# The Bell System Technical Journal

*April, 1926*


## Development and Application of Loading for Telephone Circuits[1]


By THOMAS SHAW and WILLIAM FONDILLER



SYNOPSIS: A review of the art of loading telephone circuits as practised in the United States. The introductory section briefly reviews the theory of coil loading, and summarizes the principal characteristics of the first commercial standard loading coils and loading systems, thereby serving as a background for the description of the various improvements of outstanding importance which have been made in the loading coils and loading systems during the past fifteen years to meet the new or changing requirements in the rapidly advancing communication art.

These major improvements are described in detail under the appropriate headings (1) Phantom Group Loading, (2) Loading for Repeatered Circuits, (3) Incidental Cables in Open Wire Lines, (4) Cross-talk, (5) Telegraphy over Loaded Telephone Circuits, (6) Loading for Exchange Area Cables, and (7) Submarine Cables. The discussion of these various developments sets forth the relations between the loading features and the associated phases of telephone development, such as the cables, repeaters, telegraph working, and carrier telephone and telegraph systems.

The concluding part of the paper gives some general statistics regarding the extent of the commercial application of loading in the United States, and a brief statement indicative of the large economic importance of loading to the telephone using public.


### INTRODUCTION

THE year 1926 marks the fiftieth anniversary of the birth of the telephone, and the completion of the first 25 years of the commercial application of loading to telephone circuits by means of inductance coils inserted at periodic intervals. The present time is thus peculiarly appropriate for a survey of loading developments.

The purpose of this paper is to present a review of the art of loading telephone circuits, as practised in the United States. In a paper[2] presented before the Institute in 1911 Mr. B. Gherardi described the developments in loading up to that time and gave a comprehensive statement of the results obtained. In the present paper, therefore, references to the early developments in loading may be confined to matters that are necessary to the treatment of the subsequent developments in the art.

During the period under consideration many improvements of outstanding importance have been made in the characteristics of the load-

[1] Presented at the Midwinter Convention of the A. I. E. E., New York, N. Y., February 9, 1926.

[2] "Commercial Loading of Telephone Circuits in the Bell System," B. Gherardi, Trans. A. I. E. E., Vol. 30, 1911, p. 1743.

ing coils and in the loading systems, in order to meet new or changing requirements in the rapidly advancing communication art. The more important of these improvements are listed below and will be discussed in the sequence noted:

I. Phantom Group Loading
II. Loading for Repeatered Circuits
III. Incidental Cables in Open Wire Lines
IV. Cross-Talk
V. Telegraphy over Loaded Telephone Circuits
VI. Loading for Exchange Area Cables
VII. Submarine Cables

As a basis for the discussion of the characteristics of commercial loading systems and the various developments which have been made, the elementary theory of loaded lines and a review of the first loading standards will be given. Those interested in the exact mathematical theory are referred to more complete discussions which may be found in the bibliography appended hereto.

*Theory*[3]. It is convenient to discuss the coil loaded line in terms of its corresponding smooth line, a hypothetical line in which the constants of the inductance coils are assumed to be distributed uniformly along the line.

Table I gives simplified formulas which define the important line characteristics in terms of the primary line constants, the formulas

TABLE I
*Approximate Line Formulas*

| Line Characteristics | Uniform Line Having Zero Inductance | Uniform Line Having Distributed Inductance | |
|---|---|---|---|
| $\alpha$, Attenuation constant | $\sqrt{\frac{pRC}{2}}$ | $\sqrt{\frac{R}{2Lp}}\cdot\sqrt{\frac{pRC}{2}}=\frac{R}{2}\sqrt{\frac{C}{L}}$ | (1) |
| $W$, velocity of wave propagation | $\sqrt{\frac{2p}{RC}}$ | $\sqrt{\frac{R}{2Lp}}\cdot\sqrt{\frac{2p}{RC}}=\sqrt{\frac{1}{CL}}$ | (2) |
| $Z_0$, characteristic impedance | $\sqrt{\frac{R}{pC}}\,\underline{/45^\circ}.$ | $\sqrt{\frac{Lp}{R}}\,\underline{/45^\circ}\cdot\sqrt{\frac{R}{pC}}\,\underline{/45^\circ}=\sqrt{\frac{L}{C}}$ | (3) |

In the above, $\alpha$ is the real part of the propagation constant; and $W=p/\beta$, in which $p=2\pi f$($f$=frequency) and $\beta$ is the wave length constant; *i. e.*, the imaginary part of the propagation constant. The formulas assume the leakage conductance $G$ to be negligibly small; and in the case of the line with inductance, that $R$ is small with reference to $pL$; $R$, $L$, and $C$ being the line resistance, inductance, and capacitance per unit length.

[3] This section on Theory contains a small amount of discussion not included in the paper as presented.

being so arranged as to indicate directly the nature of the changes which occur when uniformly distributed inductance is added to a uniform line initially having zero inductance.

Inspection of the formulas shows that the addition of distributed inductance:

(a) Reduces the attenuation constant and the velocity, provided that the ratio $R/2L$ is less than $p$; in practice, this limiting condition is approached only at very low frequencies which usually are of negligible importance in speech transmission.

(b) Increases the impedance, and improves the power factor.

(c) Makes the attenuation, velocity and impedance independent of frequency over the frequency range where $R$ is small with reference to $pL$; in practice, this condition holds generally, except at the low voice frequencies.

From the standpoint of the power transmission engineer, the general effect of loading in reducing the attenuation losses may be explained in terms of the changes in line impedance noted in (b) above. These impedance changes make it possible for the loaded line to transmit a given amount of power corresponding to speech sounds at a higher line potential and with a (proportionately) lower value of line current than is possible without the loading. In the non-loaded line which is inherently a low impedance line, the series dissipation losses which are proportional to the square of the line current are ordinarily very large relative to the shunt dissipation losses which are proportional to the square of the line potential. Consequently, when the line impedance is increased by a suitable amount, the reduction in series losses is much greater than the increase in shunt losses and a substantial improvement in line efficiency is obtained. The optimum impedance for minimum line losses is that which results in the shunt and series losses being equal. Ordinarily, it is not economical to apply a sufficient amount of loading to reach this condition.

In general, commercial power lines are electrically short in terms of the wave length of the transmitted frequencies and consequently the sending end impedance is very largely influenced by the receiving end impedance. This allows high impedance transmission lines to be obtained by using high ratio transformers at the receiving end to step up the terminal impedance. On the other hand, telephone lines which are of interest from the loading standpoint are electrically long and the sending end impedance is practically unaffected by the terminal impedance. Consequently, the addition of series inductance to the line is the most practical way of increasing the telephone line impedance.

Investigating the question of concentrating the line inductance at uniformly spaced intervals, Professor Pupin gave his famous solution in a paper[4] presented before the Institute in May, 1900. Dr. G. A. Campbell in his paper[5] of March, 1903, also gave a mathematical development of the loading theory along somewhat different lines.

These early investigations showed that a coil loaded line should have several coils per wave length in order to simulate a uniform line. The more closely the coils are spaced the more exact is the degree of equivalence, and when there are ten coils per wave length the equivalence is very close. On the other hand, the cost of the loading increases as the spacing is shortened. Thus, from the standpoint of commercial application, the question "What is the smallest number of coils per wave length that will give satisfactory transmission?" is very important. In the investigation which was made to determine the magnitude of the changes in attenuation, velocity and impedance, as the number of coils per wave length is reduced, abrupt changes in these characteristics were found to occur at the spacing of two coils per actual wave length. The critical frequency at which this spacing applies in a loaded line became known as the cutoff frequency, since at this frequency and higher frequencies the attenuation loss is so extremely large as to amount practically to a suppression, or cut-off effect.

At the cut-off freqeuncy the velocity of the coil loaded line is lower than the velocity of the corresponding smooth line approximately in the ratio of $2:\pi$; consequently, at the cut-off frequency there are approximately $\pi$ coils per wave length, in terms of the velocity of the corresponding smooth line.

The following expression defines the cut-off frequency in a coil loaded line having zero distributed inductance:

$$f_c=\frac{1}{\pi\sqrt{LsC}} \tag{4}$$

in which

$f_c$ = cut-off frequency,
$L$ = coil inductance,
$s$ = coil spacing,
$C$ = line capacitance per unit length.

[4] "Wave Transmission over Non-Uniform Cables and Long Distance Air Lines," M. I. Pupin, Trans. A. I. E. E., Vol. 17, 1900, p. 445. Refer also to Pupin, U. S. Patents Nos. 652, 230 and 652, 231, June 19, 1900.

[5] "On Loaded Lines in Telephone Transmission," G. A. Campbell, *Philosophical Magazine*, March, 1903.

[If the loaded line has distributed inductance, a correction is required in equation (4).]

The differences between the characteristics of a coil loaded line and its corresponding smooth line are sometimes designated "lumpiness" effects. They are due to repeated internal reflections at the points of electrical discontinuity in the line caused by the insertion of the loading coils. The lumpiness effects are usually small for the frequencies below approximately 75 per cent. of the cut-off frequency. As the frequency exceeds this value, however, the lumpiness effects increase at an accelerated rate.

Figs. 1, 2 and 3 illustrate the differences in the attenuation, velocity, and impedance characteristics of a typical telephone cable, with and without loading. The characteristics of the corresponding smooth loaded line are also given to illustrate the theoretical differences between uniform loading and coil loading. Fig. 1 includes curves

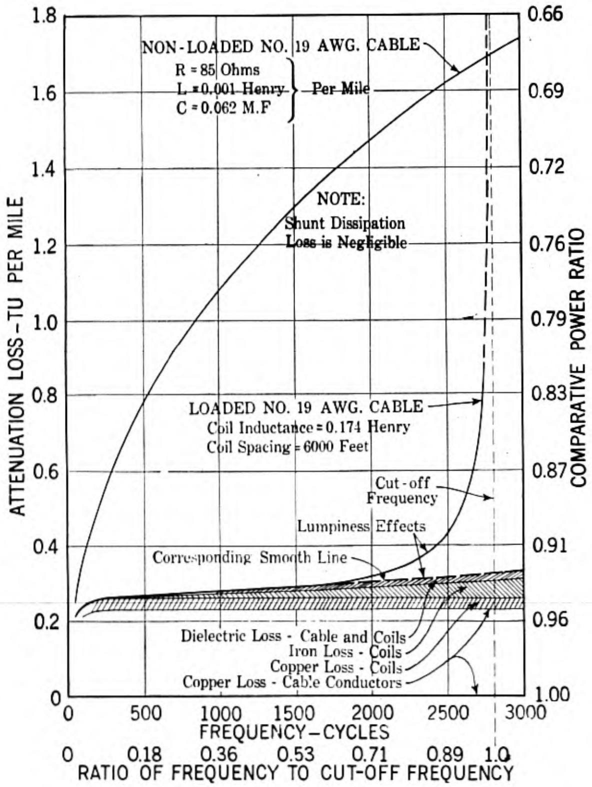



Fig. 1—Attenuation-frequency characteristics of loaded and non-loaded No. 19 A. w. g. cable

which give an analysis of the different types of line losses, (*a*) the "series" losses due to heat dissipation in the conductor and the loading coils, which are proportional to the square of the line current, (*b*) the "shunt" losses due to heat losses in the dielectrics, which are proportional to the square of the line voltage, and (*c*) the lumpiness effects due to internal reflections. The large reduction in the series losses

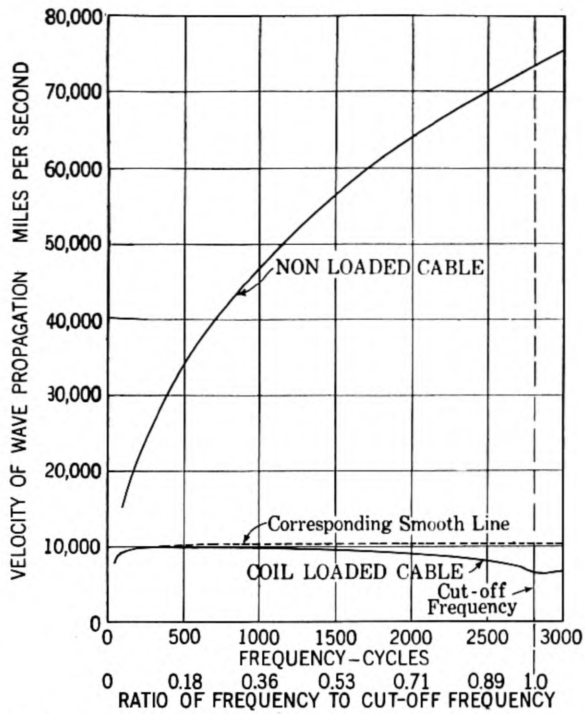



Fig. 2—Velocity-frequency characteristics of loaded and non-loaded No. 19 A. w. g. cables of Fig. 1

accomplished by the loading is clearly indicated in the diagram. A corresponding proportional increase in the shunt dissipation loss also occurs, but as previously noted this effect is small in absolute magnitude relative to the decrease in the series losses. It is interesting to note that the particular type of loading illustrated in Fig. 1 so increases the transmission efficiency of No. 19 A.W.G. cable that the loaded circuit can be used for distances about four times the permissible length of the non-loaded circuits. To obtain this increased transmission range without loading would require wires about eight times as heavy, i.e., No. 10 A.W.G.

Fig. 3 illustrates the dependency of the characteristic impedance of a coil loaded line upon the terminal condition. The most frequently used loading terminations are "mid-section" and "mid-coil." In the mid-section termination, the first loading coil is located at a dis-

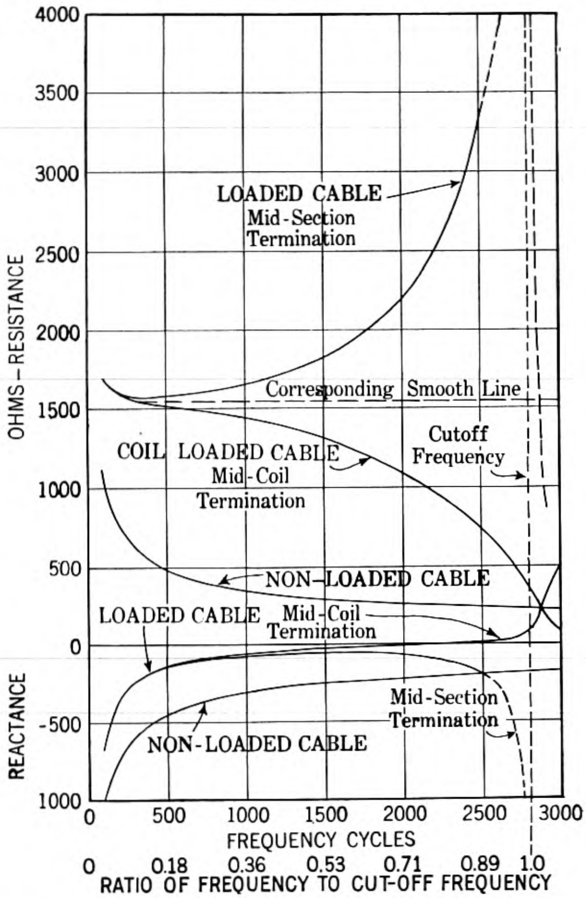



Fig. 3—Impedance-frequency characteristics of loaded and non-loaded No. 19 A. w. g. cables of Fig. 1

tance equivalent to one-half of a regular loading section from the beginning of the line. Mid-coil termination is obtained by installing at the beginning of the line, a coil having one half of the inductance of the regular coils, the first full coil being installed at the end of the first complete loading section. For mid-coil and mid-section terminations, the characteristic impedance is approximately a pure resistance, which varies with frequency as a complicated function of the ratio of the frequency to the cut-off frequency. With mid-coil termination the impedance-frequency characteristic droops with rising frequency, approaching zero at the cut-off frequency. On the other hand, the mid-section termination has a rising characteristic, approaching infinity at the cut-off frequency.

*Early Standard Loading Systems.* One of the fundamental questions involved in the early commercial development work was that of determining what range of frequencies should be transmitted in order to furnish a satisfactory grade of speech transmission. The investigation of this point resulted in the adoption of a standard cut-off frequency of about 2,300 cycles. Table II lists the other important transmission characteristics of the first loading systems standardized about 1904 for use on cables:—

TABLE II

*First Standard Cable Loading Systems*

| Loading Designation | Coil Inductance (Henrys) | Coil Spacing (Miles) | Inductance per Mile (Henrys) | Nominal Impedance (Ohms) | Attenuation Loss (TU per mile) | | |
|---|---|---|---|---|---|---|---|
| | | | | | 19 A.w.g. | 16 A.w.g. | 14 A.w.g. |
| Heavy | 0.250 | 1.25 | 0.200 | 1800 | 0.28 | 0.16 | 0.11 |
| Medium | 0.175 | 1.75 | 0.100 | 1300 | 0.39 | 0.21 | 0.14 |
| Light | 0.135 | 2.5 | 0.054 | 900 | 0.51 | 0.27 | 0.17 |
| | | (Non-loaded Cable) | | | 1.05 | 0.74 | 0.59 |

NOTE. These data apply to cables having a mutual capacitance of 0.070 $\mu f.$ per mile and assume loading coils, the electrical characteristics of which are given in Table IV. The nominal impedance is defined by the expression $\sqrt{L/C}$. The new unit of transmission loss (TU) is described in a recent Institute paper.[6]

For open wire loading, only one loading system, known as "Heavy" loading, was standardized. This involved the use of coils having an inductance of 0.265 henry at a spacing of approximately 8 miles. This loading had approximately the same cut-off frequency as the cable loading standards described in Table II. The other important line and transmission characteristics are summarized in Table III.

*Loading Coils.* The loading coils developed for use in the loading systems described in Tables II and III were of the toroidal type; i.e., they had ring-shaped cores formed by winding up a bundle of insu-

[6] "The Transmission Unit and Telephone Transmission Reference Systems," W. H. Martin, Trans. A. I. E. E., Vol. 43, 1924, p. 797; *Bell System Technical Journal*, July, 1924.

TABLE III

*First Standard Open Wire Loading*

| Wire Diameter (In.) | Loading Condition | Constants per Loop Mile | | | Nominal Impedance Ohms | Attenuation Loss TU per Mile |
|---|---|---|---|---|---|---|
| | | *R* Ohms | *L* Henrys | *C* Mf. | | |
| 0.104 | Non-loaded | 10.4 | 0.0037 | 0.0084 | 660 | 0.075 |
| 0.104 | Loaded | 11.1 | 0.037 | 0.0086 | 2100 | 0.031 |
| 0.165 | Non-loaded | 4.14 | 0.0034 | 0.0091 | 610 | 0.033 |
| 0.165 | Loaded | 4.8 | 0.037 | 0.0094 | 2000 | 0.014 |

NOTE. Transmission efficiency figures assume dry weather insulation conditions, 5 megohm-miles, or better.

lated fine wires on a suitably shaped spool. The core wire was 38 A. w. g. (0.004 in. diameter).

The wire used in the cable loading coil cores was a commercial grade of mild steel, hard drawn under conditions which gave it an initial permeability of 95. The term "initial permeability" signifies the permeability at very weak magnetizing forces; i.e., below 0.1 gilbert per cm. The core wire used in the open wire loading coils was drawn from the same stock, but differences in the drawing and annealing treatments gave it an initial permeability of about 65. This core wire had lower eddy current and hysteresis losses than the 95-permeability wire. A black enamel insulation was used on the 95-permeability wire. A celluloid-shellac compound which could be applied at a lower temperature was used on the 65-permeability wire.

As illustrating the magnitudes involved, it may be noted that in order to meet the service requirements, the coils were designed so that for telephone currents of the order of 0.001 ampere, the magnetizing force $H$ has a value of about 0.04 gilbert per cm., corresponding to a flux density of approximately $B=2$ gauss.

The winding space on the cores was divided in half by means of fiber washers, and the winding was applied in two equal sections, one being located on each half of the core. In installing the coils, one of these windings was inserted in one line wire and the other winding in the other line wire, so connected that the mutual inductance between windings aided the self-inductance for current flowing around the circuit through both windings.

The high costs of the open wire lines warranted considerable refinement in the design of the open wire coils. They were, therefore, made much more efficient and correspondingly larger than the cable coils. They were wound with insulated stranded wire and had much lower core losses. Another important difference between the open wire and cable coils was the use of high dielectric strength insulation in the open wire coils. The coils were subjected to a breakdown test at 8,000 volts (effective a-c.) and were protected in service by means of a special type of lightning arrester having non-arcing metal electrodes designed to operate at 3,500 volts direct current.

Table IV lists the principal characteristics of the loading coils

TABLE IV

*First Standard Loading Coils*

| Type Loading | Coil Code No. | Inductance (Henrys) | Average Resistance | | Overall Dimensions | |
|---|---|---|---|---|---|---|
| | | | D-C. (Ohms) | 1000-Cycle (Ohms) | Diameter (In.) | Height (In.) |
| Open Wire | 501 | 0.265 | 2.5 | 5.9 | 9 | 4 |
| Cable | 506 | 0.250 | 6.4 | 22.3 | 4⅛ | 3¼ |
| " | 508 | 0.175 | 4.2 | 13.0 | " | " |
| " | 507 | 0.135 | 3.2 | 9.1 | " | " |

NOTE. Effective resistance values apply for a line current of 0.002 ampere.

initially used in the standard loading systems listed in Tables II and III.

*Loading Coil Cases.* The cases used for potting the cable loading coils were designed so that they could be installed in underground manholes or on pole fixtures.

The general method of assembly is to dry the loading coils thoroughly and then impregnate them under vacuum with a moisture-proofing compound. The coils are then mounted on wooden spindles, adjacent coils being separated by iron washers. After carefully adjusting the individual coils to meet the electrical requirements, the spindles of coils are cabled to a short length of lead-covered cable which is referred to as a "stub" cable. Cast-iron cases with iron partitions were designed so as to provide a shielded compartment for each spindle of coils.

Commercially manufactured toroidal coils may have small irregularities in their windings resulting in a weak stray field which tends to cause cross-talk. The iron washers between coils and the partitions between spindle groups of coils provide effective cross-talk shields.

After placing the spindles of coils in the various compartments, the case is filled with a moisture proofing compound. The lead-sheathed cable stub is brought through a brass nipple in the cast iron cover of the case, and the cover is then bolted to the case. By means of a special design of case and cover joint, a double seal is provided to prevent entrance of moisture at this point. A wiped joint is made between the lead sheath of the cable and the brass nipple.

The conductors in the stub cable have an appropriate color scheme in their insulation to identify the terminals of the loading coils, thus facilitating splicing of the coils into the line circuits. A series of multi-spindle cases was standardized, ranging in capacity from 21 to 98 coils. Smaller quantities of coils were potted in a single spindle pipe type case.

Generally similar assembly and potting methods were used for the open wire coils, the important differences being first, that the open wire coils were always mounted in individual cases designed for mounting on pole fixtures, and secondly, that the coil terminals were brought out of the case in individual rubber-insulated leads.

## I. Phantom Group Loading

In Mr. Gherardi's paper reference was made to the development of means for (*a*) phantoming loaded circuits and (*b*) loading phantom circuits. The large plant economies made possible by these developments have resulted in extensive applications of these principles.

The following discussion will consider first the coil winding schemes, after which the transmission characteristics of the loading systems and the electrical characteristics of the loading coils will be briefly described.

*Loading Methods.* Fig. 4 schematically illustrates the Bell System standard method for loading phantom circuits and side circuits of phantoms.[7]

The loading problem is to introduce the desired inductance into each of the three circuits of a phantom group without causing objectionable unbalances. The method illustrated in Fig. 4 involves individual loading coils for each circuit, the design being such that the side circuit coils are substantially non-inductive to the phantom circuit, while the phantom loading coil is substantially non-inductive to the side circuits. These desirable results require close magnetic coupling between the line windings in each coil. Consequently, in

[7] U.S. Patents No. 980,021 "Loaded Phantom Circuit," G. A. Campbell and T. Shaw. No. 981,015 "Phantomed Loaded Circuit," T. Shaw.

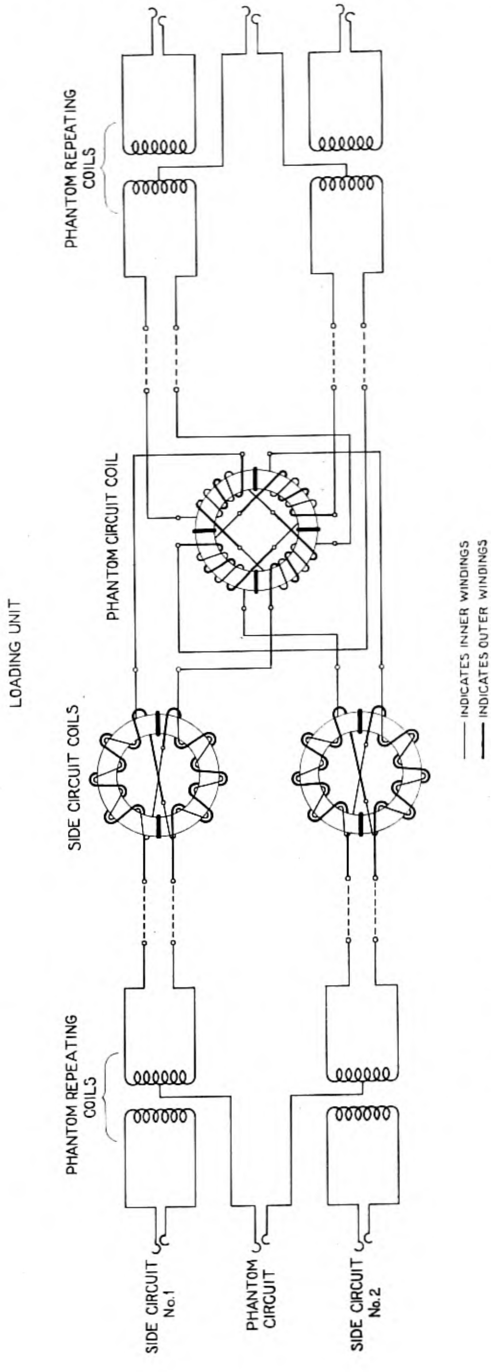



Fig. 4—Bell system standard method of loading phantom circuits and their side circuits

the side circuit coils each line winding is, in effect, distributed evenly about the entire core. The necessary high degree of symmetry required by balance considerations is obtained by dividing each line winding into two equal sections and interleaving them with the sections of the other line winding; thus each complete line winding consists of an inner section winding on one-half of the core and an outer section winding on the opposite half core. Similar design principles are applied to the phantom loading coils, with added complications, however, arising from the increased number of line windings. Each of the four line windings consists of an inner section winding located on one core quadrant and an outer section winding located on the opposite core quadrant, the two line windings associated with a given side circuit being distributed about the same pair of opposite core quadrants. In arranging the windings on the core, precautions are taken to secure a symmetrical arrangement of the direct admittances among the line windings and from the line windings to the core and the case.

It is interesting to note that the three-coil loading scheme illustrated in Fig. 4 was employed in the Boston-Neponset cable, installed in 1910, which was the first successful installation of loaded phantom circuits in the world. Other schemes of phantom group loading using two-coil and four-coil arrangements have been developed here and abroad, but none of them is considered to be as satisfactory as the scheme illustrated in Fig. 4 from the standpoint of service and cost. These other schemes are described in a recent article [8] which compares them with the scheme above described.

*Loading Systems.* In adapting the circuits to phantom working, the electrical constants of the two-wire circuits were changed as little as possible in making them suitable for use as side circuits of phantoms. In the cables, two pairs having different lengths of twist were twisted into quad formation on a still different length of twist. The required balance was obtained on open wire lines by cutting in a large number of additional transpositions.

The construction methods chosen resulted in the phantom circuits having approximately 60 per cent. greater distributed capacity than their side circuits, and a lower distributed inductance, approximately in inverse proportion. It was obviously desirable to install the phantom circuit coils at the same points as the side circuit coils; accordingly, in working to the same standard of cut-off frequency, the relative circuit constants summarized above resulted in the phantom

[8] "Commercial Loading of Telephone Cable," W. Fondiller, Electrical Communication, July 1925.

loading systems having a nominal impedance approximately 60 per cent. as high as their associated side circuit loading systems. The transmission efficiency of the phantom circuit was 20 to 25 per cent. better than that of its associated side circuits, on which basis, the phantom circuits were suitable for use over somewhat longer distances than their side circuits.

*Cable Loading.* Data regarding the general characteristics of the first phantom group loading systems standardized for use on quadded telephone cables are given in the first four rows of Table V. These loading systems were used principally on interurban toll cables. Because of the extra cost of the terminal and signaling equipment, and other factors involved in phantom working, it was not economical to use phantom circuits in the shorter lengths of loaded cable ordinarily involved in exchange area connections.

As soon as the development work on quadded toll cables and phantom group loading had progressed to a point where satisfactory commercial results were assured, active development work commenced on the Boston-New York-Washington cable project, involving the use of coarse gage quadded conductors and new types of high efficiency loading coils designed especially for use on the coarse gage wires. The Boston-Washington cable was the first link in a rapidly growing network of toll cables which now interconnects the large population centers of the Atlantic Seaboard and the upper Mississippi Valley region, providing increased reliability of service as compared with open wire lines.

It should be kept in mind that at the time under discussion (1910–1911) no commercially satisfactory type of telephone repeater was available. Accordingly, in order to assure satisfactory service between Boston, New York, Washington, and intermediate points, it was necessary to provide 10-A.w.g. and 13-A.w.g. conductors in the new cable. Cost studies showed it to be desirable to use a special weight of loading intermediate between the old heavy and medium loading systems, which was therefore designated "Medium-heavy" loading. Information regarding this special loading is given in Items 1 and 2 of Table V. In items 3 and 4, corresponding data are given on the "high-efficiency" heavy loading designed for coarse gage conductors. This heavy loading was used on certain sections of the Boston-Washington cable where plant construction reasons made it desirable to install the coils in existing loading manholes installed at heavy loading spacing.

From the last column of Table V it is seen that there is very little difference between the efficiencies of the heavy and the medium-

TABLE V

*First Loading Standards for Quadded Toll Cables*

| Item | Loading Designation | Type Circuit | Coil Inductance (Henrys) | Coil Spacing (Miles) | Nominal Impedance (Ohms) | Attenuation Loss—TU per mile | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 19 A. w. g. | 16 A. w. g. | 13 A. w. g. | 10 A. w. g. |
| 1 | Medium-Heavy | Side | 0.210 | 1.4 | 1500 | | | 0.085 | 0.050 |
| 2 | " " | Phantom | 0.130 | 1.4 | 950 | | | 0.069 | 0.040 |
| 3 | Heavy | Side | 0.250 | 1.25 | 1850 | | | 0.081 | 0.050 |
| 4 | " | Phantom | 0.155 | 1.25 | 1150 | | | 0.066 | 0.042 |
| 5 | Heavy | Side | 0.250 | 1.25 | 1850 | 0.24 | 0.14 | | |
| 6 | " | Phantom | 0.155 | 1.25 | 1150 | 0.20 | 0.12 | | |
| 7 | Medium | Side | 0.175 | 1.75 | 1300 | 0.31 | 0.17 | | |
| 8 | " | Phantom | 0.106 | 1.75 | 800 | 0.26 | 0.14 | | |

NOTES: A capacitance of 0.062 $\mu f$. per mile is assumed in side circuits and 0.100 $\mu f$. per mile in the phantom circuit. The pair capacitance value is smaller than that assumed in Table II, due to improvement in the cables.

All of the above loading systems have a cut-off frequency of about 2300 cycles.

heavy loading systems when used on 10-A.w.g. conductors. This explains the more general use of the medium-heavy loading, which was less expensive because of the greater distances between coils. The effects under discussion are due to the part played by the loading coil resistance. The loading coils themselves conformed as closely as practicable to the cost-equilibrium principle:—a condition of cost balance where a small improvement in transmission would require approximately equal expenditure whether by improving the loading or by adding copper to the cable conductors. On this basis, a somewhat less expensive grade of coil was used on the 13-A.w.g. wires than on the 10-A.w.g. wires. The grade of coils developed primarily for use on 16 and 19-A.w.g. cables, giving transmission results illustrated in Items 5–8 of Table V, was in turn less expensive than the "high efficiency" coils. In each case, since the phantom circuits were somewhat more efficient than their associated side circuits, a somewhat higher grade coil was used in the phantom circuits than in the side circuits.

*Open Wire Phantom Loading.* Phantom loading came into general use on open wire lines at about the same time as on quadded cables. In general, the methods used in applying phantom group loading to the open wire lines were used for the cable systems. The line characteristics for the side circuits were practically the same as for the original non-phantomed circuits (Table III); the principal difference being that caused by the small resistance of the phantom loading coils. The important linear and transmission characteristics of the phantom circuits are given in Table VI. The phantom loading coil had an inductance value of 0.163 henry.

*Loading Coils.* Table VII gives general information regarding the first standard side circuit and phantom loading coils used in the phantom group loading systems listed in Tables V and VI. The coils

TABLE VI

*First Standard Open Wire Phantom Loading*

| Wire Diameter (In.) | Loading Condition | Constants per Loop Mile at 1000 Cycles | | | Nominal Impedance (Ohms) | Attenuation Loss TU per Mile |
|---|---|---|---|---|---|---|
| | | *R* (Ohms) | *L* (Henrys) | *C* (Mf.) | | |
| 0.104 | Non-loaded | 5.2 | 0.0022 | 0.0141 | 400 | 0.064 |
| 0.104 | Loaded | 5.8 | 0.023 | 0.0141 | 1300 | 0.027 |
| 0.165 | Non-loaded | 2.1 | 0.0021 | 0.0154 | 400 | 0.028 |
| 0.165 | Loaded | 2.6 | 0.023 | 0.0154 | 1200 | 0.012 |

TABLE VII
*First Standard Loading Coils for Phantom Working*

| Type Line | Inductance (Henrys) | Coil Code No. | Type Circuit | Average Resistance-Ohms D-C. | Average Resistance-Ohms 1000 cycles | Overall Dimensions Diameter (In.) | Overall Dimensions Height (In.) |
|---|---|---|---|---|---|---|---|
| Open-Wire | 0.265 | 512 | Side | 5.0 | 8.4 | 9.0 | 4.0 |
| | 0.163 | 511 | Phantom | 2.5 | 4.4 | 11.0 | 4.9 |
| 10-A. w. g. Cable | 0.210 | 520 | Side | 3.8 | 6.6 | 8.5 | 3.5 |
| | 0.130 | 519 | Phantom | 1.9 | 3.4 | 10.4 | 4.0 |
| | 0.250 | 532 | Side | 4.1 | 7.8 | 8.5 | 3.5 |
| | 0.155 | 531 | Phantom | 2.1 | 3.9 | 10.4 | 4.0 |
| 13-A. w. g. Cable | 0.205 | 538 | Side | 6.0 | 9.2 | 5.7 | 2.5 |
| | 0.130 | 521 | Phantom | 3.0 | 4.5 | 7.9 | 3.0 |
| | 0.250 | 534 | Side | 6.6 | 10.7 | 5.7 | 2.5 |
| | 0.155 | 533 | Phantom | 3.3 | 5.3 | 7.9 | 3.0 |
| 16 and 19-A. w. g. Cable | 0.250 | 515 | Side | 8.9 | 23.1 | 4.6 | 2.4 |
| | 0.155 | 530 | Phantom | 4.4 | 11.9 | 5.9 | 2.9 |
| | 0.175 | 514 | Side | 5.4 | 14.4 | 4.6 | 2.4 |
| | 0.106 | 513 | Phantom | 2.7 | 7.1 | 5.9 | 2.9 |

NOTE. The resistance data apply to circuits of a complete phantom group; *i.e.*, the side circuit data include effects of the phantom coils, and phantom circuit data include effects of the side circuit coils. Effective resistance values correspond to line current of 0.002 ampere.

designed for open wire lines and for 10-A.w.g. cable had 65-permeability wire cores and stranded copper windings. The coils designed for 13-A.w.g. cables had 65-permeability wire cores and non-stranded copper windings. The other coils had 95-permeability wire cores.

*Potting Features.* The general practise for cable loading is to pot side circuit and phantom loading coils in the same case as phantom groups, since this has important installation and transmission advantages. The phantom coils, being considerably larger than the side circuit coils, are mounted in separate spindle compartments. The cross-connections between the side circuit and phantom coils are made within the case, in order to reduce the amount of splicing required in the field. Thus, the stub cable contains only the conductors to be spliced to the "east" and "west" conductors in the line cable. Quadded construction is used in the stub cable of all loading coil cases for phantom loading in order to avoid serious capacitance unbalances.

The multi-spindle cases used in potting the small size coils for 16 and 19-A.w.g. cables ranged in capacity from 12 to 24 phantom units. The larger size coils used on the coarser gage cables were potted in smaller complements.

Occasionally it is desirable to install side circuit loading alone and to install the phantom loading at a later period. Accordingly, cable loading coil cases were designed to meet these conditions. The open wire coils were potted in individual cases.

## II. Loading for Repeatered Circuits

*General.* The development of telephone repeaters to the point where they could be used for commercial service in extending the range of telephone transmission was the beginning of a new era in the communication art. In this development work, the adaptation of the lines to the requirements of repeater operation was secondary in importance only to the development of satisfactory repeater elements and circuits for associating the repeater elements with the line. The reader is referred to an Institute paper by Messrs. B. Gherardi and F. B. Jewett [9] for general information regarding telephone repeaters and to a more recent Institute paper by Mr. A. B. Clark [10] for a general discussion of subsequent developments in the application of repeaters to long telephone circuits.

The early work on the line problem was primarily concerned with obtaining a sufficiently high degree of regularity in the line impedance-frequency characteristics, so that the requisite high degree of balance could be obtained and maintained between the line and the repeater balancing network. Later on, particularly in preparing for the application of telephone repeaters to long toll cables, such as the New York-Pittsburgh-Chicago cable, it became necessary to change the fundamental transmission characteristics of the loading.

*Early Work—Reduction of Line Irregularities.* Commercial telephony, requiring two-way transmission, imposes severe balance requirements on repeater circuits over the entire band of frequencies which the repeater is designed to transmit, in order to avoid singing or distortion due to near singing. Within certain limitations, the higher the degree of balance between the line and the balancing network circuit, the higher will be the permissible amplification gain of the repeater.

[9] "Telephone Repeaters," B. Gherardi and F. B. Jewett, Trans. A. I. E. E., Vol. 38, 1919, p. 1287.

[10] "Telephone Transmission over Long Cable Circuits," A. B. Clark, Trans. A. I. E. E., Vol. 42, 1923, p. 86; *Bell System Technical Journal*, Jan., 1923.

The practical solution of this fundamental repeater-line balance problem required (*a*) the construction of lines having extremely regular impedance characteristics over the frequency band which the repeater is designed to transmit and (*b*) the development of balancing networks [11] capable of accurately simulating the sending-end impedance characteristics of the improved lines throughout this frequency range. On account of the great difficulty of getting a high degree of balance at frequencies near the cut-off frequency of the loading, partly due to line irregularity effects and partly due to network design complications, it has been found desirable to use electric wave filters [12] in the repeater sets which cut off at a frequency below the cut-off frequency of the loading. This margin of cut-off effects is usually 200 cycles or more, depending upon the repeater design and the type of loading involved.

The "regular" line referred to in (*a*) is one which is free from impedance irregularities. In the case of loaded lines, the loading coils should have very closely the same inductance values, and the sections of line between loading coils should have closely the same value of capacitance. These uniformity features should be permanent, which requires that the coils should have a high degree of stability in their inductance characteristics; i.e., they should be capable of resisting the magnetizing effects of abnormal service conditions. Some of the older types of coils did not meet this requirement. The satisfactory way in which these fundamental coil requirements are fulfilled in the newer types of coils will be described in a subsequent section.

Uniformity in the loading section capacitance values involves uniformity in cable and line capacitance values as well as precision in the coil spacing. In toll cable loading the maximum deviations from the average spacing are kept below 2 per cent., and the average deviations are in the order of 0.5 per cent. or less.

In exceptional cases where physical obstructions are encountered in reducing the spacing deviations to a sufficiently low value, use is made of "building-out condensers" or "building-out stub cables" to normalize the capacitance of loading sections.[13] Abnormally long loading sections can usually be split up into two sections, one or both of which may then be "built out" to the nominal standard capacitance values.

*Transcontinental Lines—High Stability Loading Coils.* The in-

[11] R. S. Hoyt "Impedance of Loaded Lines and Design of Simulating and Compensating Networks," *Bell System Technical Journal*, July 1924.

[12] U. S. Patents Nos. 1,227,113, and 1,227,114—G. A. Campbell.

[13] U. S. Patent No. 1,219,760—John Mills and R. S. Hoyt.

auguration of commercial transcontinental telephone service over the New York-San Francisco line in January, 1915, marked the first commercial application of these general improvements in regularity of line construction, including the use of an improved type of loading coil.

In the extensive field work which was done in preparing for transcontinental telephone service, it was found that the inductance values of a considerable percentage of the open-wire loading coils then in use (Nos. 511 and 512 types, Table VII) had changed appreciably from the nominal values to which they were adjusted at the factory prior to shipment, and that these changes were due to core magnetization caused by abnormal currents induced by lightning discharges. In some cases abnormal currents induced by power transmission lines or electric railway distribution systems were responsible for the loading coil magnetization.

The inductance changes were not sufficiently large to have serious reactions on transmission over non-repeatered circuits. Although individual coils varied in inductance from time to time, the general average of groups of coils was fairly constant. The effects of these individual variations on the impedance of the line were, however, too large to permit satisfactory operation with telephone repeaters. Some experiments made with improved lightning arresters, in an effort to reduce the coil magnetization trouble, were unsuccessful.

The solution of the problem of repeatering loaded open-wire circuits required the development of loading coils which would be stable magnetically when subjected to extreme conditions of magnetizing current in the windings. The requirement was laid down for these coils that the inductance to speech currents should not be affected more than about 2 per cent. when a magnetizing current of two amperes was passed through either line winding. In view of the fact that the extreme residual magnetizing effect of this current on the No. 511 and No. 512 loading coils was approximately 30 per cent., it will be appreciated that this imposed a very severe stability requirement.

The design adopted involved the use of air-gaps in the cores of the iron wire core loading coils.[14] Two air-gaps were employed at opposite points in the cores and suitable clamping means were provided to hold the coil halves in proper alinement. The use of only two air-gaps in the cores of the phantom loading coil brought in unbalance tendencies not present in older designs, which were corrected by special refinements in the design.

The use of a magnetic circuit having "ends," while effective for producing self-demagnetization, brought in troublesome magnetic

[14] U. S. Patents Nos. 1,289,941 and 1,433,305—Shaw and Fondiller.

leakage which necessitated special potting methods. Because of the economy of cast-iron loading coil cases, it was decided to continue their use, but to increase their dimensions sufficiently to reduce eddy-current losses in the case to a tolerable point.

The air-gap type loading coils designed for the transcontinental circuits, coded Nos. 549 and 550 for the phantom and side circuits respectively, were more generally potted as phantom loading units than as individual coils, and in such instances the cross-connections between the phantom and side circuit coils were made inside the case. Important advantages of this arrangement were that the leakage losses during periods of low line insulation were greatly reduced as well as the liability of wrong connections of windings during the installation work. Fig. 5 is a photograph of an installation of open

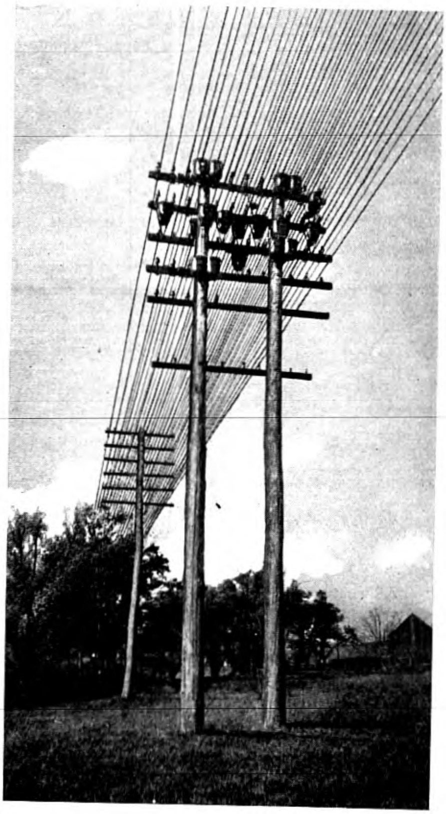


Fig. 5—Typical open wire loading installation
Showing four phantom group (3-coil) cases and nine individual coil cases

wire loading coils illustrating both the individual coil and loading unit methods of potting.

Table VIII contains data on the air-gap coils standardized for open-wire circuits. It will be noted that these coils are somewhat less efficient from the standpoint of effective resistance than the older type coils (Nos. 511 and 512) listed in Table VII, though having

TABLE VIII

*High Stability Coils Having Wire Cores with Air Gaps*

| Type Loading | Coil Code No. | Type Circuit | Inductance | Average Resistance Ohms | | Overall Dimension Inches | |
|---|---|---|---|---|---|---|---|
| | | | Henrys | D-C. | 1000-Cycles | Diameter | Hgt. |
| Open Wire | 550 | Side | 0.245 | 5.4 | 11.1 | 8.1 | 3.9 |
| " " | 549 | Phantom | 0.150 | 2.7 | 6.4 | 10.0 | 4.0 |
| 10 and 13 | 556 | Side | 0.248 | 7.0 | 14.0 | 5.6 | 2.9 |
| A. w. g. Cable | 555 | Phantom | 0.154 | 3.5 | 7.0 | 7.5 | 3.6 |
| 10 and 13 | 558 | Side | 0.200 | 6.2 | 10.9 | 5.6 | 2.9 |
| A. w. g. Cable | 557 | Phantom | 0.135 | 3.1 | 5.9 | 7.5 | 3.6 |

NOTES. Open-wire coils used in Loading Systems, Tables III and VI. Cable coils used in Loading Systems, Table V.

Resistance data apply to side circuits and phantom circuits of complete phantom groups. Effective resistance values are for 0.002 ampere line current.

marked superiority over the latter with regard to magnetic stability. To assist in getting maximum line regularity, the Nos. 549 and 550 coils were adjusted in the factory to meet ± 1 per cent. inductance precision limits. In the older types of coils ± 5 per cent. deviations had been allowed. The nominal inductance values of the Nos. 549 and 550 coils are somewhat below those of the Nos. 511 and 512 coils, the inductance difference corresponding roughly to the average magnetization effect of normal service conditions on the older types of coils.

The solution of the transcontinental line problem involved improvements in the regularity of the coil spacing as well as improvements in the magnetic stability of the coils. The line "clearing up" work usually involved a great deal of retransposing, since cross-talk considerations made it necessary to have the coils placed at balanced or neutral points in the transposition layout.

In the case of coarse gage cable circuits, such as the Boston-Washington and other toll cables installed prior to the advent of repeaters, the new requirements were met by the design of an air-gap type of wire-core coil on which data are given in Table VIII. They were somewhat smaller and not quite so expensive as the improved open-wire coils.

*Compressed Powdered Iron Core Loading Coils.* It soon became evident that the economical extension of the toll plant would involve the general introduction of telephone repeaters in cable as well as open-wire circuits. The use of telephone repeaters made it possible to supersede the coarse gage conductors by 16 and 19-A.w.g. conductors for toll connections, and this greatly increased the need for an efficient and stable loading coil of lower cost than the air-gap wire core coil.

As a result of investigations carried on over a period of several years, there was developed for commercial use early in 1916 a new magnetic material, compressed powdered iron, which has been of the utmost value in loading coil design.[15] This improved magnetic material is described in a paper presented before the Institute by B. Speed and G. W. Elmen [16] which also discusses the electrical and magnetic properties of the material.

Briefly, the method of production consists of grinding electrolytically deposited iron to the desired fineness, insulating the particles of iron, and finally compressing these insulated particles in steel dies at such very high pressures as to consolidate the mass into a ring, the specific gravity of which is substantially equal to that of solid iron. The rings are then stacked in a manner similar to laminations of sheet material to form a core of the desired dimensions. Though the separate rings are approximately 0.2 in. thick, the insulation between the individual particles is so effective that despite the use of molding pressures of 200,000 lb. per sq. in., the eddy current loss in a powdered iron core is less than that obtainable with 0.004 in. iron wire. Depending on the heat treatment and the amount of insulation, the initial permeability can be varied from approximately 25 to about 75. The specific resistance is about 20,000 times that of ordinary iron. The permeability can be controlled within comparatively narrow limits by the manufacturing processes, thus making for greater uniformity. The great advantage of this material for loading coils, however, lies in its self-demagnetizing property. The powdered iron core by virtue of its very numerous, though extremely small dis-

[15] U. S. Patents No. 1,274,952, B. Speed; 1,286,965, G. W. Elmen; 1,292,206, J. C. Woodruff.

[16] "Magnetic Properties of Compressed Powdered Iron," B. Speed and G. W. Elmen, Trans. A. I. E. E., Vol. 40, 1921, p. 1321.

tributed air-gaps, affords a means for constructing magnetically stable cores without the production of poles and their attendant magnetic leakage.

Fig. 6 gives photographs of a standard compressed iron powder core ring such as is used in the cores of toll cable loading coils; a

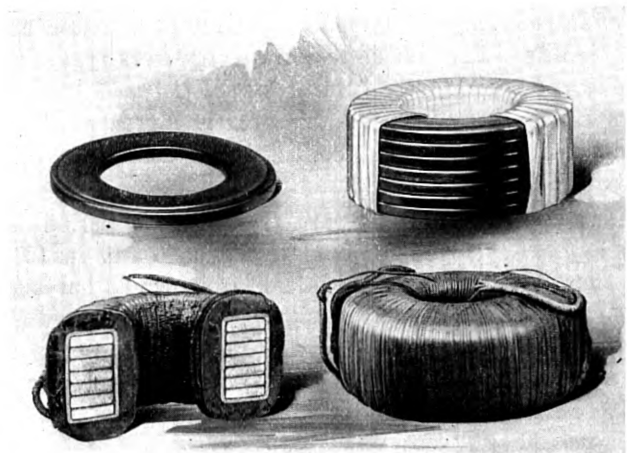


Fig. 6—Compressed powdered iron core loading coil

completely assembled core with part of the core taping removed; a completely wound coil of the side circuit type; and a coil in cross-section. Table IX gives general data regarding typical coils.

The first application of powdered iron cores was to replace some of the 95-permeability wire core loading coils in 16 and 19-A.w.g. cable. The effective permeability of the 95-permeability wire cores, making correction for air spaces and insulation, was approximately 60, and accordingly, the replacing powdered iron cores were designed to have the same effective permeability.

As a result of further developments in the direction of applying vacuum tube repeaters to loaded cable circuits, it became necessary with the extension of the length of these circuits to improve the characteristics of the loading coils. This led to the development of an improved grade of powdered iron core having an initial permeability of 35 which corresponds closely to the effective permeability of cores using iron wire having a permeability of 65. It was decided that for circuits such as interoffice trunks and short cables which would not be operated with superposed telegraph, the 60-permeability compressed iron core coils should be used; while for toll cable work involving repeatered composited circuits, 35-permeability cores should be employed. All of the compressed powder core coils intended for repeatered circuits were adjusted to meet ± 2 per cent. inductance limits.

The effective resistance-frequency characteristics of 95-permeability and 65-permeability wire core coils and 60-permeability and 35-permeability powdered iron core coils having the same inductance (0.174 henry) and the same over-all sizes are given in Fig. 7. The large improvement as to freedom from residual magnetization effects afforded by the 35-permeability powdered iron core, compared with the 65-permeability wire core is evident from the curves of Fig. 8. The effective resistance and inductance variation with current strength are shown in Fig. 9 for a 35-permeability powdered iron core coil. The remarkable property of these cores of maintaining constancy of permeability is shown by the change of only 1 per cent. in permeability as the current strength varies 400 per cent. from, say 0.001 to 0.005 ampere.

It is interesting to note that after the process had been fully worked out and production was running on a commercial scale, the cost of the improved cores was comparable with that of the wire cores which they replaced.

TABLE IX

*Typical Compressed Powdered Iron Core Loading Coils*

| Coil Code No. | Core Permeability | Inductance (Henrys) | Type Circuit | Resistance Ohms | | Dimensions Inches | |
|---|---|---|---|---|---|---|---|
| | | | | D-C. | 1000-Cycles | Diameter | Height |
| 562 | 60 | 0.245 | Side | 11.4 | 25.8 | 4.5 | 2.1 |
| 561 | 60 | 0.155 | Phantom | 5.7 | 11.7 | 6.3 | 3.0 |
| 564 | 60 | 0.174 | Side | 6.6 | 15.4 | 4.5 | 2.1 |
| 563 | 60 | 0.106 | Phantom | 3.3 | 6.7 | 6.3 | 3.0 |
| 582 | 35 | 0.245 | Side | 15.9 | 21.8 | 4.7 | 2.4 |
| 581 | 35 | 0.155 | Phantom | 8.0 | 10.0 | 6.7 | 3.1 |
| 584 | 35 | 0.174 | Side | 10.8 | 14.1 | 4.7 | 2.4 |
| 583 | 35 | 0.106 | Phantom | 5.4 | 6.6 | 6.7 | 3.1 |
| 584 | 35 | 0.174 | Side | 12.1 | 15.3 | 4.7 | 2.4 |
| 587 | 35 | 0.063 | Phantom | 6.1 | 7.0 | 4.7 | 2.8 |
| 590 | 35 | 0.044 | Side | 4.0 | 4.6 | 4.7 | 2.4 |
| 591 | 35 | 0.025 | Phantom | 2.0 | 2.0 | 4.7 | 2.8 |

NOTE. Resistance values apply to side circuits and phantom circuits of complete phantom groups. Effective resistance corresponds to 0.002-ampere line current.

These coils are used in the loading systems listed in Tables V and X.

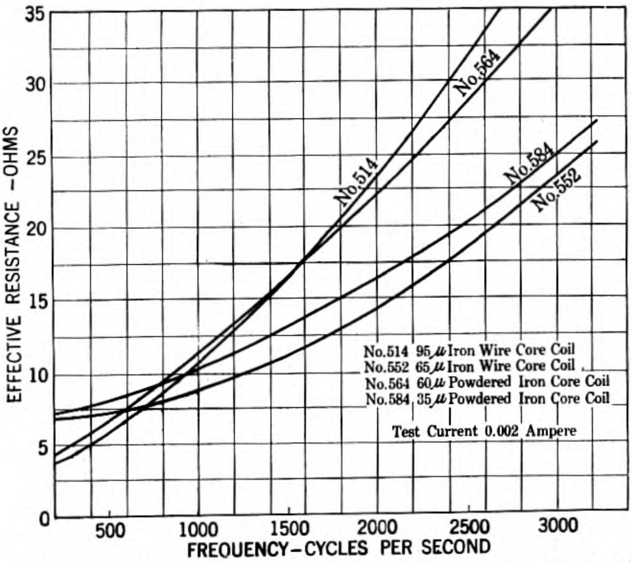



Fig. 7—Effective resistance-frequency characteristics toll cable loading coils

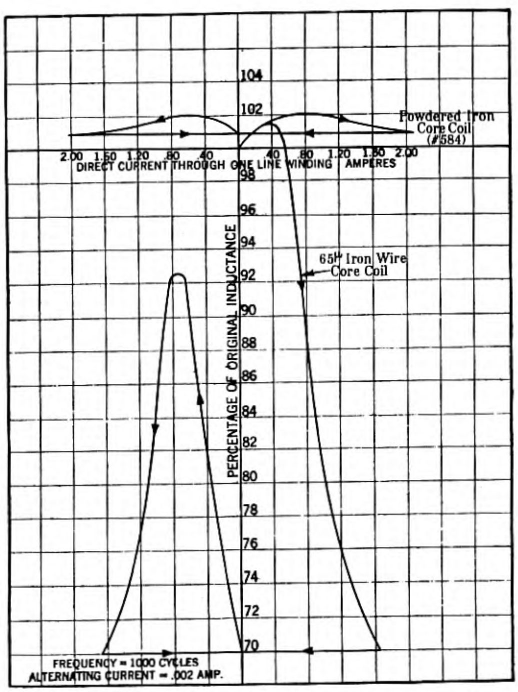



Fig. 8—Residual magnetization characteristics of compressed powdered iron core and iron wire core loading coils

In connection with the development of the new core material which was undertaken as a part of the loading coil development program, an enormous amount of work was involved which would not ordinarily be associated with loading coil design work. For instance, there were undertaken chemical studies on electro-deposition of iron

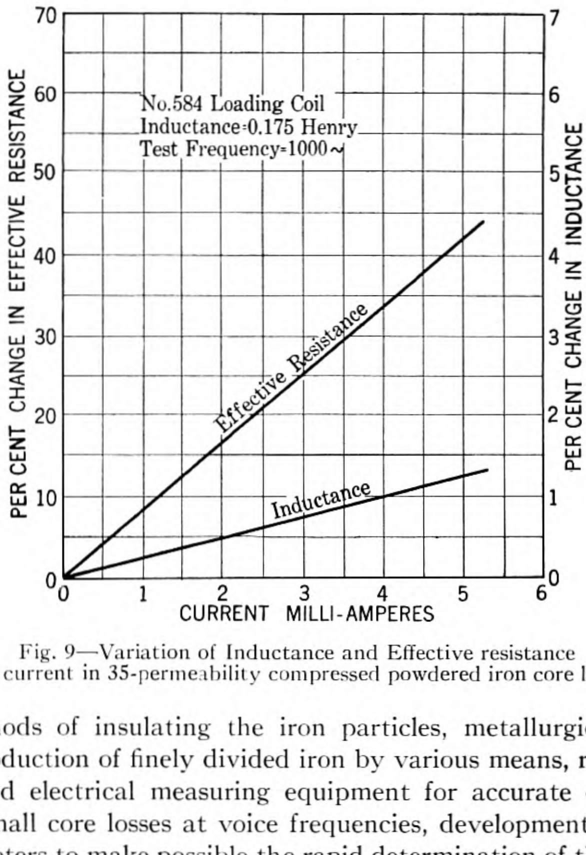



Fig. 9—Variation of Inductance and Effective resistance
With line current in 35-permeability compressed powdered iron core loading coil

and methods of insulating the iron particles, metallurgical studies of the production of finely divided iron by various means, refinements in shielded electrical measuring equipment for accurate determination of small core losses at voice frequencies, development of special permeameters to make possible the rapid determination of the permeability of rings, the design of the steel moulding dies, selection of suitable grades of alloy steel to withstand the enormous pressures, and also various other special problems. These are mentioned here as illustrative of the scope of the problem of developing this new core material.

It is of interest to note that the compressed powdered iron core loading coil has been adopted also as the international standard in Europe for repeatered circuits.[17]

[17] Minutes of Second Conference of Permanent Commission, Le Comité Consultatif Internationale de Communications Telephonique a Grande Distance, page 55—p. 119, English Version.

*New Requirements for Cable Loading Systems.* In the first commercial applications of telephone repeaters, the new features in the loading were the improved types of coils already described and the improved precision of spacing the coils. No fundamental changes were made in the loading systems then standard.

The completion of the development of a satisfactory commercial type of telephone repeater marked the beginning of a long period of experimental work for the purpose of determining the commercial possibilities of the use of repeaters over long cable circuits. When loaded cables of improved impedance regularity became available, long circuits were built up for experimental purposes by looping back and forth. As the length of these circuits was increased, phenomena not previously observed in cable circuits became increasingly troublesome, and it became apparent that it would be necessary to develop new loading systems having improved velocity and higher cut-off frequency characteristics in order to realize the full possibilities of repeaters in extending the range and reducing the cost of long distance telephone service over cables.

The disturbances above mentioned were found to be due to:

(*a*) Echo effects.
(*b*) Velocity distortion.

These phenomena originate in the lines themselves and are made more apparent by the amplifying action of the repeaters. They are present in non-repeatered circuits but not to a noticeable degree. It is the combination of the extreme length of the circuit and the use of repeaters to keep the over-all loss low that makes the disturbances troublesome.

*Echoes.* Echoes are due to unbalance currents; i.e., to the reflection of electrical energy at points of impedance irregularity in the circuits. When the circuit is so long that the time of transmission from the point of reflection to the disturbed subscriber is appreciable, there will be echo effects unless the losses in the circuit are so large as to cause the reflected energy to become inappreciably small. On such circuits it may be necessary to work the repeaters at gains well below those at which "singing" occurs or distortion due to "near singing" is experienced.

Since the time of transmission is such an important factor in echo phenomena, reductions in the harmful effects of these disturbances have been obtained in the improved loading systems which have been developed for use on long repeatered circuits, by substantially increasing the velocity of transmission. Recently there has become commercially available a device known as an "echo suppressor" which interrupts the path of the echoes without disturbing the main transmission. A description of the device and its field of application was given in a recent Institute paper.[18]

*Velocity Distortion.* In a coil loaded line the steady state velocity of wave propagation varies with frequency. At the upper frequencies the velocity change is principally due to lumpiness effects of the loading and is, therefore, a function of the ratio of the frequency under consideration to the cut-off frequency. As illustrated in Fig. 2, the departure of the actual velocity from the nominal velocity of the corresponding smooth line ($\sqrt{1/CL}$) increases as the frequency is raised, the rate of change increasing rapidly as the cut-off frequency is approached. At frequencies below approximately 0.3 of the cut-off frequency the coil loaded line has substantially the same velocity characteristics as the corresponding smooth line; when the frequency is further reduced, the departure of the actual velocity from the nominal velocity increases as a function of the ratio of the line resistance to the inductive reactance per unit length.

As a result of these velocity-frequency relations, a long loaded repeatered circuit may have seriously objectionable quality, even when the attenuation-frequency distortion is made negligible by the use of special devices at the repeater stations for correcting the attenuation-frequency distortion effects.

The velocity distortion is particularly noticeable during the building-up and dying-down periods, when it manifests itself as transient distortion. The duration of transient distortion depends, among other factors, upon the length of the line, the nominal velocity, and the cut-off frequency of the loading. In the old standard loading systems the high frequency velocity distortion caused by the lumpiness effects of the loading was more serious than the low frequency velocity distortion. Accordingly, a substantial reduction in the transient distortion has been obtained in the new standard loading systems by raising the cut-off frequency of the loading.

For further discussion of velocity distortion reference should be made to Mr. A. B. Clark's paper,[19] previously mentioned, which gives experimental results and to an earlier Institute paper by Mr. J. R. Carson [20] which gives the results of theoretical studies.

[18] "Echo Suppressors for Long Telephone Circuits," A. B. Clark and R. C. Mathes, *Jour. A. I. E. E.*, p. 618, June, 1925.

[19] Clark, Loc. Cit.

[20] "Theory of the Transient Oscillations of Electrical Networks and Transmission Systems," J. R. Carson, Trans. A. I. E. E., Vol. 38, 1919, p. 345.

*Characteristics of Improved Cable Loading Systems.* The principal electrical features of the H-44-25 and H-174-63 phantom group loading systems which have been developed primarily for use on long repeatered cables are given in Table X. Corresponding details of the older standard loading system developed for non-repeatered cables are also included in this table. Typical attenuation-frequency curves of the old and new loading systems are given in Fig. 10.

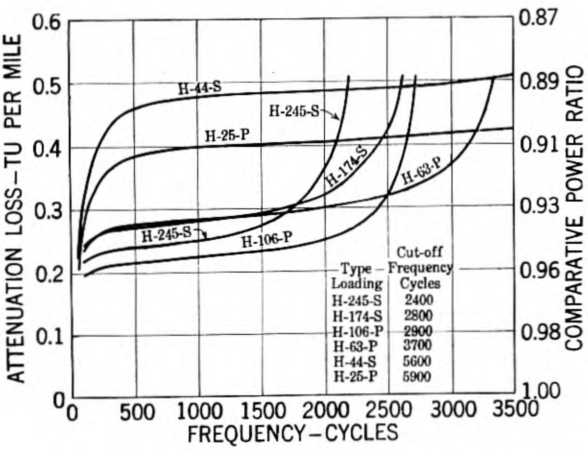



Fig. 10—Attenuation-frequency characteristics of toll cable loading

In the following discussion of detailed characteristics, the various phantom group loading systems will be referred to in terms of recently standardized designations which include a letter to symbolize the coil spacing, in combination with two numbers which correspond with the effective coil inductance values in milhenrys, the first number referring to the side circuit coils and the second number to the phantom coils. The individual side circuit or phantom circuit loading systems have designations which include a letter to symbolize the coil spacing, coupled with the inductance value of the loading coils in milhenrys, and having a letter suffix "*S*" or "*P*" indicating the type of circuit, side or phantom.

The fundamental differences between the new and the old loading systems are with respect to velocity of wave propagation and cut-off frequency, these changes having been made in accordance with the preceding discussion primarily for the purpose of reducing echo effects and transient distortion. For reasons of plant economy and

TABLE X

*Loading Systems—Small Gage Repeatered Toll Cables*

| Item | (*a*) Loading System | Circuit | (*b*) Coil Code No. | Nominal Impedance (Ohms) | Nominal Cut-off Frequency (Cycles) | Trans-mission Velocity Miles per Second | (*c*) Attenuation Loss TU per Mile at 1000 Cycles 19 A. w. g. | (*c*) Attenuation Loss TU per Mile at 1000 Cycles 16 A. w. g. | (*d*) Maximum Geographical Length (Miles) |
|---|---|---|---|---|---|---|---|---|---|
| (1) | H- 44-25 | Side | 590 | 800 | 5600 | 19000 | 0.48 | 0.25 | 5000 |
| (2) | " | Phantom | 591 | 450 | 5900 | 20000 | 0.40 | 0.21 | 5000 |
| (3) | H-174-63 | Side | 584 | 1550 | 2800 | 10000 | 0.28 | 0.16 | 500 |
| (4) | " | Phantom | 587 | 750 | 3700 | 13000 | 0.28 | 0.16 | 1500 |
| (5) | H-174-106 | Side | 584 | 1550 | 2800 | 10000 | 0.28 | 0.16 | 500 |
| (6) | " | Phantom | 583 | 950 | 2900 | 10000 | 0.22 | 0.13 | 500 |
| (7) | H-245-155 | Side | 582 | 1850 | 2400 | 8000 | 0.25 | 0.16 | 250 |
| (8) | " | Phantom | 581 | 1150 | 2400 | 8000 | 0.20 | 0.12 | 250 |

NOTES: (*a*) Nominal coil spacing is 6000 feet in cable having a capacitance of 0.062 $\mu f$/mile in the side circuits and 0.100 $\mu f$/mile in the phantom circuits.

(*b*) The loading coil data are given in Table IX.

(*c*) These attenuation values apply at 55 deg. fahr. Under extreme temperature conditions, the actual attenuation may be approximately 12 per cent larger or smaller, due principally to changes in conductor resistance with temperature. In long repeatered cable circuits these variations of attenuation with temperature require special corrective treatment by means of automatic transmission regulators. (Reference No. 10.)

(*d*) These length limitations are set by transient distortion effects; echo currents may limit circuit lengths to lower values, depending on the grade of balance of the lines and the permissible over-all loss.

flexibility, the new loading systems all have the same coil spacing of 6,000 feet.

The coil spacing being fixed, it necessarily follows that any reduction in coil inductance for the purpose of raising the cut-off frequency will also increase the transmission velocity. The attenuation improvement obtained by the loading decreases as the velocity is increased. High velocity loading is more expensive than low velocity loading, in the sense that more repeaters are required for the same over-all loss. Obviously, although high velocity loading could be used for short haul traffic, it would not be so economical as a low velocity loading. Commercial considerations thus justify a series of loading standards, graded to meet the requirements of the different lengths of circuits.

At the present time the two phantom group loading standards, H-44-25 and H-174-63, are sufficient to meet the graded requirements of commercial toll cable circuits, when used with suitable combinations of conductor sizes and repeaters. Three different general types of repeaters are used, known as the 21, 22, and 44 types.[21] The 21 type is used on two-wire circuits requiring only one repeater, under conditions where switched connections involving other repeaters are not involved. The 22 type is used on two-wire circuits requiring one or more repeaters. The 44 type is used on four-wire circuits, where one pair of wires is used for one-way transmission in one direction and the other pair of wires for transmission in the opposite direction. When phantom circuits are worked on a four-wire basis, each one-way transmission path actually uses four wires.

Table XI lists the combinations of loading, conductor gage, and type of repeater circuit which are used in meeting the wide range of commercial requirements. The position of the facility item in the table indicates the sequence of transmission excellence, Item (i) being the highest grade facility in this respect. In general, the cost of these facilities is in reverse order to the sequence of electrical excellence.

The exact limits of the field of use of a given type of facility depend upon the magnitude of the permissible over-all transmission loss, and upon the grade of repeater balance obtainable. A discussion of these features would bring in complicated engineering questions beyond the scope of the present paper. So far as loading features are concerned, it is sufficient to state that H-44-25 loading is generally used on circuits of approximately 500 miles or more. On circuits intended for switched business, it is frequently necessary to use this

[21] Gherardi—Jewett, Loc. cit.

TABLE XI

*Types of Toll Cable Facilities*

| Item No. | Length Circuit | Cable Gage | Type of Loading | Type Circuit | Type Repeater |
|---|---|---|---|---|---|
| (a) | (short) | 19 | H-174-63 | 2-wire | — |
| (b) | | 16 | " | " | — |
| (c) | | 19 | " | " | 21 |
| (d) | | 16 | " | " | 21 |
| (e) | | 19 | " | " | 22 |
| (f) | | 16 | " | " | 22 |
| (g) | | 19 | " | 4-wire | 44 |
| (h) | | 16 | H- 44-25 | 2-wire | 22 |
| (i) | (very long) | 19 | " | 4-wire | 44 |

type of loading for much shorter distances. For further discussion of the use of repeatered loaded lines reference is made to recent papers presented before the Institute by Mr. J. J. Pilliod [22] and Mr. H. S. Osborne.[23]

*H-63-P versus H-106-P Loading.* The standardization of the H-63-P loading to replace the H-106-P loading for association with H-174-S loading, is of particular interest in illustrating the reactions of repeater requirements on loading design. Phantom circuits necessarily have a lower attenuation constant than the associated side circuits, when the loading is designed to meet the same standard of cut-off frequency and the coils are spaced at the same loading points. When repeaters are used on such loaded phantom circuits, the net equivalent is practically no lower than the net equivalent of the associated side circuits, due principally to the fact that the loaded sides and phantoms have practically the same velocity and cut-off frequency characteristics.

Under present operating conditions for short small gage loaded circuits of such lengths that satisfactory transmission results can be obtained without using telephone repeaters, there is ordinarily no important advantage in having the phantom circuit more efficient than the side circuits. It is a distinct operating convenience, of course, to be able to use the phantom circuit and its associated side circuits indiscriminately for the same class of service.

Having the above situations in mind, it was decided to redesign the phantom loading so that it would have approximately the same

[22] "Philadelphia-Pittsburgh Section of New York-Chicago Cable," J. J. Pilliod, Trans. A. I. E. E., Vol. 41, 1922, p. 446; *Bell System Technical Journal*, Jan., 1922.

[23] "Telephone Transmission over Long Distances," H. S. Osborne, Trans. A. I. E. E. Vol. 42, 1923, p. 984.

attenuation constant at 1,000 cycles as the associated H-174-S loading. This resulted in the reduction of the phantom loading coil inductance to 63 milhenrys. On the basis of equal attenuation losses in the phantom circuit and its side circuits, the continued use of a higher grade coil in the phantom circuit was no longer justified from a cost standpoint. Accordingly, the new 63-milhenry phantom coil (Code No. 587, Table IX) was designed to have approximately the same d-c. resistance as the earlier standard 106-milhenry coil (Code No. 583), since this permitted a substantial reduction in the size of the loading coil and a consequent reduction in cost, without increasing the over-all losses in the associated side circuits. The design finally chosen resulted in the phantom coil having approximately the same over-all dimensions as the associated side circuit coils. This permitted the phantom coils to be mounted on the same spindles with the associated side circuit loading coils as phantom groups, thus reducing the amount of inside cabling. This gave improved electrical results, besides reducing the potting costs. The use of the smaller size phantom coil, in combination with a larger size case, made it practicable to pot a total of 45 phantom group combinations (135 coils) in a single case. Using the same size case for potting phantom group combinations involving the older large size phantom coils, the limit on the number of coils was 108 (36 phantom groups).

The reduction of the phantom coil inductance from 106 to 63 milhenrys made a substantial increase in the cut-off frequency and in the velocity of transmission, as noted in Table X. These improved characteristics made the H-63-P circuit much superior to the H-106-P circuit from the standpoint of echoes and velocity distortion characteristics. On this basis the H-63-P circuit is intermediate in transmission excellence between H-174-S and H-44-25 circuits.

It was found inadvisable to make a similar change in the H-44-25 loading system owing to cross-talk reactions following from the necessary use of higher repeater gains in the phantom circuit. These undesirable reactions, though present to a lesser degree in the case of the H-174-63 system were offset by the factors already described. The size of the H-25-P coil was, however, reduced to comform to the potting method adopted for H-174-63 loading.

From the standpoint of repeater circuits the H-174-63 system is inherently better than the H-245-155 system because of its higher velocity and higher cut-off, with resulting higher quality of transmission. Furthermore, as far as non-repeatered circuits are concerned, there is a negligibly small difference between the transmission performances, considering frequency distortion effects as well as volume efficiency effects. The standardization of the H-174-63 phantom group loading system, therefore, marked the abandonment of use in new facilities of the old standard H-245-155 phantom-group loading system.

*Attenuation—Frequency Distortion.* In addition to their improved velocity and cut-off frequency characteristics, the H-44-25 and H-174-63 loading systems have an important advantage from the standpoint of attenuation-frequency distortion effects, as is illustrated in Figs. 10 and 11. The frequency distortion effects illustrated in Fig. 10

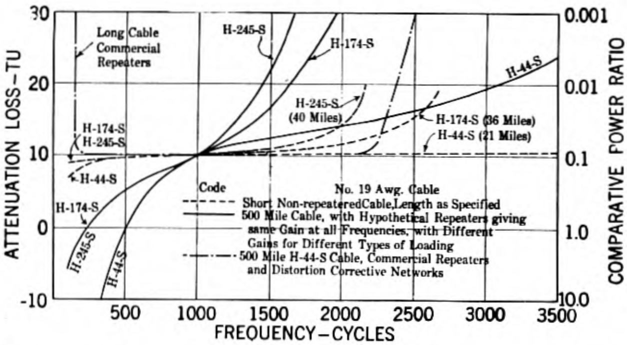



Fig. 11—Attenuation-frequency characteristics of short and long loaded toll cable circuits having a net attenuation loss of 10 TU at 1000 cycles

may become very serious in very long lines. An indication of this is given in Fig. 11. The heavy line curves in this diagram illustrate the attenuation-frequency characteristics of a 500-mile 19 A.w.g. cable circuit involving the various types of loading noted, assuming that "perfect repeaters" are used in each case to reduce the total line loss to 10 TU at 1,000 cycles. The foregoing "perfect repeater" is assumed to have the same amplification at all frequencies. Of course, in order to have the same over-all efficiency in the different types of circuits at 1,000 cycles, it is necessary to assume different total amounts of repeater gain. The dotted lines in Fig. 11 illustrate corresponding frequency characteristics of short non-repeatered cables having the same types of loading as before; in each case the length of 19 A.w.g. cable circuit being chosen so that the non-repeatered circuits have the same loss (10 TU) at 1,000 cycles. A visual inspection of the dotted and heavy line curves indicates how the line losses pile up in long connections. In the old standard low cut-off loading, the accumulated losses in very long lines amount to a suppression effect for frequencies above 1,600 cycles.

In very long lines having the newer grades of loading, the line losses are still sufficient to cause serious attenuation distortion effects if allowed to go uncorrected. The improved types of repeaters now used on long loaded circuits provide somewhat higher gains at the upper speech frequencies, thereby obtaining approximately a flat frequency characteristic over a wider frequency range. In repeaters used in conjunction with the H-44-25 loading, losses are introduced at the lower speech frequencies by auxiliaries to the repeater circuit, for the purpose of flattening the frequency characteristic at low frequencies. An indication of the improvement obtainable in the above ways is given by a dot-dash curve in Fig. 11, which illustrates the attenuation-frequency characteristic of a 500-mile H-44-S circuit having the best types of repeaters now commercially available.

In view of the difficulties brought into repeatered circuits by the use of loading, the question comes up: "Why not use more repeaters and do without the loading?" In the case of long cable circuits the answer to this question is that the coil loading substantially improves the attenuation and substantially reduces the frequency distortion at a cost which is much lower than the cost of the additional repeaters and distortion corrective networks which would be required to give the same grade of transmission without using loading.

*Long Repeatered Open Wire Lines.* In the case of the long open wire lines, the present day answer to the foregoing question is unfavorable to the use of loading. The use of improved types of repeaters now makes it possible to secure better transmission results in long repeatered circuits without loading, than can be secured in loaded repeatered lines. In this connection it should be noted that in the case of non-loaded open wire lines the distributed inductance is sufficiently large to keep the attenuation-frequency distortion low. Also the velocity of transmission is very high relative to that of a coil loaded line and there is no cut-off effect except that produced by the filters and other apparatus in the repeater sets.

These general transmission considerations are resulting in the removal of coil loading from high grade open wire lines. This dismantling work is being accelerated in order to adapt the open wire plant for a much more extensive application of carrier telephone and carrier telegraph systems.

The present expectations are that in the future new applications of open wire loading will generally be limited to isolated cases of short lines where carrier telephone or telegraph systems are not contemplated and where the maintenance and operating conditions are unfavorable to the use of telephone repeaters.

*Cable Loading Installation Features.* Cost considerations make it desirable to use aerial cable in the long toll cable installations, so this type of construction is generally used in the open country. In the vicinity of large population centers, underground cable is used.

Typical aerial cable loading installations are illustrated in Figs. 12 and 13. On the main trunk cables two-pole *H* fixtures capable of supporting four to six large coil cases are usually required. Fig. 12

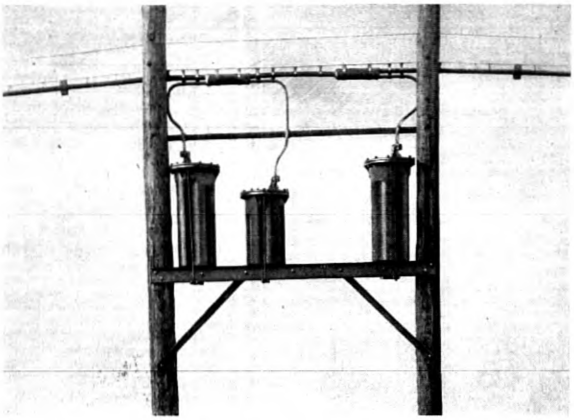


Fig. 12—Installation of aerial toll cable loading on 4-case "H" fixture

illustrates a fixture of this type designed for supporting four cases, three of which are already in place. On the smaller branch cables a single pole fixture such as illustrated in Fig. 13 is commonly used.

At the time a toll cable is installed, provision is made in the cable splices for the ultimate requirements as well as for the initial loading installation. Ordinary splices are made for the coils which are installed at the time the cable is placed, and "balloon" splices which provide the slack wire required for splicing are arranged for subsequent installations.

## III. Loading for Incidental Cables in Open Wire Lines

In the loading applications discussed in the preceding sections, the primary purpose of the loading is to reduce line attenuation losses and frequency distortion effects. In the case of incidental pieces of cable in open wire lines, however, the primary function of the loading is to give the inserted cable approximately the same impedance characteristics as the open wire line, in order to minimize reflection effects at the junction of the cable and the open wire construction. An incidental cable occurring at a line terminal is ordinarily known as a toll entrance or a terminal cable; when occurring at an intermediate point, it is known as an intermediate cable.

The reduction of junction impedance irregularities has become especially important during recent years as a result of the rapidly increasing use of telephone repeaters, since in repeatered circuits,

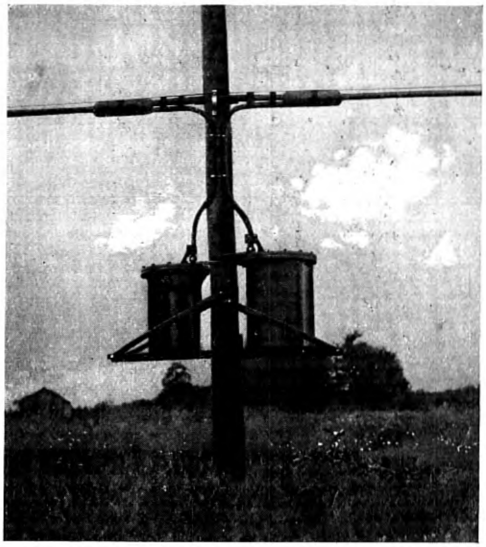


Fig. 13—Installation of aerial toll cable loading—single pole fixture for small branch cables

line impedance irregularities, by virtue of their effect upon the repeater circuit balance, may reduce the effective repeater gain and thereby impair transmission by an amount much larger than the ordinary reflection loss. Prior to the general use of telephone repeaters, satisfactory results were obtained by using some one of the standard heavy or medium weight cable loading systems on the entrance and intermediate cables associated with loaded open wire lines, and a special weight of loading was used on the incidental cables in the non-loaded open wire lines. In some cases ordinary medium loading was used, with suitable types of step-up or step-down transformers at the terminals of the inserted cable.

*Incidental Cables in Loaded Open Wire Lines.* In toll entrance and intermediate cables associated with loaded open wire lines, the primary requirements for matching impedance are that the nominal impedance and the cut-off frequency of the cable loading and of the loaded open wire line should be closely the same. To a first degree of approximation this means that the cable loading sections should have the same total mutual capacitance as the open wire loading sections, which, of course, requires a very much closer spacing. The cable loading system which was standardized for use in association with loaded open wire lines is designated "E-248-154". Its primary electrical characteristics are given in Table XII. Besides meeting

TABLE XII

*Typical Loading Systems for Toll Entrance and Intermediate Cables*

| Loading System Designation | Type Circuit | Coil Inductance Henrys | Coil Spacing Miles | Nominal Impedance Ohms | Cut-off Frequency Cycles | Attenuation Loss TU per Mile at 1000 Cycles |
|---|---|---|---|---|---|---|
| E-28-16 | Side<br>Phantom | 0.028<br>0.016 | 1.09<br>1.09 | 650<br>400 | 7200<br>7800 | 0.15<br>0.13 } (13 A.w.g.) |
| CE-4.1-12.8 | Side<br>Phantom | 0.0041<br>0.0128 | 0.176<br>1.09 | 600<br>400 | 45000<br>8500 | 0.22<br>0.19 } (13 A.w.g.) |
| M-44-25 | Side<br>Phantom | 0.044<br>0.025 | 1.66<br>1.66 | 650<br>400 | 4600<br>4900 | 0.29<br>0.24 } (16 A.w.g.) |
| E-248-154 | Side<br>Phantom | 0.248<br>0.154 | 1.09<br>1.09 | 1950<br>1200 | 2400<br>2500 | 0.081<br>0.070 } (13 A.w.g.) |

NOTE. Cable capacitance is assumed to be 0.062 $\mu f$ per mile for side circuits, and 0.100 $\mu f$ per mile for phantoms.

the impedance requirements for use in association with repeatered open wire lines, it is also very satisfactory with respect to attenuation characteristics. In placing this loading, it is customary to locate the first loading point in the cable at such a distance from the last loading point in the open wire line that the total capacitance of the junction loading section is closely the same as that in the regular open wire loading sections.

*Incidental Cables in Non-Loaded Open Wire Lines.* The problem of designing coil loading for incidental cables in non-loaded open wire lines is considerably more complicated than the case above discussed, primarily because it involves an impedance match between a smooth line and a lumpy line. Broadly stated, the first part of the problem is to design a loaded cable of such characteristics that its corresponding smooth line is closely similar to the non-loaded open wire line. The second and more complicated part of the problem is to determine the coil spacing. This usually involves some degree of compromise, because of the dependence of the impedance of a loaded cable upon the loading termination.

The first general requirement is that the ratio of inductance to capacitance to resistance per unit length in the loaded cable should be the same as the corresponding ratio in the non-loaded open wire line. Ordinarily, the loading coil resistance does not play an important part in the determination of the optimum resistance for the loaded cable, the choice of conductor gage being far more important. From this point of view, No. 13 A.w.g. is practically the best gage of conductor for entrance cable circuits connecting with 165-mil open wire lines. For the optimum impedance match on cables connecting with 104-mil open wire lines, it is necessary to use much higher resistance conductors, the choice between Nos. 16 and 19 A.w.g. conductors depending upon a number of factors which space limits do not allow to be discussed.

As noted in the discussion under "Theory" in the first part of the paper, the characteristic impedance of a uniform line is substantially a pure resistance, having the value $\sqrt{L/C}$ over the frequency range throughout which the inductive reactance per unit length is large with reference to the resistance. On the other hand, the characteristic impedance of a coil loaded cable varies over a wide range with frequency, depending upon the particular loading termination used.

Typical impedance-frequency curves for mid-coil and mid-section terminations are illustrated in Fig. 3. As will be seen from this diagram, the rising slope of the mid-section termination and the drooping slope of the mid-coil termination do not deviate greatly from a straight line relation for frequencies below approximately 0.5 of the cut-off frequency. The higher the cut-off frequency is, the more closely will the impedance-frequency characteristic of the loaded cable approach the flat characteristic of the non-loaded open wire line over the range of frequencies involved in speech transmission. In this connection, it is to be noted that the repeaters now used on open wire lines are designed to transmit frequencies between approximately 200 and 2,600 cycles. Of course, the higher the cut-off frequency, the more expensive will be the loading. Practical reasons make it desirable to space the loading coils on the cable circuits connecting with non-loaded open wire lines at the same intervals as the coils which are used on the cable circuits connected with the loaded open wire lines. This consideration in combination with the nominal impedance requirement previously mentioned, fixes the cut-off characteristics and, hence, the slope of the termination impedance-frequency characteristic.

These general considerations have led to the standardization of the E-28-16 loading system for use on entrance cable and intermediate cable conductors associated with non-loaded open wire lines. General data for this system are given in Table XII, and the computed impedance characteristics are illustrated in Fig. 14, which also gives

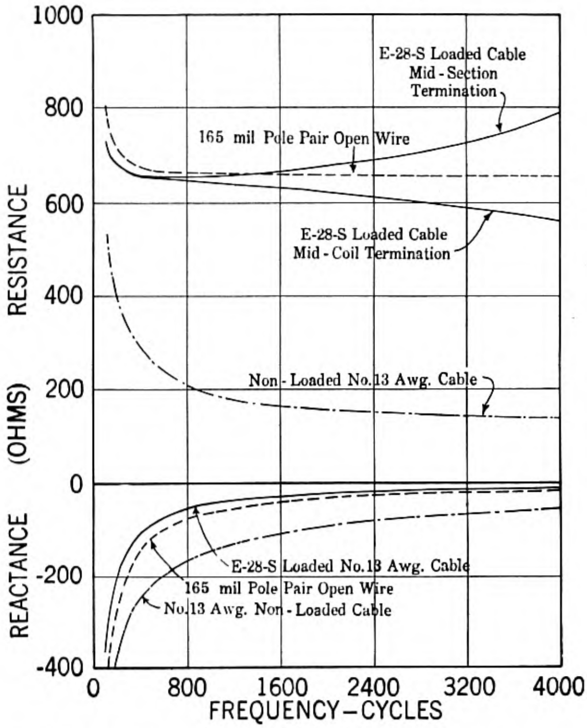



Fig. 14—Typical impedance-frequency characteristics of loaded and non-loaded entrance cable and non-loaded open wire line

the characteristic impedance curves for the non-loaded open wire line and the non-loaded cable. Since the E-28-16 loading system is a low impedance loading, the attenuation improvement is small relative to that of other types of loading system which are primarily installed for attenuation improvement.

Table XII also gives general data regarding the M-44-25 entrance cable loading system which has been used to some extent as a substitute for the E-28-16 loading system on cables connected to non-loaded open wire lines. The M-44-25 system used higher inductance loading coils and considerably longer spacing intervals than the E-28-16 system, and was consequently less expensive. The impedance characteristics, however, were not so satisfactory at the upper speech frequencies because of the greater slope of the impedance-frequency characteristic, due to the lower cut-off.

*Carrier Frequency Loading.* Special types of entrance cable loading have been developed for use on incidental cables in open wire lines on which carrier telephone or carrier telegraph systems are superposed. Loading system CE-4.1-12.8 listed in Table XII exemplifies this type of loading. The present standard carrier telephone systems operate up to frequencies of the order of 30,000 cycles.[24]

In order to get satisfactory impedance and attenuation characteristics in the loaded incidental cables, a cut-off frequency of approximately 45,000 cycles is used.

The highest working frequency in carrier loaded cables is approximately 0.75 of the cut-off frequency. The ordinary mid-coil and mid-section terminations do not give sufficiently close approximations to a flat impedance-frequency characteristic over this wide range of frequencies, so it has been necessary to use at the terminals of carrier loaded cables, a simple impedance corrective network.

Data regarding attenuation losses in a typical carrier loaded cable are given in Table XIII. For purposes of comparison similar data are given on a corresponding non-loaded cable. Effective resistance values of the carrier loading coil are also included.

The high frequency loading is used only on the side circuits, since at the present time it is not customary to operate carrier telephone systems over phantom circuits. The associated phantom circuit

TABLE XIII

*Carrier Frequency Loading*

| Frequency Kilocycles | Attenuation Loss-TU per Mile (13 A. w. g. Cable) | | Resistance-Ohms per Carrier Loading Coil |
|---|---|---|---|
| | Non-Loaded | C-4.1 Loading | |
| 1 | 0.49 | 0.23 | 1.5 |
| 5 | 0.78 | 0.27 | 1.6 |
| 10 | 0.90 | 0.33 | 1.9 |
| 20 | 1.14 | 0.52 | 4.1 |
| 30 | 1.37 | 0.90 | 8.1 |

[24] "Carrier Current Telephony and Telegraphy," E. H. Colpitts and O. B. Blackwell, Trans. A. I. E. E., Vol. 40, 1921, p. 205.

loading is designed for ordinary speech transmission. In order to transmit the high frequency carrier currents over the side circuits, it is necessary to have the side circuit loading coils spaced much more closely than for the ordinary voice frequency loading coils in the phantom circuit. On this account the theoretically best loading points for the carrier circuits frequently occur at places where it is inconvenient to locate the loading coils. The actual loading sections in such cases are made shorter than the theoretical lengths, and the deficiencies in loading section capacitance are remedied by adding lumped capacitances in the form of "building-out condensers." Recently, special types of stub cable designed specially for building out purposes have come into use as substitutes for building-out condensers.

*Loading Coils.* The design of the coils used in the E-28-16 and M-44-25 loading systems is generally similar to the toll cable loading coils having 35-permeability compressed powdered iron cores already described. The loading coils used in the E-248-154 loading system are larger coils of the air-gap type 65-permeability wire core construction listed in Table VII.

As regards the carrier loading system, CE-4.1-12.8, since this involves the transmission through the loading coils of frequencies up to 30,000 cycles or somewhat higher, special coil designs are required. The coil which loads the audio frequency phantom circuit, aside from being specially balanced for association with the side circuit coils, is generally similar in construction to the compressed powdered iron core phantom coil for toll cables.

The side circuit coil, however, is used for loading the high frequency circuit, and more severe requirements are, therefore, imposed on it owing to the multi-frequency transmission. Ordinarily the circuits are equipped to provide three or four carrier telephone channels or 10 carrier telegraph channels over a pair of wires, in addition to the ordinary audio frequency telephone and grounded telegraph channels. The primary added requirements as regards the loading coils are freedom from intermodulation between channels, and low energy losses at carrier frequencies. The most satisfactory solution as regards freedom from magnetic modulation is the avoidance of the use of ferro-magnetic core materials. The side circuit loading coils were, accordingly, designed as toroidal wood core coils, with finely stranded copper windings in order to limit the eddy-current losses. Data regarding resistance-frequency characteristics are included in Table XIII.

The air core side circuit coils have a small leakage inductance which must be allowed for in determining the phantom coil inductance. For this reason the phantom coil inductance is lower than in the E-28-16 system (Table XII.) In order to avoid impedance irregularity in the carrier circuits at the phantom loading points, it is necessary that the combination carrier-phantom loading units should have closely the same total inductance and shunt capacitance as the ordinary carrier loading coils. This requires the use of a different type of carrier loading coil at the phantom loading point from that at the non-phantom loading points, having a lower inductance and capacitance corresponding to the leakage inductance and shunt capacitance of the associated phantom coil. Other refinements of design are involved in these combination loading units.[25]

## IV. Cross-Talk

One of the greatest practical difficulties which has been encountered in extending the commercial range of long distance telephone service is that of keeping at a tolerably low value, the speech overhearing effects known as cross-talk, which occur between adjacent telephone transmission circuits whenever there is an appreciable amount of electromagnetic or electrostatic coupling between them.

From the early days of telephony great care has been exercised in plant design and construction work to avoid circuit and apparatus unbalances, but as is to be expected from the nature of the problem, it is practically impossible to obtain and maintain absolutely perfect balance. In short telephone circuits, there is no particular difficulty in keeping the unbalance effects small enough so that the over-all cross-talk is not serious. As the length of the line increases, however, there are more and more opportunities for unbalances in the lines and in the associated apparatus in the lines and offices. In repeatered lines, moreover, the repeaters amplify the cross-talk as well as the speech transmission. Thus we have the cumulative effects of cross-talk from successive sections in the long repeatered lines. From the service standpoint, moreover, it is necessary that the cross-talk in the very long lines should be within the limits set for the shorter lines.

The problem of keeping cross-talk low between a phantom circuit and its associated side circuits, and between the two associated side circuits of a phantom group, is by far the most difficult phase of the general cross-talk problem in long repeatered cables. It is present in the cables, the loading coils, the terminating apparatus and the office

[25] U. S. Patents Nos. 1,501,959, Martin and Shaw; 1,501,926, Shaw.

cabling. Of these, the cable and associated loading coils are the major sources of unbalances.

The phantom-to-side and side-to-side cross-talk unbalances in the cable quads are reduced to small values by exercisng great care both in the various manufacturing processes and in the selection of raw materials. When the cable is installed in the field, a large improvement in cross-talk conditions is secured by splicing adjacent lengths of cable together in such a way that the unbalances in one length of cable substantially neutralize the unbalances contributed by the adjacent length of cable. Usually, three such "capacity-unbalance test" splices are made at symmetrical points in each loading section and as a result the average over-all capacity unbalance in a loading section is reduced to about one-tenth of the magnitude which would hold if these test splices were not made.

In the design of the standard phantom circuit and side circuit loading coils, special care was taken to make them substantially free from inherent unbalances. Also in the manufacture of the coils, great care is exercised to realize the benefits of the inherent symmetry of the designs. In the early days before telephone repeaters came into general use on loaded lines, satisfactory results from the standpoint of self inductance and mutual inductance unbalances were obtained by adjusting the different windings to the nearest turn; *i.e.*, a condition of balance where either adding or subtracting one turn to one of the line windings would increase the cross-talk rather than reduce it.

Later when repeaters came into general use, it was found necessary to obtain much more refined adjustments. Further improvements have been worked out in manufacturing methods and processes which allow a greater degree of symmetry. As a result of these various improvements, the phantom-to-side cross-talk unbalances in the loading coils have been reduced approximately 75 per cent. or more below the values obtained before repeaters came into general use on small gage toll cable. The coil cross-talk unbalances are now nearly as low as the cross-talk unbalances in the associated cable sections after the completion of the capacity unbalance test splicing.

The loading coils used in the very long circuits having H-44-25 loading obviously are more important from the standpoint of cross-talk limitations than the coils used in the shorter circuits having H-174-63 loading, and somewhat greater care is required in their manufacture. These coils are adjusted and tested in a factory test circuit which at the cross-talk test frequency simulates the service impedance conditions. In the phantom-to-side cross-talk test, the disturbing test current is superposed on the phantom circuit, and measurements of the cross-talk are made in the side circuits, the cross-talk being expressed in millionths of the current into a transformer connected to the phantom circuit and of such ratio as to make the impedance at its input equal to that of the side circuit. As a result of the improvements previously mentioned, the average cross-talk in the coils used for the H-44-25 loading is now about 20 millionths. This corresponds to an attenuation of about 95 TU.

To assist in visualizing the real achievement which this minute value of phantom-to-side cross-talk represents, Table XIV gives information regarding the cross-talk of different elementary types of unbalance in H-44-25 loading coils:

TABLE XIV

*Cross-talk Due to Unbalance in H-44-25 Loading Coils*

| Type of Unbalance | Amount of Cross-talk |
|---|---|
| 1 ohm resistance | 400 millionths ( 68 TU) |
| 1 micro-henry inductance | 2.5 " (112 TU) |
| 1 turn of winding | 280 " ( 71 TU) |
| 1 micro-microfarad capacitance | 0.94 " (121 TU) |

These values apply at 1,000 cycles.

In the loading coils designed for H-174-63 loading, the cross-talk per unit of electromagnetic unbalance tends to be smaller and the cross-talk per unit of electrostatic unbalance larger, in rough proportion to the differences in line impedance between the H-44-25 and H-174-63 circuits.

Side-to-side cross-talk is uniformly lower than phantom-to-side cross-talk, as would be expected from the less intimate coupling between circuits. Accordingly, the special adjustments which are made are primarily for the purpose of reducing phantom-to-side cross-talk.

In the loading coils intended for H-44-25 circuits the special cross-talk adjustments are applied for minimizing "far-end" cross-talk or for minimizing "near-end" cross-talk, according as the coils are required for four-wire or two-wire repeatered circuits, respectively. The term "far-end" cross-talk applies to cross-talk heard at the distant end of the disturbed circuit, and correspondingly the term "near-end" cross-talk applies to the cross-talk heard at the end of the disturbed circuit near the talker.

Considering now the cross-talk between four-wire circuits in the same quad, it is to be noted that the directional effects of the telephone repeaters block the transmission of cross-talk in the one-way path back to the near end of the circuit, and consequently the special cross-talk adjustments on the coils for four-wire H-44-25 circuits are made primarily for reducing far-end cross-talk.

In two-wire circuits, near-end and far-end cross-talk both occur, and generally near-end cross-talk is much greater because its "average" cross-talk path has less attenuation than that of the far-end cross-talk. Consequently, the special cross-talk adjustments made in the two-wire circuit coils are for the purpose of reducing the near-end cross-talk to a minimum.

In the foregoing connection, it is to be noted that the cross-talk current caused by electromagnetic unbalances flows around the two ends of the disturbed circuit in series. On the other hand, the cross-talk current caused by electrostatic unbalances divides and flows from its point of origin in opposite directions around the two ends of the circuit in parallel. Consequently, when electrostatic and electromagnetic cross-talk currents are in phase at one end of the circuit, they will be practically in phase opposition at the other end of the circuit. The special cross-talk adjustments are made in such a way as to get the maximum benefit from the phase opposition at the particular end of the circuit where the reduction is more important.

In the four-wire type of circuit used in very long cable circuits, relatively large amplification gains are possible in the repeaters because of the characteristic circuit feature which allows the repeaters to act as one-way amplifiers. As a result of these high amplifications, there are large differences in power level on the input and output sides of the repeaters. This fact has made it desirable for cross-talk reasons to segregate the oppositely transmitting branches of the four-wire circuits. In the cables, the "east-bound" and "west-bound" branches of the four-wire circuits are in different groups. This segregation is also carried out in the loading coil pots, and in the office cabling.[26]

With loading coils as manufactured at present, the cross-talk unbalances in the loaded cables are such that the resultant over-all cross-talk is expected to be tolerable for the longest circuits now definitely planned in cable. The margin below commercial limits is much less in two-wire circuits than in four-wire circuits. At present, there is a growing tendency to use two-wire circuits for longer distances

[26] U. S. Patent No. 1,394,062—O. B. Blackwell.

than formerly, for reasons of plant economy. This trend thus increases the severity of the cross-talk requirements.

Unbalances in loaded circuits which contribute to noise due to induction from power transmission and distribution circuits are similar in nature to those contributing to cross-talk. The precautions which are taken in the design, manufacture, and installation of loaded circuits to reduce unbalances have the effect, therefore, of reducing both cross-talk and noise.

## V. Telegraphy Over Loaded Telephone Circuits

It had been the practise in the Bell System, before the advent of loading, to employ circuits for simultaneously transmitting telephone and telegraph currents. Two methods were in general use, (1) the composite system, in which each line wire of the telephone circuit



## VI. Recent Improvements in Loading for Exchange Area Cables

The developments discussed in the preceding sections were directed



The M-88 system is especially suitable for the shorter lengths of fine wire trunk cables which constitute the predominating bulk of the exchange area trunk mileage. In longer trunks, the other more expensive loading systems find their field of service. The H-175 system is limited to low capacitance cables because of the lower cut-off effects on high capacitance cables, but has considerable commercial importance because of the large number of low capacitance cables now in the plant.

Table XVI gives general transmission data on typical exchange area trunks using the new loading systems, including also non-loaded trunks. Attenuation-frequency characteristics of some of these trunks are given in Fig. 15. A dotted line curve shows the characteristics



loading on No. 22-A.w.g. cable is the preferred construction for

*Loading Coils and Cases.* As previously noted, the first important change in the coils used for exchange area loading from the early 95-permeability wire core type was the substitution of compressed powdered iron in place of wire for the cores. Initially, only coils having powdered iron cores with a permeability of 60 were designed, as this value corresponds to the effective value of the cores displaced. More recently, in order to better fit in with the requirements of the new cut-off frequency standard, coils using 35-permeability powdered iron cores have been developed. In Table XVII are listed data for the coils now used in exchange area loading.

TABLE XVII

*Coils for Loading Exchange Area Cables*

| Coil Code No. | Induct-ance (Henrys) | Core Perme-ability | Resistance-Ohms | | Over-all Dimensions Inches | |
|---|---|---|---|---|---|---|
| | | | D-C. | 1000 Cycles | Diam-eter | Height |
| 602 | 0.088 | 35 | 8.9 | 10.5 | 3.6 | 1.3 |
| 603 | 0.135 | 35 | 12.8 | 14.1 | 3.6 | 1.3 |
| 574 | 0.175 | 60 | 4.6 | 10.6 | 4.5 | 2.1 |

Effective resistance values are for a line current of 0.001 ampere.

The standardization of the small size Nos. 602 and 603 loading coils has made it possible to design containing cases and assembly methods which permit much larger numbers of coils to be enclosed in cases conforming to the dimensional limitations set by existing vault conditions. A series of cases having capacities up to 300 coils has now been developed. The use of these large potting complements will be of considerable value in reducing the space congestion encountered in the "downtown" sections of the larger metropolitan areas.

In the 300-coil case, a total of 1,200 soldered joints are required to connect the coil terminals to the stub cable conductors. It was accordingly very important that the assembly method should involve a minimum liability to open circuits, crosses, or grounds. To accomplish this, a method was devised whereby the various spindles of coils were assembled to a skeleton frame to which the cable stub containing the 600 terminal pairs is also attached. All splices to the outgoing conductors are made immediately adjacent to the individual coil terminals, after which the skeleton unit consisting of the coils and stub cable with case cover attached, is picked up with suitable tackle, and the coil unit inserted in the case. Fig. 17 illustrates this stage of the assembly. The case is subsequently filled with moisture-proof compound and sealed in the usual manner.

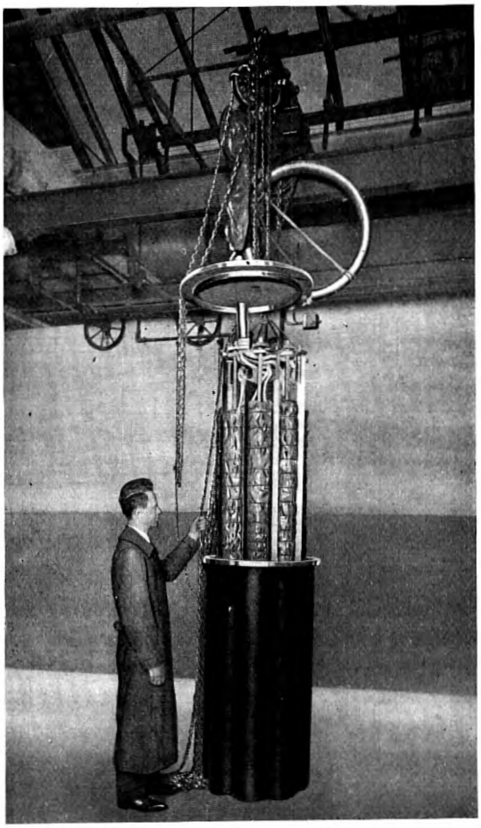


Fig. 17—Assembly of 300-coil case
Lowering loading coils into case after coil spindles have been mounted on frame and coil terminals spliced to stub cable conductors

*Installation Features.* In general, the exchange area cables on which loading is required are run in underground ducts and consequently the great bulk of the exchange area loading is installed in underground vaults. Fig. 18 shows a typical loading installation in a "double-deck" vault in New York City. The loading coil cases are placed in the lower part of the vault permitting the coil terminal stub cables to be brought up vertically behind the horizontal cable runs and spliced to the trunk cables in such a way as to minimize the difficulties of future work on the cables passing through the vault. The trunk cables enter the vault through ducts which may be seen at

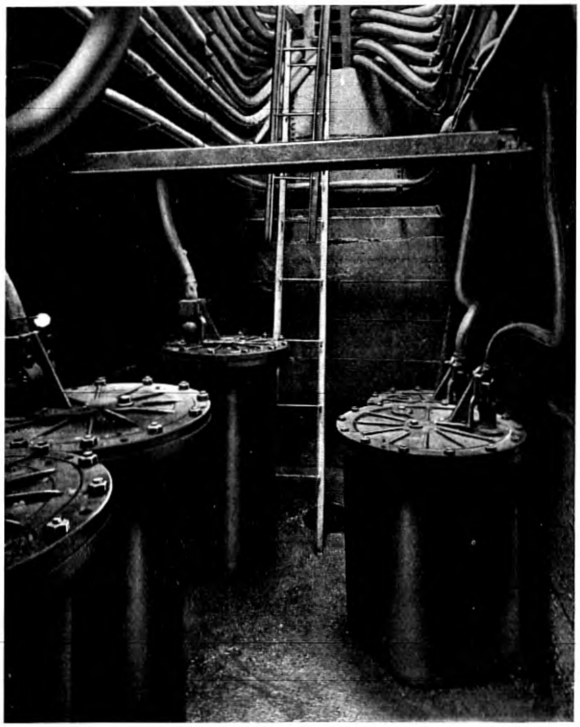


Fig. 18—Underground cable loading coil installation in Metropolitan area. Double-deck vault having ultimate capacity of 14 large coil cases

the top of the picture, and are supported on racks mounted on the upper side walls of the vault.

At present, a total of eight loading coil cases is installed in the vault illustrated in Fig. 17. Only five of these, however, appear in the picture. The cases now in place contain a total of 645 loading coils. The vault has space for six additional large cases, on which basis it is estimated that this vault will ultimately contain about 2,400 coils. Some of the largest vaults are capable of accommodating a total of 30 large cases containing a total of 9,000 coils.

## VII. Loading for Submarine Cables

*Coil Loading.* The special problem of applying coil loading to submarine cables is a mechanical one, rather than one concerning the principles of loading. The situation in the United States is such that only a few coil loaded submarine cables have been required; this, of course, does not refer to the considerable number of instances where the submarine cables are so short that ordinary types of coils installed at the terminals satisfy the transmission requirements.

To date there have been installed in the United States a total of five cables having submarine coil loading; details of which are given in Table XVIII.

TABLE XVIII

*Coil Loaded Submarine Cables*

| Location | Year of Instal-lation | Length of Cable Miles | No. of Load Points | Number of Loaded Ccts. | Spac-ing of Coils-Miles | Coil Induct-ance Henrys |
|---|---|---|---|---|---|---|
| Chesapeake Bay No. 1 | 1910 | 4.5 | 2 | 17 pr. 13 A.w.g. | 1.97 | 0.117 |
| Chesapeake Bay No. 2 | 1916 | 4.0 | 1 | 12 qd. 13 A.w.g. | 2.0 | 0.067-S<br>0.042-P. |
| Tarrytown-Nyack | 1916 | 2.7 | 2 | 37 qd. 16 A.w.g. | 0.89 | 0.250-S<br>0.155-P |
| Raritan Bay No. 1 | 1917 | 5.3 | 5 | 37 qd. 16 A.w.g. | 0.91 | 0.250-S<br>0.155-P |
| Raritan Bay No. 2 | 1918 | 5.3 | 5 | 37 qd. 16 A.w.g. | 0.89 | 0.250-S<br>0.155-P |

The Raritan Bay cables each have 37 quads loaded at five points, 111 coils at each point, constituting the largest installation of submarine coil loading in the world. The depth of water in which these cables are installed is about 35 feet.

In each of the above instances, dry core paper cables were used. The design of the coil was made to fit the special designs required for the submarine loading pots. The case design was such as to furnish complete protection of the coils against moisture penetration and adequate mechanical strength for taking up the tension in the cable. In installing the cables, the procedure was to splice the loading coil cases into the cable while the cable was coiled on a barge, and then lay the cable and coils as a continuous operation. The service record of these loaded submarine cables is excellent, thus demonstrating a satisfactory solution of the many difficult mechanical problems involved.

*Continuous Loading.* Another form of submarine cable loading, first put into practise by the Danish engineer, C. E. Krarup,[30] is to wind an iron wire or tape spirally around the copper conductor. This gives a continuous loading which has found important applications in the case of telephone and telegraph cables laid in deep water. So far as land cables are concerned, it has been found that continuous loading is uneconomical in comparison with coil loading. The only instances of continuous loading in the plant of the Bell System are the Florida-Cuba cables,[31] connecting Key West and Havana, which are the longest and most deeply submerged cables in use for telephonic communication in the world.

## VIII. Extent of Commercial Application

The following data will assist in visualizing the practical importance of the developments which have been described in this paper.

In 1911, when Mr. Gherardi addressed this Institute on the subject of loading practise in this country, there were about 125,000 loading coils in service which loaded about 85,000 miles of open wire circuits and 170,000 miles of cable circuits. Although precise figures are not yet available regarding the number of loading coils in service in the Bell System as of January 1, 1926, conservative estimates set this total at about 1,250,000 coils. These coils load about 1,600,000 miles of cable circuits and 250,000 miles of open wire. In round numbers, 500,000 coils are installed on non-quadded local area trunk cables and 700,000 in toll and toll entrance cables (the bulk of these being quadded cables). Nearly two-thirds of the total number of coils have compressed iron powder cores, all of these being installed on cable circuits. About 4500 coils having wooden cores are installed on carrier loaded entrance cables. The remainder have iron wire cores, approximately 60,000 being of the so-called "air-gap" types.

Prior to the development of satisfactory types of telephone repeaters, the principal use of loading coils was in exchange area trunk cables in large metropolitan areas such as New York, Chicago, Philadelphia, and Boston. The successful application of telephone repeaters to loaded small gage cables has greatly increased the use of loading in the telephone plant. As illustrating this trend, approximately 150,000 toll

[30] C. E. Krarup, Submarine Telephone Cables with Increased Self-Induction, *ETZ.*, 23:344, April 17, 1902.

[31] W. H. Martin, G. A. Anderegg, B. W. Kendall, "Key West-Havana Submarine Telephone Cable System," Trans. A. I. E. E., Vol. 41, 1922, p. 184.

cable coils were manufactured for the Bell System in 1925, and approximately 100,000 exchange area cable coils. Recent estimates of the loading coil requirements for the next five years indicate an annual demand at a rate which would double the total number of loading coils in service about 1930.

As regards the field of application for cable loading in terms of cable lengths, the entrance and intermediate cables represent the minimum lengths; for instance, pieces as short as 500 feet when present in carrier telephone systems may require loading. In the local exchange areas, toll switching trunks as short as two miles may require loading. On the other hand, as illustrating the longest circuit now entirely in cable, a connection between Boston and Milwaukee—via New York, Pittsburgh, Cleveland, and Chicago—typifies the possibilities in the existing repeatered loaded cable plant. The over-all length of such a circuit is approximately 1200 miles. There is no technical obstacle to the use of repeatered loaded cables for distances several times as great; i.e., in the present state of the art, this is primarily a question of economics rather than of development.

## IX. Conclusion

It will be appreciated from the foregoing account that the invention of coil loading was the beginning of an era of intensive development which has been marked by enormous advances in the design of telephone transmission lines, and that there has been no slackening of the inventional or development activity devoted to this subject. It is significant that at present more engineers and physicists in the departments represented by the authors are engaged on loading development problems than at any previous time.

In this account of the progress of the loading art during the past quarter century, the authors have endeavored to point out the relation of the loading developments to other phases of telephone development such as cables, repeaters, telegraph working, and carrier telephone and telegraph systems. In the space that is available, it would be impracticable to assign full credit to the many individuals who have been engaged in the development work on loading and the related problems. The final accomplishments should be regarded as the result of well coordinated efforts along many lines.

In conclusion, it may be of interest to note what the development and use of loading has meant to the telephone using public from an economic standpoint. Leaving out of consideration altogether loading on long toll cables—where the interdependence of repeaters and loading is such that it is impracticable to assign to each its share of the savings—and taking into consideration only the loading of interoffice trunks and toll open wire circuits, it has been estimated that the larger wires which would have been required to give the present grade of transmission if loading had not been available, taken together with the heavier pole lines and additional underground ducts, would have entailed an additional investment in Bell System telephone plant of over $100,000,000.

## BIBLIOGRAPHY


In addition to the references already cited, the following will be of interest:

Vaschy, *Annales Telegraphiques*, 1886, p. 321.

Heaviside, *Electrician* (London) June 3, 1887.

Vaschy, *La Lumiere Electrique*, 1889, 31, 83.

Heaviside, "Electromagnetic Theory," Vol. 1, 1893, p. 441.

S. P. Thompson, "Ocean Telephony," *Proceedings* Electrical Congress, World's Fair at Chicago, 1893.

M. I. Pupin, "Propagation of Long Electrical Waves," Trans. A. I. E. E., Vol. 16, 1899, 93.

W. A. J. O'Meara, "Submarine Cables for Long Distance Telephone Circuits," *Journal I. E. E.* London, April, 1910.

F. Bresig, *Theoretische Telegraphie*, 1910.

J. A. Fleming, "Propagation of Electric Currents," Van Nostrand Company, 1911·

A. E. Kennelly, "The Application of Hyperbolic Functions to Electrical Engineering Problems," University (London) Press, 1912.

A. Ebeling, *ETZ.*, 1914, p. 695 and p. 728.

K. W. Wagner, *Archiv für Electrotechnik*, Vol. 3, 1915, p. 315.

J. G. Hill, "Telephonic Transmission," London, Longmans, 1920.

K. W. Wagner, and K. Küffmüller, *Archiv für Electrotechnik*, Vol. 9, 1921, p. 461.

Sir Wm. Noble, "Long Distance Telephone System of the United Kingdom," *Journal I. E. E.*, Vol. 59, p. 389, 1921.

Das Fernsprechen Im Weitverkehr—Reichspostministerium, Berlin, 1923.

R. S. Hoyt, Impedance of Smooth Lines and Design of Simulating Networks, *Bell System Technical Journal*, April, 1923.

L. D. Cahen, Le Progres et l'Etat Actuel de la Technique des Liques Pupinisees—Bulletin de la Société francaise des Electriciens Aug.-Oct. 1924 (4 Serie), 4:755.

K. S. Johnson, Transmission Circuits for Telephonic Communication, Van Nostrand Co., 1925.

# A Static Recorder

By H. T. FRIIS

SYNOPSIS: This paper discusses different types of apparatus for recording static and also describes a new instrument in which the output of the set is kept constant by automatic control of the amplification, this amplification then being recorded as the relative measure of static. The set makes use of a fluxmeter with zero restoring torque by means of which the rectified output current arising from static interference is integrated over a period of ten seconds. The following five seconds are required to adjust the gain of the amplifier and record the change in gain from an arbitrary level. The gain is recorded in stops of 4 TU which correspond to a power amplification change by approximately a factor of 2.5. A record is shown during which the intensity of static changed by a factor of more than 10,000.

IN the following is given a general discussion of receiving sets for recording static and also a detailed description of a new instrument of this kind which is based upon the principle that the output of the set is kept constant by automatic control of the amplification, this amplification then being recorded as the relative measure of the static.

The literature on manual measurements of static is plentiful and for extensive references the reader may be referred to a paper on "Present Status of Atmospheric Disturbances," presented by L. W. Austin before the American Geophysical Union.[1] Many different methods of measuring static have been employed in obtaining the results given in this paper and it may further be added that we have found that the most reliable method of measuring the effect of static upon the intelligibility of speech signals is to introduce a local warbler signal [2] in the antenna. Unfortunately, however, all manual measurements require trained observers and therefore the cost of making continuous measurements will always be high, and besides, the human element introduced will decrease their reliability.

Very little has been published on automatic recording of static. The American Telephone and Telegraph Company and the Western Electric Company in 1923 developed an automatic static recorder which measured the high frequency currents induced in a loop antenna by amplifying them and passing them through a recording thermocouple meter. This apparatus was also equipped with a means of automatically measuring the gain of the entire receiving device so that the energy of the static could be evaluated directly. A popular account of this device was given under the caption "Getting Static's Autograph" by Austin Bailey in "Popular Radio," May, 1924. A recorder working at Aldershot, England, is mentioned in a paper by

[1] Will be published in the Proc. I. R. E. probably in the February, 1926, number.

[2] See "Radio Transmission Measurements," by Bown, Englund and Friis, Proc. I. R. E. Vol. 11, No. 2.

R. A. Watson Watt,[3] but it seems that this recorder is mainly an automatic counter of static crashes and it would therefore be of little value in U. S. A. where static is mostly a continuous rumble. The reason for the small advance which has been made to date in the automatic recording of static is probably due largely to the lack of suitable apparatus. Certainly there has never been any doubt that automatic records would be very valuable. It is just as important to know the static level as it is to know the strength of a radio signal because it is the static to signal ratio that determines the intelligibility of the signal. A static recorder connected to a rotating directional antenna system would tell us where static comes from and therefore enable the radio engineer to determine whether it is worth while to construct a directive antenna system. Also the connection between thunder-storm areas and static would make static recording valuable to the meteorological service. There is perhaps no reason why a suitable static recorder should not make it possible in a few years to obtain a daily static forecast just as we get our weather forecast now.

The question is then, what would be the best way of obtaining such a record of static? It would, of course, be very desirable to get a continuous record of the actual shape of the static wave, but we have no hope of ever realizing this and will have to be satisfied with the wave forms of a few typical static impulses as given by Watson Watt and E. V. Appleton.[4] Besides it would require a tremendous amount of labor to interpret such a record. The recorder described in this paper records the energy received within periods of 10 seconds. To be sure, such an energy curve of static does not tell the whole story due to the fact that the character of static is so variable. Thus, the same energy levels of a continuous rumbling static and of static consisting of separate clicks does not mean that these two types of static have the same effect upon the intelligibility of a speech signal. However, the shape of an energy record will indicate the general character of the static, but whether such an energy record will enable us to obtain absolute quantitative results with respect to the effect of static upon speech signals cannot yet be determined until further experimental results are available.

## Requirements of an Energy Recorder

Let us take the case of recording the rectified current through the receiver of an ordinary receiving set supplied with a local carrier

[3] "Directional Observations of Atmospheric Disturbances," by R. A. Watson Watt, Proc. Royal Soc. A, Vol. 102, page 477.

[4] "On the Nature of Static," by Watson Watt and Appleton. Proc. Royal Soc. A, Vol. 103, page 84.

oscillator. This may, for instance, be accomplished by replacing the phone by a thermocouple connected to a standard recording galvanometer. Such a recorder would probably record the average of the current squared, but only for changes of energy received of not more than fifty times. The daily variation of the energy level of static (at 60 kilocycles) is, however, generally at least 100 times and sometimes even 10,000 times, so that this method of employing a receiving set with fixed amplification is unsatisfactory. Besides it would be very difficult to prevent overloading of the set if we limit ourselves to the use of 10-watt tubes.

One important requirement of a static recorder is therefore that the *output level of the receiving set be kept constant*, or more correctly, within

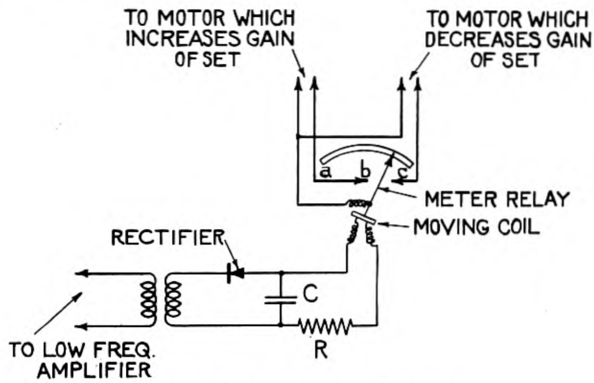



Fig. 1—Continuous recording system

certain narrow limits. In the receiver here described, the amplification of the receiving set is varied so as to satisfy this requirement, i.e., the gain is automatically cut down when the output level reaches its upper limit, and vice versa.

The output level of the set may be kept constant either by a continuously or by a discontinuously working system. The first method is mentioned here for comparison purposes only and is illustrated in Fig. 1. The figure shows that the rectified output current is sent through a moving coil relay whose pointer can only move between the contact points *b* and *c* while its real zero position is at *a*. Nothing happens if the pointer is between *b* and *c* but at the moment it touches *b* a motor will start, to increase the gain of the receiving set and will continue until the pointer is free again. Correspondingly the gain will be decreased if the pointer touches *c*. The purpose of the large resistance *R* and the condenser *C* is to prevent quick movements of the pointer caused by the individual static crashes. The time constant *RC* of the circuit would probably have to be at least five seconds and the speed at which the gain of the set is changed must be correspondingly slow. This system will, therefore, react very slowly for great changes in static level. Ordinarily static does not change very fast but if the recorder is working with a directive antenna system that is rotated say 360° in 20 minutes, then a fairly fast working recorder is desirable. The main disadvantage of the continuous system is that it will be difficult to make the meter give a true indication of the average energy level.

The method employed is therefore based upon a discontinuous system and will be described in connection with Fig. 2. The rectified output current of the set is integrated over a period of 10 seconds by means of a fluxmeter.[5] If, after these ten seconds the fluxmeter deflection is below a certain mark, then the gain of the set is increased *one* step and, vice versa, if the deflection is above a certain mark the gain is decreased *one* step. For deflections in between these marks the gain remains unchanged. To change the gain one step and to bring the fluxmeter needle back to zero takes approximately 5 seconds after which the whole process is repeated. The output energy due to static received during ten-second periods is here kept within two *definite limits.* The gain can be changed only one step after each period, but since each step corresponds to a change of 4 *TU* (1.58 times) in voltage gain it will take only one minute and a quarter for the recorder to adjust itself to a sudden change of 100 times in the energy level of static.

## The Apparatus of the Recorder

The receiving set is shown schematically in Fig. 2. It is an ordinary double detection set that requires altogether ten tubes, of which the last low frequency amplifier tube must be able to handle 10 watts in order to prevent overloading. The power supply may be rectified *AC.*

The gain control is inserted in the first intermediate frequency amplifier in order to be sure that no tubes are overloaded. The local oscillator shown is used for amplification calibration of the set and

[5] A galvanometer with negligible restoring torque, whose deflection is proportional to the coulombs sent through it. Its use for the present purpose was suggested by Mr. L. J. Sivian of the Bell Telephone Laboratories. He has employed the instrument for measurements of rectified speech and noise currents on telephone circuits. The use of a fluxmeter for similar purposes has been independently reported by Dr. E. M. Terry, of the University of Wisconsin, at the December 30, 1924, meeting of the U. R. S. I. at Washington, D. C.

requires no special shielding as its input voltage induced into the loop is comparatively large.

The selectivity of the set is determined by three separate units, viz., the antenna circuits, the intermediate frequency filter and the low frequency filter, each of which has a specific use. Carson and Zobel [6] have made the following statement.

"In filters designed to select a band of frequencies of width *w*, the ratio of energy transmitted through the network by the signal and by

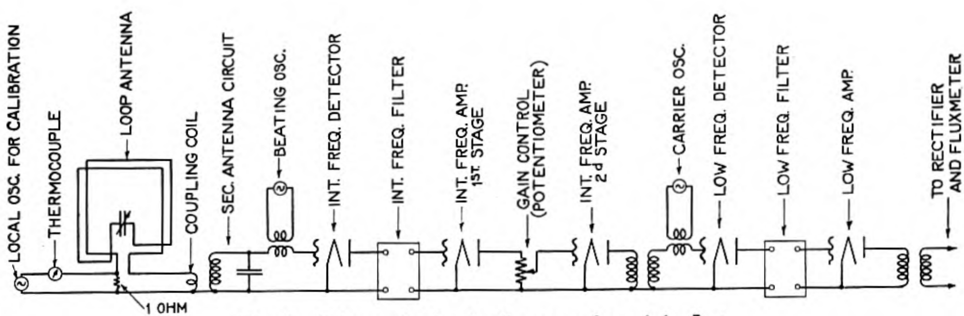



Fig. 2—Schematic circuit diagram of receiving set

random interference is inversely proportional to the band width and increases inappreciably when the number of sections is increased beyond two."

The main purpose of the filters is therefore not to define the frequency band of the set insofar as static is concerned, but to exclude continuous wave interference. It is hoped that 500 cycles wide frequency bands [7] can be maintained free of *c.w.* interference for static measurements and the simplest way to obtain such a band in the receiver is to make the low frequency filter an efficient low pass filter that cuts off every frequency above 600 cycles. More than two coupled circuits are hardly required in the antenna circuits, but the intermediate frequency filter ought to have sharper cut-off points than two coupled circuits will give. The selection of filters naturally depends upon the *c.w.* interference and it may in some cases be possible to reduce the number of filters and thereby make the recorder cheaper. The records shown later correspond to a frequency band of 2000 cycles—between 57.5 and 59.5 *k.c.*,—but it will probably not be long before *c.w.* interference makes it necessary to reduce this band

[6] "Transient Oscillators in Electric Wave-Filters"—John R. Carson and Otto J. Zobel, Bell System Technical Journal, Vol. II, No. 3, p. 27.

[7] Bands at 15, 30, 60, 120 . . . kilocycles would probably be satisfactory.

width. It is desirable to have a loud speaker connected to the output of the set and occasionally listen for *c.w.* interference.

The constant output control apparatus is shown in Fig. 3. The fluxmeter is seen in the upper right corner. Full deflection corresponds to 2 x $10^{-4}$ coulomb. The needle is normally free to move except when the cam *Z* presses the needle down until its point touches the scale *OS*. The shaft carrying the cam *Z* and the disc *N* is rotated

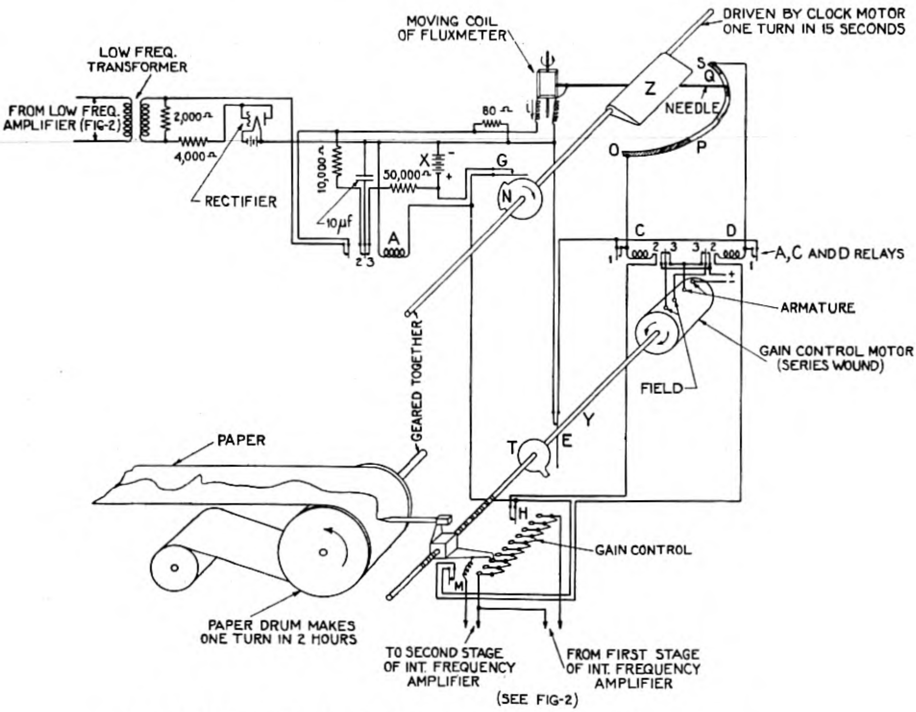



**Fig. 3—Schematic circuit of constant output control apparatus**

one complete turn in 15 seconds by a clock motor. The different elements are explained in the figure and the whole action may be understood by studying this carefully. However, it is probably worth while to go through a complete 15 second period and explain in detail the purpose of each part.

*Time in Seconds—0–10:*

Switch *G* is open, therefore relay *A* is open and no current can pass through the windings of relays *C* and *D*. (These relays start the gain control motor, which is therefore shut off.)

Contact 1 of relay *A* is closed and closes the circuit consisting of the secondary winding of the low frequency output transformer, the rectifier for the static currents and the fluxmeter. The 2000 and 4000 ohm resistances in this circuit insure distortionless input voltage to the rectifier. The fluxmeter is damped by an 80 ohm shunt. The needle, which was initially at zero, will therefore move, its deflection being proportional to $\int idt$.

*Time in Seconds—10–14:*

Switch $G$ is closed by the cam on the revolving disc $N$ and locks relay $A$.

Contact 1 of relay $A$ is opened and opens the rectifier fluxmeter circuit, thereby bringing the fluxmeter needle to a stop.

Contact 3 of relay $A$ is closed and makes the battery $X$ charge the 10 $\mu f$ condenser through the 50,000 ohm resistance.

*Time in Seconds—11–14:*

The cam $Z$ presses the needle point down on the scale $OS$. Now, one of three things will happen.

1. Static has decreased since the last period, so that the needle point will make contact with the metal strip $OP$ and close the following circuit: Battery $X$, needle of fluxmeter, winding of relay $C$, switch $H$ and switch $G$ to battery $X$. Relay $C$ is therefore closed and its closed contact 2, together with contact 3 of the open relay $D$ will start the gain control motor. After approximately half a turn of the gain control or motor shaft $Y$ the needle point is lifted from $OP$ by the rotation of the cam $Z$, but relay $C$ stays closed due to the fact that it is self-locking through its contact 1, so that the shaft $Y$ continues turning until the switch $E$ is opened by the disc $T$. This opens the self-locking circuit of relay $C$. Relay $C$ therefore opens and the gain control motor stops after the shaft $Y$ has made exactly one complete turn and increased the gain of the set one step (4 Transmission Units or 1.58 times). Notice that the opening of the needle point contact does not break any current, due to the use of self-locking relays. This preserves the needle point contact.

2. Static has not changed since the last period. The needle point will now touch the insulating strip $PQ$ and nothing else will happen, i.e., the gain of the set remains unchanged.

3. Static has increased since the last period so that the needle point now will make contact with the metal strip $QS$ and close relay $D$ and as in case 1 the motor will start and turn the shaft $Y$ one turn, but this time in the opposite direction, i.e., the gain of the set is decreased one step.

*Time in Seconds—14:*

Switch $G$ is opened again by the revolving disc $N$ and opens relay $A$. Contact 2 of relay $A$ is closed and will discharge the 10 $\mu f$ condenser through the fluxmeter, thereby bringing the needle back to zero. (Notice that the time constant of this discharge circuit is $10000 \times 10 \times 10^{-6} = 1/10$ seconds.)

*Time in Seconds—15:*

A new period has started.

The purpose of the switches $M$ and $H$ is to stop the motor when the gain control switch arm has reached the end of the scale.

The recorder is of such recent development that no comprehensive data are yet available.

Fig. 4 shows part of an actual record of static received on a set tuned to $57.5-59.5$ kilocycles. The ordinates represent the attenuation of the gain control of the set and it is to be remembered that the gain of the rest of the set is constant. The curve shows that the static power on the morning of October 30 changed more than 10,000 times. The point $B$ on the curve gives the effect of inducing a local

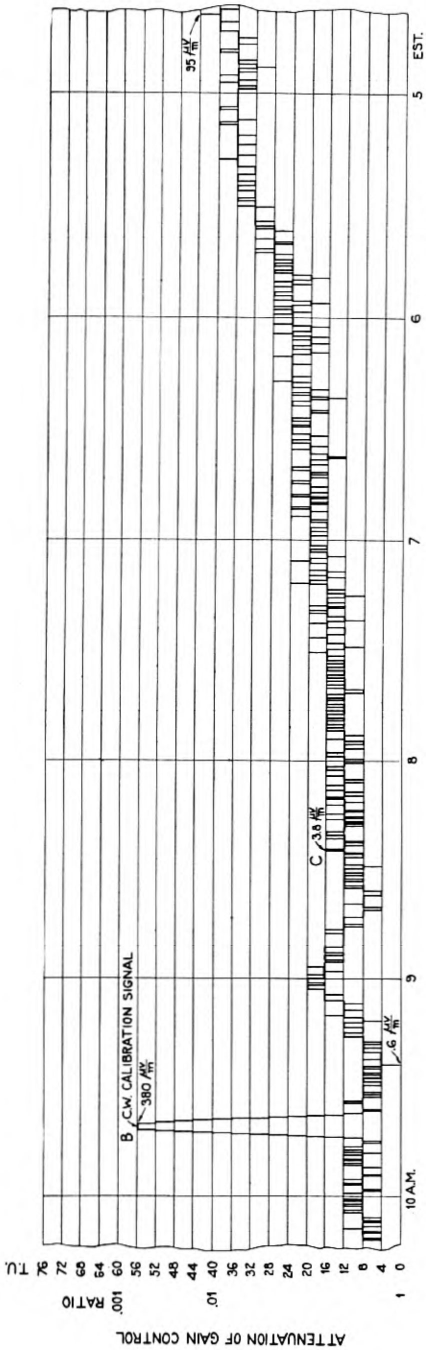



Fig. 4—Static record, morning of Oct. 30, 1925, Cliffwood, N. J., U. S. A.

signal of strength 380 $\mu v/m$ in the loop.[8] The point $C$ on the curve shows that at 8:25 A.M. the static intensity received on a 2000 cycle wide frequency band corresponded to the energy received from a *c.w.* signal of strength 3.8 $\mu v/m$. It would be practical always to relate static to such a *c.w.* signal. Experiments are now being conducted to determine whether the energy received from static is proportional to the width of the frequency band of the receiving set and if such is found to be the case then it is proposed to have the data relate to a 1000 cycle wide band. That static is, say, 7 microvolts per meter per kilocycle (7 $\mu v/mkc$) would then mean that the energy of the static received on a 1000 cycle wide frequency band is the same as the energy received from a *c.w.* signal of strength 7 $\mu v/m$.

Attempts have been made to calibrate the set by inducing in the loop, voltages of the shape shown in Fig. 5. Relating static to such signals would have the advantage of being independent of the band

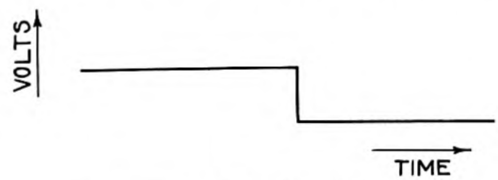



Fig. 5—Shape of impulse voltage

width of the set. Such signals were obtained by closing and opening a mercury switch, but one signal per second, or 10 impulses per period, would overload the set (the tubes) very much. At least 10 impulses per second would be required if the set should not be overloaded by each individual impulse, but this would be a difficult task to accomplish and it is therefore recommended that static be measured as explained above, by inducing a local *c.w.* singal into the loop. The fact that five static crashes in the course of 10 seconds—one period—does not overload the set while 100 impulses of the shape shown in Fig. 5 are required to prevent overloading gives us some interesting information on static. It shows that a single static crash is not a single sudden change of the field in the ether and that it cannot be represented by less than 20 consecutive impulses.

The record of Fig. 4 shows that each step on the gain control potentiometer is 4 $TU$ and the selection of such steps and of 15 seconds will now be discussed. To decrease the 4 $TU$ step to a 1 $TU$ step would

[8] It may be worth while to have such a calibration signal introduced automatically for instance once every two hours.

decrease the speed of the set, i.e., it would take four times longer for the recorder to register a sudden change in the static level which is particularly a disadvantage when the recorder is connected to a rotating directional antenna. On the other hand a step larger than 4 *TU* would not give the static level with sufficient accuracy. If the time periods are changed from 15 to 10 seconds, then the "speed" of the set is increased, but the set is then inoperative over a larger part of the period since it takes 5 seconds to change the gain of the set and bring the fluxmeter needle back to zero. Besides, such a decrease in time period would increase the probability of overloading and also it would make the energy received per period vary more irregularly especially if static consisted of separate crashes.

## Directional Static Recording

The usefulness of a static recorder will naturally be increased many times if it is able to indicate the general direction from which the static comes. A rotating loop antenna would give some results, but it would be still better to combine the rotating loop with an ordinary open antenna so as to obtain the well-known heart-shape directional characteristic. The loop should be rotated by the clock-motor of the set (see Fig. 3), say 2 complete turns in one hour, and the abscissa on the record would then require a direction scale in addition to the time scale.

# Directive Diagrams of Antenna Arrays

By RONALD M. FOSTER

SYNOPSIS: Two systematic collections of directive amplitude diagrams are shown for arrays of 2 and of 16 identical antennae spaced at equal distances along a straight line with equal phase differences introduced between the currents in adjacent antennae, assuming that each antenna radiates equally in all directions in the plane of the diagram. Three diagrams show the effect of increasing without limit the number of antennae in a given interval. Two models show the effect of distributing the antennae over an area.

## INTRODUCTION

ONE of the means proposed for obtaining directive radio effects, both in sending and in receiving, is the antenna array, consisting of a system of two or more antennae situated at specified fractions of a wave-length apart and with relations imposed upon the amplitudes and phases of the currents in the several antennae. For example, consider a sending array consisting of two vertical antennae so arranged that the currents in the antennae are equal in magnitude but a half period apart in phase, the individual antennae being identical and radiating equally in all directions in the horizontal plane. If the two antennae were placed at the same point there would be zero transmission in all directions, since the effects of the two antennae would neutralize each other. If, however, the two antennae are separated by a small fraction of a wave-length, while there will still be zero transmission in the direction perpendicular to the axis of the array, there will be transmission in all other directions. If this separation is increased to exactly one-half of a wave-length, the radiation from the array along the axis will become a maximum.

This particular type of antenna array was proposed by Brown [1] in 1899. A few years later, Stone [2] proposed a similar array with the two currents exactly in phase. This array gives maximum transmission perpendicular to the axis, zero transmission along the axis. About the same time, Blondel [3] made several suggestions, among them, two antennae placed a quarter of a wave-length apart and with a phase difference of a quarter of a period. With this arrangement a unilateral effect is obtained, there being maximum transmission in one direction along the axis, zero transmission in the opposite direction.

[1] S. G. Brown, British Patent No. 14,449 (1899).

[2] J. S. Stone, United States Patent No. 716,134 (1901).

[3] A. Blondel, Belgian Patent No. 163,516 (1902), British Patent No. 11,427 (1903).

These early suggestions have been followed by a number of more complicated arrangements proposed by Braun,[4] Bellini,[5] and others.[6]

Most books on radio communication contain one or more directive diagrams showing the variation of either the amplitude or the energy for systems of two, three, or four antennae which are separated by given fractions of the wave-length and with currents which have assigned amplitude and phase relations. Bellini,[7] Koerts,[8] and Zenneck [9] each give about a dozen such diagrams. Recently, Green [10] and Friis [11] have published directive diagrams for a pair of loops; the latter gives a curve obtained experimentally which agrees very well with the theoretical curve. Of the published diagrams, one of the most extended and systematic sets appears to have been that of Walter.[12] He showed a total of 21 diagrams for arrays of two antennae, with the three separations of 1/10, 1/4, and 1/2 wave-length and the seven phase differences of 0, 1/12, 1/6, 1/4, 1/3, 5/12, and 1/2 period.

In the present paper an effort has been made to present a more systematic and comprehensive collection of directive diagrams for arrays consisting of 2 and of 16 antennae, respectively, spaced at equal distances along a straight line or axis, with currents of equal amplitude in all the antennae, and with equal phase differences introduced between the currents in adjacent antennae. These diagrams are polar diagrams showing the relative amplitude of the field of the radiation at a great distance in a plane through the array, assuming that each antenna radiates equally in all directions in this plane. The unit circle shown in each diagram represents the amplitude of the radiation if all the antennae were made coincident in space and in phase.

These directive diagrams may be used to obtain the directive diagram in any plane through an array made up of antennae which

[4] F. Braun, Electrician, 57, pages 222–224, 244–248, 1906.

[5] E. Bellini, Electrician, 74, pages 352–354, 1914.

[6] For extensive bibliographies see L. H. Walter, Directive Wireless Telegraphy, London, 1921, pages 119–121; H. H. Beverage, C. W. Rice, and E. W. Kellogg, Journal of the A. I. E. E., 42, pages 736–738, 1923; A. Koerts, Atmosphärische Störungen in der drahtlosen Nachrichtenübermittlung, Berlin, 1924, pages 149, 150; J. Zenneck and H. Rukop, Drahtlose Telegraphie, fifth edition, Stuttgart, 1925, Chapter XIII.

[7] E. Bellini, Jahrbuch der drahtlosen Telegraphie und Telephonie, 2, pages 381–396, 1909.

[8] A. Koerts, *loc. cit.*, pages 101, 102, 104, 105, 110, 111, 130, 131, 133.

[9] J. Zenneck and H. Rukop, *loc. cit.*, pages 412–415, 419, 421, 423, 428, 432.

[10] E. Green, Experimental Wireless, 2, pages 828–837, 1925.

[11] H. T. Friis, Proceedings of the I. R. E., 13, pages 685–707, 1925.

[12] L. H. Walter, Electrician, 64, pages 790–792, 1910.

do not radiate equally in all directions in this plane, but which satisfy the other conditions named above; the total directive effect is the product of the individual effect multiplied by the group effect. Thus, since the amplitude of the radiation in the horizontal plane from a single vertical loop varies as the cosine of the angle between the direction of transmission and the plane of the loop, the directive diagram in the horizontal plane of an array of loops, all loops being oriented the same, is the corresponding directive diagram of an antenna array as presented in this paper, with the radius vector multiplied by a cosine factor.

The present discussion has been stated in terms of transmission, but the directive diagrams apply equally well to the case of reception by an array from a distant source.

Each of the two sets of diagrams is presented in a rectangular arrangement so as to exhibit the effect of changes both in the separation between adjacent antennae, specified in wave-lengths, and in the phase difference introduced between the currents in adjacent antennae, specified in periods. These drawings were originally made at the suggestion of Dr. G. A. Campbell to illustrate the application of antenna arrays as a means for reducing the ratio of static to signal.

The antenna array is analogous to the optical diffraction grating. By eliminating the transmission wires connecting together the individual antennae of the array and utilizing instead re-radiation from suitably designed antennae, the radio system would correspond more closely with this optical analogue. With this arrangement, however, the phase difference cannot exceed a value in periods numerically equal to the separation in wave-lengths, a restriction to which the ordinary ruled grating is also subject. The retardation grating proposed by Rayleigh [13] and the echelon spectroscope of Michelson [14] offer more complete analogies to the antenna array in that there is no theoretical limitation on the separation and phase difference.

### Two Antennae, Fig. 1

A total of 90 directive diagrams for an array of two antennae is shown by Fig. 1. The separation between antennae varies from 0 to 2 wave-lengths, in steps of 1/8 wave-length; the phase difference between antennae varies from 0 to 1/2 period, in steps of 1/8 period; an additional set of diagrams is included with a separation of 4 wave-lengths. These curves were carefully drawn with a unit circle ten

[13] Rayleigh, Collected Papers, 3, pages 106–116.

[14] A. A. Michelson, Astrophysical Journal, 8, page 37, 1898.

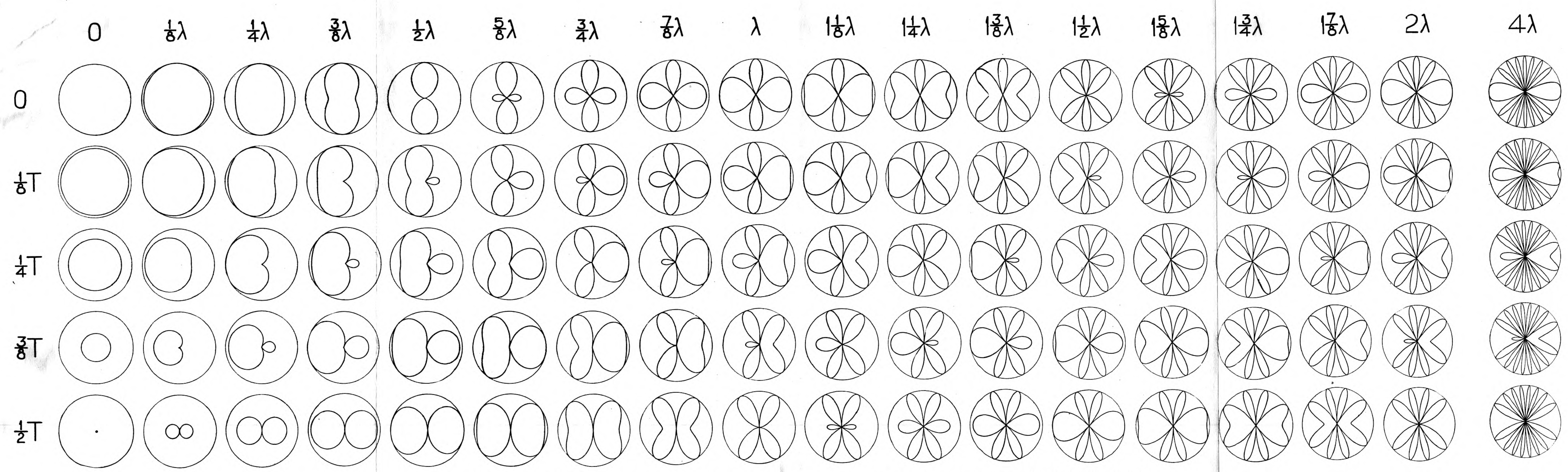



Fig. 1—Directive amplitude diagrams for an array of two antennae; separation in wave-lengths (λ) along the top, phase difference in periods ($T$) at the left

inches in diameter, so that, on the reduced scale of reproduction, the accuracy should leave nothing to be desired. For a sending array the specified phase difference is the lag of the current in the right-hand antenna behind the current in the left-hand antenna; for a receiving array it is the lag introduced in the current from the right-hand antenna. In each case the line of the array is parallel to the horizontal axis of the diagram.

Reversing the sign of the phase difference reflects the directive diagram about its vertical axis, that is, the right and left sides are interchanged. With increasing phase difference the diagrams repeat cyclically, those from $\frac{1}{2}T$ to $\frac{3}{2}T$ being the same as those from $-\frac{1}{2}T$ to $\frac{1}{2}T$, and so on.

In the first column, that is, for zero separation, all diagrams are circles, since the two antennae are coincident, and thus the array radiates uniformly in all directions. For zero phase difference the directive diagram is the unit circle, since the radiations from the two antennae reenforce each other without interference. As the phase difference is increased, this circle grows smaller, due to increasing interference, until, for a phase difference of a half period, the two radiations completely neutralize each other, the directive diagram shrinking down to a null circle.

The diagrams in the first row, that is, for zero phase difference, are symmetrical about the vertical axis in addition to being symmetrical about the horizontal axis. In every case the amplitude is unity along the vertical axis, that is, in a direction perpendicular to the line of the array. As the separation is increased from zero, the amplitude along the horizontal axis diminishes, until it reaches zero for a separation of $\frac{1}{2}\lambda$, it then increases to unity at $\lambda$, it diminishes to zero at $1\frac{1}{2}\lambda$, it reaches unity at $2\lambda$, and so on.

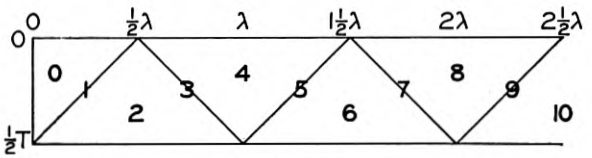
The diagrams in the bottom row, that is, for a phase difference of a half period, are also symmetrical about the vertical axis. In every case the amplitude is zero along the vertical axis. For small separations, the directive diagram is approximately a pair of tangent circles, which increase in size as the separation is increased. When the separation reaches $\frac{1}{2}\lambda$, the amplitude along the horizontal axis reaches unity, it then falls off to zero as the separation is increased to $\lambda$, it rises to unity at $1\frac{1}{2}\lambda$, it falls to zero at $2\lambda$, and so on.

The diagram for $(\frac{1}{4}\lambda, \frac{1}{4}T)$ is particularly interesting in that there is a single direction of unit amplitude with zero amplitude in the opposite direction. This array was proposed by Blondel, as stated above, and it is the basis of the Alexanderson barrage.[15] The diagrams

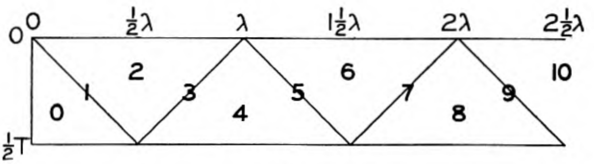
[15] E. F. W. Alexanderson, Proceedings of the I. R. E., 7, pages 363–378, 1919.

situated on the line from $(0\lambda, \frac{1}{2}T)$ to $(\frac{1}{4}\lambda, \frac{1}{4}T)$ have a similar property: each has a relative maximum in a single direction, with zero in the opposite direction. The diagrams on the line from $(0\lambda, 0T)$ to $(\frac{1}{2}\lambda, \frac{1}{2}T)$ have a maximum along the horizontal axis to the left, with an amplitude to the right decreasing from unity at $0T$ to zero at $\frac{1}{4}T$, and then increasing to unity at $\frac{1}{2}T$.

The number of lobes tends to increase as the separation is increased, as shown by (a) of Fig. 2. A zigzag starting at $(0\lambda, \frac{1}{2}T)$




(a)




(b)

Fig. 2—(a) Number of null directions (which is also the number of lobes) for two antennae. (b) Number of unit directions (directions of absolute maximum amplitude) for any number of antennae, in terms of separation and phase difference between adjacent antennae

and made up of lines sloping up and down at an angle of 45° divides the rectangular arrangement of diagrams into sections with 0, 2, 4, 6, . . . null directions in each diagram, respectively. On these lines the number of null directions is 1, 3, 5, 7, . . . , respectively, with the intermediate numbers at the junction points. The number of lobes is, of course, equal to the number of null directions.

Part (b) of Fig. 2 is a diagram specifying the number of unit directions (directions of absolute maximum amplitude) in terms of the separation and the phase difference between adjacent antennae, and it holds regardless of the number of antennae, that is, the number and position of the main lobes are not changed by increasing the number of antennae, provided the same separation and phase difference are preserved between adjacent antennae.

It is interesting to observe the variation in the diagrams along any line in this rectangular arrangement of Fig. 1, whether horizontal, vertical, or diagonal. A lobe starts as a small bud, it grows in size until it reaches the unit circle, it then becomes dented; the two prongs of the lobe separate more and more until a division into two lobes takes place; then these lobes separate as a new lobe starts to grow

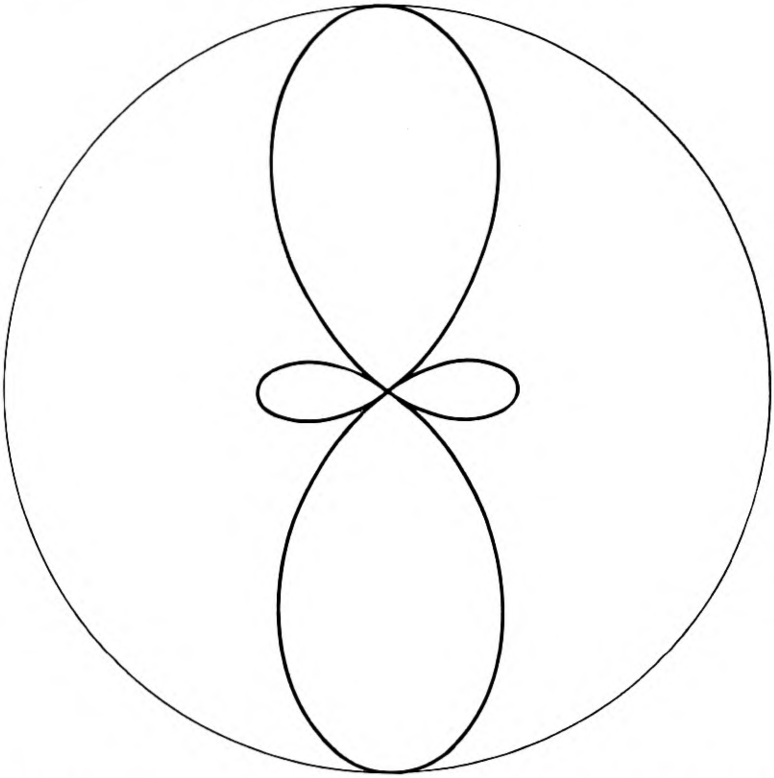


Fig. 3—Directive amplitude diagram for two antennae (0.6098λ, 0$T$) having the minimum area (0.2986) relative to the unit circle

between them. The additional column of diagrams for a separation of 4λ still further illustrates the way in which the lobes multiply and narrow as the separation between the two antennae in increased.

The area of the polar directive diagram of an array relative to the area of the unit circle is a measure of the reduction in the energy ratio of random static to signal for that array, assuming that the signal comes from a direction in which the radius vector of the diagram is unity while the static is uniformly distributed. All the diagrams for a phase difference of 1/4 period have an area of 1/2.

The area of the other diagrams oscillates about 1/2 and approaches it as a limit upon increasing the separation and keeping the phase difference constant. The minimum area (for a diagram in which the radius vector reaches its maximum of unity [16]) is 0.2986, obtained

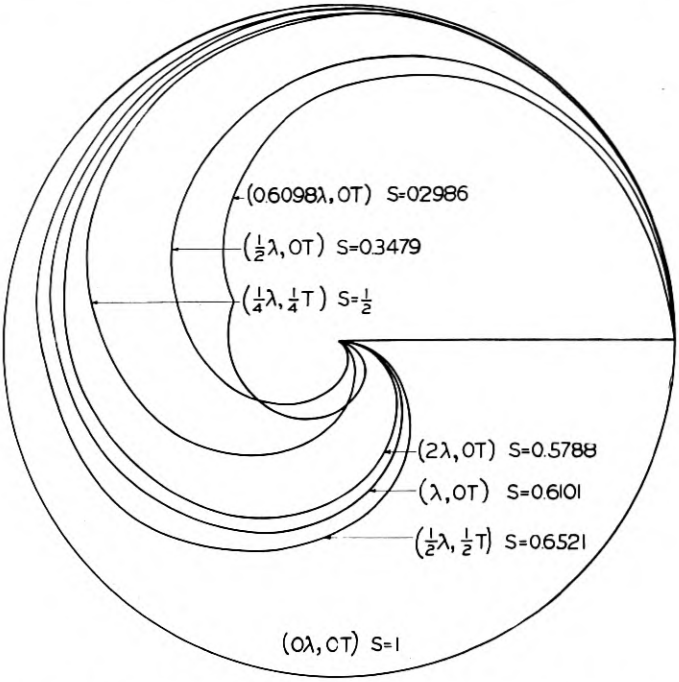



Fig. 4—Cumulative amplitude diagrams for two antennae

by the array (0.6098λ, $0T$). The directive diagram for this case is shown by Fig. 3.

Cumulative amplitude diagrams are shown by Fig. 4 for six selected arrays, including the unit circle, which is the cumulative diagram for the array (0λ, $0T$). In this figure, the angle corresponding to any value of the radius vector is equal to the total angle of the directive diagram of the array throughout which the relative amplitude

[16] A. Koerts, *loc. cit.*, pages 104, 105. In those cases in which the radius vector does not reach the unit circle, in order to obtain a measure of the reduction of the energy ratio of random static to signal, the unit circle should be replaced by the circle with a radius equal to the maximum radius vector. The absolute minimum 0.2986 is not changed upon including these cases but a relative minimum 1/3 occurs for the array ($a\lambda$, $bT$) upon letting $a$ and $b$ approach 0 and 1/2 in such a manner that $a+2b=1$. If the two antennae are loops, with planes parallel to the axis of the array, the area approaches the relative minimum 3/14 upon letting $a$ and $b$ approach the same limits 0 and 1/2 but in such a manner that $3a+4b=2$.

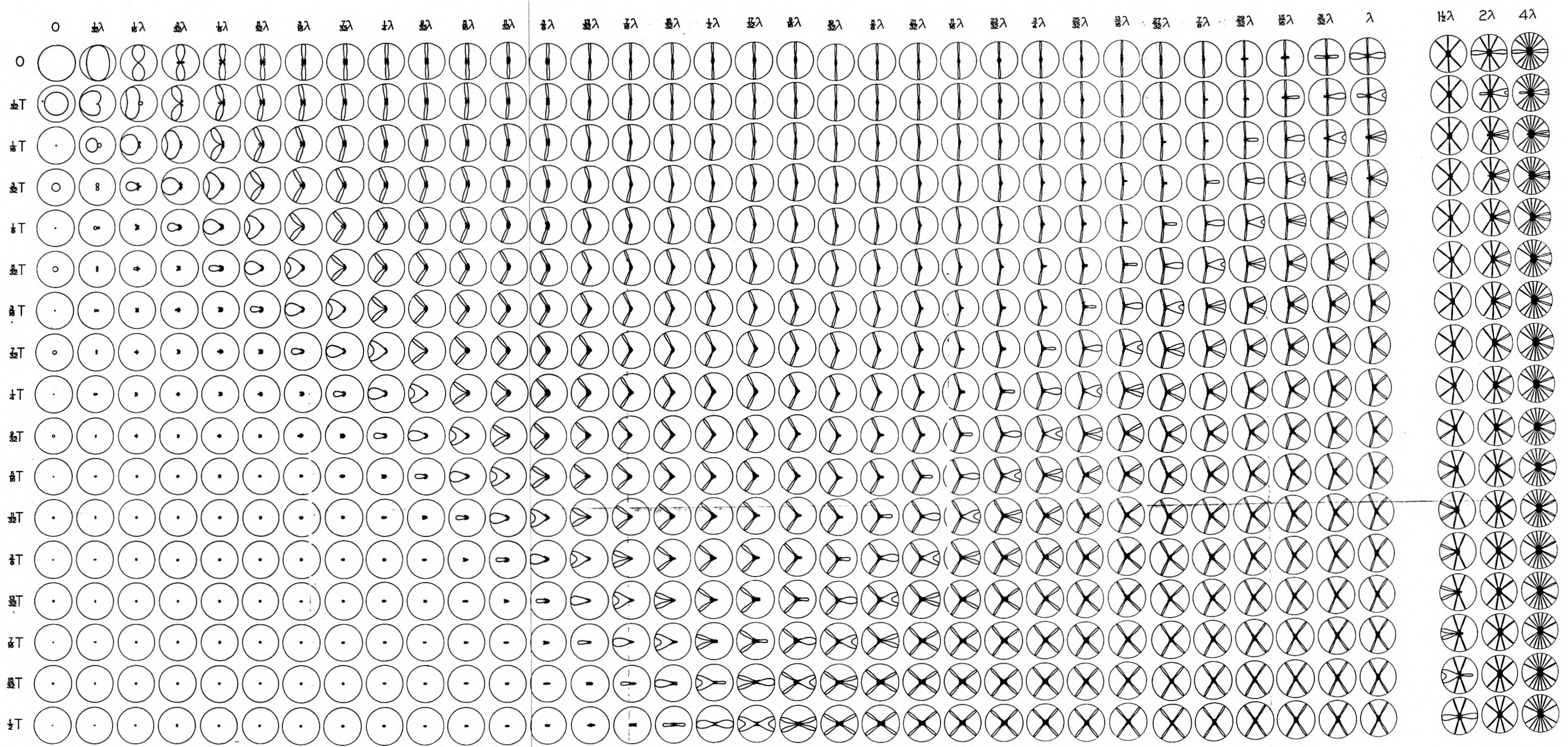


Fig. 5—Directive amplitude diagrams for an array of sixteen antennae; separation in wave-lengths (λ) along the top, phase difference in periods (T) at the left

is equal to or greater than this value. The area ($S$) relative to the unit circle is given in Fig. 4 for each of the cumulative diagrams, this area being equal to the area of the corresponding directive diagram.

### Sixteen Antennae, Fig. 5

A total of 612 directive diagrams for an array consisting of sixteen antennae is shown by Fig. 5. The separation between adjacent

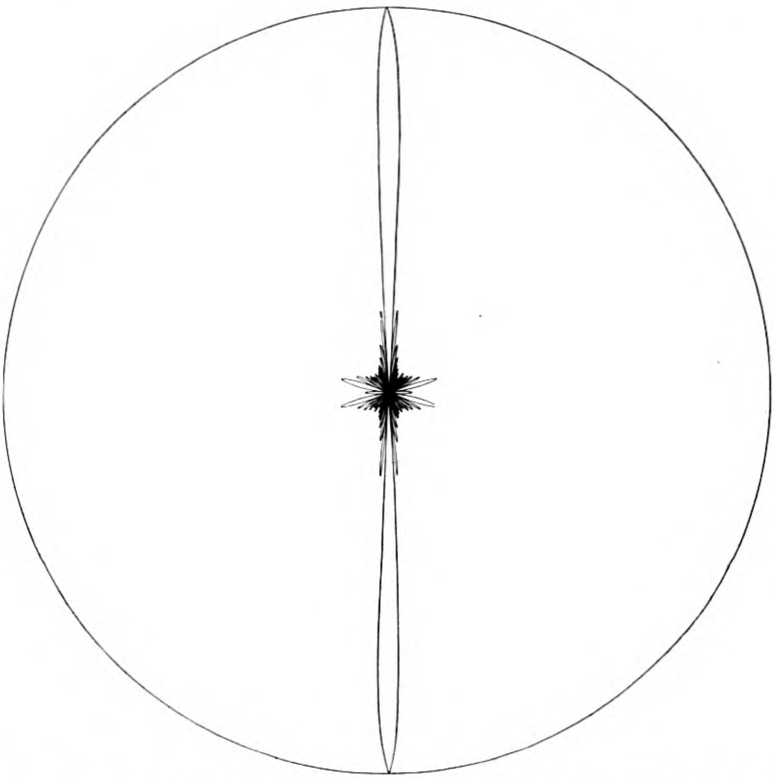


Fig. 6—Directive amplitude diagram for sixteen antennae (0.8825λ, $0T$) having the minimum area (0.0254) relative to the unit circle

antennae varies from 0 to 1 wave-length, in steps of 1/32 wave-length; the phase difference between adjacent antennae varies from 0 to 1/2 period, in steps of 1/32 period; additional sets of diagrams are included with separations of $1\frac{1}{2}$, 2, and 4 wave-lengths. The specified phase difference is the lag of the current in one antenna behind the current in its left-hand neighbor. The diagrams are reflected about the vertical axis upon changing the sign of the phase difference, and they repeat cyclically with increasing phase difference. These curves were copied from the original drawing; detailed accuracy is not claimed, but the arrangement and relative sizes of the lobes are approximately correct.

Comparison of Figs. 1 and 5 shows that, for the same set of parameters, the main features of the two diagrams are similar, that is,

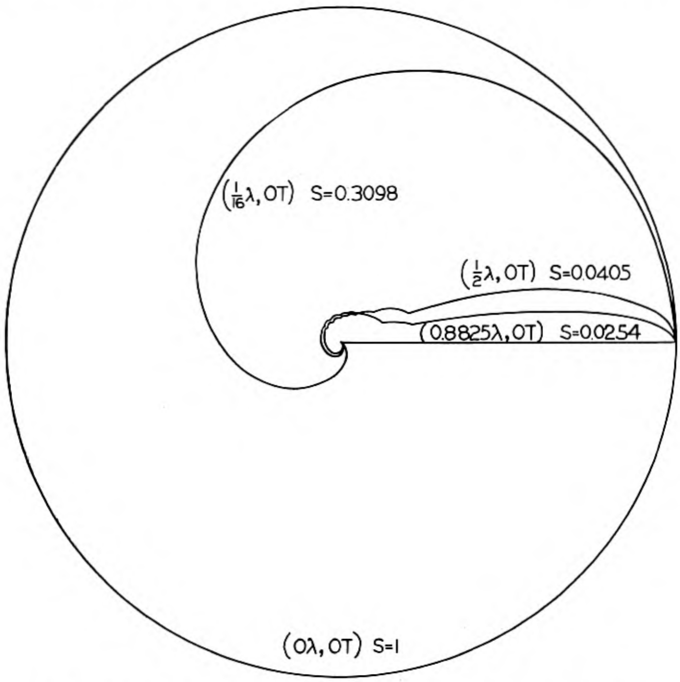



Fig. 7—Cumulative amplitude diagrams for sixteen antennae

the main lobes are located in the same positions, but with 16 antennae the main lobes are much narrower and in addition a multiplicity of small lobes occurs in many cases.

The area of the directive diagrams for 16 antennae oscillates about 1/16 and approaches it as a limit upon increasing the separation, keeping the phase difference constant. The minimum area (for a diagram in which the radius vector reaches its maximum of unity) is 0.0254, obtained by the array $(0.8825\lambda, 0T)$. The directive diagram for this case is shown by Fig. 6. Cumulative amplitude diagrams for 16 antennae are shown by Fig. 7 for three selected arrays in addition to the array $(0\lambda, 0T)$, the area of each diagram being given on the drawing.

## Infinite Antenna Arrays

In view of the points of similarity between the diagrams for 2 and for 16 antennae, in particular for those pairs of diagrams for which the parameters of Fig. 1 are eight times the parameters of Fig. 5, provided the latter are relatively small, the question naturally arises as to the effect of increasing the number of antennae without limit.

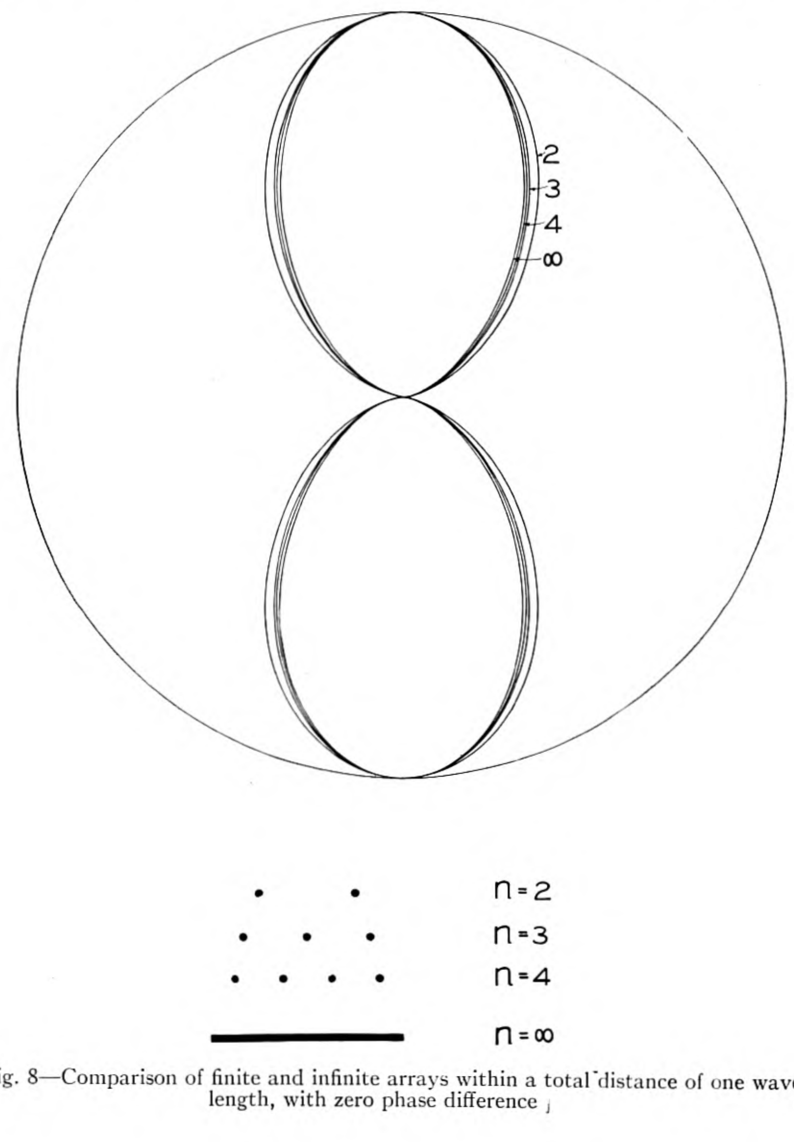



Fig. 8—Comparison of finite and infinite arrays within a total distance of one wave-length, with zero phase difference

The similarity among the diagrams for arrays of antennae within a given fixed interval is illustrated by Figs. 8, 9, and 10.

In Fig. 8 are shown diagrams for 2, 3, and 4 antennae situated within a total distance of one wave-length with separations of 1/2, 1/3, and 1/4 wave-length, respectively, and with zero phase difference between adjacent antennae. The curve ($\infty$) gives the limit of the family of directive diagrams for arrays of $n$ antennae with a separation of $1/n$ wave-length, as $n$ becomes infinite.

Fig. 9 gives similar curves for arrays of $n$ antennae within an interval of two wave-lengths, with the parameters $\left(\frac{2}{n}\lambda,\ 0T\right)$. Fig. 10 shows a similar set of diagrams for arrays within an interval of one wave-length and within a total phase interval of one period, that is, for arrays of $n$ antennae with the parameters $\left(\frac{1}{n}\lambda,\ \frac{1}{n}T\right)$.

For any interval $(A\lambda,\ BT)$ a similar family of curves can be obtained for arrays of $n$ antennae with the parameters $\left(\frac{A}{n}\lambda,\ \frac{B}{n}T\right)$. As the number $n$ is increased without limit, the directive diagram approaches a limiting curve. This limiting curve never has more than two directions of unit amplitude. There are zero, one, or two such directions, depending upon whether $A$ is less than, equal to, or greater than $B$. The diagrams for the infinite case are reflected about the vertical axis upon changing the sign of the phase difference, but they do not repeat cyclically with increasing phase difference.

The rapidity with which the diagrams approach this limiting curve as $n$ is increased is well illustrated by Figs. 8, 9, and 10. On the scale of these drawings, the curves for 16 antennae would be indistinguishable from the limiting curves for the infinite case. The upper left-hand corner of Fig. 5 may thus serve as a chart of the directive diagrams for the infinite case if the column and row headings are multiplied by the factor 16 to give the total separation and phase difference of the interval. For larger values of these parameters, however, the curves for the infinite case depart more and more from those of Fig. 5.

The diagrams with $A=B$ are of particular interest since these are unilateral, with the main lobe growing narrower as the total separation and phase difference are increased. In the case of the Beverage antenna, the ideal system [17] consists essentially of a long loop, which we may think of as the limiting case of a succession of a large number of narrow loops. The directive diagram of such an antenna system

[17] H. H. Beverage, C. W. Rice, and E. W. Kellogg, *loc. cit.*, pages 372, 373.

would, therefore, be the product of the group curve for an infinite number of antennae in the given interval multiplied by a cosine factor for the individual narrow loop.

## Space Characteristics

When the antennae are not confined to a straight line but are distributed over an area, a surface with its radius vector propor-

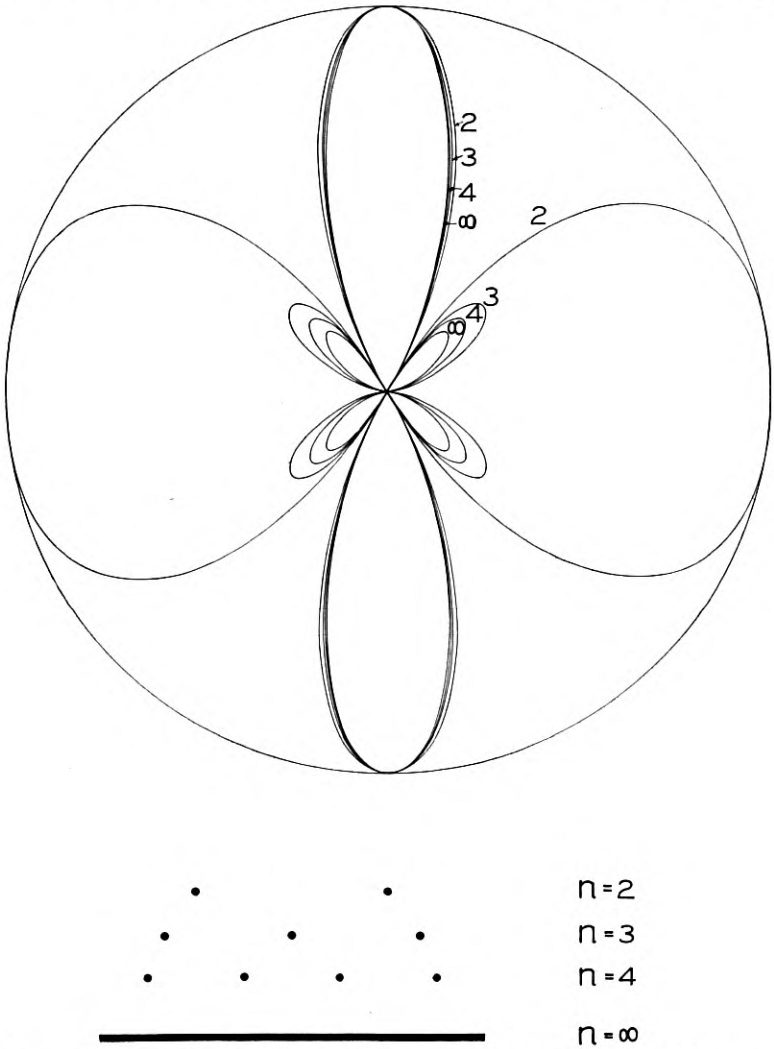



Fig. 9—Comparison of finite and infinite arrays within a total distance of two wavelengths, with zero phase difference

tional to the amplitude of the field of the radiation at a great distance from the array in the direction of the radius vector is required. Two particular cases will be illustrated in order to give some idea of the surface which shows the group effect of the array; for actual antennae, the radius vector of this surface must be multiplied by the corresponding radius vector of the space characteristic of the individual antenna in order to obtain the actual characteristic of the array.

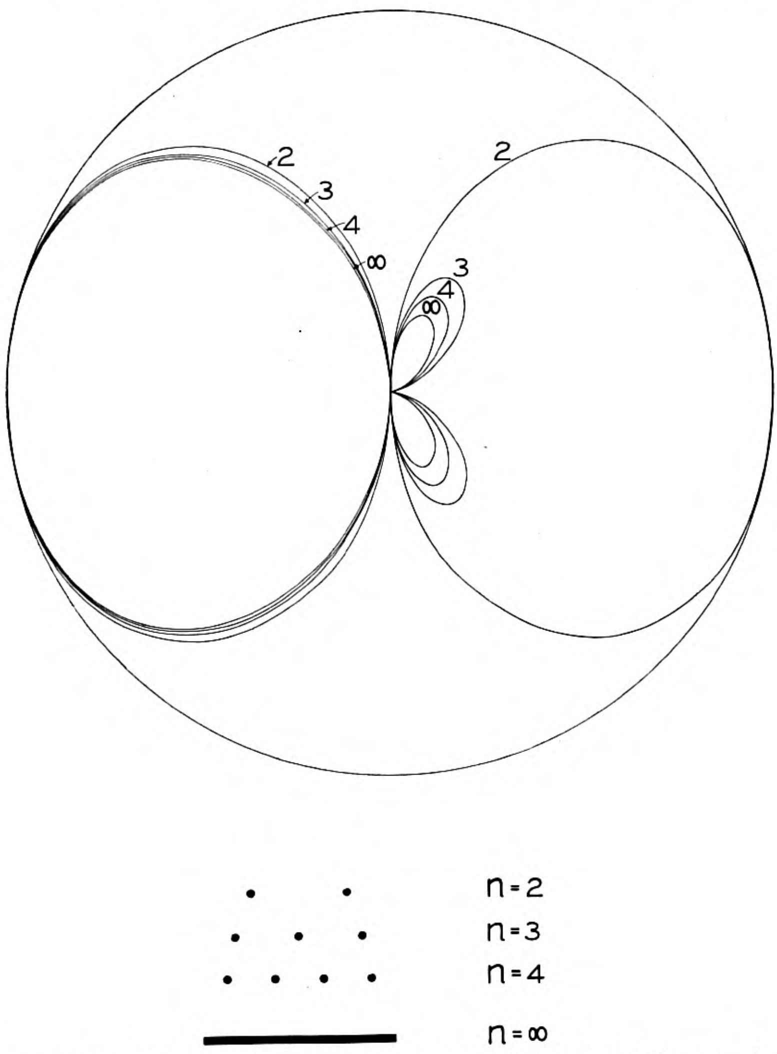



Fig. 10—Comparison of finite and infinite arrays within a total distance of one wave-length, and within a total phase interval of one period

A model of the upper half of the space characteristic of an array of four antennae located at the corners of a square is shown by Fig. 11, each side of the square having the parameters ($\frac{1}{2}\lambda$, $0T$). Below the model is shown the directive diagram in the plane of the array, which

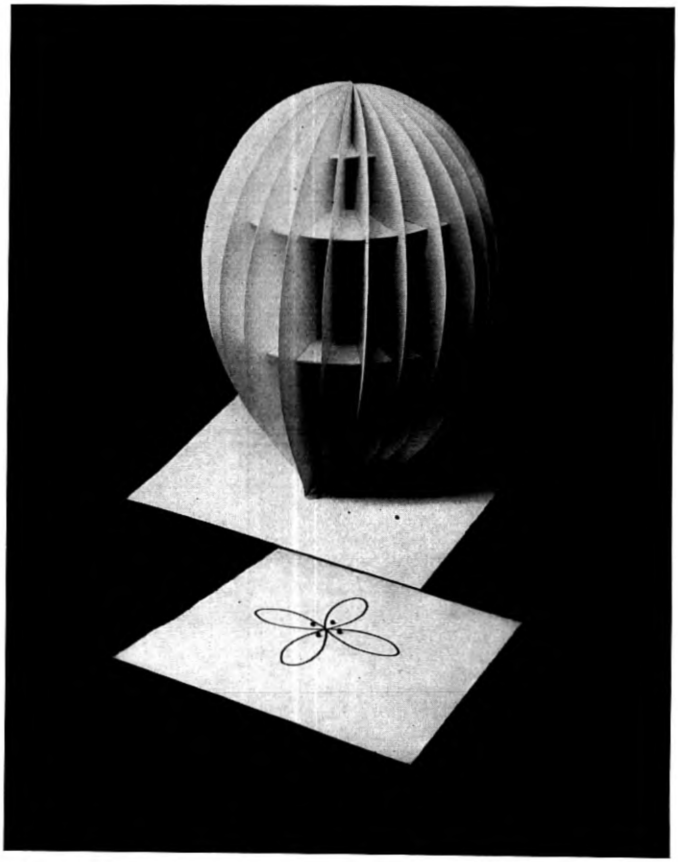


Fig. 11—Model of the space characteristic for an array of four antennae located at the corners of a square, with a separation of one-half wave-length between antennae on each side of the square, and with zero phase difference

is identical with the base of the model, together with a representation of the array itself. In the horizontal plane the maximum amplitude is slightly less than 1/5, occurring along the diagonals of the square; the amplitude reaches its absolute maximum of unity only in the vertical direction.

Fig. 12 shows a model of the upper half of the space characteristic of an array of 32 antennae located along the diagonals of a square, with the parameters ($\frac{1}{2}\lambda$, $0T$) in each diagonal. This space characteristic is a complicated surface, the main features of which are shown by the model, the smaller lobes not being shown clearly in detail. In

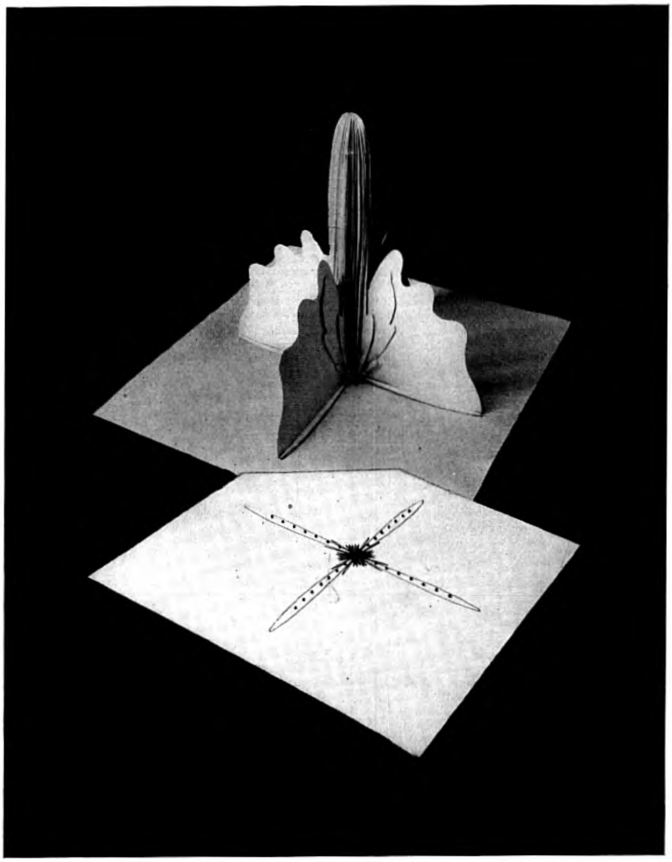


Fig. 12—Model of the space characteristic for an array of 32 antennae located along the diagonals of a square, with a separation of one-half wave-length between adjacent antennae in each diagonal, and with zero phase difference

the horizontal plane the maximum amplitude is 1/2, occurring along the diagonals of the square; the amplitude reaches a pronounced maximum of unity, however, in the vertical direction.

I am greatly indebted to Dr. Louisa E. Townshend for supervising the preparation of the drawings and models, and expecially for the accuracy attained in redrawing Fig. 1.

## Appendix

Formulae for the directive diagrams of this paper are conveniently expressed in terms of polar coordinates, as follows: For a linear array of $n$ antennae with the parameters $(a\lambda, bT)$,

$$r = \left| \frac{\sin n\,(\pi a \cos\theta + \pi b)}{n \sin\,(\pi a \cos\theta + \pi b)} \right|, \tag{1}$$

where $\theta$ is measured from the axis of the array. The area of this diagram, relative to the unit circle, is

$$S = \frac{2}{n^2}\left(\frac{n}{2} + \sum_{k=1}^{n-1} (n-k) J_0(2\pi ka) \cos\,(2\pi kb)\right). \tag{2}$$

For the special case $n=2$, Fig. 1, formula (1) reduces to

$$r = |\cos\,(\pi a \cos\theta + \pi b)|, \tag{3}$$

and formula (2) for the area to

$$S = \tfrac{1}{2}(1 + J_0(2\pi a) \cos\,(2\pi b)). \tag{4}$$

For the special case $n=\infty$ in a total interval $(A\lambda, BT)$, the limit of formula (1) for $a=A/n$ and $b=B/n$, as $n$ becomes infinite, is

$$r = \left| \frac{\sin\,(\pi A \cos\theta + \pi B)}{\pi A \cos\theta + \pi B} \right|. \tag{5}$$

For the array of Fig. 11,

$$r = |\cos\,(\tfrac{1}{2}\pi \cos\theta \cos\phi) \cos\,(\tfrac{1}{2}\pi \sin\theta \cos\phi)|, \tag{6}$$

where $\phi$ is the angle which the radius vector makes with the plane of the array, and $\theta$ the angle which the projection of the radius vector in this plane makes with one side of the square. For the array of Fig. 12,

$$r = \left| \frac{\sin\,(8\pi \cos\theta \cos\phi)}{32 \sin\,(\frac{1}{2}\pi \cos\theta \cos\phi)} + \frac{\sin\,(8\pi \sin\theta \cos\phi)}{32 \sin\,(\frac{1}{2}\pi \sin\theta \cos\phi)} \right|, \tag{7}$$

where $\phi$ and $\theta$ are the same as for formula (6) except that the latter is measured from one diagonal of the square.

# Correction of Data for Errors of Averages Obtained from Small Samples

By W. A. SHEWHART

SYNOPSIS: Recent contributions to the theory of statistics make possible the calculation of the error of the average of a small sample—something that cannot be done accurately with customary error theory. Obviously, these contributions are of very general importance, because experimental and engineering sciences alike rest upon averages which in a majority of cases are determined from small samples, and because an average cannot be used to advantage without its probable error being known.

The present paper attempts to show in a simple way why we cannot use customary error theory to calculate the error of the average of a small sample and to show what we should use instead. The points of interest are illustrated with actual data taken for this purpose. The paper closes with applications of the theory to four types of problems involving samples of small size for each of which numerous examples arise in practice. These types are:

1. Determination of error of average.
2. Determination of error of average difference.
3. Determination of most probable value of the root mean square deviation of the universe when only one sample of $n$ pieces has been examined.
4. Determination of most probable value of the root mean square deviation of the universe when several samples of $n$ pieces each have been examined.

## USEFUL THEORY OVERLOOKED: WHY?

PRACTICALLY everyone uses averages—research workers and engineers in particular. Moreover, all of us have long appreciated the fact that an average is often only of value when we know its probable error. Naturally, we turn to the theory of errors to guide us in calculating the probable error. Naturally, because from 1733 to 1908 there was nothing else that we could turn to. Since 1908 the recognition has been gradually making headway that to use customary error theory for determining the probable errors of averages of small samples is a mistake.

The story of how to calculate the probable error of a small sample was originally told in Biometrika, a journal for the statistical study of biological problems—a veritable mine of useful information. The truth was given in equations involving terms familiar only to statisticians and hence was concealed from many. The story, however, with the aid of such experimental results as are used in this paper can be told in a simple manner: it is of interest to all of us who, for one reason or another, cannot make large numbers of observations on every quantity that we measure, but must nevertheless estimate the probable errors of our results. In this discussion, diagrams will be used instead of equations, and, because of this rather popular presentation, many readers may want to consult, as the original sources, the intensely interesting mathematical contributions of "Student",[1] Professor Karl Pearson,[2] and R. A. Fisher.[3]

## Case Where Customary Theory Applies

We start, as in customary error theory, with the assumption that the probability distribution of errors is normal. This simply means that the probability of the occurrence of an error within any range is assumed to be equal to the area under the so-called normal curve[4] (such a curve is shown in Fig. 1) between the limits of the same range.

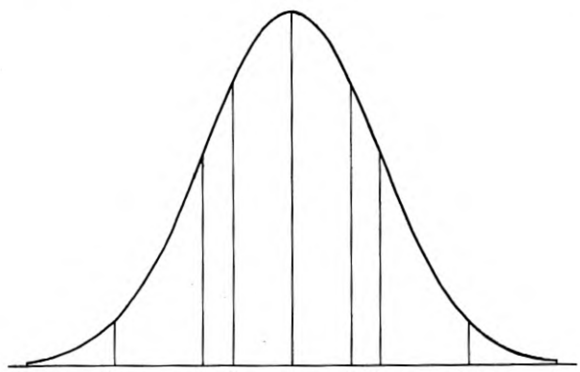



Fig. 1—Customarily assumed law of error curve—normal law

50.00000% of area within $0 \pm .67449\sigma$
68.26894% of area within $0 \pm 1\sigma$
95.44998% of area within $0 \pm 2\sigma$
99.73002% of area within $0 \pm 3\sigma$

The total area under the curve is, of course, unity. This curve is plotted with the origin at the true value and with the errors measured in units of the root mean square error $\sigma$. The fractions of the area bounded by certain multiples of the root mean square error are shown for reference.

Let us make an experiment and see how far customary error theory

[1] Biometrika, Vol. VI, 1908, pp. 1–15. Vol. XI, 1917, pp. 416–417.

[2] Biometrika, Vol. X, 1915, pp. 522–529.

[3] Biometrika, Vol. X, 1915, pp. 507–521. Proc. Camb. Phil. Soc., Vol. XXI, 1923, pp. 655–658.

[4] The equation for this has recently been traced back to Abraham De Moivre (1733) by Professor Pearson. See Biometrika, Vol. XVI, 1924, pp. 402–404.

carries us, see where it breaks down, see why it breaks down, and then avail ourselves of the new theory—a powerful tool of great value, because it makes possible for the first time the solution of many practical problems. Here is the experiment. Take 998 small circular chips, 499 green and 499 white. Mark 20 white ones with 0, 40 white ones with 0.1, 39 white ones with 0.2, etc., in accordance with the normal law. Do the same for the green chips except that all numbers on the chips are minus. Put the 998 chips in a bowl, mix thoroughly, draw out one and record it. Replace the chip, again mix thoroughly, and repeat the process until 4000 values are observed. A little reflection shows that this experiment is equivalent to making 4000 measurements of a quantity by a method subject to a normal law of error with a root mean square error of approximately unity.

Let us group these 4000 values into 1000 groups of 4, and determine the average for each group, taking the first four observations as the first group, the second four as the second group and so on. This gives

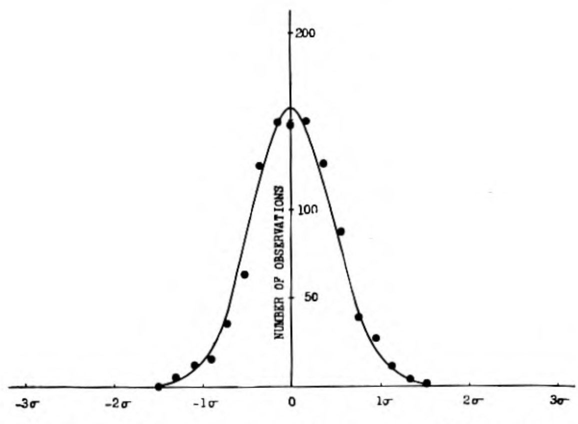



Fig. 2—Curve showing customary error theory to be satisfactory on one condition not often met in practice; *i.e.*, $\sigma$ is known

• Distribution of 1000 averages of 4

– Normal law with root mean square error $\frac{\sigma}{\sqrt{4}}$

us 1000 averages. Suppose we subtract the true value $m$ (in this case zero) from each average and divide this result by the root mean square error of the frequency distribution of values within the bowl. This gives us 1000 observations of the error of the average of 4 observations measured in terms of $\sigma$. Customary error theory shows that these averages should be distributed normally as indicated by the smooth curve in Fig. 2 with a root mean square error of $\frac{\sigma}{\sqrt{4}}$ or one half that in Fig. 1. The dots show the experimental results.[5]

So far the customary error theory is satisfactory. But we do not often have this case in practice; that is, we do not know the root mean square error $\sigma$, and instead know only the observed root mean square error $s$ of the sample.[6]

## Case Where Customary Theory Does not Apply

Let us next recall just the way we use the customary theory in practice and then see what mistake we usually make. Take the results of drawing the first sample of 4 in the experiment previously cited. The four observed values are .6, $-.2$, 1.1, $-2.0$, the average $\overline{X}$ of these

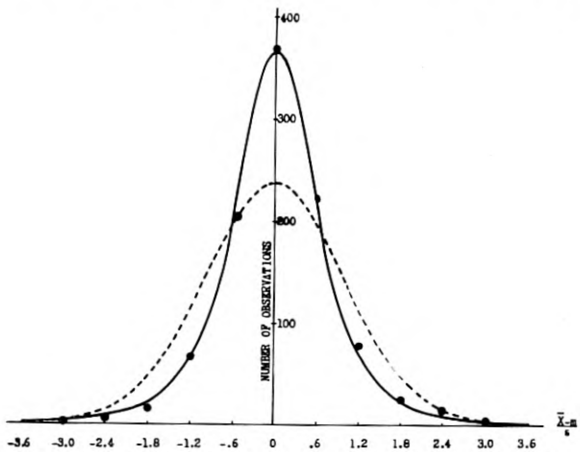



Fig. 3—Curves showing inaccuracy of customary error theory in finding error of average in terms of the observed standard deviation $s$

- - - - - Customary theory
———— New theory
• Distribution of 1000 $z$'s

is $-.125$, and the observed root mean square deviation $s$ is 1.177. Assuming no knowledge of the root mean square error $\sigma$ of the distribution from which the sample of 4 was taken and using customary theory, we should assume the probable or 50% error to be $.6745 \frac{1.177}{\sqrt{4}}$.

[5] I am indebted to Miss Victoria Mial and Miss Marion Cater for securing the experimental results, making all necessary calculations, and drawing the curves given in this paper.

[6] Customarily we do not know the true value $m$, hence instead of knowing the root mean square errors we know the root mean square or standard deviations.

This follows from the fact that the observed values of the ratio $z = \frac{\bar{X} - m}{s}$ where $m$ is the true value, are customarily assumed to be distributed normally. *Here we come to the crux of the discussion: these observed values of the ratio are not distributed normally.* "Student"[7], in 1908, was the first to show how they are distributed.

Let us look at the observed frequency distribution of the 1000 $z$'s given by the above experiment (dots Fig. 3). To be normally distributed, as customarily assumed, these dots would have to lie on the dotted normal curve. Obviously they do not. Instead they lie on a much more peaked curve (solid line) than the normal. This was calculated with the aid of "Student's" theory. We must therefore conclude: the probability that the mean of a sample of $n$, drawn at random from a normal distribution, will not exceed (in the algebraic sense) the mean of that distribution by more than $z$ times the root mean square deviation of the sample cannot be found from the normal law when $n$ is small. We must use the tables provided by "Student" in the two papers referred to above.

## Why the Customary Theory Fails to Give the Error of the Average in Case of Small Samples

Let us look a little further into the reason why the $z$'s are not distributed normally, before we consider the question as to the magnitude

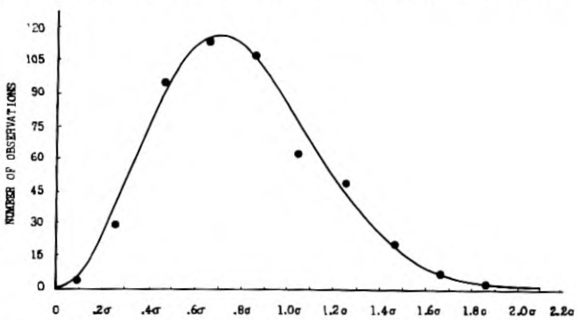



Fig. 4—Data furnishing a clue to reason for inadequacy of customary error theory
• Observed distribution of standard deviations of 1000 samples of four
– Theoretical curve of asymmetrical type

of the difference between the probable error determined from one theory and that determined from the other.

Let us look at the distribution of the 1000 standard deviations, the $s$'s, Fig. 4, for here we shall find the secret revealed: The distri-

[7] Loc. cit.

bution of $s$'s, as we might expect, is asymmetrical; the most probable standard deviation $s$, to be observed is not the average $s$. Of course, the customary theory assumes that the average $s$ is the most probable $s$, and that the distribution of $s$ is normal. We should therefore expect to find the $z$'s distributed normally for values of $n$ such that the dis-

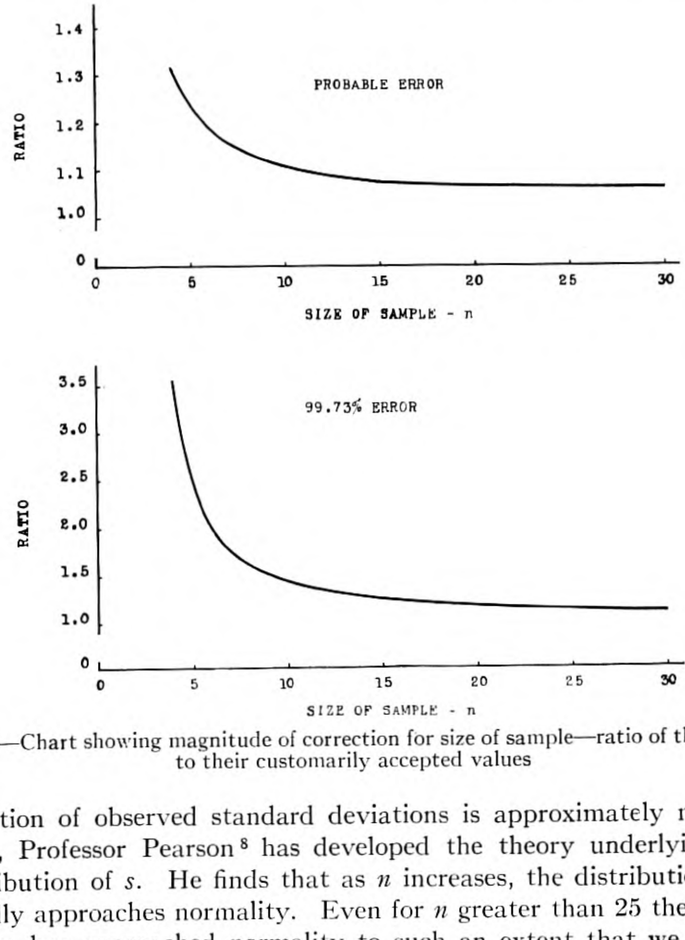



Fig. 5—Chart showing magnitude of correction for size of sample—ratio of the errors to their customarily accepted values

tribution of observed standard deviations is approximately normal. Now, Professor Pearson [8] has developed the theory underlying the distribution of $s$. He finds that as $n$ increases, the distribution of $s$ rapidly approaches normality. Even for $n$ greater than 25 the distribution has approached normality to such an extent that we should expect the $z$'s to be distributed approximately in normal fashion. The study of the distribution of $z$ shows this to be true, as we shall see below.

In passing, we should note how closely the theoretical curve, Fig. 4, fits the observed points and also note two other checks between theory

[8] Loc. cit.

and observation furnished by the new data given herein. According to theory, the modal and mean values of $s$ for samples of size 4 expressed in units of $\sigma$ should be .707 and .798 respectively. The experimental results are .717 and .801.

## How Much Larger Are the Probable and 99.73% Errors of an Average Than the Customarily Accepted Values?

The difference between the error of an average and its customarily accepted value increases as the number of observations $n$ (or size of

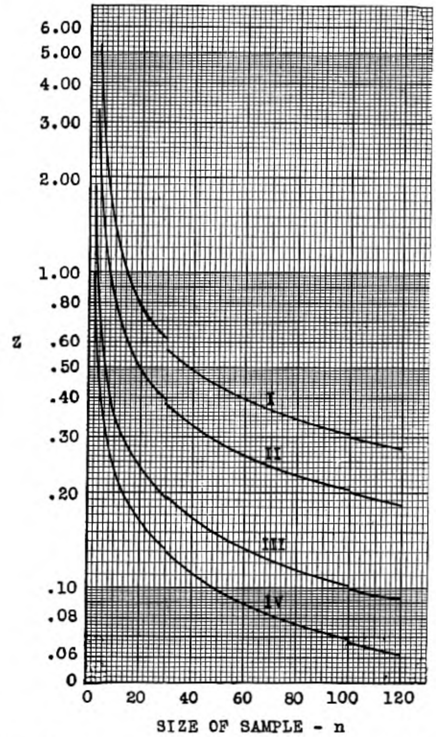



Fig. 6—Errors of averages of samples of size $n$
I −99.73002% error
II −95.44998% error
III−68.26894% error
IV −50.00000% error
$z$=the ratio of the error of the average to the observed standard deviation

sample) decreases. This fact is illustrated in Fig. 5. This figure shows the ratios of the errors to their customarily accepted values plotted for values of $n$ from 4 to 30.

Curves showing the most frequently used errors of averages measured in terms of $z$ (*i.e.* in terms of the ratio of the error to the observed standard deviation) are given in Fig. 6. The error curves for $n$ less than 30 have been obtained with the aid of "Student's" original tables, those for $n$ between 30 and 100 have been obtained from the normal law integral tables using the standard deviation of $z$; *i.e.* $\frac{s}{\sqrt{n-3}}$ as given by "Student." For $n$ greater than 100, customary error theory has been used.[9]

## Typical Practical Applications

But few, if any, recent developments of statistical theory are of more general application in most fields of scientific research and engineering than the one herein described.[10] This follows because the theory herein discussed must be used in calculating the required probable error (or other measure of dispersion) of the averages obtained from small numbers of observations. The number of applications of this character is legion.

## Problem Type 1, Determination of Error of Average

*Example* 1:

Five samples of granular carbon taken from a crucible show resistances of 47.5, 49.4, 43.2, 48.0 and 46.2 ohms respectively. What are the probable and 99.73% errors of the average of these resistances?

*Solution:*

The observed values of average resistance $\overline{X}$, and standard deviation $s=\sqrt{\frac{\Sigma(\overline{X}-X)^2}{n}}$ are 46.9 ohms and 2.097 ohms respectively. Hence from Fig. 6 we see that the probable and 99.73% errors are respectively $.372s=.780$ ohms and $3.33s=.699$ ohms respectively whereas from customary theory they would be $.302s=.633$ ohms and $1.34s=270$

[9] For the curves in this figure as in the preceding one, I have assumed the customary theory for the case where the true value of $X$ is known so that the root mean square error of the average $\overline{X}$ of sample of size $n$ is the ratio $\frac{s}{\sqrt{n}}$. Of course, as we know from customary error theory, if we assume no knowledge of the true value of $X$, we should use $\frac{s}{\sqrt{n-1}}$.

[10] Since this paper was written, a very interesting article, "Statistics in Administration," has appeared in *Nature* (V. 117, pp. 37–38, Jan. 9, 1926), calling attention to the importance of the theory of small samples.

ohms respectively. The true probable and 99.73% errors are 23% and 148% higher respectively than those calculated by customary theory, as is evident from Fig. 5.

*Discussion of Type* 1:

Examples of this type of problem are obviously so numerous that further illustrations need not be given. They occur every day in practically every science. We see that in such cases it is certainly necessary to allow for the effect of the small size of sample.

## PROBLEM TYPE 2, DETERMINATION OF ERROR OF AVERAGE DIFFERENCE

*Example* 1:

Five instruments are measured for some characteristic $X$, first on one machine and then on another, giving two sets of values $X_{11}$, $X_{12}$, $\ldots X_{15}$, and $X_{21}$, $X_{22}$, $\ldots x_{25}$ respectively. Calculate the 5 differences $X_{11}-X_{21}=x_1$, $X_{12}-X_{22}=x_2$, $\ldots X_{15}-X_{25}=x_5$. Assume that the average difference is $\bar{x}$ and the standard deviation of the differences is $s$. Assuming the two machines give the same results except for random variations, what is the probability that the observed difference would occur? Are we justified in the assumption that the machines give the same results?

*Solution:*

The true difference is zero on this assumption. The observed difference is $z=\dfrac{\bar{X}-0}{s}$, and "Student's" tables may be used to evaluate this probability.[11] If this probability is very small, let us say .001 or less, it may be taken as indicating that the machines do not give the same results.

*Example* 2:

We wish to compare the depth of penetration obtained from two different methods of preserving chestnut telephone poles. We choose $n$ poles for test. A sample from each pole is treated by one process, and a sample from each pole is treated by another process. The depths of penetration are measured. Are we justified in assuming the two methods to give significantly different results?

[11] Approximate values can be obtained from the curves in Fig. 6.

*Solution:*

If $n$ is small, we proceed as in the previous case, to find the probability of occurrence of the observed difference. If this probability is small, we conclude that the difference is significant; *i.e.*, the two methods of preservation give different results.

*Example* 3:

Three-bolt guy clamps are used for clamping the guy wires on telephone poles. These are supplied from different sources. Those from one source fail to hold the wire as well as those from another and inspection shows that these same clamps fail to meet a certain specified dimension. The force required to slip the wire in each of 10 clamps from this source is measured. These clamps are then modified to meet the specified dimension and the force required to slip the wire in each clamp is again measured. Are we justified in attributing the failure to hold the wire to the fact that these clamps did not meet the specification?

*Solution*:

The solution follows the same line as in the first case.

*Discussion of Problems of Type* 2:

Problems of this type are very numerous. It is obvious that significant differences calculated as above indicated are always larger than those calculated by customary theory.

## Problem Type 3, Determination of Most Probable Value of the Root Mean Square Deviation of the Universe When Only One Sample of $n$ Pieces Has Been Examined

*Example* 1:

Five tool-made models are tested for their efficiency, giving values $X_1, X_2, \ldots X_5$. What is the most probable value of the range within which the efficiencies of product instruments may be expected to lie approximately 99.7% of the time, assuming that a manufacturing process can be developed which is the same as that used in producing the tool-made models?

*Solution*:

Customary practice would answer: the average of the five values plus or minus 3 times their standard deviation. The better answer is: the average plus or minus $\frac{3}{.7746}$ times the standard deviation.

This follows from Professor Pearson's work previously quoted. He has shown that the most probable observed standard deviation $\breve{s}$ of

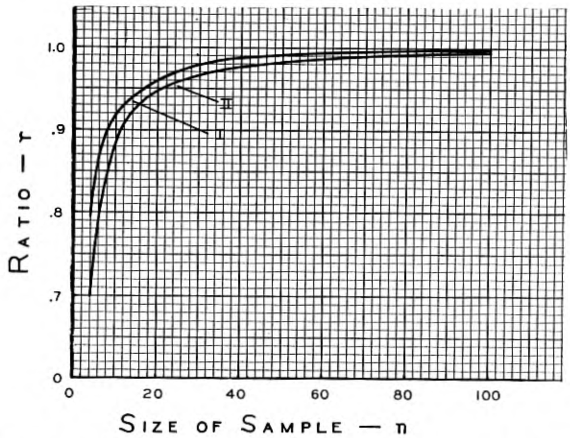



Fig. 7—Curves giving the most probable value of the true standard deviation $\sigma$
I When the average $\bar{s}$ of standard deviations of many samples is known. $r\sigma = \bar{s}$
II When the standard deviations of one sample is known. $r\sigma = \breve{s}$

a sample of $n$ from a normal distribution with standard deviation $\sigma$ is $\breve{s} = \sqrt{\frac{n-2}{n}}\,\sigma$. Substituting the value $n=5$ in this equation we get $\breve{s} = .7746\sigma$.

A curve of the values of $\frac{\breve{s}}{\sigma}$ vs. $n$ is presented in Fig. 7 for reference in solving problems of this character.

## Problem Type 4, Determination of Most Probable Root Mean Square Deviation of the Universe When Several Samples of $n$ Pieces Each Have Been Examined

*Example* 1:

One thousand transmitters, known to have different efficiencies, have been tested five times each for efficiency. Find the standard deviation of the machine method of measurement.

*Solution:*

Calculate the standard deviation of the five tests for each transmitter. Find the average value of these 1000 values and divide it by .8407. This follows from the fact that the average $\bar{s}$ of the observed standard deviation for a series of samples of size $n$ drawn from a normal distribution with standard deviation $\sigma$ is

$$\bar{s}=\sqrt{\frac{2}{n}}\,\frac{\left|\underline{\frac{n-2}{2}}\right.}{\left|\underline{\frac{n-3}{2}}\right.}\,\sigma.$$

where the symbol $\left|\underline{X}\right.$ is equivalent to $\Gamma\ (X+1)$

Thus for $n=5$ we get [12] $\bar{s}=.8407\sigma$.

Fig. 7 also presents the values of the ratio $\frac{\bar{s}}{\sigma}$ for reference and with sufficient accuracy for solving problems similar to the example cited. Greater accuracy than that afforded by the curves can be secured by direct substitution in the equations for $\breve{s}$ and $\bar{s}$ or by referring to the original tables.

[12] We will recall with interest how closely the observed average, $\bar{s}=.798\sigma$, of the 1000 values of $s$ corresponding to the 1000 samples of four herein presented checked the theoretical average of $.801\sigma$.

# The Alkali Metal Photoelectric Cell

By HERBERT E. IVES

## Introduction

IN the development of the commercial system of picture transmission now in operation over certain of the Bell System lines, one of the initial problems was the choice of a method of transforming the light and shade of the picture to be transmitted into properties of an electric current. There are in general two methods of accomplishing this. The first, which we may term the photo-mechanical method, utilizes some photographic process to produce a mechanical structure, which may be used either to make and break contact, or to produce mechanical movement of some element whose motion produces a variable electric current. The second method consists in the utilization of some light sensitive device which produces or varies an electric current.

An indispensable requirement in the electrical transmission of pictures is *speed* in conveying the picture from one point to another. The choice of a method of transforming light and shade into an electrical current will therefore, other things being equal, be that method which requires the least time for the transformation. It is on this basis that the photo-mechanical methods were not favorably considered in this development. The preparation of the line or dot structure image, similar to the half tone plate, or the preparation of a photo-relief, are processes which cannot be completed in less than one to two hours, and involve a delay which in many cases would seriously detract from the advantages of electrical transmission over other means now available, such as the airplane.

In choosing a photo-sensitive device for this purpose, certain requirements had to be met. The light responsive device should be as nearly as possible instantaneous in its action. The response should also be proportional to the light intensity. These requirements cannot be met by any photo-sensitive devices of the group whose resistance changes under the action of light, such as selenium. The field was therefore limited to the *photoelectric cell*, of the type in which the effect of light is to release electrons from the surface of the light sensitive element and so cause an electric current to flow in the space between the light sensitive surface and another electrode. Photoelectric cells are considerably less sensitive than the best variable photo-resistances, but while this characteristic would have made them difficult to utilize in the earlier days of efforts at picture transmission, the development of vacuum tube amplifiers admirably fitted for amplifying photoelectric currents has remedied this deficiency. A further requirement, that the light sensitive device should preferably be sensitive to visible radiation, ruled out the use of those sensitive materials sensitive chiefly to infra-red or ultra-violet radiation. All of these requirements pointed to the *alkali metal photoelectric cell* as developed by the work of Elster and Geitel and others.

## General Characteristics of Photoelectric Cells

The typical photoelectric cell consists of a hermetically sealed glass bulb containing an atmosphere of gas at a low pressure, and provided with two electrodes, one of which is the light sensitive material. In the schematic cell shown in Figure 1, $K$ is the photo-sensitive material

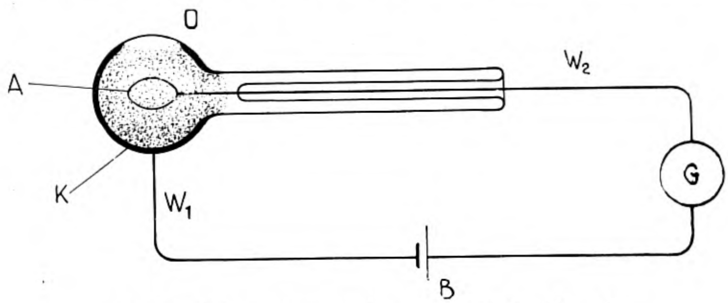



Fig. 1—Schematic central anode photoelectric cell.

(cathode), for instance an alkali metal such as potassium, which is spread upon the inside wall of the glass bulb and is connected with the exterior of the bulb by a sealed-in wire, $w_1$, $A$ is the other electrode (anode). As here shown, it is a simple metallic ring connected with a second wire, $w_2$, carried through the stem of the bulb. The two electrodes are shown connected together through a battery, $B$, and galvanometer, $G$. The operation of the cell consists in letting light fall upon the cathode through the window, $O$. The resulting current may then be measured by the galvanometer, or utilized to operate suitable apparatus.

A complete study of the photoelectric cell resolves itself into obtaining knowledge of the effect of varying a number of factors which enter into its construction and use. Of these we may note: the material which is used for the light sensitive surface, and the treatment to which this material is subjected; the composition and pressure of the gaseous atmosphere; the shape and disposition of the various elements, that is, the *structure* of the cell. We must investigate the relationship between the quality of the light falling upon the cell and the electric current produced. We must in addition consider certain other physical variables which must be met with in practice, notably temperature.

## Characteristics of Cells of Various Structures

For purposes of discussion we may classify photoelectric cells in regard to structure as *central cathode cells* and *central anode cells*. As the terms imply, these two extreme types of cells differ in the position of the photoelectric material. The central anode type of cell is shown in Figure 1. The sensitive material entirely covers the walls, so that the cathode is of relatively large area. The central cathode cell is illustrated in Figure 2, in which the symbols are the same as in Figure 1.

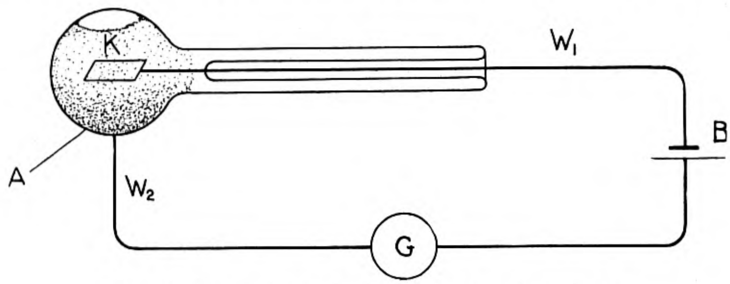



Fig. 2—Schematic central cathode photoelectric cell.

In this the walls of the cell are covered with a non-light sensitive material (e.g. silver), and the sensitive material is coated upon a relatively small centrally placed electrode.

## Central Cathode Cells

Central cathode cells possess certain decided advantages for the theoretical study of photoelectric phenomena and have consequently been used in many of the more important photoelectric investigations. The simplest case to consider first is that of the high vacuum cell, that is one containing no appreciable amount of gaseous atmosphere. When a constant light is incident on the light sensitive cathode, and a series of voltages are applied to the terminals of the cell, voltage-current relationships are obtained of the character shown by any one of the curves of Figure 3. Several significant points are to be noted about these characteristic curves. We find that the photoelectric current starts at a definite positive value of the voltage. This voltage is called the "stopping potential." It varies with the wave length of the exciting light. This is shown in the figure by the several curves for different wave lengths, varying from $\lambda_1$, representing short wave energy, such as blue light, to $\lambda_3$, long wave energy such as yellow. The shorter the wave length (the higher the frequency), of the exciting light, the higher must be the positive potential necessary to prevent or stop the emission of electrodes under illumination. As the positive potential is reduced, the photoelectric current increases, until the applied field (or the effective field if contact potential differences are present) becomes zero. At this point, the current becomes *saturated*, that is increase of voltage in the negative direction fails to increase the current. This means that the applied field does not penetrate to any appreciable depth into the photoelectric material.

Characteristic curves of the type shown in Figure 3, have played a

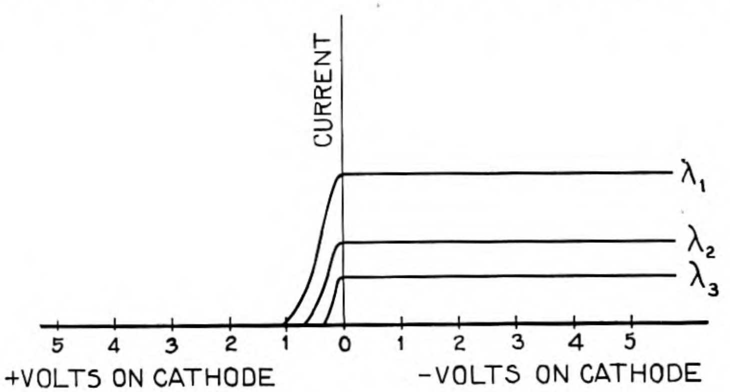



Fig. 3—Voltage-current curves for typical central cathode vacuum photoelectric cell.

very important part in the development of photoelectric theory, and particularly of the quantum theory. If $V$ is the voltage applied to the cell, $e$ the charge on an electron, $h$ the quantum constant, $\nu$ the frequency of the exciting light, Einstein predicted and Millikan has shown experimentally that the following relationship holds: $eV = h(\nu - \nu_0)$, where $\nu_0$ is the limiting frequency corresponding to the long wave length limit, beyond which the photoelectric emission does not occur. If $m$ is the mass of the electron, and $v$ its velocity, the above relation can be written $1/2\ m\ v^2$ (Velocity$^2$) $= h(\nu - \nu_0)$. From this expression it is evident that the greater the interval between the frequency of the light used, and the limiting frequency, the higher is the velocity of emission of the photoelectrons.

When instead of being highly exhausted, the cell has an atmosphere of gas at a low pressure (a few tenths of a millimeter of mercury) the condition of saturation typical of the high vacuum cell for high negative voltages no longer holds. Instead the photoelectric current is increased by the occurrence of ionization, by collision of the electrons initially produced with the molecules of gas. The current increases with applied voltage in the manner shown in Figure 4, until at some value characteristic of the kind of gas in the cells, the gas breaks down and a visible electrical discharge takes place. The amplifying effect of the gaseous atmosphere increases with the pressure of the gas up

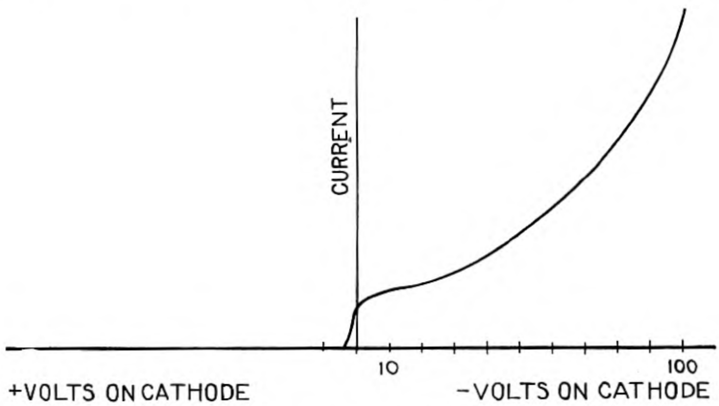



Fig. 4—Voltage-current curves for typical central cathode gas filled photoelectric cell.

to a maximum and then decreases. The value of this optimum pressure depends on the kind of gas and the dimensions of the tube. The best pressure is usually a few tenths of a millimeter of mercury.

As the illumination of the cell is changed, the current changes in exact proportion, that is the illumination-current relationship is rectilinear. This relationship holds for both the vacuum and gas cells provided there are no free glass surfaces on which charges may accumulate. If the window is made too large it may become charged and cause an appreciable curvature of the illumination-current relationship.

## Central Anode Cells

In cells with a relatively small centrally placed anode, the voltage-current relationship differs from that of the central cathode cells most noticeably in that high applied voltages are necessary in order to insure saturation. Typical voltage current curves for short ($\lambda_1$) and long ($\lambda_2$) wave length energy, for a central anode cell consisting of a spherical anode and concentric spherical cathode are shown in Figure 5. The rate at which saturation is approached with voltage varies with the wave length of the exciting light. The longer the wave length, the slower the electrons, as shown above, and the more quickly are they captured by the central anode.

When gas is introduced into a central anode cell, we again have ionization by collision, and the voltage-current curve is turned upward

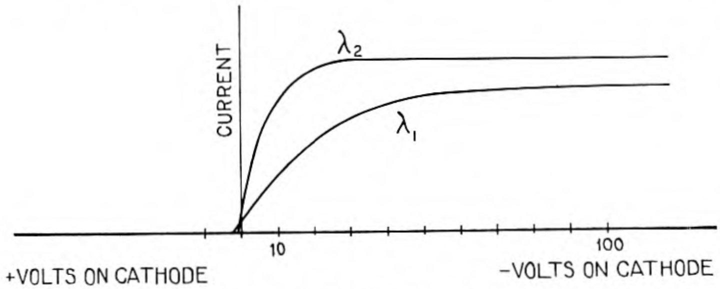



Fig. 5—Voltage-current curves for typical central anode vacuum cell.

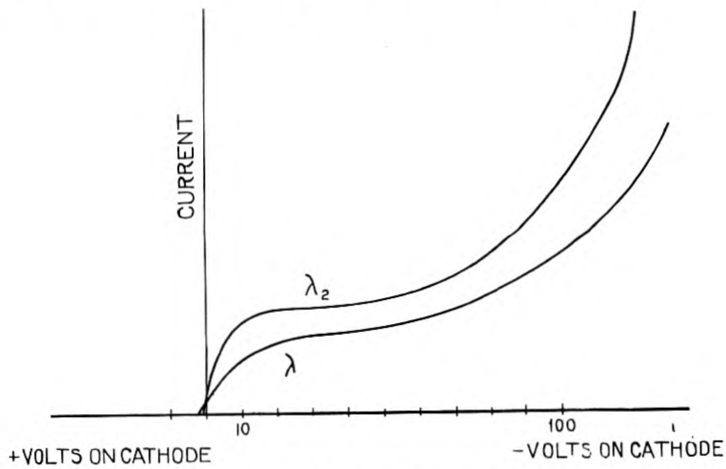



Fig. 6—Voltage-current curves for typical central anode gas filled cell.

from the voltage axis in the manner shown in Figure 6. As in the case of the central cathode cells, the current increases with voltage until the critical potential for the gas is reached.

The illumination-current relationship is rectilinear in the central anode cells, as it was in the central cathode cells, provided the precautions as to avoiding free glass surfaces, already mentioned, are observed.

### Infulence of the Nature of the Irradiation

The magnitude of the photoelectric current depends upon the angle of incidence, the plane of polarization and the color or wave length of the light used. This dependence is closely interlinked with the choice of the photo-sensitive material and the state of its surface. For the purpose of separating out the effects of the several variables in the incident light, it is necessary to study the properties of optically plane or specular surfaces of the photo-sensitive material. Such surfaces can only be obtained with the alkali metals by raising them above their melting points, by forming alloys, or by depositing extremely thin films on a polished underlying metal surface such as platinum. Specular surfaces of the aklali metals obtained in these several different ways exhibit differences in their behavior, but it will be sufficient for the present purposes to disregard these secondary differences and to speak merely of the photoelectric current from specular surfaces when the incident light varies in wave length in certain typical ways, or is polarized.

### Influence of the Plane of Polarization

When light is incident at a steep angle on a specular surface, the two extreme conditions of polarization are those in which the electric vector lies in the plane of incidence and that in which it lies perpendicular to the plane of incidence. In the first case, the electric vector has a component perpendicular to the surface. In the latter case the electric vector lies parallel to the surface. It has long been known that the amount of light absorbed by a metal surface is, in general, greater when the electric vector is in the plane of incidence. Consequently, since photoelectric emission must be due primarily to the absorption of the energy from the incident light, it is to be expected that the photoelectric current will be greater for light polarized with the electric vector in the plane of polarization. Such is actually the case, but while the ratio of absorption of light, for the two planes of polarization, at say 60° incidence, never rises above a value of four for any of the alkali metals, the ratio of the photoelectric currents under the same conditions may mount to a very high value, such as 20 or 30 to one. This effect is particularly noticeable in the liquid alloy of sodium and potassium, and in the case of all the four alkali metals, sodium, potassium, rubidium and caesium, when these spontaneously deposit in a high vacuum upon a polished surface. It is very much less marked in the case of the pure alkali metals in the molten condition. Typical examples of the influence of the plane of polarization on the photoelectric effect are shown in Figure 7, where the symbol || indicates that the electric vector is in the plane of incidence, the symbol ⊥ that it is perpendicular to this plane.

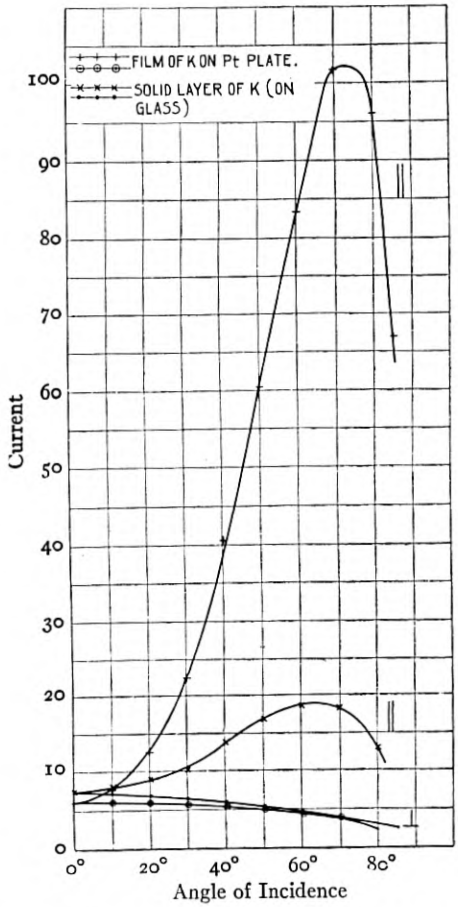



Fig. 7—Photoelectric emission from a specular surface of solid potassium, and from a thinly coated platinum plate, at various angles of incidence; electric vector in plane of incidence (||); electric vector perpendicular to the plane of incidence (⊥).

## Distribution of Response According to Wave Length

When the exciting light is incident either perpendicularly on a specular alkali metal surface, or at a high angle of incidence with the plane of polarization such that the electric vector is parallel to the surface, the response for equal intensities of monochromatic radiation through the spectrum is as shown by the curve marked ⊥ in Figure 8. Photoelectric emission is entirely absent on the long wave side of a certain wave length which is known as the *long wave length limit*. From this wave length, the emission rises gradually and uniformly toward the short wave or blue end of the spectrum.

When the incident light is polarized so that the electric vector has a component perpendicular to the surface, the wave length distribution of response follows the general character shown in the curve marked || in Figure 8. Correlating the wave length distribution of response with

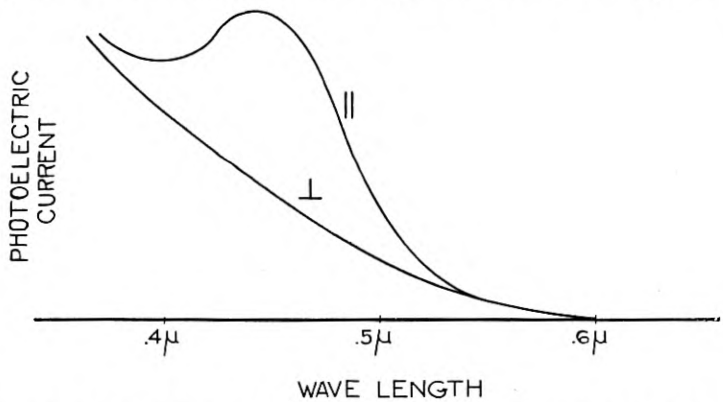



Fig. 8—Wave-length distribution of response from specular alkali metal surface; electric vector in plane of incidence (||); electric vector perpendicular to plane of incidence (⊥).

the variation of emission with the plane of polarization considered in the last section, the general conclusion may be drawn that the enhanced emission for light with the electric vector in the plane of incidence is due largely to radiation falling within a narrow spectral region. The magnitude of this wave length peak varies greatly with different materials. The maximum occurs at a wave length different not only for the different alkali metals, but also for different modes of securing the specular surface. This wave length maximum has the appearance of being due to some resonance phenomenon, and its variation in position through the spectrum according to the method of preparation of the surface is connected in some unknown way with the state of binding of the alkali metal atom on the surface with the body of material beneath.

### The Photoelectric Current from Rough Surfaces

With rough surfaces of alkali metal, the plane of polarization of the incident light no longer has meaning. We would therefore expect no significant difference in the total emission from a really rough surface when the plane of polarization of the incident light is changed; and such in fact is the case. We would also expect that the distribution of response according to wave length would be a mixture of the effects of the two planes of polarization. This also is found to be so. Rough surfaces show the wave length maximum characteristic of light polar-

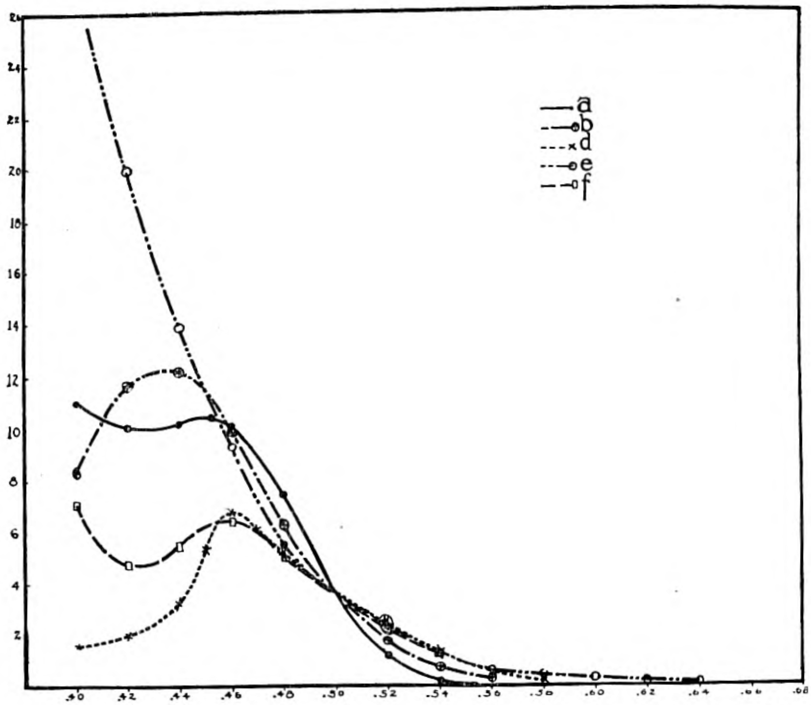



Fig. 9—Wave-length distribution of emission from rough surface potassium photoelectric cells, showing variation from cell to cell depending on difference of treatment.

ized with the electric vector in the plane of incidence on a specular surface. Depending on the degree of roughness and the method of preparation of the surface, the wave length maximum is different in size and also in position. In Figure 9 are shown several wave length distribution curves for potassium cells in which the surface is roughened and colored by a hydrogen glow discharge. These cells differ both in respect to their absolute sensitiveness, in the position of their maximum sensitiveness, and the extent of their sensitiveness toward the red end of the spectrum. It has not as yet been found possible to prepare photoelectric cells with properties uniform from one cell to another.

## Influence of the Photoelectric Material and its Treatment

The alkali metals differ in their photoelectric sensitiveness in a perfectly definite order, which is that of their degree of electro-positiveness, as shown by their position in the periodic table of the elements. The variation in sensitiveness is correlated with the extension of sensitiveness in the spectrum. This progresses regularly from sodium, which in its pure state is not photoelectrically sensitive beyond about $.58\mu$, through potassium and rubidium, to caesium, which is photoelectrically sensitive in the near infra-red. The exact terminations of sensitiveness in the spectrum depend upon the character of the surface and its treatment, and have not been exactly correlated with any other properties of the material.

In order to attain the greatest sensitiveness with the alkali metals, these are commonly subjected, in the preparation of the photoelectric cell, to what is called the *coloring* process, discovered by Elster and Geitel. This consists in subjecting the surface to a glow discharge in an atmosphere of hydrogen. The result is to color the otherwise silvery alkali metal a rather deep blue-purple or blue-green. The exact cause of this color is not known, but it has every appearance of being due to the production of small (colloidal) particles of alkali metal. The greater sensitiveness is probably due to the increased effective surface presented by the colloidal particles rather than the increased absorption coefficient of the darker color. Similar colors may be obtained by distilling the alkali metal in a very thin layer on glass and the color of the surface changes when observed by polarized light in much the same manner as do colloidal surfaces of other sorts.

After the completion of the coloring process, it is necessary to remove all the hydrogen from the cell by pumping. Otherwise the surface will revert to its original uncolored form. In order to obtain the amplifying effect of a gaseous atmosphere, it is customary to introduce an inert gas, such as argon or helium, into the cell.

Cells made in the manner just outlined are reasonably permanent in their important characteristics. Elster and Geitel have made potassium cells in this manner, which when connected with a delicate electrometer exhibited a degree of sensitiveness approximately that of the human eye. According to what has gone before, the most sensitive cells should be obtained if rubidium or caesium are used in place of potassium. It is found however by experiment that rubidium, and particularly caesium, do not lend themselves so well to the coloring process, probably because of their lower melting points, and hence cells made of potassium may represent practically the best performance which is now attainable.

## Effect of Temperature

It has long been held that the photoelectric effect is independent of temperature. Recent experiments, however, have shown that the alkali metals are affected in their photoelectric response by variation of temperature. A typical set of data for potassium is shown in Figure 10. It will be noted that the influence of temperature is small

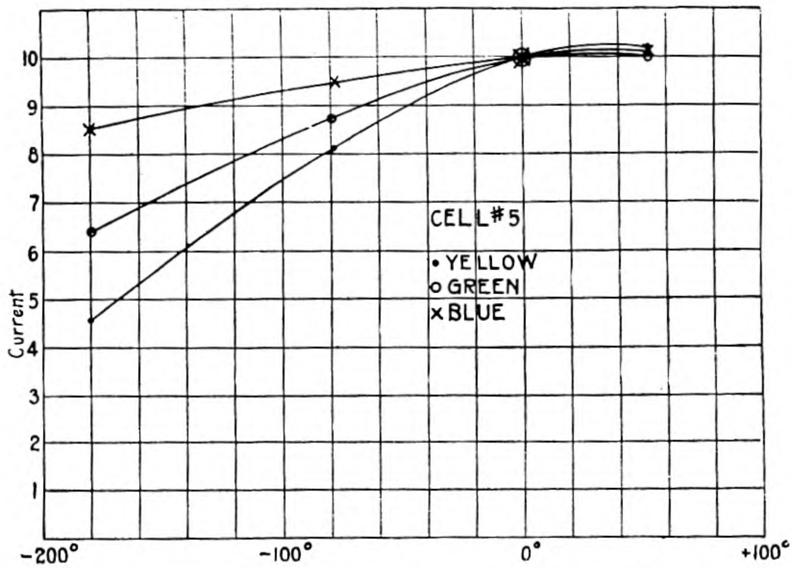



Fig. 10—Variation of photoelectric sensitiveness of potassium with temperature.

for the shorter wave lengths of light, but considerable for long wave excitation. These data were secured in highly exhausted cells of pure alkali metal. When gas is present or a large amount of alkali metal vapor can deposit on the cooled surface, the effect of decreased temperature may be to increase the photoelectric current. It will be noted that the changes shown in Figure 10 are insignificant over the ordinates range of room temperatures, and for practical purposes, particularly for picture transmission, the effects of temperature on the performance of photoelectric cells may be taken as negligible.

## Practical Features of Cells as Used

The photoelectric cells as used for picture transmission are classified, according to the above discussion, as central anode, gas filled, colored cells. The shape of the cells is that shown in Figure 1, and also in the photograph, Figure 11, with which is an accompanying scale. The cells are made of pyrex glass, which is chosen because it is highly resistant to corrosion by potassium during the distillation stages. The very long neck of the cells is dictated partly by the space into which the cells are placed in the picture transmission apparatus, in

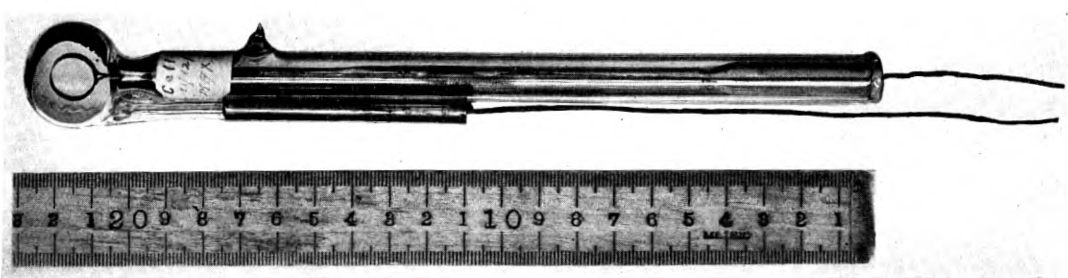


Fig. 11—Photograph of photoelectric cell of type used in picture transmission.

part by the desirability of having as long an insulating space as possible. Where, as in earlier types of photoelectric cells, the alkali metal is in close proximity to the other electrode, leakage currents over the glass surface greatly interfere with accurate results. (In working with extremely small currents it is desirable to have in addition to the considerable length of glass insulating path, a metallic guard ring in the stem of the cell, which may be earthed.)

The alkali metal ordinarily used is potassium. This is introduced by distillation on the pump. The cell is first baked to a temperature of 400° C. for several hours while on the pump in order to drive out all traces of water vapor. The potassium for use in making up the photoelectric cells is first of all distilled in a vacuum into long glass tubes. In this preliminary distillation, the greater part of the absorbed gaseous impurities are removed. After the cell has been baked out on the pump, a piece of the glass tube containing potassium is broken off and introduced into the pump system. Between the point of introduction and the cell are a series of bulbs. The potassium after melting in vacuo is distilled successively through these bulbs and into the photoelectric cell, where it is condensed on the walls of the bulb. A window is then made in the cell by applying a small flame on the appropriate part. The next step is to introduce a small amount of pure hydrogen gas, which is permitted to enter from a reservoir on the system. This hydrogen gas goes through the system of bulbs through which the potassium has been distilled, which still contain a large amount of potassium, and is thereby cleaned of all traces of gases or vapors which might react on the potassium in the cell. A glow discharge is then

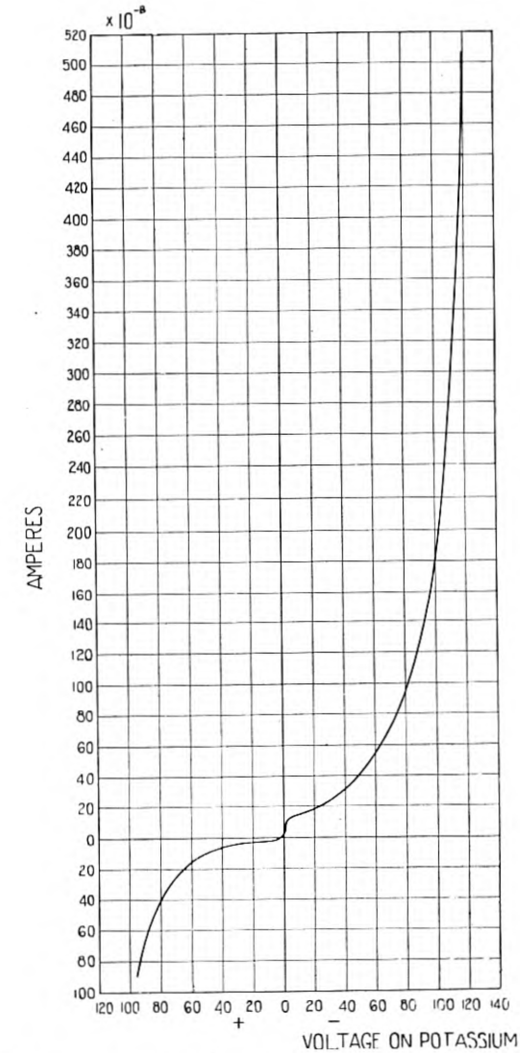



Fig. 12—Typical voltage current characteristic of potassium photoelectric cell as used in picture transmission. Incident luminous flux = .015 lumen.

passed from a high voltage source, until, by illuminating the alkali metal surface and reading the current on a sensitive galvanometer, it is found that a maximum of sensitiveness has been attained. The hydrogen is then completely removed by long continued pumping. The final step in the preparation of the cell consists in the introduction of a small quantity of carefully purified argon. The argon for this purpose is held in a reservoir in which there is a pool of sodium-potassium alloy. By passing an electric discharge from this pool to an electrode through the gas, the argon is purified of all active impurities. It is introduced into the cell through the same series of potassium coated bulbs already mentioned, the potassium in the meantime having been vigorously heated to drive off all occluded hydrogen, so that the gas when it finally reaches the photoelectric cell is entirely inert. The gas pressure is carefully adjusted while the cell is still on the pump so as to give an optimum effect, after which the cell is sealed off.

Typical voltage-current characteristics of the cells thus made are as shown on Figure 12, where the currents indicated are those obtained from an illumination of 100 meter candles from a vacuum tungsten lamp, the aperture of the cell being 1.5 sq. cm. It will be noted that, unlike the ideal characteristic shown in Figure 6, the actual cells shows a small current in the opposite direction for positive voltages applied to the sensitive surface. This is because practically it is very difficult to prevent some alkali metal from depositing on the anode, which thus becomes light sensitive, and responds to the scattered and reflected light in the cell.

For use in the picture transmission apparatus, the cells are mounted in tubular metal cases, from which they are insulated by hard rubber rings attached to the glass stem, by sealing wax. The cells in their cases are handled as units; and are sufficiently rugged to be readily shipped from place to place. Their characteristics remain practically unchanged indefinitely.

## SELECTED BIBLIOGRAPHY

*Books*

Photo-electricity, H. Stanley Allen, 2nd Ed. 1925 (containing bibliography of articles up to 1925).

Photo-electricity, A. L. Hughes, 1914.

Die Lichtelektrische Erscheinungen, Pohl and Pringsheim, 1914.

National Research Council Report on Photo-electricity, A. L. Hughes, 1921.

*Recent Articles*

The Transmission of Pictures over Telephone Lines, H. E. Ives, J. W. Horton, R. D. Parker and A. B. Clark, Bell System Technical Journal, April, 1925.

Photo-electric Properties of Thin Films of Alkali Metals, H. E. Ives, Astrophysical Journal *60*, p. 209, 1924.

The Normal and Selective Photo-electric Effects in the Alkali Metals and their Alloys, H. E. Ives and A. L. Johnsrud, Astrophysical Journal *60*, p. 231, 1924.

Variation with Temperature of the Photo-electric Effect in Potassium Photo-electric Cells, H. E. Ives, Optical Society of America Journal *8*, p. 551, 1924.

The Influence of Temperature on the Photo-electric Effect of the Alkali Metals, H. E. Ives and A. L. Johnsrud, Optical Society of America Journal *11*, p. 565, 1925.

The Voltage-Current Relation in Central Anode Photo-electric Cells, Ives and Fry, Astrophysical Journal *66*, p. 1, 1922.

# Electric Circuit Theory and the Operational Calculus[1]

By JOHN R. CARSON

## CHAPTER IX

### *The Finite Line with Terminal Impedances*

So far in our discussions of wave propagation in lines and wave-filters, we have confined attention to the case where the impressed voltage is applied directly to the infinitely long line. We have found that, by virtue of this restriction, the indicial admittance functions of the important types of transmission systems are rather easily derived and expressible in terms of well known functions, and the essential phenomena of wave propagation clearly exhibited. In practice, however, we are concerned with lines of finite length with the voltage impressed on the line through a terminal impedance $Z_1$ and the distant end closed by a second terminal impedance $Z_2$. We now take up the problem presented by such a system.

Let $K=K(p)$ denote the characteristic operational impedance of the line, and $\gamma=\gamma(p)$ the operational propagation constant of the line. We have then

$$\begin{aligned} V&=Ae^{-\gamma x}+Be^{\gamma x}, \\ I&=\frac{1}{K}Ae^{-\gamma x}-\frac{1}{K}Be^{\gamma x}, \end{aligned} \tag{240}$$

where $A$ and $B$ are so far arbitrary constants. To determine these constants we assume an e.m.f. $E$ impressed on the line at $x=0$ through a terminal impedance $Z_1$ and the line closed at $x=s$ by a second terminal impedance $Z_2$. At $x=s$ we have therefore

$$Z_2 I = V$$

whence from (240)

$$\frac{Z_2}{K}e^{-\gamma s}A-\frac{Z_2}{K}e^{\gamma s}B=Ae^{-\gamma s}+Be^{\gamma s}$$

and

$$B=-\frac{1-\rho_2}{1+\rho_2}e^{-2\gamma s}A \tag{241}$$

where $\rho_2=Z_2/K$.

[1] Concluded from the issue of January, 1926.

At $x=0$ we have

$$V=E-Z_1I$$

whence

$$A+B=E-\frac{Z_1}{K}A+\frac{Z_1}{K}B,$$

$$(1+\rho_1)A+(1-\rho_1)B=E, \tag{242}$$

where $\rho_1=Z_1/K_1$.

From (241) and (242) we get

$$A=\frac{1+\rho_2}{(1+\rho_1)(1+\rho_2)-(1-\rho_1)(1-\rho_2)e^{-2\gamma s}}E$$

$$B=\frac{-(1-\rho_2)e^{-2\gamma s}}{(1+\rho_1)(1+\rho_2)-(1-\rho_1)(1-\rho_2)e^{-2\gamma s}}E$$

and finally

$$I_x=\frac{E}{K+Z_1}\frac{e^{-\gamma x}+\dfrac{1-\rho_2}{1+\rho_2}e^{-\gamma(2s-x)}}{1-\dfrac{1-\rho_1}{1+\rho_1}\dfrac{1-\rho_2}{1+\rho_2}e^{-2\gamma s}}. \tag{243}$$

If we replace $E$ by a unit e.m.f. we get the operational formula for the indicial admittance $A_x$; thus

$$A_x=\frac{\lambda}{K}\frac{e^{-\gamma x}+\mu_2e^{-\gamma(2s-x)}}{1-\mu_1\mu_2e^{-2\gamma s}}=\frac{1}{Z_x(p)} \tag{244}$$

where

$$\lambda=K/(K+Z_1),$$

$$\mu_1=\frac{1-\rho_1}{1+\rho_1}=\frac{K-Z_1}{K+Z_1},$$

$$\mu_2=\frac{1-\rho_2}{1+\rho_2}=\frac{K-Z_2}{K+Z_2}.$$

$K,\gamma,Z_1,Z_2,\mu_1$ and $\mu_2$ are, of course, functions of the operator $p$.

The integral equation corresponding to the operational formula (244) is

$$\frac{1}{pZ_x(p)}=\int_0^\infty e^{-pt}A_x(t)dt. \tag{245}$$

Now by (244) we can expand $1/Z_x(p)$; it is

$$\begin{aligned}\frac{1}{Z_x(p)}=\lambda\frac{e^{-\gamma x}}{K}+\lambda\mu_2\frac{e^{-\gamma(2s-x)}}{K}\\ +\lambda\mu_1\mu_2\frac{e^{-\gamma(2s+x)}}{K}+\lambda\mu_1\mu_2^2\frac{e^{-\gamma(4s-x)}}{K}\\ +\lambda\mu_1^2\mu_2^2\frac{e^{-\gamma(4s+x)}}{K}+\ldots.\end{aligned} \qquad (246)$$

Now we observe that $e^{-\gamma x}/K$ is simply the operational formula for the indicial admittance at point $x$ of an infinitely long line with unit e.m.f. impressed directly on the line at $x=0$. This will be denoted by $a_x(t)$. Similarly $e^{-\gamma(2s-x)}/K$ is the operational formula for the indicial admittance at point $(2s-x)$ with unit e.m.f. impressed directly on the line at $x=0$. This will be denoted by $a_{2s-x}(t)$, etc.

Recognition of this fact allows us to derive a formal solution in terms of a series of reflected waves. For let a set of functions $v_o,v_1,v_2,v_3,\ldots$ satisfy and be defined by the operational equations

$$\begin{aligned}v_o&=\lambda(p)=\lambda\\ v_1&=\lambda\mu_2\\ v_2&=\lambda\mu_1\mu_2\\ v_3&=\lambda\mu_1\mu_2^2, \text{ etc.}\end{aligned} \qquad (247)$$

It then follows from the preceding and theorem II that

$$A_x(t)=\frac{d}{dt}\int_0^t d\tau\left\{\begin{array}{c}v_o(t-\tau)a_x(\tau)+v_1(t-\tau)a_{2s-x}(\tau)\\ +v_2(t-\tau)a_{2s+x}(\tau)+\ldots\end{array}\right\}. \qquad (248)$$

If, therefore, we know the indicial admittance of the infinitely long line with unit e.m.f. directly applied and if we can solve the operational equations (247), then $A_x(t)$ is given by (248) by integration. This solution may well present formidable difficulty in the way of computation. It is, however, formally straightforward and the numerical computation is entirely possible, the only question being as to whether the importance of the problem justifies the necessary expenditure of time and effort. Without any computations, however, the solution (248) admits of considerable instructive interpretation by inspection. The first term represents the current at point $x$ of an infinitely long line in response to a unit e.m.f. impressed at $x=0$ through an impedance $Z_1$; $v_o=v_o(t)$ is the corresponding voltage across the line terminals proper. The second term is a reflected wave from the other terminal due to the terminal irregularity which exists there. The third term is a reflected wave from the sending end terminal, etc. The solution is therefore a wave solution and is expanded in a form which corresponds exactly with the sequence of phenomena, which it represents.

The solution takes a particularly instructive form when $Z_1=k_1K$ and $Z_2=k_2K$ where $k_1$ and $k_2$ are numerics. Then

$$\begin{aligned} v_o &= \frac{1}{1+k_1} \\ v_1 &= \frac{1}{1+k_1}\,\frac{1-k_2}{1+k_2} \\ v_2 &= \frac{1}{1+k_1}\,\frac{1-k_2}{1+k_2}\,\frac{1-k_1}{1+k_1}, \text{ etc.} \end{aligned} \tag{249}$$

and

$$A_x(t)=\frac{1}{1+k_1}\left\{\begin{array}{l} a_x(t)+\dfrac{1-k_2}{1+k_2}\,a_{2s-x}(t) \\ +\dfrac{1-k_1}{1+k_1}\,\dfrac{1-k_2}{1+k_2}\,a_{2s+x}(t)+\ \ldots\ . \end{array}\right\}. \tag{250}$$

If $k_1=0$, $k_2=1$ we have the case of the e.m.f. impressed directly on the sending end of the line and the distant end closed through its characteristic impedance; the solution reduces to

$$A_x(t)=a_x(t)$$

as, of course, it should be by definition.

If $k_1=0$ and $k_2=\infty$, we have the case of the line open-circuited at the distant end, and the solution reduces to

$$A_x(t)=\left\{a_x(t)-a_{2s-x}(t)-a_{2s+x}(t)+a_{4s-x}(t)+\ \ldots\right\}. \tag{251}$$

Finally, if both $k_1$ and $k_2$ are zero, the line is shorted and

$$A_x(t)=\left\{a_x(t)+a_{2s-x}(t)+a_{2s+x}(t)+a_{4s-x}(t)+\ \ldots\right\}. \tag{252}$$

The operational equations (247) admit of further interesting and instructive physical interpretation without computation. Consider a circuit consisting of an impedance $Z_1$ in series with an impedance $K$. Let a unit e.m.f. be applied to this circuit and let $v_o$ be the resultant voltage across the impedance $K$. Then, operationally,

$$v_o = \frac{K}{K+Z_1} = \lambda$$

so that $v_o$, thus defined in physical terms, is the $v_o$ of equations (247).

Now let this voltage be impressed on a circuit consisting of an impedance $2Z_2$ in series with an impedance $K-Z_2$ so that the total impedance is $K+Z_2$. Let the resultant voltage drop across the impedance element $K-Z_2$ be denoted by $v_1$; then operationally

$$v_1 = \frac{K}{K+Z_1} \cdot \frac{K-Z_2}{K+Z_2} = \lambda\mu_2$$

which agrees with $v_1$ as given by equation (247).

Similarly if voltage $v_1$ is applied to a circuit consisting of an impedance $2Z_1$ in series with an impedance $K-Z_1$ and if $v_2$ denote the voltage drop across impedance $K-Z_1$, then

$$v_2 = \lambda\mu_1\mu_2$$

We can thus see physically what the voltages $v_o, v_1, v_2$ . . mean in terms of simple circuits consisting of $K$ and $Z_1$ in series and $K$ and $Z_2$ in series respectively.

I shall now work out a specific problem exemplifying the preceding theory. The example is made as simple as possible for two reasons. First because its simplicity makes it more instructive than when the phenomena depicted and the essentials of the mathematical methods are obscured by complicated formulas and extensive computations. Secondly while the general method of solution illustrated is thoroughly practical we cannot hope to arrive at the numerical solutions of the complicated problems without a large amount of laborious computations. Problems involving transmission lines with complicated terminal impedances are among the most difficult, as regards actual numerical solution, of any which present themselves in mathematical physics. On the other hand, the formal solution (248) gives at a glance the essential character of the phenomena involved.

The specific problem we shall deal with may be stated as follows: A unit e.m.f. is directly impressed on the terminals of a transmission line of length $s$, the distant end of which is closed by a condenser $C_o$. The line is supposed to be non-dissipative, its constants being inductance $L$ and capacity $C$ per unit length. Required the current at any point $x(x<s)$ of the line.

We write $\sqrt{L/C}=k$, $1/\sqrt{LC}=v$: then by virtue of the preceding analysis of transmission line propagation the indicial admittance $a_x$ of the *infinitely long line* is given by

$$a_x=0, \text{ for } t<x/v,$$

$$=\frac{1}{k}, \text{ for } t\geqq x/v.$$

The operational characteristic impedance is, of course, $k=\sqrt{L/C}$, and the terminal impedances $Z_1$ and $Z_2$ are given by

$$Z_1=0,$$

$$Z_2=1/pC_o.$$

Referring now to equation (244) we have:—

$$\lambda=1, \quad \mu_1=1,$$

$$\mu_2=\frac{k-1/pC_o}{k+1/pC_o}=\frac{kC_op-1}{kC_op+1}.$$

Consequently, referring to equations (247), we have, operationally,

$$v_o=1$$

$$v_1=v_2=\frac{kC_op-1}{kC_op+1}$$

$$v_3=v_4=\left(\frac{kC_op-1}{kC_op+1}\right)^2$$

$$v_5=v_6=\left(\frac{kC_op-1}{kC_op+1}\right)^3.$$

In order to determine these functions we have therefore to solve the general operational equation

$$V_n=\left(\frac{kC_op-1}{kC_op+1}\right)^n$$

where $V_n$ denotes either $v_{2n-1}$ or $v_{2n}$.

In order to eliminate the coefficient $kC_o$, we make use of theorem VIII, and write

$$\phi_n=\left(\frac{p-1}{p+1}\right)^n.$$

In accordance with that theorem

$$V_n(t)=\phi_n(t/kC_o).$$

We therefore start with the operational equation

$$\phi_n = \left(\frac{p-1}{p+1}\right)^n.$$

Now the solution of this operational equation is very easy and can be expressed in a number of ways. We require it expressed in the form most easily computed. The following appears best adapted for our purposes. Consider the auxiliary operational equation:—

$$\sigma_n = \left(\frac{p-2}{p}\right)^n$$

$$= \left(p^n - 2\frac{n}{1!}p^{n-1} + 2^2\,\frac{(n)(n-1)}{2!}\,p^{n-2} + \ldots + (-1)^n 2^n\right)\frac{1}{p^n}.$$

The explicit solution is gotten by replacing $1/p^n$ by $t^n/n!$ and $p^n$ by $d^n/dt^n$, whence

$$\sigma_n(t) = \left(\frac{d^n}{dt^n} - 2\frac{n}{1!}\frac{d^{n-1}}{dt^{n-1}} + \ldots + (-1)^n 2^n\right)\frac{t^n}{n!},$$

$$= 1 - \frac{n}{1!}\frac{2t}{1!} + \frac{n(n-1)}{2!}\frac{(2t)^2}{2!} + \ldots + (-1)^n\frac{(2t)^n}{n!}.$$

But writing

$$\sigma_n = \left(\frac{p-2}{p}\right)^n = \frac{1}{H(p)}$$

it follows that

$$\phi_n = \left(\frac{p-1}{p+1}\right)^n = \frac{1}{H(p+1)}$$

$$= \frac{p+1}{p}\cdot\frac{p}{p+1}\,\frac{1}{H(p+1)}$$

$$= \left(1+\frac{1}{p}\right)\cdot\frac{p}{p+1}\,\frac{1}{H(p+1)}.$$

Referring now to theorem VII, we see that

$$\phi_n(t) = \left(1 + \int_0^t dt\right)\sigma_n(t)\cdot e^{-t}.$$

Since we have already solved for $\sigma_n(t)$, this determines $\phi_n(t)$ and hence $V_n(t)$. The functions $v_o, v_1, v_2 \ldots$ are therefore determined.

Now refer back to equation (248) giving the required current in terms of $v_o, v_1, v_2 \ldots$ and the admittances $a_x(t), a_{2s-x}(t), \ldots$. It follows at once by substitution of the preceding that

$$A_x(t) = \frac{1}{k}\left\{ v_o\left(t-\frac{x}{v}\right) + v_1\left(t-\frac{2s-x}{v}\right) + v_2\left(t-\frac{2s+x}{v}\right) + \ldots \right\}$$

the functions $v_o$, $v_1$, $v_2$ being zero for negative values of the argument. This result may possibly require a little explanation.

Consider the expression

$$\frac{d}{dt}\int_0^t f(t-\tau)\cdot 1(\tau)d\tau$$

where $1(t)$ denotes a function which is zero for $t<t_o$ and unity for $t \geq t_o$. It is evidently identical with the admittance $a_x(t)$ provided the proper value is assigned to $t_o$.

Now since $1(t)=0$ for $t<t_o$ and unity for $t \geq t_o$, the preceding may be written as zero for $t<t_o$, and

$$\frac{1}{k}\frac{d}{dt}\int_{t_o}^t f(t-\tau)d\tau \qquad \text{for } t \geq t_o$$

which is equal to $f(t-t_o)$.

If we set $x=0$, we get the current entering the line; thus

$$\begin{aligned} A_o(t) &= \frac{1}{k}\left\{ v_o(t) + v_1\left(t-\frac{2s}{v}\right) + v_2\left(t-\frac{2s}{v}\right) \right. \\ &\qquad \left. + v_3\left(t-\frac{4s}{v}\right) + v_4\left(t-\frac{4s}{v}\right) + \ldots \right\} \\ &= \frac{1}{k}\left\{ 1 + 2V_1\left(t-\frac{2s}{v}\right) + 2V_3\left(t-\frac{4s}{v}\right) \right. \\ &\qquad \left. + 2V_5\left(t-\frac{6s}{v}\right) + \ldots \right\}. \end{aligned}$$

This has been computed for the case where $\sqrt{sLC_o}=10$ and is shown in Fig. 26. Referring to this figure we see that the current jumps at $t=0$ to the value $\sqrt{C/L}=1/k$, and keeps this constant value for a time interval $2s/v$. At this instant the first reflected wave arrives and the current takes another jump, of $2/k$. Thereafter it begins to decrease very slowly until time $t=4s/v$ at which time it takes another jump of $2/k$. Thereafter we have a series of jumps of $2/k$ at time intervals $2s/v$, the current decreasing between successive jumps. The smooth curve is the indicial admittance of an oscillation circuit consisting of an inductance $sL$ in series with a capacity $C_o$. We see therefore, that the current in the line oscillates with discontinuous

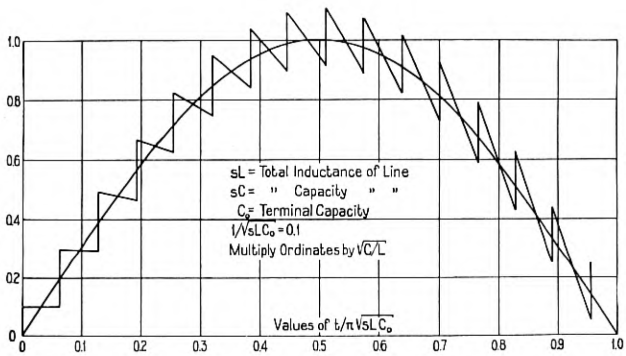



Fig. 26—Current entering non-dissipative line terminated by capacity $C_o$ unit E.M.F. applied to line

jumps about the current in the corresponding oscillation circuit. Since the whole circuit contains no resistance, the oscillations never die away, but continue to oscillate, as shown, about the curve

$$\sqrt{\frac{C}{L}}\sin\left(\frac{t}{\sqrt{sLC_o}}\right)$$

which is the indicial admittance of the corresponding oscillation circuit.

I shall now discuss a method of solving circuit theory problems, quite generally applicable to complicated networks, and particularly useful in dealing with transmission lines terminated in impedances. I have found it particularly useful in arriving at numerical solutions where other methods prove far more laborious. It is also of mathematical interest, as it applies another type of integral equation to the problems of electric circuit theory.

Suppose that we have a network with two sets of terminals as shown in Fig. 27.[7] Now suppose that terminals 22 are short circuited and a unit e.m.f. inserted between terminals 11. Let the resultant current flowing between terminals 11 be denoted by $S_{11}(t)=S_{11}$ and that

[7] Regarding conventions as to signs, see the Appendix to this chapter.

between terminals 22 by $S_{21}(t) = S_{21}$. $S_{11}$ is the driving point indicial admittance with respect to terminals 11 and $S_{21}$ the transfer indicial admittance of terminals 22 with respect to 11 under short circuit conditions.

Similarly if terminals 11 are shortcircuited and a unit e.m.f. inserted

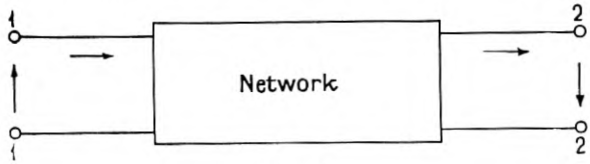



Fig. 27

between terminals 22 the current flowing between terminals 22 is denoted by $S_{22}(t) = S_{22}$ and that flowing between terminals 11 by $S_{12}(t) = S_{12}$. If the network is passive, i.e., contains no internal source of energy, it follows from the reciprocal theorem that $S_{21} = S_{12}$. As far as the two sets of terminals are concerned, the network is completely specified by the indicial admittances $S_{11}, S_{22}, S_{21} = S_{12}$.

Now let a voltage $V_1(t) = V_1$ be inserted between terminals 11, and a voltage $V_2(t) = V_2$ between terminals 22. The current flowing between terminals 11, denoted by $I_1$ is

$$I_1(t) = \frac{d}{dt}\int_0^t V_1(\tau)S_{11}(t-\tau)d\tau + \frac{d}{dt}\int_0^t V_2(\tau)S_{12}(t-\tau)\,d\tau \qquad (253)$$

while the corresponding current between terminals 22 is

$$I_2(t) = \frac{d}{dt}\int_0^t V_2(\tau)S_{22}(t-\tau)d\tau + \frac{d}{dt}\int_0^t V_1(\tau)S_{21}(t-\tau)\,d\tau \qquad (254)$$

Now consider two networks of indicial admittances $S_{11}$, $S_{22}$, $S_{12} = S_{21}$ and $T_{11}, T_{22}, T_{12} = T_{21}$ respectively and let them be connected in tandem as shown in Fig. 28 to form a compound network.

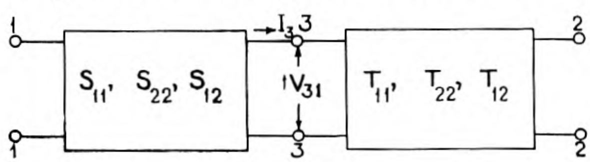



Fig. 28

We require the indicial admittances of the compound network in terms of the indicial admittances of the component networks.

Short circuit terminals 22 of the compound network and insert a unit e.m.f. between terminals 11. Let $V_{31}(t)$ denote the resultant voltage between terminals 33 measured in the direction of the arrow, and $I_3$ the current flowing between the networks. We have then the two following expressions for the current $I_3$.

$$I_3=S_{21}(t)-\frac{d}{dt}\int_0^t V_{31}(\tau)S_{22}(t-\tau)d\tau \tag{255}$$

and

$$I_3=\frac{d}{dt}\int_0^t V_{31}(\tau)T_{11}(t-\tau)d\tau. \tag{256}$$

Equating we get

$$\frac{d}{dt}\int^t V_{31}(\tau)\,[S_{22}(t-\tau)+T_{11}(t-\tau)]\,d\tau=S_{21}(t). \tag{257}$$

By precisely similar reasoning, if terminals 11 are short circuited and a unit e.m.f. inserted between terminals 22, and the corresponding voltage across terminals 33 denoted by $V_{32}$, we have [8]

$$\frac{d}{dt}\int_0^t V_{32}(\tau)\,[S_{22}(t-\tau)+T_{11}(t-\tau)]\,d\tau=T_{12}(t). \tag{258}$$

Equations (257) and (258) are integral equations of the Poisson type which completely determine $V_{31}$ and $V_{32}$ in terms of the indicial admittances $S$ and $T$. We shall discuss the solution of these equations presently.

If $U_{11}, U_{22}, U_{21}=U_{12}$ denote the indicial admittances of the compound network we have at once

$$U_{11}=S_{11}(t)-\frac{d}{dt}\int_0^t V_{31}(\tau)S_{12}(t-\tau)d\tau \tag{259}$$

$$U_{22}=T_{22}(t)-\frac{d}{dt}\int_0^t V_{32}(\tau)T_{21}(t-\tau)d\tau \tag{260}$$

and

$$\begin{aligned} U_{21}=U_{12}&=\frac{d}{dt}\int_0^t V_{31}(\tau)T_{21}(t-\tau)d\tau \\ &=\frac{d}{dt}\int_0^t V_{32}(\tau)S_{12}(t-\tau)d\tau. \end{aligned} \tag{261}$$

If, therefore, equations (257) and (258) are solved for $V_{31}$ and $V_{32}$, the required indicial admittances of the compound network are given

[8] $V_{32}$ being opposite to $V_{31}$ in direction.

by (259), (260) and (261) in terms of the indicial admittances of the component networks.

A simple example will now be worked out illustrating the method of solution just discussed. Suppose that a unit e.m.f. is impressed on a transmission line (infinitely long) of distributed constants $R,L,C,$ through a terminal resistance $R_o$. Required the terminal line voltage $V$.

The operational equation of this problem is gotten in the usual manner. The current entering the line is

$$V\sqrt{\frac{Cp}{Lp+R}}.$$

It is also obviously equal to $\frac{1}{R_o}(1-V)$: equating the two expressions, and rearranging we get:—

$$V=\frac{\sqrt{\frac{Lp+R}{Cp}}}{R_o+\sqrt{\frac{Lp+R}{Cp}}}.$$

Writing $R/2L=\rho$ and setting $R_o=\sqrt{L/C}$, this becomes

$$V=\frac{\sqrt{1+2\rho/p}}{1+\sqrt{1+2\rho/p}}. \qquad (262)$$

This operational equation can, of course, be solved in a number of ways, though, as a matter of fact, its numerical solution is quite troublesome. This point will be returned to later: we shall first formulate the problem in accordance with the method just discussed.

The indicial admittance of the line is known; it is

$$\sqrt{\frac{C}{L}}e^{-\rho t}I_0(\rho t)=A(t).$$

Consequently the current entering the line is explicitly

$$\frac{d}{dt}\int_0^t V(\tau)A(t-\tau)d\tau.$$

But the current is also equal to $\frac{1}{R_o}(1-V(t))$; equating, we get

$$V(t)=1-R_o\frac{d}{dt}\int_0^t V(\tau)A(t-\tau)d\tau.$$

Performing the indicated differentiations

$$V(t) = 1 - R_o A(o)\, V(t) - R_o \int_0^t V(\tau) A'(t-\tau) d\tau.$$

Now $A(o) = \sqrt{\frac{C}{L}}$ and

$$A'(t) = \rho e^{-\rho t}(I_1(\rho t) - I_o(\rho t)) \sqrt{\frac{C}{L}}$$

and $R_o = \sqrt{L/C}$; therefore the equation becomes

$$V(t) = \frac{1}{2} + \frac{\rho}{2} \int_0^t V(t-\tau)[I_o(\rho\tau) - I_1(\rho\tau)] e^{-\rho\tau} d\tau.$$

As a matter of convenience we change the time scale to $\rho t$, and get

$$V(t) = \frac{1}{2} + \frac{1}{2} \int_0^t V(t-\tau)[I_o(\tau) - I_1(\tau)] e^{-\tau} d\tau \tag{263}$$

$$= \frac{1}{2} - \frac{1}{2} \int_0^t d\tau\, V(t-\tau) \frac{d}{d\tau} e^{-\tau} I_o(\tau),$$

where it is understood that $t$ is actually $\rho t$. This is the integral equation of the problem and is in the canonical form of Poisson's integral equation.

Before solving this equation numerically I shall show how a simple approximate solution is obtainable immediately; an advantage often attaching to this type of integral equation.

The function $\frac{d}{dt} e^{-t} I_o(t)$ is equal to $-1$ for $t=0$ and converges rapidly to zero. $V(t)$ has, as we know from the operational equation, the initial value 1/2 and the final value 1. Neither function changes sign. It follows from the mean value theorem that the equation can be written as

$$V(t) = \frac{1}{2} - \frac{1}{2} V(t) \int_0^{\alpha t} \frac{d}{dt} e^{-t} I_o(t) dt$$

where $\alpha \leq 1$. Integrating

$$V(t) = \frac{1}{2} - \frac{1}{2} V(t)\, [e^{-\alpha t} I_o(\alpha t) - 1]$$

and

$$V(t) = \frac{1}{1 + e^{-\alpha t} I_o(\alpha t)}. \tag{264}$$

The correct initial and final values of $V(t)$ result for all final values of $\alpha \leq 1$; so that approximately

$$V(t)=\frac{1}{1+e^{-t}I_o(t)}.$$

This equation, while not exact, except for $t=0$ and $t$ very large, shows faithfully the general character of $V(t)$ and the way it approaches its final value unity. For large values of $t$

$$e^{-t}I_o(t)=1/\sqrt{2\pi t}$$

whence

$$V(t)=\frac{1}{1+1/\sqrt{2\pi t}}, \quad t\geq 8. \tag{264-a}$$

Approximations of the foregoing type are not always possible and may not be of sufficient accuracy. I shall therefore give next a method of numerical solution which is generally applicable to integral equations of this type and works quite well in practice. We shall write the integral equation in the more general form

$$u(x)=f(x)+\int_0^x u(x-y)k(y)dy \tag{265}$$

where $f(x)$ and $k(y)$ are known and $u(x)$ unknown. The method depends on the numerical integration of the definite integral. Let us divide the $x$ scale into small intervals $d$ and for convenience write

$$u(nd)=u_n$$

$$f(nd)=f_n$$

$$k(nd)=k_n.$$

Now from the integral equation we have at once

$$u(o)=u_o=f_o,$$

$$u(d)=u_1=f_1+\int_0^d u(d-y)k(y)dy.$$

Now if $d$ is taken sufficiently small

$$\int_0^d u(d-y)k(y)dy=\frac{d}{2}[u_1k_o+u_ok_1],$$

whence

$$u_1=f_1+\frac{d}{2}[u_1k_o+u_ok_1]$$

and

$$u_1=\frac{1}{1-k_od/2}[f_1+u_ok_1d/2]$$

which determines $u_1$ since $u_o$ is known. Similarly

$$u_2=f_2+d\,[\tfrac{1}{2}u_ok_2+u_1k_1+\tfrac{1}{2}u_2k_o]$$

which determines $u_2$. Proceeding in the same manner

$$u_3=f_3+d\,[\tfrac{1}{2}u_ok_3+u_1k_2+u_2k_1+\tfrac{1}{2}u_3k_o], \text{ etc.}$$

In this way we determine the value of $u(x)$, point by point from the recurrence formula

$$u_n=\frac{f_n+d\,[\tfrac{1}{2}u_ok_n+u_1k_{n-1}+u_2k_{n-2}+\;..\;+u_{n-1}k_1]}{1-\tfrac{1}{2}k_od}. \tag{266}$$

The result of the application of numerical integration, in accordance with formula (266), to the integral equation (263) is shown in Fig. (29). The dotted curve is a plot of the approximate solution as

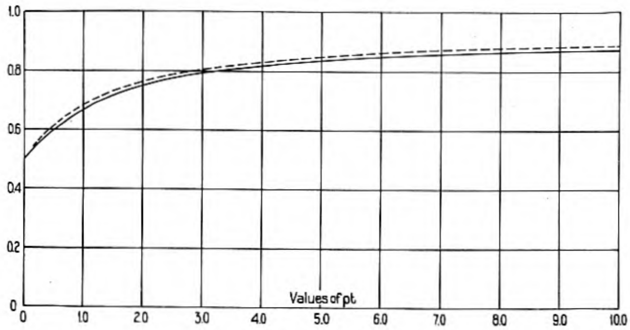



Fig. 29—Line terminal voltage unit E.M.F. impressed on line through resistance $R_0=\sqrt{L/C}$

given by equation (264), for $\alpha=1$. We see that the voltage starts with the value 1/2 and slowly reaches its ultimate value, unity, its approach to unity, for large values of $t$, being in accordance with the formula

$$V(t)=\frac{1}{1+1/\sqrt{2\pi t}}\,.$$

The application of the foregoing method to the transmission line problem proceeds as follows. Let $S_{11}(t)$, $S_{22}(t)$ and $S_{12}(t)$ be the short indicial admittances of the line. $S_{11}(t)$ is the current entering the line (at $x=0$) with unit e.m.f. directly impressed and the distant end short circuited. $S_{12}(t)$ is the current at $x=s$ under the same circumstances. Consequently from (252)

$$\begin{aligned} S_{11}(t) &= a_o(t) + 2a_{2s}(t) + 2a_{4s}(t) + \dots \\ S_{12}(t) &= 2\{a_s(t) + a_{3s}(t) + a_{5s}(t) + \dots\}. \end{aligned} \tag{267}$$

$S_{22}$ is clearly equal to $S_{11}$ by symmetry.

Now suppose that an e.m.f. $E = E(t)$ is impressed on the line at $x=0$, $t=0$, through a terminal impedance $Z_1$, and the distant end $(x=s)$ closed through an impedance $Z_2$. We suppose these terminal impedances and the actual impressed e.m.f. replaced by the actual line voltages $V_1$ and $V_2$, impressed directly on the line at $x=0$ and at $x=s$ are

$$\begin{aligned} I_o(t) &= \frac{d}{dt}\int_0^t S_{11}(t-\tau)\, V_1(\tau)d\tau \\ &\quad - \frac{d}{dt}\int_0^t S_{12}(t-\tau)\, V_2(\tau)d\tau, \end{aligned} \tag{268}$$

$$\begin{aligned} I_s(t) &= -\frac{d}{dt}\int_0^t S_{12}(t-\tau)\, V_1(\tau)d\tau \\ &\quad + \frac{d}{dt}\int_0^t S_{22}(t-\tau)\, V_2(\tau)d\tau. \end{aligned} \tag{269}$$

But the current at $x=s$ is also equal to the current in the terminal impedance $Z_2$ in response to the terminal voltage $V_2$: denoting by $\alpha_2(t)$ the indicial admittance of $Z_2$ it is

$$I_s(t) = \frac{d}{dt}\int_0^t \alpha_2(t-\tau)\, V_2(\tau)d\tau. \tag{270}$$

Similarly the current entering the line at $x=0$ is the current flowing in the terminal impedance $Z_1$ in response to the e.m.f. $E - V_1$. Denoting by $\alpha_1(t)$ the indicial admittance of $Z_1$, it is

$$I_o(t) = \frac{d}{dt}\int_0^t \alpha_1(t-\tau)\{E(\tau) - V_1(\tau)\}d\tau. \tag{271}$$

Equating equation (268) and (271) and (269) and (270) we eliminate $I_o(t)$ and $I_s(t)$ and get

$$\begin{aligned} \int_0^t [S_{11}(t-\tau) + \alpha_1(t-\tau)]\, V_1(\tau)d\tau - \int_0^t S_{12}(t-\tau)\, V_2(\tau)d\tau \\ = \int_0^t \alpha_1(t-\tau)E(\tau)d\tau, \end{aligned} \tag{272}$$

$$-\int_0^t S_{12}(t-\tau)\, V_1(\tau)d\tau + \int_0^t [S_{22}(t-\tau) - \alpha_2\,(t-\tau)]\, V_2(\tau)d\tau = 0. \tag{273}$$

*These two equations are simultaneous integral equations of the Poisson type in $V_1$ and $V_2$, which completely determine these voltages provided the admittances and the impressed voltages are known. They therefore represent the application of a new type of integral equation to the problem of electric circuit theory.*

The numercial solution of the general case, either by (248) or (272–273) is necessarily laborious when the terminal impedances are complicated and is only justified when the technical importance of the problem is considerable. I wish, however, to emphasize two points in this connection: the numerical solution is always entirely possible and, compared with other and older forms of solution, enormously simpler. One has only to inspect the classical forms of solution of problems of the type to realize the truth of this last statement.

I shall now give two applications of equations (272–273) to specific problems, in one of which an approximate solution of the integral equation can be gotten, and in the other of which numerical integration is applied.

*Problem I.* Given a non-inductive cable of distributed constants $C$ and $R$ and length $s$, with unit e.m.f. applied at $x=0$, while at $x=s$ the cable is closed by a condenser $C_o$. Required the terminal voltage $V(t)$ across the condenser $C_o$.

We first write down the short-circuit indicial admittances of the cable; from equation (168) of a preceding section and equation (267) they are:—

$$S_{11}(t)=S_{22}(t)$$
$$=\sqrt{\frac{C}{\pi Rt}}\left\{1+2e^{-\frac{4\beta}{t}}+2e^{-\frac{16\beta}{t}}+2e^{-\frac{36\beta}{t}}+\ldots\right\}, \tag{274}$$

$$S_{12}(t)=S_{21}(t)$$
$$=2\sqrt{\frac{C}{\pi Rt}}\left\{e^{-\frac{\beta}{t}}+e^{-\frac{9\beta}{t}}+e^{-\frac{25\beta}{t}}+\ldots\right\}, \tag{275}$$

where $\beta=s^2RC/4$.

Now the current at $x=s$ is equal to

$$S_{12}(t)-\frac{d}{dt}\int_0^t V(\tau)S_{22}(t-\tau)d\tau.$$

It is also the condenser current due to the voltage $V(t)$; that is

$$C_o\frac{d}{dt}V(t).$$

Equating the two expressions and integrating we get

$$C_o V(t) = \int_0^t S_{12}(\tau)d\tau - \int_0^t V(\tau)S_{22}(t-\tau)d\tau \tag{276}$$

which is the integral equation of the problem. In order to get an approximate solution without detailed computation we assume that the cable is long. In this case the leading terms of (274) and (275) are large compared with the terms following: Furthermore $S_{12}(t)$ builds up very slowly while $S_{22}(t)$ is a rapidly varying function. A good approximation therefore results if we take $V(\tau)$ outside the integral sign in (276) and write

$$C_o V(t) = \int_0^t S_{12}(\tau)d\tau - V(t)\int_0^t S_{22}(\tau)d\tau$$

whence

$$V(t) = \frac{1}{C_o}\frac{\int_0^t S_{12}(t)dt}{1+\frac{1}{C_o}\int_0^t S_{22}(t)dt}. \tag{277}$$

This approximation is quite good for long cables and shows the way $V(t)$ builds up quite truthfully. We see that $V$ is initially zero, and builds up ultimately to unity. For large values of $t$, it becomes

$$V(t) = \frac{\int_0^t S_{12}(t)dt}{\int_0^t S_{22}(t)dt}. \tag{278}$$

This is the approximate formula also for the open circuit voltage, as may be seen by setting $C_o=0$ in (277).

In electric circuit problems, it is often sufficient, as implied above, to know qualitatively the behavior of an electric system without going through the labor of detailed computation. For this purpose the formulation of the problem as a Poisson Integral Equation is particularly well adapted. A simple example will be given, which can be checked from the known solution. Suppose that we require the voltage $V$ at point $x$ of an infinitely long transmission line $(L,R,C)$ in response to a unit e.m.f. impressed at $x=0$. This is, of course, known from formula (211-a): we shall here be concerned, however, with approximate solutions from the Poisson integral equation of the problem.

If $a_x(t)$ denote the indicial admittance of the line at point $x$, then the current at point $x$ is simply $a_x(t)$, which is given by formula (210-a). But if $V(t)$ is the voltage at point $x$, the current is also given by

$$\frac{d}{dt}\int_0^t V(\tau)a_o(t-\tau)d\tau.$$

Equating these two expansions, we get the integral equation of the problem

$$\frac{d}{dt}\int_0^t V(\tau)a_o(t-\tau)d\tau = a_x(t).$$

Now if we write $T = \rho t - A$ where $\rho = R/2L$ and $A = \frac{xR}{2}\sqrt{\frac{C}{L}}$, then

$$a_x = \sqrt{\frac{C}{L}}e^{-(T+A)}I_o\sqrt{T(T+2A)}, \quad T \geq 0,$$

and in terms of the relative time $T$, the integral equation is reducible to

$$\frac{d}{dT}\int_0^T V(T-\tau)e^{-\tau}I_o(\tau)d\tau = e^{-(T+A)}I_o(\sqrt{T(T+2A)})$$

while the exact formula for $V$ is by (211-a)

$$V(T) = e^{-A} + Ae^{-A}\int_0^T \frac{e^{-\tau}I_1\,(\sqrt{\tau(\tau+2A)})}{\sqrt{\tau(\tau+2A)}}d\tau.$$

From the integral equation it is easy to establish superior and inferior limits for $V(T)$; it is

$$V(T) \leq e^{-A}\,\frac{I_o(\sqrt{T(T+2A)})}{I_o(T)} = V_a(T),$$

$$\geq \frac{\int_0^T e^{-T}\,I_o(T)\,V_a(T)dT}{\int_0^T e^{-T}\,I_o(T)dT} = V_b(T).$$

Both formulas give the correct initial and final values of $V$; namely $e^{-A}$ and unity. Since $V$ lies between $V_a$ and $V_b$, the mean value $(V_a + V_b)/2$ also has correct initial and final values and should be a better approximation than either. The table given below shows the orders of approximation obtainable from the case where $A = 3$. It is evident from this table that the foregoing approximate formulas exhibit the form of $V(T)$ qualitatively in a quite satisfactory manner.

| $T$ | $V_a$ | $V_b$ | $\frac{1}{2}(V_a+V_b)$ | $V$ |
|---|---|---|---|---|
| 0 | 0.05 | 0.05 | 0.05 | 0.05 |
| 2 | 0.25 | 0.12 | 0.18 | 0.17 |
| 4 | 0.39 | 0.19 | 0.29 | 0.26 |
| 6 | 0.50 | 0.23 | 0.36 | 0.32 |
| 8 | 0.57 | 0.27 | 0.42 | 0.37 |
| 10 | 0.64 | 0.31 | 0.47 | 0.41 |
| 12 | 0.69 | 0.34 | 0.51 | 0.44 |
| 15 | 0.74 | 0.38 | 0.56 | 0.48 |
| 18 | 0.78 | 0.41 | 0.60 | 0.52 |

*Problem II.* Our second illustrative problem may be stated as follows:—A unit e.m.f. is impressed on a transmission line of length $s$ and distributed constants $L,R,C$. At $x=s$ the line is closed by a resistance $R_o$ in parallel with an inductance $L_o$. Required the current in the terminal resistance. If $V(t)$ denotes the terminal voltage, the current at $x=s$ is given by

$$S_{12}(t)-\frac{d}{dt}\int_0^t V(\tau)S_{22}(t-\tau)d\tau.$$

It is also equal to the current flowing into the terminal impedance; that is

$$\frac{1}{R_o}V(t)+\frac{1}{L_o}\int_0^t V(\tau)d\tau.$$

Equating and rearranging

$$\left[\frac{1}{R_o}+S_{22}(o)\right]V(t)=S_{12}(t)-\int_0^t V(\tau)\left[\frac{1}{L_o}+S'_{22}(t-\tau)\right]d\tau. \quad (279)$$

Now the short circuit admittance $S_{22}$ and $S_{12}$ are given by formula (210-a) of a preceding chapter, and $S_{22}(o)=\sqrt{C/L}$. In order to apply numerical integration to (279), numerical values must be assigned to the constants. We take

$$R_o=\sqrt{L/C}=1935 \text{ ohms},$$

$$L_o=0.4 \text{ henry},$$

$$\frac{R}{2L}=\rho=292,$$

$$v=1/\sqrt{LC}=1.105\times 10^4,$$

$$s=100.$$

The results of the numerical evaluation of equation (279), with these values inserted, is shown in Fig. 30. The voltage is identically zero until $vt=100$; $t=100/v$ is the time of propagation of the line. At that instant it jumps to the value $e^{-\rho t}=e^{-100\rho/v}$ and then begins to die away rapidly due to the draining action of the inductance.

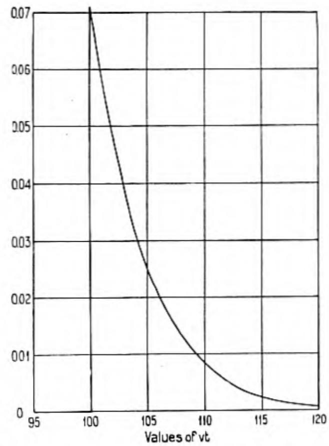



Fig. 30—Voltage across terminal impedance on smooth line

The effect of secondary reflection is insignificant and therefore not shown. The current in the terminal resistance is $V/R_o$ so that it is given by the same curve.

I have reserved until the last the exposition of the *expansion theorem solution* as applied to transmission lines with terminal impedances, for the reason that it is the least powerful and the most restricted, although most closely resembling the classical form of solution. Furthermore, it does not represent the sequence of physical phenomena, in fact it is not a *wave* solution, but a solution in terms of normal or characteristic vibration. In practical application its usefulness is restricted to the non-inductive cable.

It will be recalled that the expansion theorem solution is formulated as follows:—

If

$$A=1/Z(p)$$

is the operational equation of the problem, then the explicit solution is

$$A(t)=\frac{1}{Z(o)}+\sum_1^n \frac{e^{p_k t}}{p_k Z'(p_k)}$$

where $p_1, p_2 \ldots$ are the roots of the equation $Z(p)=0$.

Let us apply this formula to the case of a line of length $s$, with unit e.m.f. directly applied at $x=s$, and line short circuited at $x=s$. Referring to equation (244) and putting $\lambda=\mu_1=\mu_2=1$ we get

$$A_x=\frac{1}{K}\frac{\cosh \gamma(s-x)}{\sinh \gamma s}=\frac{1}{Z_x(p)} \tag{280}$$

as the operational formula of the problem. This can be written as

$$A_x=(Cp+G)\frac{\cosh \gamma(s-x)}{\gamma \sinh \gamma s}=\frac{1}{Z_x(p)} \tag{281}$$

where in the general case,

$$\gamma=\sqrt{(Lp+R)(Cp+G)}. \tag{282}$$

The values of $\gamma$ for which $Z_x(p)$ vanishes are the roots of the transcendental equation

$$\sinh \gamma s=0$$

excluding zero. These roots are infinite in number: Let $\gamma_m$ be the $m^{\text{th}}$ root; then

$$\gamma_m=i\frac{m\pi}{s}, \quad m=1,2, \ldots \infty. \tag{283}$$

The corresponding values of $p_m$ are then gotten by solving (282) for $p$ and writing $\gamma=\gamma_m$.

The explicit solution of the operational equation (281) is then

$$\begin{aligned} A_x(t) &= \frac{1}{Z_x(o)}+\sum\left(C+\frac{G}{p_m}\right)\frac{\cosh \gamma_m(s-x)}{s\gamma_m\dfrac{d\gamma_m}{dp_m}\cosh \gamma_m s}e^{p_m t}, \\ &= \frac{1}{Z_x(o)}+\sum (C+G/p_m)\frac{\cosh \gamma_m x}{s\gamma_m\dfrac{d\gamma_m}{dp_m}}e^{p_m t}. \end{aligned} \tag{284}$$

Let us apply this to the non-inductive, non-leaky cable in which $L=G=0$ and $\gamma=\sqrt{RCp}$, so that

$$p_m=\gamma_m^2/RC=-\frac{m^2\pi^2}{s^2RC},$$

and

$$\gamma_m\frac{d\gamma_m}{dp_m}=\frac{RC}{2},$$

Also $Z_x(o) = sR$. We thus get

$$A_x(t) = \frac{1}{sR} + \frac{2}{sR} \sum_{m=1}^{\infty} \cos \frac{m\pi}{s} x . e^{-\frac{m^2\pi^2}{s^2RC}t}. \tag{285}$$

This is a thoroughly practical formula for computation, owing to the rapid convergence of the series. In fact, for this particular line termination chosen, it is probably the simplest and most easily computed form of solution. These advantages depend, however, strictly on two facts. First, the fact that the line is taken as non-inductive and secondly that the terminations chosen are those of a short circuit. In fact, as we shall see, it is only in the case of the non-inductive cable that this type of solution is of any practical value.

There is one other point which should be carefully observed in connection with this solution (285). This is that it is not expressed in terms of a series of direct and reflected waves, corresponding to the sequence of physical phenomena, but in terms of *normal* or *characteristic vibrations*. This point will be returned to later.

Let us now attempt to apply this type of solution to the transmission line, $L,R,C,G$. Writing

$$\rho = \frac{R}{2L} + \frac{G}{2C},$$

$$\sigma = \frac{R}{2L} - \frac{G}{2C},$$

$$v = 1/\sqrt{LC}.$$

We have

$$\gamma^2 = \frac{1}{v^2}[(p+\rho)^2 - \sigma^2]$$

whence

$$p_m = -\rho \pm v\sqrt{\gamma_m^2 + \frac{\sigma^2}{v^2}}$$

$$= -\rho \pm iv\sqrt{\left(\frac{m\pi}{s}\right)^2 - \frac{\sigma^2}{v^2}}, \qquad m = 1, 2, \ldots$$

$$\gamma_m \frac{d\gamma_m}{dp_m} = \frac{1}{v^2}(p_m + \rho)$$

$$= \pm \frac{i}{v}\sqrt{\left(\frac{m\pi}{s}\right)^2 - \frac{\sigma^2}{v^2}}.$$

Setting $G=0$ for simplicity and substituting in (284) we get, after easy simplifications,

$$A_x(t)=\frac{1}{sR}+\frac{2vC}{s}\sum\frac{\cos\left(\frac{m\pi}{s}x\right)}{\sqrt{\left(\frac{m\pi}{s}\right)^2-\frac{\rho^2}{v^2}}}\sin\left(vt\sqrt{\left(\frac{m\pi}{s}\right)^2-\frac{\rho^2}{v^2}}\right)e^{-\rho t}. \quad (286)$$

If we write

$$\sqrt{\left(\frac{m\pi}{s}\right)^2-\frac{\rho^2}{v^2}}=\mu_m\frac{m\pi}{s}$$

(286) can be written as

$$A_x(t)=\frac{1}{sR}+\frac{vC}{s}\sum\frac{e^{-\rho t}}{\mu_m\frac{m\pi}{s}}\left\{\sin\frac{m\pi}{s}(\mu_m vt-x)+\sin\frac{m\pi}{s}(\mu_m vt+x)\right\}. \quad (287)$$

This type of solution is often referred to as a *wave* solution and the component terms of the series regarded as travelling waves. As a matter of fact it is a solution in terms of normal or characteristic vibrations, each of which is to be regarded as instantaneously produced at time $t=0$. The solution in terms of true waves has been fully discussed in the preceding.

Formula (287) is practically useless for computation on account of the slow convergence of the series (the series are only conditionally convergent), and cannot be interpreted to bring out the existence of the actual direct and reflected waves and the physical character of the phenomena it formulates. In fact, as stated above, this form of solution is useful only in connection with the non-inductive cable.

In the cases considered above we have taken the simplest possible terminations--these of short circuits in which case the roots of $Z(p)$ are easily evaluated. If, however, the line is closed by arbitrary impedances, the case is quite different, and the location of the roots becomes, except for simple impedances, and then only in the case of the non-inductive cable, practically impossible. While, therefore, the expansion theorem solution can be formally written down, its actual numerical evaluation is a practical impossibility, except in a few cases. For this reason it will not be considered further here.

The physically artificial character of the expansion solution, as applied to transmission lines, may be seen from the following considerations. When a wave is sent into the line, for a finite time equal to the time of the propagation of the line, it is independent of the character of the distant termination. Yet in the expansion solution every term involves and is dependent upon the impedance constituting the distant termination. Evidently, from physical considerations, the series of component vibrations making up the complete solution must therefore so combine as to annihilate the effect of the distant termination for a finite time. The solution is, therefore, mathematically correct but physically artificial.

*Note on Integral Equations.*

An integral equation is defined as an equation in which the unknown function occurs under a sign of integration; the process of determining the unknown function is called solving the equation.

Integral equations are of great importance in mathematical physics and in recent years very considerable work has been done on them from the standpoint of pure analysis.

The types of integral equations with which we are concerned in the present work are *Laplace's Equation*

$$F(p) = \int_0^\infty e^{-pt} f(t) dt$$

and *Poisson's Equation*

$$\phi(x) = f(x) + \int_0^x \phi(y) K(x-y) dy.$$

But little work has been done on Laplace's Equation from the standpoint of pure analysis; its most extensive and useful applications appear to be in connection with the Operational Calculus. Practical methods of solution are extensively discussed in the text.

We shall now briefly discuss the solution of Poisson's Equation.

The formal series solution, which is absolutely convergent, is obtained by successive substitution. Thus suppose we write

$$\phi(x) = \phi_o(x) + \phi_1(x) + \phi_2(x) + \ldots.$$

and define the terms of the series in accordance with the scheme

$$\phi_o(x) = f(x),$$

$$\phi_1(x) = \int_0^x \phi_o(y) K(x-y) dy,$$

$$\phi_2(x) = \int_0^x \phi_1(y) K(x-y) dy, \text{ etc.,}$$

the resulting series satisfies the integral equation and is absolutely convergent. It is, however, practically useless for computation or interpretation.

A power series solution, when it exists, can be gotten by repeated differentiation; thus

$$\phi(o) = f(o),$$

$$\phi'(x) = f'(x) + \phi(o)K(x) + \int_0^x \phi'(x-y)K(y)dy,$$

$$\phi'(o) = f'(o) + \phi(o)K(o)$$

In this way all the derivatives at $x=0$ are calculable; let them be denoted by $\phi_o, \phi_1, \phi_2 \ldots$ Then

$$\phi(t) = \phi_o + \phi_1\frac{x}{1!} + \phi_2\frac{x^2}{2!} + \ldots .$$

This form of solution, also, is of limited practical usefulness, except for small values of $x$.

A number of mathematicians, including Wittaker and Bateman, have studied the question of numerical solution and suggested other processes. After quite extensive study of the question, however, the writer is of the opinion that point-by-point numerical integration like that discussed in the text is, in general, the most practical, rapid and accurate method of numerical solution. This judgment is confirmed by G. Prasad who, in a paper on the Numerical Solution of Integral Equations delivered before the International Mathematical Congress (Toronto, 1924), discusses the whole question and arrives at the same conclusion.

In the text, numerical integration is carried out in accordance with Simpson's Rule. It is possible, of course, to employ more complicated and refined formulas for approximate quadrature. It is the writer's opinion that this is hardly justified in practical problems and that the required accuracy is more simply obtained by employing smaller intervals.

### *Appendix to Chapter IX. Note on Conventions as to Signs in Networks*

In the network shown on page 196 the arrows indicate the directions chosen as positive in the network itself, quite regardless of the presence of any e.m.fs. and currents.

The sign attributed to a current, an e.m.f., or a voltage is positive if the current, e.m.f., or voltage is in the positive direction; otherwise the sign is negative.

Stated more fully:

A current at a specific point (at a specific instant of time) is positive if it is flowing in the positive direction; negative if flowing in the negative direction.

An e.m.f. or a voltage between two points is positive if the potential increases in the positive direction between the two points; negative if the potential increases in the negative direction. (It may be noted that this convention makes the sign of a voltage the same as the sign of that e.m.f. which could be inserted between the two points without producing any effects in the network.)

## CHAPTER X

### Introduction to the Theory of Variable Electric Circuits[9]

In the preceding chapters it has everywhere been assumed that the networks are *invariable*: that is to say, that the constants and connections of the network do not vary or change with time. In many important technical problems, however, we wish to know, not merely what happens when an electromotive force is applied to an invariable network, but the effect of suddenly changing a circuit constant or of introducing a variable circuit element. In the present chapter we shall show that this type of problem can be dealt with by a simple extension of the methods discussed in the preceding chapters.

The simplest and at the same time one of the most technically important problems of this type is the effect of sudden short circuits and sudden open circuits on an energized network or system. This type of problem will serve as an introduction to the more general theory.

*The Sudden Short Circuit*

Consider the network shown in Fig. 31.

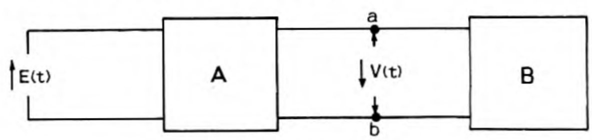



Fig. 31

This network, which for generality is supposed to consist of two parts A and B, indicated schematically, is energized by an electromotive force $E(t)$ which produces a voltage $V(t)$ between the points *ab*. The voltage $V(t)$ is calculable by usual methods from $E(t)$ and the constants and connections of the network, supposed to be specified.

[9] The material in this chapter is largely taken from a paper by the writer on "Theory and Calculation of Variable Electrical Systems," Phys. Rev. Feb. 1921.

We now suppose that, at reference time $t=0$, a short circuit is suddenly placed across $ab$; and require the effect of this short circuit on the distributions of currents in the network. The solution of this problem is based on the following proposition:

*The effect of the short circuit is precisely the same as the insertion at time $t=0$ of a voltage $-V(t)$, equal and opposite to $V(t)$, between points a and b.*

The resultant currents in the system for $t \geq 0$ are then composed of two components:—

(1) The currents which would exist in the invariable network, in the absence of the short circuit, due to the impressed source $E(t)$. These are calculable by usual methods.

(2) The currents due to the electromotive force $V(t)$ inserted at time $t=0$, between the points $a$ and $b$. These are also calculable by usual methods, since $V(t)$ is itself known from the primary distribution of currents and charges.

By the preceding analysis we have succeeded, therefore, in reducing the problem of a sudden short circuit, to the determination of the currents in an *invariable* network in response to a suddenly impressed electromotive force: that is, the problem to which the preceding chapters have been devoted.

*The Sudden Open Circuit*

The problem of a sudden open circuit in any part of a network can be dealt with in a precisely analogous manner, although the actual calculation of the resultant current and voltage distribution is mathematically more complicated. Consider the network shown in Fig. 32.

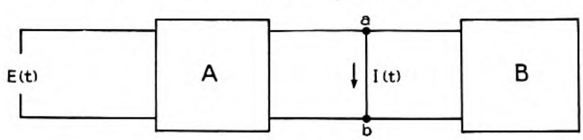



Fig. 32

Here the network is supposed to be energized by an electromotive force $E(t)$ which produced a current $I(t)$ in the *invariable* network in branch $ab$. We require the effect of suddenly opening this branch. The solution of this problem depends on the following proposition.

*The effect of opening branch ab at reference time $t=0$ is the same as suddenly inserting at time $t=0$, a voltage $V(t)$ which produces in branch ab a current $-I(t)$ equal and opposite to the current which would exist in the branch in the absence of the open circuit.*

While this proposition is precisely analogous to the corresponding proposition in the case of a sudden short circuit, it does not *explicitly* determine the voltage $V(t)$, which must be calculated as follows:

Let the driving point indicial admittance of the network, as seen from branch *ab* be denoted by $A_{ab}(t)$. Then, from the preceding proposition, it follows at once that $V(t)$ is given by

$$\frac{d}{dt}\int_0^t V(\tau)A_{ab}(t-\tau)d\tau = -I(t),\ t \geq 0.$$

This is a Poisson integral equation in $V(t)$, from which $V(t)$ is calculable. With $V(t)$ determined, the currents in any part of the network are calculable by usual methods, and consist of two components:—

(1) The current distribution in the network due to the impressed source $E(t)$ *in the absence of the open circuit.*

(2) The current distribution due to the electromotive force $V(t)$ inserted in branch *ab* at time $t=0$.

As in the case of the sudden short circuit, we have thus reduced the problem of a sudden open circuit to the determination of the current distribution in an *invariable* network in response to a suddenly impressed electromotive force.

*Variable Circuit Elements*

In the preceding cases of sudden open and short circuits it will be observed that the network changes discontinuously from one invariable state to another. A more general case, and one which includes the preceding as limiting cases, is presented by a network which includes a variable circuit element: that is, a circuit element which varies, continuously or discontinuously, with time. A network which includes such a variable circuit element will be called a *variable network*. Variable circuit elements of practical importance are the microphone transmitter, which consists of a variable resistance, varied by some source of energy outside the system; the condenser transmitter, which consists of a condenser of variable capacity; and the induction generator, in which the mutual inductance between primary and secondary, or stator and rotor, is varied by the motion of the latter. The case of a variable resistance will serve as an introduction to the general theory of such variable networks.

Consider a network, energized by a source $E(t)$ in branch 1, and containing a variable resistance element $r(t)$ in branch $n$. The functional notation $r(t)$ indicates that the resistance $r$ varies with time. Let $I_n(t)$ denote the current in branch $n$, and assume that the network is in equilibrium prior to the reference time $t=0$. The mathematical theory of this network depends on the following proposition:—

*The network described above can be treated as an invariable network by eliminating the variable resistance element $r(t)$ and inserting an electromotive force $-r(t)I_n(t)$: that is, an electromotive force equal and opposite to the potential drop across the variable resistance element.* Consequently the current in the variable resistance branch is determined analytically by the integral equation

$$I_n(t)=\frac{d}{dt}\int_0^t E(\tau)A_{1n}(t-\tau)d\tau-\frac{d}{dt}\int_0^t r(\tau)I_n(\tau)A_{nn}(t-\tau)d\tau. \qquad (288)$$

The first component is simply the current $I_o(t)$ which would exist in the variable branch if the variable element were absent; hence, dropping the subscript $n$ for convenience, the current in the variable branch is given by the integral equation

$$I(t)=I_o(t)-\frac{d}{dt}\int_0^t r(\tau)I(\tau)A(t-\tau)d\tau \qquad (289)$$

and the voltage across the variable element by

$$v(t)=r(t)I(t). \qquad (290)$$

Having determined $I(t)$ and $v(t)$ from this integral equation, the distribution of currents in the network is calculable as that due to a source $E(t)$ in branch 1 and a source $v(t)$ in branch $n$ of the *invariable network*: that is, the network with the variable resistance element eliminated.

A very simple example will serve to illustrate the foregoing:—

Into a circuit of unit resistance, and inductance $L=1/a$, in which a steady current $I_o$ is flowing, a resistance $r$ is suddenly inserted at time $t=0$: required the resultant current $I(t)$. In this case we have:

$$A(t)=\text{indicial admittance of unvaried circuit}$$
$$=1-e^{-at},$$
$$r(t)=r,$$

and the integral equation of the problem is:

$$I(t)=I_o-r\frac{d}{dt}\int_0^t (1-e^{-ay})I(t-y)dy$$
$$=I_o-ra\int_0^t I(t-y)e^{-ay}\,dy.$$

If the solution is carried out as indicated by (291) below, and if the notation $at=x$ is introduced, we get without difficulty

$$I(t)=I_o\{1-r(1-e_1(x)e^{-x})+r^2(1-e_2(x)e^{-x})-r^3(1-e_3(x)e^{-x})+\ldots\}.$$

where the function $e_n(x)$ is defined as:

$$e_n(x)=1+x/1!+x^2/2!+x^3/3!+\ \ldots\ +x^{n-1}/(n-1)!$$

$$=\text{first } n \text{ terms of the exponential series.}$$

For all finite values of the resistance increment $r$ the series can be summed by aid of the identity

$$1-e_n(x)e^{-x}=\int_0^x dxe^{-x}x^{n-1}/(n-1)!$$

Substitution of this identity gives

$$I(t)=I_o\left(1-r\int_0^x e^{-(1+r)x}dx\right)$$

$$=I_o\frac{1+re^{-(1+r)x}}{1+r}.$$

Equation (289) is an integral equation of the Volterra type, which includes the Poisson integral equation as a special case. Its formal series solution is obtained as follows:—Assume a series solution of the form

$$I_n(t)=I_o(t)-I_1(t)+I_2(t)-I_3(t)+\ \ldots \tag{291}$$

and define the terms of the series by the scheme

$$I_1(t)=\frac{d}{dt}\int_0^t r(\tau)I_o(\tau)A(t-\tau)d\tau$$

$$\text{- - - - - - - - - - - - - - - - - - - - -} \tag{292}$$

$$I_{k+1}(t)=\frac{d}{dt}\int_0^t r(\tau)I_k(\tau)A(t-\tau)d\tau.$$

Direct substitution shows that this series satisfies the integral equation. Furthermore, it is easily shown that it is absolutely convergent.

While this series solution is not, in general, well adapted for numerical calculations, it throws a good deal of valuable light on the ultimate character of the oscillations in the important case where $E(t)$ and $r(t)$ both vary sinusoidally with time. In this case, if the frequency of the applied e.m.f. be denoted by $F$ and that of the resistance variation by $f$, it is easy to show that the current $I_o(t)$ in the unvaried circuit is ultimately [10] a steady state current of frequency $F$. This follows from the fact that the definite integral which defines the current $I_o(t)$ is resolvable into the ultimate steady state current corresponding to an applied force of frequency $F$, and the accompanying transient oscillations which ultimately die away. The fictitious e.m.f. which may be regarded as producing the component current $I_1(t)$ is $rf(t)I_o(t)$; this is ultimately the product of the two frequencies $F$ and $f$, and therefore resolvable into two terms of frequency $F+f$ and $F-f$ respectively. Carrying through this analysis, it is easy to show that each component current is ultimately a steady-state but poly-periodic oscillation, as indicated in the following table:

| Component Current | Frequency |
|---|---|
| $I_0$ . . . . . . . . . . . . . . . . . | $F$, |
| $I_1$ . . . . . . . . . . . . . . . . . | $F+f, F-f$, |
| $I_2$ . . . . . . . . . . . . . . . . . | $F+2f, F, F-2f$, |
| $I_3$ . . . . . . . . . . . . . . . . . | $F+3f, F+f, F-f, F-3f$, |
| $I_4$ . . . . . . . . . . . . . . . . . | $F+4f, F+2f, F, F-2f, F-4f$. |

It is of importance to observe that the component currents involve, from a mathematical standpoint, multiple integrals of successively higher orders, the $n$th component $I_n(t)$ involving a multiple integral of the $n^{\text{th}}$ order with respect to $I_0(t)$. Consequently the successive currents require longer and longer intervals of time to build up to their proximate steady-state values, so that the time required for the resultant steady-state to be reached cannot be inferred from the time constant of the unvaried circuit.

From the preceding table it will be seen that the ultimate steady-state current is obtained by rearranging the series $I_o+I_1+I_2$ and is of the form

$$\sum_{n=-\infty}^{+\infty} A_n \cos(\Omega+n\omega)t+B_n \sin(\Omega+n\omega)t$$

where $\Omega=2\pi F$ and $\omega=2\pi f$.

It is interesting to note that this series comes within the definition of a Fourier series only when $F=0$ or an exact multiple of $f$. The steady-state solution is of very considerable importance and is considered in more detail in a succeeding chapter.

From the foregoing we deduce an outstanding distinction between the variable and invariable networks. In the latter the currents are

[10] It hardly seems necessary to remark that the reference time $t=o$ is purely arbitrary and that the resistance variation may start at such a time thereafter that $I_o(t)$ may be regarded as steady state during the entire time interval in which we are interested. Going farther, if we confine our attention to sufficiently large values of $t$, the whole process may be treated as steady state.

ultimately of the same frequency as the impressed e.m.f., whereas in the former they are ultimately of an infinite series of frequencies.

In the preceding example, the variable impedance element is a resistance $r(t)$. If the variable element is taken as an *inductance* $\lambda(t)$ the voltage, corresponding to equation (290) is

$$\frac{d}{dt}\lambda(t)I(t).$$

The case of a variable capacity element is handled as follows: Let $1/C=S$ and assume that $S$ is variable: thus, $S=S_o+\sigma(t)$. The drop across the variable condenser element is then

$$v(t)=\sigma(t)\int_0^t I(t)dt.$$

Similarly a variable mutual inductance $\mu(t)$ between branches $m$ and $n$ produces the voltages

$$\frac{d}{dt}\mu(t)I_n(t)$$

in branch $m$, and

$$\frac{d}{dt}\mu(t)I_m(t)$$

in branch $n$. This case may be illustrated by:

*The Induction Generator Problem*

In a sufficiently general form, this problem, which includes the fundamental theory of the dynamo, may be stated as follows:

Given an invariable primary and secondary circuit with a variable mutual inductance $Mf(t)$ which is an arbitrary but specified time function, and let the primary be energized by an e.m.f. $E(t)$ impressed in the circuit at the reference time $t=0$: required the primary and secondary currents.

In operational notation the problem may be formulated by the equations:

$$Z_{11}I_1-pMf(t)I_2=E(t),$$
$$-pMf(t)I_1+Z_{22}I_2=0,$$

in which $Z_{11}$ and $Z_{22}$ are the self impedances of the primary and secondary respectively; $Mf(t)$ is the variable mutual inductance; $E(t)$ is the applied e.m.f. in the primary; and $p$ denotes the differential operator $d/dt$. By aid of the fundamental formula these equations may be written down as the following simultaneous integral equations:

$$I_1(t) = \frac{d}{dt}\int_0^t dy A_{11}(t-y)\left(E(y) + M\frac{d}{dy}[f(y)I_2(y)]\right)$$

$$I_2(t) = M\frac{d}{dt}\int_0^t dy A_{22}(t-y)\frac{d}{dy}[f(y)I_1(y)].$$

In these equations, $A_{11}(t)$ and $A_{22}(t)$ denote the indicial admittances of the primary and secondary circuits respectively (when $M=0$): that is, the currents in these circuits in response to a unit e.m.f. (zero before, unity after time $t=0$). We assume, of course, that they are known or can be determined by usual methods.

It follows at once that the formal solution of these equations is the infinite series:

$$I_1(t) = X_o(t) + X_2(t) + X_4(t) + \ldots + X_{2n}(t) + \ldots$$

$$I_2(t) = Y_1(t) + Y_3(t) + Y_5(t) + \ldots$$

in which the successive terms of the series are defined as follows:

$$X_o(t) = \frac{d}{dt}\int_0^t dy A_{11}(t-y)E(y) = I_o(t),$$

$$Y_1(t) = M\frac{d}{dt}\int_0^t dy A_{22}(t-y)\frac{d}{dy}[f(y)X_o(y)],$$

$$X_2(t) = M\frac{d}{dt}\int_0^t dy A_{11}(t-y)\frac{d}{dy}[f(y)Y_1(y)],$$

$$Y_3(t) = M\frac{d}{dt}\int_0^t dy A_{22}(t-y)\frac{d}{dy}[f(y)X_2(y)], \quad \text{etc.}$$

In the light of formula

$$I(t) = \frac{d}{dt}\int_0^t f(y)A(t-y)dy$$

the physical interpretation of the series solutions follows at once: Thus, $X_o(t)$ is equal to the current $I_o(t)$ flowing in the *isolated* primary in response to the applied e.m.f. $E(t)$; the first component current $Y_1(t)$ in the secondary is equal to the current which would flow in the isolated secondary in response to the applied e.m.f. $M(d/dt)f(t)X_o(t)$; $X_2(t)$, the second component current in the primary, is equal to the current in the isolated primary in response to the applied e.m.f. $M(d/dt)$

$f(t)Y_1(t)$; etc. The resultant currents are thus represented as built up by a to-and-fro interchange of energy between primary and secondary, or by a series of successive reactions. In the important case where the applied e.m.f. and the variation of mutual inductance are both sinusoidal time functions, of frequency $F$ and $f$ respectively, it is easy to show that each component current becomes ultimately equal to a set of periodic steady-state currents. Thus the component $X_o$ is ultimately single periodic, of frequency $F$; $Y_1$ is ultimately doubly periodic, of frequencies $F+f$ and $F-f$; $X_2$ triply periodic, of frequencies $F+2f$, $F$ and $F-2f$; $Y_3$ quadruply periodic, of frequencies $F+3f$, $F+f, F-f, F-3f$; etc.

*The Solution for the Steady-State Oscillations*

For the very important case of periodic applied forces and periodic variations of circuit elements we are often concerned exclusively with the ultimate steady-state of the system, and not at all with the mode in which the steady-state is approached: that is, attention is restricted to the periodic oscillations which the system executes after transient disturbances have died away. In this case, if the periodic variations of circuit elements are sufficiently small, the required steady-state is obtained in the form of a series by replacing each term of the complete series solution by its ultimate steady-state value; a process which is very simple in view of the physical significance of each term of the latter series. The appropriate procedure will be briefly illustrated in connection with the variable resistance element. In view of the fact that we are concerned only with the ultimate steady-state oscillations, we can base the solutions on the symbolic equation

$$I = I_o - \frac{r(t)}{Z} I. \tag{293}$$

Here $r(t)$ is the variable resistance element; $I_o$ is the current which would flow in the absence of the resistance variation; and $Z$ is a generalized impedance of the network, as seen from the variable branch. Its precise significance and functional form is given below.

We now suppose that $I_o$ is given by

$$I_o = J_o e^{i\Omega t} \qquad \text{(real part)} \tag{294}$$

$$= \tfrac{1}{2}\left(J_o e^{i\Omega t} + \bar{J}_o e^{-i\Omega t}\right) \tag{295}$$

where the bar indicates the conjugate imaginary of the unbarred symbol, so that (295) is entirely real. Correspondingly the variable resistance will be taken as

$$r(t) = \frac{r}{2}(e^{i\omega t} + e^{-i\omega t})$$
$$= re^{i\omega t} \qquad \text{(real part)} \tag{296}$$
$$= r \cos \omega t.$$

Here $r$ is taken as a pure real quantity, which fixes the size of the resistance variation. No loss of generality is involved in this, since it merely involves referring the time scale to the zero of the resistance variation.

The symbolic impedance $Z$, as employed in the theory of alternating currents, will depend on the frequency and is, in general, a complex quantity. Its value at frequency $\Omega/2\pi$ will be denoted by

$$Z(i\Omega) = Z_o$$

while its value at frequency $(\Omega + n\omega)/2\pi$ will be written as

$$Z(i(\Omega + n\omega)) = Z_n.$$

We now assume a series solution of (293) of the form

$$I = I_o + I_1 + I_2 + \ldots$$

where the terms of the series are defined by the symbolic equations

$$I_1 = -\frac{r(t)}{Z} I_o,$$
$$\text{-----------} \tag{297}$$
$$I_{n+1} = -\frac{r(t)}{Z} I_n.$$

Substitution shows that this series formally satisfies the equation.

Starting with the first of (297) and substituting (295) and (296) we get

$$I_1 = -\frac{r}{4Z}(e^{i\omega t} + e^{-i\omega t})(J_o e^{i\Omega t} + \bar{J}_o e^{-i\Omega t})$$

$$= -\frac{r}{4Z}\left\{ J_o e^{i(\Omega+\omega)t} + \bar{J}_o e^{-i(\Omega+\omega)t} + J_o e^{i(\Omega-\omega)t} + \bar{J}_o e^{-i(\Omega-\omega)t} \right\}, \tag{298}$$

or

$$I_1 = -\frac{r}{2} J_o \left\{ \frac{e^{i(\Omega+\omega)t}}{Z_1} + \frac{e^{i(\Omega-\omega)t}}{Z_{-1}} \right\}. \tag{299}$$

In (299) it is to be understood that the real part is alone to be retained.

Proceeding in a similar way with the equation

$$I_2=-\frac{r(t)}{Z}I_1$$

we get

$$I_2=\left(\frac{r}{2}\right)^2 J_o\left\{\frac{e^{i(\Omega+2\omega)t}}{Z_1Z_2}+\frac{e^{i(\Omega-2\omega)t}}{Z_{-1}Z_{-2}}+\frac{e^{i\Omega t}}{Z_o}\left(\frac{1}{Z_1}+\frac{1}{Z_{-1}}\right)\right\}. \tag{300}$$

In this way the steady-state series solution is built up term by term, the component currents being poly-periodic as indicated in a previous table.

For sufficiently small impedance variations this method of solution works very well, and leads to a rapidly convergent solution. In other cases, however, the solution so obtained may be divergent, even when the complete series solution from which it is derived is absolutely convergent. The explanation of this lies in the fact that the steady-state series so obtained is the *sum of the limits* (as $t$ approaches infinity) of the terms of the complete series solution, whereas the actual steady-state is the *limit of the sum.* These are not in general equal; in particular the former may be and often is divergent when the latter is convergent.

In view of the foregoing considerations it is of importance to develop another method of investigating the steady-state oscillations which avoids the difficulties in the formal series solution. The following method has suggested itself to the writer and works very well in cases where the previous form of solution fails. It should be stated at the outset, however, that the absolute convergence of the solution to be discussed, while reasonably certain in all physically possible systems, has not been established by a rigorous mathematical investigation, which appears to present very considerably difficulties.

We start with the problem just discussed and, in view of the results of the formal series solution there obtained, assume a solution of the form:

$$I=\frac{1}{2}\sum_{-N}^{N}A_m e^{i(\Omega+m\omega)t}+\bar{A}_m e^{-i(\Omega+m\omega)t} \tag{301}$$

$$=\sum_{-N}^{N}A_m e^{\,i(\Omega+m\omega)t} \quad \text{(real part).} \tag{302}$$

Here the series is supposed to extend from $m=+N$ to $m=-N$. Ultimately, however, $N$ will be put equal to infinity. As before, the bar indicates the conjugate imaginary of the unbarred symbol and (301) is therefore entirely real.

If we now substitute (301) in the symbolic equation (293) we get, by (295) and (296),

$$\frac{1}{2}\sum\left\{A_m e^{i(\Omega+m\omega)t}+\bar{A}_m e^{-i(\Omega+m\omega)t}\right\}=\tfrac{1}{2}J_o e^{i\Omega t}+\tfrac{1}{2}\bar{J}_o e^{-i\Omega t}$$

$$-\frac{r}{2Z}\left(e^{i\omega t}+e^{-i\omega t}\right)\sum\left\{A_m e^{i(\Omega+m\omega)t}+\bar{A}_m e^{-i(\Omega+m\omega)t}\right\}.$$

Simplifying this equation and dropping the conjugate imaginaries gives:—

$$\sum A_m e^{i(\Omega+m\omega)t}=J_o e^{i\Omega t}-\frac{r}{Z}\sum A_m e^{i(\Omega+(m+1)\omega)t} \tag{303}$$

$$-\frac{r}{Z}\sum A_m e^{i(\Omega+(m-1)\omega)t}.$$

Finally, if we write

$$Z(i(\Omega+m\omega))=Z_m$$

and

$$r/Z_m=h_m \tag{304}$$

and equate terms of the same frequency on the two sides of the equation, we get

$$\left.\begin{aligned}A_N&=-h_N A_{N-1}\\ A_m&=-h_m(A_{m-1}+A_{m+1}) \qquad 0<|m|<N\\ A_o&=J_o-h_o(A_{-1}+A_1).\end{aligned}\right\} \tag{305}$$

It will be observed that, by (305), starting with $A_N$ each coefficient is determined in terms of the coefficient of the next lower index. Thus:

$$A_N=-h_N A_{N-1}$$

$$A_{N-1}=-h_{N-1}(A_{N-2}+A_N)$$

$$=-\frac{h_{N-1}A_{N-2}}{1-h_{N-1}h_N}.$$

Similarly

$$A_{N-2}=-\frac{h_{N-2}A_{N-3}}{1-h_{N-2}h_{N-1}\dfrac{1}{1-h_{N-1}h_N}}.$$

Continuing this process it is easy to show that, for positive indices ($m$ positive),

$$A_m = -h_m C_m A_{m-1} \tag{306}$$

where $C_m$ designates the continued fraction

$$\frac{1}{1-h_m h_{m+1}}\;\frac{1}{1-h_{m+1}h_{m+2}}\;\cdots\;\frac{1}{1-h_{N-1}h_N}$$

The procedure for the coefficient $A_{-m}$ is precisely similar. For convenience we write $A_{-m}=A'_m$, $Z_{-m}=Z'_m$, and $r/Z_{-m}=h'_m$. In this notation we get by precisely similar procedure

$$A'_m = -h'_m C'_m A'_{m-1} \tag{307}$$

where $C'_m$ designates the continued fraction

$$\frac{1}{1-h'_m h'_{m+1}}\;\frac{1}{1-h'_{m+1}h'_{m+2}}\;\cdots\;\frac{1}{1-h'_{N-1}h'_N}$$

We now put the index $N$ equal to infinity and the continued fractions $C_m$ and $C'_m$ become infinite instead of terminating fractions.

Collecting formulas we now have

$$A_m = -h_m C_m A_{m-1}$$

$$A'_m = -h'_m C'_m A'_{m-1}$$

and

$$A_o = J_o - (h_o A_1 + h'_o A'_1)$$

whence

$$A_o = \frac{J_o}{1-h_o h_1 C_1 - h_o' h_1' C_1'}.$$

The coefficients are thus all determined in terms of $J_o$.

The practical value of this method of solution will depend, of course, on the rate of convergence of the continued fractions. While no rigorous proof has been obtained, it is believed that they are absolutely convergent for all physically possible systems, but this question certainly requires fuller investigation. Nevertheless any doubt regarding the convergence of the solution need not prevent the use of the method in a great many problems where physical considerations furnish a safe guide. For example this method of solution, when applied to the problem of the induction generator, discussed above, leads to the usual simplified engineering theory of the induction generator and motor, besides exhibiting effects which the usual treatment either ignores or fails to recognize.

*Non-Linear Circuits*

In the previous examples discussed, the variations of the variable circuit elements are assumed to be specified time functions, which is the same thing as postulating that these variations are controlled by ignored forces which do not explicitly appear in the statement and equations of the problem. We distinguish another type of variable circuit element, where the variation is not an explicit time function but rather a function of the current (and its derivatives) which is flowing through the circuit. For example, the inductance of an iron-core coil varies with the current strength as a consequence of magnetic saturation. The equation of a circuit which contains such a variable element (provided it is a single valued function) may be written down in operational notation

$$ZI+\phi(I)=E(t),$$

or

$$ZI=E(t)-\phi[I(t)]. \tag{311}$$

In this equation $Z$ is, of course, to be taken as the impedance of the invariable part of the circuit, the indicial admittance of which is denoted by the usual symbol $A(t)$.

Equation (311) may be interpreted as the equation of the current $I(t)$ in a circuit of invariable impedance $Z$ when subjected to an applied e.m.f. $E(t)-\phi[I(t)]$; consequently, by aid of our fundamental formula, $I(t)$ is given by

$$I(t)=\frac{d}{dt}\int_0^t A(t-y)E(y)dy-\frac{d}{dt}\int_0^t A(t-y)\phi[I(y)])dy.$$

The first integral is simply the current in the invariable circuit of impedance $Z$ in response to the applied e.m.f. $E(t)$; denoting this by $I(t)$, we have

$$I(t)=I_0(t)-\frac{d}{dt}\int_0^t A(t-y)\phi[I(y)]dy.$$

This is a *functional integral equation*, the solution of which is gotten by some process of successive approximations. For example, provided the sequence converges, $I(t)$ is the limit as $n$ approaches infinity of the *sequence*

$$I_o(t), I_1(t), I_2(t), \ldots, I_n(t),$$

where the successive terms of the sequence are defined by the relations:

$$I_1(t) = I_o(t) - \frac{d}{dt}\int_0^t A(t-y)\phi[I_o(y)]dy,$$

- - - - - - - - - - - - - - - - - - - - - - - - - -

$$I_{n+1}(t) = I_o(t) - \frac{d}{dt}\int_0^t A(t-y)\phi[I_n(y)]dy.$$

We shall not pursue the discussion of non-linear circuits further, in view of their mathematical complexity and their relatively specialized technical interest. The reader who is interested may, however, consult the writer's paper on Variable Electrical Systems,[11] for a fuller treatment of the subject.

## CHAPTER XI

### The Application of the Fourier Integral to Electric Circuit Theory

The application of Fourier's series in electrotechnics is a commonplace; the use of the Fourier integral, however, has largely remained in the hands of professional mathematicians. An outstanding distinction between the series and the integral, from which the greater power of the latter may be inferred, is that the series represents only a periodic regularly recurrent function, whereas the integral is capable of representing a non-periodic function: in fact all types of functions, subject to certain mathematical restrictions which are usually satisfied in physical problems.

Before taking up the application of the Fourier Integral to Electric Circuit Theory, we shall very briefly review the elementary mathematics of the series and integral; for a fuller treatment the reader is referred to Byerly, Fourier's Series and Spherical Harmonics.[12]

Consider a function $\phi(t)$, which in the region $0 \leq t \leq T$ is finite, single-

[11] Phys. Rev. Feb., 1921.

[12] In this chapter the Fourier Integral is approached from the view-point of its physical application and no completeness or rigour is claimed for the treatment. The mathematical theory of the Fourier integral is, of course, completely developed in treatises on the subject. The object of this chapter is merely to outline some of its applications.

valued and has only a finite number of discontinuities or of maxima or minima. In this region it can then be expressed as the Fourier series

$$\phi(t)=\tfrac{1}{2}\,A_o+\sum_{1}^{\infty}\left\{A_n\cos\left(\frac{2\pi n}{T}t\right)+B_n\sin\left(\frac{2\pi n}{T}t\right)\right\} \tag{312}$$

where

$$A_n=\frac{2}{T}\int_0^T\phi(t)\cdot\cos\left(\frac{2\pi n}{T}t\right)dt,$$
$$B_n=\frac{2}{T}\int_0^T\phi(t)\cdot\sin\left(\frac{2\pi n}{T}t\right)dt. \tag{313}$$

An equivalent series is

$$\phi(t)=\tfrac{1}{2}\,F_o+\sum_{1}^{\infty}F_n\cos\left(\frac{2\pi n}{T}t-\Theta_n\right) \tag{314}$$

where

$$F_n=\sqrt{A_n^2+B_n^2},$$
$$\Theta_n=\tan^{-1}(B_n/A_n). \tag{315}$$

This expansion is valid in the region $0\leq t\leq T$, irrespective of the form of the function elsewhere. Let us, however, assume that the function repeats itself in the period $T$: that is

$$\phi(t\pm kT)=\phi(t),\qquad k=1,\,2,\,3\,\ldots\,N.$$

Then the expansion represents the function in the region $-NT\leq t\leq NT$. Finally if $N$ is made infinite, the function is truly periodic and the Fourier series represents it for all positive and negative values of time.

It follows from the foregoing that, if the Fourier series represents the function for all positive and negative values of time, the function must be periodic for all positive and negative values of time; otherwise the expansion is valid only over a restricted range of time.

Now let us suppose that $\phi(t)$ is non-periodic. For convenience, in connection with subsequent applications we shall suppose that it is zero for all finite *negative* values of time, that it converges to zero as $t\rightarrow\infty$, and that

$$\int_0^\infty\phi(t)\;dt$$

exists. Such a function obviously cannot be represented by the usual Fourier series for all finite positive and negative values of time; it can be represented, however, by the limiting form assumed by the series as the fundamental period $T$ is made infinite. That is, we can assume that the function is periodic in an infinite fundamental period and this will not affect the expansion for finite positive and negative values of time. Proceeding in this way and putting the fundamental period $T$ equal to infinity in the limit, the Fourier series (314) becomes an infinite integral and we get

$$\phi(t)=\frac{1}{\pi}\int_0^{\infty} F(\omega)\cdot\cos\left(\omega t-\theta(\omega)\right)d\omega \tag{316}$$

where

$$F(\omega)=\left\{\left[\int_0^{\infty}\phi(t)\cos\omega t\,dt\right]^2+\left[\int_0^{\infty}\phi(t)\sin\omega t\,dt\right]^2\right\}^{\frac{1}{2}} \tag{317}$$

and

$$\tan\theta(\omega)=\int_0^{\infty}\phi(t)\sin\omega t\,dt\div\int_0^{\infty}\phi(t)\cos\omega t\,dt. \tag{318}$$

This is the *Fourier integral* identity of the function $\phi(t)$ and is valid for all finite positive and negative values of time.

In physical applications, particularly those to electric circuit theory, it is often convenient to employ exponential instead of trigonometric functions. The required transformation follows easily from the relation

$$e^{i\theta}=\cos\theta+i\sin\theta, \qquad i=\sqrt{-1}.$$

Thus if we write $2\pi/T=\omega_o$ the Fourier series (312) is easily reduced to the form

$$\phi(t)=\sum_{-\infty}^{+\infty}F(in\omega_o)e^{in\omega_o t} \tag{319}$$

where

$$F(in\omega_o)=\frac{1}{T}\int_0^{T}\phi(\tau)\,e^{-in\omega_o\tau}\,d\tau. \tag{320}$$

In precisely similar manner the Fourier integral (316) can be written as

$$\phi(t)=\int_{-\infty}^{\infty}F(i\omega)\cdot e^{i\omega t}d\omega \tag{321}$$

$$=\frac{1}{2\pi}\int_0^{\infty}\phi(\tau)d\tau\int_{-\infty}^{\infty}e^{i\omega(t-\tau)}d\omega. \tag{322}$$

*Applications to Electric Circuit Theory*

Let us assume that at time $t=-NT$, an electromotive force $E(t)$, periodic in fundamental period $T$, is impressed on a circuit of complex impedance $Z(i\omega)$, where $\omega$ denotes $2\pi$ times the frequency. Required the resultant current $I$.

For values of $t>-NT$ the electromotive force (see formula (319)) can be expressed as the Fourier series

$$E(t)=\sum_{-\infty}^{\infty} F(in\omega_o)e^{in\omega_o t}$$

where

$$F(in\omega_o)=\frac{1}{T}\int_0^T E(\tau)e^{-in\omega_o\tau}d\tau.$$

The resultant current for $t>-NT$ is therefore

$$I=\sum_{-\infty}^{\infty}\frac{F(in\omega_o)}{Z(in\omega_o)}e^{in\omega_o t}+\left\{\begin{array}{l}\text{transient oscillations}\\ \text{initiated at time}\\ t=-NT.\end{array}\right\}.$$

If we are concerned with the current for values of $t\geq 0$, and if $NT$ is made sufficiently large, the initial transients will have died away and the *complete current* for $t\geq 0$, will be given by

$$I=\sum_{-\infty}^{\infty}\frac{F(in\omega_o)}{Z(in\omega_o)}e^{in\omega_o t}. \tag{323}$$

This formula implies the periodic character of $E(t)$ for sufficiently large negative values of time. If, however, $E(t)$ is zero for negative values of time, we can employ the Fourier integrals (321) and (322) in precisely the same way and get, as the *complete* expression for the current for positive or negative values of time:—

$$I=\int_{-\infty}^{\infty}\frac{F(i\omega)}{Z(i\omega)}e^{i\omega t}\,d\omega \tag{324}$$

$$=\frac{1}{2\pi}\int_0^{\infty}E(\tau)d\tau\int_{-\infty}^{\infty}\frac{e^{i\omega(t-\tau)}}{Z(i\omega)}\,d\omega. \tag{325}$$

*The infinite integrals* (324) *and* (325) *formulate the current in the network, specified by the impedance function* $Z(i\omega)$, *in response to an electromotive force* $E(t)$ *impressed at time* $t=0$; *they therefore mathematically formulate, by aid of the Fourier integral identity, the fundamental problem dealt with in the preceding chapters and solved by aid of the operational calculus.*

No attempt will be made here to discuss the solution of the infinite integral (325), which is usually a problem presenting formidable difficulties, even to the professional mathematician. The general method of solution is by contour integration in the complex plane and the calculus of residues. By this process it has been successfully applied to the solution of special problems, and also to deriving some general forms of solution such as the expansion theorem solution.[13] Compared, however, with the operational calculus, it has no advantages from the standpoint of rigour, and lacks entirely the remarkable simplicity and directness of the Heaviside method.

In the direct solution of circuit problems, therefore, it is believed that the application of the Fourier integral is attended by few if any advantages, and presents formidable mathematical difficulties. On the other hand, there are certain types of problems encountered in circuit theory, where the Fourier integral is a powerful tool. These will be briefly discussed.

### *The Energy Absorbed from Transient Applied Forces*

In many technical problems, the complete solution for the instantaneous current due to suddenly applied electromotive forces, although formally straight-forward, involves a prohibitive amount of labor. In yet others, the applied forces may be random and specified only by their mean square values. In such problems a great deal of useful information is furnished by the mean power and mean square current absorbed by the network, and to the calculation of these quantities, the Fourier integral is ideally adapted. Its application depends on the following proposition, due to Rayleigh (Phil. Mag., Vol. 27, 1889, p. 466), and its corollary.

Let a function $\phi(t)$, supposed to exist only in the epoch $0 \leq t \leq T$, be formulated as the Fourier integral

$$\phi(t) = \frac{1}{\pi}\int_0^\infty |f(\omega)| \cdot \cos\,[\omega t - \theta(\omega)]\,d\omega$$

where

$$f(\omega) = \left\{\left[\int_0^T \phi(t)\cos\omega t\,dt\right]^2 + \left[\int_0^T \phi(t)\sin\omega t\,dt\right]^2\right\}^{\frac{1}{2}}$$

$$\tan\theta(\omega) = \int_0^T \phi(t)\sin\omega t\,dt \div \int_0^T \phi(t)\cos\omega t\,dt.$$

[13] Bush, "Summary of Wagner's Proof of Heaviside's Formula." Proc. Inst. of Radio Engineers. Oct., 1917. Fry. "The Solution of Circuit Problems." (Phys. Rev. Aug., 1919).

Then

$$\int_0^T [\phi(t)]^2\,dt = \frac{1}{\pi}\int_0^\infty |f(\omega)|^2\,d\omega,$$

whereby the time integral is transformed into an integral with respect to frequency.

A corollary of this theorem is as follows:

If two functions $\phi_1(t)$, $\phi_2(t)$ supposed to exist only in the epoch $0 \le t \le T$, are formulated by the Fourier integrals

$$\phi_1(t) = \frac{1}{\pi}\int_0^\infty |f_1(\omega)| \cdot \cos[\omega t - \theta_1(\omega)]\,d\omega,$$

$$\phi_2(t) = \frac{1}{\pi}\int_0^\infty |f_2(\omega)| \cdot \cos[\omega t - \theta_2(\omega)]\,d\omega,$$

then

$$\int_0^T \phi_1(t)\phi_2(t)dt = \frac{1}{\pi}\int_0^\infty |f_1(\omega)| \,.\, |f_2(\omega)| \,.\, \cos(\theta_1 - \theta_2)d\omega.$$

The applications of these theorems to circuit theory proceeds as follows:—

If an electromotive force $E(t)$, supposed to exist only in the epoch $0 \le t \le T$, is applied to a network of complex impedance $Z(i\omega) = |Z(i\omega)|e^{i\beta(\omega)}$ we know from the preceding discussion of the Fourier integral, that the electromotive force $E(t)$ and current $I(t)$ are expressible as the Fourier integrals

$$\begin{aligned} E(t) &= \frac{1}{\pi}\int_0^\infty |f(\omega)| \,.\, \cos(\omega t - \theta(\omega))d\omega, \\ I(t) &= \frac{1}{\pi}\int_0^\infty \frac{|f(\omega)|}{|Z(i\omega)|}\cos(\omega t - \theta(\omega) - \beta(\omega))d\omega. \end{aligned} \tag{326}$$

It follows at once from Rayleigh's theorem that

$$\int_0^\infty I^2 dt = \frac{1}{\pi}\int_0^\infty \frac{|f(\omega)|^2}{|Z(i\omega)|^2}d\omega. \tag{327}$$

Now let $I_n$ be the current absorbed in branch $n$; let $z(i\omega) = |z(i\omega)|e^{i\alpha(\omega)}$ be the impedance of that branch and let $E_n(t)$ be the potential drop across that branch. It follows at once from the corollary to Rayleigh's theorem that

$$W = \int_0^\infty E_n(t)I_n(t)dt = \frac{1}{\pi}\int_0^\infty \frac{|f(\omega)|^2}{|Z(i\omega)|^2}|z(i\omega)|\cos\alpha(\omega)d\omega. \tag{328}$$

Formulas (327) and (328) formulate the mean square current and mean power absorbed by the branch of the network under consideration, and enable us to calculate these quantities, even in the case of complicated networks, with a minimum of labor. Formula (327) is particularly well adapted to computation because the integrand is everywhere positive, permitting, in most problems, of easy numerical integration, whereas the analytical solution may be complicated.

Formulas (327) and (328) have been applied to the theory of selective circuits, to the problem of interference from random disturbances, including static, and to the theory of the Schrotteffekt. For the details of such applications, which will not be entered into here, the reader is referred to the following papers.

Transient Oscillations in Electric Wave Filters, Bell System Technical Journal, July, 1923.

Selective Circuits and Static Interference, Trans. A. I. E. E., 1924.

An Application of the Periodogram to Wireless (Burch & Bloehmsma), Phil. Mag., Feb., 1925.

The Theory of the Schrotteffekt (Fry), Journal Franklin Institute, Feb., 1925.

### *The Building-Up of Alternating Currents*

Another application of the Fourier Integral, which may be briefly mentioned, is to the building-up of alternating currents in response to suddenly impressed sinusoidal electromotive forces. The investigation of this problem is of great importance to the communication engineer, since the excellence of a signal transmission system is to a considerable extent determined by the duration and character of the building-up phenomena.

In long transmission systems the calculation of the building-up current as a time function is extremely complicated and laborious if not practically impossible. Furthermore we are usually not concerned with the current as an instantaneous time function, but rather with its *envelope*. The envelope of the current can be formulated and calculated by modified Fourier integrals, by the following process.

Suppose that an e.m.f. $E \cos \omega t$ is suddenly applied, at reference time $t=0$, to a network of transfer impedance

$$Z(i\omega)=|Z(i\omega)|\, e^{iB(\omega)}.$$

The resultant current $I(t)$ may always be written as:

$$I(t) = \frac{1}{2}\frac{E}{|Z(i\omega)|}\left\{ (1+\rho)\cos(\omega t - B) + \sigma \sin(\omega t - B) \right\}$$

$$= \frac{1}{2}\sqrt{(1+\rho)^2+\sigma^2}\,\frac{E}{|Z(i\omega)|}\cos(\omega t - B(\omega) - \theta)$$

where

$$\theta = \tan^{-1}\frac{\sigma}{1+\rho}$$

Evidently the functions $\rho$ and $\sigma$, which it is our problem to determine, must be $-1$ and $0$ respectively for negative values of $t$, and approach the limits $+1$ and $0$, respectively, as $t \to \infty$.

In an engineering study of the building-up process we are principally concerned with the *envelope* of the oscillations: hence with

$$\frac{1}{2}\sqrt{(1+\rho)^2+\sigma^2}.$$

Our problem is therefore to determine the functions $\rho$ and $\sigma$ and to examine the effect of the applied frequency $\omega/2\pi$ and of the characteristics of the circuit, on their rate of building-up and mode of approach to their ultimate steady values.

The functions $\rho$ and $\sigma$ can be formulated as the Fourier integrals

$$\rho = \frac{1}{\pi}\int_0^\infty [P_\omega(\lambda) + P_\omega(-\lambda)]\sin t\lambda .\,\frac{d\lambda}{\lambda}$$

$$-\frac{1}{\pi}\int_0^\infty [Q_\omega(\lambda) - Q_\omega(-\lambda)]\cos t\lambda\,\frac{d\lambda}{\lambda}$$

$$\sigma = \frac{1}{\pi}\int_0^\infty [Q_\omega(\lambda) + Q_\omega(-\lambda)]\sin t\lambda\,\frac{d\lambda}{\lambda}$$

$$+\frac{1}{\pi}\int_0^\infty [P_\omega(\lambda) - P_\omega(-\lambda)]\cos t\lambda\,\frac{d\lambda}{\lambda},$$

where

$$P_\omega(\lambda) = \frac{|Z(i\omega)|}{|Z(i\omega)+i\lambda)|}\cdot\cos[B(\omega+\lambda) - B(\omega)],$$

$$Q_\omega(\lambda) = \frac{|Z(i\omega)|}{Z(i\omega+i\lambda)|}\cdot\sin[B(\omega+\lambda) - B(\omega)].$$

These formulas are directly deducible from the fact that the applied e.m.f., defined as zero for $t<0$ and $E\cos\omega t$ for $t\geq 0$, can itself be expressed as

$$\frac{E}{2}\cos\omega t\left[1+\frac{2}{\pi}\int_0^\infty \sin t\lambda\frac{d\lambda}{\lambda}\right].$$

For important types of transmission systems, including the periodically loaded line, these formulas have been successfully dealt with and solutions of a satisfactory approximate character obtained. For further details, the reader is referred to a paper on "The Building-Up of Sinusoidal Currents in Long Periodically Loaded Lines" (Bell System Technical Journal, October, 1924).

The foregoing must conclude our very brief account of the Fourier Integral and its applications in Electric Circuit theory; an adequate treatment of this subject would require a treatise in itself, and is beyond the scope of the present work. All that has been attempted is to give a very brief introduction to its significance in physical problems and a few of its outstanding applications in circuit theory. The reader who is interested in pursuing this subject further is referred to a paper by T. C. Fry on "The Solution of Circuit Problems" (Phys. Rev., Aug., 1919), which gives a rigorous discussion of the solution of the Fourier Integral by contour integration, together with some general forms of solution of the circuit problem.[1]

[1] It was planned to include in this paper a bibliography of the important papers bearing on the Heaviside operational method. This, however, has not been completed, but plans call for its publication in the next issue.—Editor.

# Abstracts of Recent Technical Papers from Bell System Sources

*Cipher Printing Telegraph Systems for Secret Wire and Radio Telegraphic Communications.* G. S. VERNAM.[1] This paper describes a printing telegraph cipher system developed during the World War for the use of the Signal Corps, U. S. Army. This system is so designed that the messages are in secret form from the time they leave the sender until they are deciphered automatically at the office of the addressee. If copied while en route, the messages cannot be deciphered by an enemy, even though he has full knowledge of the methods and apparatus used. The operation of the equipment is described, as well as the method of using it for sending messages by wire, mail or radio.

The paper also discusses the practical impossibility of preventing the copying of messages, as by wire tapping, and the relative advantages of various codes and ciphers as regards speed, accuracy and the secrecy of their messages.

*Methods of High Quality Recording and Reproducing of Music and Speech Based on Telephone Research.* J. P. MAXFIELD and H. C. HARRISON.[2] The paper deals with an analysis of the general requirements of recording and reproducing sound, with the nature of the inherent limitations where mechanical records are used, and a detailed description of a solution involving, first, the use of electrical equipment for the purposes of recording and, second, the use of mechanical equipment based on electric transmission methods for reproducing.

Probably the most useful feature of the paper is the complete description of the application of electrical transmission theory to mechanical transmission systems. A detailed analysis is made of the analogies between the electrical and the mechanical systems.

*Electrical and Photo-Electric Properties of Thin Films of Rubidium on Glass.* HERBERT E. IVES and A. L. JOHNSRUD.[3] Films which spontaneously deposit on glass surfaces in a highly exhausted cell containing rubidium are electrically conducting, and photo-electrically active. A study of the photo-electric properties of a rubidium coated

[1] *A I. E. E. Journal*, Vol. 45, pp. 109–115, Feb., 1926.
[2] *A. I. E. E. Journal*, Vol. 45, pp. 243–253, Mar., 1926.
[3] *Astrophysical Journal*, Vol. 52, pp. 309–319, Dec., 1925.

plane glass surface shows the normal and selective effects less well differentiated than for the similar coatings which form on metal plates. A rubidium film formed on the inside of a glass cylinder is found to exhibit, in the dark, a pure ohmic resistance. This decreases under illumination in a manner which appears to be explained as due to the liberation of photo-electrons which under a potential gradient form an added current along the tube.

*The Influence of Temperature on the Photo-Electric Effect of the Alkali Metals.*[4] HERBERT E. IVES and A. L. JOHNSRUD. Special cells having a hollow central cathode were immersed in liquid air for an extended period to insure that any gases, if present, were condensed on the outer alkali metal coated walls. The temperature of the cathode was controlled by a stream of evaporating liquid air, whereby all temperatures between +20 and −180° C. could be attained and held constant and be measured. In these cells the variation of photoelectric current with temperature in sodium, potassium, and rubidium is continuous, without abrupt changes. The effect is relatively small for sodium, showing hardly at all for blue light or white light, but clearly for yellow light. The behavior of rubidium is similar to that previously reported for potassium.

In a second form of cell, potassium was collected in a deep pool. By slowly cooling the metal from the molten condition, smooth crystalline surfaces were obtained. With these annealed potassium surfaces, the variation of photoelectric current with temperature is represented by curves varying systematically in shape with the color of the light, and the effect is far greater than previously reported, amounting, for yellow light, to a variation of 10 to 15 times between room and liquid air temperature. When the surface is roughened curves of the previously reported type are obtained. Small pools give erratic effects, showing changes in opposite directions for different portions of the temperature range. It is concluded that the variation of photoelectric effect is intimately connected with the strains produced in the surface by expansion and contraction with temperature.

*Positive Rays in Thermionic Vacuum Tubes.*.[5] HERBERT E. IVES. Thermionic tubes in which a quantity of alkali metal is present exhibit not only the normal electron current from the heated filament, but a positive current, which at low filament temperatures may be

[4] *Journal of the Optical Society of America & Review of Scientific Instruments*, Vol. 11, No. 6, pp. 565–579, Dec., 1925.

[5] *Journal of the Franklin Institute*, Vol. 201, pp. 47–69, Jan., 1926.

many times larger than the negative current. The electron current is in general reduced by the positive rays, but at higher filament temperatures the reduction of space charge by the positive causes a considerable increase of the current over a limited voltage range. By immersing the tube in liquid air the positive ray effects are almost eliminated, indicating that the alkali metal vapor is the source of the rays, which are probably produced by contact of metal atoms with the hot filament.

*A New Directional Receiving System.*[6] H. T. FRIIS. Reduction of static interference, or to state it more correctly, reduction of the ratio of static to signal, has been, almost since the beginning of the radio art, the most important problem in radio engineering. It is now well known that static disturbances have definite points of origin and that the impulses which are detected at a receiving station have definite directions of propagation. A receiving system having no directional selectivity is, therefore, affected by static impulses from all directions and, in spite of many inventions, it has not yet been possible to improve its signal-static ratio except by limiting the frequency band transmitted. A system which, however, is so designed as freely to receive waves arriving from a limited range of directions is susceptible only to static disturbances propagated within that range, and large improvements in signal static ratio have been claimed for different types of directive antenna systems during the past few years.

A directional receiving system for radio telephony in which directional selectivity is obtained by combining the output voltages from two antennas is described in this paper. The main feature of the system is the arrangement for controlling the output voltages of the antennas, so that they may be combined to neutralize each other or to reinforce each other as desired. A double detection (super-heterodyne) receiver is employed and the output voltages, which are combined so as to produce the directional characteristic, are the intermediate frequency currents due to the waves received by the antennas and the beating oscillator currents. The control of these output voltages is effected by operating upon the beating oscillator currents.

*High-Power Metallography—Some Recent*[7] *Developments in Photomicrography and Metallurgical Research.* FRANCIS F. LUCAS. The

[6] Proceedings of the Institute of Radio Engineers, Vol. 13, No. 6, pp. 685–707, Dec., 1925.

[7] *Journal of the Franklin Institute*, Vol. 201, No. 2, pp. 177–216, Feb., 1926.

usual conception of high-power metallography seems to be great enlargement, indistinct definition and lack of resolution. Such results, generally, have been classed under the heading "empty magnification" because they have failed to show more detail than has been shown at lower magnifications and with objectives of less resolving ability. Oftentimes the pictures would be unintelligible taken by themselves, but the reason they are recognized at all is because the same structures have been seen and identified by low- or medium-power methods. Such high-power results are like an elastic band which has been stretched unduly. As the band is stretched it becomes more and more attenuated and finally snaps. If the optical image is stretched by enlargement the details of the image become less and less distinct and finally the image breaks down altogether, so that the detail and the background blend together into a hazy outline of what formerly was a sharp image.

High-power metallography as presented in this article consists of so preparing metallurgical specimens that crisp, brilliant images may be obtained and photographed at high powers and of achieving approximately the potential resolving possibilities of splendid objectives.

By improvements in the method of preparing metallurgical specimens and in the technique of manipulating the apparatus, "empty magnification" is no longer synonymous with high-power photomicrography.

It is the object of this contribution to show the application of this new tool for metallurgical research to the study of metal structures which heretofore have not been resolved and the nature of which has led to much speculation and to wide differences of opinion. A clear understanding of the current conceptions of magnification and resolution is essential and a knowledge of the limitations which were regarded for many years as restricting the employment of high powers will prove of value in the interpretation of the results obtained. For this reason a brief discussion follows which not only shows the method of approach in the present development, but indicates the path along which we may work to secure a higher order of resolution. By resolution is meant that property of a lens system which enables it to distinguish or "resolve" as separate and distinct units fine structural details spaced very close together.

*Research and Engineering.*[8] E. B. CRAFT. Research in industry—which the author mentions is of comparatively recent origin—is defined as the application of methods of systematic and logical deduc-

[1] Address before the Engineers' Club, Phila., Oct., 1925. *Engs. and Engg.*, Jan. 1926, Vol. 43, pp. 11–19.

tions to our every-day industrial and technical problems. Such research necessarily is of a highly specialized nature and requires special training. What is equally important, as is pointed out by the author, is the need of properly organizing and directing this group of specialized workers. Since research is a creative process and hence particularly individualistic, one of the important problems in what the author calls "organized research" is the supplying of such an atmosphere that the worker realizes his own welfare and advancement to be adequately cared for in this system of group working. A number of examples of organized research are mentioned (radio and wire telephony, telephotography, ocean telegraphy, speech and hearing, artificial speech, phonograph recording and reproducing) as apropos to the point in question. The close relationship between engineering and research and the impossibility of the one getting along without the other is made clear. For the worker, there is pointed out the necessity of management and for those in charge the soundness of industrial research as a business proposition. Industrial research far from being a luxury has become a necessity.

## Contributors to this Issue

Thomas Shaw, S.B., Massachusetts Institute of Technology, 1905; American Telephone and Telegraph Company, Engineering Department, 1905–1919; Department of Development and Research, 1919—. Mr. Shaw's major activities have been devoted to development problems in loading telephone circuits, including the loading apparatus.

W. Fondiller, B.S., College of the City of New York, 1903; E.E., Columbia University, 1909; M.A., 1913; Engineering Department, Western Electric Company, 1909–25; Bell Telephone Laboratories, 1925—. Mr. Fondiller is in charge of the General Development Laboratory and has been engaged in the design of loading coils and electric filters, and directing studies of the properties of materials and analysis of manual and machine switching apparatus design.

H. R. Friis, E.E., Royal Technical College in Copenhagen, 1916; Columbia University, 1919–20; Research Department, Western Electric Company, 1920–24; Bell Telephone Laboratories, 1925—. Mr. Friis' work has been largely in connection with radio reception methods and measurements. He has published papers on vacuum tubes as generators, radio transmission measurements and static interference.

Ronald M. Foster, S.B., Harvard, 1917; American Telephone and Telegraph Company, Engineering Department, 1917–19; Department of Development and Research, 1919—.

Walter A. Shewhart, A.B., University of Illinois, 1913; A.M., 1914; Ph.D., University of California, 1917; Engineering Department, Western Electric Company, 1918–24; Bell Telephone Laboratories, Inc., 1925—. Mr. Shewhart has been engaged in the study of the relationship between the microphonic and physicochemical properties of carbon.

Herbert E. Ives, B.S., University of Pennsylvania, 1905; Ph.D., Johns Hopkins, 1908; assistant and assistant physicist, Bureau of Standards, 1908–09; physicist, Nela Research Laboratory, Cleveland, 1909–12; physicist, United Gas Improvement Company, Philadelphia, 1912–18; U. S. Army Air Service, 1918–19; research engineer, Western Electric Company, 1919–24; Bell Telephone Laboratories, 1925—. Dr. Ives' work has had to do principally with the production, measurement and utilization of light.

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's work has been along theoretical lines and he has published many papers on theory of electric circuits and electric wave propagation.