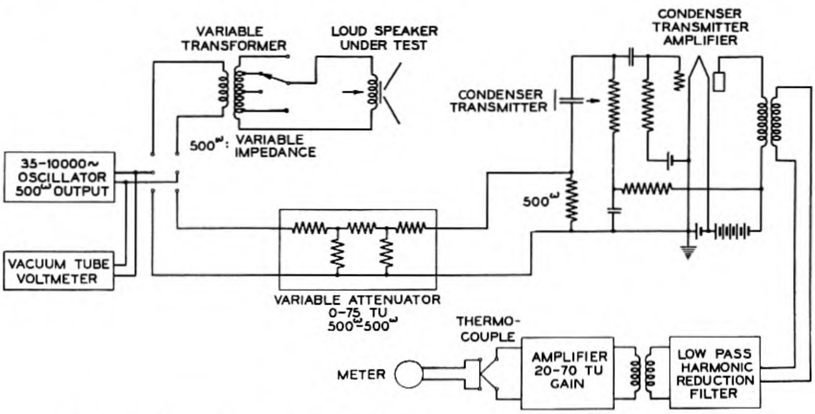


Fig. 1—Schematic circuit of loud speaker response measuring system.

indicating system as shown to insure that only the fundamental frequency of the output from the condenser transmitter or attenuator was indicated. The measuring procedure was as follows: The output or terminal voltage of the oscillator when open-circuited or connected to the attenuator was kept constant at all frequencies by means of a vacuum tube voltmeter. With the loud speakers considered the sound output over a wide magnitude range was proportional to the oscillator voltage so that the magnitude of this voltage was governed entirely by the sound pressure in the medium most suitable for making measurements. With the oscillator connected to the loud speaker (through the transformer) the sensitivity of the voltage indicating system was adjusted at each frequency until a mid-scale deflection of the meter was obtained as a result of the voltage generated by the condenser transmitter. After each adjustment the oscillator was then switched from

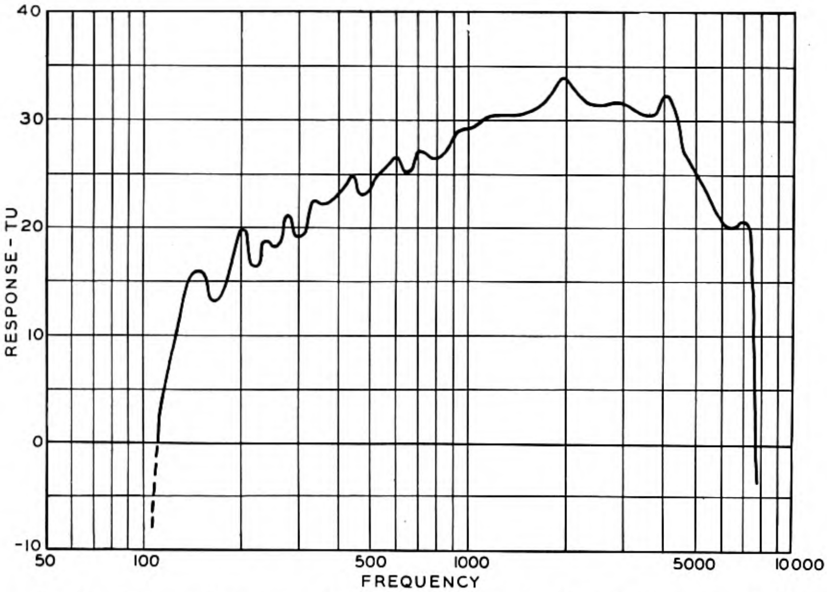


Fig. 2—Response-frequency characteristic of 115 cycle cut-off exponential horn with moving coil type receiver. Measured outdoors at a distance of 12′ from horn mouth on the axis.

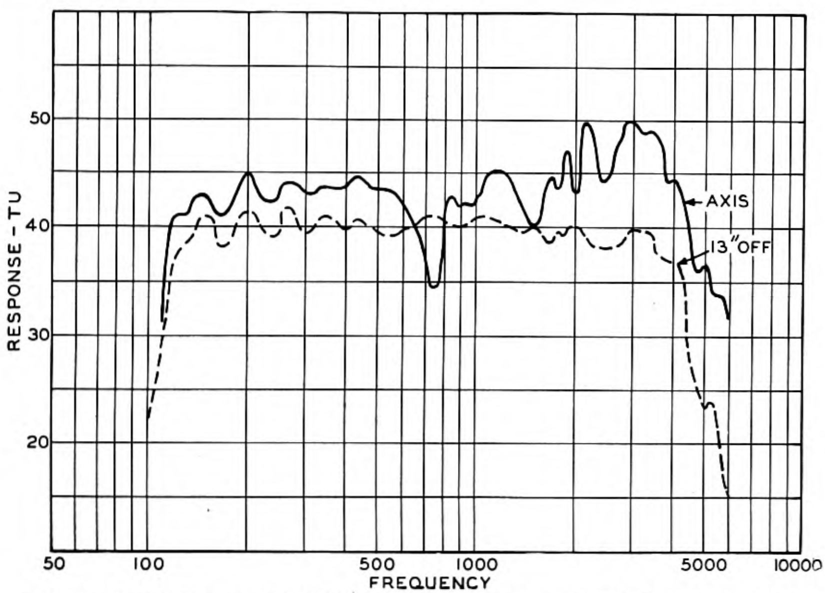


Fig. 3—Response-frequency characteristics of 115 cycle cut-off exponential horn with moving coil type receiver. Measured outdoors 2″ from plane of horn mouth with center of condenser transmitter diaphragm on horn axis and 13″ from axis.

the loud speaker to the input terminals of the attenuator and the attenuator adjusted to give the same meter deflection. The variations in the attenuator settings with frequency showed the variations in the performance of the loud speaker in TU. When the ratio of the open-circuit voltage of the oscillator to the square root of its output impedance equalled 1, the setting of the attenuator which gave a voltage between the attenuator output terminals equal to the voltage across the condenser transmitter terminals with a pressure of one bar on the diaphragm, gave the reference zero of one volt, one ohm, and one bar

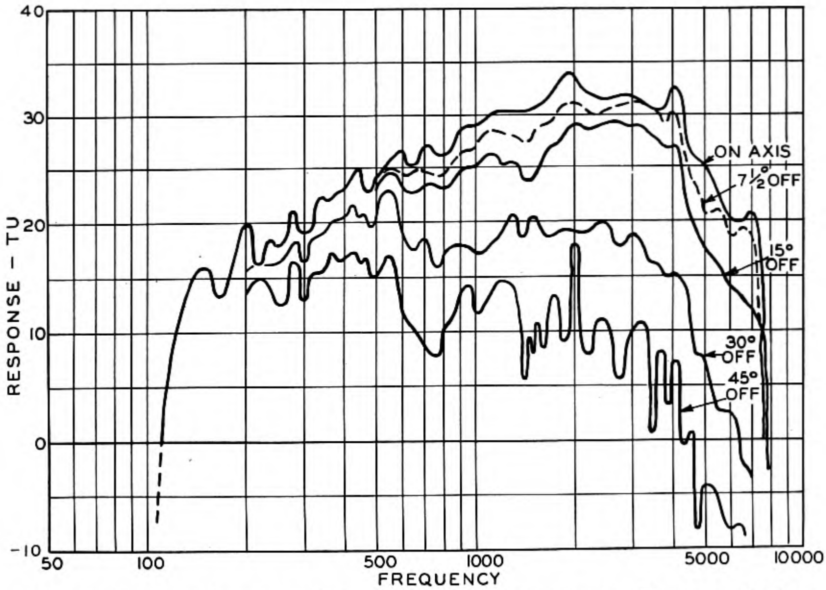


Fig. 4—Response-frequency characteristics of 115 cycle cut-off exponential horn with moving coil type receiver. Measured outdoors 12 feet from horn mouth at the specified angles with the axis.

described above. The response of the loud speaker was then read directly from the attenuator setting.

This measuring method has the commendable feature that the results obtained are independent of any variations from day to day in the sensitivity in the voltage indicating system. The voltage indicating system simply serves to compare the voltage produced by the condenser transmitter with the voltage across the output terminals of the attenuator and any variations in battery voltage, tubes, etc., or any variations in the amplification with frequency can in no way affect the accuracy of the measurements. Aside from the condenser transmitter, the only element in the measuring system which must be calibrated and maintain its calibration closely, is the attenuator which involves only a group of resistances.

Outdoor Measurements.—As an illustration of those acoustic effects involved in loud speaker measurements in a reflectionless medium, the following response data obtained outdoors in an open field will be of

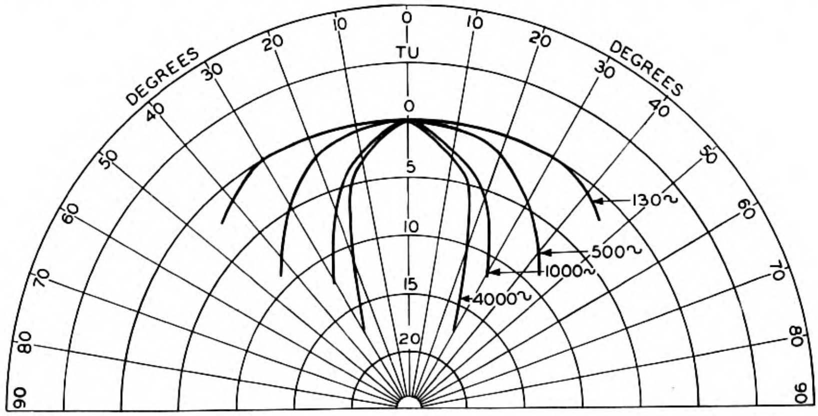


Fig. 5—Polar curves showing response (expressed relative to the axis response) of 115 cycle cut-off exponential horn with moving coil type receiver at various angles from horn axis and 12 feet from the mouth.

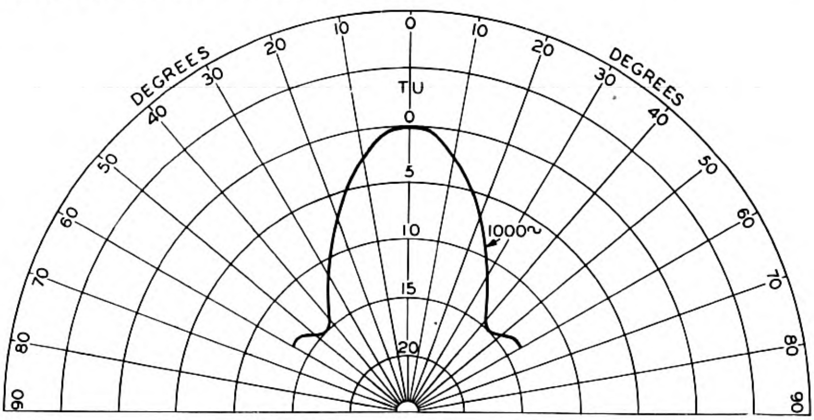


Fig. 6—Polar curve showing response (expressed relative to the axis response) of 115 cycle cut-off exponential horn with moving coil type receiver at different angles from axis and at a distance of 12 feet from center of horn mouth.

interest. Two loud speakers having uniform sound power output over a wide frequency range so that the dynamic characteristics would not obscure the acoustic effects were used. The loud speakers were placed at the edge of a skeleton platform approximately 15′ above the ground and the condenser transmitter suspended at various positions in the sound field. Care was taken to suspend the transmitter and its small associated amplifier between small poles in such a way that any possible reflections from such objects in the sound field would not reach the transmitter position. As for reflections from the ground, the distance of the loud speaker from the ground with the consequent sound divergence from the radiating surface, and also the absorption and diffraction at the ground, caused by the magnitude of the sound reflected to the transmitter position to be quite undetectable.

One of the loud speakers was a 115 cycle cut-off exponential horn with a moving coil type receiver.[6] The mouth of the horn was located at the platform edge with the axis making an angle upward from

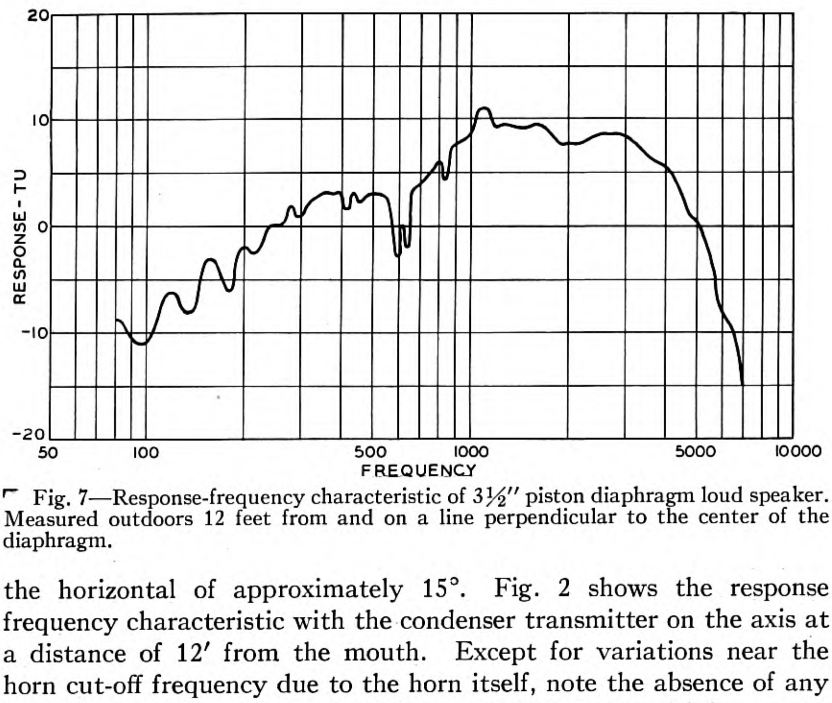


Fig. 7—Response-frequency characteristic of 3½″ piston diaphragm loud speaker. Measured outdoors 12 feet from and on a line perpendicular to the center of the diaphragm.

the horizontal of approximately 15°. Fig. 2 shows the response frequency characteristic with the condenser transmitter on the axis at a distance of 12′ from the mouth. Except for variations near the horn cut-off frequency due to the horn itself, note the absence of any large irregularities and also the rising trend of the curve with frequency. The diameter of the horn mouth was 30″ so that 12′ is greater than the distance $\frac{D^2f}{4500}$ feet discussed above.

Fig. 3 shows response-frequency characteristics of the same loud speaker measured under exactly the same conditions except that the condenser transmitter was located on the axis only 2″ from the horn

[6] This type of receiver was described by Wente and Thuras in *The Bell System Technical Journal*, for January, 1928.

mouth. This curve differs considerably from the one obtained at a distance of 12′. The marked depression in the curve at 750 cycles checks very closely the first interference frequency as calculated for a piston radiator approximately 30″ in diameter and allowing for a slight contraction of the radiating surface as the frequency is increased (which assumption would be quite reasonable for the horn), the irregularities at the higher frequencies are also explained in the same manner. For the piston, however, the minimum pressure point would be zero, which fact indicates that the wave-front at the horn mouth either is not plane or is not of uniform intensity over the radiating surface. Below 1,000 cycles the average trend of this curve is very nearly parallel to the axis of abscissæ while as noted for the curve

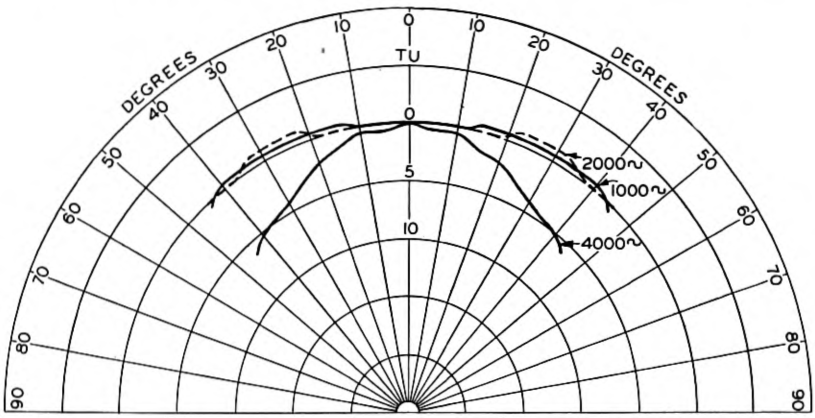


Fig. 8—Polar curves showing response (expressed relative to axis response) of 3½″ piston diaphragm loud speaker at various angles from perpendicular to center of diaphragm and 12 feet away.

obtained at a distance of 12′, there is a very definite downward slope. This is as would be expected if there were an increasing concentration of the sound field about the axis as the frequency increased. The fact that there is such a varying concentration is shown by the data on Fig. 4. These curves were obtained with the condenser transmitter at a distance of 12′ in each case, but with a line from the center of the horn mouth to the center of the transmitter making various angles with the horn axis as specified. In making these measurements, the transmitter remained fixed and the horn was rotated upward in a vertical plane about the center of the mouth. It is apparent from these curves that as the angle is increased the response at the higher frequencies becomes lower, while at lower frequencies the change is slight. The irregularities in the 45° curve are probably due to interference. Note that a curve of almost any desired trend may be obtained by locating the condenser transmitter at the proper position.

On Fig. 5 the data of Fig. 4 are plotted on polar coordinate paper to show more clearly the approximate manner in which the sound field varies. On these curves the magnitude of the response is expressed relative to that on the axis and the approximate distribution at each of four frequencies is shown. At the larger angles if a sufficiently large number of measurements are made an irregular interference pattern is obtained like that shown on Fig. 6 for 1,000 cycles.

The second loud speaker measured consisted of a $3\frac{1}{2}''$ piston diaphragm (inertia control) mounted in one side of a cubical box approxi-

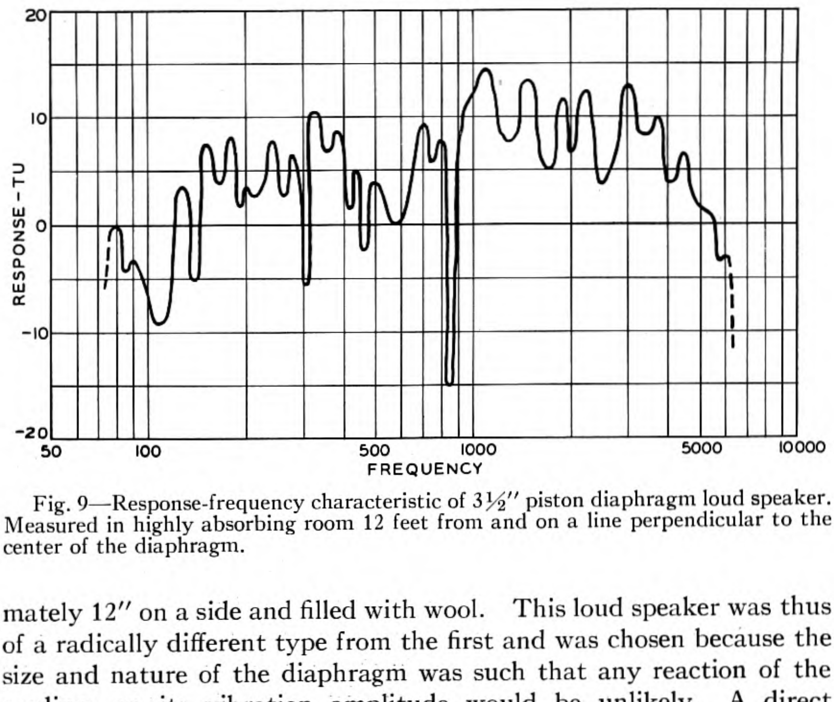


Fig. 9—Response-frequency characteristic of $3\frac{1}{2}''$ piston diaphragm loud speaker. Measured in highly absorbing room 12 feet from and on a line perpendicular to the center of the diaphragm.

mately 12″ on a side and filled with wool. This loud speaker was thus of a radically different type from the first and was chosen because the size and nature of the diaphragm was such that any reaction of the medium on its vibration amplitude would be unlikely. A direct comparison of its performance under one medium condition with that under another would therefore be justifiable. The response-frequency characteristic of this loud speaker as measured outdoors with the condenser transmitter at a distance of 12′ on the diaphragm center perpendicular is shown on Fig. 7. The irregularity in this curve at 600 cycles has been shown by other tests to be due to poor dynamical characteristics of the loud speaker itself.

On Fig. 8 are polar coordinate curves for the piston diaphragm loud speaker similar to those shown on Fig. 5 for the horn type loud speaker. Note for the same frequency the greater concentration of the sound field in the case of the horn. This is due to its larger radiating surface. From these data it might be inferred that if the sound field of a loud speaker of either of these types is to be the same at all frequencies, the size of its radiating surface must decrease as the frequency increases.

Indoor Measurements.—Making use of the above described outdoor measurements as standards of comparison, response measurements were made indoors on the same loud speakers and at the same relative positions in the sound field in order to determine the magnitude of the effect of reflections on such measurements. The room available for

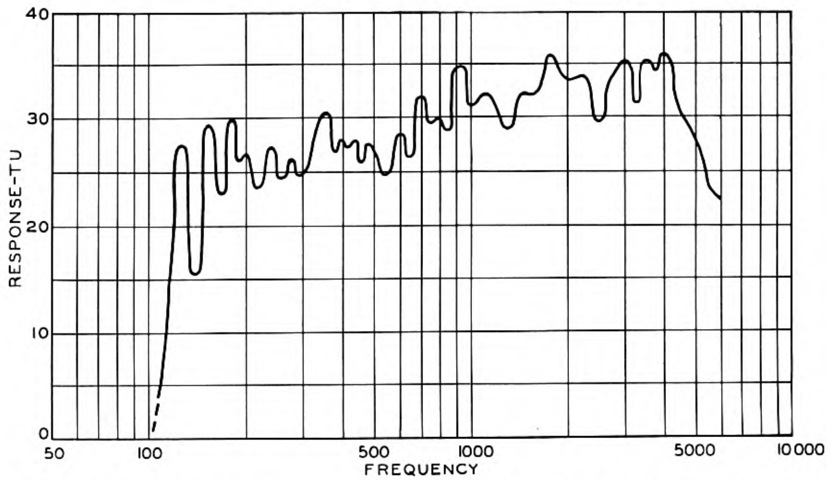


Fig. 10—Response-frequency characteristic of 115 cycle cut-off exponential horn with moving coil type receiver. Measured in highly absorbing room 12 feet from horn mouth on axis.

this work was approximately 34′ × 18′ × 9′. The six bounding surfaces of the room were covered with ½″ asbestos hair felt with heavy monks cloth curtains loosely and irregularly draped over the walls. The condenser transmitter and the loud speaker being measured were located along the major axis equi-distant from and on opposite sides of the center of the room and suspended about mid-way between the ceiling and the floor. A moderately large room with the usual precautions to eliminate reflections was thus used.

Fig. 9 shows the response frequency characteristic of the 3½″ piston diaphragm loud speaker as obtained in this room with the condenser transmitter located on the diaphragm center perpendicular at a distance of 12′. A comparison of this curve with that of the same loud speaker obtained outdoors with the condenser transmitter at the same relative position in the medium gives an idea of the magnitude of the effect of room reflections even under comparatively favorable indoor measuring conditions. The same frequencies were measured in both the indoor and outdoor curves and no attempt was made to locate all the irregularities in the indoor curve.

Fig. 10 shows an indoor curve of the 115 cycle cut-off exponential horn measured in the same room and under the same conditions at a distance of 12′. The variations in this latter curve as compared to the outdoor curve shown on Fig. 2 appear to be less than in the case of the piston type loud speaker, probably because the horn is more directive,

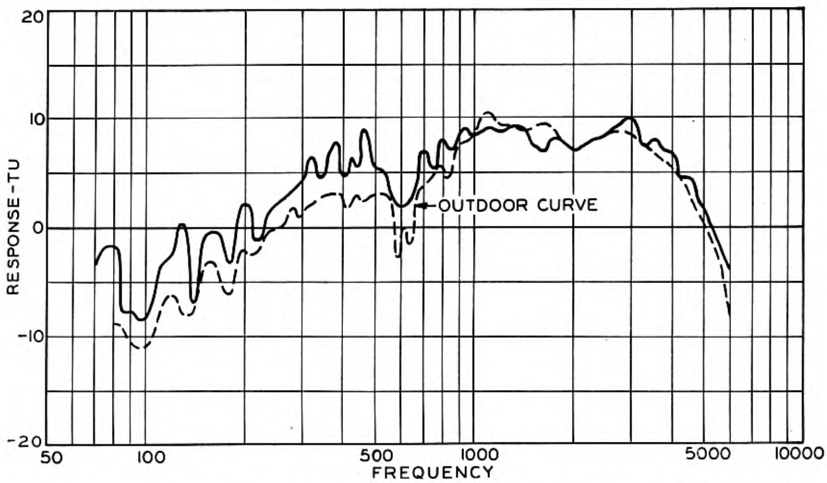


Fig. 11—Response-frequency characteristic of 3½″ piston diaphragm loud speaker. Measured in highly absorbing room 12 feet from diaphragm with rotating condenser transmitter.

resulting in a larger difference between the direct and the reflected sound energy at the transmitter position.

The variations in these two indoor response-frequency characteristics resulting from reflections could have been reduced by making the measurements in a much larger room treated with sound absorbing material in the same manner. Such a room for loud speaker measuring purposes is not usually available, but if sufficiently large and of a suitable shape does afford the most satisfactory indoor measuring conditions especially at very low frequencies.

Another method of obviating the effects of reflections is to measure the mean square pressure throughout a suitable volume as discussed previously. A practicable means of making such a mean square pressure measurement and one which has been found to be quite satisfactory, consists in approximating the volume measurement by rotating the condenser transmitter in a circle 69″ in diameter, inclined 45° to the horizontal. A mechanism so arranged that the plane of the condenser transmitter diaphragm always remains perpendicular to the loud speaker axis is used and the condenser transmitter is connected to the same voltage indicating system described above. As noted, this indicating system involves a thermocouple and as a result, the meter deflection is very nearly proportional to the square of the input

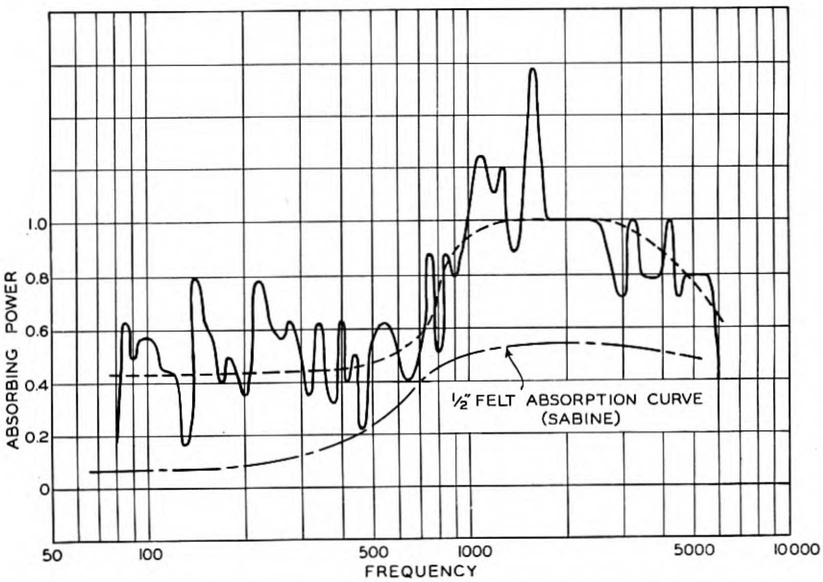


Fig. 12—Curve showing variation with frequency in the effective absorbing power of a felt lined room with respect to a region near the center and relatively near the sound source. Determined from loud speaker measurements in this room and outdoors as shown on Fig. 11.

voltage. As the condenser transmitter is rotated about the periphery of the circle, therefore, the average meter deflection is proportional to the average of the squares of the transmitter terminal voltages or the average of the squares of the pressures throughout the revolution.

Fig. 11 is a response-frequency characteristic of the piston diaphragm loud speaker measured in the same room and under the same conditions as the curve in Fig. 9 except with the rotating condenser transmitter. The center of the circle was located at the same point as the condenser transmitter for Fig. 9.

While rotating the transmitter in this manner does not average the squares of the pressures throughout a volume, it does give a measure which is closely proportional to the power density. This is apparent from a comparison of Fig. 11 with a curve on the same loud speaker measured outdoors shown on Fig. 7 and replotted on Fig. 11 for comparison. Note that these two curves very closely coincide between 1,000 and 3,000 cycles. Below 1,000 cycles and above 3,000 cycles the uniformly greater response indoors can be explained in the following manner.

As given by equation (8) above, the average energy density in a room resulting from a loud speaker emitting sound power at a constant

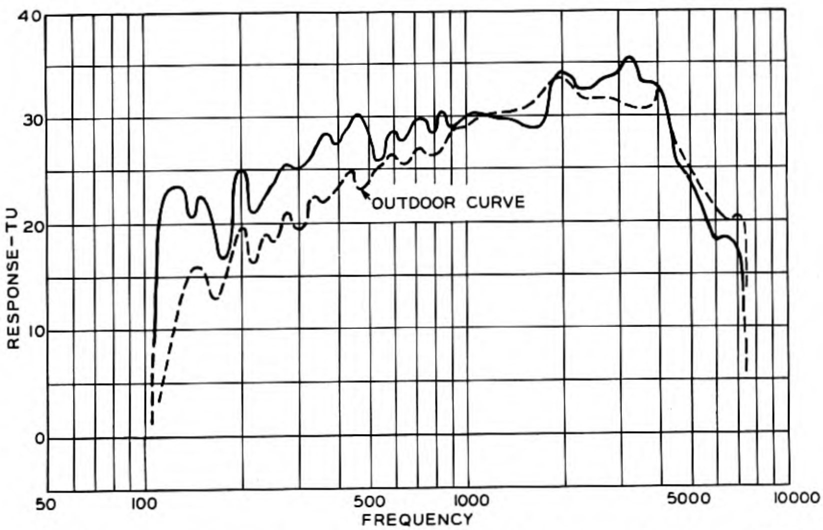


Fig. 13—Response-frequency characteristic of 115 cycle cut-off exponential horn with moving coil type receiver. Measured in highly absorbing room 12 feet from horn mouth with rotating condenser transmitter.

rate is inversely proportional to the absorbing power of the room. Inasmuch as the two curves in Fig. 11 were obtained with the same loud speaker with the condenser transmitter located at the same distance and at approximately the same relative position in the medium, the difference in the two curves would appear to be due only to the difference between the indoor and outdoor absorption. Assuming for the outdoor case a fictitious bounding surface making an enclosure of the same shape and size as the indoor room and that the energy striking this fictitious surface is completely absorbed (as would be the case) the area element of the factor "a" in equation (8) becomes the same for the indoor and outdoor tests, and the ratio of the indoor energy to the outdoor energy would therefore bear some simple relation to the average absorption coefficient of the sound absorbing material of the indoor measuring room. From the ratios of the outdoor to indoor energy densities at different frequencies as determined from Fig. 11, the solid curve on Fig. 12 is obtained. The irregular character of this curve is probably due to the fact that the rotating indoor transmitter did not give an exact measurement of the energy density at each frequency. The trend of this curve, however, is quite definite as is indicated by the dotted curve which is an average curve obtained by inspection. A comparison of this mean curve with the dot-dash curve showing the approximate absorption power of a $\frac{1}{2}''$ layer of asbestos hair felt [7] indicates an interesting correlation in the trend of the two

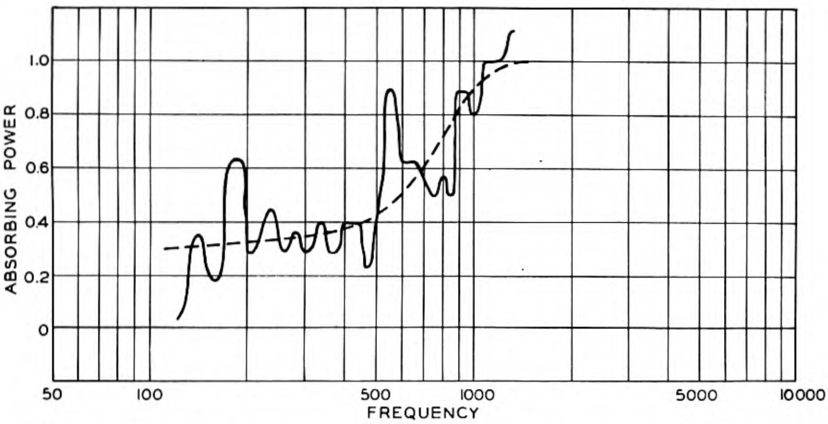


Fig. 14—Curve showing variation with frequency in the effective absorbing power of a felt lined room with respect to a region near the center and relatively near the sound source. Determined from loud speaker measurements in this room and outdoors, as shown on Fig. 13.

curves. The difference in magnitude is probably due to sound diffusion at the walls resulting in the ratio of the energy reflected to the energy direct from the source being smaller than would be the case were the measuring region located close to the absorbing surface.

The solid curve in Fig. 13 is a response-frequency characteristic of the 115 cycle cut-off exponential horn obtained with the rotating condenser transmitter under the same conditions as the curve in Fig. 11. The center of the horn mouth was placed at the same position as was the center of the diaphragm of the $3\frac{1}{2}''$ diameter piston radiator. The dotted curve in Fig. 13 is the outdoor characteristic obtained with the condenser transmitter on the horn axis at a distance of 12′. It will be noted that these indoor and outdoor curves diverge at the low

[7] See "Collected Papers on Acoustics," W. C. Sabine, p. 213, Fig. 4.

frequencies. From this divergence the absorption curve shown in Fig. 14 was obtained. Due to the relatively small angle subtended by the sound field of the horn at the higher frequencies, the indoor data where the condenser transmitter was rotated throughout a relatively large region in front of the horn, are not comparable with the outdoor data where the transmitter was located in one position on the horn axis.

Conclusion

From the above considerations it is obviously quite impossible to interpret the significance of response measurements on loud speakers in general unless such measurements are qualified by statements regarding the acoustic measuring conditions. Especially must information be given as to the position of the condenser transmitter relative to the loud speaker when the measurements were made, the method of measurement (pressure measured at one position or averaged within a region), and the size and nature of the medium. In general response measurements to be most indicative of the capabilities of the loud speaker should be made with the condenser transmitter at a distance from the loud speaker commensurate with or equivalent to the most likely listening distance of an observer.

To determine which of two loud speaker response-frequency characteristics is the better involves in addition to the above discussed acoustic considerations, an interpretation of the physiological significance of the magnitude and position in the frequency spectrum of departures in the curves from a straight horizontal line. Such an interpretation involves many physiological factors, the discussion of which is not within the scope of this paper. It should also be borne in mind that the response-frequency characteristics described in this paper are determined from steady state amplitude measurements and that they therefore give little information as to transient or phase distortion. However the cause of transient or phase distortion (the storage and release of energy in the reactive elements of the loud speaker) is also a cause of poor dynamical characteristics so that the peaks and depressions in a response-frequency characteristic may also be an indication of the phase and transient distortion. On the whole the response-frequency characteristic even though complicated by such a wide variety of factors has been found to be the most significant single criterion upon which to base a judgement of the merits of a loud speaker.

Recent Advances in Wax Recording [1]

By HALSEY A. FREDERICK

SYNOPSIS: This paper considers chiefly the frequency-response characteristics and limitations of the lateral cut "wax" record. It shows that the frequency range from 30 to 8,000 cycles can be recorded and reproduced from the record with practically negligible deviation from a flat frequency-response characteristic. The paper brings out the ease with which the record can immediately be replayed from the "wax" as an aid in assisting the artist to obtain the best results. A brief description is given of commercial processing methods including both plating and pressing. These methods give essentially a perfect copy of the original "wax." The time required for this work has been considerably reduced of late so that a test pressing can be obtained within three hours of the cutting of the original "wax."

IN the recording and reproducing of sound by the so-called "electric" method with the "wax" disc, the process may be considered as consisting of eleven steps. In order, these are: (1) studio, with its acoustic conditions, (2) microphone, (3) amplifier, (4) electromechanical recorder, (5) "wax" record, (6) copying or reproducing apparatus, (7) hard record or "pressing," (8) electric pickup, (9) amplifier, (10) loud speaker, (11) auditorium.

With this chain of apparatus the chief problem is that of making the reproduced sound in the auditorium a perfect copy of that in the studio. This is a matter of quality or fidelity of reproduction. There are other problems of cost, reliability, time required, etc., which are important but secondary to that of fidelity. While it may be necessary or convenient to introduce distortion in one of these links to compensate for such unavoidable distortion as may occur in other links, experience shows that it is desirable for the sake of simplicity, reliability and flexibility to reduce such corrective warping to a minimum and to make each step in the process as nearly perfect as possible. Perfection of a complete system may be judged by the practical method of listening to the overall result. It is necessary, however, to analyze each element of the complete system. To do this, other more analytical methods of test and standards of performance must be used. One of the most useful of these is the response-frequency curve. In order that all frequencies be reproduced equally and that the ordinary faults of resonance be avoided, this must be flat and free from sharp peaks. Good reproduction requires that frequencies from 50 to 5,000 cycles be included without discrimination. If, however,

[1] Presented before Society of Motion Picture Engineers at Lake Placid, New York, September 26, 1928.

the low frequency range be lowered to 25 or 30 cycles, a noticeable improvement will be obtained with some classes of music, whereas if the upper limit be increased to 8,000 or even 10,000 cycles, the naturalness and smoothness of practically all classes of reproduction will be noticeably improved.

A second important requirement in the judgment or analysis of any such system is that the ratio of output to input shall not vary over the range of currents or loudnesses (as well as frequencies) from the minimum up to the maximum used. If this requirement is not met, sounds or frequencies not present in the original reproduction will be introduced. This type of distortion has probably been heard by all of us in listening to an overloaded vacuum tube amplifier and is often referred to as "non-linear" distortion.

A third requirement not entirely disassociated from the first two is that any shifts in the phase relations shall be proportional to frequency.

Our judgment of the degree of perfection needed in sound reproduction systems is changing and growing more critical, so that what seemed excellent yesterday may be only fair today and tomorrow may seem intolerable. It is therefore necessary that our consideration and analysis be continually more searching and fundamental.

Of the eleven links in the chain of apparatus required for electric "wax" recording and reproduction, only five are peculiar to the "wax" method. These are the electromechanical recorder, "wax" record, the copying apparatus, the "pressing" and the pickup or reproducer. The extent to which the "wax" method is capable of the highest quality of reproduction will be disclosed by an examination of these five links. Any consideration of the practical advantages or disadvantages of the method can logically follow this examination into the quality possibilities.

The consideration which follows refers to the so-called "lateral" cut record; that is, a record in which the groove is of constant depth and oscillates or undulates laterally about a smooth spiral. This is the type used in the Western Electric Company disc record type of synchronized motion picture system. Some, but not all of the considerations and conclusions might apply to the "hill and dale" type record. It is not the purpose of this paper to consider the relative characteristics of "hill and dale" and "lateral" "wax" records.

Electromechanical Recorder

It is the task of the electromechanical recorder to take power from the amplifier and drive a mechanical recording stylus. The present-day recorder is a highly developed apparatus based on extensive experimental as well as theoretical studies. A diagrammatic view is given in Fig. 1.[2] Recorders which have been supplied by the Western Electric Company have been designed to operate over a range of frequencies from 30 to 5,500 cycles. A typical frequency characteristic

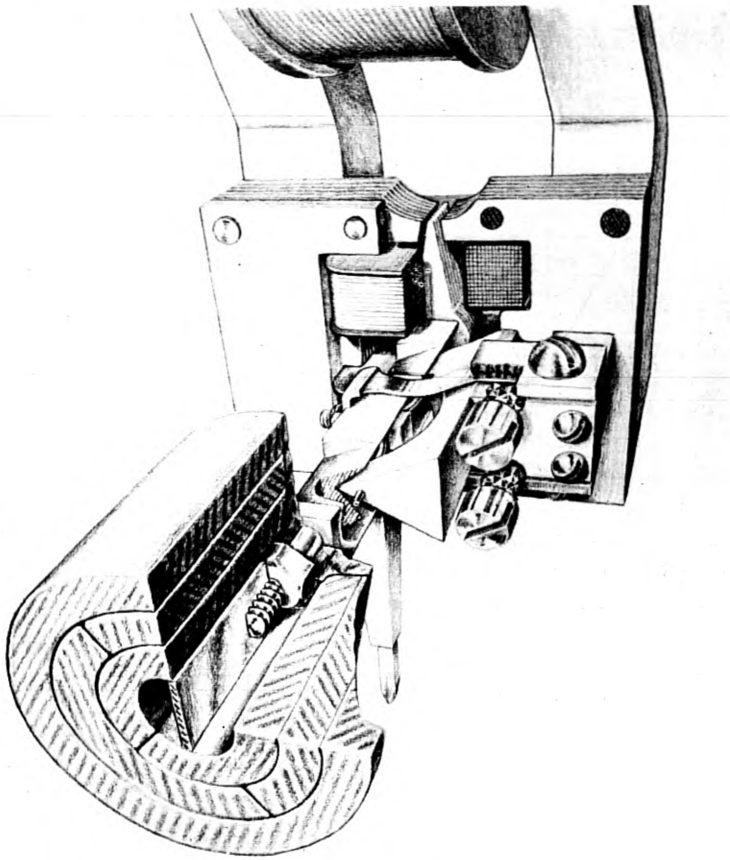


Fig. 1—Diagrammatic view of the electromechanical recorder.

is shown in Fig. 2. The device operates in linear fashion over the range of amplitudes involved in speech and music. As is seen, the response falls off below about 250 cycles. This falling characteristic is necessary in order that the maximum loudness be obtained from a record for a given spacing between grooves without cutting over

[2] "High Quality Recording and Reproducing of Music and Speech," by J. P. Maxfield and H. C. Harrison, presented at 14th Midwinter Convention of the American Institute of Electrical Engineers, New York, N. Y., Feb., 1926.

from groove to groove. In order that a lateral oscillation in a groove may represent constant intensity of sound or a constant energy over a range of frequencies, not the amplitude of the oscillation but the velocity, which is proportional to the product of the amplitude and the frequency, must be maintained constant. With the characteristic shown with these recorders, constant velocity is obtained from about

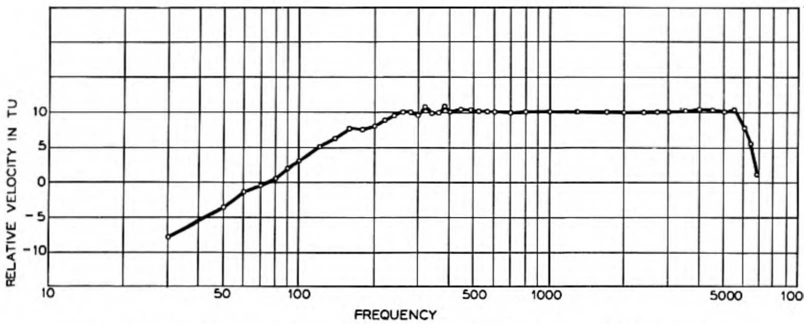


Fig. 2—Typical frequency characteristic of a commercial recorder.

250 cycles to 5,500 cycles. Below 250 cycles an approximately constant amplitude is obtained. If, therefore, sounds of constant absolute intensity are to be recorded over this range of 30 to 250 cycles, there is equal tendency for sounds of the different frequencies in this range to over-cut the record groove. It may be corrected in reproduction by a suitable electric network. Such a network will increase the subsequent amplification required but, as this additional

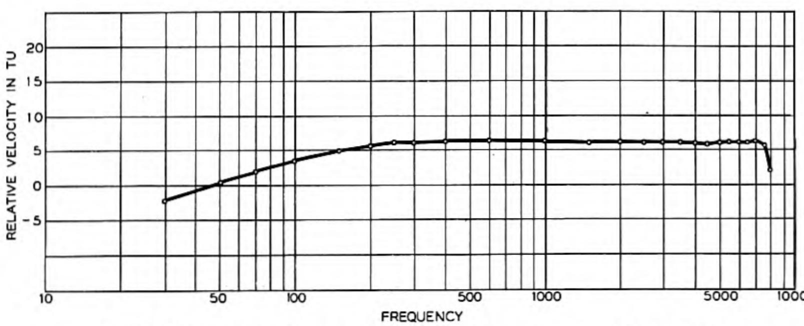


Fig. 3—Frequency characteristic of a laboratory model recorder.

amplification occurs in the first stages, it is not expensive. Practically it has not been found necessary or desirable to introduce such a corrective network since the correction has been largely cared for by the characteristics of the pickups used.

Recent development studies have established the possibility of flattening the response at the low frequency end and of raising the high frequency cut-off of the recorder. Fig. 3 shows a characteristic obtained with such a laboratory model. This shows uniform performance within ± 1 TU from 250 to 7,500 cycles and within ± 4 TU from 30 to 8,000 cycles. Although its immediate practical value might be limited by other portions of the system, this device is of great interest in that it establishes beyond question the fact that an extremely broad range of frequencies can be successfully recorded in the "wax."

The broad, flat characteristic obtained with electric recorders has been made possible by so designing their elements that they constitute correctly designed transmission systems. In such a transmission system, whether it be an electrical recorder or a long telephone line, a correct terminating impedance is required. The load imposed by the "wax" is somewhat variable but fortunately is rather small. It has been found desirable to make the other impedances in the recorder relatively large so as to dominate the system and thus minimize the effects of any changes in the impedance imposed on the stylus by the "wax." The mechanical load used as a terminating impedance and to control the device has consisted of a rod of gum rubber 25 cm. long. Torsional vibrations are transmitted along this rod. The rate of propagation is about 3,000 centimeters per second so that its length is equivalent to an ideal electrical line of about 1,500 miles. The dissipation along this rubber rod is such that a vibration is substantially dissipated by the time it has travelled down the line and back. It thus constitutes substantially a pure mechanical resistance, the magnitude of this resistance being approximately 2,500 mechanical abohms, referred to the stylus point as its point of application.

"Wax" Record [3]

In recording the usual procedure is to use a disc from 1 in. to 2 in. thick and from 13 in. to 17 in. in diameter, composed of a metallic soap with small amounts of various addition agents to improve the texture. This is shaved to a highly polished surface on a lathe. This polished disc or so-called "wax" is placed in a recording machine. In Fig. 4 is shown what is essentially a high grade lathe arranged to rotate the "wax" in a horizontal plane at a very uniform speed in a definite relation to the film with which it is synchronized. The recorder with its cutting tool or stylus is moved relative to the surface of the disc, common phonograph procedure being to record from the outer edge of the disc towards the center, whereas in the Western

[3] "Some Technical Aspects of the Vitaphone," by P. M. Rainey, presented at the meeting of the Society of Motion Picture Engineers at Norfolk, Va., April, 1927.

Electric Company theater system the direction of cutting is reversed. After a record has been cut into the "wax," the "wax" may be handled and with proper precautions readily shipped from place to place.

The shape of the groove varies somewhat in commercial practice. The groove and stylus most commonly used with Western Electric apparatus are shown in Fig. 5. The groove is approximately .006 in. wide and .0025 in. deep. The pitch of the groove is between .010 in. and .011 in. so that the space between grooves is about .004 in. Thus

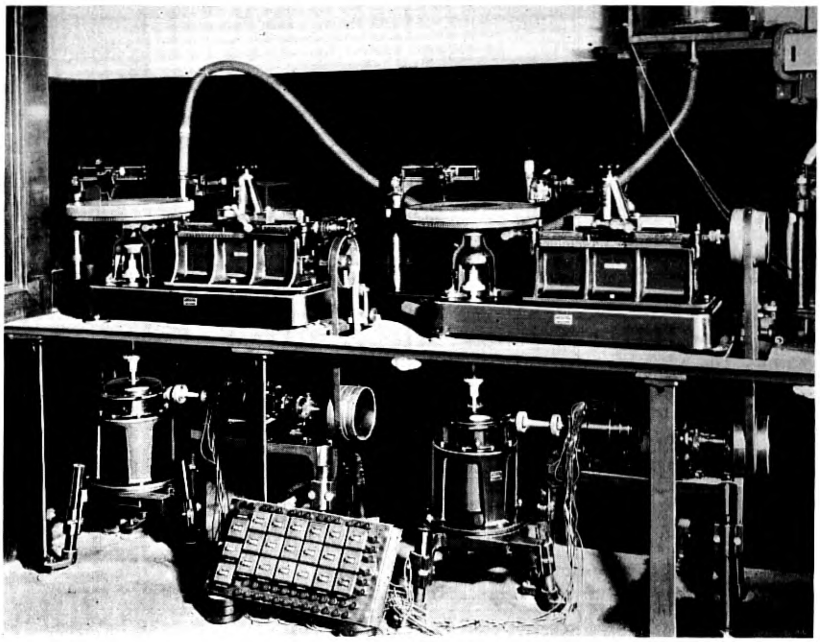


Fig. 4—The recording machine.

the maximum safe amplitude is about .002 in. If this occurs at 250 cycles the corresponding amplitude at 5,000 cycles, assuming constant absolute intensity of sound over this range, would be .0001 in.

It is important that a smooth groove be cut as any roughness in the walls introduces extraneous noise in the reproduced sound. To insure a truly smooth groove the surface of the "wax" must be shaved to a high polish. The texture of the "wax" must be fine and homogeneous which requires not only that the "wax" composition be correct, but that it be operated at the proper temperature. "Waxes" may be obtained commercially which will operate satisfactorily over the ordinary range of room temperature. The "wax" must be levelled in the recording machine with reasonable care. The stylus must be sharp and so ground that the cut will be very clean. The "wax" shaving is removed as cut by air suction. The operator is aided in maintaining the correct depth of cut by the use of a so-called "advance" ball which rides lightly on the "wax" and serves to maintain uniform depth of cut in spite of small inaccuracies of leveling of the "wax" or deviations from planeness. The "advance" ball is adjusted relative to the stylus by observing the cut with a calibrated microscope. A satisfactory operation of the recording machine requires an ordinarily skilled mechanic with reasonable experience.

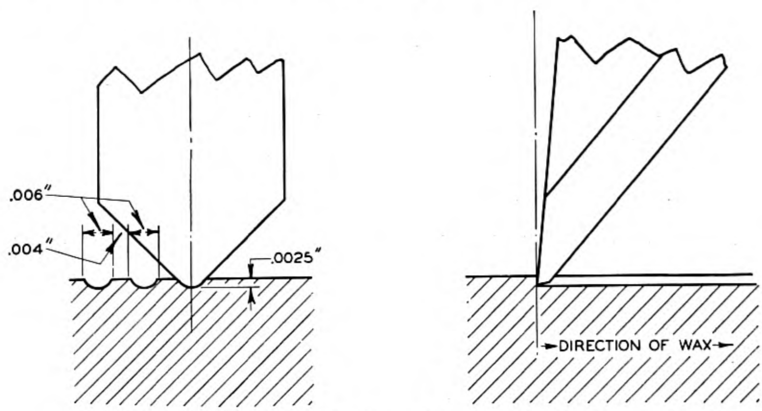
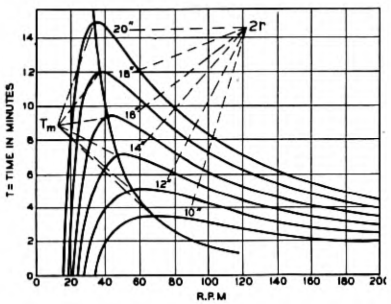


Fig. 5—Recorder stylus.

The rate of rotation is dependent upon the diameter of the record groove which is determined primarily by the length of time which it is desired to have covered by a single disc. The controlling element is the linear speed of the groove past the recorder or reproducer. In the Western Electric system the speed varies from 70 ft. to 140 ft. per minute, in other words, of the same order of magnitude as with the film record. The wave-lengths also are about the same as for a sound record on a film. At the minimum linear speed the half wave-length for a 5,000 cycle wave is .0014 in. If the minimum linear speed is fixed at 70 ft. per minute and the groove spacing is fixed, there is an optimum relation between the size of the record, the rate of rotation and the playing time. This is illustrated graphically in Fig. 6.

After a record has been cut, the sound may be reproduced directly from the "wax" by using a suitable pickup or reproducer. Ordinary reproducers or pickups rest much too heavily on the records to be used on ordinary "wax." That this would be so is obvious from the fact that the vertical pressures between the point of the needle and the record in an ordinary phonograph are of the order of 50,000 pounds per square inch. Obviously any such pressures would destroy a groove cut in soft "wax." These high pressures have been necessary in order that the groove might properly drive the needle point of the reproducer. Reduction of this pressure requires reduction of the impedance offered by the needle point to transverse vibration.



$$T_m = \frac{\pi N R^2}{24 \overline{V}_o}$$

T_m = MAX. PLAYING TIME IN MINUTES
N = GROOVES PER INCH
R = OUTSIDE GROOVE RADIUS-INCHES
$\overline{V}_o$ = MIN. LINEAR SPEED-FT. PER MINUTE
r = INSIDE GROOVE RADIUS-INCHES

Fig. 6—Relation between playing time and rate of rotation of disc for various values of $R (R = 2r)$.

The design of a suitable "wax" "playback" requires reduction of both the mass and the stiffness of the reproducing system to a minimum. In the past such "playbacks" have failed to reproduce the higher and lower frequencies with much satisfaction. The device shown in Fig. 7 represents a large advance toward ideal reproduction.

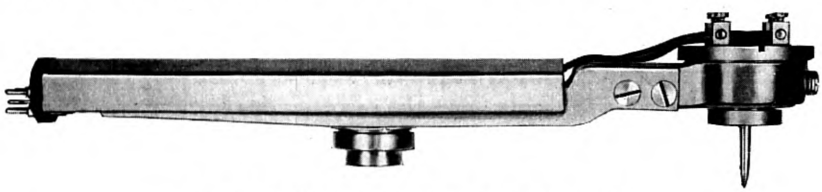


Fig. 7—Playback pickup for "wax" records.

The response of such a device when driven by a "wax" record recorded at constant velocity over the frequency range is shown in Fig. 8. This reproduction is not widely dissimilar from that obtained from finished records with the best electric pickups now commercially available and is sufficiently good to serve as a very valuable criterion in judging the quality of the record. The record may be played a number of times without great injury. The extent of the injury is indicated by the frequency characteristics obtained on successive playings shown in Fig. 9. They show little change in the low frequency response and a loss of about 2 TU per playing at high frequencies. The practical value in studio work of being able to let an artist immediately hear and criticize the results of his own efforts can hardly be overestimated.

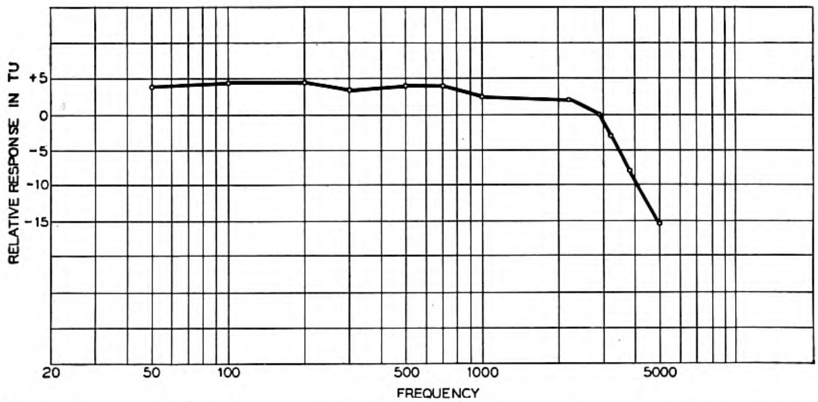


Fig. 8—Response of a "wax playback" driven by constant velocity wax record.

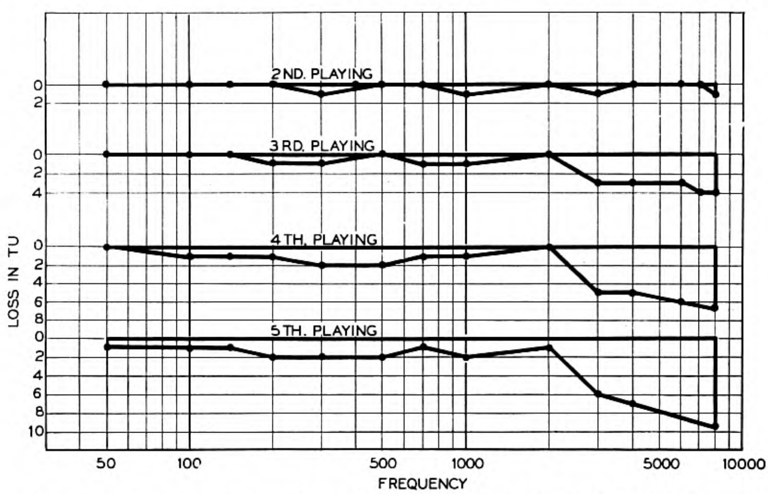


Fig. 9 Loss in response on successive playings of a wax playback driven by constant velocity wax record.

Copying (Processing) of the "Wax" Record

After a groove has been satisfactorily cut into the "wax" record, the usual procedure is to render the surface of the "wax" conducting by brushing into it an extremely find conducting powder. It is then electroplated. The technique in this step varies somewhat with the different companies doing such work, although not in any fundamental manner. The negative electroplate thus made may be used to hot-press a molding compound such as shellac containing a finely ground filler. This first electroplate is called a "master." From it two test pressings are usually made. If satisfactory the matter is then electroplated with a positive, being first treated so that this positive plate may be easily removed. This positive is sometimes called an "original." From this in turn is plated a metal mold or "stamper." From these, duplicate "originals" may be plated and from them, duplicate "molds" or "stampers." These processes involve no measurable injury to the quality of the record and are comparatively simple and extremely safe in practice. By this practice of making a number of duplicates it is possible to safeguard the "master" and insure against any accident which might destroy a valuable record. From a single "stamper" it is not unusual to make a thousand finished pressings. The time required for these operations is such that test pressings are commonly obtained from the "wax" in 12 hours. Recent refinements in the art have reduced the time required so that finished records may, if necessary, be obtained in 3 hours after delivery of the "wax."

Hard Record or "Pressing"

Various materials have been used in making the hard record or "pressing." In some cases the material has been made homogeneous and in others the surface is of a different material from that used in the body of the record. Some have used a laminated structure. There has not, however, been much latitude allowed the experimenter concerned with materials for the hard record. The material has had to be quite hard and, in order to show a reasonable life, it has had to contain sufficient abrasive to grind the needle quickly to a good fit. At the beginning of the run of a new needle due to the small bearing surface, the pressures are very high. They rapidly decrease so that with an ordinary loud steel needle after one minute's wear in the ordinary phonograph, the bearing area is increased to such an extent that the pressure is only about 50,000 pounds per square inch. As the needle continues to wear to a larger bearing surface, the pressure obviously continues to decrease. These high pressures and necessary abrasive characteristics of the record have introduced irregularities which are responsible for most of the extraneous noise commonly known as "surface" or "needle scratch."

The "pressing" copies the "wax" record with a very high degree of accuracy so that if our attention be confined to frequency characteristics alone, the "pressing" shows almost complete perfection.

Moreover, it is cheap and durable, and reproduction of the sounds from this record calls for no fine adjustments or intricate apparatus as has been long evidenced by the broad use of the ordinary phonograph.

The major part of the extraneous or "surface" noise found with this method of reproduction comes from the material of the finished record. Recent progress has been made in reducing this noise. As a result of this, together with refinement in the plating processes, records used with Western Electric Company theater equipment during the last two years have shown a reduction of 3 to 6 TU in "surface" noise. This corresponds to eliminating 50 per cent to 75 per cent of that previously present. It is not necessary to reduce the level of "surface" noise to the zero point but merely to the threshold of audibility under the conditions of minimum auditorium noise which are of interest. This noise masks the surface. Moreover, it is not the absolute amplitude of the imperfection giving rise to "surface" noise but the relative magnitude in comparison with the useful sound amplitudes which counts. Thus, an effective reduction in "surface" could be made if we were willing to use larger records or if we were willing to reduce the playing time of the present records by increasing the spacing of the grooves and the amplitude at which the grooves are cut. Any large reduction in "surface" noise made by a reduction in the irregularities in the record material would open the door to increasing the playing time of a record of given size. There is no known absolute or fundamental reason why further improvements in record materials may not be expected to reduce further the amount of "surface" noise. Moreover, large advances in pickup design open distinctly new possibilities as to reductions in "surface."

It has sometimes been thought that in order to reproduce high frequencies properly, the linear record speed would have to be increased or the size of the needle point reduced. At present the diameter of the bearing portion of a representative needle is about .003 in. whereas, as mentioned before, the half wave-length for a 5,000 cycle wave is .0014 in. The factor determining whether a needle will follow the undulation of the groove is not any consideration of the relative diameter of the needle point and the undulation of the groove but rather the radius of curvature of the needle and the bend of the groove. As indicated before, the amplitude at 5,000 cycles would be only about .0001 in. if sounds of that frequency were as intense as those of lower frequencies (.002 in. at 250 cycles). As a matter of fact, sounds of 5,000 cycles or more in speech or music are characterized by lower intensity than those of lower frequency. If, however, we assume an amplitude of .0001 in. at 5,000 cycles and assume a linear record speed of 70 ft. per minute, then the minimum radius of curvature of the undulation of the groove is .00193 in.[4] With the foregoing assumption, the radius of curvature of the undulation of the groove becomes equal to that of the needle point at about 7,000 cycles. Taking into account the lower intensities of sounds encountered at these high frequencies, it is obvious that present commercial needle points are quite capable of following the high frequency undulations of the groove up to frequencies of at least 10,000 cycles. The limitations of high frequency reproduction commonly found in the past are associated with limitations in the design of the pickup or reproducer and relate either to inability of the record groove to drive the needle point, with resultant chatter, or inability of the pickup structure to transmit high frequency motions from the needle point to the armature.

Electric Pickup

Large advances have been made within the last two or three years in designing electric reproducing structures. The mechanical im-

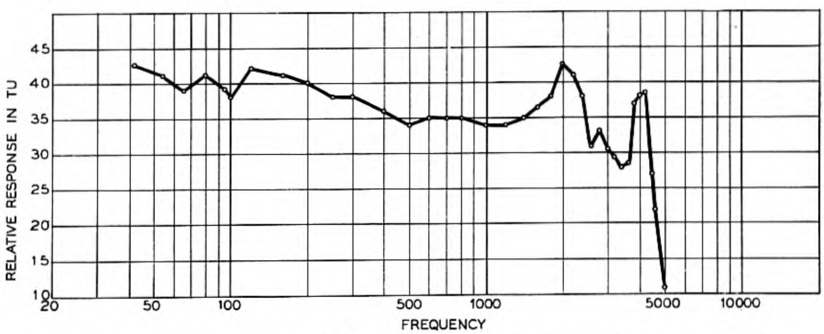


Fig. 10—Response of a 2-a pickup driven by constant velocity pressings.

pedance at the needle point has been reduced so that the needle point truthfully follows the undulations in the groove without necessitating excessive and somewhat destructive bearing pressures. At the

[4] The minimum radius of curvature is computed by the formula

$$R_c = \frac{V^2}{100\pi^2 A f^2},$$

where:

R_c = minimum radius of curvature in inches.

V = linear speed in feet per minute.

A = amplitude of vibration in inches.

f = frequency in c.p.s.

same time the transmitting structure has been so designed that a very broad range of frequencies is properly transmitted from the needle point to the armature. Moreover, proper mechanical loads have been provided so that the motions after transmission are absorbed and hence not reflected back. This is another way of stating the fact

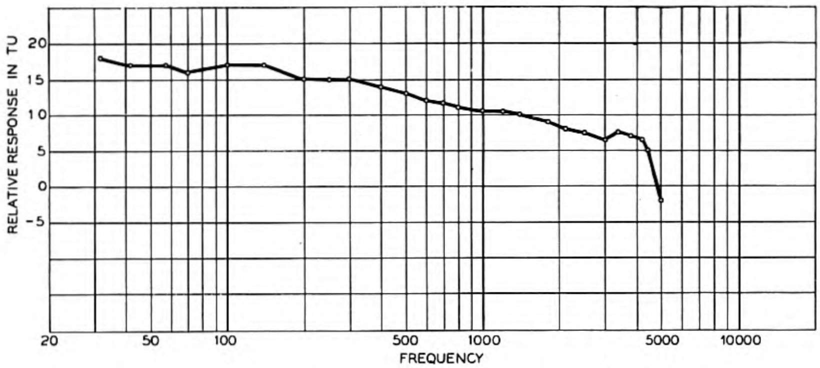


Fig. 11—Response of a 4-a pickup driven by constant velocity pressings.

that resonance as ordinarily considered has been eliminated from these structures.

The curves shown in Figs. 10 to 12 illustrate the steps which have been taken. The pickup shown in Fig. 11 is free from the resonances shown in that of Fig. 10. The resonances present in the earlier

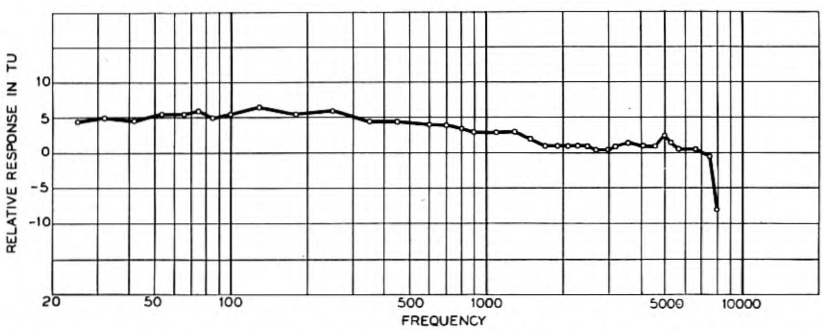


Fig. 12—Response of experimental pickup driven by constant velocity pressings.

pickup involved high needle point impedances in the region of these resonances. These high impedances involved large driving forces destructive both to needle and record. Certain records were injured after only a few playings with this reproducer. The later reproducer is characterized not only by considerably reduced *average* needle point impedance, but, as shown by the curves, the resonance is practically eliminated and hence there is an even greater reduction from the maximum impedance which occurred in the earlier reproducer at resonance. Both needles and records have a relatively long life with the later type pickup, which has been in commercial use for some months. As is seen, the higher frequencies are reproduced in considerably better fashion. A third curve is given in Fig. 12. This was obtained with a more recent experimental model in which a further large reduction in the needle point impedance had been effected and in which, in addition, a very much more rigid, though a lighter structure served to connect the needle point with the armature. This model shows further reduction in wear and tear on the record and very greatly improved reproduction at the high frequency end of the scale.

The application of the processes of sound recording on "wax" to the synchronized film has involved meeting a number of conditions not previously encountered in the phonograph field. One of the most important of these relates to editing, cutting and rearranging of the picture. Various methods have long been used to copy or "dub" a disc record. The prime requirement is that there be no sacrifice in quality. To attain this end records have sometimes been copied at very low speed. This method appears unnecessarily laborious and slow and the results obtained are not altogether satisfactory in the light of possibilities presented by pickups and recorders of the characteristics shown above. Rearrangement of material on records is entirely practicable, portions may be deleted or new portions added either as a whole or the new sounds added to those already on a record—in fact any changes of this type may be made which can be made in the picture.

The detailed technique of "dubbing" appears to offer no serious technical difficulties. The refinement reached and the extent of its future use may be expected to be governed by the demand in the synchronized motion picture field.

Sound Recording with the Light Valve [1]

By DONALD MACKENZIE

SYNOPSIS: The light valve developed by Bell Telephone Laboratories is an electromagnetic shutter consisting of a loop of duralumin tape formed into a slit at right angles to a magnetic field. Sound currents from the microphone and amplifier flow in this loop causing it to open and close in accordance with the current variations.

The slit is focussed by a lens on the sound negative film. An incandescent ribbon filament is focussed on the light valve, and the light passed by the undisturbed slit appears on the film as a line at right angles to the direction of the film travel. As the valve aperture is modulated by sound currents, the film receives a varying exposure and a sound record of the variable density type is obtained.

For talking pictures such a sound film is made on a separate recording machine synchronized with the camera and is printed alongside the picture on the finished positive. The prints are displaced so that the sound is advanced over the corresponding picture. This is in order that the sound may be projected at a point of continuous film motion below the picture gate.

THE sound records I am about to describe are of the variable density type, and the method of making them is that developed by Bell Telephone Laboratories.

It is not difficult to specify the requirements of this type of sound film. So far as possible the exposure of the negative must be kept within the straight line portion of the Hurter and Driffield curve for the emulsion chosen, and the print must be timed with the same restriction. The development of the negative and of the print must result in a positive where the transmission of each element of length is proportional to the exposure of the corresponding element of the negative. The light modulator must be supplied with undistorted power from the recording microphone and amplifier. When the positive is projected, the striations of the sound track must be enabled to modulate the illumination of a photo-sensitive cell to retranslate the photographic effect into electrical current which shall be a fair copy of the microphone current generated by the original sound. From this point on the problem is the familiar one of sound re-enforcement, the film and cell having taken the places of the sound source and microphone.

Fig. 1 shows a photograph of the light valve, invented in 1922 by E. C. Wente of the Bell Telephone Laboratories. Essentially, it consists of a loop of duralumin tape suspended in a plane at right

[1] Presented before Society of Motion Picture Engineers at Lake Placid, New York, September 25, 1928.

angles to a magnetic field. The tape, 6 mils wide and 0.3 mil thick, is secured to windlasses A and A' and stretched tight by the spring held pulley B. At points C and C' insulated pincers confine the central portions of the tape between windlasses and pulley to form a slit 2 mils wide. Supporting this loop and adjusting devices is a slab of metal with central elevation D, which constitutes the armature of an electromagnet. The central portions of the loop are supported on insulating bridges to lie 3 mils above the face of D; here the sides of the loop are centered over a tapered slot, 8 mils wide by 256 mils long in this plane, opening to 204 mils by 256 mils at the outside face of the armature. Viewed against the light, the valve appears as a slit 2 mils by 256 mils.

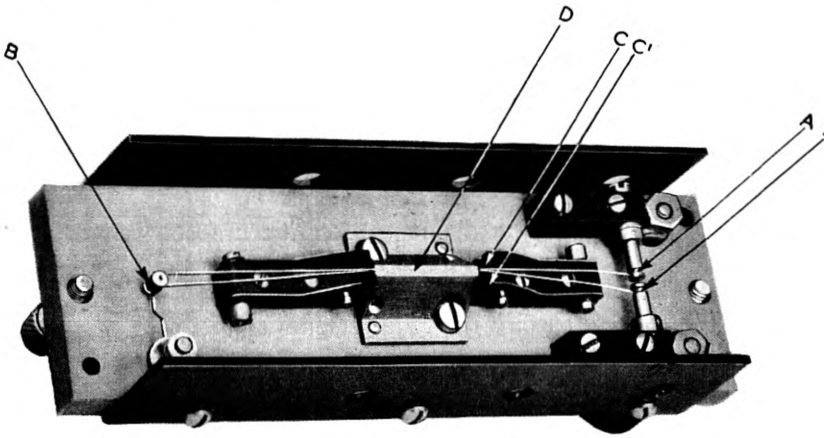


Fig. 1—The light valve.

The electromagnet core has a similar elevation opposing D across an air gap of 8 mils which closes to 7 mils when the magnet is energized from a 12 volt battery. A tapered slot in the magnet core begins 8 mils wide by 256 mils long and opens with the same taper as the slot in the armature. When the assembly of magnet and armature is complete, the valve constitutes a slit 2 mils by 256 mils, its sides lying in a plane at right angles to the lines of force and approximately centered in the air gap. The windlasses A and A', one of which is grounded, are connected to the output terminals of the recording amplifier. If the magnet is energized and the amplifier supplies a sine wave current from an oscillator, the duralumin loop opens and closes in accordance with the current alternations.

When one side of the wave opens the valve to 4 mils and the other side closes it completely, full modulation of the aperture is accomplished. The natural frequency of the valve is set by adjusting the tension applied by the pulley *B*; for reasons which involve many considerations the valve is tuned to 7,000 cycles per second. Under these circumstances about 10 milliwatts of A.C. power are required for full modulation at a frequency remote from resonance; about one one-hundredth of this power at the resonant frequency. The impedance of the valve with protecting fuse is about 12 ohms.

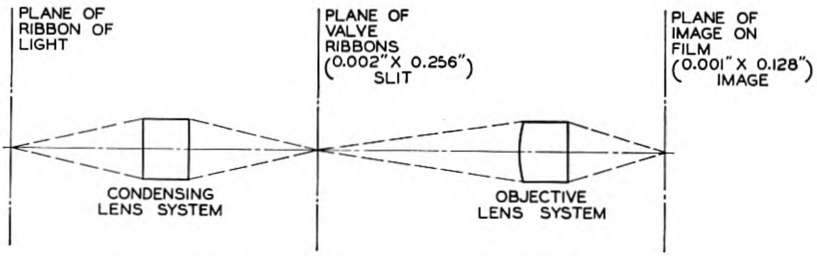


Fig. 2—Diagram of the optical system for studio recording.

If this appliance is interposed between a light source and a photographic film we have a camera shutter of unconventional design. Fig. 2 shows a diagram of the optical system for studio recording. At the left is a light source, a ribbon filament 18 ampere projection lamp, which is focussed on the plane of the valve. The light passed by the valve is then focussed with a 2 to 1 reduction on the photographic film at the right. A simple achromat is used to form the image of the filament at the valve plane, but a more complicated lens, designed to exacting specifications by Bausch and Lomb, is required for focussing the valve on the film. The undisturbed valve opening appears on the film as a line 1 mil by 128 mils, its length at right angles to the direction of film travel. The width of this line varies with the sound currents supplied to the valve, so that the film receives a varying exposure: light of fixed specific intensity through a varying slit.

Fig. 3 shows a studio recording machine with the door of the exposure chamber open. In this machine the film travels at 90 feet per minute, and the sound track is made at the edge away from the observer. The line of light, the image of the valve, overruns the perforations by 6 mils, extending toward the center of the film 122 mils inside the perforation line. The right-hand sprocket serves to draw film from the feed magazine above and to feed it to the take-up magazine below; this sprocket is driven from the motor shaft through a worm and worm-wheel. The left-hand sprocket engages 20 perforations and is driven through a mechanical filter from a worm and worm-wheel similar to that driving the feed sprocket. The mechanical filter enforces uniform angular velocity of the left-hand sprocket which carries the film past the line of exposure: the focussed image of the valve; balancing of the flywheel which forms part of this mechanical filter holds the angular velocity constant to one-tenth of one per cent, despite the imperfections of the driving gears.

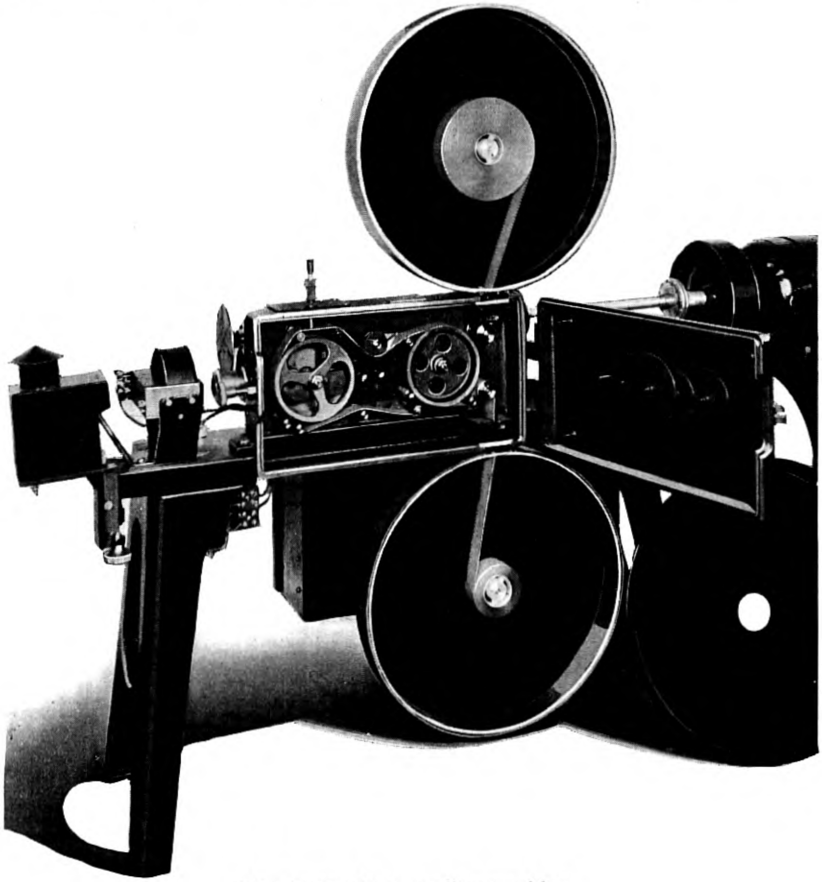


Fig. 3—Studio recording machine.

So far we have provided a means for driving the film and a means for modulating the light thereon, but we have not chosen the average illumination about which the modulation is to take place. The maximum exposure corresponds to the maximum opening of the valve and is therefore double the average.

Choose now the contrast to which the negative sound record is to be developed and draw the Hurter and Driffield curve for this contrast for the emulsion chosen for the negative sound record. The maximum exposure should correspond to the beginning of over-exposure, the average should be half this. The Hurter and Driffield curve will give the density of the over-exposure point for the chosen contrast and the density for half this exposure. Let the machine be run to expose film to light through the unmodulated valve for several values of the lamp current. Develop the film and measure the densities due to the various values of lamp current. Select, by interpolation if necessary, the lamp current which corresponds to half over-exposure. With this current in the lamp the machine is ready to make a sound record, since the focussing of the valve has already been done and manufacturing specifications insure that the line of illumination shall lie, within 3 minutes of arc, at right angles to the direction of film travel.

Consider at this point the procedure in the recording studio. Adding sound to the picture introduces no complication of technique other than to require sufficient rehearsing to make sure of satisfactory pick-up of the sound: microphone placement must be established and amplifiers adjusted to feed the light valve currents which just drive it to the edge of overload in the fortissimo passages of music or the loudest utterances of speakers.

In Fig. 3 the photograph shows a photoelectric cell mounted inside the left-hand sprocket, which carries the film past the line of exposure. Fresh film transmits some 4 per cent of the light falling on it, and modulation of this light during the record is appreciated by the cell inside the sprocket. This cell is connected to a preliminary amplifier mounted below the exposure chamber, and with suitable further amplification the operator may hear from the loud speaker the record as it is actually being shot on the film. Full modulation of the valve implies complete closing of the slit by one side of the wave of current; this modulation should not be exceeded or photographic overload will abound.

One or more cameras and one or more sound recording machines are driven by motors electrically synchronized from a common distributor. Speed control and synchronization of these motors are described in Mr. Stoller's paper. At the beginning of the day's work a check is made of the operation of the driving motors, and the tuning-and-spacing of the valves is verified.

Fig. 4 is a schematic diagram of the studio equipment for sound recording. Provision is made for combining if desired the contributions of several microphones on the set. This combination is under the control of the mixer operator in the monitoring room, viewing the set through a double window in the studio wall. The mixer controls also the gain of the amplifiers for the recording machines.

The diagram shows relays which permit the mixer to connect the horn circuit either directly to the recording amplifier or to one or the other of the monitoring photoelectric cells in the film recorders. The direct connection is used in preparing the sound pick-up in the studio: the program is rehearsed until satisfactory arrangement of microphones and of amplifier gain is effected. The electrical charac-

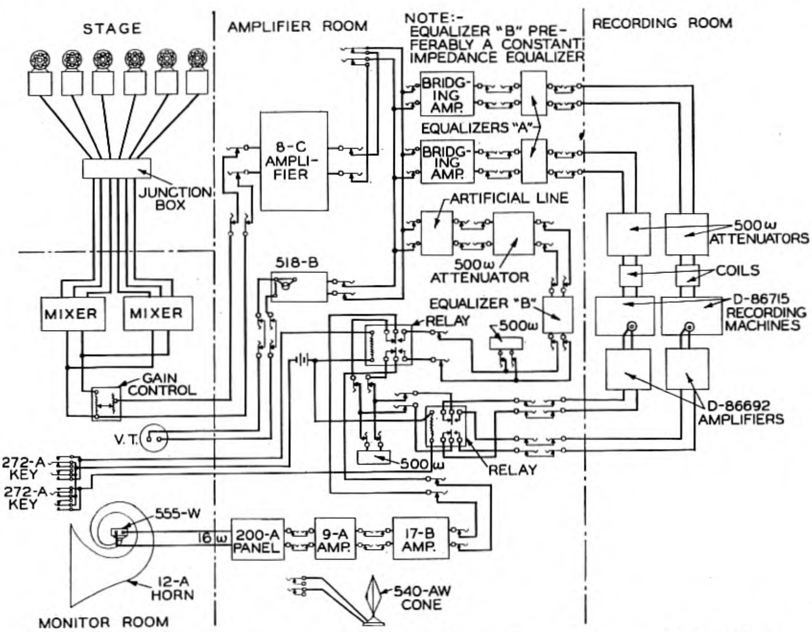


Fig. 4—Schematic diagram of the studio equipment for sound recording.

teristic of this direct monitoring circuit is so designed that the sound quality heard in the horns shall be the same as the quality to be expected in the reproduction of the positive print in the theater. Acoustic treatment of the walls of the monitoring room secures the reverberation characteristic of the theater, and the monitoring level is so adjusted that the mixer operator hears the same loudness that he would wish to hear from the theater horns. It is capitally important that the operator judge his pick-up on the basis of sound closely identical in loudness and quality with that to be heard later in theater reproduction.

After the pick-up has been established on the direct monitoring circuit, the output of the recording amplifier is applied to the light valves and the monitoring horns are connected to the photo cell amplifiers on the recording machines. With no film in the machine and at a convenient lamp current a complete rehearsal is made to verify the operation of the valves at the proper level. Film is then loaded, cameras and sound recorders are interlocked and starting marks made on all films by punches or light flashes.

A light signal from the recording room warns the studio, which after lighting up signals back its readiness to start. The machine operator starts the cameras and sound recorders, brings up the lamp current to the proper value, and when the machines are up to speed signals the studio to start. During the recording, the mixer operator monitors the record through the light valves, thereby assuring himself that no record is lost.

In the choice of emulsion for the sound negative, the usual designation of speed may be disregarded, because it is desired to make the exposure of the unmodulated track many times the under-exposure of the emulsion used. The advantages of positive emulsion for the sound negative have come to be generally recognized; positive has been used by Bell Telephone Laboratories since 1924. The scale of Eastman positive film is about 20 to 1; we adjust the recording lamp current to give an illumination on the film for the unmodulated track of 10 times the under-exposure. After one lamp has been calibrated as described before it may be replaced when necessary by another in which the wattage in the ribbon filament is the same; the light emission is very closely correlated with the wattage. Where the unmodulated or average exposure is ten times the under-exposure minimum, 90 per cent modulation of the light can be permitted without running into under-exposure on the faint side of the wave. For sound currents reaching 100 per cent modulation of the light, 90 per cent of the wave is free from distortion; if the average light were halved, still 80 per cent would be free from distortion. There is therefore considerable latitude in the average exposure, and the negative is satisfactory if the transmission of the unmodulated track lies between fairly wide limits.

The choice of the negative sound gamma is determined by the practice of the laboratory in regard to picture development. It is usual to see on the screen pictures whose overall gamma considerably exceeds unity. On the sound track the overall gamma should equal unity, and the development of sound negatives should be uniform, though that of picture negatives is left to the judgment of the finisher.

Theoretically, it should be immaterial what combination of reciprocal values is chosen for the negative and positive sound gammas. Practically, we have to recognize the existence of ground noise in all records and take precaution to minimize it. No matter how excellently we reproduce the fortissimo passages, our record is unsatisfactory unless the ground noise is low enough for a wide volume range, that is, a wide range in level between fortissimo and pianissimo. Whether our negative sound record is made on negative or positive emulsion, there is always the danger that in reproduction we shall encounter variations in transmission from point to point due to local variations in the celluloid base, to local action of the developing agent, or to a developer excessively granular in action. The photoelectric cell is able to recognize variations of 1/10 of 1 per cent, whereas the eye ignores contrasts under 2 per cent. These local variations in transmission, continued to the positive print, constitute the ground noise.

The remedy is, in part, to choose a developer as little granular in its effect as possible. In part, to insist on machine development of the sound film with thoroughly agitated developer. Further, to carry the sound development to a high gamma; this obviates to a large extent flow marks of the developer, and goes a long way to escape local variations in the base by developing the negative striations to be conspicuous in comparison.

In 1924 we concluded that the optimum choice was positive emulsion developed to unit gamma for both sound negative and sound print. This is feasible for sound records separate from pictures, but a compromise must be made for the combination of sound and picture in a single positive print. Here the positive development required for a satisfactory picture is always to a gamma far above unity.

It is customary to develop picture negatives by inspection, having in mind the uniform positive development to be undergone by the prints from these negatives. The gamma of these positives need never exceed 1.8; the sound negative then should be developed to 0.55. In order not to disturb the practice of the film laboratory, we ask that the positive development be standardized and its gamma ascertained, the reciprocal of this gamma then arranged for in the standardized negative development. A negative gamma above 0.5, together with the precautions of careful handling, permits the realization of an adequate volume range.

It is beyond the scope of this paper to discuss the details of manipulation and of choice of developer, but I wish to acknowledge the cooperation of Mr. J. W. Coffman in the solution of such problems. The problem is the reduction of ground noise, and its seriousness is not to be diminished by choosing a different recording method.

In printing the sound negative, a uniform density for the print of the unmodulated track is desired. The volume of reproduced sound for a given reproducing light source, varies directly with the average transmission and the per cent modulation of this average. This average density should be on the straight line portion of the positive Hurter and Driffield curve, far enough to keep the denser negative portions from reaching the under-exposure region. For Eastman positive film a suitable transmission of the unmodulated portion of the sound print is 35 per cent, referred to air, for the usual values of positive gamma: 1.4 to 1.8. At this average transmission only the peaks of the recorded sound will encroach on the region of under-exposure. For the reciprocally developed negative track the region of under-exposure will have been reached by occasional peaks on the other side of the wave, and such photographic distortion as exists will be balanced between positive and negative.

Here we appropriately consider the photographic distortion as it occurs in variable density records. If the entire negative exposure has been confined to the under-exposure region of the emulsion chosen, a huskiness will result in the reproduction which can not be corrected by any known technique. But if the unmodulated negative transmission, for a gamma of 0.55, is about 16 per cent referred to air, 90 per cent of the wave will be clear of under-exposure, and experience shows that the ear detects no distortion. In telephonic terms, everything at a level 1 TU below full modulation will be free from distortion, and the peaks will be substantially perfect. The same may be said of the positive printed to an average transmission of 35 per cent, provided the overall gamma approximates unity.

It has been calculated that if the overall gamma departs from unity by 0.2 in either direction, a harmonic of 5 per cent amplitude of the fundamental will be introduced. Experimentation has shown that a 5 per cent harmonic is the least detectible. We state then the tolerance on the overall gamma for the sound track as 0.8 to 1.2. Variation of corresponding amount in the contrast of a picture print is intolerable; therefore greater latitude in contrast is permissible in the sound record than could be tolerated in the accompanying picture.

In printing these sound negatives in combination with pictures for projection in the theater, it is customary at the present time to print one negative, masking the space needed for the other, then run the positive again through the printer with the other negative, masking now the space already printed. In printing the picture negative, light changes are made as usual; for the sound negative the light is regulated to result in 35 per cent transmission of the unmodulated track after positive development. Provision of suitable masks in the camera has been made to show in the finder and expose on the film only the portion which will be available for picture projection.

In the theater projector, the sound gate is located 14.5 inches below the picture gate, in order to project the sound record at a point where the film is in continuous motion. Therefore in the printing it is arranged to print the sound negative displaced along the length of the positive enough to bring the sound 14.5 inches ahead

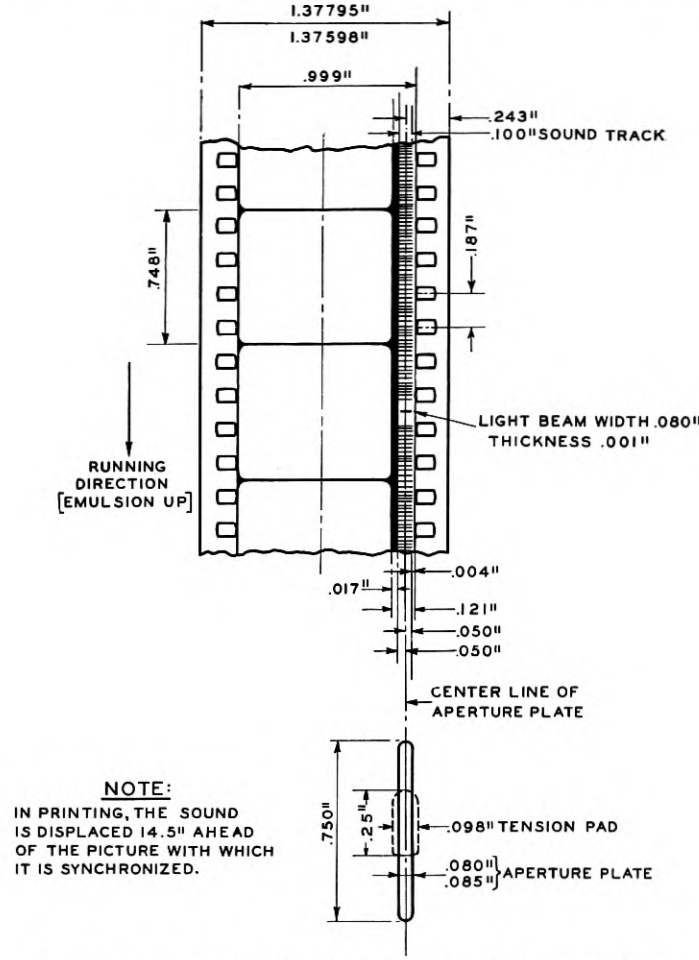


Fig. 5—Picture and sound track dimensions of synchronized sound film for standard 35 mm. positive stock.

of the corresponding frame. The printer apertures are chosen to give a dark no man's land 17 mils wide between picture and sound track; the latter at the outside is separated 4 mils from the inner perforation edge.

Fig. 5 exhibits the present practice for the finished positive. It will be seen that the sound track covers 100 mils clear, and is illuminated in the projector by a line of light 80 mils long, 1 mil wide, centered on the striations. This gives a margin of 10 mils at each end of the reproducing line, an allowance for lateral shifting of the film on the sprocket teeth.

In conclusion, let me estimate the quality of the sound record to be expected. Assume that the recording lamp current has been set to within 5 per cent of the theoretical optimum, the overall gamma held between 0.8 and 1.2, and the final average positive transmission is between 32 per cent and 38 per cent. Then the distortion of wave form due to photographic handling is so small that the ear can not distinguish the record from a theoretically perfect one. The frequency-amplitude characteristic of the reproduced sound remains to be stated.

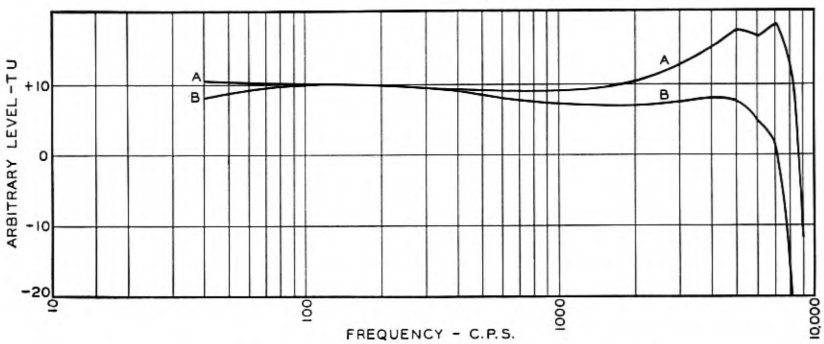


Fig. 6—Characteristic curves of sound recording.

Due to the fact that the element of illumination, both in recording and in reproducing, is 1 mil wide instead of infinitely narrow, the final print will not reproduce the higher frequencies as efficiently as the lower. For example, at the standard speed of 90 feet per minute, the line of illumination covers on the film an entire cycle length of the frequency of 18,000 cycles. This frequency is therefore extinguished completely. The drooping characteristic resulting from this effect, called the film transfer loss, may be largely offset by judicious choice of electrical characteristics and by taking advantage of the mechanical tuning of the light valve.

In Fig. 6 is shown in curve A the light modulation by the valve in recording for constant sound pressure of various frequencies at the transmitter; in curve B the overall characteristic of the reproduction in terms of electrical power delivered to the loud speaker for constant sound pressure at the transmitter in the studio. Experience shows that curve B is close enough to flat; the success of the record, as of the picture, depends on the director.

Synchronization and Speed Control of Synchronized Sound Pictures [1]

By H. M. STOLLER

SYNOPSIS: The reproduction of the synchronized sound picture of today presents no serious problem of synchronization, for this factor has been practically eliminated by the perfection of electrical means for reproducing sound with equipment which may be coupled mechanically to the picture projector.

The important problem of the present day, in connection with the reproduction of synchronized sound pictures, is the provision of suitable means for maintaining a constant speed of the sound reproducing mechanism in order that the pitch of the sound being reproduced may not suffer any sudden change which would be sensed by a good musical ear. Control circuits using vacuum tubes with a frequency bridge as a speed standard with provision for manual variable speed control are described and explained for use with both A.C. and D.C. motors. Remote synchronization permitting the recording of pictures and sound simultaneously on equipment located some distance apart is obtained by a modification of the Michalke electric gear system.

WHEN Thomas A. Edison gave a demonstration of his talking motion pictures nearly sixteen years ago one of his chief problems was proper synchronization between his acoustic phonograph and the motion picture projector. It was then necessary to locate the phonograph behind the screen in order to make the sound appear to come from the picture. A system of belts and pulleys running from one end of the theater to the other was used to secure synchronization with the projector in the booth.

The development of the electrical reproducer has made it possible to locate the turntable and reproducing mechanism in the projection booth permitting a direct mechanical coupling between it and the projector. The horns are located behind the screen and electrically connected by wires with the electrical reproducer.

Thus there is no problem of synchronization in reproducing except to set the needle on the disc at the proper point before starting. However, such mechanical coupling between the projector and sound recorder (either of the disc or film type) makes it necessary to provide very close speed regulation on the projector motor, since variations in speed produce proportional changes in the pitch of the sound.

This paper will describe the speed regulating system employed in reproducing and the synchronization system used in recording.

[1] Presented before Society of Motion Picture Engineers at Lake Placid, New York, September 24, 1928.

Speed Regulation Requirements

A good musical ear while having a sense of absolute pitch of only about 3 per cent is extremely sensitive to sudden changes in pitch. It has been found that a sudden change in pitch as small as one half of 1 per cent may be noticed if made abruptly. In order to properly take care of this requirement, therefore, the speed regulation or change in speed of the motor drive over normal variations in line voltage and load should be held within 2/10 of 1 per cent.

The absolute speed must also be held near these limits since at the end of a film it is necessary to switch from one projector to another with minimum change in the pitch of the sound reproduction.

Voltage, Frequency and Load Variations

A study of the voltage variations in power supply systems indicated a range from 100 to 125 volts. At a particular location the normal variation of voltage was found to be 5 per cent above or below the mean value with occasional momentary variations of as much as 10 per cent above and below mean value.

An investigation of variations in frequency of the supply voltage showed that in the large cities the frequency was held very accurately at 60 cycles. In New York City for example the frequency stays within one quarter of 1 cycle and does not change rapidly. However, in some small power systems the frequency varied as much as 5 cycles and in some cases was subject to rapid changes in frequency.

The load of the motor is due mainly to mechanical friction in the projector and take-up mechanism. This load was found to be on the average 1/10 of a horse power but subject to wide variations. In the case of a new machine with a stiff adjustment of the take-up mechanism, the load was found to be as high as one-fifth of a horse power.

Speed Control Circuit

A consideration of the variables just discussed imposes rather severe requirements of speed control, the two extremes being (1) the combination of low line voltage, low frequency and heavy load, (2) the combination of high line voltage, high frequency and light load. Ordinarily it might be possible to compromise and not provide for such an extremely unfavorable combination of requirements. However, it must be borne in mind that in the case of a musical program the failure of the speed regulating system for even as short a time as a fraction of a second would be a very serious matter causing the music to sound off pitch similar to a phonograph which has run down while in operation.

An examination of the standard commercial types of speed control indicated that there was nothing exactly suitable. The nearest approach to a suitable governor is the standard type of phonograph governor but this friction brake type of governor has serious objections if applied to a motor of considerable power output. In order, therefore, to have a control system which would be free from maintenance, it was necessary to develop a special form of control system for the purpose. Fig. 1 shows a photograph of the A.C. motor and its control cabinet.

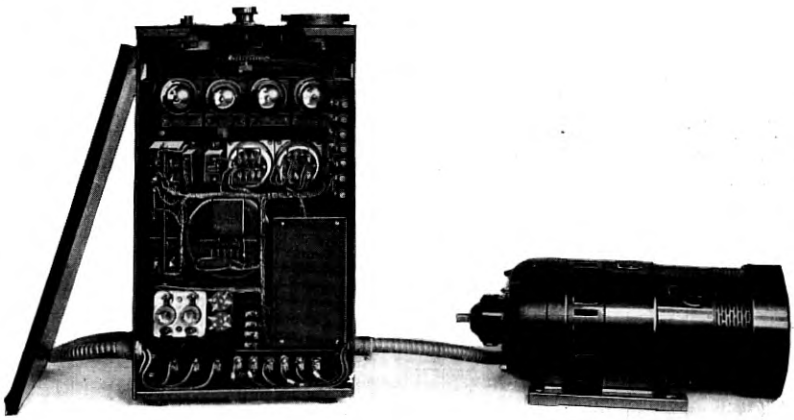


Fig. 1—A. C. Motor with control cabinet.

A.C. Motor Control Circuit

Fig. 2 shows the A.C. motor circuit which consists of a repulsion type of motor coupled to a small auxiliary alternator providing a frequency of 720 cycles which through a control circuit is made to operate a variable reactor across the armature terminals of the motor. If the speed of the motor is too high, the control circuit produces a maximum impedance in the reactor L_1 thereby reducing the armature current of the motor and causing it to slow down. While if the speed is too low the control circuit causes the reactor L_1 to have a minimum of impedance increasing the armature current and causing the motor to speed up.

This reactor L_1 is of the D.C. saturating type having two outer legs with A.C. windings and the middle leg with a D.C. winding. The A.C. flux circulates around the two outer legs. The D.C. flux flows from the middle leg and returns through the outer legs in parallel. When D.C. flux is sent through the middle leg it saturates the outer legs thereby reducing their impedance to A.C. This type of reactance is old and was employed by Alexanderson as a magnetic modulator in his early radio work.

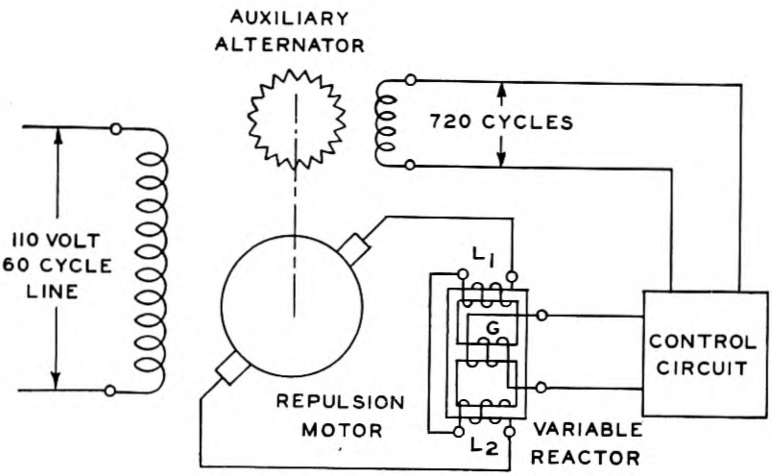


Fig. 2—A. C. Motor circuit diagram.

Fig. 3 shows one element of the speed control circuit. This consists of a bridge circuit having one variable arm and three fixed arms. The variable arm comprises a tuned circuit consisting of the in-

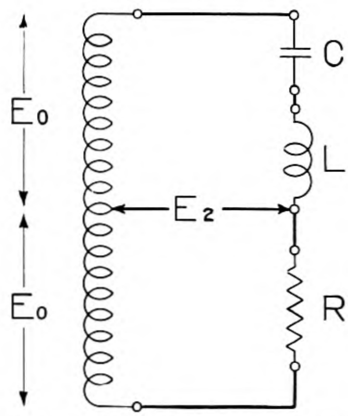


Fig. 3—Bridge circuit in speed control system.

ductance L and the capacity C which are designed to tune at exactly the frequency corresponding to the desired motor speed. When this circuit is in tune it has a resistive impedance which is balanced by the fixed arm R of the bridge. The other two fixed arms on the left are windings of a transformer with a mid tap. If a voltage $2E_0$ having a frequency of 720 cycles (which is the frequency corresponding to the desired speed of 1,200 R.P.M.) is supplied to the bridge it will be apparent that the output voltage E_2 will be zero. If, however, the speed is low the tuned circuit will have a condensive reactance while if the speed is high it will have an inductive reactance. The output voltage E_2 will, therefore, change abruptly 180 electrical degrees from a speed below 1,200 to a speed above 1,200. This characteristic is shown in Fig. 4.

The use of the above described bridge circuit gives a very sharp characteristic due to the fact that the effective resistance component of the tuned circuit is balanced out by the adjacent resistance arm of

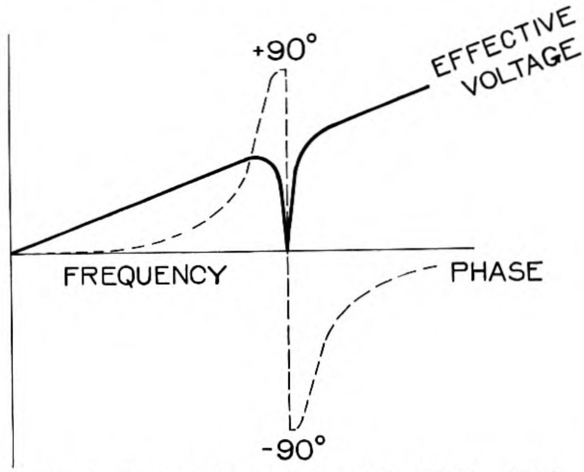


Fig. 4—Output voltage characteristic of bridge circuit.

the bridge. In this way an overall characteristic of the desired sharpness is secured using a comparatively small and inexpensive coil and condenser.

Fig. 5 shows the complete control circuit. The output from the bridge circuit is supplied to the grid of tube V_4 which is called the detector tube. The plate voltage of this tube comes from the 720-cycle generator through the step-up transformer T_4. The phase of this voltage, therefore, remains constant. The phase angle of the grid voltage, however, comes from the bridge output circuit through the step-up transformer T_3 and as previously explained suffers a sudden reversal of phase as the speed passes through 1,200 R.P.M. Fig. 6 shows the resulting current characteristic through tube V_4. This current flows through coupling resistance R_1 which drives the grids of tubes V_1 and V_2 negative. This in turn reduces the plate current through tubes V_1 and V_2 and hence through the D.C. winding of the inductance L_1 controlling the armature current of the motor.

Tube V_3 is a rectifier tube supplying excitation to the field of the

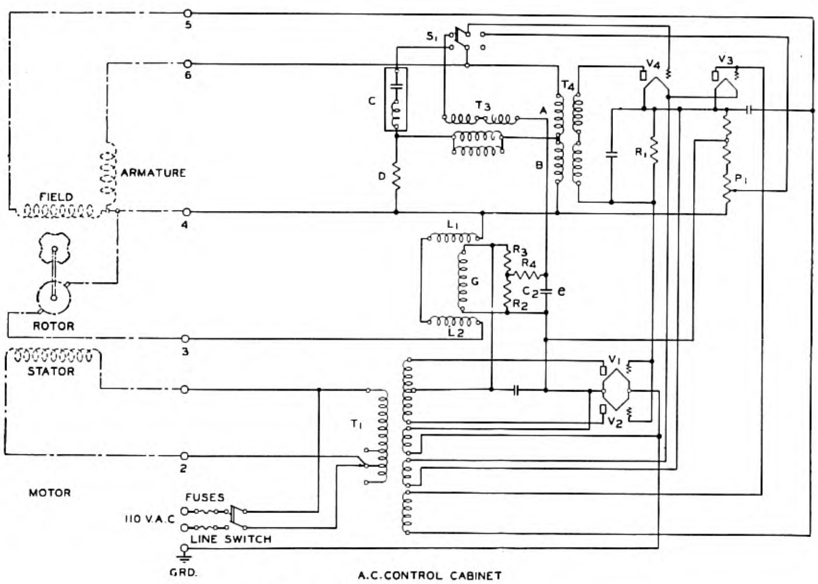


Fig. 5—A. C. control circuit diagram.

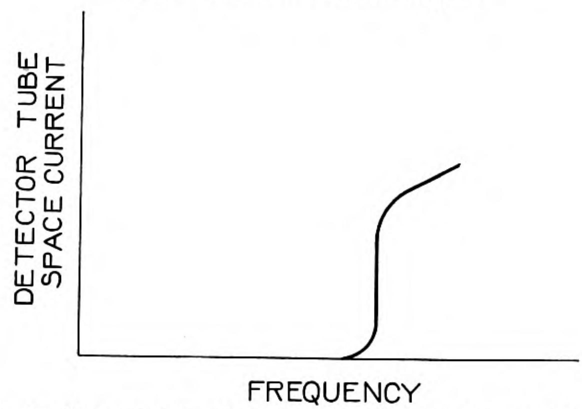


Fig. 6—Characteristic of space current through detector tube V_4.

720-cycle alternator and the negative "*C*" voltage along potentiometer P_1.

Fig. 7 shows the performance characteristics of the motor. It will be noted that the actual speed characteristic is practically flat. This flat characteristic is secured by a compensating network consisting of the resistances R_2, R_3 and R_4 and the condenser C_2. This compensating network feeds back on the grid of tube V_4 a portion of the voltage drop across the D.C. winding of inductance L_1 thereby correcting for the "static fluctuation" of the control circuit. By a suitable adjustment of this compensating resistance the control circuit may be arranged to give flat regulation, under regulation or even over regulation if desired. Fig. 7 has been drawn with line voltage as the variable. A similar characteristic is also obtained with load as the variable instead of voltage.

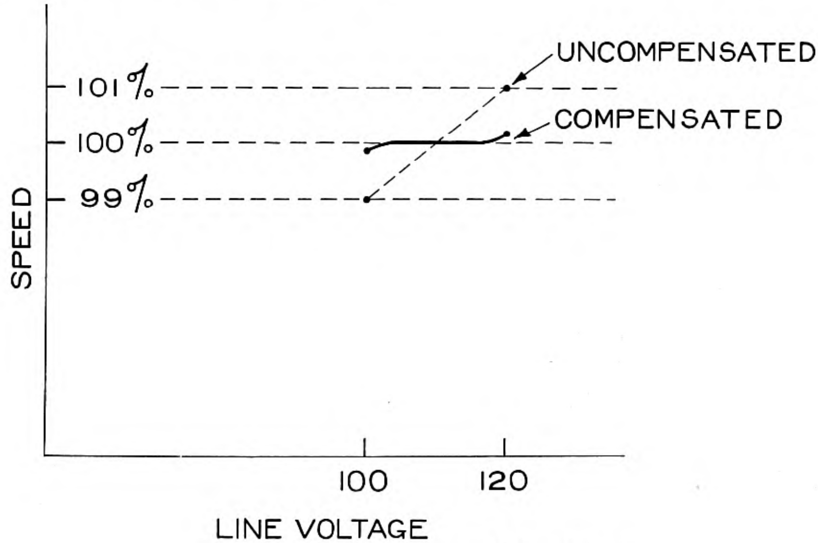


Fig. 7—Performance characteristics of motor.

An interesting point in connection with this compensation circuit is the necessity for avoiding hunting or surging of the speed. It is a well-known property of all forms of governors that if they are adjusted to too great a sensitivity the speed instead of remaining constant will fluctuate up and down about a mean value. The simplest method of preventing such speed fluctuations is to decrease the sensitivity of the governor allowing a bigger change in speed with load (or voltage) and then compensating for this change of speed or "static fluctuation" by means of a delayed action compensator. This phenomenon is well-known in the mechanical governor art and is described by Trinks in his book "Governors and the Governing of Prime Movers." The electrical equivalent of this mechanical system is obtained by introducing the condenser C_2 in series with the high resistance R_4. When a change in current through the regulating reactance L_1 occurs the corresponding change in voltage drop is not transmitted to the condenser C_2 immediately, but C_2 changes its voltage after a certain time lag (approximately 1 second), required to charge the condenser through the resistance R_4. The introduction of this time lag restores the precision of the circuit to the flat characteristic desired without introducing hunting.

Variable Speed Operation

By throwing the switch S_1 to the right the operator can disconnect the tuned circuit control and substitute a potentiometer P_1 as a

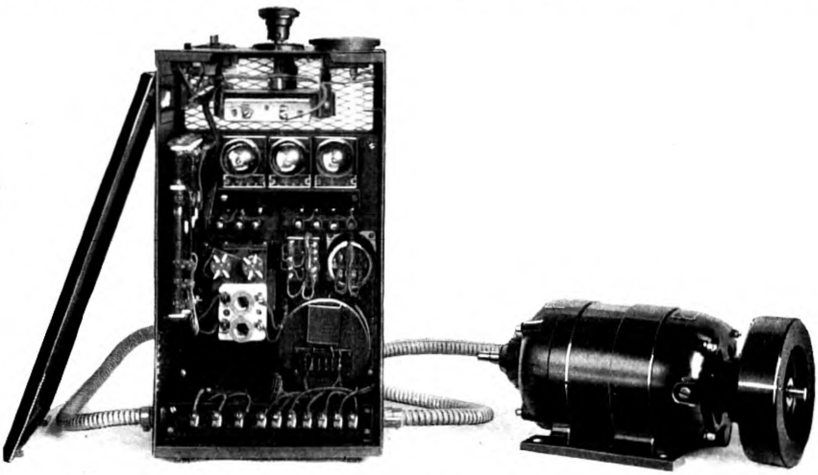


Fig. 8—D. C. motor with control cabinet.

source of grid voltage for tube V_4. By means of this potentiometer the operator can adjust the speed of the motor at any speed from 900 to 1,500 R.P.M. corresponding to 68 to 112 feet of film per minute. This feature is employed for ordinary motion picture work where it is unnecessary to synchronize the picture with the sound. The regulation of the circuit under these conditions is sufficiently good for ordinary motion pictures.

An interesting feature in this connection is that theaters in many cases have preferred to use the regulated speed position for ordinary motion pictures as well as synchronized pictures. The reason for this being that with the speed of the projector precisely controlled the orchestra leader is better able to keep his orchestra in step with the picture indicating apparently that closer speed regulation than is at present provided would be desirable for ordinary motion pictures as well as synchronized pictures.

D.C. Control Circuit

A circuit very similar to the one just described is employed in the case of the D.C. motor. Fig. 8 shows a photograph of this motor and its control cabinet. The circuit is shown in Fig. 9. It differs from the A.C. circuit in that an auxiliary regulating field winding is employed on the motor instead of a variable reactor. The source of power for the plates of the vacuum tubes is obtained from the auxiliary 720-cycle generator instead of from a 60-cycle transformer as in the

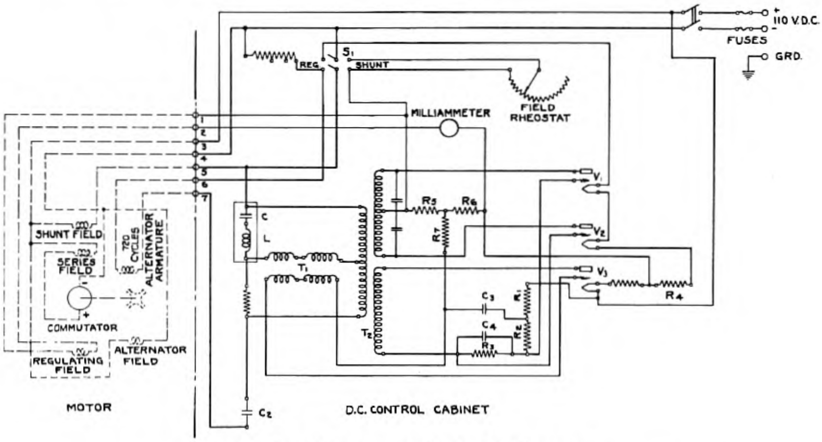


Fig. 9—D. C. control circuit diagram.

A.C. circuit. Since a strengthening of the field of the D.C. motor is required in order to reduce the speed it is necessary to reverse the phase relationship of the transformer T_1, so that the current in the detector tube decreases at speeds above 1,200 instead of increasing as in the case of the A.C. circuit.

The operation of the circuit is as follows: When the line switch is first thrown the motor acts as an ordinary D.C. shunt motor and accelerates. At low speeds the output from the 720-cycle generator is low and consequently there is no plate voltage supplied to the tubes and no current through the auxiliary field winding. The field is, therefore, weak and the motor speeds up. This condition is maintained until the equilibrium speed of 1,200 R.P.M. is approached. The phase angle of the voltage supplied to the grid of the tube V_3 is then in phase with the voltage supplied to the plate so that the grid of the tube goes positive at the same time that the plate is positive.

This causes a current to flow through the coupling resistances R_1 and R_2 which drives the grids of tubes V_1 and V_2 negative, thereby keeping down the current through these tubes and hence maintaining a weak motor field. The motor, therefore, continues to accelerate until a speed of 1,200 R.P.M. is reached. At this point as previously explained under the description of the bridge circuit, the phase of the output suddenly reverses whereupon the grid of the detector tube goes negative at the same time that the plate goes positive, thereby cutting off the current through the detector tube V_3 and reducing the negative C voltage on the grids of tubes V_1 and V_2. This increases the plate current through the regulating field thereby stiffening the field of the motor and checking its rise in speed. In practice the current through the detector tube is neither at one extreme nor the other but reaches an equilibrium at the speed of 1,200 R.P.M. A feedback network having the delay feature for prevention of hunting is included in the same manner as previously described for the A.C. circuit. The characteristic curves for the D.C. motor are similar to those shown in Fig. 7 for the A.C. motor.

For the operation of ordinary motion pictures the motor is changed to a simple shunt D.C. motor by the switch S_1 and the speed varied by means of the field rheostat.

Motor Drive of Recording System

It might appear that the simplest method of securing synchronization in recording work would also be mechanical connection between the recording machine and the camera. It has been found desirable, however, from a practical standpoint to have the camera movable with respect to the recording machine as the recorder has to be accurately lined up and adjusted and is not essentially a portable machine whereas the camera in ordinary motion picture work must be a portable piece of equipment. It has been necessary, therefore, to develop a motor drive equipment which will satisfactorily interlock the camera and the recording machine but leave the camera unit portable. It is essential that the interlock should hold not only during normal conditions but during acceleration and deceleration. In other words, the system must be the full equivalent of a mechanically geared system. The principle employed is old being disclosed in a patent issued to Michalke in 1901. In Fig. 10, A and B are two units which it is desired to interlock. Each unit has a three phase stator and a three phase rotor, the latter provided with slip rings. Magnetizing current for the system is supplied from an independent three phase, 60-cycle source. If the rotors of A and B are in exactly the same positions with respect to the stators it is evident that the e.m.f.'s produced in them by transformer action will be identical as to voltage and phase. Consequently there will be no flow of current over the rotor leads and hence no torque developed. If, however, unit A is turned through a small angle then the phase of the e.m.f.'s produced in the rotor circuits will differ from that in B and a current will flow in the rotor circuits producing a torque which will tend to make unit B assume the same position as A. If A is rotated continuously B will follow it up to synchronous speed of the stator field at which point the torque will drop to zero since no e.m.f. is induced in the rotor of either machine.

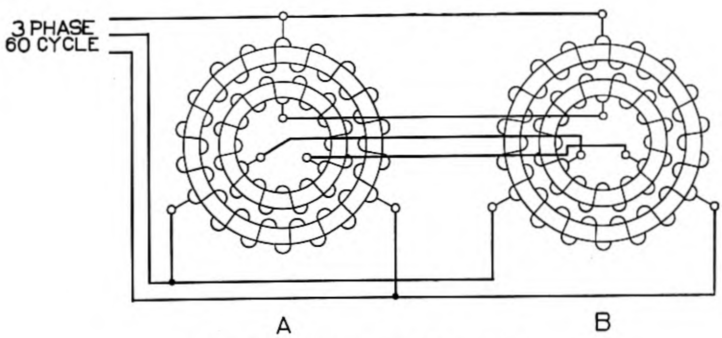


Fig. 10—Electrical driving gear.

Complete Recording Circuit

The portion of the circuit shown in Fig. 10, is merely the equivalent of a mechanical gear, neither unit tending to rotate as a motor by itself. In order to produce such rotation, therefore, a distributor set is added as shown in Fig. 11, the distributor acting, to use a mechanical analogy, as the driving gear of the system and each of the individual units of the system as driven gears. The distributor is itself driven by a D.C. motor provided with the speed control circuit previously described and shown in Fig. 9. The speed of the system is thus solely dependent on the D.C. driving motor and independent of the 60-cycle excitation frequency.

In practice the system is controlled by an operator at the distributor set and by means of switches any desired number of cameras, recording machines, or projectors may be employed. The projecting machines are used in case it is desired to make up a sound record to accompany an ordinary motion picture film which has previously been recorded without sound accompaniment. It has been found that the system operates very satisfactorily and requires very little maintenance.

When starting up for the first time it is necessary that the various units should line up properly as to phase otherwise there will be a local flow of current in the rotor circuits, which will cause the motors to operate as induction motors and run away. Under running conditions the system is very stable showing no tendency to hunt or surge between units, for the reason that being polyphase each phase as it becomes inactive (when the induced e.m.f. passes through zero) acts as a damping winding for the other two active phases.

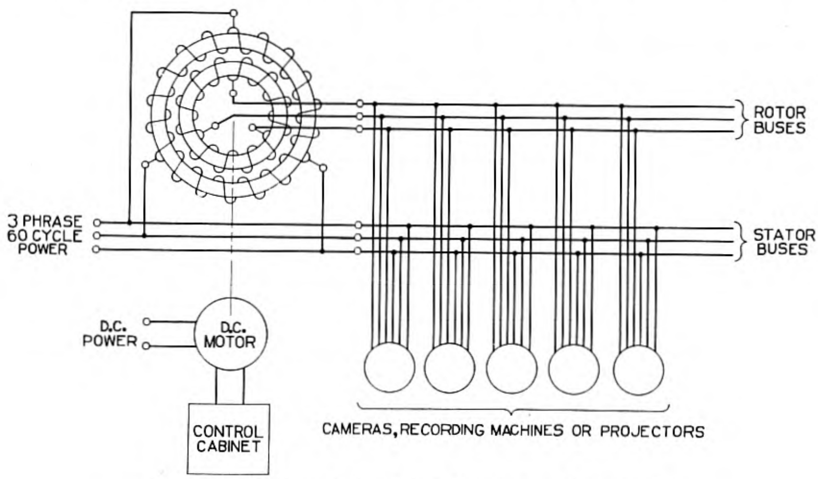


Fig. 11—Diagram of synchronizing system for recording.

If the load on a particular unit of this system is varied there will be a variation in the phase angle between this unit and other parts of the system in the same manner as in the case of a synchronous motor of the ordinary type. The magnitude of this phase angle, however, does not vary more than 30 electrical degrees or 15 mechanical degrees and is sufficiently small so that it produces an inappreciable effect on the synchronization.

Discussion

The above described motor equipment with its associated control circuits has been in practical use for over a year both in recording and reproducing work and there have been practically no troubles in service.

In the design of this equipment, first consideration has been given to its precision and reliability in operation and the provision of adequate margins to care for all variations in service conditions. As a result it has been possible to maintain a high standard of quality in music and speech reproduction.

A Sound Projector System for Use in Motion Picture Theaters[1]

By E. O. SCRIVEN

SYNOPSIS: The general problem involved in the design of a system suitable to be used to record and reproduce sounds such as are required for "talking" motion pictures is outlined. The general method of attack is indicated. There follows a description of the several pieces of apparatus which comprise the theatre equipment, including a discussion of some of their salient features and of the part each plays in the sound projector system.

IN order to reproduce in a theater the pictorial record of events accompanied by the sound associated with those events, it is, of course, necessary to add equipment to that installed to produce only the silent motion picture. It is the purpose of this paper to outline and discuss briefly the major items of such equipment as developed by Bell System engineers.

In the design of sound equipment a primary requisite is that there shall be freedom from distortion. Distortion may be of the sort which is independent of load and is evident in that the intensity of some portion or portions of the sound spectrum is increased or decreased in comparison with the rest; or there may be the distortion which is a function of the level at which the device is operated and is characterized by the reduction of a pure tone into fundamental and one or more harmonics. This latter condition is most often the consequence of operating a vacuum tube amplifier above its proper energy handling capacity.

It is the resonances of vibrating strings or reeds or air columns or vocal cords that give us the music we record but we are careful that a minimum of the resonances of the recording system itself shall go into the record, and that any resonances of the reproducing system shall not appear in the output of the sound projectors. Aside from the effects of overloading, the prevention of distortion is largely a matter of getting away from resonance phenomena since it is the characteristic of the resonant system to respond with disproportionate amplitude to stimuli in the region of its own natural period. The whole story of the passage from sound energy through the various recording and reproducing devices back to sound energy again is one of contest with this fundamental physical phenomenon.

[1] Presented before Society of Motion Picture Engineers at Lake Placid, New York, September, 1928.

There are in general two things one can do to avoid the harmful effects of resonance in vibration transmitting or transforming apparatus: (1) the period of resonance of each piece of equipment can be moved outside the range of frequencies one wishes to transmit, at the same time providing damping means to minimize free vibrations; (2) the distortion produced by resonance in one piece of apparatus can be compensated for or equalized by similar and opposite distortion

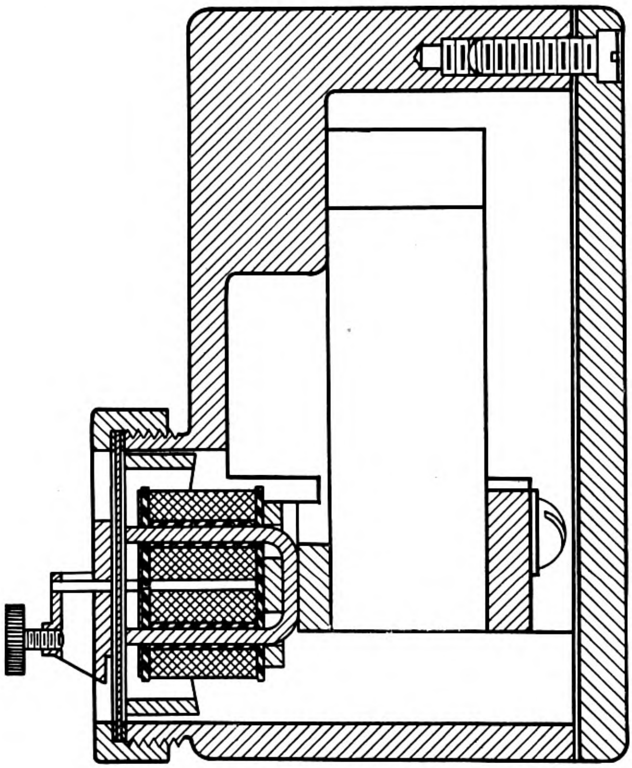


Fig. 1—Reproducer for disc record.

in some associated apparatus. The first is not always easy of practical accomplishment in any particular device and generally results in an instrument of very low inherent efficiency; the second usually involves loss of energy. In both cases increased amplification is required.

The sound record comes to the theater either as a wavy groove in a composition disc or as a striated track of varying density at one side of the picture film. It is the function of the apparatus being considered to derive from these records an electric current in which all the variations in pitch and loudness are accurately represented, to suitably amplify this current, to effect its conversion into sounds approximating those from which the records were made, and to so direct those sounds as to reasonably create the illusion that sound and picture are cognate.

The disc records do not differ essentially from those used in the ordinary phonograph except that they are considerably larger and run at a much slower speed so that a single record will play throughout an entire reel. The reproducer used is in some ways similar to that

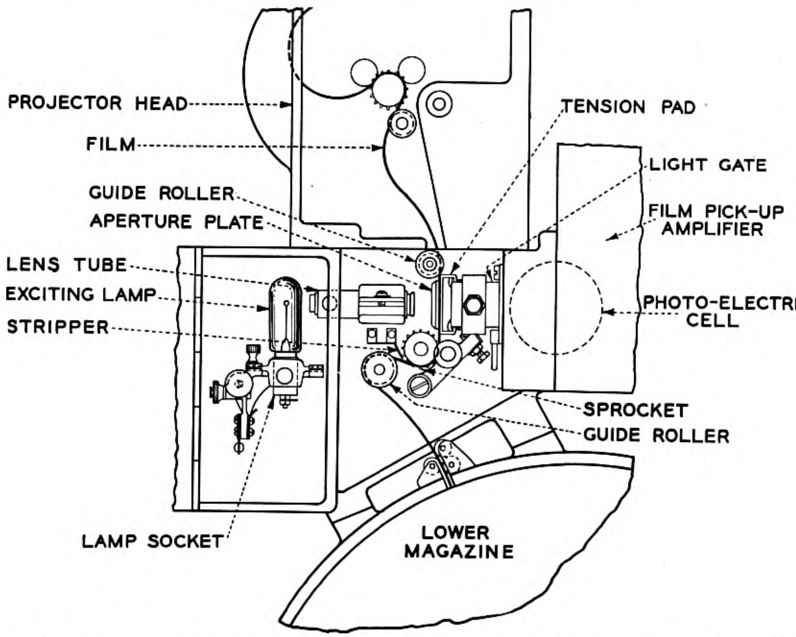


Fig. 2—Diagram of motion picture projector equipped for reproducing sound from film.

used on the acoustic phonograph, the needle holder being connected to a clamped diaphragm. This diaphragm is of highly tempered spring steel and to it there is fastened an armature made of a special high permeability alloy and so arranged that as the diaphragm vibrates the flux in the air-gap of a permanent magnet varies correspondingly, thereby inducing in appropriately placed coils currents which are the electric representation of the wavy groove which the needle travels. This reproducer is shown in Fig. 1. Although the energy delivered by this instrument is comparatively low it has a very uniform response over a wide frequency range. This result is largely brought about by moving all resonances out of the working range and by filling the magnet chamber back of the diaphragm with a heavy damping oil. The film used with the disc record, called a synchronized film, differs from ordinary film only in that one frame at the beginning is specially marked to give the starting point.

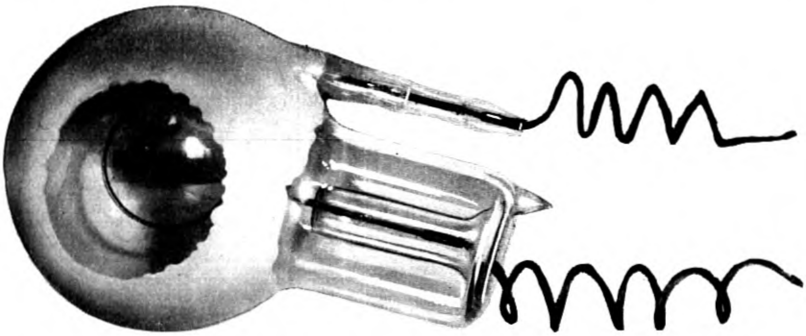


Fig. 3—Photoelectric cell.

The film sound record, as has been said, consists of a track of varying density running along one side of the picture. This sound track is 1/10″ wide. Differences or changes in intensity of sound are represented by differences in the density of the record, while pitch

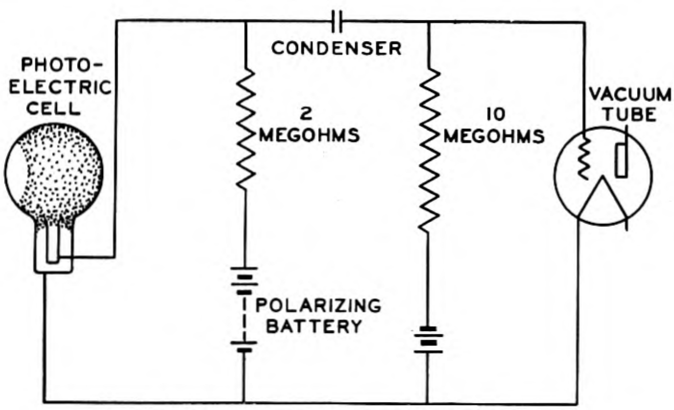


Fig. 4—Photoelectric cell circuit.

is represented by the number of changes from dark to light and back again in a given length of track. This sound record is converted into a corresponding electric current by arranging that a narrow high intensity beam of light shall pass through it and fall upon a photoelectric cell. The arrangement is shown in Fig. 2. The light from the bright filament of the exciting lamp is focused as a very narrow line upon the film by passing through a system of lenses and an aperture plate. The lamp filament is focused upon a slit of dimensions .0015″ x 3/16″. The image of this slit is then brought to focus upon the film as a .001″ line whose length has been reduced in passing through the aperture plate to .080″. This reduction in length allows .010″ on either side for variations in position of the .100″ sound track. The position and focus of the lens tube are fixed, but the carriage of

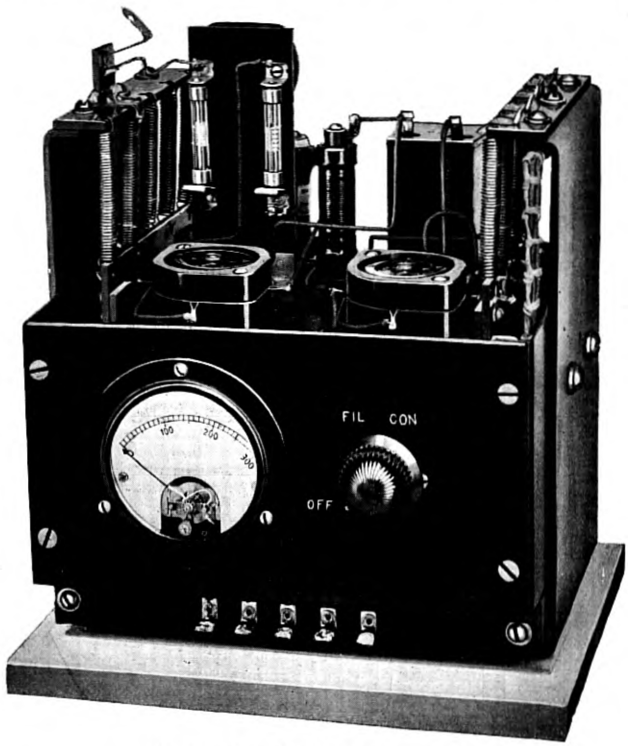


Fig. 5—Amplifier used at projector.

the exciting lamp is movable so that when replacing lamps the filament may be properly brought on focus.

A photoelectric cell of the type used is shown in Fig. 3. The characteristic of this device is that when it is polarized by a proper voltage and is used within proper limits the current through it is proportional to the incident light. The circuit is shown in Fig. 4. It is to be noted that the polarizing voltage is supplied to the photo cell through a very high resistance and there is, therefore, obtained across this resistance a voltage which is proportional to the light falling upon the cell and accordingly bears a direct relation to the varying density of the sound track interposed between the exciting lamp and the cell.

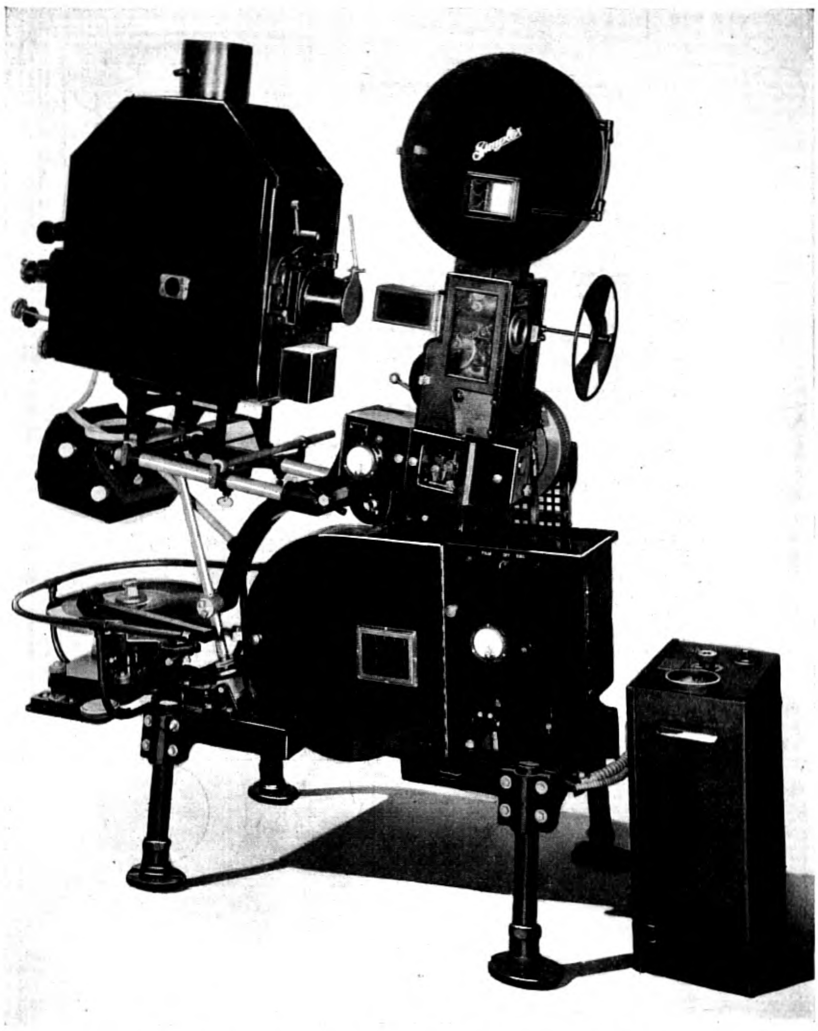


Fig. 6—Sound projector equipped with Simplex head.

The photo cell circuit is inherently one of high impedance. In such a circuit there are two matters which require attention. (1) Local interference—"static," to use the radio expression, is most readily picked up, and at this point where the energy level is low, may be appreciable in comparison with the sound currents themselves; (2) also the shunting effect of capacity between the electrical conductors becomes noticeable, particularly at the higher frequencies.

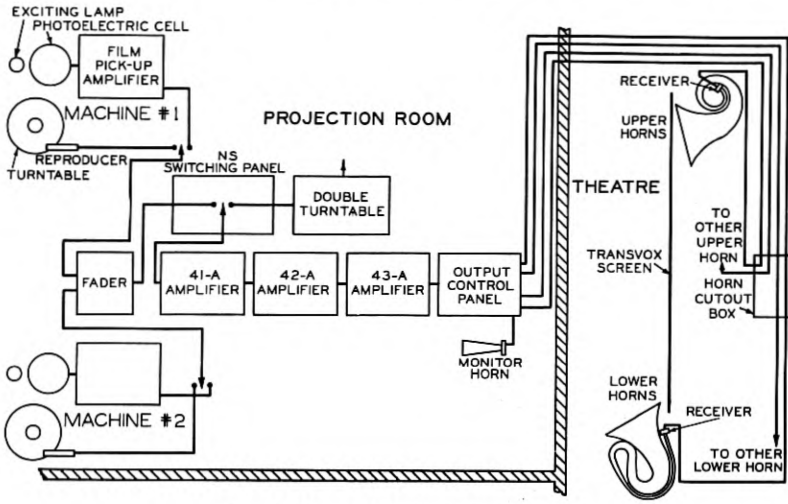


Fig. 7—General layout of equipment.

Hence a vacuum tube amplifier, which serves both to increase the energy and to make that energy available across a low impedance circuit, is closely associated with the cell upon the projector itself.

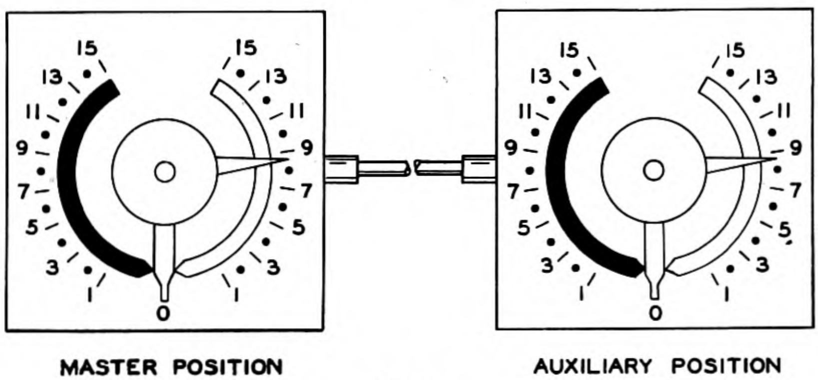


Fig. 8.

The cell and amplifier are enclosed in a heavy metal box or shield which is made fast to the frame of the projector and the projector itself is carefully grounded. This amplifier is shown in Fig. 5. It is designed to bring the level of the electric counterpart of the film sound record up substantially to the same energy value as that obtained from the magnet coils of the disc reproducer. The filaments are heated from a 12-volt storage battery. Small dry batteries supply its plate

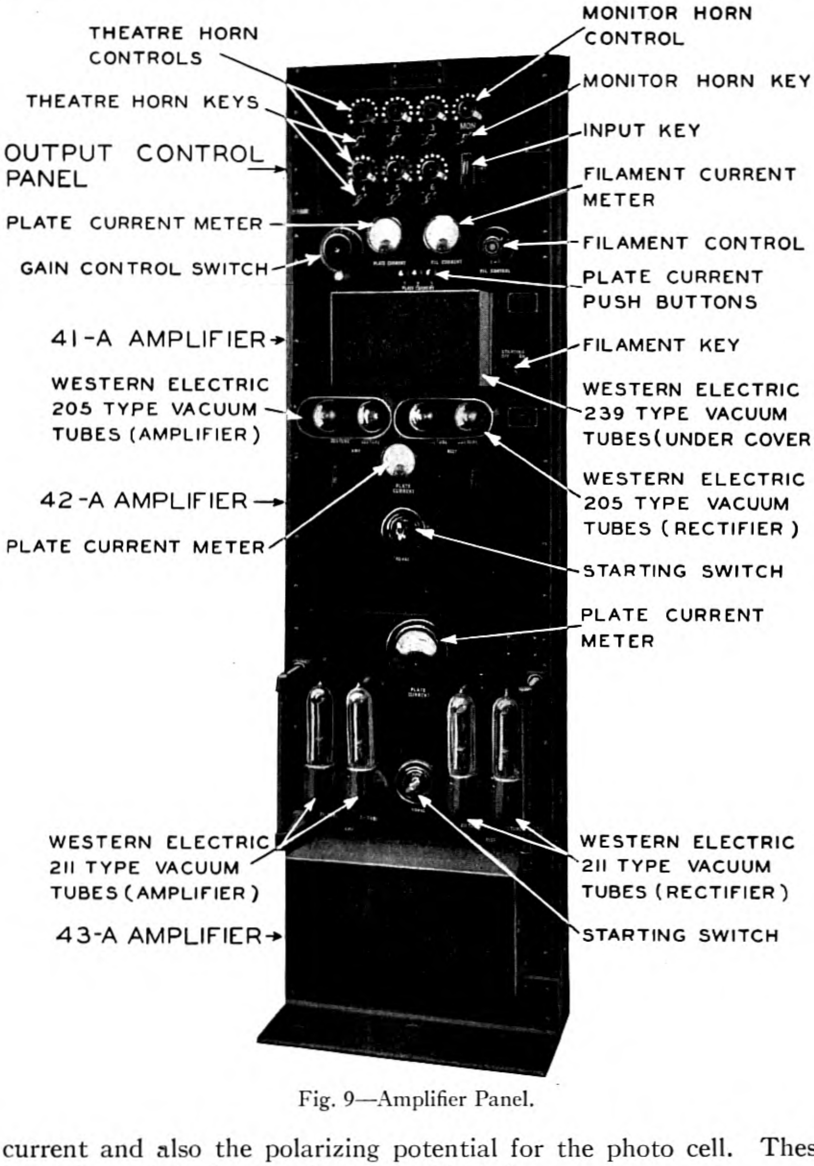


Fig. 9—Amplifier Panel.

current and also the polarizing potential for the photo cell. These batteries and the battery leads are shielded.

Vibration of a vacuum tube often produces sufficient motion of its elements with respect to each other to effect changes in the stream of electrons which appear when sufficiently amplified as noise from a loud speaker. In spite of all precautions there is a certain amount of vibration of the projector when in operation and it has therefore been necessary to design a rather elaborate shock-proof mounting for the photo cell amplifier.

It is evident from the relative location of apparatus as shown in Fig. 2 that it is not feasible to print the film sound record directly beside the picture to which it applies. As a matter of fact, there is a spacing of $14\frac{1}{2}''$ between picture and corresponding sound record and a certain amount of slack is allowed between the sprocket which carries the picture with an intermittent motion before the picture projection lens and the sprocket which must carry the sound record

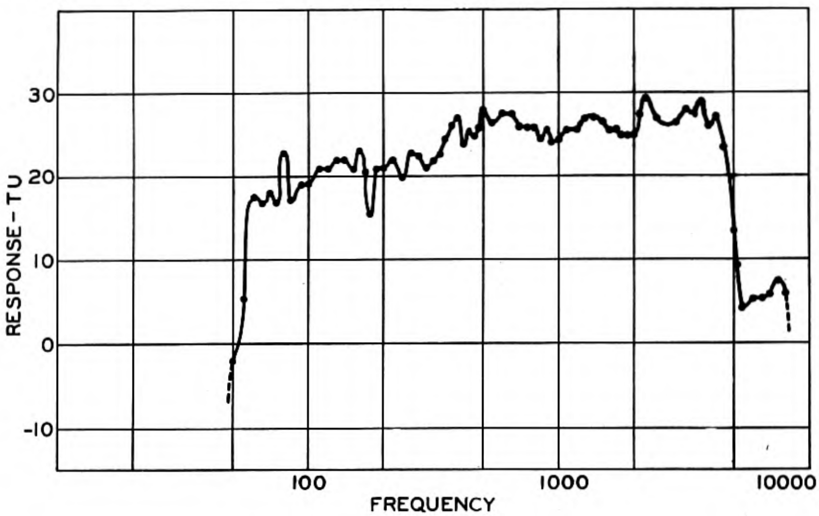


Fig. 10.

with a uniform motion in front of the photoelectric cell. In this connection it is noteworthy that special precautions are necessary in order to prevent vibrations and speed fluctuations due to either varying supply voltage or varying load from affecting the uniformity of rotation of this sound sprocket. This is taken care of by the very effective means of automatically controlling the speed of the driving motor and by means of a mechanical device interposed between the sound sprocket and the rest of the moving equipment of the projector which effectively opposes the transmission of any abrupt change of speed to this sprocket.

The control box which contains the apparatus for governing the speed of the driving motor is arranged to hold the record speed the same as that at which the records are made, i.e. 90 feet per minute in the case of synchronized sound and picture productions. By throwing a switch the automatic feature may be cut out and the speed of the machine may then be manually controlled by the operator.

This completes the apparatus associated directly with the projector. The general arrangement of the latest type of projection machine, equipped with a Simplex head, is shown on Fig. 6. Incidentally this projector is also arranged to be fitted with the Powers or the Motiograph head. Fig. 7 shows a typical layout of a sound projector system as installed for use with talking motion pictures.

As in ordinary pictures, in order to run a continuous program, it is necessary to use two projectors alternately. As the picture from one machine is faded imperceptibly into that on the other so the sound record may be faded from one machine to the other without the audience being aware that a change has been made. At the end of each record or sound film the music overlaps the beginning of the next and a device called a fader is employed in making the transition. All that is necessary is to turn the fader knob when the incoming machine is started. This fader is in fact a double potentiometer. In the upper or normal operating range the change in volume in moving from one step to the next is hardly more than perceptible whereas in the lower range used only in fading the steps are large and the volume decreases to zero on one machine and builds up on the other very rapidly. By choosing the proper step in the upper range one can obtain any volume of sound desired within reasonable limits and thereby equalize the level obtained from different sound records. The fader is ordinarily installed with one or more auxiliary dials and handles interconnected so that it may be operated from any projector position. In connection with the fader there is provided a switch for changing from the film to the disc input system and also a key for switching a spare projector in place of either of the regular machines. Fig. 8 shows a fader with one auxiliary position.

Following the fader, we come to the main amplifier which raises the energy of the feeble electric currents to a level adequate to supply the loud speakers with sufficient volume to serve the particular theater. Fig. 9 shows a typical amplifier panel. This combination is capable of an energy amplification of about 100,000,000 times and is so designed that all frequencies in the range from 40 to 10,000 cycles are amplified practically equally. A potentiometer is provided on the amplifier but while its handle is readily accessible it is ordinarily not used after having once been set at the time of installation to give proper results in the particular theater. Necessary adjustments are made on the fader. The amplifier shown consists of three units. The first consists of three low power tubes in tandem, resistance coupled, and requiring a 12-volt battery delivering 1/4 ampere to heat their filaments. The second consists of a single stage of two medium power tubes, connected in push-pull arrangement with filaments

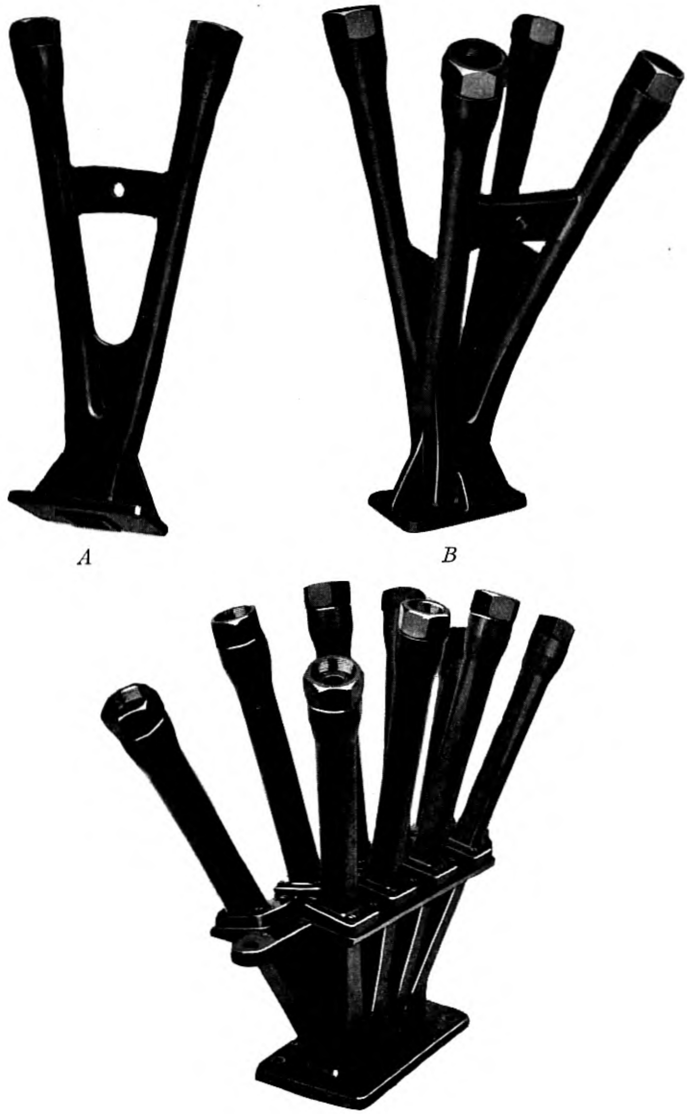


Fig. 11.

heated by low voltage alternating current. Two similar tubes in this unit operate as a full wave rectifier and supply rectified alternating current for the plate circuits of the amplifier tubes of both the first and second units. The third unit has a single stage of high power push-pull amplifier tubes and push-pull rectifier tubes and also operates entirely on alternating current.

These three types are capable of arrangement into combinations to meet the particular need. For small theaters only No. 1 and No. 2 are required. In the larger houses the high power unit No. 3 is added, while to meet exceptional conditions two or more of the high power amplifiers may be operated in parallel from the output of No. 2.

Following the amplifier there is an output control panel. This consists of an auto-transformer having a large number of taps, the taps being multipled to a number of dial switches, to which the sound projectors or loud speakers are connected. By means of this panel, it is possible to match the impedance of the amplifier output to the desired number of horns in order to obtain the most efficient use of the power available and also to adjust the relative volume of the individual horns.

The ordinary theater installation employs four horns, two mounted at the line of the stage and pointed upward toward the balconies and two mounted at the upper edge or above the screen and pointed downward. This combination has been found to give good distribution throughout the house.

The loud speaker unit used with the horns in theater equipments is essentially that recently described by Messrs. Wente and Thuras.[2] As brought out in this article, this unit shows extremely high efficiency; about 30 per cent of the electrical power supplied is radiated in the form of sound. This is important since the higher the loud speaker efficiency, the smaller the power capacity of the amplifier needed in the system. The frequency-response characteristic of a typical receiver and horn is given in Fig. 10. An individual horn may be equipped with two, four or nine loud speakers by using the throats shown in Fig. 11. The power capacity for continued safe operation of the horn with one, four and nine throats is approximately 5, 20 and 45 watts, respectively (electrical input). The number of horns used is dependent upon the particular installation and is related to the directive characteristic of the horn. If it is necessary to disperse the sound over a large angle, more horns are needed than when it is desired to concentrate over a comparatively small angle. This directive characteristic of the horn is important in talking motion pictures as it is responsible for the illusion of the sound coming directly from the mouth of the horn; that is, from the screen. If the horn is replaced by a loud speaker of otherwise identical characteristics but which radiates its sound over a very wide angle, there is a tendency for the sound to appear to come from a point some distance back of the screen, thus tending to destroy the illusion.

The power supply equipment has been fairly completely covered in discussing various parts of the system. Under ordinary conditions the requisite power is obtained from the electric mains in the theater except for the 12-volt battery required for some of the vacuum tube filaments and for the electromagnets in the loud speakers and the dry cells used with the photo cell and photo cell amplifier. Where 110-volt D.C. only is available there is a projector-driving equipment which operates on this voltage, but a D.C. motor driving a 60-cycle generator is required for supplying the amplifiers. Where 110-volt A.C. is available, it is only necessary to connect the projector motor and the amplifiers to this supply.

[2] *Bell System Technical Journal,* January, 1928—"A High Efficiency Receiver of Large Power Capacity for Horn-type Loud Speakers," by E. C. Wente and A. L. Thuras.

Abstracts of Bell System Technical Papers Not Appearing in this Journal

The Communication System of the Conowingo Development.[1] W. B. BEALS and E. B. TUTTLE. This paper describes the communication system which has been installed to serve the power plant at Conowingo, Maryland, and its associated transmission line.

The important features to be considered in designing a telephone system for a power plant are pointed out. The types of telephone switchboard and telephone instruments chosen in this case to meet the special requirements of the generating station, together with the layout and cabling arrangement, are outlined.

The paper also discusses the possible ways of providing for the needs of the load dispatcher and the plan adopted at Conowingo; the facilities provided the patrolmen for calling from points along the transmission line; the connection from the private branch exchange to the general telephone system; and the special electrical protection installed on the long lines leaving the power house.

Reflection and Refraction of Electrons by a Crystal of Nickel.[2] C. J. DAVISSON and L. H. GERMER. This is a report of further observations on the regular reflection of electrons from the surface of a nickel crystal; an earlier report was published in the same journal.[3] In the present report data are given of the selectivity of reflection for angles of incidence from 10 to 50 degrees, and for electrons of wave-lengths 0.6 to 1.5 Å. The previously found result is confirmed that to explain the occurrence of the intensity maxima of the reflected beam it is necessary to assume that electron waves are refracted on passing into the crystal. The data are used for calculating indices of refraction for nickel for electrons of various speeds or wave-lengths, and a dispersion curve is constructed. This curve displays a feature suggestive of the optical phenomenon of anomalous dispersion.

Optical Experiments with Electrons.[4] L. H. GERMER. A semi-popular account of a series of experiments performed by C. J. Davisson and the author upon the scattering of electrons by single crystals of

[1] *Journal of the A. I. E. E.*, October 1928, pp. 737–741.

[2] *Proceedings of the National Academy of Sciences*, August 1928, pp. 619–627.

[3] *Proceedings of the National Academy of Sciences*, April 1928, pp. 317–322.

[4] *Journal of Chemical Education*, Part I, Sept. 1928, pp. 1041–1055. Part II, Oct. 1928, pp. 1255–1271.

nickel. These experiments establish the fact that under certain conditions moving electrons behave like trains of waves. In the interaction of these waves with a single crystal the optical phenomena of diffraction, reflection and refraction have been observed. Scientific accounts of these experiments are contained in the following papers: *Nature*, 119, 558 (1927); *Phys. Rev.*, 30, 705 (1927); *Proc. Nat. Acad. Sci.*, 14, 317 (1928); *Proc. Nat. Acad. Sci.*, 14, 619 (1928). Although the present paper is of a popular nature it aims to be quite comprehensive. It attempts to represent the status of this series of experiments in August 1928.

Rubber Compression Testing Machine.[5] C. L. HIPPENSTEEL. This paper gives a brief account of a new compression test developed at the Bell Telephone Laboratories for more reliably judging the ability of rubber insulation on metallic conductors to withstand certain service conditions to which it is subjected. A recording compression testing machine, which has been built for applying the test, and typical results are illustrated. Other possible test uses for the machine are suggested.

New Languages from Old—How Secrecy is Gained by the Inversion of Speech Sounds.[6] C. R. KEITH. The inversion of speech sounds may be accomplished with the aid of methods used in radio broadcasting and in carrier telephony. Among the possible applications, it is illustrative of methods used to achieve secrecy in electrical communications.

The character of speech sounds is determined by the frequencies and amplitudes of the component waves into which the sound may be resolved. The process of inversion consists effectively in altering the frequency distribution of these components so that low tones appear as high tones, while high tones appear as low tones. To the untrained observer, inverted speech is unintelligible, although the characteristic cadence is preserved. Inversion of the frequency scale is produced by modulating speech with a carrier wave which lies just above the highest speech frequency which is to be transmitted, and selecting the lower sideband. For practical reasons connected with undesired distortion, it is more desirable to break up the modulating process into two distinct steps. The original speech sounds may then be regained by repeating the process which led to its inversion.

[5] *India Rubber World*, Sept. 1928, pp. 55–56.
[6] *Scientific American*, October 1928, pp. 310–311.

Joint Pole Use with Power Companies.[7] D. E. LOWELL. The relations between the telephone company and the other wire using companies, especially the power companies operating in the same area, are discussed in this paper. It recognizes the responsibility of the telephone company as well as that of the power company for good operating conditions in areas where both types of line are involved and also points out the necessity of close cooperation between Connecting and Bell Telephone Companies. The considerations involved in the joint use of poles by telephone and power companies are given with particular mention of the general joint use agreement. The importance of mutual advance notice of plans is developed. The reports of the Joint General Committee of the N. E. L. A. and Bell System form the background of the talk and are recommended to those who have not already read them.

Adsorption of Gases by Graphitic Carbon. II—X-ray Investigation of the Adsorbents.[8] H. H. LOWRY and R. M. BOZORTH. This paper is supplementary to one by Lowry and Morgan appearing in the *Journal of Physical Chemistry* in 1925 [9] and gives direct evidence that the adsorbents studied were graphitic carbon. The X-ray data show that carbon prepared by the explosion of graphitic acid is graphitic in structure and that the individual particles are flakes averaging approximately 50 atom diameters in breadth and 10 atom layers in thickness. The significance of this finding is discussed in relation to current views of the nature of active carbon adsorbents.

Recent Toll Cable Construction and its Problems.[10] H. S. PERCIVAL. One of the outstanding developments in the Bell System has been the rapid extension of toll cables. This has required the development of new methods and apparatus. Material is carried into rough right of way and installed through the use of tractors, with equipped trucks and various types of automotive equipment. The development of permalloy now allows the complete loading of a full-sized cable in two pots where six were required before. Crossings over rivers are made in submarine cable or by long span construction with catenary suspension. Cables are tested before completion for sheath damage, defective splices, etc., which might cause service failures, by means of dry gas under pressure.

[7] *Telephony*, September 8, 1928, pp. 22–24.
[8] *Journal of Physical Chemistry*, October 1928, pp. 1524–1527.
[9] *Journal of Physical Chemistry*, Vol. 29 (1925), p. 1105.
[10] *Telephone Engineer*, September 1928, pp. 31–33.

Quality Control by Sampling.[11] W. L. ROBERTSON. A discussion of the application of the mathematical theory of sampling to commercial shop inspection. Also gives tables illustrating numerically the results obtained from the various sampling plans in use.

Problems in Power Line Carrier Telephony and Recent Developments to Meet Them.[12] J. D. SARROS and W. V. WOLFE. Power transmission lines as commonly encountered present relatively complex networks having irregular and unstable attenuation-frequency characteristics within the 50–150 K.C. band employed for power line carrier telephony. The high frequency noise may be very high.

A single side band carrier suppressed system operating on a single frequency duplex basis has been developed to overcome these transmission difficulties.

A comparison of this system with other types shows its superiority.

The initial installation of this equipment was made on the 220 K.V. lines of the Pacific Gas and Electric Company.

The Planning of Telephone Exchange Plants.[13] W. B. STEPHENSON. This paper discusses procedures followed in planning future extensions to telephone exchange plants to care for increased demand for telephone service. An outline is given of the methods employed in forecasting future demand for telephone service and in determining the most efficient design of the plant to meet the service requirements. The uses made of engineering comparisons in solving the economic phases of various kinds of telephone engineering problems are discussed, with particular reference to location and size or extent of major items of plant as well as the time when they should be ready to give service. Emphasis is placed upon the importance of those factors less readily evaluated, such as service factors, practicability from a construction and operating standpoint, flexibility, etc.

The Effect of the Acoustics of an Auditorium on the Interpretation of Speech.[14] E. C. WENTE. Studies of speech sounds in the Bell Telephone Laboratories have shown that 60 per cent of the acoustic energy in speech lies below 500 c.p.s., although the intelligibility of individual speech sounds is reduced by only 2 per cent if all the

[11] *Factory and Industrial Management*, pp. 503–505, Sept. 1928; pp. 724–726, Oct. 1928.

[12] *Journal of the A. I. E. E.*, October 1928, pp. 727–731 (abridgment).

[13] *Journal of the A. I. E. E.*, July 1928, pp. 500–503 (abridgment).

[14] *The American Architect*, August 20, 1928, pp. 259–261.

energy below this frequency is completely suppressed. These results indicate that the sound absorption coefficient of materials placed in an auditorium for reducing the reverberation time should be high for tones of low frequency and low for those of high frequency. Most porous materials commonly used for this purpose have absorption characteristics quite the reverse. Rooms that have been treated with a rather large amount of such materials are therefore often unsatisfactory for speaking purposes, although the adjustment for reverberation time may have been carried out according to accepted standards.

Contributors to this Issue

W. H. MARTIN, A.B., Johns Hopkins University, 1909; S.B., Massachusetts Institute of Technology, 1911. American Telephone and Telegraph Company, Engineering Department, 1911–19; Department of Development and Research, 1919–. Mr. Martin's work has related particularly to transmission of telephone sets and local exchange circuits, transmission quality and loading.

H. S. OSBORNE, B.S., Massachusetts Institute of Technology, 1908; Austin Research Fellow in Engineering, 1908–10; Eng.D., 1910; American Telephone and Telegraph Company, Engineering Department, 1910–19; Department of Development and Research, 1919–20; Department of Operation and Engineering, 1920–. Mr. Osborne is Transmission Engineer and as such is responsible for assisting the Associated Companies in connection with telephone and telegraph transmission and protection matters.

G. W. ELMEN, B.S., University of Nebraska, 1902; M.A., 1904; Research Laboratories of the General Electric Company, 1904–06; Engineering Department of the Western Electric Company, 1906–25; Bell Telephone Laboratories, 1925–. Mr. Elmen's principal line of work has been magnetic investigations. He is the inventor of the permalloy alloys. For this he was awarded the John Scott Medal in 1926 and the Elliott Cresson Medal in 1928.

H. O. SIEGMUND, B.S., University of Illinois, 1917; E.E., University of Illinois 1926; Instructor, U. S. Army School of Military Aeronautics, 1917–1918; Assistant Professor of Electrical Engineering, Drexel Institute, 1918–1919; Engineering Department, Western Electric Company, 1919–1925; Bell Telephone Laboratories, 1925–. Mr. Siegmund has been engaged in apparatus development work and has made contributions relating to telephone power plant apparatus and circuits, to provide more quiet transmission.

KARL K. DARROW, S.B., University of Chicago, 1911, University of Paris, 1911–12, University of Berlin, 1912; Ph.D. in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, Inc., 1925–. Mr. Darrow has been engaged largely in writing studies and analyses of various fields of physics and the allied sciences. Some of his earlier articles on Contemporary Physics form the nucleus of a recently published book entitled "Introduction to Contemporary Physics" (D. Van Nostrand Company).

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919–. Mr. Carson is well known through his theoretical transmission studies and has published extensively on electric circuit theory and electric wave propagation.

Edward C. Molina, Engineering Department of the American Telephone and Telegraph Company, 1901–19, as engineering assistant; transferred to the Circuits Design Department to work on machine switching systems, 1905; Department of Development and Research, 1919–. Mr. Molina has been closely associated with the application of the mathematical theory of probabilities to trunking problems and has taken out several important patents relating to machine switching.

W. P. Mason, B.S., University of Kansas, 1921; M.A., Columbia, 1924; Ph.D., Columbia, 1928. Engineering Department, Western Electric Company, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Mason's work has been largely in transmission studies.

L. G. Bostwick, B.S. in E.E., University of Vermont, 1922; American Telephone and Telegraph Company, Development and Research Department, 1922–1926; Bell Telephone Laboratories, Inc., Research Department, 1926–. While with the Development and Research Department, Mr. Bostwick's work involved general problems on systems for the high quality transmission of speech and music; since then his work has been largely on loud speakers and loud speaker measuring methods.

H. A. Frederick, B.S., Princeton, 1910, E.E., Princeton, 1912; Engineering Department, Western Electric Company, 1912–1925; Bell Telephone Laboratories, 1925–. Mr. Frederick is in charge of researches and engineering on telephone transmission instruments.

Donald MacKenzie, Ph.D., Johns Hopkins University, 1914; Assistant in Astronomy, Johns Hopkins, 1914–17; Ensign, National Naval Volunteers, 1917–1918; Bureau of Standards, 1918–1920; Engineering Department, Western Electric Company, 1920–1925; Bell Telephone Laboratories, 1925–. Engaged since 1922 in the development of a system of sound recording by photographic means.

H. M. Stoller, E.E., Union College, 1913; M.S. in electrical engineering, 1915; Engineering Department of Western Electric Company, 1914 and 1916–1925; Bell Telephone Laboratories, 1925–. Most of Mr. Stoller's work has dealt with special problems connected with electrical power machinery, particularly voltage and speed regulators. He designed a multi-frequency generator which is now employed in the voice frequency carrier telegraph system.

E. O. Scriven, B.S., Beloit College, 1906; instructor, Fort Worth University, 1906–08; S.M., Massachusetts Institute of Technology, 1911; Engineering Department, Western Electric Company, 1911–25; Bell Telephone Laboratories, 1925–. Mr. Scriven has been identified with the development of apparatus employing vacuum tubes, e.g., amplifiers, oscillators, carrier current equipment, etc. At present his work is largely confined to public address and related audio frequency systems.